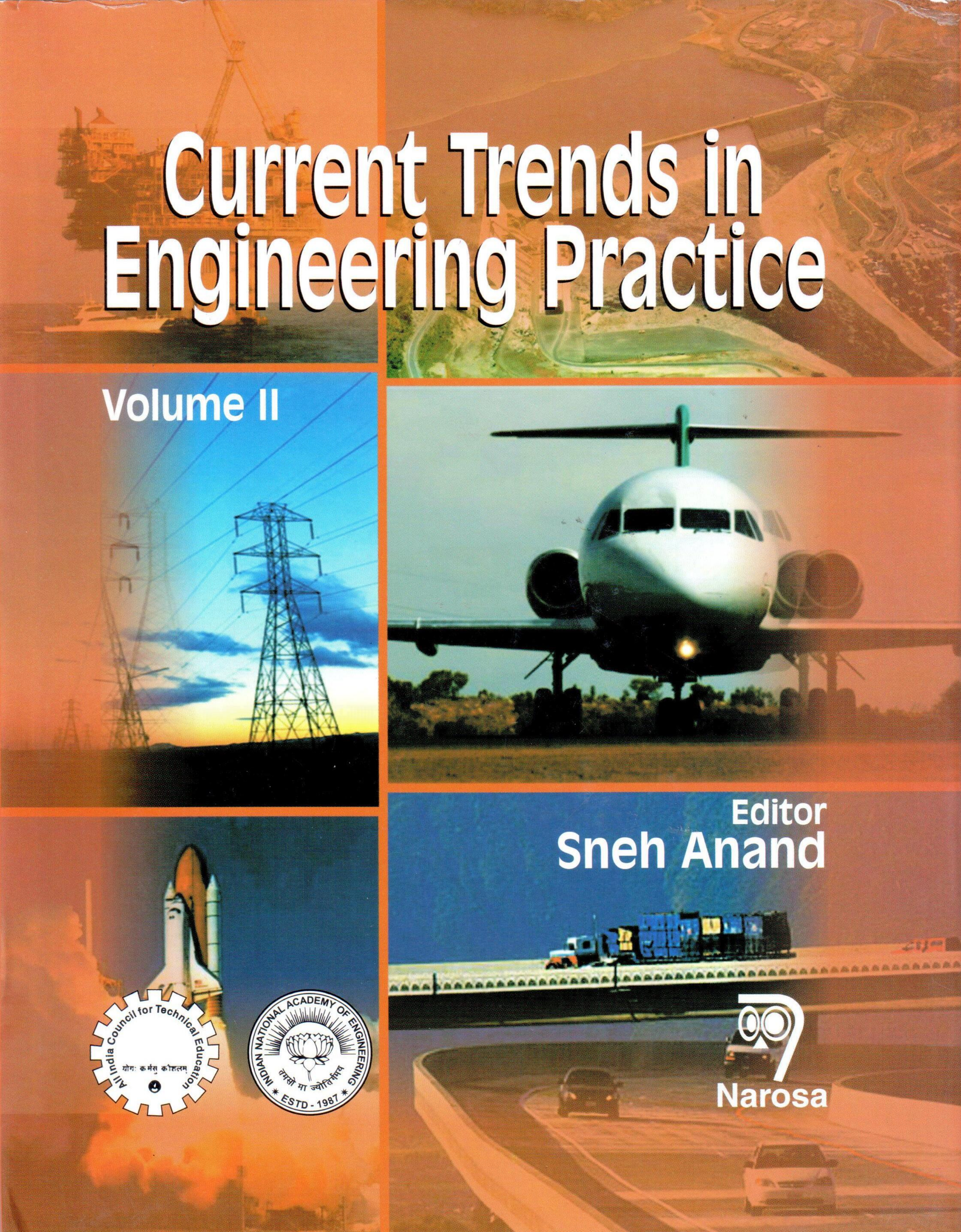
Current Trends in Engineering Practice

Volume II



Editor
Sneh Anand

Narosa

Current Trends in Engineering Practice

Volume II

Current Trends in Engineering Practice

Volume II

Editor

Sneh Anand

Narosa Publishing House

New Delhi Chennai Mumbai Kolkata

Current Trends in Engineering Practice Volume II
732 pgs. | 384 figs. | 75 tbls.

Editor
Sneh Anand
Professor
Centre for Biomedical Engineering
Indian Institute of Technology Delhi
New Delhi

NAROSA PUBLISHING HOUSE PVT. LTD.

22 Delhi Medical Association Road, Daryaganj, New Delhi 110 002
35-36 Greams Road, Thousand Lights, Chennai 600 006
306 Shiv Centre, Sector 17, Vashi, Navi Mumbai 400 703
2F-2G Shivam Chambers, 53 Syed Amir Ali Avenue, Kolkata 700 019

www.narosa.com

ISBN 978-81-8487-019-0

Published by N.K. Mehra for Narosa Publishing House,
22 Delhi Medical Association Road, Daryaganj, New Delhi 110 002

Printed in India

FOREWORD

The Indian National Academy of Engineering (INAE) launched a Distinguished Visiting Professorship (DVP) Scheme jointly with AICTE in 1999. The Scheme envisages promotion of industry-institute interaction by facilitating the dissemination of knowledge through the expertise of experienced and knowledgeable persons from industry to integrate their rich industrial experience with technical education. The Scheme has received very enthusiastic response from industry and engineering research institutions over the years. It has helped in strengthening the symbiotic relationship between industry and academic institutions and has built bridges between the industry experts and students and faculty of engineering colleges.

The first volume of "Current Trends in Engineering Practice" a compilation of papers based on the lectures delivered by industry experts in engineering colleges under the AICTE-INAE Distinguished Visiting Professorship scheme was brought out in the year 2006. This highlighted the expertise of the visiting professors and was well received by persons from academia and industry.

The second volume of this series is being brought out. Needless to say this has been possible due to the efforts made by the industry experts who have devoted their valuable time in delivering lectures, guiding student projects, participating in revision of curricula and keeping the academia abreast of the state-of-the-art industrial practices and techniques. It is hoped that this book will be well received in the academic and industrial realms.

Baldev Raj
Vice-President, INAE
Chairman, Steering Committee on
AICTE-INAE Distinguished Visiting
Professorship Scheme

PREFACE

Under the AICTE-INAE Distinguished Visiting Professorship scheme, industry experts are encouraged to give series of lectures at an educational institution in their proximity for a specific time period. This scheme has become popular among industry experts as well as educational institutions. Thirteen Industry Experts were selected during the year 2000; eighteen each during the years 2001 and 2002; fourteen during the year 2003; ten during the year 2004; thirteen during the year 2005; fourteen during the year 2006; fifteen during the year 2007; eleven during the year 2008 and eighteen during the year 2009, by a high level selection committee of experts from Academia, Industry and representatives from All India Council for Technical Education (AICTE) and Confederation of Indian Industry (CII).

The scheme has been a great success and has been running effectively during the last ten years. During their visits, the visiting professors besides delivering lectures on the state-of-art industrial practices and sharing their industrial experience with the faculty/ students of the affiliated engineering institutions are also assisting in updating the curriculum. They also have made significant contributions in guiding and evaluating UG and PG projects, and formulating certain useful collaboration projects between the concerned industry and institution. This scheme no doubt helps in bridging the gap between Industry and Academia and serves as a model for beneficial industry-institute interaction.

The first volume of "Current Trends in Engineering Practice" which was brought out in the year 2006 was based on lectures delivered by the industry experts and dealt with certain engineering practices adopted in various projects in different engineering disciplines and specializations. We are happy that the second volume of this series is being brought out.

I wish to express my sincere thanks to all the industry experts who have devoted their precious time in contributing to this book. I also grateful to all the Distinguished Visiting Professors and Heads of Departments & Programme Coordinators of the concerned engineering institutions who have made this scheme an astounding success.

Sneh Anand
Project Coordinator
AICTE-INAE Distinguished
Visiting Professorship Scheme

FROM THE PRESIDENT'S DESK

The first volume of "Current Trends in Engineering Practice" a compilation of papers based on the lectures delivered by industry experts in engineering colleges under the AICTE-INAE Distinguished Visiting Professorship scheme was brought out in the year 2006.

I am happy to note that the second volume of this compilation is being brought out. I am sure that this volume will be well received by the engineering institutions and industry and that the efforts will continue to promote industry-institute interaction.

P. S. Goel
President, INAE

CONTENTS

NUCLEAR POWER AND ENERGY TECHNOLOGIES

INTERDISCIPLINARY ENGINEERING AND TECHNOLOGY

CIVIL ENGINEERING

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

HIGH PERFORMANCE CONCRETE

PRABIR C. BASU*

1. INTRODUCTION

Concrete is a composite material consisting of binding medium within which aggregates are embedded. Mixture of hydraulic cement and water forms the binding medium in case of hydraulic cement concrete [1, 2]. Strength and durability of concrete are governed by its microstructure formation. Concrete can be considered as a non-homogenous composite material with different phases. It is a three phase composite, first two phases are aggregates and bulk hydrated cement paste (hcp) and the third one is the transition zone [2, 3]. Fig. 1 depicts the diagrammatic representation of microstructure of concrete composite [2].

Tracing back the history of development of HPC, it could be seen that work in this field was initially limited to high strength concrete. Compressive strength has been considered as the most important parameter to classify concrete since its inception. However, it has been realized at a later date with experience that the strength is not the only important parameter other attributes such as durability, workability of concrete are also important for successful engineering of a concrete structure. This led to the evolution of the concept of HPC.

What is HPC? American Concrete Institutes defines HPC as, Concrete, which meets special performance and uniformity requirements that can not always be achieved routinely by using only conventional materials, and normal mixing, placing and curing practices. The requirements may involve enhancement of characteristics, such as placement and compaction without segregation, long term mechanical properties, early age strength, toughness, volume stability, or service life in severe environment [4]. The attributes, which generally need to be directly taken care during engineering of the HPC mix, are broadly of three categories: strength, durability and rheology.

There is no unique specification for HPC. The performance requirements or properties for which HPC is to be engineered are specified from its intended use. The intended use of concrete is again derived from the functional roles of buildings/structures for which it is engineered. Specifications of HPC for different uses are different. For example, in the case of high ways, Strategic Highway Research Programme of USA has specified HPC depending upon the requirements of highways [5]. This specification is different than that of HPC for nuclear power plant [6]. Apart from the intended

*Civil & Structural Engineering Division, Atomic Energy Regulatory Board, Mumbai

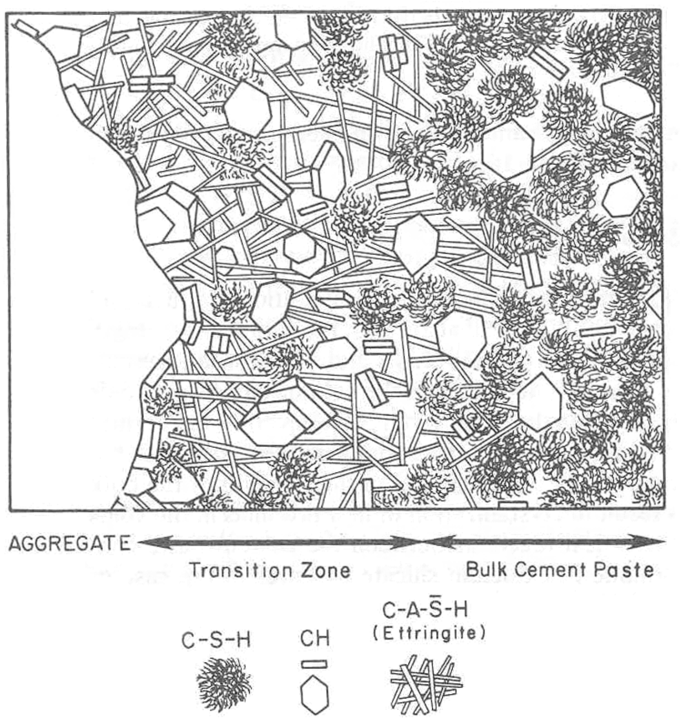


FIG. 1 DIAGRAMMATIC REPRESENTATION OF THE TRANSITION ZONE AND BULK CEMENT PASTE IN CONCRETE (REPRODUCED FROM REF. 2)

use, other aspects such as design features (shape of structural elements, congestion, etc.), construction methods also have bearing in outlining the specification of HPC.

Recent edition of IS 456 [7] formally accepted the use of HPC in the construction of concrete structures in India. Silica fume [4,8,9,10], fly ash [11,12,13,14], ground granulated blast furnace slag [15,16] are commonly used mineral admixture for HPC. High reactive metakaolin [17,18] and rice husk ash [19] are also reported to be suitable for HPC. IS 456 allows the use of the reactive mineral admixtures, silica fume (SF), pulverized fly ash (FA), high reactive metakaolin (HRM), rice husk ash (RHA). Though there exist Indian specifications on SF and PFA, hardly any guidelines on specification for characterizations for HRM or RHA are available in India.

Probably the construction of the primary containment dome of atomic power project at Kaiga, Unit-1 (Kaiga-1) is the first concrete structure in India constructed with engineered HPC [20]. The primary containment dome of Rajasthan Atomic Power Project, Units-3&4 (RAPP-3&4) was constructed with HPC [21]. Complete primary containment structure of Tarapur Atomic Power Project Units-3&4 (TAPP-3&4) [7] had also been constructed with HPC. Recently JJ fly over at Mumbai has also been constructed using HPC [22]. Bridges of Bandra-Worli sea link will be constructed with HPC [23]. At present many structures are being constructed with HPC.

High performance concrete engineered with SF, HRM, FA and GGBS is the subject matter of the present article. Mix proportioning of HPC and other construction aspects are discussed. Comparative study on the behaviour of HPC developed using these mineral admixtures are also presented.

2. MECHANISM OF HPC

How HPC is different from the conventional normal strength concrete (NSC)? According to Neville, “High performance concrete is concrete selected so as to be fit for the purpose for which it is required. There is no mystery about it, no unusual ingredients are needed, and no special equipment has to be used. All we use is an understanding of the behaviour of the concrete, and will to produce a concrete mix within closely controlled tolerances” [24]. Therefore, for optimum engineering of concrete structures using HPC, it is necessary to dwell upon broad aspects related to the mechanism of HPC.

The phenomena that the properties of materials can be tailored by suitable engineering of its structure have been widely applied in developing a number of materials, composites, etc. [25]. Like any other composites, key to developing the composite of HPC is engineering the microstructure of concrete suitably. A detailed discussion on mechanism of HPC is presented in reference 26.

Concrete has extremely complex microstructure consisting of solid phases, pores and water with a high degree of heterogeneity [2]. The solid phases; aggregates, bulk “hcp” and transition zone. Transition zone is the interface between aggregates and bulk “hcp”; which itself is “hcp” and exists as a thin shell around the aggregates. Transition zone is a mixture of hydrated product, un-reacted particles of cementitious material, pores and water. Micro cracks are present in this phase. Transition zone is the weakest. It is considered as strength limiting phase and has also significant influence on other attributes of concrete such as durability.

Interfaces in the microstructure of concrete are not only those between aggregates and “hcp” (i.e. transition zone) but also may exist otherwise, for example, between anhydrous particles of cementitious materials and bulk “hcp”. The enhancement properties such as high strength, low permeability of HPC are achieved principally by improving the microstructure of the interfaces. The mechanism, which leads to the improvement of microstructure of concrete, especially at the interfaces, has three basic components [26],

- Reaction mechanism among the ingredients
- Physical process
- Curing

Reaction mechanism is based upon chemical reaction among the ingredients and related physical phenomena; principally the hydration of cementitious materials. Ordinary Portland Cement (OPC) consists essentially of tricalcium and dicalcium silicates, C_3S and C_2S. Other main compounds are interstitial phases like tricalcium aluminates (C_3A), tetracalcium aluminoferrites (C_3AF) and small amount of gypsum (calcium sulphate). OPC also contains minor compounds and impurities like alkali, sulphates, free lime, unreacted silica and magnesia. Principal product of cement hydration process, i.e. the reaction between water and C_2S and C_3S, is C-S-H gel, which is the main contributor of the strength of hydration product,

$$2C_3S + 6H = C\text{-}S\text{-}H + 3CH$$

$$2C_2S + 4H = C\text{-}S\text{-}H+CH$$

Reactive silica (SiO_2) of mineral admixtures reacts with the CH generated from primary hydration in presence of water and further produces C-S-H when pH of the hydration system is 13:

$$S + CH + H = C\text{-}S\text{-}H$$

This chemical reaction is known as secondary hydration or pozzolanic reaction. The pozzolanic action results in grain refinement of the transition zone. Mineral admixtures also act as filler materials. By virtue of filler action they fill the voids between the cement particles resulting in well-packed concrete mix. However, for maximum reduction of voids, appropriate size of aggregates and suitable grading of solid ingredients starting from coarse aggregates to the finest one is necessary.

Physical process consists of mixing of ingredients for production of concrete mix, transportation, placements etc [27]. This process results in creating conducive condition for the reaction mechanism to take place appropriately for imparting the desired attributes of the concrete mix both in fresh and hardened state. Curing maintains the satisfactory condition so that the reaction mechanism is completed to the desirable state [28]. Therefore, reaction mechanism is the basic component of mechanism for developing HPC, while physical process and curing are supportive components.

3. ENGINEERING OF STRUCTURES USING HPC

Improvement of microstructure is achieved by using both chemical admixture (superplasticiser) and reactive mineral admixtures (pozzolans) along with low water cementitious material (w/c_m) ratio. Modification of the microstructure of transition zones requires suitable mix proportion with desired characteristics of ingredients and other interfaces depends on suitable aggregate size, appropriate grading of solid materials starting from coarse aggregate to the finest one, optimum content of cementitious materials and water along with chemical admixture are important for it. This is achieved by appropriate engineering activities

Successful approach for engineering of structure, when HPC is used, should basically aim at providing all possible measures so that reaction mechanism, i.e. reaction between the ingredients takes place in in-situ condition without hindrance. This could be achieved by developing concrete mix with ingredients of desirable characteristics, adopting suitable physical process, which includes suitable process for production of concrete (mixing of ingredients) and placement (transportation, pouring, compaction, etc.), and providing necessary curing regime. Adequate curing regime is necessary for maintaining the suitable ambiance so that reaction mechanism takes place appropriately. Systematic approach is necessary for successful construction of structures using HPC is achieved implementing these concepts. Important activities of construction of structure using HPC are,

- Mix proportioning
- Mixing method for production of concrete
- Transportation

- Placement
- Curing

Another important aspect is quality assurance program. HPC is an engineered material and very sensitive to the quality of the above activities. Well laid down procedure should be followed in each step and provisions should be made to check the quality of work in the steps. Characterization of ingredients, adopting relevant Indian codes and codes of other countries when Indian codes are not available for a particular ingredients, is very important in this context [26, 29,30].

4. MIX PROPORTIONING

Creation of conducive condition of hydration of cementitious materials by minimizing flocculation of fine particles, modification and reduction of interfaces in the structure of concrete, and filling up of pores are important considerations in developing the mix proportion of HPC. It is true that chemical and mineral admixtures augment the chemical reactions to change the structure of concrete. However, a quantity of any one of these ingredients in the mix more than required, may result in undesirable interfaces that could be detrimental to the properties of concrete. Optimization of the quantity of ingredients, especially cement, mineral admixture and water, helps in this regard.

When water is mixed with cementitious material (cement plus active mineral admixtures), fine particles of these materials tend to flocculate [31]. This would result in the risk of a portion of flocculated cement particles getting hydrated leaving behind a substantial portion of unhydrated and high degree of voids resulting unnecessary interfaces. Adopting two pronged measures in mix proportioning minimizes this risk. Firstly, water content is optimised and secondly by use of high range water reducing admixtures, superplasticiser, in the concrete mix. Superplasticiser creates conducive condition for complete hydration of cement by deflocculating the lumps of cementitious materials and making the mixtures a well dispersed system. This improves the pore structure during hydration process by bringing almost all particles of cementitious materials fully in contact with water.

Concrete mix proportioning method described by Aitcin is reported to be suitable for HPC [30]. No specific method of mix proportioning method for NSC, presently prevailing in India, was found suitable for the mix design of HPC mix [32]. Main difficulty lies in non-applicability of water – cement ratio. No guideline is specified for determination of suitable quantity of superplasticiser.

Following considerations are useful in proportioning of HPC mix. Air content for concrete may be assumed to be 1.0%. Unit water content may not be less than about 150 kg/m^3. However, this depends on the initial slump considered in the mix design. It is always prudent to taken in to account a target slump at time of trial mix. Target compressive strength of the mix may be taken as per IS 456, but this needs not to be mandatory. The mix should satisfy the codal requirements of acceptance criteria.

4.1 Mix Proportioning of HPC based on Absolute Volume Method

The approach based on absolute volume method was adopted in proportioning of the HPC mix for containment domes of Kaiga-1&2, RAPP-3&4 and TAPP-3&4. This is discussed briefly to illustrate the approach for proportioning the quantities of cementitious materials, superplasticiser and w/cm ratio for HPC mix with silica fume. The unit water content was judged to achieve the target slump as well as requisite strength of hardened concrete. The quantity of ordinary Portland cement (OPC) and silica fume first and then w/c_m were by trials. Superplasticiser dose was fixed through compatibility study and the quantities of coarse and fine aggregates are determined adopting absolute volume method.

Table 1. *Performance Requirements of KAIGA-1&2, RAPP-3&4 and TAPP-3&4 Containment Structure*

Parameters	*Kiaga-1&2*	*RAPP-3&4*	*Tapp-3&4*
Comp. Strength	60 MPa	60 MPa	60 MPa
Split tensile strength	3.87 MPa	3.87 MPa	3.48 MPa
Permeability (Water)	5.0mm	5.0mm	5.0mm
Slump	175±25mmm	175±25mm	175±25mm; 100 mm after 1hr.

Table 1 contains the performance requirements of HPC for these nuclear power plants. Amongst the mixes, HPC mix proportion for Kaiga was developed first. Example of Kaiga HPC is cited in the discussion for better understanding. All ingredients were characterized following well-established specification.

4.1.1 Cementitious materials

Determination of optimum quantity of cement and mineral admixture is extremely important in designing efficient HPC mix. Quantity of cementitious material in excess of required quantity may leave behind unhydrated particles that may result in undesirable interfaces. In Kaiga, trials were carried out to determine required quantity of silica fume along with cement with respect to strength taking the value of slump = 175 ± 25 mm as the reference parameter. It was observed that for silica fume quantity of about 7.5 percent by weight of cement, the mix attains maximum strength and beyond 10 percent the strength is reduced, Fig. 2. Quantity of cement was taken as 475 kg/m^3 and that of silica fume as 7.5 percent by weight of cement. Similar quantities of cement and silica fume were taken for HPC of RAPP-3&4 inner containment domes and TAPP-3&4 containment rafts. However, after gaining sufficient experience, the quantity of cement was further reduced from 475 to 425 kg/m^3 in the HPC mix of containment wall of TAPP-3&4.

4.1.2 Superpalsticiser

The optimum dosage of superplasticiser can be evaluated by a flow consistency test using Marsh Cone. The point at which the slope changes correspond to the optimum dosage of the admixture. The result of flow consistency test carried out during the development of Kaiga HPC is plotted in Fig. 3. The superplasticiser (ASTM F-type) admixture dosage of about 2 percent by weight of cement was found to be optimum for Kaiga HPC. Dosage of retarder is determined by trial mix, optimum dosage was selected as 0.1 percent by weight of cement.

Determination of saturation dosage of superplasticiser is not easy when the compatibility between cement and plasticiser is not good. Nkinamubanzi, P.C. et al. [29] summarized the

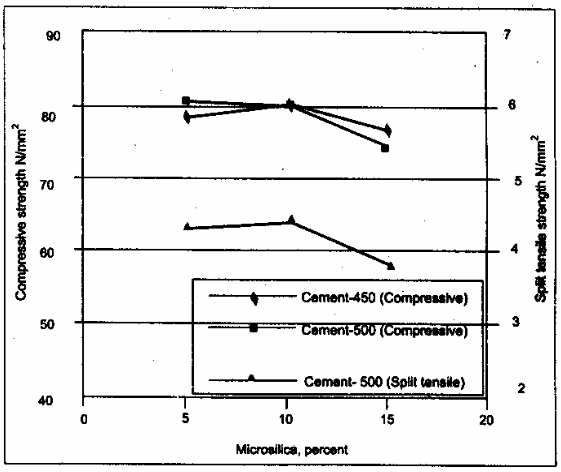


FIG. 2 PLOT OF STRENGTH VERSUS SILICA FUME CONTENT (KAIGA 1&2)

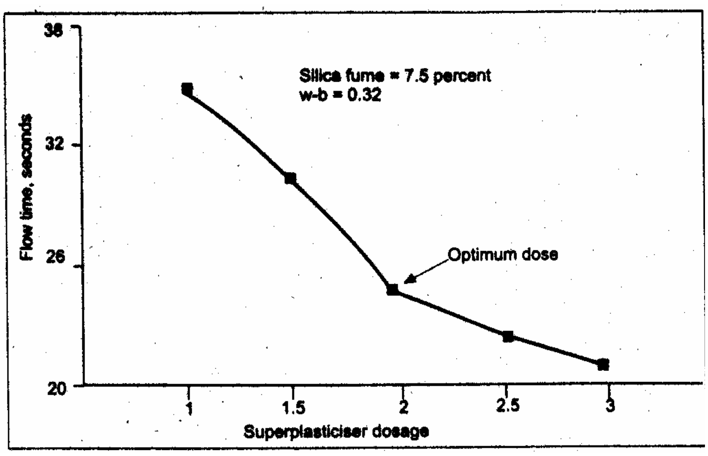


FIG. 3 SELECTION OF OPTIMUM SUPERPLASTICISER DOSAGE (KAIGA-1&2)

anomaly in case of SNF based superplasticiser. No difference between the flow time at 5 minutes and 60 minutes of a superplasticised cement grout was observed for some cements; again flow time increased with higher dosage of SNF for some other cement. The increase of SNF dosage beyond the saturation point may in some cases result in high slump for a long time, known as robust combination. However, there exists evidence of non-robust combination when increase in SNF dosage beyond saturation point leads to bleeding and segregation.

4.1.3 w/c_m ratio

Strength versus w/c_m curves for the cementitious materials proposed to be used in the concrete mix are generally required to be established in order to select the desirable w/c_m value to achieve target

Table 2. *HPC Mix Proportions for KAIGA-1&2, RAPP-3&4 and TAPP-3&4 Containment Structure*

Ingredients	*Mix proportion*			
	Kaiga 1&2	*RAPP 3&4*	*TAPP-3&4*	
	KM1	*RM1*	*TM1 **	*TM2 ***
Cement (Kg)	475	475	475	425
Silica Fume (Kg)	35.6	36	36.62	32
Water/Ice (Kg)	163	152	152	151
Coarse Aggregate (Kg)	1092	1047	1133	1017
Fine aggregates (Kg)	695	730	721	852
Super plasiciser (Kg)	8.4	9.63	9.53	9.563
w/c	0.343	0.32	0.32	0.35
w/c_m	0.32	0.3	0.3	0.33

Note: *TM1 for raft construction, **TM2 for wall and dome construction

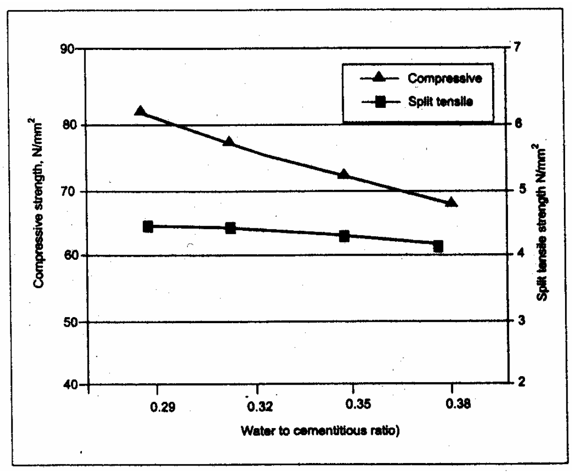


FIG. 4 STRENGTH VS. WATER CEMENTATIONS RATIO, W/C$_M$ (KAIGA-1&2)

strength of the mix. Such curves were established in developing the mix design of Kaiga HPC. It could be seen from Fig. 4 that for compressive strength requirement of 69 N/mm^2, w/c_m ratio of 0.35 was sufficient. However, w/c_m ratio 0.32 was used to meet the target split tensile strength of 4.37 N/mm^2.

In addition, unit water content is also an important parameter for achieving HPC with high slump, especially when fresh concrete is placed by pump. From the laboratory trials at Kaiga, it was established that each kg addition of silica fume in the concrete mix increases the unit water content by 0.4 to 0.61 when all other parameters are kept constant. Use of superplasticiser results in the reduction of water content by nearly 28 to 30 percent. Approximate unit water content for 175 mm slump concretes is163 liter.

4.1.4 Aggregate quantities

After finalizing the w/c_m, unit water content, and percentage content of silica fume and superplasticiser, the quantities of aggregates are proportioned absolute volume method [32].

4.1.5 Tests results and properties

The silica fume based HPC mix developed for Kaiga-1&2, RAPP-3&4 and TAPP-3&4 are given in Table 2. A number of tests are carried out to ascertain that these HPC mixes attained the requisite properties as its specification,

- laboratory tests
- field tests
- tests to qualify the mix

Laboratory test results verified the parameters such as slump, strength, and permeability, with corresponding specified values and/or compliance with the general codes [21,27]. By carrying out field tests, attempts were made to ensure that in-situ strength of the mix would meet the structural design requirements of the IC dome. Third category of tests confirms robustness of the mix with respect to the variability imparted in the concrete manufacturing process. These tests were carried out following well laid down quality assurance procedure. Table 3 contains laboratory test results of the mixes.

Table 3. *HPC Mix Proportions for KAIGA-1&2, RAPP-3&4 and TAPP-3&4 Containment Structure*

Parameters	*Performance standard*	*KM1*	*RM1*	*TM1/TM2*
Slump, mm	IS 1199	185	175	
Air content, %	IS 1199	0.95	1.3	0.7
Setting time, Hr:Min	IS 8142	6:20	6:20	7:00
Comp. Strength, MPa	IS 516	75.9	76.42	74.2
Split tensile strength , MPa	IS 516	4.36	4.70	4.35

4.2 HPC Mixes with Different Mineral Admixtures

HPC mixes of M60 grade concrete using SF, HRM, FA, GGBS were developed adopting the mix design method due to Aitcin [34]. Limiting value of target strength considered was 69 N/mm^2. Target initial slump was taken as 150 mm. Ingredient used were charracterised following relevant codes of practices. The mix proportions are given in Table 4. The mixes MSF and MHRM are of similar type. Same observation could be made between MFA and MGGBS. Laboratory test results are given in Table 5. All the mixes satisfy the requirements of target strength and slump.

However, initial slump of MHRM is less than that of MSF even though water content for both the mixes is same with the later one having less superplasticiser. This may be due to the fact that MHRM has more angular particle shape compared to SF. 28-day strength of MHRM is more than that of MSF though split tensile strength is almost same. Both compressive and tensile strength of MGGBS are higher than that of MFA. This is due to the fact that GGBS possess primary

Table 4. *HPC Mix Proportions with different mineral admixture*

Mix	*MA*	*Coarse Aggr. (Kg/m³)*	*Fine Aggr. (Kg/m³)*	*Cementitious material (kg/m³)*		*Water (Kg/m³)*	*w/c_m*	*CA (%)*
				Cement	*MA*			
MCM	–	1176.8	672.9	500.0	–	162.5	0.325	2.00
MFA	FA	1164.3	665.7	333.3	166.7	150.0	0.300	0.70
MGGBS	GGBS	1010.5	577.8	300.0	375.0	185.6	0.275	1.25
MSF	SF	1230.6	703.6	400.0	30.0	150.5	0.350	2.50
MHRM	HRM	1234.4	705.8	400.0	30.0	150.5	0.350	2.50

Note: MA = Mineral admixture, CA = Chemical admixture

Table 5. *Properties of HPC Mixes with different mineral admixture*

Properties	*Mixes*				
	MCM	*MHRM*	*MSF*	*MFA*	*MGGBS*
Fresh Concrete					
Density (kg/m³)	2579	2545	2531	2499	2534
Slump (mm)	150	165	170	200	180
Hardent Concrete (properties at 28 days)					
Strength (MPa)	73.5	77.0	70.5	69.3	73.3
Split tensile strength (MPa)	–	4.50	4.55	3.68	4.36
RCPT value (Coulomb)	1502	796	–	1443	1361

cementitious property while MFA does not. 28 day RCPT result indicates lowest permeability in case of MHRM and highest for MFA. Permeability of fly ash and GGBS based concrete falls drastically with age. This indicates that pozzolanic action of HRM and SF starts at quite early stage, while that of FA after some time.

5. PHYSICAL PROCESS

Three important components of physical process for construction of any concrete structure are,

i) Mixing of ingredients for production of concrete mix
ii) Transportation
iii) Placement

These three elements are discussed highlighting the special features when structure is constructed with HPC.

5.1 Mixing method

Objectives of mixing method is principally two fold; homogeneous mixing of ingredients, especially the cementitious materials within water, and avoidance of fine particle to flocculate.

These are important for hydration process to continue in desired way, modification of interfaces and avoidance of secondary type of interfaces. Sequence and time of mixing during production of the fresh concrete are two important parameters of the mixing method.

Kakizaki, et al. [3] found that suitable mixing procedure enhancing the properties of concrete indicating its influence on reaction mechanism to take place in desired way. Effects of mixing method on HPC were examined by producing the five mixes MCM, MFA, MGGBS, MSF and MHRM by five different multistage mixing methods [34]. Table 6 contains the outline of these methods. All the mixes, MCM, MFA, MGGBS, MSF and MHRM, were produced adopting these five different mixing methods and following properties were tested to examine the impact of these methods:

- Fresh concrete: density, slump and rheology
- Hardened concrete: Cube compressive strength, rapid chloride permeability test (RCPT) [35] and pH [36] value at 28 days.

Mixing of concrete ingredient adopting these methods was done in a laboratory version of pan mixture having blade speed of about 32 rpm. In an actual construction site, concrete after being produced in a batching mixing plant is generally transported by transit mixer to the construction site, which agitates the fresh concrete continuously with low frequency of rotation. In order to simulate this situation at laboratory, concrete after its mixing in the pan mixer, is kept in a stationary condition. The fresh concrete was put in the pan mixer and mixed for one-minute at an interval of 15 minutes up to one hour before taking the readings on slump.

Table 6. *Details of different mixing methods*

Mixing method	*Sequence of mixing*		
	Stage-1	*Stage-2*	*Stage-3*
I	CAg+FAg+C+W/3 (60 second)	W/3+MA (60 seconds)	W/3+CA (150 second)
II	CAg+FAg+C+MA+W/3 (60 second)	(2/3)W+CA (210 second)	
III	CAg+FAg+C+MA (60 second)	W/3 (60 seconds)	(2/3)W+CA (150 second)
IV	a) MSF & MHRM		
	CAg+FAg+C+W/3 (30 second)	W/3+MA (60 seconds)	W/3+CA (60 second)
	b) MCM, MFA & MGGBS		
	CAg+FAg+C+MA+W/3 (30 second)	(2/3)W+CA (120 second)	
V	a) MSF & MHRM		
	CAg+FAg+C+W/3 (90 second)	W/3+MA (60 seconds)	W/3+CA (60 second)
	c) MCM, MFA & MGGBS		
	CAg+FAg+C+MA+W/3 (90 second)	(2/3)W+CA (180 second)	

Note: CAg: Coarse aggregate; FAg: Fine aggregate; C: Cement
W: Water; MA: Mineral admixture; CA: Chemical admixture; HRM:High reactivity metakaolin; FA: Fly ash; GGBS: Ground granulated blast furnace slag; SF: Condensed silica fume

Mixing methods do no influence densities of the mixes. All mixing methods generally produced initial slump greater than 110 mm for all the five mixes except mixing methods I and II for mix MFA. Mixing methods III, IV, I, III & V, and II yielded maximum slump for mixes MCM, MFA, MGGBS, MSF and MHRM respectively.

Slump retention of mixes MFA and MGGBS is poor in case of all mixing method excepting the method V for MGGBS. From rheological point of view, mixing method II found to be most suitable for MCM, MSF and MHRM; and methods IV and V for MFA and MGGBS respectively. The mixing method does not seem to have much influence on segregating and bleeding.

The variation of 28-day cube compressive strength with mixing methods is plotted in Fig. 5. In general, mixing method II yielded good mixes in terms of compressive strength. Mixing method has impact on the strength of concrete mixes with or without mineral admixture and irrespective the type of admixtures.

Similar observation was observed incase of RCPT values at 28 days, Fig. 6. For mix MCM, high and moderate RCPT value is observed for mixing method I and IV. However, mixing method has little influence on the pH value, varies within a narrow range of 11.30 and 12.54, Table 7. The results also indicate that insignificant change occurs in pH value of concrete mixes due to inclusion of mineral admixture.

Among the five mixing methods, two stage methods are found to be suitable for SF based HPC. At Kaiga-1&2 and RAPP-3&4, two-stage mixing was found to be efficient. Silica fume was mixed dry along with aggregates, in the first stage for 5 seconds, followed by final mixing of 45 seconds after adding all the ingredients. Total cycle time for each batch of 0.5 m^3 worked out to 80 seconds, which is nearly one, and half times of that of normal strength concrete. The multistage mixing process has been found suitable for the mixes, which are reported in the present paper.

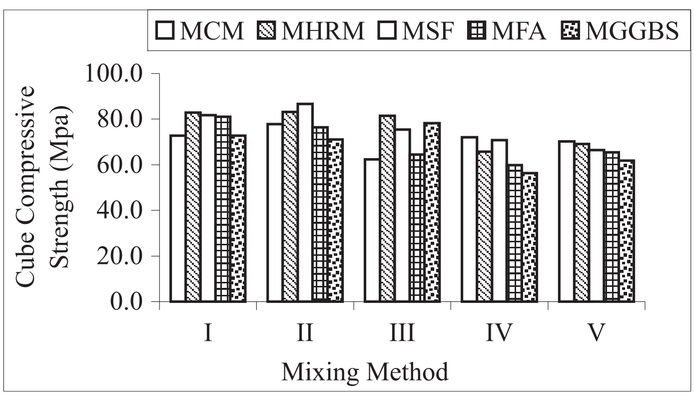


FIG. 5 VARIATION OF CUBE COMPRESSIVE STRENGTH (MPA) WITH MIXING METHOD (28 DAYS)

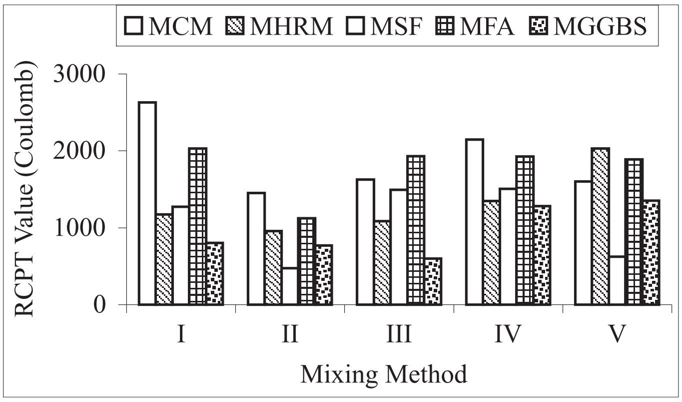


FIG. 6 VARIATION OF RCPT VALUE WITH MIXING METHOD (28 DAYS)

Table 7. *Variation of pH value for different mixes with mixing method*

Mixing method	*Mix*				
	MCM	*MFA*	*MGGBS*	*MSF*	*MHRM*
I	12.22	11.96	11.85	12.15	11.90
II	12.01	12.20	12.10	12.29	12.20
III	11.30	12.05	12.05	12.54	11.29
IV	12.38	11.98	12.20	12.29	11.93
V	11.38	11.48	11.60	12.15	11.94

5.2 Transportation

Fresh HPC mix should not start settling down, setting process needs to be avoided and temperature to be maintained within specified limit during transportation. Strictly, these are applicable for all type of concrete mix. All these are necessary so that fresh concrete can maintain desired rheology at the time of placement, and also help in keeping the hardened concrete free from potential interfaces of secondary nature.

Concrete mix should not be kept in static condition during long duration transportation mode, otherwise fresh concrete would tend to settle down causing undesirable bleeding. This is avoided by gentle agitation with long period motion. Transportation of rotary drum transit mixture is common mode of transportation. Another important consideration during transportation is that the temperature of the fresh concrete should not be changed. This is achieved by covering the rotary drum surface of transit mixture by insulating material. Re-dosing of retarder may become necessary some time to avoid initial setting in case of log duration between the time of production and placement.

5.3 Placement

Two important activities associated with placement of concrete are pouring and compaction. Objective of good placement process is to ensure that fresh concrete fills all the spaces encapsulate by shuttering. Reinforcement and other embedded elements offer obstruction of flow of fresh concrete for filling the spaces. Good workability is essential for this purpose, but appropriate pouring and compaction are also required. Otherwise, void will be left behind that would be potential source of weakness of hardened concrete in service condition.

Properly proportioned HPC mix is more cohesive than NSC and does not tend to segregate. But it loses slump rather rapidly compared to NSC and becomes sticky. This makes HPC more sensitive to the temperature, both intrinsic to concrete (such as heat of hydration) and ambient temperature. In hot weather, cooled concrete may be required specially when there is limitation on the placement temperature. For the PC Dome of Kaiga-1&2 and RAPP-3&4, production temperature was estimated at 15°C from the specified placement temperature at 23°C. Aggregates were pre-cooled and ice flakes replaced about 90% of water in the mix [20,27].

High slump HPC mix may create wrong impression that compaction is not necessary. Experience of Kaiga-1&2 and RAPP-3&4 suggests that good vibration after placement of fresh mix is essential especially for structural elements having high congestion with reinforcement and embedded parts. HPC mix if not engineered for self-compacting capability [37] shall be well compacted by means of external energy such as vibration.

6. CURING

Performance of HPC at hardened state is rather more sensitive to curing than that of NSC. Curing is the most important construction activity in case of HPC and also more elaborate as compared to NSC. Detail discussion of HPC curing is given in reference – 32. Distinct features of HPC mix are high cement content, mineral admixture and low water cement ratio. For a given level of workability, HPC mix has lesser value of w/c_m compared to NSC resulting in almost no bleeding in case of HPC. Loss of moisture from the exposed surface of fresh concrete at early age would cause plastic shrinkage.

Protection against moisture loss from fresh HPC is crucial for the development of strength as well as for durability. Again wet curing of HPC cannot be done at the very early stage, as this will increase w/c_m ratio at the surface. In view of this, curing of HPC is carried out in two stages, initial curing and final curing.

Curing compound was not found efficient for the initial curing. Spreading of opaque color plastic sheet over the fresh HPC contacting the exposed concrete surface is the efficient approach for initial curing of surface. Initial curing should be continued couple of hours beyond the final setting time of the concrete mix. Finishing work of surface, if any, should be completed prior to initial setting of the HPC.

Conventional methods like pounding water; using wet burlap and sprinkling water are suitable for curing of HPC with SF. Duration of about 10 days for wet curing is sufficient for HPC. Wet curing by means of wet burlap for10 days was adopted both at Kaiga-1&2 and RAPP-3&4. For fly ash based HPC, especially high volume fly ash concrete, wet curing is recommended till the age of 14 days.

7. CONCLUDING REMARKS

1) High performance concrete (HPC) is primarily normal strength concrete (NSC) with improved microstructure, especially at the interfaces like transition zone. Adopting low water cementitious material (cement and minerals admixtures) ratio, w/c_m, use of chemical and mineral admixtures, selection of optimum quantity of cementitious materials and suitable grading of all solid materials result in achieving this.
2) Reaction mechanism between the ingredients, physical process and curing are principally responsible for the modification of microstructure. Reaction mechanism, principally the hydration of cementitious material, is the basic element for development of HPC, while physical process and curing are the supportive ones.
3) Major steps for successful engineering of a concrete structure with HPC are mix proportion, mixing of ingredients for production of concrete, placement and curing. These steps are involved with certain additional special features and are more elaborate activities in case of HPC as compared to NSC. Quality of every stage of work needs to be assured.
4) Quality control of ingredients by characterization following well laid down specification is essential. Good compatibility of cementitious material and superplasticiser is also important.
5) Determination of optimum quantity of cementitious materials, i.e. cement and reactive mineral admixtures, and water cementitious material ratio (w/c_m) is key to the success of proportioning of an optimum HPC mix.
6) The mineral admixtures like silica fume (SF), high reactive metakaolin (HRM), fly ash (FA) and ground granulated blast furnace slag (GGBS) are found suitable for developing HPC using Indian cement. HPC mixes with SF and HRM are similar, and that of GGBS and PFA are similar.
7) Mixing method to produce fresh concrete should be such that all the ingredients, especially the cementitious materials are mixed homogeneously. Multistage mixing method has been found to be suitable for HPC. Density of fresh concrete appears to be insensitive to mixing methods. Mixing method does not have noticeable influence on bleeding, segregation, stickiness and cohesiveness on the concrete mixes, but it has influence on the initial slump and slump retention. It has also influence on the compressive strength of the mixes and RCPT value.
8) No noticeable influence of mixing method observed was on the pH value irrespective of NSC and HPC. Use of mineral admixture has very little influence on the pH value of the mix.
9) Fresh HPC mix should not start settling down, setting process needs to be avoided and temperature to be maintained within specified limit during transportation.

10) Placement process, pouring and compaction, should ensure that fresh concrete fills all the spaces encapsulate by shuttering. HPC mix, if not engineered for self-compacting capability, shall be well compacted by means of external energy such as vibration. Otherwise, void will be left behind that would be potential source of weakness of hardened concrete in service condition.
11) Curing maintains the sufficient moisture content and required temperature within the concrete mass so the reaction mechanism can progress satisfactorily and be completed to a desirable state. Curing has significant influence on development of properties of HPC. HPC is cured in two stages; initial curing (without using water and continued up to one or two hours beyond the final setting time of concrete) and wet curing (commenced immediately after initial curing and duration of about 10 days.).

ACKNOWLEDGEMENTS

The results presented on MCM, MSF, MHRM, MPFA, MGGBS are outcome of AERB sponsored research project on durability of HPC being undertaken at the Construction Engineering Department, Jadavpur University, Kolkata. Author is grateful to Professor Subhajit Saraswati of Jadavpur Unversity for providing the information on these HPC mixes. He is also grateful to Shri Subrata Chowdhury, Ultratech Cement Ltd, Mumbai and A.D. Roshan, AERB, Mumbai for their valuable suggestions.

REFERENCES

[1] “Cement and concrete terminology”, SP. No. ACI 116, “American Concrete Institute”, Detroit, USA, pp 68.
[2] Mehta P.K., Monteiro Paulo J.M., “Concrete; microstructure, properties and materials”, Indian edition, Indian Institutes, Chennai, India, 1997.
[3] Neville A.M., “Properties of concrete”, Third edition, ELBS with Longman, Singapore, 1995.
[4] “State of the art report on high strength concrete”, ACI 363R-84, American Concrete Institute, 1984.
[5] “High performance concrete – a state of the art report”, No. C/FR-91/113, Strategic Highway Research Programme (SHRP), USA.
[6] Basu Prabir C., “Performance requirements of HPC for Indian NPP structure”, The Indian Concrete Journal, September 1999, Vol. 73, No.9, pp. 539-546.
[7] “Indian Standard: Plain and Reinforced concrete Code of Practice”, IS 456 (fourth revision), July 2000, Bureau of Indian Standard, New Delhi, India.
[8] Basu Prabir C., “NPP containment structures: Indian experience in silica fume based HPC”, The Indian Concrete Journal, Volume 75, No.10, October 2001, pp.656-666.
[9] “Condensed silica fume”, FIP state of the art report, Thomas Telford House, London, 1988.
[10] O.E. Gjorv, “Silica fume: an overview”, Advances in Concrete Technology, ed. by V.M. Malhotra.

[11] Wesche K, “Fly Ash in concrete, properties and performance”, RILEM: E&FNSPON.

[12] Celik Ozyilidrium and Woodrow J Halstead, “Improved concrete quality with combination of fly ash and silica fume”, ACI, Materials Journal, Vol. 92, No.3, May-June, 1995.

[13] Malhotra, V. M, and Ramezanianpour, A. A. Fly Ash in concrete, 2nd edition, CANMET, 1994.

[14] Basu Prabir C., Saraswati Subhajit,” High Volume Fly Ash Concrete with Indian Ingredients”, The Indian Concrete Journal, Vol. 80, No.3, March 2006, pp. 37-48.

[15] Swamy, R. N., Role of slag in the development of durable, Sustainable High Strength Concrete, International Symposium on Concrete technology for Sustainable Development in the Twenty-First Century, Hyderabad, India, 1999.

[16] Saraswati Subhajit, Basu Prabir C.,” Concrete composites with ground granulated blast furnace slag”, The Indian Concrete Journal, Vol. 80, No. 6, June 2006, pp. 29-40.

[17] Michael A. Caldrone, Karen A. Guber, and Ronald G. Burg, “High reactivity metakaolin: A new generation mineral admixture”, Concrete International, November 1994, pp. 37-40.

[18] Shreeti Mavinkurve, Basu Prabir C., Kulkarni Vijay R.,” High Performance Concrete Using High Reactivity Metakaolin”, The Indian Concrete Journal, Vol. 77, May 2003, No. 5, pp. 1077-1085.

[19] Cook D.J., “Rice hush ash”, Cement replacement materials; R.N.A. Swamy, 1986, pp.171-196.

[20] Mittal Amit, Basu Prabir C, “Development of HPC for P.C. dome of NPP, Kaiga”, The Indian Concrete Journal, September 1999, Vol. 73, No. 4, pp. 548-556.

[21] Bapat S.G., Agarwal Neeraj, “Experience in the development and production of HPC for RAPP-3&4”, The Indian Concrete Journal, September 1999, Vol. 73, No. 9, pp. 582-586.

[21] Saini Sunil, Dhuri S.S., Kanhere D.K., Momin S.S., “High performance concrete for an urban viaduct in Mumbai”, The Indian Concrete Journal, Vol. 75, No.10, October 2001, pp. 634-640.

[23] Joshi N.G., “Bandra-Worli seal link: Evaluation of HPC mixes containing silica fume”, The Indian Concrete Journal, Vol. 75, No. 10, October 2001, pp. 627-633.

[24] Neville A.M., “Maintenance and durability of concrete structures”, Key note address, Proceedings, International Conference on Maintenance and Durability of Concrete Structures, March 4-6, 1997, J.N.T. University, Hyderabad, 1997.

[25] Shackelford James F., “Introduction to material science for engineers”, 3rd edition, Maxwell Macmillan International Editions, 1992.

[26] Basu Prabir C., “High performance concrete: mechanism and application”, ICI Journal, Vol. 2, No. 1, April-June 2001, pp. 15-26.

[27] Jagdish H.T., Ram Mohan V., Suresh M.S., Rao Ramesh and Kamath Radhakrishna R., “Field trials of PC dome of NPP, Kaiga”, The Indian Concrete Journal, September 1999, Vol. 73, No. 9, pp. 571-579.

[28] Basu Prabir C., “Curing of high performance concrete, Indian Experience”, The Indian Concrete Journal, Vol. 73, No. 9, 1999, pp. 571-581.

[29] Nkinamubanzi, P.C.; Kim, B.G; Aitcin P.C., “Some key cement factors that control the compatibility between naphthalene based superplasticiser and ordinary port land cements”, Industrial Italiana del cemento, V.iiC, No. 2, Feb.2002, pp. 192-202.

[30] Aitcin, P.C., “High performance concrete”, E&FNSPON, London 1998.

[31] Mahalingam, N.P., “Chemical admixture in concrete: side effects and compatibility problems”, The Indian Concrete Journal, Vol. 73, No. 6, June 1999, pp. 367-374.

[32] “Recommended Guidelines for Mix Proportioning”, IS 10262, Bureau of Indian Standard, New Delhi.

[33] M. Kakizaki, H-Edahiro, T. Tochigi and T. Niki, “Effects of mixing method on mechanical properties and pore structures of ultra high strength concrete”, SP 132-54.

[34] PCB & SS paper on durability of HPC at Chicago

[35] “American standard test method for electrical indication of concrete ability to resists chloride ion penetration”, ASTM C 1202, 97.

[36] “Corrosion and corrosion protection of prestressed concrete bridges in Marine environment”, Special Report No.18, IRC Highway Research Board, New Delhi, 1996.

[37] Skarendahl, A., Petersson, O., “Self-Compacting Concrete”, State of the art Report of RILEM Technical Committee 174-SCC, Report 23, 2000.

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

ESTIMATION OF COST OF DREDGING

M.M. KAMATH*

1. INTRODUCTION

The estimation of cost of a dredging job is not a straight forward exercise as in the case of civil, electrical and mechanical works, in view of the complexities involved in the calculation. No two estimates of works to be carried out will be the same, due to the variable nature of the components involved in the projects. Hence it is not possible to standardize the cost estimate format.

The dredging industry is highly capital intensive and often involves deployment of few main pieces of equipment on each project. The variable costs such as labour, fuel and materials are minor compared to the fixed costs, arising from capital investment. This is the fundamental difference between dredging and most other civil engineering works as can be seen from the following table.

Table 1. *Comparative costs of dredging vs other civil engineering works*

(In percentage)

Costs	*Dredging*	*Road, Rail, Water works*	*Building*
Variable Cost	35	69	80
Fixed Cost	65	31	20
Investment in Equipment	14.5	5	1

Source: Netherlands Central Bureau of Statistics, 1962

Further, it is a multi-disciplinary exercise. No two estimate of any work will be the same, due to the alternative methods involved in tackling the job.

2. ELEMENTS OF COST

The elements that constitute the cost of a dredging work are:

- Mobilisation and Demobilisation of plant & equipment
- Operating cost of equipment, materials, consumables, labour etc.

*Shree Matha, Mizar Ramakrishna Pai Compound, Mannagudda, Mangalore, India 575 003
E-mail: mmkamath40@yahoo.com

- Financing and Insurance
- Taxes, Duties etc
- Miscellaneous costs, such as costs of sub contractors, agents, cost of facilities for the employer etc.
- Overheads

3. PRICE OF DREDGING

The price and cost of dredging are different.

The price of a dredging job will be derived based on the cost as worked out above, taking in to consideration the following additional factors.

- Profit
- Risks
    - Technical
    - Financial
    - Political
- Status of order book
- Future job potentialities
- Overall market situation

The above factors are mostly commercial secrets, which vary from company to company and hence is difficult to generalize, as the dredging industry operates in a global marketplace.

4. FACTORS AFFECTING UNIT COSTS

The unit costs are dependent on number of factors and they can be grouped into two main categories

- Directly related to time
- Independent of contract duration

The costs are also dependent on the dredging, production rate. In the case of trailer suction dredging, normally the cost is inversely proportional to the dredger capacity.

4.1 Time Related Cost

The cost related to time are

- Depreciation of the plant
- Interest on capital
- Maintenance and repairs
- Fuel and lubricants
- Establishment and crew charges
- Insurance and
- Overheads

4.1.1 Depreciation

The depreciation is mainly dependent on the following factors.

- Original cost of the dredger
- Life span of the dredger
- Average utilization of the dredger and
- Residual value of the dredger

4.1.2 Interest on Capital

This will depend on the source of finance.

There are two methods of charging Depreciation and Interest as adopted by international dredging companies.

- Linear method
- Annuity method

4.1.3 Maintenance and Repairs

The maintenance and repair charges depends upon the installed machinery, site conditions etc.

The Federation of Major Contractors NIVAG Netherlands, had published the ‘Depreciation and Interest’ and ‘Maintenance and Repairs’ figures for most of the construction equipments including dredgers. Subsequently, these were published in Dutch as “VGBouw, Operating Costs Standards for Construction Equipment” issued by the major construction companies. However, these figures are to be used with caution as these are based on average circumstances in Netherlands.

Subsequently, CIRIA published “Cost Standards for Dredging Equipment – 2005” which is a further revision of NIVAG, and latterly VGBouw. This has been audited by an independent panel and being used by the dredging industry.

The figures relating to cost of spares and maintenance, would be available based on the past experience and include paint, steelwire ropes, welding rods, gases, normal running and repairs of machineries etc. A special item which will have to be considered is the case of cutter suction dredgers, is the cutter teeth and adapters specially when working in hard soils like rocks etc. Each cutter teeth costs about Rs. 1,000/- and there have been instances when 45 cutter teeth had to be changed, every twenty minutes, in the case of rock dredging.

The wear on the pumps and the pipeline is a source of high cost to the dredge operators, specially when working in abrasive and coarse materials and long pipe lines. Unfortunately, good reliable information on wear of pumps and pipelines is scarce. Normally, operators will allow their pipes to wear until a specified minimum wall thickness is reached, which is normally 6 mm, depending upon the pressure in the line. The following information is relevant with reference to wear in the dredge pipelines.

- A.N. Bray, in his book on “Dredging, A Handbook for Engineers” (1997), has stated that pipe wear rates may vary from 1,00,000 cum of dredged material per mm of wear, to well over 1,00,00,000 cum for the same wear.

- TM Turner, in his book on "Fundamentals of Hydraulic Dredging" (1996), has developed a mathematical model for the life of pipe lines as under:

$$L = \frac{K \times B \times C \times S}{W \times (d_{50})^{0.8} \times A}$$

Where
L = Life in cubic yards of solids pumped
K = Best available prediction of life
B = Brinnel Hardness
C = Concentration of slurry
S = Size factor (inside diameter of pipe)
W = Specific gravity
d_{50} = Median diameter of the soil particle
A = Angularity or sharpness of soil particles

- The Bureau of Ports and Harbours, Ministry of Transport, Japan International Cooperation Agency, has recommended based on the field experience of operation of a 4000 horse power, cutter suction dredger as under:

Table 2. *Wear in the dredge pipelines*

Type of Soil	*Rotation of Pipe(in cum)*	*Renewal of Pipe(in cum)*
Medium Sand	2,50,000	6,00,000
Coarse Sand	-	3,00,000
Sand with Gravel	-	1,50,000
Gravel	-	60,000

From the above it may be seen that the wear and tear of the pipe line is dependent on many factors and is site specific.

4.1.4 Fuel and Lubricants

This is one of the major element in the cost, especially the fuel and is mainly related to the installed Horse Power. The cost of fuel is about 25 to 30% of the total cost. The fuel consumption will depend upon

- sailing distance to dumping ground, in the case of trailer suction dredgers
- type of material and discharge pipe length, in the case of cutter suction dredgers

The average rate of fuel consumption is equal to 0.18 liter per horse power per hour. Normally for cutter suction dredgers, the fuel requirement is about 1.5 liters per cum of soil dredged with booster and 1.3 liter per cum without booster.

The lubricant cost is normally taken as 10% of the fuel cost.

4.1.5 Establishment and Crew Charges

This will depend upon the number of crew employed on the dredger. This again is governed by local customs and agreement with unions etc. This would also include expenses towards food, accommodation, travel costs, social security etc.

4.1.6 Insurance

Normally the insurance premium are fixed for dredging equipment and is about 1.6% per annum of the cost of the dredger. However, whenever jobs are be carried out in special danger prone areas, special insurance cover may be required e.g. war zones, earlier mined areas, exposed conditions etc.

4.1.7 Overheads

This will depend upon the company policy and normally is in the range of 10 to 15% of the total cost, in the absence of detailed calculations.

4.2 Cost Independent of Contract Duration

The costs which are independent of the job duration are

- Surveys
- Mobilization and Demobilisation
- Contract related Costs

4.2.1 Surveys

This will depend upon the site conditions, environment, availability of pre bid data and its quality etc. Surveys will also have to be carried out, on completion of the job. The standard adopted by U.S. Corps. of Engineers is, 0.7 to 2.0% of the total cost.

4.2.2 Mobilization and Demobilisation

Mobilization costs for dredging jobs could be quite high, depending upon the type of equipment, their volume, weight, geographical location etc. and have a significant bearing on the cost of the work.

The dredgers are normally transported by

- Ships
- Dock ships
- Submersible pontoons
- Barge towed by a tug
- Direct towing by a tug

In addition to the above, the starting up cost, including setting up of site office, initial preparation of dredger and other equipments, reclamation bunds etc have also to be considered.

The demobilization costs normally includes the winding up costs. These are very difficult to estimate at the time of tendering for the job, as this will depend upon the destination of the equipment, after completion of the work etc. Hence, normally a figure of about 50% of mobilization cost is included in the cost, in the absence of specific information to the contrary.

4.2.3 Contract Related Costs

The various contract related costs are

- Bid Preparation Cost: - In some cases, bid preparation will include site visits, preliminary soil investigation etc. An average value of 1% of the project cost is normally considered
- Negotiation:- Sometimes the contracts are negotiated and even awarded on cost plus basis.
- Final claims:- In some cases, the contractor may claim additional costs due to the change in the site conditions etc.

5. BASIC OPERATING COSTS AND COST PER UNIT TIME

From the above, the basic operating costs can be worked out. To work out the cost per unit time, annual utilization of the dredger is required and the average figure will depend upon the company policy. The average working hours per week are normally taken as 144, for cutter suction dredgers. This will again vary from dredger to dredger and the site conditions. These figures are commercial figures which are not normally disclosed by dredging companies.

6. PRODUCTION ESTIMATES

The production estimates are required to calculate the deployment time of each type of dredger to complete the job. The starting point for the estimates is the type of soil to be dredged. Based on the soil investigation data, schematic soil profile is drawn for the specified location and based on the above, the quantities of soil to be dredged in each category is quantified. The next step is to select the type of dredger to be deployed.

The dredging cost is highly sensitive to the classification of soil to be dredged and also suitability of a particular type of dredger, as can be seen from Table 3

Table 3. *Comparative costs of various types of dredgers vs materials*

Type of Dredger *Cost Factor*	*Cutter Suction*	*Trailing Suction* *Material*	*Bucket*
1.00	Fine Silt	Sand	Soft Clay
1.30 – 2.00	Sand	Fine Silt	Sand
2.30 – 2.70	Soft Clay	Soft Clay	Medium Gravel
2.70 – 3.50	Medium Clay Medium Gravel	Medium Gravel	Coarse Gravel Stiff Clay
3.50 – 4.00	Hard ClayChalk	Coarse Gravel	ChalkHard Clay
6.00 – 10.00	Soft Rock	Stiff Clay	Soft Rock
15.00 – 25.00	Hard Rock (with blasting)	Hard Rock (with blasting)	Hard Rock (with blasting)

Source: Anthony Bates – Dredging + Port Construction (October 1978)
Soil Investigations for Dredging Works.

To proceed further, the realistic output of the dredger, in similar types of soil, will have to be obtained from the data bank maintained by the company from comparable jobs. This will require the knowledge regarding the dredgeability of the soil, and its effect on the equipment, in terms of wear and tear, actual operating hours likely to be achieved etc. This in turn will depend upon the environmental constraints, shipping movements, repairs and maintenance etc.

From the above data, the basic operating cost is determined from the cost per hour, the productivity per hour and the likely number of productive hours in the operating period. Summarizing, the various steps involved in the cost estimation are

- Calculation of dredging quantities for each type of soil
- Selection of the dredger type
- Estimation of the output for various soils
- Optimization of the dredger type / size
- Estimation of the total time required to carry out the job
- Estimation of cost viz. total time x unit cost
- Estimation of cost of mobilization and demobilization

From the above, the complexities involved in the estimation of costs can be gauged and no standardized format can be adopted for working out the costs. The working out of the price is still more complex, in view of the uncertainties involved.

7. CONCLUSION

- Dredging occupies an important position in any port development project.
- Dredging is a highly capital intensive activity and hence the correct estimation of dredging costs is a pre-requisite for correct evaluation of port projects.
- Estimation of costs of dredging, is not a straight forward exercise, as in the case of civil, electrical and mechanical works.
- The various steps involved in the cost estimation are
    - Calculation of dredging quantities for each type of soil
    - Selection of the dredger type
    - Estimation of the output for various soils
    - Optimization of the dredger type / size
    - Estimation of the total time required to carry out the job
    - Estimation of cost viz. total time x unit cost
    - Estimation of cost of mobilization and demobilization
- The price of dredging job is derived based on the costs and other components such as overheads, profits, risk factors etc.

REFERENCES

[1] Anthony Bates, Soil Investigations for Dredging Works, Dredging + Port Construction,1978

[2] Bray, R.N., "Dredging" *A Handbook for Engineers,* Edward Arnold, London, 1979

[3] Bray R.N., Cost Standards for Dredging Equipment, CIRIA, 2005

[4] Federation of Major Contractors NIVAG Netherlands, Depreciation and Interest & Maintenance and Repairs

[5] Japan International Cooperation Agency, Bureau of Ports & Harbours, Dredger & Dredging Work, 1990

[6] Kamath M. M., "Under Water Blasting & Dredging", *Port Technology International* London, Twelfth Edition, 2000

[7] Kamath M. M., "Maintenance Dredging Practice at New Mangalore Port", *Port Technology International London*, Fifteenth Edition. 2001

[8] Kamath M. M., "Dredging an Overview", *Seminar on Capital & Maintenance Dredging Practices,* New Mangalore Port, 2005

[9] Turner T.M., "Fundamentals of Hydraulic Dredging", Second Edition, *American Society of Civil Engineers*, U.S.A., 1996

[10] VGBouw, Operating Costs Standards for Construction Equipment, Eleventh Revised Edition, The Netherlands, 1995

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

ROCK DREDGING - A STATE OF THE ART REVIEW

M.M. KAMATH*

1. INTRODUCTION

The present day trend towards bigger ships such as super tankers demand improved port facilities which call for greater depths involving deepening the existing channels etc. In the process, harder soil formations are encountered which cannot be expeditiously and economically cleared by ordinary methods of dredging. The dredging of rock is the most expensive type of dredging carried out in normal maritime engineering. However, powerful cutter suction dredgers with installed cutter horse power in excess of 6000 equipped with heavy duty cutters and anchors specially designed for rock cutting have been developed. These dredgers are capable of directly dredging hard rock formations having average compressive strength of about 40 Mpa without pre-treatment. In harder formations pre-treatment by the use of explosives and underwater blasting is necessary. Under water blasting is also used for rock removal for foundation of marine structures, trenching for sub sea pipelines etc. When large volumes of rock are involved, the removal by normal underwater blasting and clearance by divers is an expensive and time consuming process. In such situations under water drilling and blasting will have to be employed.

2. INVESTIGATIONS

Since, rock dredging and pre-treatment equipment are so specialized and expensive, it is vital that detailed investigations are carried out to determine the type and quantum of rock to be dredged. In rock dredging, the investigations assume great importance since its results not only significantly affect the design and cost of the work, but may also cast serious doubts as to the desirability of the carrying out of the work at all.

The two important investigations required to be carried out to delineate the rock profile are –

2.1 Seismic Profile Survey

Seismic profile surveys are now being extensively used in the dredging industry. There are a number of systems available, but all of them are based on the same principle. A high energy acoustic

*Shree Matha, Mizar Ramakrishna Pai Compound, Mannagudda, Mangalore, India 575 003
E-mail: mmkamath40@yahoo.com

source is triggered periodically just below the water surface. The acoustic energy travels outwards in the form of a sphere. The sea bed & horizon lying beneath the survey vessel reflect this signal back towards the surface where it is received by a hydrophone. The returning signal can then be recorded on a paper chart recorder, depicting a continuous record of the sub bottom profile. A general lay out of the system is shown in Figure 1. Correlation boreholes should later the sunk at selected positions to prove the seismic profile.

2.2 Jet Probing and Drilling

Water jet emitting from the end of a nozzle is used to determine the layer of non-penetration. Since there is no recovery of samples, the material cannot be identified. This is supplemented by drilling. Rotary drilling is carried out to obtain rock cores varying from 25 to 100 mm. dia. Core examination is an important guide to the rock quality etc. The rock cores enable a detailed engineering and geological assessment to be carried out of the strata. The classification of rocks to be dredged are classified as specified in PIANC 'Classification of Soils and Rocks to be Dredged – Report of a Working Group of Permanent Technical Committee II (1984)'.

3. METHODS OF ROCK BREAKING

Rock breaking can be broadly classified under the following heads, depending upon the method used.

3.1 Mechanical

3.1.1 Percussion Rock Breaking

The fragmentation of rock by means of drop chisels is the time proven conventional method. Number of chisels are mounted on a pontoon and the same are allowed to fall on the same spot several times till the desired penetration is reached. These have normally a weight of 5 to 30 tonnes and a fall height of 25 to 30 metres. A typical rock breaking chisel mechanism is shown in Figure 2.

3.1.2 Rock-Ripping

Hydraulic and pneumatic hammers which can be conveniently mounted on the boom of a hydraulic back acting machines are also now available. These are able to break rock having compressive strength upto 8 Mpa. A typical rock ripping mechanism is shown in Figure 3.

3.1.3 Direct Dredging

Generally, Igneous and Metamorphic rocks cannot be dredged directly without pretreatment, whereas most of the sedimentary rocks can be dredged. This will however depend upon the strength and structure of the rock.

There are five types of dredgers, which can be deployed for direct dredging of rock without pretreatment. These are

- back hoe
- bucket

- cutter suction
- dipper, and
- grab dredgers

However, the universally used equipment is the cutter suction dredger when large volumes of rock are involved. Powerful cutter suction dredgers with 6000 H.P. on the cutter and having a total output of about 22,000 H.P with ladder weights of over 100 tonnes, equipped with special cutters and heavy anchors specially designed for rock cutting are available. Direct rock dredging has been reported by the use of cutter suction dredgers in rocks up to an average compressive strength of 40 MPa. However this does not depend upon only on the compressive strength, but also on layering, fracture etc. The choice between direct dredging and dredging after pretreatment is governed by economic considerations rather than technical limitations.

3.2 Chemical

The breaking of rocks by chemical methods involve the use of explosives. The explosives can be used in different forms.

3.2.1 Surface Blasting

3.2.1.1 Slab Charges

The explosive is directly placed on the rock in a suitable grid pattern and sufficient head of water of about 8 m. is required to confine the explosive energy. The individual charges are in the range of 10 to 25 kg. The efficiency depends upon, the depth of water and the contact between the explosive and the rock and the shattering effect of the detonation wave. The gases produced are largely dissipated in the water. This water borne shock wave has destructive effect on the marine life and also vulnerable structures within the immediate working area. A typical slab charge is shown in Figure 4.

3.2.1.2 Shaped Charges

The efficiency of surface blasting can be improved by the use of ‘Shaped Charges’. The principle is based on ‘Monroe-Effect’ which focuses the explosive shock wave in a specific direction giving greater efficiency of blasting. The charge mainly consists of a metal or a plastic cylindrical canister with a conical bottom filled with a suitable explosive and placed on the rock to be blasted in a suitable grid pattern. The bottom of the canister is normally fitted with a concrete ring to provide weight from stability considerations. There are number of patented devices available in the market.

The advantages of shaped charge over the slab charge are

- increased output,
- improved fragmentation,
- reduced vibration and shock pressures and
- no limitations of water depth as it can be used in depths from 0 to 300 m.

However, shaped charges are economical for removal of rock up to a thickness of about 0.5 to 1 meter. The principle of operation and a typical shaped charge container is shown in Figure 5. A steel frame work is normally used to place the charges in a predetermined pattern. This steel frame work is placed in position with the help of a crane assisted by divers.

4. DRILLING AND BLASTING

The drilling and blasting method is suitable when the head of water is low and the depth of rock to be excavated is relatively high.

4.1 Factors Affecting the Performance of Drilling and Blasting Work

- Depth of water and other environmental parameters such as current, waves, tidal range, ice etc.
- Depth and type of overburden
- Nature and formation of the rock
- Location of work such as navigational constraints etc.
- Quantum of rock to be removed
- Environmental factors such as noise, vibration etc.

4.2 Methods of Drilling

Drilling in rock can be carried out by any one of the following methods.

- Divers using handheld drills.
- Divers using heavy mechanical equipment.
- Above water with a floating pontoon
- Above water with a jack up platform.

The configuration of the drilling system can be either

- Drills cantilevered over side
- Drills mounted over a central well

A combination of the above two systems is sometimes adopted depending upon the job requirement and site conditions.

A typical four tower floating drilling pontoon with drills cantilevered over side is shown in Figure 6.

A floating drilling platform with a central well is shown in Figure 7.

4.3 Drilling Procedure

The 'Over Burden Drilling' (OD) is the commonly used method for under water drilling work, as in most of the cases, the rock is overlaid by overburden. The 'OD' equipment consists of an outer casing tube with a ring bit of cemented carbide at the lower end. This casing tube encloses an inner drill string of standard drill steel with a cross or button bit. The casing tube and the inner drill steel are connected to the rock drill by a special shank adapter, which transfers both impact and rotary forces to the casing as well as the drill rods.

The drilling cycle can be broken down into the following sub-cycles:

- Setting up of drilling rig
- Lowering outer casing to bed and collaring into rock
- Drilling through rock with inner drill steel to design depth
- Charging the hole with explosives with the use of pneumatic charger
- Lifting of outer casing and retrieving the detonator/fuse
- Connecting the wires and testing
- Moving the drill rig into safe position
- Blasting

A typical drilling cycle is shown in Figure 8. Figure 9 shows the details of drilling and charging of explosives. The average time required for setting up of rig, drilling and charging of a typical pattern of 6 rows of 10 holes each for one setting of the platform equipped with four drilling towers is about 20 to 24 hrs. This however will depend upon depth of water, environmental conditions, position fixing system, depth of overburden, depth of rock to be drilled, hardness of the rock, and joints/fissures in the rock etc. The drillability of a rock depends on many factors such as composition, strength, structure and homogeneity of rock etc.

4.4 Types of Explosives

The explosive suitable for under water blasting should have

- Safety in handling
- Largest practical diameter
- High density
- High bulk strength
- Remain resistant to flash-over and impact during the period of submersion
- Remain effective form the time of charging, till blasting, withstanding high hydrostatic pressures
- Preferably suitable for use with a pneumatic charger

The explosive with a high Nitro-Glycerin content can remain under water with almost undiminished ability. The high resistance to water is more of a disadvantage than advantage as the high Nitro-Glycerin content also increases the flash-over sensitivity. A Nitro-Glycerin content of 30 to 40% is normally recommended for under water blasting work. It is always advisable to consult the manufacturers when selecting the type of explosives for a particular job.

4.5 Charging of Explosives

One of the main aims of under water blasting is to attain a high loading concentration in the drill hole and many times is a difficult process. For the work on ground, tamping rods can be used. But for under water work normally, compressed air chargers are used. The charging hose should have anti-static property with longitudinal fluting on the inside. This will facilitate increased pressure in the hose in the event of a cartridge jams inside. A nozzle is fitted at the end incorporating knives which cut the paper round the cartridge when it is forced out through the nozzle. Flexible stemming rods for checking and loading of explosives into the drill hole are also used.

5. METHODS OF DETONATION

The initiation of explosives is carried out by one of the following methods.

5.1 Detonating Cord

This normally consists of a thin core of high explosive protected by suitable layers of protective materials. The detonating fuse should withstand high head of water for long periods having high abrasion resistance. The detonating fuse is used in conjunction with a suitable booster or primer.

5.2 Electric Detonators

These are of special design and robust construction. Two detonators are normally used in each hole connected in different circuits to increase the reliability of complete detonation. The great advantage of electric detonation is that each detonator and the entire round can be checked before the round is fired. Further the moment of detonation can also be controlled. To increase fragmentation and reduce the ground vibrations, millisecond electric delay detonators are used.

5.3 Toroidal Detonators

These are electric detonators whose wires are coupled to ferrite rings. These are significantly safe compared to the conventional electric detonators.

5.4 Nonel System

This is a patented system. This consists of a flexible plastic tube of 3 mm in dia. whose inside surface is coated with a special substance which maintains the propagation of a shock wave up to 2000 m/sec. The shock waves does not affect the ** tube and will not initiate any explosives column through which it passes. The tubes can be bunched together at a special bunch connector which in turn can be connected to a common initiation point. Delay devices can also be incorporated in the system.

6. TESTING OF CIRCUIT

After completion of charging and connection, tests for current leakage and resistance should be invariably carried out, when electric detonators are used. The insulation resistance to earth is checked by an instrument called ‘Current Leakage Tester’. After the above test, the circuits are tested for internal resistance with an ‘Ohm Meter’. The resistance to earth is also measured and it must be ensured that this is at least twice as high as the internal resistance of the circuit.

7. EXPLOSIVE BLAST RATIOS

The breaking characteristics of a rock will depend upon its behaviour towards shock and other dynamic loads inter-related to strength, structure and homogeneity of rock. Normally, in underwater blasting the spacing of drill holes is equal to the burden. The drill holes are to be drilled deeper than the final design level to a depth known as the 'Sub Grade Depth' which is approximately equal to the spacing of holes or burden, whichever is larger. This additional depth will ensure proper angle of breakage, fracturing the rock below the required depth etc.

The smaller spacing of drill holes are not desirable from safety considerations and larger spacing increases the size of blasted rock and un-blasted patches pose difficulties in dredging. The normal blast ratios for the land blasting work is about 0.6 kg/cum. The blast ratios for underwater surface blasting depend upon the rock type and are in the range of 3.5 to 7 kg/sqm. However the blast ratios for under water blasting is comparatively high for the following reasons.

- Higher fragmentation required
- Compensation for the extra head of water
- Compensation for the over burden, if any

The blast ratios for under water drilling and blasting varies from 0.6 to 3.0 kg/cum. depending upon the type of rock and type of dredger to be deployed for the work.

The blast ratios are normally selected based on experience and a rough guide of blast ratios depending on the fragmentation required for different types of dredging equipment are shown in Table 1.

Table 1. *Blast ratios for various degree of fragmentation*

Type of Dredger	*Fragmentation required in mm. based on average equipment*	*Range of blast ratio kg/cum.*	
		Soft Rock	*HardRock*
Cutter Suction < 300	1.25	3.00	
Grab < 450 1.00	3.00		
Bucket < 600	0.90	3.00	
Dipper < 750 0.60	1.75		

Source: Abraham J.L. – Under Water Drilling and Blasting for Rock Dredging.

Normally it is not possible to drill and blast rock in depths greater than three metres in a single operation and for deeper depths the operation may have to be repeated. Hence, normally there is a steep rise in cost of blasting for rock thickness in excess of three metres.

The criteria for planning of drilling and blasting operations will primarily depend upon the –

- Type of dredger and the fragmentation required
- Type of rock
- Physical properties of rock such as porosity, layering, hardness etc. and
- Available time to complete the work

8. DREDGING OF PRE TREATED ROCK

The dredgers normally used for clearance of pre-treated rock are –

8.1 Grab Dredger

The grab dredger is the most common dredger used for dredging of pre-treated rock. The fragmentation required is about 400 mm or less. Grabs are available in different types and sizes. Sizes of grab vary from 1 to 25 cum.

8.2 Back Hoe Dredger

The fragmentation required is about 300 mm or less. The size of rock bucket vary from 1 to 6 cum and is capable of dredging up to a depth of 18 m.

8.3 Bucket Dredger

The fragmentation required is about 600 mm. or less. The size of bucket varies from 100 to 1000 liters. The maximum dredging depth is limited to about 35 m.

8.4 Dipper Dredger

The fragmentation required is about 750 mm or less. Bucket capacity vary from 1 to 9 cum. with a dredging depth varying from 6 to 12 m. However, dipper with a bucket size of 12 cum capable of dredging up to 20 m are in use.

8.5 Cutter Suction Dredger

The fragmentation required is about 300 mm or less. The size of the dredger is specified by the discharge pipe diameter and varies from 150 to 1200 mm. The maximum dredging depth is about 30 m. Specially designed cutter suction dredger with two or three times the normal horse power on the cutter with a heavy ladder and specialised anchoring system are in use for direct dredging of rock. These powerful cutter suction dredgers have about 6000 HP on the cutter and about 22000 total installed Horse Power.

8.6 Trailing Suction Hopper Dredger

The application of trailer suction dredger for dredging blasted rock is limited due to the higher fragmentation of the order of less than 300 mm. required for efficient dredging. However fragmentation in the range of 50 to 100 mm is ideal for economical dredging. Hopper capacity of 500 to 3500 cum are common but dredgers with hopper capacity of 35000 cum are being built, capable of dredging up to 35 m. depth

9. ENVIRONMENTAL EFFECTS

The principal environmental effects due to blasting are

- Vibration
- Water shock and
- Flying rock

9.1 Vibrations

When explosives are detonated, energy is transmitted from one particle to another as elastic wave, some of which travel through the ground and others through water. Based on the statistical data various criteria such as amplitude, acceleration, kinetic energy and particle velocity have been suggested by various codes etc. However, amplitude and particle velocity appear to be the most reliable criteria. Various types of instruments are now available for measuring the ground vibrations.

9.2 Amplitude

Amplitude of vibrations which is half of the total displacement recommended by ** Mr. Morris is a reasonably reliable criterion for estimating damage. He suggested various levels of displacement which should not be exceeded depending on the type of structure under consideration. This criterion however does not consider the frequency which is an important factor, as the damage potential increases in the higher frequency range.

9.3 Peak Particle Velocity

The peak particle velocity of a vibration is now accepted as the most reliable criterion for assessing the damage potential for vibrations. This factor takes frequency into account and is a fairly accurate indication of the 'Permissible' value of the ground movement. The studies carried out in America, Sweden and in the U.K. have confirmed that a particle velocity of 50 mm/sec. represents a safe level of vibration for non residential structures.

9.4 Water Shock

As the detonation shock wave reaches the rock surface a pressure front is created in the water. Surface blasting creates more pressure than drill hole blasting.

The effect could be reduced by the use of 'Air Bubble Curtain'. The system consists of a perforated pipe laid on the sea bed through which compressed air is pumped. The air bubbles rising from the tube forms a curtain between the working area and the area to be protected. The curtain absorbs and remits the shock wave energy depending upon the quantum of air. Reduction up to 60% have been reported.

9.5 Flying Rocks

Are not a problem in water depths exceeding 3 m and wire rope nets can be used to reduce the damage due to flying rock pieces in shallow depths.

10. CONCLUSION

- Rocks with compressive strength in excess of 40 MPa will normally require pre-treatment before dredging.
- The cost of underwater blasting and dredging fluctuates within wide limits due to many variables involved in the process and may vary from Rs.500 to Rs.3,000 per cum.
- The most economical pay depth is in the region of 2.5 to 3 m and the costs increasing disproportionately beyond the 3 m depth.
- Cost of explosives and detonator account for about 10% of the total cost of blasting.
- Drilling and charging of holes cover nearly 75% of the total cost of blasting.
- Spacing of drill holes varies from 1 to 2 m depending upon the dia. of hole.
- Sophisticated survey system is an essential prerequisite for the precise positioning of the drill holes and also safety of operations.
- Blast ratios for under water drilling and blasting vary from 0.6 to 3 kg/cum depending upon the type of rock and required fragmentation.
- Ground vibration could be minimised by the use of delay detonators.
- Fragmentation can be improved by the use of delay detonators.
- Water shocks could be minimised by the use of 'Air Bubble Curtain'.

REFERENCES

[1] ABRAHAMS J.L., "Underwater Drilling and Blasting", *Proceedings of World Dredging Conference*, WODCON, 1968

[2] ABRAHAMS, J.L., "Underwater Drilling and Blasting for Rock Dredging", *Journal of Institution of Civil Engineers*, London,1974

[3] BRAY R.N. et. al.– Dredging, *"A Hand Book for Engineers"*, Edward Arnold, 1997

[4] B.S. 6349. Part 5, "Code of Practice for Marine Structures", *Recommendations for the Dredging of Waterways and for Land Reclamation*, 1991

[5] CLAUDY BROWN AND CHARLES GRENDY, "Shaped Charges Blasting Techniques Simplified, Speeds Dredging Operations", *World Dredging and Marine Construction*, 1973

[6] ELL WERENSKIOLD, K.J., "Rock Dredging Described in Norway", *World Dredging and Marine Construction*, 1975

[7] GUNNAR RYDE BRANT, "Modern Rock Drilling and Blasting Techniques for Underwater Excavations" Proceedings of WODCON, 1968

[8] HIGNETT, H.J. & BAUKS, D.C., "Current Issues in Rock Dredging" *Proceedings of the Conference Dredging 1984*, American Society of Civil Engineers, Newyork, November, 1984

[9] Indian Explosives Ltd., Rock Blasting.

[10] KAMATH, M.M., Under Water Blasting & Dredging, *Seminar on Hydrography in Exclusive Economic Zones Demarcation and Survey of its Wealth Potential,* Calcutta Port Trust. 1981

[11] KAMATH, M.M., "Capital Dredging at New Mangalore Port", *National Seminar on Dredging Indian Institute of Port Management*, Calcutta, 1986

[12] KAMATH, M.M., "Soil Investigation for Dredging Works", *First National Conference on Dock & Harbour Engineering*, Indian Institute of Technology, Bombay, 1985

[13] KAMATH, M.M., "Underwater Blasting & Dredging – State of the Art Report", Third National Conference on Dock & Harbour Engineering, Karnataka Regional Engineering College, Surathkal, 1989

[14] KAMATH, M.M. et.al., "Execution of Underwater Blasting Projects", A Case Study – *National Seminar on Drilling and Blasting for the Next Millennium*, Bhuvaneshwar, 2000

[15] KAMATH, M.M., "Underwater Blasting and Dredging", *Port Technology International London*, Twelfth Edition, 2000

[16] KAMATH, M.M., "Capital Dredging at New Mangalore Port", *Terra et Aqua IADC*, Netherlands, 2001

[17] KAMATH, M.M., "Underwater Blasting for Dredging in Ports" A Case Study of a Major Port in Southern India, Technology Exchange Programme on Safe and Efficient Handling of Explosives, *Karnataka Regional Engineering College*, Surathkal, 2001

[18] KAMATH, M.M., "Underwater Drilling, Blasting and Dredging at New Mangalore Port" National Seminar Cum Workshop on Application of Advanced Blasting Techniques for *Rock Dredging* – Tuticorin Port Trust, Tuticorin., 2003

[19] MORRISON, W.R., "Under Water Drilling and Blasting History, Equipment and Methods", World Dredging and Marine Construction, 1974.

[20] MORRISON, W.R., "Drilling and Blasting Techniques", Dock and Harbour Authority, 1973

[21] Noble Explosives Ltd., U.K., Under Water Blasting, 1973

[22] Paavo Kolari, "Rock Dredging" – Paper No. 1516. Haka U.K. Ltd.,1991

[23] PASTOORS, W.B.C., "Marine Trenching", *International Dredging & Port Construction*. 1975

[24] Peter Francis & Gorden Raynor, "Explosive Engineering" *Journal of Institute of Explosives Engineering*, December, 1993

[25] PIANC, Classification of Soils and Rocks to be Dredged - Report a Working Group of Permanent Technical Committee II. Supplement to Bulletin No.47, 1984

[26] Proceedings of Short Term Course on Small Harbour Engineering, I.I.T., Bombay, 1980.

[27] RUNE GUSTAFFON, Sweedish Blasting Techniques, SPI Gothenburg, Sweeden, 1973

[28] Roberts R.D.G., "Blasting for Underwater Rock Excavation", Technical Paper No.4, *Explosives Engineering Hand Book*,1989

[29] ROCLOFSE, F. & BRAY, R.N., Drilling and Blasting of Submarine Rock in Relation to Dredging, Proceedings of WODCON, 1968

[30] VOLBEDA, I.R.J.M., Where Direct Dredging Stops, Proceedings of International Course on Modern Dredging, 1971

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

DREDGING CONTRACTS

M.M. KAMATH*

1. INTRODUCTION

In most of the port development projects, the cost of capital dredging covers about 40 to 50% of the total developmental cost. In the case of maintenance dredging, in most of the ports, the cost of maintenance dredging, is about 25% of the total operational expenses. In spite of the above, necessary importance for drawing up appropriate specifications and framing a proper contract is lost sight of, resulting in unsatisfactory completion of the work and claims etc. from the contractor. It is advisable to have a good consulting engineer with a dredging background so as to facilitate drafting of a contract document to provide good understanding between the Client the and the Contractor.

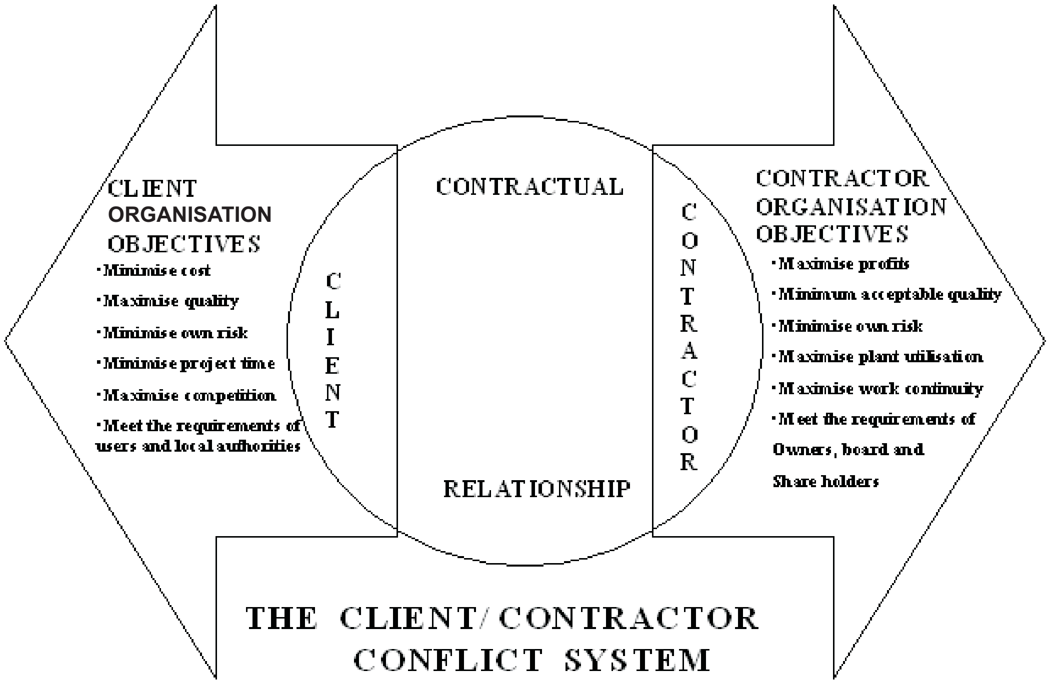
The dredging work involves application of very sophisticated techniques and use of capital intensive equipments. The work has normally to be carried out in a high risk environment. Hence, a comprehensive dredging contract has to be evolved to share the risk evenly between the client and the contractor. A contract is an agreement which describes the work to be undertaken and the method by which payment will be made. It should be the mutually agreed means by which the client and contractor optimize the achievement of their objectives. Its primary purpose is to resolve the conflict between the client and the contractor. Such conflicts arises naturally from their respective objectives as shown in the following Figure.

In view of the divergent objectives of the client and contractor, the minimization of the cost versus the maximization of the profit, a type of contract to suit the work must be evolved so that the risk is distributed evenly between both the parties. The following Figure indicates the relationship between Risk vs Price.

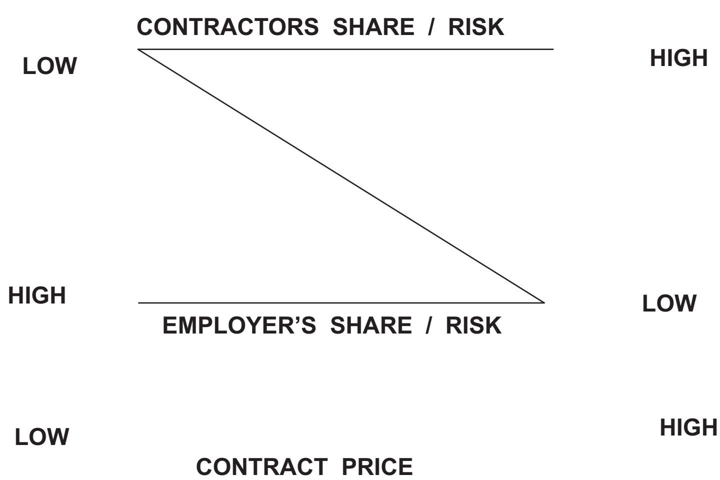
2. RISK ELEMENTS IN DREDGING CONTRACT

The Risk elements in dredging contracts are quite complex and inter-related. The most important among them are

Shree Matha, Mizar Ramakrishna Pai Compound, Mannagudda, Mangalore, India 575 003
E-mail: mmkamath40@yahoo.com



THE MORE THE RISK FOR THE CONTRACTOR, THE HIGHER THE CONTRACT PRICE



- Macro Economic
    - Sudden swings in demand
    - Currency fluctuations
    - Cost inflation mainly due to fuel

- Financial
    - Idle cost of equipment
    - Infructious marketing and site research expenses
    - Mobilization and transportation
    - Bankruptcy of the employer
- Circumstantial
    - Insufficient site date
    - Incorrect soil data
    - Breakdown, collisions and war
    - Inadequate design
    - Siltation
- Institutional
    - Protectionist legislation
    - Competition from public authorities
    - Environmental restrictions
    - Political instability
    - Tax changes
    - Rules & regulations
    - Local labour

3. METHODS OF BIDDING

The following methods are generally used in the competitive bidding process.

- Single Stage – Single Envelope
- Single Stage – Two Envelope
- Two Stage – Two Envelope
- Prequalification

4. TYPES OF CONTRACTS

The type of a particular contract is selected based on the

- Magnitude,
- Nature
- Special design needs
- Complexity of the work and
- Owner's preference

The common types of contracts are

- Lump sum
- Item rate

- Cost plus percentage
- Cost plus fixed fee
- Turnkey

5. BIDDING DOCUMENTS

In the United Kingdom, dredging works are commonly carried out under the ICE Conditions of Contract for Works of Civil Engineering Constructions (5^{th} Edition 1976). These contract conditions have been drafted in consultation with the Institution of Civil Engineers, the Federation of Civil Engineering Contractors and the Association of Consulting Engineers. However, in other parts of Europe, the FIDIC (Federation Internationale Des Inqenieurs – Conseils) Conditions of Contract for Works of Civil Engineering Construction. (Fourth Edition 1987 – Reprinted 1988 with editorial amendments and Reprinted 1992 with further amendments) which are similar to ICE Conditions of Contract, are commonly used for International Dredging Contracts.

The clauses of general application have been grouped together and are referred to as "Part I – General Conditions". They have been printed in a form which will facilitate their inclusion as printed, in the contract document normally prepared. The Part – I General Conditions, has 72 Clauses with forms of Tender and Agreement. The General Conditions are linked with the "Condition of Particular Application", referred to as Part II, by the corresponding numbers of the clauses, so that Part I & II together, comprise the Conditions governing the rights and obligations of the parties. Part II must be specially drafted to suit each individual contract.

The FIDIC during 1999 has published first edition of four new standard forms of contract as under:

- Conditions of Contract for Construction
- Conditions of Contract for Plant and Design – Build
- Conditions of Contract for EPC / Turnkey Projects
- Short Form of Contract

The "Conditions of Contract for Construction" are recommended for building or engineering works designed by the employer. This document mainly deals with engineering works but not dredging and reclamation works. However, this document is being used with suitable modifications to suit dredging and reclamation works.

However, FIDIC has recently developed a form of contract for "Dredging and Reclamation Works" First Edition 2001, in close collaboration with the International Association of Dredging Companies (IADC), which is recommended for dredging and reclamation work. There are fifteen clauses in the General Conditions. The General Conditions are expected to cover the majority of contracts. Nevertheless, users will be able to introduce Particular Conditions if they wish to cater for special cases or circumstances. The General Conditions and the Particular Conditions will together comprise the conditions governing the rights and obligations of the parties. However, this document is yet to become popular with the dredging industry.

The Government of India, have approved the Standard Bidding Documents (SBD) for the projects funded by the World Bank, Asian Development Bank etc. These documents are basically based on FIDIC (1987) Fourth Edition, duly modified for individual works. In fact, World Bank and ADB have also drafted the Part II – Conditions of Particular Applications, to suit their requirements. Recently, Govt. of India has drafted a standard bidding document based on FIDIC document "Dredging and Reclamation Works" First Edition 2001. However, based on this document no tenders have been finalized so far.

6. STRUCTURE OF BIDDING DOCUMENT

The Bidding Documents are normally structured as under:

- Invitation to Bid
- Instructions to Bidders
- Part – I General Conditions of Contract
- Part – II Conditions of Particular Application
- Technical Specifications
- Bill of Quantities
- Drawings
- Form of Bid including Appendix
- Form of Agreement
- Contract Forms and
- Schedules

6.1 Invitation to Bid

The first requirement in a Bidding Process is to invite the prospective Bidders to Bid.

6.2 Instructions to Bidders

Instructions to Bidders is a comprehensive document. This section normally contains all that a bidder has to know about the bidding procedure and the method of preparation of bid. The important among them are

- Scope of work
- Construction Period
- Language of the Bid
- Bid Form
- Bid currency
- Bidder's Qualification
    - Financial
    - Technical
- Bid Security
- Clarification of Bids

- Deadline for submission of Bids
- Withdrawal of Bids
- Award of Contracts
- Signing of Agreement
- Performance Security

6.3 Part – I General Conditions of Contract

As indicated earlier, Part –I – General Conditions of Contract of any standard bidding document are adopted as General Condition of Contract. These are normally printed in a form which will facilitate their inclusion as printed, in the contract document.

6.4 Part – II Conditions of Particular Application

The General Conditions are linked with the Conditions of Particular Application, referred to as Part –II, by the corresponding numbering of the clauses, so that, Part –I and Part –II together comprise the conditions governing the rights and obligations of the parties. This is more relevant in the case of dredging contracts.

6.5 Technical Specifications

The technical specifications for the work should define the work required in terms of dimensions, tolerances and any restrictions to the methods employed. The technical specifications generally to be covered in the dredging and reclamation works are:-

- Description of the work
- Material to be dredged
- Method/sequence of dredging
- Method and placement of soil disposal
- Method of measurement
- Dimensions of dredged areas
- Side slopes
- Tolerances, both horizontal and vertical
- Over dredging restrictions if any
- Natural siltation
- Siltation resulting from the contractor's operations
- Measurement of tide levels
- Setting out and position control
- Method and density of surveys
- Dredging in the vicinity of structures
- Records of dumping
- Environmental Regulation

6.6 Bill of Quantities

The objectives of Bill of Quantities are

- To provide sufficient information on quantities of works to be performed to enable bids to be prepared efficiently and accurately , and
- When a contract has been entered into, to provide a priced Bill of Quantities for use in the periodic valuation of works executed.

The Bill of Quantities should be divided generally into four sections namely:

- Preamble
- Work items

The work items are grouped in to part bills depending upon the nature of work namely

- Bill No.1 – General Items
- Bill No. 2 - Dredging and offshore dumping
- Bill No.3 – Dredging and Reclamation
- Bill No.4 – Dredging of Rock etc.
    - Day work schedule and
    - Summary

6.7 Drawings

The drawings define the nature of work and the dimensions, dredging areas, reclamation areas if any, spoil disposal grounds etc. The drawings, even if not fully developed, must show sufficient details to enable bidders to understand the type and complexity of the work involved and to price the Bill of Quantities.

6.8 Form of Bid including Appendix

The Form of Bid is addressed to the Employer, singed by the Bidder, offering to carry out the work, including remedying of any defects therein, in conformity with the Conditions of Contract, Specifications, Drawings, Bill of Quantities and Addenda if any for the specified sum noted therein.

The Appendix summarises the various clauses of the Condition of Contract and the Bidders should fill in, appropriate blank spaces specified in the document. The bidders are required to sign, each page of the Appendix to Bid.

6.9 Form of Agreement

The Form of Agreement is the document entered in to between the employer and the contractor for the successful completion of the work.

6.10 Contract Forms

These normally include

- Form for Bid Security (Bank Guarantee/ Bond)

- Performance Bank Guarantee
- Bank Guarantee for advance payment for works etc.

6.11 Schedules

The schedules incorporated in bid document, vary from work to work and also depends upon the employer /borrower. The schedules normally included in the bid document are

- Composition of Applicant's Organization
- Details of Joint Venture's if any
- General Experience Record
- Details of Similar Experience
- Current Contract Commitments / Works – in –progress
- Declaration of Sub Contractors
- Proposed Site Organization
- Details of Supervisory / Technical staff
- Schedule of Plant and Equipment
- Financial Capabilities
- Litigation History

7. STANDARDS FOR DREDGING WORKS

There are no Indian standards relating to dredging.

B.S.6349: Part -5 : 1991 British Standard – Maritime Structures - Code of Practice for Dredging and Land Reclamation, is the standard for dredging and reclamation works.

EAU 1996 – Recommendation of the Committee for Water Front Structures issued by the Society for Harbour Engineering and the German Society for Soil Mechanics and Foundation Engineering, gives certain recommendations regarding dredging in front of water front structures.

The Rotterdam Public Works Engineering Department, Port of Rotterdam, the Netherlands Association of Dredging, Shore and Bank Protection Contractors (VBKO), International Association of Dredging Companies, have published in March 2001, a useful guide on 'Construction and Survey Accuracies for the Execution of Dredging and Stone Dumping Works'. Outlining the methodology, survey accuracy etc.

The Permanent International Association of Navigational Congresses (PIANC) has brought out a good publication titled – "Site Investigation Requirements for Dredging Works- Report of Working Group No. 23 of the Permanent Technical Committee II (2000) – Supplement to Bulletin No. 103", which covers the various aspects of Site Investigation relating to dredging works.

8. INFORMATION TO BE SUPPLIED

Adequate and precise information concerning the work under consideration is very essential, as otherwise the contractor may anticipate conditions which are more favourable than those actually encountered, leading into a claim against the employer for additional costs involved. Conversely, the contractor may quote a price which is unnecessarily high. Hence, the need for precise and detailed information relating to dredging works. The important among them are

8.1 Soil Information

The characteristic of the soil to be dredged is one of the most important factors which will determine the cost of dredging. The cost of dredging for different types of soil varies within wide limits, sometimes as high as 1:25 depending upon the type of soil and the dredger as shown in Table 1. The International Association of Navigational Permanent Congresses (PIANC) in 1972 published a report on the 'Classification of Soils for Dredging Purposes'. The classification as gained general acceptance by the majority of the contractors involved in dredging. The working group of PIANC (Permanent Technical Committee – II) has revised the classification and has published as a supplement to PIANC Bulletin No.47 – 1984. The soil information and the tests to be carried out are detailed in the above document.

Table 1. *Comparative costs of various types of dredgers vs materials*

Type of Dredger *Cost Factor*	*Cutter Suction* *Material*	*Trailing Suction*	*Bucket*
1.00	Fine Silt	Sand	Soft Clay
1.30 – 2.00	Sand	Fine Silt	Sand
2.30 – 2.70	Soft Clay	Soft Clay	Medium Gravel
2.70 – 3.50	Medium Clay Medium Gravel	Medium Gravel	Coarse Gravel Stiff Clay
3.50 – 4.00	Hard Clay Chalk	Coarse Gravel	Chalk Hard Clay
6.00 – 10.00	Soft Rock	Stiff Clay	Soft Rock
15.00 – 25.00	Hard Rock (with blasting)	Hard Rock (with blasting)	Hard Rock (with blasting)

Source: Anthony Bates – Dredging + Port Construction (October 1978)
Soil Investigations for Dredging Works.

8.2 Bathymetry

The water depth will influence the size and type of the dredging plant which can be deployed for the job. It is important therefore to supply adequate information concerning bed levels not only within the area to be dredged but within any areas over which the dredging plant must travel in order to gain access to the dredging or to the disposal areas. Tidal information and the relationship between the local datums should be clearly specified.

8.3 Meteorological Information

All dredging plants will be effected by the meteorological parameters.

The meteorological parameters which have important bearing on dredging process are

- Wind
- Cyclonic storms and depressions
- Rainfall
- Air temperature
- Relative humidity and
- Visibility

8.4 Oceanographic and Hydraulic Information

The Oceanographic information cover a wide range of natural phenomenon and the important inputs required for planning of dredging works are

- Waves
- Tides
- Currents
- Sea water parameters

8.5 Shipping Movements

The frequency of shipping movement within the operational area may affect the contractor's dredging operations. The average number of movements per day and restrictions such as prohibition of the passing in channels of certain types of vessels should be clearly specified. Where the frequency and duration of delay to the contractor cannot be reasonably quantified, then, provision for payment of suspension of work should be made in the contract.

8.6 General Information

Normally, a relatively short time is allowed to the contractors tendering for the work who will have to familiarize with the area of operation and the facilities available. Hence, it could be preferable to list out important facilities such as bunkering, water supply, towing, diving, communication, power and telephone connection and any other facilities which are likely to be of interest to tenderers in assessing the logistics and the costs of the work under consideration.

9. ENVIRONMENTAL IMPACT OF DREDGING

The increasing concern with regard to the environmental impact of dredging in the marine environment, specially the contaminated/polluted materials has led to the imposition of restrictions etc. The choice of appropriate dredging technology and selection of dredged material disposal sites is a matter of concern for the planners & engineers.

The World Bank which is associated with number of dredging projects has recommended certain environmental guidelines in the formulation and execution of the projects. There are

- To determine whether the materials to be dredged are contaminated and if so the extent to which they might cause pollution.
- To carry out a detailed characterization of all of the sediments within the port limits and to depths to which they are likely to be disturbed by the dredging process.

- To select a suitable dredging equipment to be employed for each zone.
- To select the method of disposal, including transportation.
- To select the disposal site, including studies & long term monitoring.

The environmental impacts of dredging and its mitigative measures have to be suitably structured in the contract document.

10. RECOMMENDATIONS OF WORKSHOP ON CONTRACT MANAGEMENT FOR DREDGING AND MARITIME CONSTRUCTION

Successful maritime construction projects require skilled contractors, consultants and clients, bound by a fair contract, who are willing and able to work together to resolve technical and management difficulties as they arise. Keeping this in view a workshop was organized on "Contract Management for Dredging and Maritime Construction" jointly by

- International Association of Dredging Companies (IADC)
- Central Dredging Association (CEDA)
- Institution of Civil Engineers (ICE) at Hilton Docklands, London, UK on 12-13 October 2006.

Number of topics on pre-tender information contract management etc. were discussed.

The summary of the recommendations topic wise is presented in Annexure A.

11. CONCLUSION

- In most of the port development projects, the cost of capital dredging covers about 40 to 50% of the total developmental cost. In the case of maintenance dredging, in most of the ports, the cost of maintenance dredging, is about 25% of the total operational expenses. In spite of the above, necessary importance for drawing up appropriate specifications and framing a proper contract is lost sight of resulting in unsatisfactory completion of the work and claims etc. from the contractor.
- The dredging work involves application of very sophisticated techniques and use of capital intensive equipments. The work has normally to be carried out in high risk environment. Hence, a comprehensive dredging contract has to be evolved to share the risk evenly between the client and the contractor.
- The following four methods of bidding are normally adopted in dredging contracts.
    - Single Stage – Single Envelope
    - Single Stage – Two Envelope
    - Two Stage – Two Envelope
    - Prequalification
- The common types of dredging contracts are
    - Lump sum
    - Item rate

- Cost Plus percentage
- Cost plus fixed fee
- Turnkey

- Conditions of Contract for Works of Civil Engineering Construction. (Fourth Edition 1987 – Reprinted 1988 with editorial amendments and Reprinted 1992 with further amendments) which are similar to ICE Conditions of Contract, are commonly used for International Dredging Contracts.

REFERENCES

[1] Anthony, D. Bates, "Soil Investigations for Dredging Works", *Dredging + Port Construction*, Intec Press, England, October 1978

[2] Bray, R.N., "Dredging", *A Hand Book for Engineers, Second Edition Edward Arnold*, London, 1997

[3] B.S.6349 : Part -5 : 1991 British Standard, Maritime Structures –Part-5, Code of Practice for Dredging and Land Reclamation

[4] EAU 1996, Recommendation of the Committee for Water Front Structures issued by the Society for Harbour Engineering and the German Society for Soil Mechanics and Foundation Engineering.

[5] ESCAP, Seminar on Dredging Contract, Bangkok, September 1988

[6] Gupta M.P., Competitive Bidding, Procedures and Practices – Tata McGrawHill Publishing Company Ltd, New Delhi, 1998

[7] Kamath M.M., Soil Investigation for Dredging Works, First National Conference on Dock & Harbour Engineering, IIT, Bombay, December 1985

[8] Kamath M.M, "Site Investigation for Dredging Works", Dredging for Better Environment, Dredging Symposium 199, Eastern Dredging Association (India), Hyderabad, June 1995

[9] Kamath M.M., Dispute Resolution in Construction Contracts, Workshop on Problems in Construction Engineering Project Management, MAHE, Manipal, October 2004

[10] Kamath M.M., Dredging Contracts, Workshop on Global Tenders, National Institute on Port Management, Madras, July 1998

[11] Kamath M.M., "Site Investigation for Dredging Works", Dredging Seminar 2005, Dredging Corporation of India, Mumbai, October 2005

[12] Permanent International Association of Navigation Congresses. Report of a Working Group of the Permanent Technical Committee –II (Supp to PIANC Bulletin No.47), 1984

[13] PIANC on Site Investigation Requirements for Dredging Works, Report of Working Group No. 23 of the Permanent Technical Committee II, Supplement to Bulletin No. 103, 2000

[14] Rotterdam Public Works Engineering Department, Port of Rotterdam Construction and survey accuracies for the execution of Dredging and Stone Dumping Works, March 2001

[15] World Bank, Annual Review of Project Performance Audit Results, Washington. Feb. 1978

ANNEXURE - A

EXTRACT FROM
REPORT ON WORKSHOP
CONTRACT MANAGEMENT FOR
DREDGING AND MARITIME CONSTRUCTION

Conducted By
International Association of Dredging Companies (IADC)
Central Dredging Association (CEDA)
Institution of Civil Engineers (ICE)

Hilton Docklands, London, UK
12-13 October 2006

INTRODUCTION

The following topics were discussed and the summary of recommendations of each topic is extracted and presented below.

- Pre-tender information
- Environmental issues during preparation stage
- The balance between technical and functional requirements
- Choice of form of contract
- Tender procedures
- Project finance
- Liability issues and
- Dispute settlement

A.1 Pre-Tender Information

- Consultants should have a bigger role in the pre-tender process. Consultants' qualifications are key and the consultant in charge should be experienced in the dredging process on every level.
- Unfortunately, one cannot always trust that people have honest intentions when starting tender procedures for dredging projects.
- Perhaps it is not really necessary for the client to spend great sums of money to study the site and then have the contractor do the same. The contractor could instead send a representative.
- Although ground investigation data always needs to be interpreted, from the client's view, certainty of price should be achieved regardless of the conditions at site.
- There was consensus that time allowed for tendering should be feasible and risk analysis should be part of the site survey report.

A.2 Environmental Issues during Preparation Stage

- Dialogue and lobbying needs to be intensified amongst contractors, consultants, clients, legislators and the general public.
- Dialogue with legislators in particular must be strengthened. Such communication should aim to help clients navigate in a fast-changing field and inform legislators on the technical feasibility of legislative measures. In fact, contractors should rethink their role in the environmental debate and not leave all the initiative with legislators.
- Ways in which contractors could affect the system could include a voluntary or enforced code of conduct or practise, specifying that tenders without EIAs would be unacceptable.
- In general, knowledge about the environmental impact of dredging and maritime construction must be deepened. Existing knowledge must be better harnessed and utilised through more open sharing of expertise amongst clients, consultants and contractors.

A.3 The Balance between Technical and Functional Requirements

- A correct balance between technical and functional requirements in a project should be found, and a more constructive working relationship amongst clients, consultants and contractors should be promoted. A clear definition of rules and objectives from the start helps avoid conflicts later on. The process of tendering should be seen as a dialogue.
- Clients are an extremely varied group of people. Contractors may find some clients in certain countries do not feel very committed to their projects and are happy to hire a firm of consultants and hand over all responsibility to them. Inexperienced clients will often include overly narrow technical requirements in their project specifications instead of focusing on functional rather than technical requirements in project specifications. Giving contractors more freedom to decide which technical solutions and equipment to use should make for better proposals and lower prices.
- Overall success depends upon forging a real partnership amongst client, consulting engineer and contractor. Continuity is very important so clients should make sure they employ the same consultants from early stages through to implementation. Ideally all three partners, including pre-qualified contractors, should be involved in the project process from the outset, granted that in some countries this is not allowed.
- Consultants should strive to be as independent as possible and if a project is not technically feasible, they should say so. Clients should aim to pass on their local knowledge such as the availability of local materials or which other projects are underway in the area to contractors.
- Potential challenges for clients are: Prevailing market conditions favour contractors and mean they can pick and choose their projects. This has made them reluctant to take excessive risks. Also, low fees and fierce competition between consultants are making it harder for them to attract and retain good technical personnel.
- All players should be prepared to think out of the box and look for innovative solutions. Clients should concentrate on reaching timely decisions rather than aiming for total perfection if this will mean long delays in the tendering process.

- There is a need for an industry-wide code of conduct. For contractors, this should cover minimum rates to avoid the practice of offering a lower price and then relying on subsequent claims as compensation. The IADC should draw up a set of minimum requirements for clients.

A.4 Choice of Form of Contract

- Alliance contracts (AC) offer many advantages, but are not the sole form of contract for the future. AC can be overkill in case of straightforward jobs. Also, it may not work if the contractor is client for third party with different obligations.
- AC can save time, because of "true cooperation", but it has to be tailor made. It is not a formal form of contract; key elements are trust, a cooperative mentality and a certain mindset.
- In all forms of contract, beware of copying the past and realize that while trust is good, it should be supported by a good and transparent contract. Agree in advance on procedures to avoid claims and legal procedures and include consultants in the process.
- Price and working methods are still the main selection criteria. Presently, environmental requirements are driving up prices. Dilemma: will clients accept higher prices or go to more lenient contractor?
- Managing risk means identifying risks at an early stage; sharing risks if possible; and making clear agreements on risks that can be or should not be shared. Or should all risks go to the contractor? If the client is willing to pay that would be fine, but this is a dilemma for client. They may pay for risks which do not happen and this drives the price up as well.
- Generally speaking, never forget that dredging is always part of larger projects with more parties. It is important to try to control the "unknown" at an early stage by identifying problems and risks in an early stage and by sharing knowledge and information. When there is real cooperation possible, all parties benefit.
- Time and money are the only ways to enforce a contract. In case of complex and extensive projects: establish a board of experts, strive for trust instead of distrust, solve problems in the most efficient way regardless of blame, and share the benefits.

A.5 Tender Procedures

- Registration database systems for pre-qualification already exist in other industries. It would be good to check on their experiences and learn from this.
- Should alternative solutions be evaluated, and if so, how and based on what criteria? Should the client take the initiative in the tender documentation? Can alternative solutions be copyrighted to protect the contractor?

A.6 Project Finance

- Clients often do not start looking for funding early enough in the project cycle. There is as well a growing tendency for clients to ask contractors to get involved in funding, often by asking for funding packages in the tender. Contractors do not see this as part of their core business. Clients should instead do their homework on financing, preferably as part of the feasibility study. In any case, they should have secured funding before the project is put out to tender.

- Securing funding early has benefits for all; it reduces uncertainty for contractors preparing their bids. Contractors are then able to offer clients lower prices.
- If they are not knowledgeable about securing finance, clients should consider using financial consultants the way they use consulting engineers. This is not common practice today and could be a new service to be offered by consultants. Although contractors are not keen on providing long-term financing for projects, short-term modalities such as bridge loans may be possible. However, sharing risk in a dredging project, has financial implications and can result in higher bids. Sharing information can reduce cost, but in a competitive arena, the players are usually reluctant to do this. For instance, clients often do not give enough historical data, while contractors do not want to provide production rates. Contractors often do not have sufficient ground information to identify and manage the risks, yet carrying out their own site investigations may not be feasible. Should clients provide the maximum amount of information on ground conditions to contractors preparing a bid? Some people suggested that clients could create a more level playing field for contractors by hiring independent consultants to carry out site investigations and then making this information available to all interested parties via a data room. Others do not believe this is a realistic suggestion. Hiring the right team of sensible advisors could help all parties to adopt a reasonable attitude towards sharing risk.

A.7 Liability Issues

- It is difficult to make general statements about liability. Many factors are involved, such as ownership (of vessels or equipment or damaged property), local laws, contracts and policy. Consultants are in a vulnerable position as they will often be blamed for failing investigations. In addition, the insurer for the owner of the damaged property or the insurer for the contractor will often lay a claim on the consultant. But do consultants actually have enough coverage? Liability for clients may suggest too much knowledge on their part, but they should be more open although there is the danger of this backfiring: the insurer may retract the policy if client tells all. Sometimes contracts and policies contradict each other so the dilemma is whether to go uninsured or go elsewhere. But is that really a good option? Generally speaking, all eventualities should be covered. However, the differences in size between a contractor and a consulting firm can be substantial. That should be taken into account. On the other hand, contractors are not insurance brokers. When considering insurers, be aware that not all insurers have enough knowledge of the industry and that all dredging projects are more or less unique. A few recommendations included creating more risk awareness by bringing in the right people at an early stage; investigating the possibility of sharing risks and liability; synchronising policies of contractors, consultants and clients to eliminate too much double coverage. Each partner should take some responsibility and the partner with the most knowledge in a certain field should do the assessment. For instance, the risk for the rig or equipment should be for the owner as they can best assess the needed coverage. The goal should be to keep the insurance coverage and liability simple, but be wary of paste and copy policies. Overall there should be better communications between contracting parties and an effort to avoid conflicts of interest. Perhaps the solution is to try to establish an industry framework for insurance coverage.

A.8 Dispute Settlement

- DABs (Dispute Adjudication Boards) are an important step in the right direction but specifications of their modus operandi must be refined in future FIDIC contract models, and the enforcement of their decisions deserves particular attention. DAB enforcement can be supported by way of the NY Convention through ICC arbitration. This construction must be supported by future FIDIC model contracts. The 28-day notice period phrasing in FIDIC contracts needs to be clarified, as should vague phrasing, such as the infamous "conditions reasonably foreseeable by an experienced contractor". A specified and priced bill of unforeseeable conditions may be a helpful requirement to tendering parties. The cost-driven procurement of consultants risks resulting in more disputes and claims. Contract administration must be improved on all sides, and detailed and complete records should be kept in order to be prepared in case a dispute cannot be settled in an amicable way. In the end, prevention is far preferable over any form of dispute settlement and so using all means in the pre-tender phase to reduce potential conflicts is advisable.

Final – \\server\e\book work
3-06-10

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

DREDGING - AN OVERVIEW

M.M. KAMATH*

1. INTRODUCTION

Dredging is an ancient art but a relatively young science. With the advent of industrial revolution which transformed many arts into sciences, dredging was also subjected to great scientific analysis. Today dredging is treated as a science covering not only the design of dredgers but also the dredging methods and their effect on the environment etc.

Dredging is defined as 'the process of disloading, raising handling and transporting mainly soil under water from the layers of the earth'.

Dredging is normally used for one of the following purposes:

- Creation of artificial depths for ports and harbours
- Maintaining the depths in the existing ports and harbours
- To replace unsuitable foundation materials with suitable materials
- Construction of dams, sea walls, dykes etc.
- Construction of artificial islands in sea
- To improve or maintain the discharge or flow capabilities of rivers by deepening natural water depths for the purpose of flood control.
- To provide fill materials for protection and replenishment of beaches
- Mining of aggregates, sand, gravel and minerals such as tin, zinc, lead, gold etc.

2. CLASSIFICATION OF DREDGING

Dredging is mainly classified under the following two heads:

2.1 Capital Dredging

The dredging process when applied to the construction of new harbours, basins, entrance channels, trenches for foundation and pipelines, for deepening the harbour, etc. is classified as capital dredging.

*Shree Matha, Mizar Ramakrishna Pai Compound, Mannagudda, Mangalore, India 575 003
E-mail: mmkamath40@yahoo.com

2.2 Maintenance Dredging

The dredging process when carried out for removing the siltation in the existing harbours, rivers, channels and clearance of siltation in lakes, etc., the same is termed as maintenance dredging.

3. DREDGING PROCESS

The dredging process can be split up into the following four sub systems.

- Pre-treatment
- Excavation
- Transportation and
- Disposal

3.1 Pre-treatment

Pre-treatment consists of treating the ground surface before the excavation process. This is normally carried out by one of the following methods –

- Mechanical and
- Chemical

Both the methods are applied normally for rock. For normal soils, no pre-treatment is required and the material is disintegrated at the time of excavation.

3.1.1 Mechanical method

In the conventional mechanical method, the chisel is mounted on a pontoon and is allowed to fall on the same spot several times till the desired penetration is reached. The chisels normally weigh about 15 to 30 tonnes and have a fall height of about 25 to 30 m.

3.1.2 Chemical method

The pre-treatment by chemical methods involve the use of explosives. Explosives can be used either on the surface or in drilled holes. The efficiency of surface blasting can be improved by the use of ‘shaped charges’.

3.2 Excavation

The excavation process is a combination of two operations. The primary operation involves disintegration of the soil and the secondary operation involves the movement of the soil. The former is performed either mechanically or hydraulically.

3.3 Transportation

Transportation consists of various methods of movement of the dredged materials from the dredging site to the disposal site.

4. DREDGING COST COMPARISON

The dredging industry is highly capital intensive compared to other Civil Engineering industries. The variables costs such as fuel consumables and labour are of a minor proportion compared to the fixed costs of equipment, as can be seen from Table 1.

Table 1. *Comparative costs of dredging vs other civil engineering works*

(In percentage)

Costs	*Dredging*	*Road, Rail,Water works*	*Building*
Variable Cost	35	69	80
Fixed Cost	65	31	20
Investment in Equipment	14.5	5	1

Source: Terra ET Aqua No : 7 – Netherlands Central Bureau of Statistics 1962

From the above it is clear that, the selection of the appropriate fleet configuration for the job in hand is of paramount importance.

5. SOIL CLASSIFICATION

Soil classification is the arrangement of soils into groups that have similar properties. The soils to be dredged can be broadly classified into three main groups.

- Cohesive soils such as clay, silt etc.
- Non Cohesive soils such as sand, gravel etc., and
- Consolidated soils namely, rock varying from soft to hard

Between these main groups lie broad areas of transition.

The Permanent International Association of Navigational Congresses (PIANC) has produced a soil classification, specific to dredging. The Working Group of PIANC (Permanent Technical Committee II) has published a report, as a supplement to PIANC Bulletin No. 47-1984. The classification has gained general acceptance by the majority of Organisations involved in dredging.

6. STANDARDS FOR DREDGING WORKS

There are no Indian standards relating to dredging.

B.S.6349: Part -5 : 1991 British Standard – Maritime Structures - Code of Practice for Dredging and Land Reclamation, is the standard for dredging and reclamation works.

EAU 1996 – Recommendation of the Committee for Water Front Structures issued by the Society for Harbour Engineering and the German Society for Soil Mechanics and Foundation Engineering, gives certain recommendations regarding dredging in front of water front structures.

The Permanent International Association of Navigational Congresses (PIANC) has brought out a good publication titled – "Site Investigation Requirements for Dredging Works- Report of

Working Group No. 23 of the Permanent Technical Committee II (2000) - Supplement to Bulletin No. 103”, which covers the various aspects of Site Investigation.

The World Bank Publication “Environmental Considerations for Port and Harbour Developments – World Bank Technical Paper No. 126”, is a very good reference material on environmental issues of port and harbour developments including dredging.

7. CLASSIFICATION OF DREDGERS

The dredgers are mainly classified into the following four heads:

- Mechanical
- Hydraulic
- Pneumatic and
- Special dredging equipments

7.1 Mechanical Dredgers

Mechanical dredgers use the mechanical power for excavation. The various types of dredgers under the mechanical type are:

- Grab dredger
- Dipper dredger
- Bucket dredger
- Rock breaker and
- Back-hoe dredger

7.1.1 Grab Dredger

The grab dredger is the most common among the mechanical dredgers and a number of types of them are in use. The dredger essentially consists of a crane fitted with the grab mounted on a pontoon or a self-propelled hopper barge. The crane is of a slewing type which lowers and hoists the grab into and out of water. The grab hopper dredger is normally provided with more than one grab crane. Up to four grab cranes can be conveniently installed in a dredger. The grab dredger is an universal tool and is used in almost all types of soil. Unfortunately its limitation is the low output. However, it is specially suitable for working in confined areas such as docks, alongside berths etc. Number of designs of grabs have been developed for different types of soils ranging from loose silt to blasted rock.

Grab dredgers are normally rated by their bucket capabilities. Fig. 1 shows a typical grab dredger. Normal size of grabs vary from one to five cum. Some of the modern grabs have as high a capacity as 200 cum. The output of the grab dredger will depend upon the number of cycles that could be achieved in a particular situation and outputs of 300 to 400 cum per hour are not un-common.

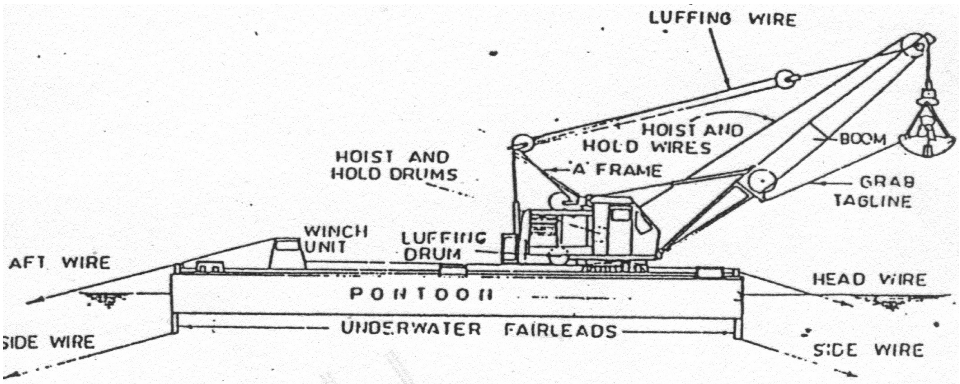


FIG. 1. TYPICAL DETAILS OF A GRAB DREDGER

7.1.2 Dipper Dredger

The dipper dredger is basically a power shovel operating from a pontoon. The bucket is attached to the extremity of a hinged rigid arm and the digging power is supplied by a forward leading hoist wire. Since large horizontal forces have to be applied to the ground by the bucket, it is necessary for the pontoon to have positioning spuds in order that the reaction forces are not transferred to the other wires. It is ideal for dredging of hard material such as blasted rock etc. The normal size of the bucket varies from one to nine cum with the dredging depths varying from 6 to 12 m Dipper dredgers with a capacity of about 6 to 12 cum capable of dredging up to 20 m are in use and the outputs are in the range of about 200 cum per hour. Fig. 2 shows a typical dipper dredger.

7.1.3 Bucket Dredger

The bucket dredger is the oldest type of dredger in existence and still continues to occupy an important place in any dredging fleet. . The most characteristic feature of a bucket dredger is an endless chain consisting of a large number of buckets and links by which they are joined. The chain is suspended from an upper tumbler and is guided and supported by a ladder at the lower extremity of which is a lower tumbler. The bucket chain is driven by a suitable prime mover either an electric or hydraulic motor. As the buckets reach the lower tumbler, the soil is dislodged by the rim and fills the bucket which then commences its journey up the ladder. On reaching the top the material is automatically discharged either into the self-contained hopper or to a chute for delivery into an alongside barge. In order to achieve more or less continuous dredging process, the vessel in swung from side to side. The movement is obtained with the aid of winches installed in the dredger and the anchors. The average size of the buckets are in the range of 200 to 900 litres. Normal dredging depths vary from 5 to 18 m. However, by suitable extension of the ladder, dredging depths up to 30 m can be achieved. The output varies with the speed of the bucket which in turn is related to the horse power of the motor, type of soil etc. and are normally in the range of 250 to 1000 cum. per hour. Fig. 3 shows the details of a bucket dredger.

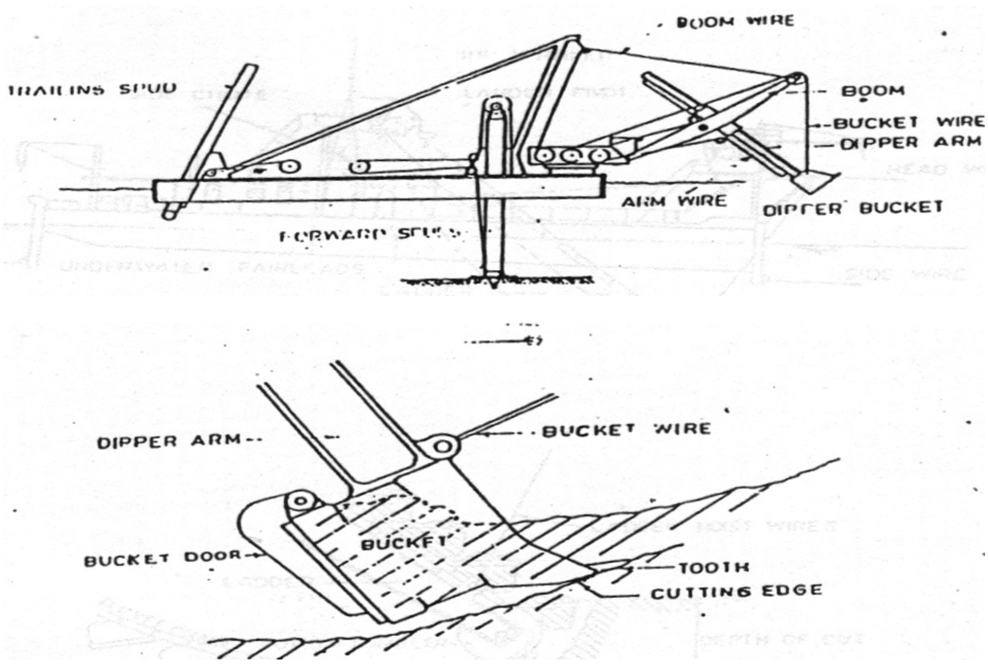


FIG. 2. TYPICAL DETAILS OF DIPPER DREDGER AND THE DIPPER BUCKET

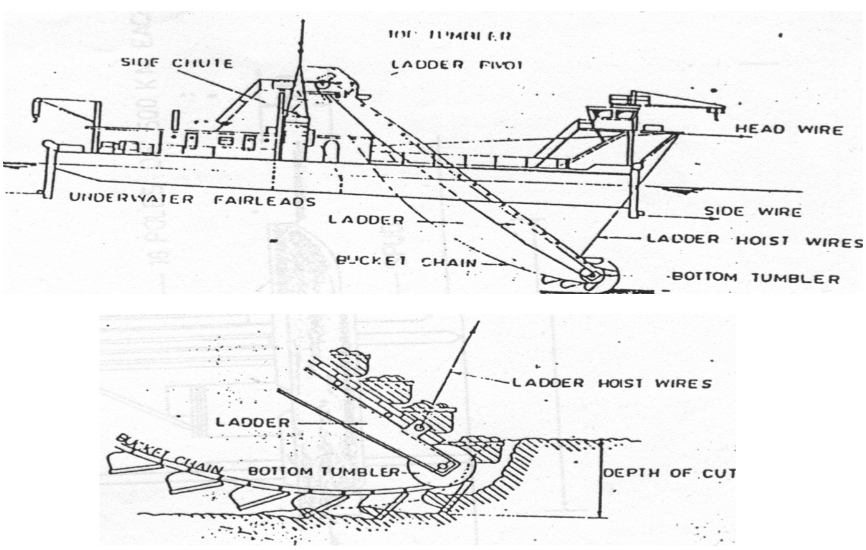


FIG. 3. TYPICAL DETAILS OF A BUCKET DREDGER AND THE DETAILS OF BUCKET

7.1.4 Rock Breaker

The rock breaker consists of a heavy pointed chisel some of them as heavy as 30 tonnes mounted on a suitably designed pontoon. The chisel can be hoisted and dropped vertically on the rock to be broken. The chisel has a heavy cast steel point. Modern version of the rock breakers have pneumatic or hydraulic hammers which break the rock with a frequency of 1.2 to 2 blows per second. With a normal 15 tonne rock chisel and in average type of rock, outputs in the region of 10 cum per hour can be achieved.

Recently a rock breaker was installed on a pontoon of size 60 x 20 4.5 m consisting of sixteen 50 tonne steel spuds arranged in two rows. By a suitable system of pulley and winches, the spuds were raised and dropped from a height of about 5m. This system known as 'RAMMELAAR' was used in the development of the Jubail Industrial port for beaking cap rock covering an area of about 15,000 sqm of one m. overall thickness. The general arrangement of the system is shown in Fig. 4

7.1.5 Back-hoe Dredger

This dredger is basically a back acter land based excavating machine mounted on a pontoon. Back hoes are powered by line pull or direct hydraulic linkage. The outer arm of the back hoe has cutting edges and the teeth are fitted to increase the point pressure on the material to be dug. To resist the high break-out force, the supporting pontoon is firmly held in place normally with three spuds. The two main spuds are mounted at the front or the machine end to provide the holding power. The rear single spud is used to move the dredger over the working area either by tilting or using a spud carriage. This dredger is ideal for excavation of trenches, dredging of blasted rook etc. The bucket capacities vary from 1 to 6 cum and are capable of dredging to a depth of 18 m. The outputs are in the range of 100 to 400 cum per hour. A typical back-hoe dredger is shown in Fig. 5.

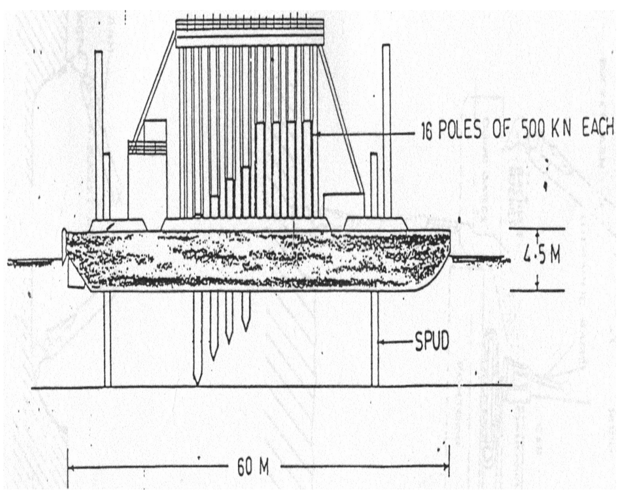


FIG. 4. MECHANICAL ROCK BREAKER

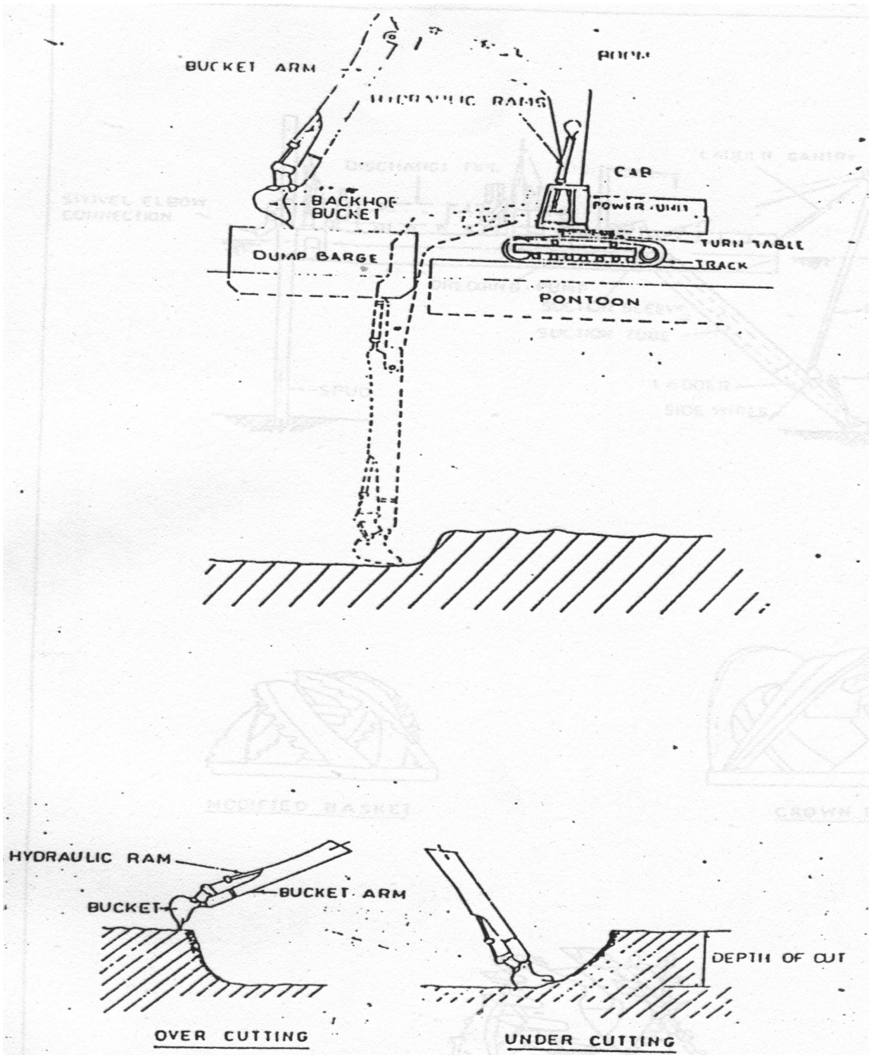


FIG. 5. TYPICAL DETAILS OF A BACK HOE DREDGER AND METHOD OF WORKING

7.2 Hydraulic Dredgers

The hydraulic dredgers employ suction technique in which the soil is drawn up by a centrifugal pump. The various types of hydraulic dredgers are –

7.2.1 Plain Suction Dredger

These dredgers are equipped with a centrifugal pump for raising the mixture of water and soil to deliver the material into the transport system. This is suitable for loose or free flowing materials. The maximum dredging depths are in the region of 60 to 75m.

7.2.2 Cutter Suction Dredger

The cutter suction dredgers made its appearance towards the 19^{th} century. These dredgers are versatile and can deal with a wide variety of materials including clay, silt, sand, gravel and some grades of rock. The dredger mainly consists of a rotating cutter head which is mounted in front of the suction inlet which performs a rotary motion, there by dislodging the soil and depositing in a position from where it can be drawn into the suction pipe. The cutter and the pipe are respectively mounted on a pivoting rugged framed ladder which is attached to the pontoon of the dredger. The free end of the ladder is suspended from a gantry and can be raised or lowered with the aid of a winch. The dredger is anchored by means of spud about which it is swung by manipulating the anchor wire leading from the port and the star board sides. Forward motion is achieved by the shifting position of the spud. Mixture is drawn up through the suction pipe passing through the pump and enters the discharge pipe. This is connected to a floating pipeline, stern of the dredger, which leads to the disposal ground through a shore pipeline. The cutter suction dredgers are ideally suited for land reclamation, hydraulic fill projects, construction of dykes, excavation of new harbours and channels, etc. Specially designed dredgers can handle hard rock upto 500 Mpa without pre-treatment by explosives, etc. Cutter suction dredgers are rated by their discharge pipe and vary in diameter from 150 to 1050 mm. Some of the smaller dredgers are built in sections so that they can easily transported from one site to another. The main limitation of the dredger is the sea condition and it can only work in light to moderate condition of about one m, beyond which the spuds are liable to damage etc. In rough sea conditions, the spuds are substituted by a 'Christmas Tree' arrangement.

The present day cutter suction dredgers have pipe dia of 1200 mm and have pump capacity of about 5000 H.P. and total installed H.P. in the region of 21000 capable of pumping the material to a distances of about 6 km. The dredging depths are in the region of 20-25m. The horse power of normal cutter suction dredgers vary from 100 to 1000. However specially designed dredgers for direct dredging of rock have atleast three to four times the normal power on the cutter. . In case the reclamation areas are located at far away places outside the reach of the pump/pumps installed in the dredger, booster stations are installed in the system to maintain the pressure in the pipeline. There have been instances when the dredged materials have been pumped to a distance as long as 25 km. The modern cutter suction have a computer which provides for the full automatic control of the dredging process. Typical output of the dredgers vary from 200,000 to 500,000 cum / week. Fig. 6 indicates the general arrangement of a cutter suction dredger and also the various types of cutters in use.

7.2.3 Wheel Dredger

The principle employed in the land based wheel excavator has been extended to dredging industry. In the wheel dredger, the dredging wheel replaces the conventional cutter used in the cutter suction dredgers. The plane of rotation is at right angles to that of the cutter, providing smoother dredging process with higher output of solids. The dredging wheel combines the high controllable dredging process of a bucket dredger and an efficient hydraulic transport system of a cutter suction dredger. The dredging wheel essentially consists of two rings connected by the buckets. The bottomless buckets with sloping sides fitted with patented lips have solved the problem of adhesion of soil. The

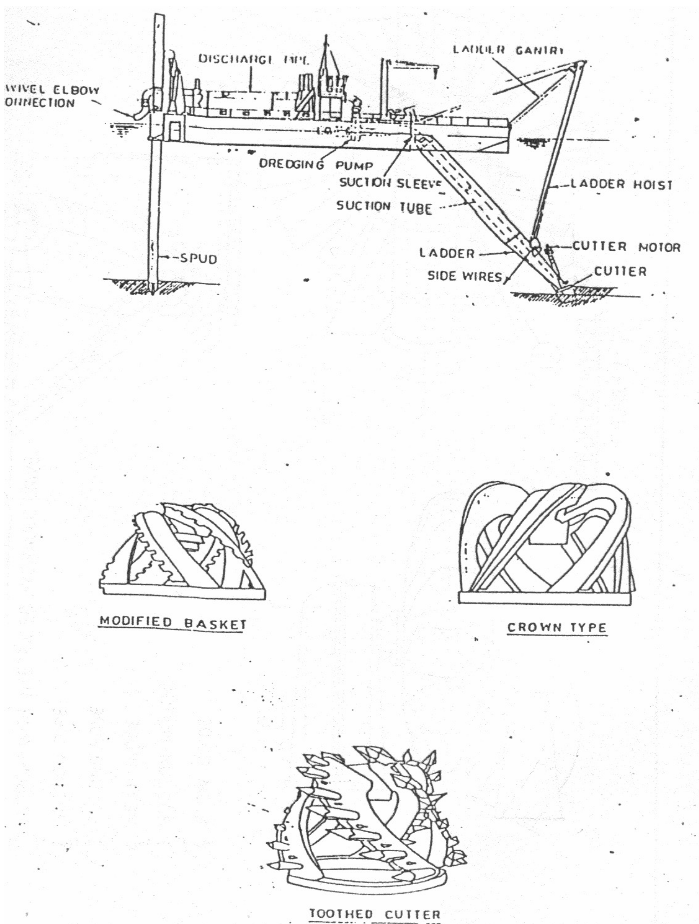


FIG. 6. GENERAL ARRANGEMENTS OF A CUTTER SUCTION DREDGER AND VARIOUS TYPES OF CUTTERS

wheel cuts the soil and transports it to the suction inlet. The wheel is supported at one end and is hydraulically powered. It can be used virtually in all types of soils including sticky clay. The equipment is very well suited for alluvial mining. A mixture concentration of 70% has been reported. The normal dredgers are in the power range of 20-550 kw with outputs of 200-3600 cum /hr. However specially built dredgers have drive powers up to 1100 kw and outputs up to 5000 cum /hr. A typical dredging wheel is shown in Fig. 7.

7.2.4 Barge unloading Dredger

The dredger consists of a suction pipe which draws the dredged mixture from a barge moored along side or to the jetty. The mixture is pumped through a pipeline to the point of reclamation, via a dredge pump installed in the pontoon. The size of the dredger is expressed in terms of the dia of the suction pipe and varies from 250 to 1000 mm.

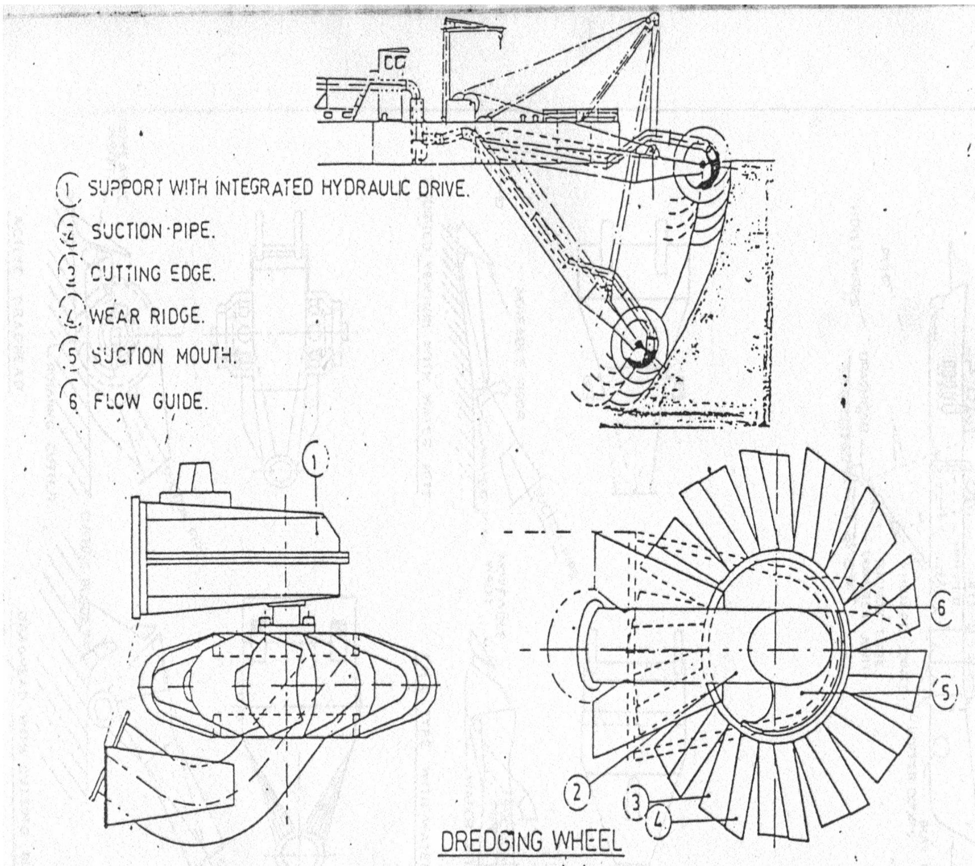


FIG. 7. DREDGING WHEEL

7.2.5 Trailer Suction Dredger

The self-propelled hopper dredger made its appearance in 1875. It was derived from the stationary hopper dredger which was first developed in U.S.A. during 1857. Prior to 1960, there were only a few vessels of this type in service but since 1960, the trailing suction dredger has developed into a powerful dredging tool with wide range of applications. Today trailing suction dredgers of 30,000 cum capacity capable of dredging upto a depth of 35m and possessing a laden speed of 17 knots are in use.

The generally accepted definition of a trailer suction/hopper dredger is a self-propelled, self loading and self-discharging sea going vessel with one or more flexible suction pipes equipped with special suction heads. As the name implies, the material is dislodged and sucked in, as the dredger moves forward with the help of a drag head attached to a suction pipe suspended on the side of the dredger. Different type of drag heads are in use depending upon the type of material to be dredged. The various types of drag head in use are shown in Fig. 8 along with the general details of the dredger. The dredged material is sucked up by the centrifugal pump and the material is discharged

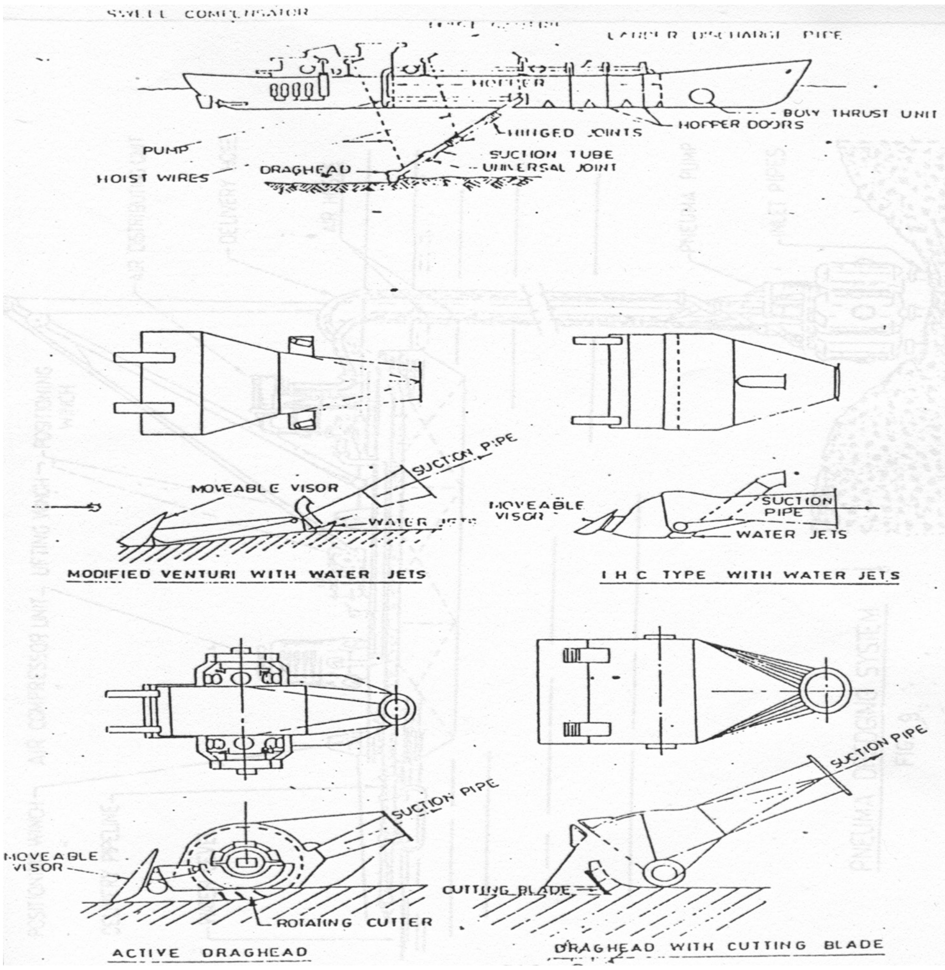


FIG. 8. GENERAL ARRANGEMENTS OF A TRAILER SUCTION HOPPER DREDGER AND VARIOUS TYPE OF DRAGHEADS

into the hopper contained within the dredger. The dredger is suitable for all types of normal soils except for very fine sand and stiff clay.

The hoppers of the dredger are normally provided with a discharge system such as the bottom opening gates, conical valves, lyster valves or sliding doors, etc. However the material can also be used for reclamation by a system of pumps by emptying the material from the hopper and pumping into a land line at a specially designed reclamation berth.

The trailer suction dredgers are rated by their hopper capacities which vary from 300 to 30,000 cum and majority of the dredgers are in the range of 500 to 3,500 cum. The normal dredging depths vary from 10 to 30 m.

7.3 Pneumatic Dredger

The dredger in this category is known as the 'Pneuma System'. The Pneuma system was developed by M/s. Pneuma International S.A., Itlay and has been successfully used in Holland, Japan and Libya.

The system can be operated both from the land and from the traditional barge. It can also be used for long distance transportation.

A typical barge mounted system is shown in Fig. 9.

7.4 Special Dredging Equipments

Recently many dredging equipments have been developed to tackle special problems to meet the requirement of accessibility to site, environmental considerations, etc. The following are the descriptions of some of the equipments in use.

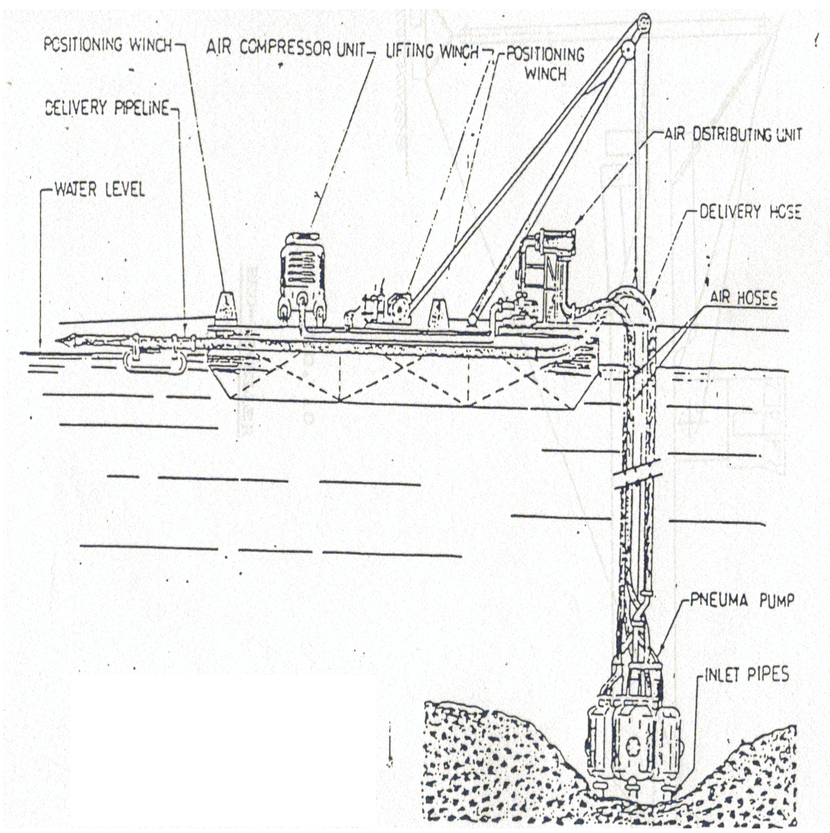


FIG. 9. PNEUMA DREDGING SYSTEM

7.4.1 Hover Dredger

This is a hybrid of a hover craft and a pontoon mounted dredger. In view of its construction, it is amphibious and is designed to negotiable fairly soft soils. The dredging equipment normally consists of either a back-hoe or a grab.

7.4.2 Amphi Dredger

The Amphi dredger was first developed in Holland for working in areas of very low bearing capacity or in places where the work is carried out in land-water interface. The equipment mainly consists of a dredger which can be a grab, back-hoe or cutter suction unit mounted on a super-structure fitted with specially designed wheels. Since the super – structure is mounted on a steel hull, it has its own floatation.

7.4.3 Amphibious Bull Dozer

Japanese have built an Amphibious bull dozer which is probably the first of its kind in the world. The bull dozer can be used for leveling river beds, digging trenches in fairly hard soils etc. The bull dozer in controlled by a direct cable or radio system from the shore. This is ideally suited for working in restricted areas where normally dredging equipments cannot be used. The equipment has been used in depths upto 7 m.

7.4.4 Under Water Bull Dozer

The under water bull has also been developed by the Japanese and can work in depths upto 60 m. It can be controlled by a diver or operated from an interesting ship board control which simulates the movements of the bull dozer. A sonar equipment is fitted to the bull dozer to carry out survey of under water topography.

7.4.5 Toyo Pump

Is a submersible sand pump with a patented built in agitator which pikes up material from rest, agitating it into slurry and pumps it at high densities at long distances. It has number of applications and when used as a dredge pump, it can either be suspended or mounted on a suitable ladder. The power is supplied to the motor through a cable.

7.4.6 Mud Cat

This is a small suction dredger fitted with horizontal angers, which dislodge the material and feed the suction pipe from where it passes through the dredging pump to the pipeline.

7.4.7 Crawl and Roll Cat

Crawl cat is an amphibious dredger with four hydraulic spuds fitted with a crawler at the lower end. The swinging cutter ladder is provided with disc cutters.

Roll cat is a miniature version of the crawl cat without spuds. It is normally provided with drum wheels.

7.4.8 Bed Leveller

This is a lowest cost dredging technique. The bed leveling technique uses an underwater plough/shovel that is towed at a controlled height via two cables.

The method has been successfully applied in the port of Rotterdam, Nantes and ports in the United Kingdom, where it is applied to clear the mud from the front of quay walls. It has also been tried out at Antwerp, where it has been effectively used to clean the access channels to the locks situated on the river Scheldt. A typical Bed-leveller installation is shown in Fig. 10.

7.4.9 Water Injection Equipment

This method consists of the re-fluidization of the settled mud by injection of air or water in the upper mud layers. Thus the re suspended mud can be taken away by the existing currents. The injection can be performed by either:

- Fixed injection-installation to be mounted alongside a quay wall or on the bottom.
- Mobile injection-installation or water-injection dredger such as “JetSed”.

A typical installation is shown in Fig. 11.

8. WORLD WIDE DISTRIBUTION OF DREDGERS

Based on a recent worldwide survey carried out, the various broad types of dredgers in use are as under:

		Percentage
• Cutter suction dredgers	-	50
• Trailer suction hopper dredgers	-	30
• Grab and Back hoe dredgers	-	10
• Bucket dredgers	-	5
• Other dredgers	-	5

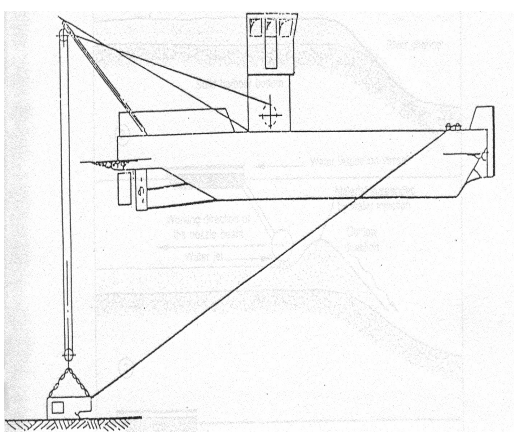


FIG. 10. BED-LEVELLER

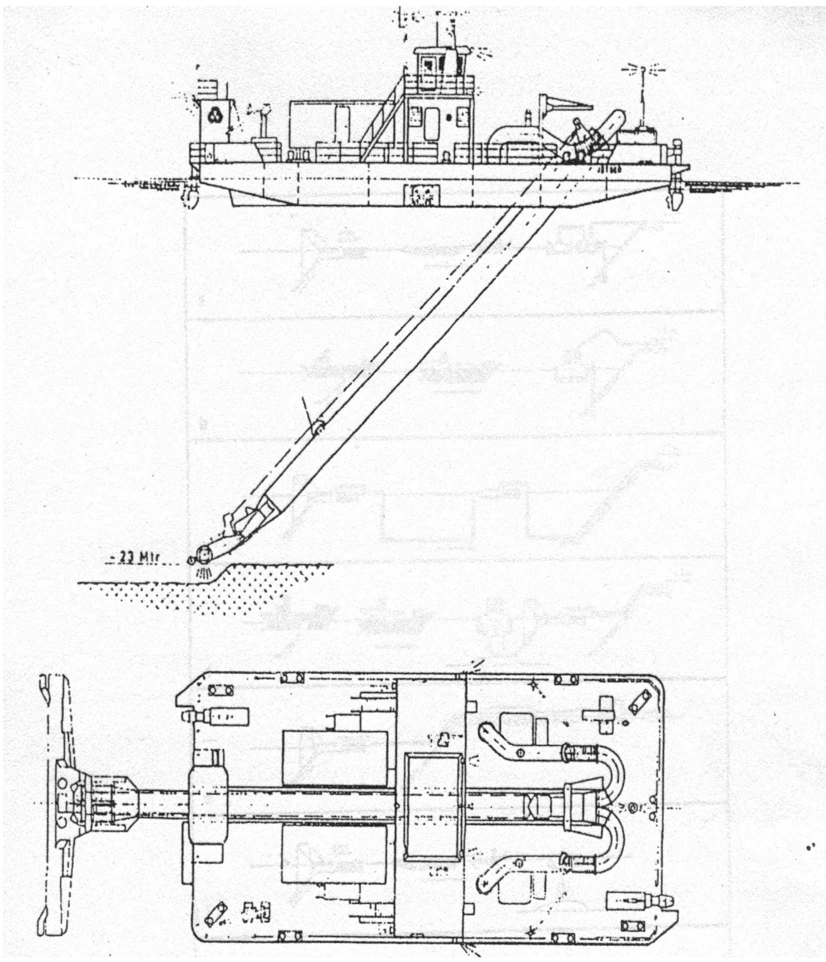


FIG. 11. TYPICAL INSTALLATION OF 'JETSED' DREDGER

9. SELECTION OF DREDGING EQUIPMENT

The equipment required to satisfactorily and economically complete a given dredging job will depend upon

- Size of the Project
- Nature of the material
- Volume of the material
- Topography of the area with reference to accessibility etc.
- Distance of dumping ground from dredging area
- Environmental factors at the site

- Accuracy of the work required
- Availability of equipment

All the above factors will have to be considered before taking a final decision in the matter.

Precise knowledge of soil and their dredgeability along with the environmental factors is very crucial in deciding the type of equipment to be used, the wear and tear to which the equipment will be subjected to etc. Accurate information regarding the type of soil to be dredged is important, as the variation in prices between silt and hard rock to be dredged can be 1:25 as shown in Table 2, representing the relative costs for different types of materials and type of dredgers. This in some cases may go even up to 1:100.

Table 2. *Comparative costs of various types of dredgers vs materials*

Type of Dredger *Cost Factor*	*Cutter Suction*	*Trailing Suction* *Material*	*Bucket*
1.00	Fine Silt	Sand	Soft Clay
1.30 – 2.00	Sand	Fine Silt	Sand
2.30 – 2.70	Soft Clay	Soft Clay	Medium Gravel
2.70 – 3.50	Medium Clay Medium Gravel	Medium Gravel	Coarse Gravel Stiff Clay
3.50 – 4.00	Hard ClayChalk	Coarse Gravel	ChalkHard Clay
6.00 – 10.00	Soft Rock	Stiff Clay	Soft Rock
15.00 – 25.00	Hard Rock (with blasting)	Hard Rock (with blasting)	Hard Rock (with blasting)

Source: Anthony Bates – Dredging + Port Construction (October 1978)
Soil Investigations for Dredging Works.

10. DREDGING IN INDIAN CONTEXT

India has a long coast line of about 7640 km out of which 2650 km is on the East Coast and 3360 km on the West Coast and the balance in Andaman Nicobar and Lakshadweep islands. There are 12 major ports. 6 on the West Coast and 6 on the East Coast. In addition there are over 185 non major ports. The capital and maintenance dredging requirement of these ports involve huge expenditure. These ports play an important role in the maritime trade of the country which is vital for the Indian economy. For efficient functioning of the ports, adequate depths are to be maintained at all times. The rate of siltation varies from port to port. The siltation is mainly caused by-

- Silt brought by the river
- Salt water/sweet water interaction
- Turbulence of the waves offshore during the monsoon and
- Littoral drift.

The actual quantities of Capital dredging carried out in the Indian Major Ports during the Ninth plan (1997-2002), the Tenth Plan (2002-2007) and the projected requirement for the Eleventh plan (2007-2012) are presented in Table 3.

Table 3. *Actual and projected capital dredging quantities of indian major ports*

Plan Period	*Quantity (in million cum)*
Ninth Plan (1997-2002) (Actual)	36
Tenth Plan (2002-2007) (Actual)	15

The actual quantities of Maintenance dredging carried out in the Indian Major Ports during the Ninth Plan (1997-2002), the Tenth Plan (2002-2007) and the projected requirement for the Eleventh Plan (2007-2012) are presented in Table 4.

Table 4. *Actual and projected maintenance dredging quantities of indian major ports*

Plan Period	*Quantity (in million cum)*
Ninth Plan (1997-2002) (Actual)	285
Tenth Plan (2002-2007) (Actual)	235
Eleventh Plan (2007 - 2012) (Projected)	380

The maintenance dredging quantities vary from year to year. The average annual quantity and the approximate cost of maintenance dredging in Indian major ports is presented in Table 5.

Table 5. *Average quantity and cost of maintenance dredging in indian major ports*

Port	*Quantity (in million cum)*	*Cost (in Rs. million)*
Kandla	4.5 500	
Mumbai (Average)	4.0	220
JNPT	1.5	150
Murmugao	3.5	120
New Mangalore	5.0	300
Cochin	11.0	350
Kolkata	21.0	3000
Paradip	2.5	200
Visakhapatnam	1.0	150
Ennore	0.3	50
Chennai	1.5	150
Tuticorin	Nil	Nil
Total	55.8	5010

11. DISPOSAL GROUNDS FOR DREDGED MATERIAL

One of the major problems encountered by many of the ports is the proper disposal of the dredged spoil. All the material obtained by dredging cannot be easily disposed off on the land and has to be dumped off shore. It therefore becomes imperative that the material disposed off shore does not find its way back to the navigational channels of the harbour. The cost of dredging will considerably increase in case the dumping ground is located too far away in the sea.

In the case of Indian Ports, these locations are situated about 4 to 10 km away from the dredging area. The location of the dumping ground is selected based on the sediment movement. These problems are normally studied on a movable bed model and confirmed by actual field studies. The following two methods are used:

- Fluorescent Tracer Studies
- Radio Active Tracer Studies

12. ROCK DREDGING

With the advent of very powerful and sophisticated cutter suction dredgers, it is possible to dredge rocks having a compressive strength of about 40 to 50 Mpa. The compressive strength of rock alone is not the sole criteria for direct dredging of rock. The tensile strength, rock quality designation, hardness, brittleness also play a vital role in direct dredging of rock. Specially designed heavy duty cutter suction dredgers equipped with specially designed water jets with cutter motor of about 8000 horse power and total installed horse power of about 36000 are being deployed for direct dredging of rock. The cost of direct dredging of rock may vary from Rs.200 to Rs.1000 per cum, compared to about Rs.100 per cum. for ordinary soils. Rocks having compressive strength above 40 MPa normally require pretreatment by underwater blasting before dredging.

13. UNDERWATER BLASTING AND DREDGING

The rock underwater can be removed by any one of the following methods

- Mechanical

  The fragmentation of the rock is carried out by means of a drop chisel or specially designed hydraulic rock hammers.
- Direct Dredging

  Universally used equipment for dredging of rock directly underwater is the cutter suction dredger equipped with powerful cutters. Heavy duty Backhoe dredgers are also deployed for ripping of rock.
- Chemical

  The chemical method of removal of rock underwater is by the use of explosives. The following three methods are used depending on the type, depth and area of rock to be removed.
  - Surface charge
  - Shaped charge and
  - Underwater drilling and blasting

  Once the rock is fragmented, any of the dredgers may be deployed depending upon the fragmentation of the rock.

14. DEEP SEA MINING

Dredging technology is being extended for deep sea nodule mining. The Poly metallic nodules are located between 4000 to 6000 m below covering an area of about 46 million square km and as such presently it is not economically viable compared to the extraction of metals from the land source. However, a great deal of interest has been generated in the recent years about the possibilities of extraction of minerals from the marine environment. This complex technology development is planned with multi institutional and multinational participation to derive the maximum benefit of the existing technology, resources and potential.

15. ENVIRONMENTAL IMPACT OF DREDGING

Environment can be defined as all the physical, chemical, biological and social factors likely to have an affect directly or indirectly, immediately or later on all living beings. Any alteration to this system is defined as an 'Impact' and 'Environmental Impact Surveys' should aim at identifying the areas of impact, types of impact, measurement and its control.

'Environmental Impact Assessment' forms an essential part of dredging and must be integrated in the whole sequence of processes that constitute the port project appraisal, planning, design and operation process. Beginning with the Oslo Convention of 1974 and the Paris Convention 1978, the European nations sought to limit the input of contaminants to the adjacent marine waters. The conventions addressed international waters. It was accepted that, the disposal of dredged materials could occur provided the materials contained only 'trace quantities of contaminants'. The intergovernmental convention of the dumping of wastes at sea is commonly known as the London Dumping Convention (LDC). The LDC and the rules and regulations made there under are now providing the basic framework to be followed. The London Convention on the prevention of marine pollution (Convention on the Prevention of Marine Pollution by Dumping Wastes and other Matters, London) was negotiated in London in November 1972 and came into force on 30th August 1975.

The working group of Permanent International Association of Navigation Congress (PIANC) in their report on 'Disposal of Dredged Material at Sea' (1986) had observed that "Out of the total volume of dredged material 90-95% is uncontaminated and not different from the material which is brought into the Sea by natural riverine processes". Thus the regulation will apply only to the remaining 5-10%.

16. CONCLUSION

- Dredging is an ancient art but a relatively young science.
- Dredging is broadly classified under
    - Capital and
    - Maintenance
- Soils to be dredged are classified as
    - Cohesive Soils

  - Non Cohesive Soils and
  - Consolidated Soils such as rock etc.
- Dredgers are broadly classified under
  - Mechanical
  - Hydraulic
  - Pneumatic and
  - Special Dredging Equipments
- In dredging projects the term “Site Investigation” assumes a greater significance, since its result not only significantly affect the cost of the work but may also cause serious doubts as to the desirability or otherwise of carrying out the work at all.
- Accurate information regarding the type of soil to be dredged is important, as the variation in prices between silt and hard rock to be dredged can be 1:25 to 1: 100.
- Maintenance dredging quantities of Indian Major Ports involve a quantity of about 56 million cum per annum involving a financial out lay of about Rs. 5000 million.
- Environmental considerations have assumed considerable importance in the recent years specially for dredging works. An Environmental Impact Assessment will be required during the planning stages of the capital dredging as well as during the maintenance dredging phase.

REFERENCES

[1] Bray, R.N., “Dredging”, *A Handbook for Engineers*, Edward Arnold, London, 1979

[2] “Dredging + Port Construction”, *Industrial and Marine Publications Ltd*. U.K.

[3] Herbich John, B., “Coastal and Deep Ocean Dredging”, *Gulf publishing Co. Texas*, 1975

[4] Kamath, M.M., “Recent Developments in Dredging Technology”, *Second Indian Conference in Ocean Engineering*, CWPRS, Pune, 1983

[5] Kamath, M.M., “Under Water Blasting and. Dredging”, *Port Technology International London*,Twelfth Edition, 2000

[6] Kamath, M.M., “Capital Dredging at New Mangalore Port” *Terra et Aqua IADC Netherlands*, Number 82, 2001

[7] Kamath, M.M., “Maintenance Dredging Practice at New Mangalore Port” *Port Technology International London*, Fifteenth Edition, 2001.

[8] Kamath, M.M., “Rock Dredging – A State of the Art Review” *Dredging Seminar*, *Ministry of Shipping, Government of India, Ministry Flanders, Belgium*, Dredging Corporation of India Ltd., Bombay, 2003

[9] Kamath, M.M., “Sediment Transport along the Indian Coastline” Dredging for Development, *Dredging Symposium – 2003, EADA (India),* Mumbai, 2003

[10] Kamath, M.M., “Radio Tracer Applications for Selection of Disposal Sites for Dredged Materials”, Nuclear Technology and Societal Needs - *15th Annual Conference of Indian Nuclear Society (INSAC – 2004)*, BARC, Mumbai. 2004

[11] Kamath, M.M., “Technological Developments in Dredging” *Third Indian National Conference on Harbour and Ocean Engineering (INCHOE 2004)* – National Institute of Oceanography, Goa, 2004

[12] Kamath, M.M., “Site Investigation For Dredging Works” *Ministry of Shipping, Government of India, Ministry Flanders, Belgium*, Dredging Corporation of India Ltd., Mumbai, 2005

[13] Kamath, M.M., “Maintenance Dredging Practices in The Major Ports Of India” *New Developments in Dredging Equipment and Planning – EADA (India) 2005*, Mumbai, 2005.

[14] Proceedings of the International symposium on Dredging Technology -I to III BHRA. England,1975 – 1980

[15] Proceedings of the World Dredging Conferences, WODCON I to XI, 1967 – 1986

[16] Proceedings of the International Course on Dredging, Institution of Civil Engineers, London, 1967.

[17] “Port & Dredging”, *IHC*, Holland.

[18] Report of the Dredging Task Force, Dredging for development, Committee on Port & Ship Safety, Environment and Construction, The International Association of Ports & Harbour, 1991

[19] *Terra ET Aqua*, International Association of Dredging Companies, Netherlands.

[20] World Dredging & Marine construction, California, USA.

[21] VLIEGER. H. de., “Dredging in cohesive sediments Areas” *Coastal, Esturaial and Harbour Engineers* Reference Book Edited BY: Abbott M.B. and Price W.A. Chapman & Hall, London, 1993

[22] “Report of the Working Group on Dredging Facilities for Tenth Five Year Plan 2002 – 2007”, Planning Commission, Govt. of India.

COMPUTER AND INFORMATION TECHNOLOGIES

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

GRID COMPUTING[1]

P.S. DHEKNE*

1. INTRODUCTION

Grid computing has been identified as one of the main "ingredients" of e-Science: a mean of linking research communities, developing partnerships between countries and promoting worldwide science, education and innovation, thus truly enabling a genuine virtual global research space. Also by integrating distributed computing environments using low cost Internet based technology, now it has become possible to provide supercomputing power in the hands of individual users that until now could only dream of affording such power without having to make an exorbitant capital investment. Thus Computational Grid, in a way, is next logical step to Cluster/parallel computing, and allows to be used as a single, unified resource for solving large-scale computation and data-intensive computing applications.

Grid technology has changed the way advanced research is being conducted today. In India too, the main driver for introduction of GRID activity has been participation in High Energy Physics (HEP) experiments at CERN, Geneva. This paper provides as a real-world example of a production quality operational Worldwide LHC Computing Grid (WLCG). Also different Grids available in the country are described and future challenges are also discussed.

2. GRAND CHALLENGE COMPUTE PROBLEM OF CERN

European Organization for Nuclear Research (CERN) is building Large Hadron Collider (LHC), the largest accelerator in the world, for searching Higgs particle leading to the understanding of the origin of masses of fundamental particles and unification of fundamental forces of nature. The LHC represents a leap forward in particle beam energy, density and collision frequency. The LHC is scheduled to be operational in mid 2008. It will initially accelerate proton beams up to 7 TeV (7000 Gev) and study p-p collision and latter on study collision of Pb nuclei, each of 1150 TeV. Latter experiment is expected to give evidence for the quark-gluon plasma, the new phase of matter. Four large experiments, ALICE, ATLAS, CMS and LHCb [Fig.1] are being planned for the detection of

*Raja Rammana Fellow, BARC,Trombay, Mumbai 400085
[1]This paper is based on the lectures delivered at TIFAC-Core, SASTRA, Tanjavur

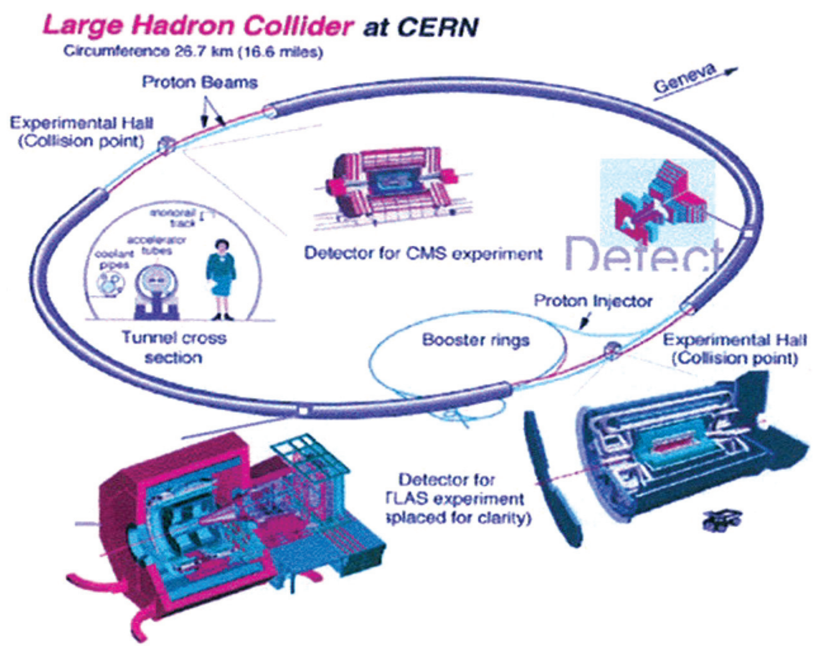


FIG. 1

subsidiary particles generated during the collision of the proton beams. Each collision is expected to generate up to 2000 subsidiary particles, which will be observed through 180,000 channels. At a collision rate of 40 MHz, the expected data generation rate out of the experiments is about 7- 8 Giga Bytes per second, amounting to 8-10 Petabytes (1000, 000 Gigabytes) of data generated in a year of experiment time.

The extraction of results from the LHC experiments will present a number of challenges in terms of computing, due to the unprecedented complexity and rates of the data, the length of time of the programme and the geographically distributed scientific communities that will coherently need to operate on these data. The large (Presently 1800 Physicists, 150 Institutes, 32 Countries) Computing challenges in the LHC lies in the real time storage of the huge amount of data, re-construction of tracks of particles released during collision and computational simulation for physics experiments. The storage requirements are about a million times the presently available storage on desktop personal computers. The expected performance required for most rudimentary simulations is about 20 Teraflop sustained speed, which is equivalent to 40,000 personal computers. The required computational performance is about 10 times the highest sustained speed so far achieved on the fastest super computer.

Clustering a large number of the processors has been the standard practice being followed for the past several years for meeting the very large computational requirements beyond the capacity of the available processors. In the ranking of **supercomputers** according to the LINPACK benchmark in Top500 supercomputer list more than 70% supercomputers are constructed using clustering technology. At BARC, developments of ANUPAM series of supercomputing clusters have served as main workhorse for high-end computing, storage and visualization [1,5,9,10] requirements of scientists & engineers for the past several years. However clustering beyond a few hundred processors is extremely expensive and it has many problems due the huge requirement of space, electricity, air-conditioning and staffing thus making it beyond the capacity of many large laboratories.

To extract meaningful physics results within a reasonable span of time from peta-bytes of data of unprecedented complexity and to include effective participation of collaborating institutions spread out world wide, the only way practically feasible was to develop GRID based distributed computing. CERN is meeting the computing challenge of LHC by using Grid Computing technology with the objective of exploiting widely dispersed Large National Computing facilities located in various countries. Of course integration of various distributed environments in different administrative domains having varied security policies poses new challenges in data security and data sharing.

3. WHAT IS THE GRID?

The Grid refers to an infrastructure that enables the integrated, collaborative use of high-end computers, networks, databases, and scientific instruments owned and managed by multiple organizations. The Computing Grid (usually just called "the Grid") is a powerful concept that provides a unifying principle for many activities in – and infrastructure plans for – computational science and technology.

Grid computing allows access to geographically distributed resources and offers consistent and inexpensive access to resources irrespective of their physical location or access point. It enables sharing, selection, and aggregation of a wide variety of geographically distributed computational resources (such as supercomputers, compute clusters, storage systems, data sources, instruments). Today's Internet as an infrastructure has increasing bandwidth, advanced services, higher storage capacity (Terabytes per site), increased availability of compute resources (clusters, supercomputers, etc.), advanced applications (simulation based design, advanced scientific instruments). The Internet infrastructure will connect multiple grids, creating a universal source of pervasive and dependable computing power that will support dramatically new classes of applications.

This technology involves development of Middleware and tools, on top of standard Internet software to facilitate the operation of a number of very large computing facilities interlinked through very high speed dedicated links, as a single virtual facility. The required computational performance for LHC is supported through a hieratical 3 Tier distributed computing architecture via world wide LHC Computing Grid (WLCG) [3]. The WLCG [12] is being implemented with CERN as Tier 0 centre (only 20 % computing power comes from Tier 0) surrounded by very large computational facilities located at major participating organizations in European and USA as Tier 1 centres (about 70 % computing power) forming the core of the grid and smaller computing facilities at various institutes/universities as Tier 2 centres (about 10% computing power) [Fig.2]. The desktops with individual physicists are the Tier 3 stations on the grid.

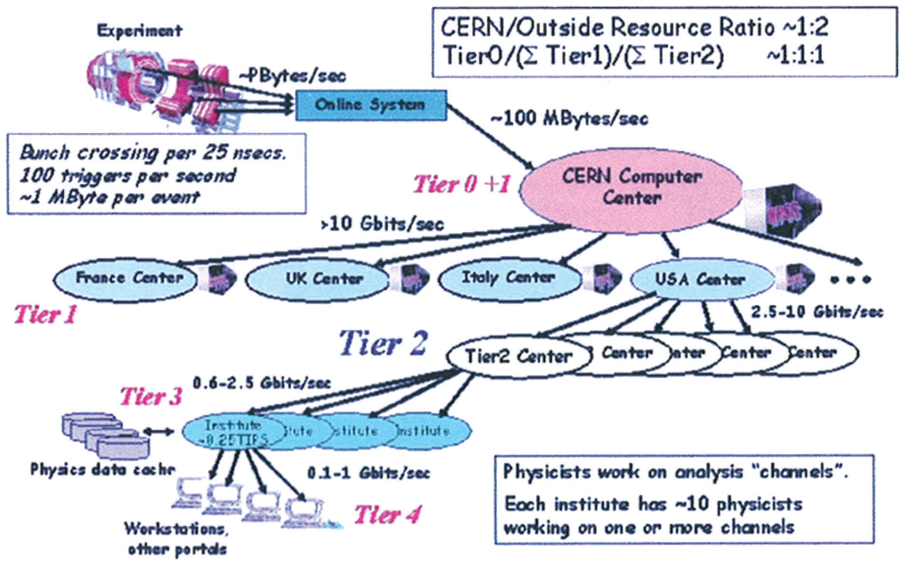


FIG. 2

3.1 Grid Middleware

The Grid technology involves development of Middleware and tools to facilitate the trouble free operation of a number of very large computing facilities interlinked through very high speed dedicated links, as a single virtual facility. For any Grid computing effort, middleware is a crucial component. Grid Middleware are software stacks designed to present disparate compute and data resources in a uniform manner, such that these resources can be accessed remotely by client software without needing to know *a priori* the systems' configurations.

The **Globus Toolkit** is an open source software toolkit used for building Grid systems and applications. It is being developed by the Globus Alliance and many others all over the world constructed from a number of components which make up a toolkit. This toolkit provides client, server and development components for the three Globus *"pillars"* of Grid Computing: Resource management, Information Management and Data Management. Currently WLCG machines use the **gLite** middleware deployed by the Enabling Grids for E-sciencE (EGEE) project. Distributed under a business friendly open source license, gLite integrates components from the best of current middleware projects, such as Condor and the Globus Toolkit, as well as components developed for the LCG project [6, 7].

Access to both flavors of middleware is based upon user owned X509 certificates and allows the use of GSI (Grid Security Infrastructure) proxy impersonation certificates. Both middleware Globus & gLite allow for single sign-on through a concept of virtual organization (VO).

The both middleware are compatible with schedulers such as PBS, Condor and LSF, built with interoperability in mind and providing foundation services that facilitate the building of Grid applications from all fields. However in absence of any particular standard actual implementations are varying to an extend causing interoperability & compatibility problems in certain areas.

4. WORLD WIDE SCENARIO

A very capable network infrastructure will be required than available presently to support the anticipated data flows in the future. Fortunately, such a network infrastructure now popularly known as e_infrastructure, is emerging in many countries mainly due to High Energy Physics collaborations and many emerging applications in weather, bio, material sciences.

The exponential use of the web by industry and the general population led commercial carriers to install a prodigious amount of optic fiber and related equipment, with far more capacity than the current demand. Those excess capacities, coupled with advances in optical network technology (such as dense wave division multiplexing) have resulted in steeply declining network prices, although in India prices have remained high. Furthermore, largely due to the adoption of Grids by the global high-energy physics community, transoceanic networks for research are becoming much faster; in 2003 there was at least one transatlantic network running at 2 Gb/s, faster than most networks within continents.

With European Union's political goal of equal opportunities for researchers throughout the European Research Area, they have created the GEANT network, which provides a European "backbone" network for connecting National Research & Education Networks (NREN) within European countries with 10 Gb/s bandwidth presently and with firm plans for further upgrades in the near future. GEANT connects individual country high-speed research networks (such as RENATER/France, GRNET/Greece, GARR/Italy, FCCN/Portugal, REDIRIS/Spain, SuperJANET/United Kingdom and ACONET/Austria). In total, GEANT and the networks it connects reach almost four thousand institutes in 33 countries.

EU-IndiaGrid a part of EGEE is set up, funded by European Commission with the aim to disseminate and spread Grid technology. Other world regions have or are putting in place high-speed research networks (e.g., in the United States the Teragrid network (40 Gb/s backbone, 30 Gb/s to individual sites), the Internet2 and Lightrail networks, the high-speed networks created by the CANARIE organization in Canada, CERNET in China, Academia Sinica Grid Centre, Taiwan, Academic and Research Network in Indonesia, Japan Gigabit Network and SInet and NIRI in Japan, KOREN in Korea, Research Networks in Malaysia, PHNET and PREGINET in the Philippines, SingAREN in Singapore, Thailand and APAN, the Asia Pacific Advanced Network. Equally important, high-speed transoceanic links are bridging these networks so that there will soon be a global research network infrastructure fast enough and with sufficient connectivity to support LHC data transfer needs.

5. REGIONAL WLCG IN INDIA

DAE is actively involved in the design, development and supply of LHC magnets, Power supplies, LHC control software, Data Acquisition System, detectors at CERN for the past few years. Indian scientists are taking active part in CMS and ALICE experiment. To support researchers with required infrastructure, India has also committed to setup a regional Tier-2 of CERN. These centers would provide access for various Universities and Indian collaborators, LCG Data Grid services from India. There are two Tier-2 centers in India, one at TIFR for CMS experimentalist, Mumbai with 1 Gbps/300 Mbps direct link to CERN, which in turn has a routing for ALICE Tier II center at VECC Kolkata. There are number of TIER-3 centres located at user's premises [Fig.3]. The funding for this is mainly coming from DAE and DST. During the international conference CHEP06 in Feb 2006 at TIFR, Mumbai a 622 Mbps link that was set up between TIFR, Mumbai and worldwide Grid in USA site on an experimental basis to demonstrate few Grid tools.

To make IT systems Grid-enable one has to develop Grid Middleware Software in the area of Grid Fabric management, Grid Data management, Data Security, Grid workload scheduling and monitoring services, fault tolerant systems etc. Starting in 2002 DAE has developed number of Grid based Tools [5,6] in the area of Fabric management, AFS file system and Data Management, which are being deployed by CERN in their phase-I of LHC Grid operations. BARC has a test-bed setup of 32 nodes Cluster, each node running Linux with Grid middleware like OpenAFS, Globus Grid software, OpenPBS, EDT tools connected to CERN for software development work of CERN. Currently BARC engineers are working on the development of Extremely Large Fabric Monitoring, Grid View and Grid Deployment & Operations software tools [4].

6. COMPUTING GRID AT BARC

BARC has very rich Computing and Communication facilities spread across 10-12 geographically distributed computer centers consisting of high speed compute servers, parallel computers and high end graphics systems, files servers connected by high speed fiber optics campus wide network.

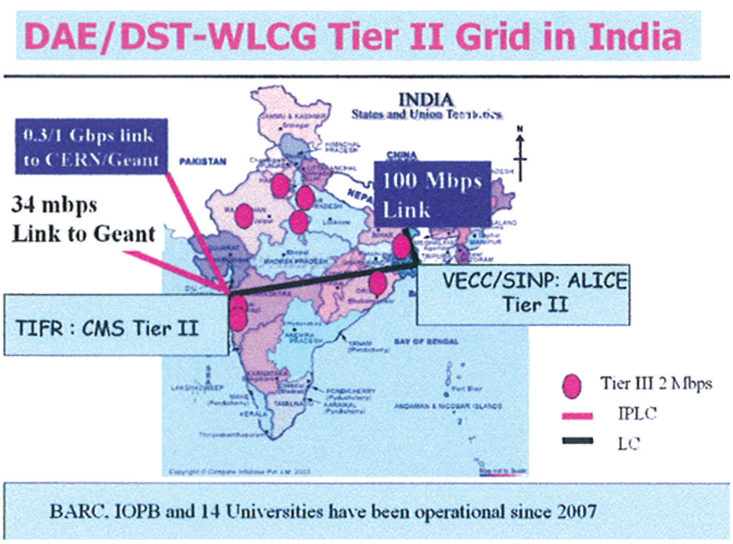


FIG. 3

However users have to connect to a particular computing resource to use them. Computer Division, BARC had as a proof-of-concept phase-I implementation of Grid technology to organize local computing resources at BARC in a computing grid fashion, which allow users to submit their jobs without worrying about the availability of resources like computing speed, memory or storage space, network bandwidth etc.

BARC has developed on experimental basis Computing Grid Infrastructure [8] by connecting distributed computing centers within BARC, to allow transparent & user-friendly access to computing, storage and network resources, which is first of its kind in the country [Fig.4]. Aims and objectives of this work was:

- Provide latest Grid technology to BARC users
- Use existing Computing infrastructure at BARC
- Gradually develop and add grid services as usage and technology matures to make computing systems more automated and transparent to users

7. DAE GRID

Based on the experience gained during WLCG and BARC Grid project a DAE wide Grid project was initiated by DAE in 10th five year plan. Currently DAE has operational DAE-Grid for exclusive use of DAE users. To start with various R&D centers of DAE such as BARC-Mumbai, VECC-Kolkatta, CAT-Indore and IGCAR-Kappakam have been interconnected (2/4 Mbps) via dedicated links. It uses **gLite** middleware components & supports both sequential as well as MPI based parallel jobs. To start with four major applications such as Grid-Portal for sharing High-end computers, Collaborative Tools for interactive working, remote backup and Portal for Information Sharing are being developed [Fig.5]. In the next phase all other DAE locations will be brought under DAE-Grid.

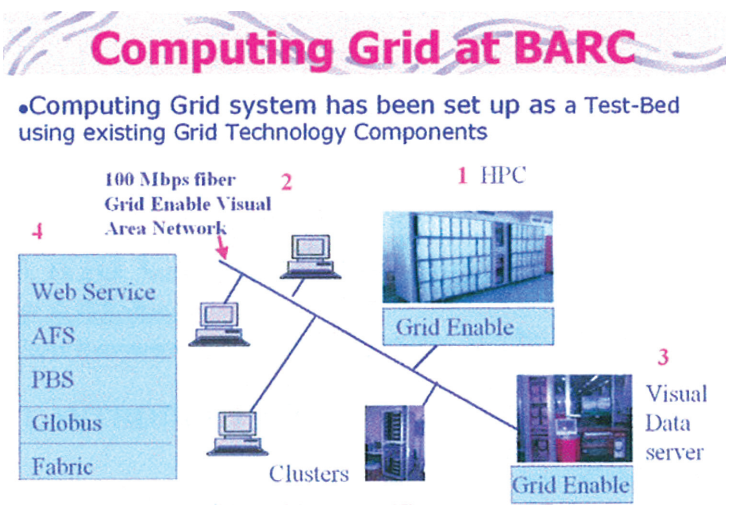


FIG. 4

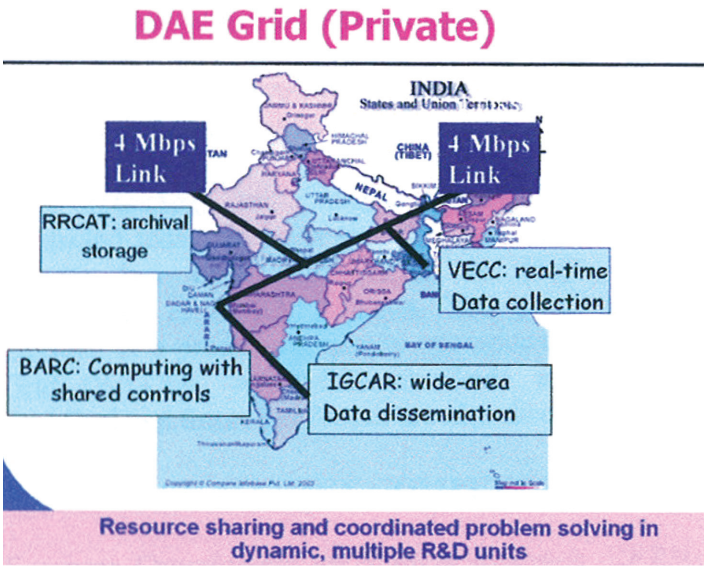


FIG. 5

8. C-DAC GARUDA GRID

The Centre for Development of Advanced Computing (http://www.cdac.in) is an autonomous scientific society of the Department of Information Technology, Ministry of Communication and Information Technology (MCIT), Government of India. It is primarily an R&D institution involved in the design, development and deployment of electronics and advanced information technology (IT) products and solutions.

Recently, C-DAC has initiated the "National Grid Computing Initiative: GARUDA" project connecting 17 cities across the country in its Proof of Concept (PoC) phase with an aim to bring grid networked computing to research labs and industry. GARUDA aims at strengthening and advancing scientific and technological excellence in the area of Grid and Peer-to-Peer technologies. To achieve its objective, GARUDA brings together a critical mass 45 higher learning educational and research institutions in the country on a 100 Mbps backbone. Initially, the institutions are able to access computer resources of each of the institutions and computer facilities installed at C-DAC. The Garuda uses Globus 2.0 middleware and other tools from multiple sources.

9. ERNET-GEANT GRID

There are a substantial number of initiatives in Europe involved in deploying and operating the European-wide e-Infrastructure. EU-IndiaGrid has leveraged on the existing infrastructure provided by the research networks such as: the Gigabit Pan-European Research & Education Network (GEANT) and the Indian Education Research Network (ERNET) to exploit the initiatives of high speed intercontinental network connections, and work in synergy with them for the optimisation of network usage in agreement with the objectives of the Communication Network Development scheme.

ERNET with European Grid – GEANT, has a programme (being pursued in consultation with European Union) to provide high speed Geant-connectivity for research and educational communities in India. Recently many IIT's and NIT's have initiated few R&D projects in Computing Grids and have implemented open source Grid software on lab scale to facilitate student projects and to gain in-depth knowledge in Grid technology.

10. EU-INDIAGRID

EU-IndiaGrid [11] aims to make available a common, interoperable Grid infrastructure to the European and Indian Scientific Community, in order to support existing EU-Indian collaborations in eScience and promoting new ones. In line with the support goals of the Research Infrastructures activity area of the European Union Sixth Framework Program, EU-IndiaGrid has enabled the interconnection between the most relevant European Grid infrastructure, EGEE, and the Indian Grids Garuda & regional WLCG, thus implementing an interoperable Grid infrastructure between Europe and India [2,3]. Already operating collaborations between Europe and India in specific application areas, characterized by strong requirements in terms of data processing will immediately take advantage from the availability of the new common infrastructure, providing a larger pool of computing and storage resources than available in non-Grid Environment.

The partners of this project are prominent actors on the European and Indian eInfrastructures scene: Italian National Institute for Nuclear Physics (INFN, project coordinator), Metaware SpA, Consortium GARR - the Italian Academic and Research Network (GARR), the Abdu Salam International Centre for Theoretical Physics, the Cambridge University, the Indian National Centre for Biological Sciences (NCBS), The Indian Education and Research Network (ERNET), the University of Pune, the SAHA Institute Calcutta, the Centre for Development of Advanced Computing (C-DAC), the Bhabha Atomic Research Centre, the TATA Institute for Fundamental Research Mumbai. The major emphasis was to support Grid applications from diverse communities

- High Energy Physics…………. DAE units
- Condense Matter Physics .... Pune Univ, BARC
- Material Science…..................TIFR, BARC
- Bio-Sciences………………….NCBS
- Earth Sciences …………….. IITM, CDAC, Pune

The EU-IndiaGrid project is playing a key role in fostering the cooperation between the GARUDA and regional WLCG Projects and EGEE, the major European Grid Initiative, and has achieved significant progress so far. This "bridging" role between European and Indian grid infrastructure has now the opportunity to improvise, thanks to the recently approved plan for a multi-gigabit, low latency, e-infrastructure: National Knowledge Network (NKN).

11. NATIONAL KNOWLEDGE NETWORK

To support Nations growing needs for Computing, Communication and encourage collaborative research, a hierarchical Cyber Environment supporting Grid Computing Model has been suggested as a part of National Knowledge Network. The Computational Grid will provide innovative approach for linking storage & computing resources that are widely distributed worldwide or nationwide. It is planned to explore new modes of computing by extending the concept of clusters to that of wide-area grids of supercomputers allocated dynamically to a common problem over both wide distance and multiple organizations.

A global computing layer built upon the collection of Campus Clusters inter-connected to a multi-gigabit all-optic-Network (AON) backbone of NKN that would allow fast, easy access and sharing of work between any groups across the country.

The desktops, servers, parallel systems and application software will run on Grid related open source suit of software such as Linux, MPI-CH, Globus, gLite, PBS, Condor, VO box and other tools to facilitate development of home grown products. The NKN will follow standard information security implementation and will have single sign on facility.

In the first phase of the project a demonstration of pilot applications in Biology and High Energy Weather/Climate and Condensed Matter Physics will be attempted on the newly implemented infrastructure in order to validate it and will furthermore be a first set of case histories. It is expected that more institutes from major cities would join NKN grid initiative in the phase II where it is expected to provide high-speed connectivity to various major places. Any institute who would like to join NKN must commit certain resources.

12. GRID CHALLENGES

The success of this strategy will depend upon how effectively it will be implemented and primarily on developing software which will help to solve various issues as given below: The Grid is also quite relevant for commercial applications and many are being pursued. However this technology is still evolving and many issues (few are given below) need to be tackled properly.

How to achieve high reliability and 99 % uptime on widely distributed systems operated under different administrative domains?

How to provide efficient access to shared resource without causing undue delays?

How to give end user confidence that his data is well protected & secure?

How to provide access to multiple applications to a single user?

How to meter resources and charge users?

It is becoming apparent that the use of Grids will be an enabler for major advances and new ways of doing science. Grids have the potential to integrate as never before the triad of scientific methods – theory, experiment, and computation – and to do so, on a global scale. This integration can be accomplished by providing a unified environment in which one can execute simulations using models based on theory, access relevant experimental data, perhaps obtain instrument data in real time under control of the simulation, and compare the computational and experimental results.

Grids also provide a way to greatly increase the interaction between number of individuals, who analyze observational data, to facilitate tele-collaboration, and to provide broader access to unique experimental or computational facilities.

13. ACKNOWLEDGEMENTS

This article is based on the series of lectures delivered by the author at TIFAC-CORE, SASTRA, Engineering College, Tanjavur (TN), under Distinguished Visiting Professorship award scheme with AICTE and INAE support. Author wishes to thank Prof K. Ushadevi, staff members & students from SASTRA College for their interest & encouragement. Thanks are also due to INAE for the sponsorship and pursuance to convert lecture material into an article. Author also wishes to thank DAE authorities for the encouragement and support to carryout this work.

14. REFERENCES

1. A.G. Apte, P.S. Dhekne et al., "The Evaluation of ANUPAM Supercomputers", BARC NEWS Letter, issue 291, April 2008, pp.2-13
2. P.S. Dhekne, Keynote talk "Status of EU-IndaGrid", during 3rd IEEE International Conference on e-Science and Grid Computing, held at the Grand Hotel Ashok, Bangalore, India, in conjunction with the EU-IndiaGrid Conference during December 10-13, 2007 and December 14, 2007.
3. P.S. Dhekne, "Grid Status and EU-IndiaGrid", an invited paper presented during International workshop on Worldwide LHC Computing Grid Tier-2 and EU-IndiaGrid on High Energy Physics Applications at T.I.F.R.,Mumbai, December 1-4, 2006
4. P.S. Dhekne, "Status of Grids In India", an invited paper presented during International Symposium on Grid Computing, ISGC-2006 at in Academia Sinica Grid Computing Centre, Taipei, Taiwan during March-April 2006
5. Dinesh M. Sarode, S.K. Bose, Venkata P.P. K., P.S. Dhekne, Comupter Division, BARC, "High End Visualization with Scalable Display System", The Proceedings of International Conference on computing in High Energy and Nuclear Physics during 2006.
6. Rajesh Kalmady, Digamber Sonvane, Kislay Bhatt, Phool Chand, James Casey, Zdenek Sekera, "Gridview - A visualization tool for LCG", Proceedings of the CHEP06 an International Conference on Computing in High Energy and Nuclear Physics, Feb 13-17, 2006, T.I.F.R., Mumbai, India
7. Rohitashva Sharma, R.S. Mundada, P.S. Dhekne, Sonika Sachdeva, Tony Cass, Olof Barring, Helge Mainhard, "Measuring Quality of Service on Nodes in a Cluster", *CHEP06*, 2006 an International Conference on Computing in High Energy and Nuclear Physics February 13-17, 2006, T.I.F.R., Mumbai, India
8. P.S. Dhekne, "Emergence of Grid Computing Technology", presented during International conference Computers and Devices for Communication (CODEC 2004) The Institute of Radio Physics and Electronics, University of Calcutta at Hyatt Regency Hotel, Kolkata, during January 1-3, 2004

9. Dinesh M. Sarode, S.K. Bose, Venkata P.P.K, Soumen Dey, P.S. Dhekne, Computer Division, BARC, "Visual Environment with High Resolution Tiled Display and PC Rendering Cluster", published in BARC NEWS LETTER, N0:248, 2004
10. P.S. Dhekne, "High End Computing & Communication Systems", fellowship paper presented during 69th Indian National Academy of Engineering Annual Function held on December 12-13, 2003 at ECC Convention Centre, L&T Ltd. Chennai.
11. www.euindiagrid.eu
12. http://lcg.web.cern.ch/LCG/

MECHANICAL ENGINEERING

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

CHARACTERIZATION OF EFFERVESCENT ATOMIZATION

K. RAMAMURTHI*

1. INTRODUCTION

Effervescent atomizers have been employed for a variety of applications during the last decade. These include high-speed combustion devices, piston engines, industrial painting, drug preparation and agricultural sprays. The atomization process in these atomizers is brought about by the rapid growth of the gas bubbles in a two phase mixture of liquid and gas. Fine droplets with good penetration characteristics are reported to be obtained by this type of atomizer. The small droplets formed in the spray provide faster evaporation of fuel in combustion chambers while the deep penetration gives better mixing, thus making the effervescent atomizer a promising option for high speed combustion devices such as scramjets.

Figure1 gives a schematic of a typical effervescent atomizer. It comprises of an aerator for generating gas bubbles in a flowing liquid, a mixing chamber or manifold, in which the gas bubbles are mixed with the stream of the liquid and a nozzle from which the two phase mixture is discharged through an orifice into the ambient. The atomizing gas is admitted in the form of fine bubbles. The gas bubbles in two phase mixture grow rapidly on being discharged from the orifice and shatter the surrounding liquid into ligaments and droplets.

The conditions of the two phase flow in the discharge orifice viz., the diameter of the gas bubbles, the distribution of bubble sizes, the number density of the bubbles and the density of the liquid and gas control the degree of expansion of the bubbles and hence govern the atomization process. The two-phase flows are generally catalogued as bubbly flow, slug-bubbly flow, annular flow and reverse annular flow. Depending on the regime or nature of the two-phase flow through the orifice, the atomization process will vary. Early studies of Lefebvre et al. [1] showed that good atomization was achieved for very small gas flow rate for which bubbly flow prevails in the orifice. The atomization characteristics in the different flow regimes have not been investigated earlier.

The effervescent atomizer is generally designed to operate in the bubbly flow regime for which the Gas to Liquid Ratio (GLR) is typically less than 0.02 [1]. However, depending on the

*ex- ISRO Trivandrum. Presently Department of Mechanical Engineering, IIT Madras, Chennai 600036

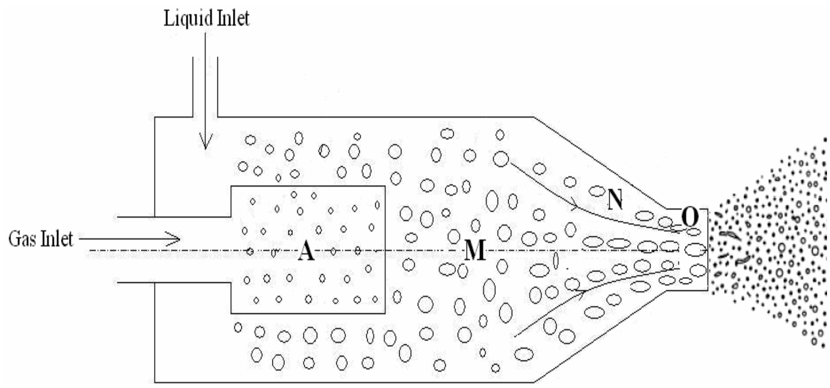


FIG. 1. SCHEMATIC OF EFFERVESCENT ATOMIZER

application, higher values of GLR have been used. Panchagnula et al. [2] use GLR between 0.02 and 0.1, whereas Peterson et al. [3] operated with GLR up to 0.5. The bubbly flow regime will not be operational for values of GLR exceeding about 0.02 and slug flow and annular type of two phase flows are formed. The atomizers using aqueous polymer solutions also use GLR up to 0.5. Commercially available effervescent atomizers for agricultural sprays, industrial painting and engines provide catalogue information of their products for GLR up to 0.10. The principle of atomization in bubbly flow and annular flow regimes are distinctly different and it is essential to determine the regime of operation and the changed characteristics of atomization with changes in the nature of the two phase flow in the discharge orifice of the atomizer.

The research work in effervescent atomizer has traditionally focused primarily on relating spray characteristics to operating parameters such as GLR, injection pressure and fluid physical properties such as viscosity, density and surface tension. Although these contribute in a major way to the practical development of effervescent atomizers, the basic differences in the mechanism of formation of the structure of the spray in the different flow regimes are not determined in the studies. The flow in the atomizer and the characteristics of spray formed in the bubbly, slug and annular flow regimes were recently experimentally investigated through a series of flow visualization experiments and spray measurements [4]. The findings are reported in the following. Section 2 gives the experiments done to study the structure of the spray in the different flow regimes. The observed changes in the structure of the spray and the driving mechanisms are discussed in section 3. The sizes of droplets and the distribution of droplet sizes in the spray in the different regimes are given in section 4.The major conclusions are summarized in the section 5.

2. EXPERIMENTS

2.1 Flow Visualization Experiments

The atomizer comprised a central mixing tube of 8 mm inner diameter and 160 mm long surrounded by an annular gas manifold (Fig. 2). Liquid was admitted in the mixing tube at pressures between

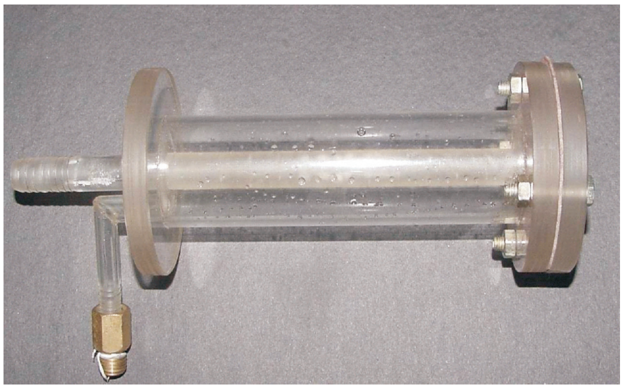


FIG. 2. OUTSIDE-IN EFFERVESCENT ATOMIZER

0.125 and 0.30 MPa. The mixing tube was surrounded by a gas manifold 35 mm in diameter. The gas supply pressure to the chamber was kept slightly higher than liquid injection pressure to ensure gas bubbling in the stream of liquid flowing in the mixing chamber. This configuration constitutes the outside-in type atomizer described by Chen and Lefebvre [5]. Sixty air injection holes of 0.5 mm diameter were drilled in the mixing tube. The choice of 0.5mm holes for bubbling gas into the mixing chamber was arrived at from consideration of force balance between the drag force on a growing bubble and the surface tension force retaining the bubbles at the injection holes. The diameter of the bubbles formed was seen to be between 0.5 and 1.2 mm for liquid velocities between 0.5 to 2 m/sec. The holes were organized in 15 rows with 4 holes in each row. A mixing length of 40 mm was provided downstream of the last row of injection holes. The mixing chamber was followed by a 45° conical nozzle to a discharge orifice 3.0 mm diameter. The length to diameter ratio of the orifice was 0.5. The atomizers were made of transparent acrylic and a typical atomizer is shown in Fig.2.

Water was used as the test liquid and air was used as the bubbling gas. The evolution of the bubbles and change in shape and agglomeration of bubbles in the atomizer was photographed using high-speed CMOS type digital camera at 1000 frames per second. The atomizer was illuminated with diffused backlighting. A translucent screen was placed at the opposite end of the camera as done by Catlin et al. [6] and illuminated with a 1000W halogen lamp. The experiments were carried out for injection pressures between 0.125 MPa and 0.3 MPa. The net variation of GLR in the experiments was between 0.002 and 0.20. The spray was photographed using the same high-speed digital camera with backlighting but at higher frame speed of 5000 frames per second.

A digital camera, Nikkon-D1X with an exposure time of 10 milliseconds was used to take still photographs of the spray. Illumination was provided by a stroboscope (1538-A Strobotac) with a peak flash intensity of 200,000 candela at 150,000 flashes per minute. Photographs were taken at a flashing rate of 10000 rpm in order to provide higher flash intensities.

3. STRUCTURE OF SPRAY IN DIFFERENT FLOW REGIMES

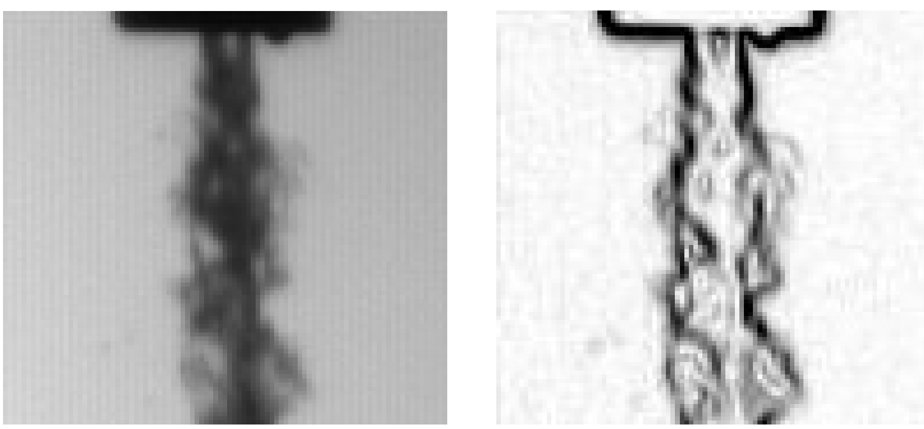
3.1 Bubbly Flow Regime

Figure 3a shows high-speed image of the spray obtained for a low injection pressure of 0.125 MPa and GLR of 0.002. This image after edge detection using image processing software ImageJ is shown in Fig. 3b. The Sobel edge detection method was used which performs a 2-D spatial gradient measurement on an image and emphasizes regions of high spatial frequency that correspond to edges. Figure 3b shows bubbles expanding at the core at section AA' very near to the exit of the atomizer. Thereafter, bubbles move towards the edge. The liquid column is seen to get ruptured by the expansion of the multiple bubbles at the edges and leads to the formation of the ligaments and holes. These holes further tear the liquid into additional ligaments and subsequently into droplets.

The above observations are different from the findings of the earlier researchers in that a sequence of individual bubbles expanding out to form ligaments in the liquid was put forth as the mechanism for spray formation. The difference can be attributed to the fact that earlier researchers such as Santangelo et al. [7], Catlin et al. [6] used comparatively smaller diameter discharge orifice and low liquid flow rates. The bubble sizes were therefore comparable with the discharge orifice diameter and the expansion of individual bubble initiated the atomization.

The movement of multiple bubbles in the atomizer nozzle and discharge orifice is shown in Fig. 4 for a typical value of GLR of 0.005 at an injection pressure of 0.15 MPa. The bubbles are seen to get elongated prior to the discharge orifice due to the acceleration of the flow in the nozzle.

Figures 5 and 6 show the photographs of spray at an injection pressure of 0.15 MPa and GLR of 0.005 and 0.015 respectively. Bubbly flow prevails in these cases. It is observed from Fig. 5 that atomization is initiated by the expansion of the gas bubbles outside the atomizer and the ligaments formed subsequently break-up into small droplets. The scheme of events is shown in Fig. 5. The number of bubbles present in the flow is small at the GLR of 0.005 (Fig. 5) and the small number of



a. Unprocessed spray picture   b. Edge-detected photograph

FIG. 3. SPRAY AT AN INJECTION PRESSURE = 0.125 MPA AND GLR = 0.002

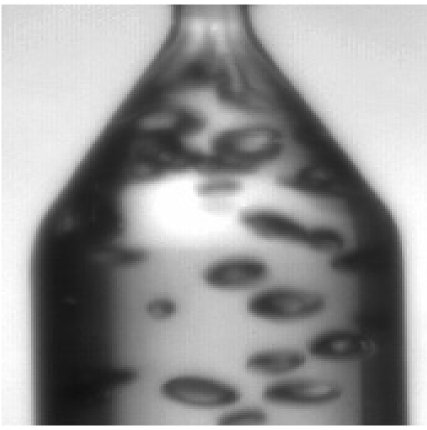


FIG. 4. BUBBLY FLOW INSIDE THE NOZZLE AT INJECTION PRESSURE = 0.15 MPA AND GLR = 0.005

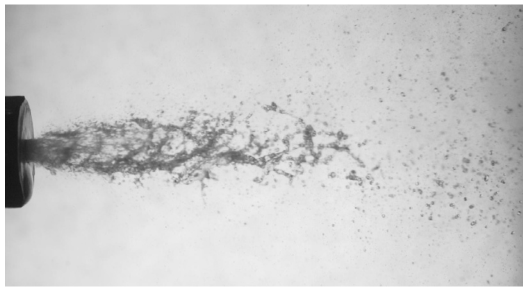


FIG. 5. SPRAY STRUCTURE FOR INJECTION PRESSURE = 0.15 MPA AND GLR = 0.005

bubbles leads to significantly larger ligaments. As the GLR is increased, the number of bubbles present in the flow increase giving rise to smaller ligaments. These small ligaments break-up more rapidly. This is illustrated in Fig. 6. The length scale required for ligaments to atomize in droplets is seen to reduce significantly from 5 cm to 2.4 cm as GLR increases from 0.005 to 0.015. The length scale at which the bubbles induce break-up is however small and reduces only slightly with increase in GLR as shown in Figs. 5 and 6.

Figure 7 shows the structure of the spray under bubbly flow regime at the same GLR of 0.005 but at higher injection pressure of 0.3 MPa. It is observed from Fig. 7 that the characteristic length at which break-up occur increases drastically from 0.5 cm at low injection pressure to 1.7 cm at the higher injection pressure of 0.3 MPa (Fig. 7). The increase is attributed to the fact at lower injection pressure, the size of the bubble is comparatively higher, so that the expansion of a small number of

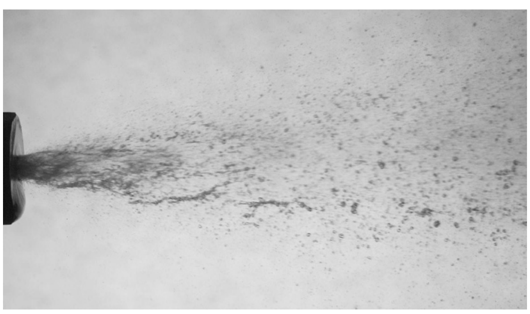


FIG. 6. SPRAY STRUCTURE FOR INJECTION PRESSURE = 0.015 MPA AND GLR = 0.015

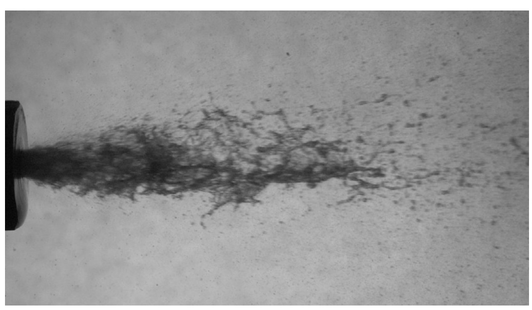


FIG. 7. SPRAY STRUCTURE FOR INJECTION PRESSURE = 0.3 MPA AND GLR = 0.005

large size bubbles ruptures the liquid column very near to the discharge orifice. This observation is consistent with the previous investigation of Santangelo et al. [7].

At the higher injection pressures, smaller diameter bubbles are formed in the flow. As a result, a dense column of liquid is observed to persist over a longer distance as shown in Fig. 7. But as the GLR is increased, the length of this dense liquid column reduces drastically as shown in the Fig. 8 for a GLR of 0.013. With a further increase in GLR to 0.0225, smaller bubbles present in the flow coalesce to form bigger diameter bubbles. This is shown in Fig. 9.

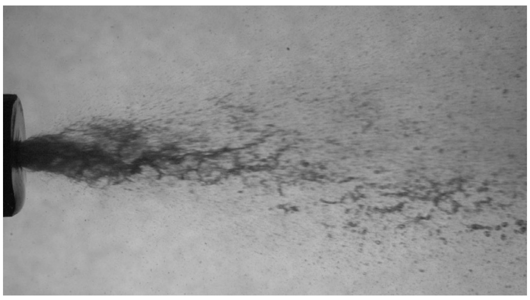


FIG. 8. SPRAY STRUCTURE FOR INJECTION PRESSURE = 0.3 MPA AND GLR = 0.013

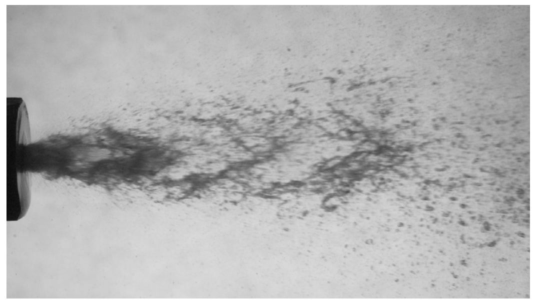
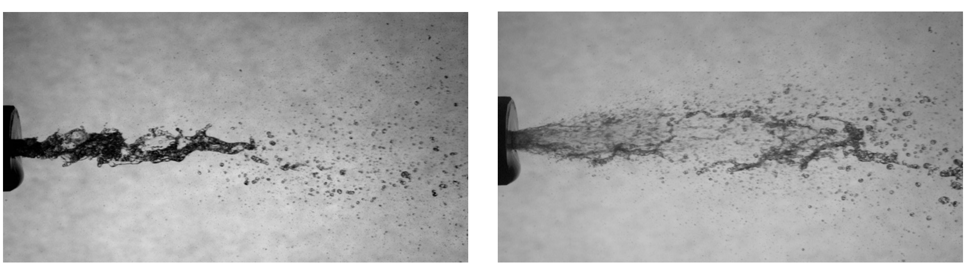


FIG. 9. SPRAY STRUCTURE FOR INJECTION PRESSURE= 0.3 MPA AND GLR= 0.0225

3.2 Slug Flow Regime

The characteristics and structure of the spray is significantly different in slug flow regime as compared to bubbly flow. Slug flow is associated with the intermittent flow of big bubbles comparable to the size of the discharge orifice. As a result, slug flow is inherently unsteady and fluctuating. Figure 10 shows the spray structure for an injection pressure of 0.15 MPa and GLR of 0.03 obtained at two different instants of times 5 ms apart. Large bubbles comparable to discharge orifice diameter are seen to emanate from the discharge orifice. The expansion of these bubbles leads to the very fine atomization in some portion of spray (Fig. 10b). The atomization, however, is not uniform and certain regions of large sized ligaments and droplets are seen in Figs. 10a and 10b.



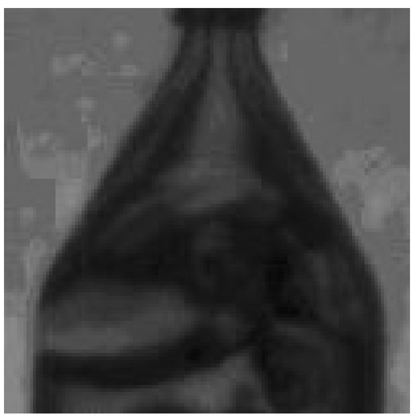
a) time 't_o' b) time 't_o'+ 5 ms

FIG. 10. SPRAY STRUCTURE FOR INJECTION PRESSURE OF 0.15 MPA AND GLR OF 0.03

3.3 Annular Flow Regime

At higher value of GLR, the expansion and elongation of the cluster of bubbles leads to the coalescence of the bubbles to form gas core in the nozzle. This is shown in Fig. 11 for a GLR of 0.08 and injection pressure of 0.15 MPa. A thin annular film of liquid flow takes place through the discharge orifice.

The structure of spray formed in annular flow regime is totally different from that in the bubbly and slug flow regime. A wave mode type of break-up is observed as shown in Fig. 12. The liquid emanates as thin hollow conical sheet, which becomes unstable and wavy due to interaction with the gas moving at high relative velocity inside the core of the annular liquid column. The image processing of this photograph of spray by the edge detection procedure, outlined earlier, showed the amplitude of the wave to progressively increase with distance from the orifice. This is illustrated in Fig. 13 and agrees with the nonlinear break-up of thin liquid sheets by the Kelvin Helmholtz instability [8]. Very fine droplets are formed in the spray.



GLR = 0.008

FIG. 11. FLOW INSIDE THE NOZZLE AT INJECTION PRESSURE OF 0.15 MPA

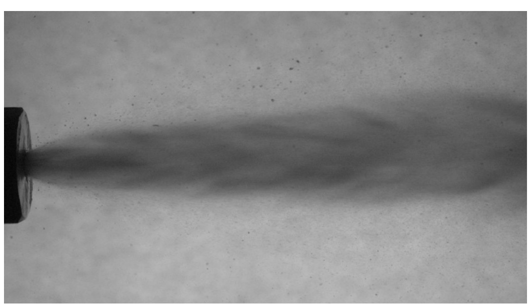


FIG. 12. SPRAY STRUCTURE FOR INJECTION PRESSURE OF 0.15 MPA AND GLR OF 0.07

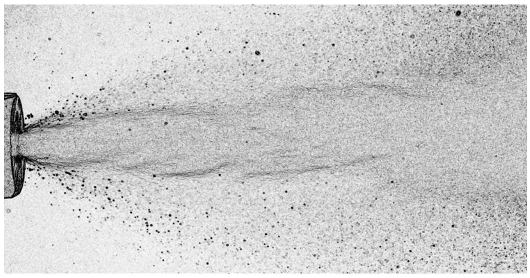


FIG. 13. EDGE-DETECTED PHOTOGRAPH OF SPRAY STRUCTURE FOR INJECTION PRESSURE OF 0.15 MPA AND GLR OF 0.07

4. SIMULATION OF SPRAY STRUCTURE

Flow visualization experiments are carried out to determine the structure of the sprays formed for bubbly and annular flow in the atomizer. The experimentally determined structures are seen to be qualitatively reproduced in the numerical simulations. Simulation of spray for the two values of GLR in bubbly flow regime is shown in Fig. 14. The qualitative features are similar to the experimentally determined spray structure given in Figs. 5, 6 and 7. The expansion of bubbles tears the edge of the liquid to form ligaments and droplets as in the experiments. At the small GLR of 0.005, larger ligaments are observed (Fig. 14) which is also in agreement with the experiments. As

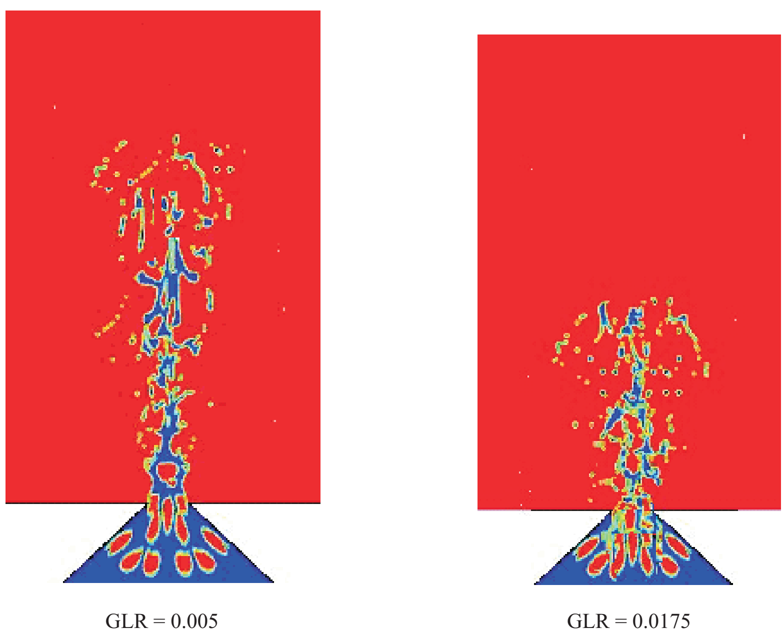


FIG. 14. SIMULATION IN A BUBBLY FLOW REGIME FOR INJECTION PRESSURE = 0.3 MPA

the GLR is increased to 0.0175, finer ligaments are formed leading to the smaller droplet sizes in spray (Fig. 14). The length scale required for the ligament to break-up reduces with increase in GLR as was observed in the experiments.

The simulation of the spray in the annular flow regime is shown in Fig. 15. The GLR considered is 0.08. It is observed that liquid comes out as thin sheet with gas inside it. This liquid sheet breaks into very fine droplets through wave motion as seen in the experiments.

5. DROP SIZES AND THEIR DISTRIBUTION

5.1 Drop Sizes in Spray

The SMD measured at a distance 30 cm from the orifice for different value of injection pressure at GLR between 0.005 and 0.15 is shown in Fig. 16. It is seen that SMD decreases rather rapidly with increase of GLR for small value of GLR for which bubbly flow exists. As GLR increases, the rate of change in SMD with GLR decreases. In particular, the variation of SMD due to change in GLR in the annular flow regime is small compared to the bubbly flow region. In the bubbly flow region, changes in injection pressure significantly influence the drop size.

A decrease in SMD with an increase of injection pressure has been catalogued in the earlier investigations. In the present work the pressure dependence is observed essentially in the bubbly flow region. Catlin et al. [18] in their work on slug and annular flow regime also found an insignificant role of injection pressure on SMD.

The variation of SMD due to changes in injection pressure and GLR in the bubbly flow region can be expressed in terms of a combined parameter involving GLR and Reynolds number. The data of SMD given in Fig. 16, if plotted as a function of the product of GLR and Re for small GLR

FIG. 15. NUMERICAL SIMULATION OF THE SPRAY STRUCTURE FOR ANNULAR REGIMES FOR AN INJECTION PRESSURE OF 0.3 MPA AND GLR OF 0.08

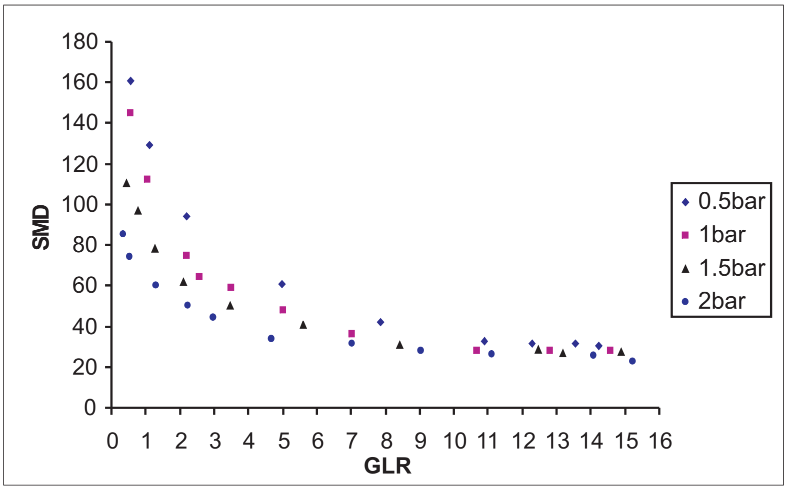


FIG. 16. SMD VARIATION WITH GLR AT DIFFERENT INJECTION PRESSURE

(< 0.04), for which bubbly flow exists, shows that the data for all pressures fall on a single curve. This is shown in Fig. 17. A curve fit given by the equation,

$$\text{SMD} = 4 \times 10^{-6}[\text{Re}.(\text{GLR})^{0.5}]^{-1.14} \tag{7}$$

fits the data with a correlation coefficient of 0.9814.

Lefebvre [9] has proposed for twin fluid atomizers the non-dimensional representation of SMD to be given by,

$$\frac{SMD}{do} = [A.(We)^{-0.5} + B.(oh)^{0.5}]\,[1 + GLR^{-1}] \tag{8}$$

For a given fluid and orifice diameter the above expression can be represented in the form

$$SMD = f[GLR^{-1} \times Re^{-0.25}] \tag{9}$$

A plot of SMD as a function of $\text{GLR}^{-1}\ \text{Re}^{-0.4}$ for values of GLR for which annular flow regime is shown in Fig. 18. A curve fit of experimental data gives

$$\text{SMD} = 1300\ [\text{GLR}^{-1}\ \text{Re}^{-0.4}]^{0.64} \tag{10}$$

with a correlation parameter R of 0.94.

5.2 Drop Size Distribution

Figure 19 shows the typical droplet size distribution in the spray obtained at an injection pressure 0.25 MPa for an axial distance of 30 cm from the atomizer. As the GLR increases and the different regimes of bubbly, slug and annular flow are encountered; the symmetric distribution about a mean gets skewed leading bimodal and trimodal type of distributions. The departure from a symmetric distribution is probably due to the wave mode type of sheet disintegration as the flow departs from bubbly flow.

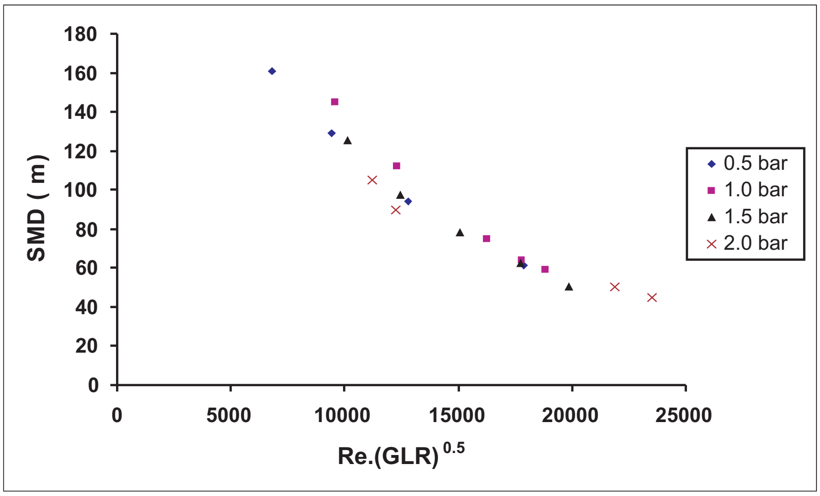


FIG. 17. VARIATION OF DROP SIZE IN BUBBLY FLOW REGIME

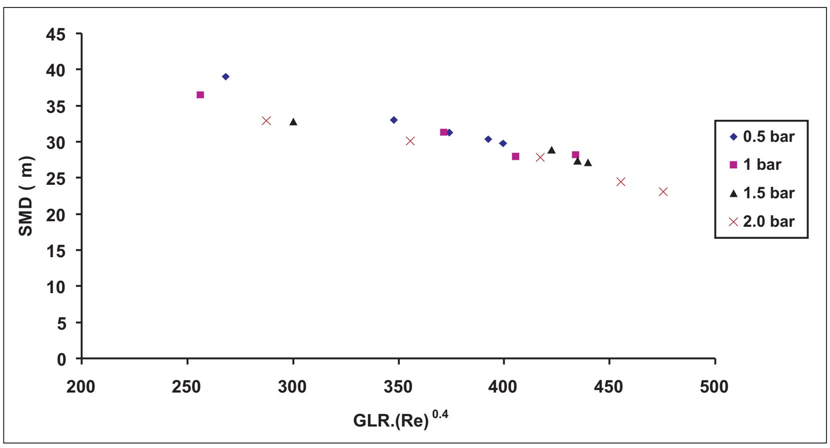


FIG. 18. VARIATION OF DROP SIZE IN ANNULAR FLOW REGIME

6. CONCLUSIONS

Experiments and numerical simulations show different characteristics of the spray formed in the various flow regimes in an effervescent atomizer. The simulations reproduce the experimental findings. The following are the main conclusions.

1. In the bubbly flow region expansion of bubbles at the edges of the spray promote atomization. As the injection pressure increases, for a given gas to liquid ratio, the break-up length increases due to the small diameter of bubbles formed. With increase in gas to liquid ratio the break-up length drastically decreases.

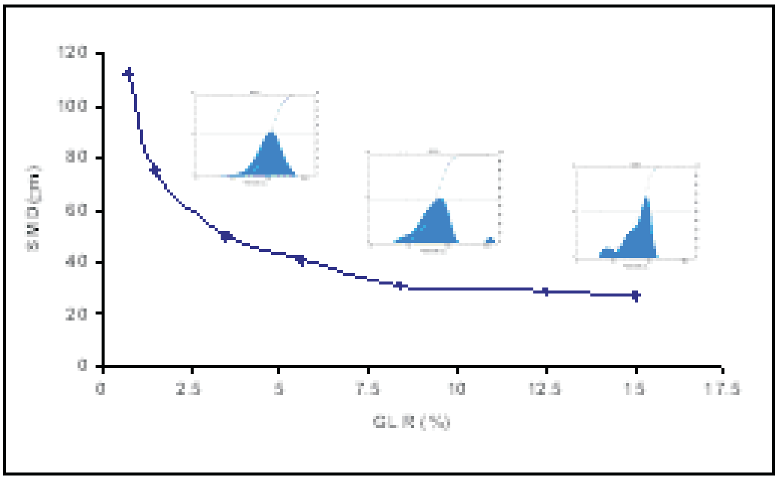


FIG. 19. DROP SIZE DISTRIBUTION AT 0.15 MPA INJECTION PRESSURE AND AT AXIAL DISTANCE OF 30 CM FOR VARYING GLR

2. The slug flow is associated with unsteadiness in flow and results in the formation of regions of fine and coarse atomization. The agglomeration of bubbles to form slug flow and the heterogeneity in the spray is captured in the experiments and simulations.
3. The wavy nature of atomization is seen for annular flows with the amplitude of wave motion increasing with distance from the discharge orifice. Very fine droplets are found in the spray in the annular regime.
4. The droplet sizes and the distribution of droplet sizes in the effervescent spray depend on the regime of two phase flow in the atomizer. Correlations are given for Sauter mean diameter in the bubbly and annular flow regimes. The symmetric monomodal distribution of droplet sizes in the bubbly flow region changes to bimodal and trimodal distributions for slug and annular flows.

ACKNOWLEDGEMENTS

Mr. U.K. Sarkar, M. Tech student IIT Madras and Professor B.N. Raghunandan of IISc, Bangalore contributed to the work reported in this paper.

REFERENCES

[1] A.H. Lefebvre, X.F. Wang and C.A. Martin, Spray Characteristics of Aerated-Liquid Pressure Atomizers, *AIAA J. Prop Power*, vol. 4, pp. 293-298, 1988.

[2] M.V. Panchagnula and P. E. Sojka, Spatial Droplet Velocity and Size Profiles in Effervescent Atomizer-Produced Sprays, *Fuel,* vol.78, pp.729–741,1999.

[3] F.J. Peterson, O. Worts, T. Schaefer and P.E. Sojka, Design and Atomization properties of an Inside-out Type Effervescent Atomizer, Drug Dev. Ind. Pharm., vol. 30, pp. 319-326, 2004.

[4] U.K. Sarkar, K. Ramamurthi and B.N. Raghunandan, Performance Characteristics of Effervescent Atomizer in Different Flow Regimes, Submitted to *Atomization and Spray,* April 2007.

[5] J.S. Chin and A.H. Lefebvre, A Design Procedure for Effervescent Atomizers, *ASME J. Engg Gas Turbines Power*, vol.117, pp. 266-271, 1995.

[6] C.A. Catlin and J. Swithenbank, Physical Processes Influencing Effervescent Atomizer Performance in the Slug and Annular Flow Regimes, *Atomization and Sprays*, vol.11, pp. 575-595, 2001.

[7] P.J. Santagelo and P.E. Sojka, A Holographic Investigation of the Near-Nozzle Structure of an Effervescent Atomizer-Produced Spray, *Atomization and Sprays*, vol. 5, pp. 137-155, 1995.

[8] T.J. Tharakan and K. Ramamurthi, Nonlinear Break-up of Thin Liquid Sheets, *Acta Mechanica*, vol. 156, pp. 29-46, 2002.

[9] A.H. Lefebvre, *Gas Turbine Combustion*, Hemisphere Pub., New York, 1983.

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

CHARACTERISTICS OF SPRAYS FORMED IN SWIRLED FLOWS

K. RAMAMURTHI*

1. INTRODUCTION

Coaxial gas-assist injectors of shear and swirl type are used in gas turbine combustors and cryogenic propellant rockets. The gas, in these injectors, is admitted through an annular gap located coaxially with the inner cylindrical post. The ratio of the mass flow rates of the liquid to the gas in the injectors of cryogenic propellant rockets is about an order of magnitude higher than for the gas turbine injectors and this brings about changes in the configuration. The mechanism of formation of droplets is, however, similar with the gas shear being primarily used for disintegrating the liquid. The high velocity gas provides the momentum for breaking up the relatively slow moving liquid stream.

The gas assist injectors of gas turbines have been extensively studied and are termed as air-blast atomizers. Two types of air blast atomizers comprising plain jet and pre-filming types are more popular. In the pre-filming type, liquid fuel is spread into a thin continuous annular sheet which is sandwiched between two high velocity air-streams. In the plain jet atomizer, the fuel is discharged in the form of one or more circular jets in a single high velocity air-stream. The early designs of pre-filming atomizers provided swirl to the liquid fuel to spread it out. The swirl was generated by tangential entry or through use of vanes and helical swirlers. Swirl was also provided to the high velocity air in an effort to ensure more intimate mixing with the fuel. Generally a marked improvement was obtained in the quality of the spray in the presence of air swirl.

Coaxial injectors of cryogenic propellant rockets are basically similar to the plain jet and pre-filming type of air blast atomizers. The much larger flow rates of liquid in the rockets gives rise to the scheme of injection being different. Methods of improving the atomization and combustion by recessing of the central liquid post within the outer annular element and provision of step and flare at the exit of the central element have been explored. Swirl has also been provided to the liquid or to the gas or to both liquid and gas as in the early pre-filming type of atomizers. A schematic of the

*Ex. Liquid Propulsion Systems Centre, Trivandrum [Presently, Department of Mechanical Engineering, Indian Institute of Technology Madras, Chennai 600 036]

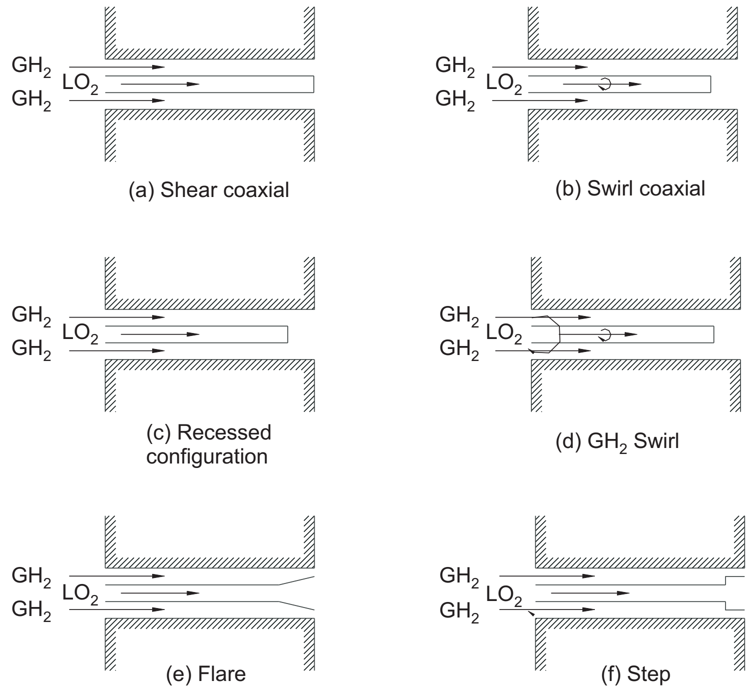


FIG. 1. SCHEMATIC OF FLOW CONFIGURATIONS

different configurations is given in Fig. 1. Cold flow experiments show that with gas swirl a diverging spray shown in Fig. 2.a becomes a largely dispersed spray under certain conditions. The configuration of the dispersed spray is shown in Fig 2.b. The gas swirl, instead of improving combustion, could therefore lead to a deterioration of the performance. The spray structure from different configurations of the coaxial injector was therefore studied with specific experiments on flow visualization of the gas jets since they significantly influence the spray pattern. The results from these experiments are discussed in this article.

2. SHAPE OF LIQUID SHEET FORMED BY INJECTOR IN ABSENCE OF GAS FLOW

A series of experiments was done by varying the degree of swirl at different injection velocities and ambient pressure with different dimensions of the nozzle [1] to characterize the divergence behavior in the absence of coaxial gas flow. These showed the formation of a well-defined conical sheet disintegrating into droplets at higher values of injection pressures. At lower values of injection pressures, a tulip-shaped sheet is formed. These two shapes are shown in Fig. 3.

The conical divergent sheet is not sensitive to small variations in injection pressures and ambient pressures unlike a tulip-shaped sheet. The cone-shaped sheet is formed above a certain threshold value of Weber number typically around 100 [2]. The Reynolds number is not very influential even

FIG. 2a. CONICAL DIVERGING SPRAY

FIG. 2b. DISPERSED SPRAY

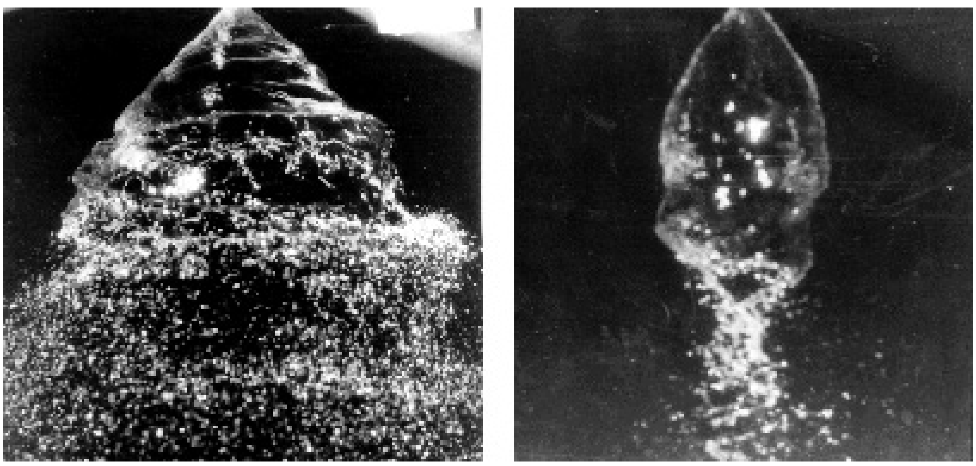


FIG. 3. CONICAL AND TULIP-SHAPED LIQUID SHEET FORMED

though the transition is usually seen with turbulent values of Reynolds number as shown in Fig. 4. The desirability of generating a thin liquid sheet, such as formed in the diverging cone, is important in that the mean droplet sizes vary as the sheet thickness.

The variation of the divergence angle, non-dimensionalized with a characteristic helical angle generating the swirl, plotted as a function of the ratio of the centrifugal forces in the liquid sheet to the surface tension forces at the exit from the injector is given in Fig. 5. It shows that a constant

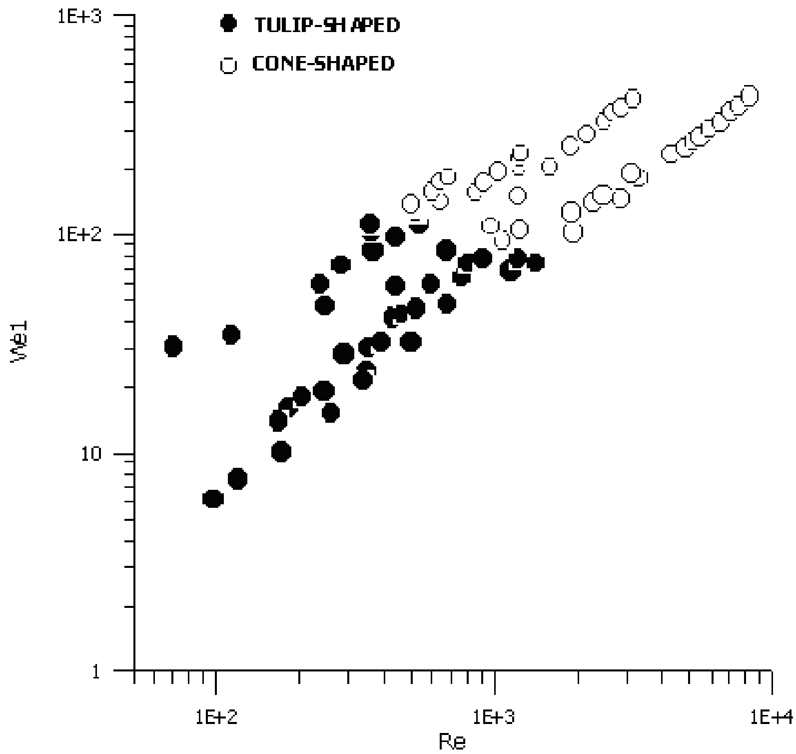


FIG. 4. TRANSITION WEBER NUMBER

divergence angle is achieved when the centrifugal force exceeds the surface tension force by about two orders of magnitude. The centrifugal force causes the sheet to diverge out whereas the surface tension force brings about an inward collapse. Figure 6 shows that a conical shaped sheet, once formed, does not lead to any significant change in the divergence angle nor the discharge coefficient when the injection pressure is varied. There is a need therefore to use a strong conical structure for the sprays.

Large values of ambient pressure modify the shape of the spray especially in the far field. On leaving the injector, the initial divergence angle would not be influenced by the ambient pressure and the values measured in atmospheric conditions would hold good for a high pressure environment. In the far-field, however, a closing of the spray boundaries is to be expected and coalescing of the droplets is likely. In the cryogenic propellant rockets, combustion proceeds by the atomization of the liquid oxygen, its vaporization and mixing with hydrogen from the annular flow. Flame is anchored in the immediate vicinity of the injector exit and the flame sheet extends in the stoichiometry region of the mixing layer. The far-field of the spray and the oxygen droplet sizes therein would therefore not be important.

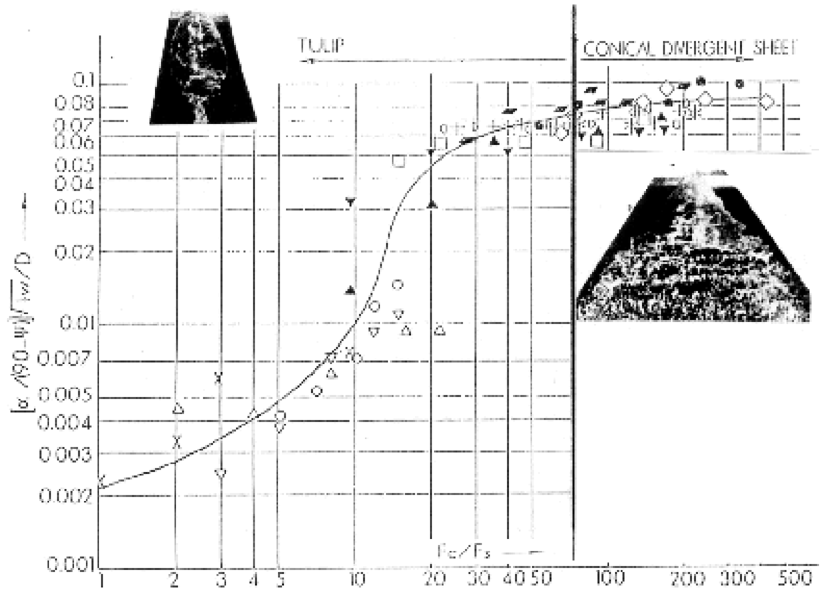


FIG. 5. NON-DIMENSIONAL REPRESENTATION OF TRANSITION

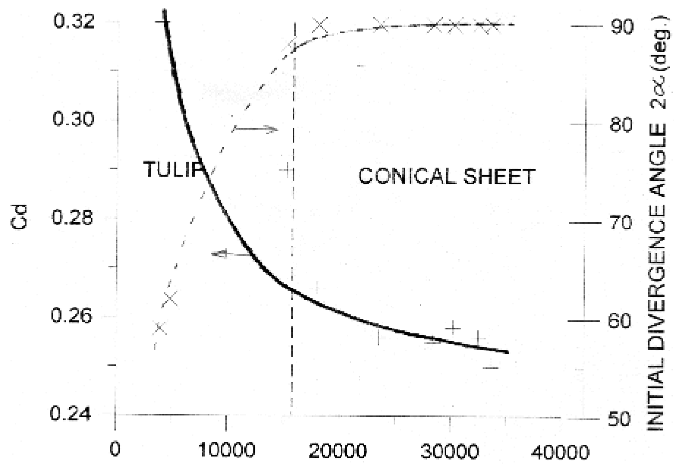


FIG. 6. INSENSITIVITY OF CONICAL STRUCTURE

3. ROLE OF GAS FLOW

The spread of the spray, shown in Fig. 2.b is not due to the liquid flow alone but could be governed by the flow pattern of the gas emanating from the annular nozzle. Experiments carried out with swirled and non-swirled jets at different values of Reynolds number showed not only higher

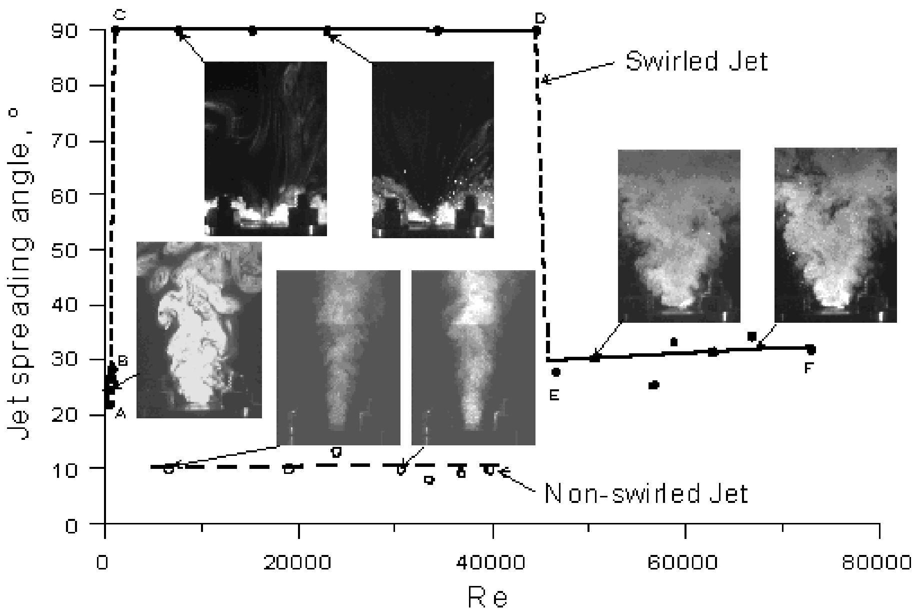
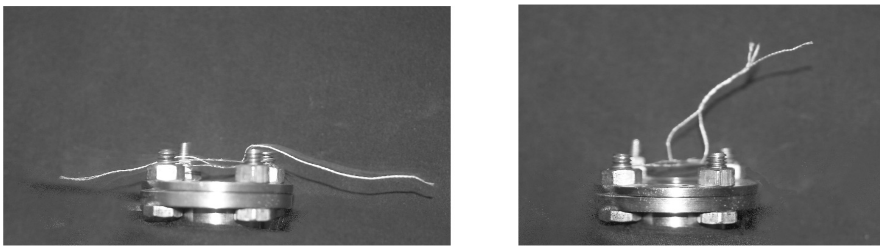


FIG. 7. DIVERGENCE OF GAS JETS

diverging angles for the swirled jet but also the transition into a wall jet under certain conditions. This is shown in Fig. 7. The gas jets were visualized by seeding the gas with a fine mist of oil and observing them with a planar sheet of high intensity laser light. The laser used was a Nd-Yag laser (532 nm) with an energy level of 25 mJ.

When the gas flow through the annulus is not swirled, smaller values of divergence are obtained as shown in Fig. 7. The flow, in this case, never gets attached to the injector base. A near-constant divergence angle of about 9^o is observed. In the case of a swirled jet, the divergence angle increase rapidly with Reynolds number at small values of Reynolds number (Portion AB in Fig. 7). Eddies are seen to be formed which are not symmetrical due the swirl and these grow. Once the swirled jet reaches a jet divergence angle of about 28°, the jet abruptly gets attached to the base of the injector to produce a wall jet. This is shown by the portion CD in Fig. 7. This wall-jet persists till about a Reynolds number of 46000 is reached at which point the jets reverts back to a divergent jet. This is shown by the region EF in Fig. 7. The different characteristics result from the interaction of eddies with the injector base. The two regimes could be distinctly seen in the flow picture of the jet taken by placing tufts in the flow (Fig. 8).

The noise spectrum in the jet flows also shows certain characteristic frequencies when the swirled jet gets attached to the wall. The broad-band spectrum, associated with turbulent jets is no longer obtained. The characteristic frequencies are seen to come from the vortex and scale with the flow velocity.



Wall Jet   Diverging Axial Jet

FIG. 8. MOTION OF TUFT SHOWING THE WALL JET AND THE DIVERGING JET

4. NOISE SPECTRUM OF THE DIFFERENT CONFIGURATIONS OF THE SWIRLED JET

Figure 9 gives the sound pressure level measured at an axial distance of 50 mm from the nozzle exit and at a radial distance of 50 mm from the axis of the nozzle. At Reynolds number less than about 750, very insignificant sound pressure level is observed considering that the mean kinetic energy of the jet is itself small. At higher value of Reynolds numbers, sufficient amplitude of sound pressure level is seen in both the wall jet region and the subsequent axial jet region. The frequency at which the sound pressure level is a maximum is also observed to increase with Reynolds number. This is seen in Fig. 9, wherein the wall jet region is illustrated for a Reynolds number of 19,000 and 30,000 and the axial jet regime for a Reynolds number of 48,500 and 79,000. The variation of the dominant frequency with increase of Reynolds number in the wall jet and the axial jet regimes is shown in Fig. 10. A reduction in the scaling of the frequencies with respect to velocities is observed when the jet becomes axial. The Strouhal number based on the diameter of the nozzle changes from 0.8 for the wall jet to 0.66 for the axial jet. The dominance of the shear layer of the axial jet could lead to a reduction of vortex shedding frequencies [3]. The frequency spectrum for the wall jet and axial jet, given in Fig. 9, shows that the sound pressure level at the dominant frequency abruptly drops once the jet changes to the axial diverging jet. This is brought out in Fig. 11 wherein the sound pressure level at the dominant frequency is plotted. A rapid decrease of the sound pressure level at the dominant frequency is observed when the flow changes from a wall jet to an axial jet. The sound pressure level in the jet gets diffused over a range of frequencies to give a relatively broader band spectrum. The line spectrum formed with wall jets is suggestive that it is essentially driven by the vortex and computations did show such vortex formation. The interaction of the vortex with the wall generates the flow along the wall. When the flow reverts back to axial, the sound pressure level becomes more diffused due to the difficulty of forming an organized vortex in a turbulent flow field.

5. NOISE SPECTRUM OF NON-SWIRLED JETS

In the case of non-swirled flow, the spot frequency from the vortex is missing and a very broad band frequency spectrum shown in Fig. 12 is obtained. The peaking of amplitude is seen to take place at frequencies of about 7000 Hz and 10,000 Hz and these dominant frequencies do not change with

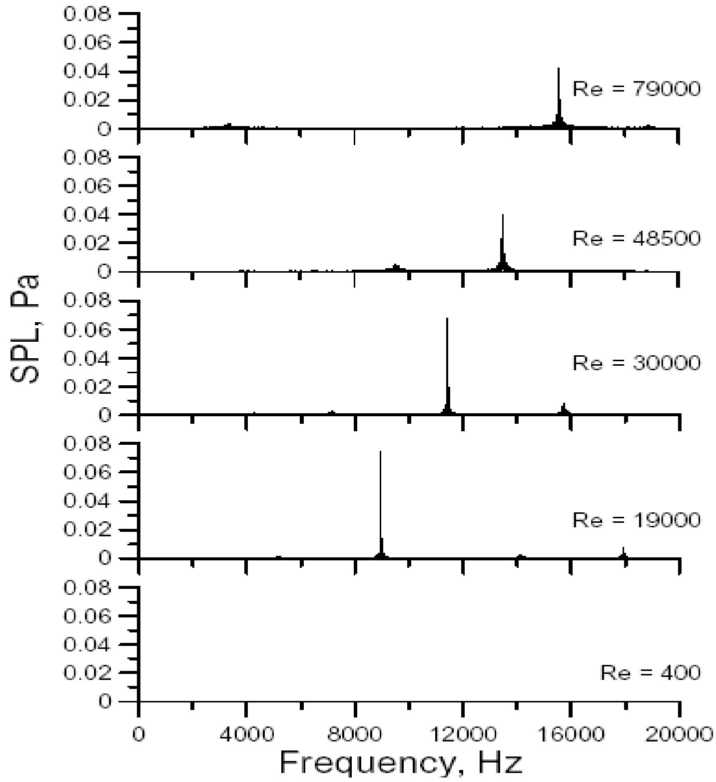


FIG. 9. SOUND PRESSURE SPECTRUM FOR SWIRLED JETS AT DIFFERENT REYNOLDS NUMBERS

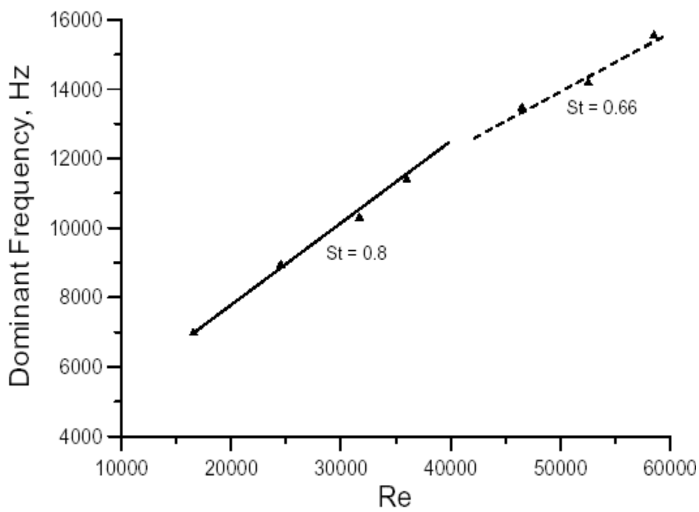


FIG. 10. VARIATION OF THE DOMINANT FREQUENCY

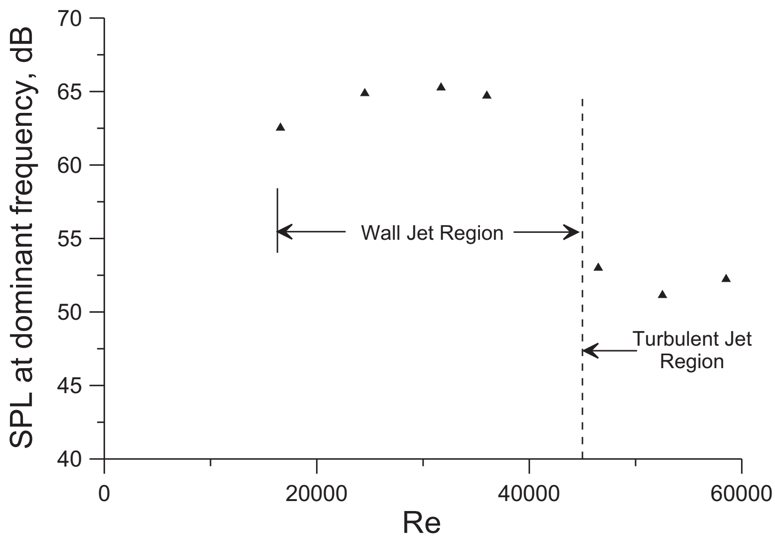


FIG. 11. SOUND PRESSURE LEVEL IN WALL JET AND AXIAL JET

increase of Reynolds number. They do not therefore come from flow but rather from the characteristic acoustic frequency of the nozzle used. The length of 23 mm of the nozzle gives a fundamental frequency in the longitudinal standing wave mode to be 7170 Hz which is near to the observed frequency of 7000 Hz. The frequency of the first tangential mode of standing wave is about 30,000 Hz which is very much higher than the value of 10,000 Hz. It appears likely that a sub-harmonic of the tangential mode is present from the radial injection of the gas. In general, compared to the swirled jet which has a line spectrum at the vortex shedding frequencies, the non-swirled jet has a broad band spectrum with peak amplitude at the characteristic acoustic frequencies. A wall jet is also never formed with the non-swirled jets.

6. SPREAD AND NOISE CHARACTERISTICS OF ANNULAR JETS

Injectors used in cryogenic propellant rockets make use of annular gas jets which could either be swirled or non-swirled. The annular gas jets were simulated with central cylindrical blockage of 4 mm diameter in the nozzle. The central blockage was also submerged within the nozzle to give a recess of 1.5 mm in some of the experiments. Figure 13 gives the measured values of the jet spread angle for the annular swirling and non-swirling jets. It is observed that in the presence of the central blockage, a wall jet is no longer formed as in the case of the swirled jets formed without the central blockage. The jet spread angle has also about the same value as obtained in the axial jet regime. The spread remains reasonably constant with increase of Reynolds number and this characteristic was also seen to hold good in the turbulent swirled jet regime. The provision of a recess in the annular nozzle also did not give the region of the wall jet. The structure of the jet even at very low Reynolds number (these are inset in Fig. 13) shows the absence of laminar mixing regions observable at the low Reynolds numbers. The near constant spread angle and a vigorous mixing structure with the

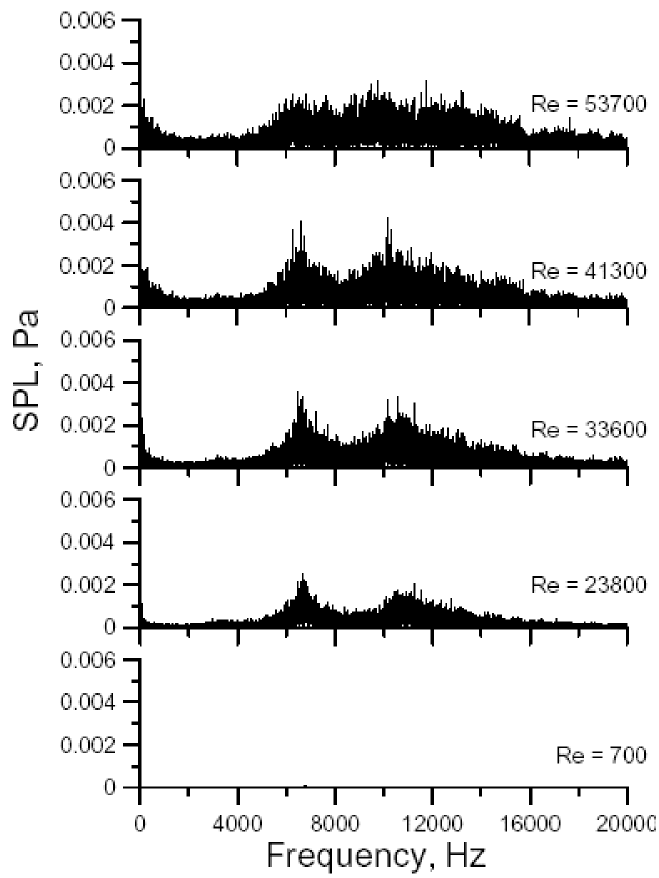


FIG. 12. FREQUENCY SPECTRUM FOR NON-SWIRLED JETS

incorporation of the central blockage is suggestive of formation of turbulent flow probably from the tangential entry interacting with the central blockage. Computations carried out at different values of Reynolds number assuming turbulent flow in the nozzle did not produce the wall jet. The jet spread angle obtained with a non-swirling flow (Fig. 10) is higher at 15° compared to 9° without the central blockage. The same trend is observed when the central blockage is recessed in the nozzle. The generation of vorticity in the narrow annular flow passages would have provided the enhanced divergence. The sound pressure level in the swirled and non-swirled jets is given in Fig. 15. Both give a broad-band spectrum. The frequencies corresponding to the maximum sound pressure level are higher at about 14,000 Hz.

7. CONCLUSIONS

Flow visualization experiments of swirled gas jets formed by tangentially admitting gas in cylindrical and annular nozzles show that a diverging axial jet, whose spread angle increases

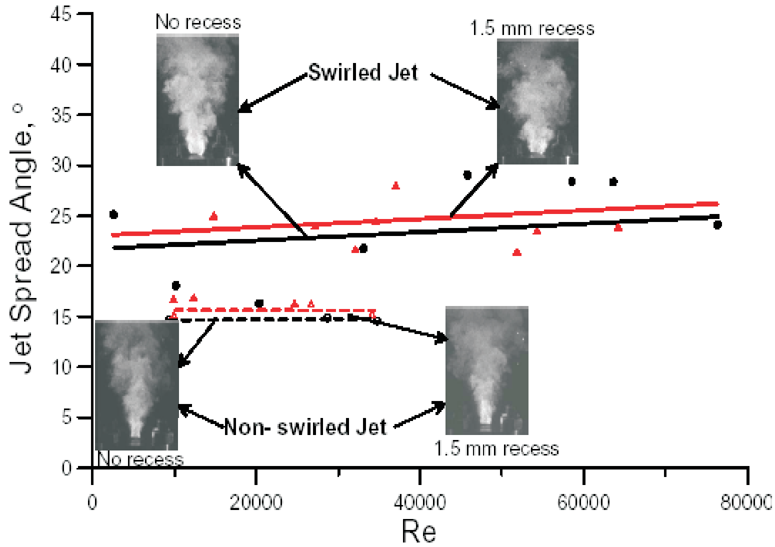


FIG. 13. SPREAD AND SHAPE OF ANNULAR JETS

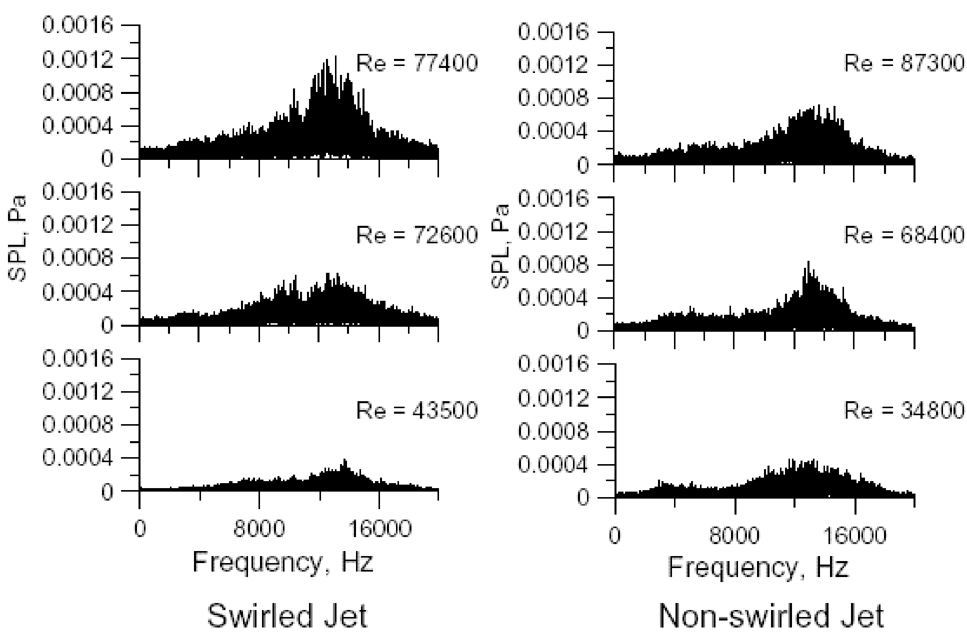


FIG. 14. SOUND PRESSURE LEVEL SPECTRUM OF SWIRLED AND NON-SWIRLED ANNULAR JETS

rapidly with increase of Reynolds number, is formed at low values of Reynolds number. This axial jet abruptly transits to a wall jet for laminar flow conditions at a Reynolds number wherein a vortex is formed near to the base of the nozzle. The interaction of the vortex with the base of the nozzle drives the flow along the base. The wall jet persists over a large range of Reynolds number as long as the flow through the nozzle is laminar. It transits back to an axial diverging jet when the flow through the nozzle becomes turbulent. The wall jet, driven by the vorticity, generates noise comprising of pure tones compared with the relatively broad band spectrum for the turbulent axial jets. Annular swirled jets, formed by improvising a central blockage in the cylindrical nozzle does not have the wall jet regime due to the turbulence generated from the central blockage. The study brings to focus the importance of the choice of Reynolds number, laminar or turbulent flow conditions and geometry of the nozzle in order to obtain a desired configuration of the gas jet and the spray.

REFERENCES

[1] K. Ramamurthi and T.J. Tharakan, An experimental study of liquid sheets formed in Coaxial swirl injector, J. Propulsion and Power, Vol. 11, 1995, pp. 1103-1109.

[2] K. Ramamurthi and T.J. Tharakan, Flow transition in swirled liquid sheets, AIAA J, Vol. 36, 1998, pp. 420-427.

[3] P. Huerre, Open shear flow instabilities, in Perspectives in Fluid Dynamics, Ed. G.K. Batchelor, H.K. Muffat, M.G. Worster, Cambridge University Press. (2000) pp. 159-230.

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

COMBUSTION INSTABILITY AND CONTROL

K. RAMAMURTHI*

1. INTRODUCTION

Combustion in any system is never perfectly smooth. Fluctuations occur in pressure, temperature, velocities and species present in the combustion products. The fluctuations are not only caused by combustion but also arise from turbulence, flow separation and other fluid mechanical processes. The fluctuations or disturbances, under certain conditions, grow with time to form limit cycle oscillations with large amplitudes. The growth of the oscillations is due to the coupling between the unsteadiness associated with the combustion process and the fluctuation of the different parameters [1]. The phenomenon is spoken of as combustion instability. Large amplitude oscillations are undesirable since they could lead to rapid enhancement in heat transfer and excessive vibrations and to component melting, mechanical failure, high levels of noise and increased burn rates.

A large number of investigations have been carried out on combustion instability in solid propellant and liquid propellant rockets and in afterburners of gas turbines. Passive methods of suppressing the instability have been studied and implemented in the design of combustion chambers of solid and liquid propellant rockets and afterburners of gas-turbines.

The desire to burn lean fuels in gas turbine combustors has led to a fresh wave of problems in combustion instability and this has motivated the search for new and innovative methods of actively and passively controlling the instability. Passive means of suppressing the combustion instability, implemented in the earlier solid and liquid propellant rockets and gas turbine afterburners, would not be adequate in situations wherein the operating regimes of a combustor keep varying.

The configuration and geometric dimensions of the combustor contribute to determine the frequency and amplitude of the oscillations. The combustion instability problem is essentially *system-driven*. Small changes in the configuration and dimensions have a significant influence and a generalization of the strategy used for suppressing i.e., “fixing” the instability is difficult. In this review, the characteristic features of the mechanisms driving the instability are examined in the context of two instances wherein combustion instability was encountered during the testing of

---

*Ex. Liquid Propulsion Systems Centre, Trivandrum 695 547 [Presently Department of Mechanical Engineering, Indian Institute of Technology Madras, Chennai 600 036]

liquid propellant rockets in ISRO. A few fixes for overcoming the problem of combustion instability were also studied experimentally and these are discussed. Some experiments carried out in lifted non-premixed jet flames offer scope for giving new directions on methods of suppressing combustion instability and these are examined.

2. COMBUSTION INSTABILITY IN A SMALL CRYOGENIC ROCKET

A small thrust rocket burning Liquid Oxygen (LO2) at a temperature of about 85K and gaseous Hydrogen at 200K showed consistently large amplitude oscillations at frequencies around 110 Hz. The thrust level was about 2 KN and the chamber pressure was 4 MPa. Figure 1 gives a sketch of the rocket along-with the feed-lines and the measurements. Figure 2 shows the large amplitude oscillations getting initiated at about 1.8 s after ignition and sustaining thereafter throughout the test. The spectral analysis of the oscillating pressures gave a spot frequency of 110 Hz. Disturbances are seen to grow into a stable limit cycle oscillation.

Low frequencies are associated with the feed system and it was initially suspected that the lower pressure drop across the LO2 injector port may not be providing adequate gain stabilization. The pressure drop across the LO2 injector port was varied by increasing the flow by 25% (i.e., increasing the pressure drop by about 5%). However, the oscillations persisted at about the same frequency. Reducing the LO2 pressure drop also led to oscillations at the same frequency. It was

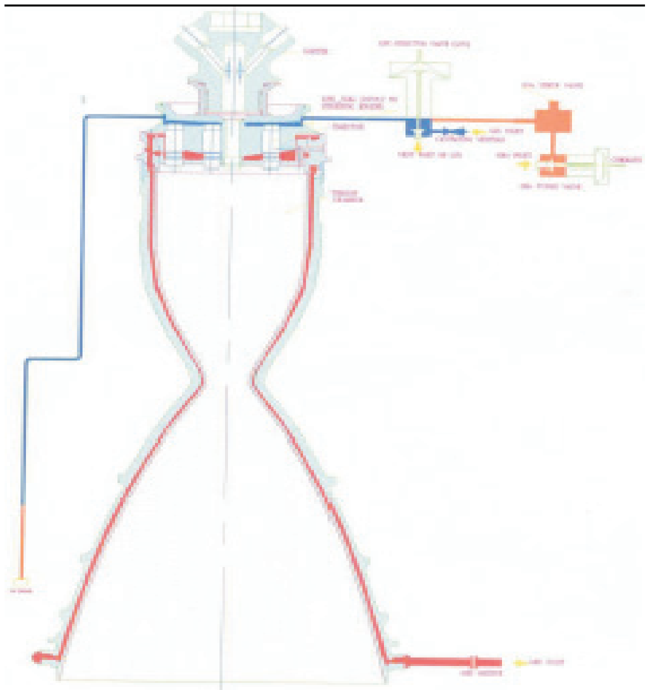


FIG. 1. CONFIGURATION OF ROCKET SYSTEM

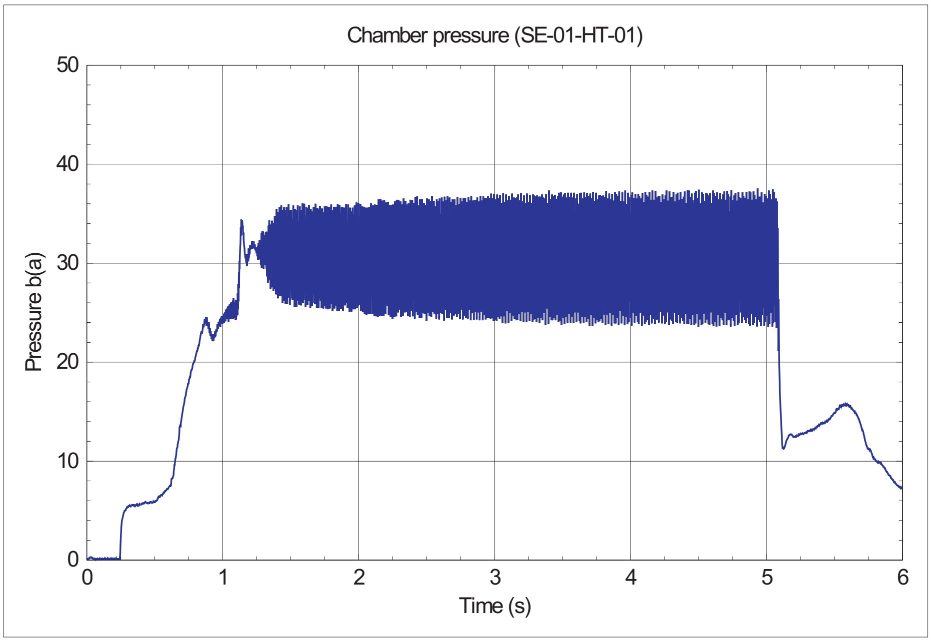


FIG. 2. OBSERVED LIMIT CYCLE OSCILLATIONS

therefore not the pressure drop which is alone responsible for the oscillations. If the gain stabilization due to the lower injection pressure drops was singly responsible, the frequency of oscillations should have given a trend of an increase in frequency at reduced pressure drops. This is because of the larger combustion delays associated with larger quantities of flow rates into the chamber. If the rate of mass accumulation in the chamber is expressed in terms of the difference between the hot gas generation in the chamber after a certain delay time (τ) after the propellant is admitted, it is possible to show that:

$$\omega\tau = \pi - \tan^{-1}\sqrt{P^2 - 1}$$

where $P = p_c/(2\Delta p)$, p_c being the chamber pressure and Δp the pressure drop across the injector. ω is the circular frequency and τ is the ignition delay. This assumes infinite capacitance upstream of injection. As Δp decreases, P would increase leading to a fall in the value of $\omega\tau$. In practice, only the sensitive part of the ignition delay would contribute to combustion instability. Even if the ignition delay does not change, the frequency should have decreased due to increase of P. This was not observed.

The length of the LOX feedline was 5 m giving the fundamental frequency of the standing waves in the line to be 100 Hz when the ends of the feedline corresponded to pressure antinodes. It was possible that the flow perturbations into the chamber are at this line frequency and could contribute to the observed pressure oscillations at about 110 Hz. The line length was drastically decreased to 0.55 m in some of the experiments in order to push the frequency up to about 910 Hz. However, the high amplitude oscillations were invariably present and persisted at the same frequency of 110 Hz.

The LO2 supply pressure in the above tests was measured with a canalization tube in order to isolate the pressure sensor from the low-temperature oxygen. This is shown in Fig. 1. The length of the canalization was 700 mm and was initially filled with room temperature helium gas. Depending on the initial gas pressure in the chamber, the canalization tube gets filled with varying length of the oxygen during the test. It was observed in the different tests that when the initial pressure of helium in the canalization tube is higher, the observed frequency of oscillation is smaller. A higher helium pressure in the canalization implies a shorter length of column of oxygen. The natural frequency of the canalization was around 100 Hz and a decreasing trend of frequency was also computed as the length of oxygen column in the canalization tube decreased. This suggested that the source for the observed frequency could be the canalization tube introduced for the measurements. It is the oscillation in the flow rate of propellant which couples to the combustion to produce limit cycle oscillations. When the canalization tube was removed (the measurement of the LO2 supply pressure was eliminated in the test), the oscillations in the chamber pressure vanished.

The above example shows the importance of the frequency content of the different elements of the feed system in causing the instability. The combustion process in this case should have had a component at about 110 Hz and this might have come from the recirculating eddies during the injection process. The coupling of this frequency with a measurement column in the feedline caused the combustion instability.

3. COMBUSTION INSTABILITY IN AN UPPER STAGE EARTH-STORABLE PROPELLANT ROCKET

Figure 3 shows the pressure-time trace in one of the earlier tests of an upper stage rocket of the PSLV launch vehicle. There is a dip in pressure followed by an increase in the chamber pressure. During this transient, oscillations in chamber pressure were observed. Several tests were conducted at ground level to determine the driving mechanism for the instability. The cause was seen to arise from flow separation in the injection orifice as given in the following.

Figure 4 shows the discharge coefficient (Cd) measured for a sharp-edged orifice at different values of Reynolds number. The orifice had a length to diameter ratio of 1.8. It is observed that as the flow through the orifice increases, the Cd value remains reasonably constant as shown by the portion AB in Fig. 4. At B, the Cd suddenly drops as shown by line BC due to onset of cavitation and the flow separating from the walls of the orifice. For subsequent increase in flow the Cd remains at the lower value (line CD in Fig. 4) due to the separated nature of flow [2].

Once the smaller value of Cd is reached, and the flow is reduced, the value of Cd remains consistently at a lower value as shown by line CE in Fig. 4 and does not return to the higher values. A zone of hysteresis is thus formed and two flow rates can be obtained for a given pressure drop corresponding to the two values of Cd at a given Reynolds number. These values depend on whether the flow is attached as obtained during the ascending mode or detached as during the descending mode.

The hysteresis can lead to perturbations in the flow and has been described in literature as hydraulic flip. Any hysteresis phenomenon could be a source of oscillation since two states are possible. In the upper stage liquid engine which showed the combustion anomaly and oscillations during the anomaly (Fig. 3) , the orifices used for injecting the fuel into the combustion chamber

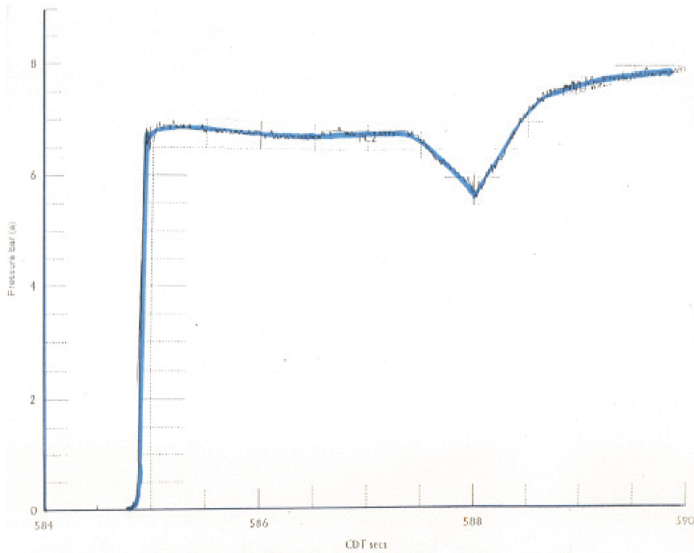


FIG. 3. OBSERVED PRESSURE-TIME TRACE

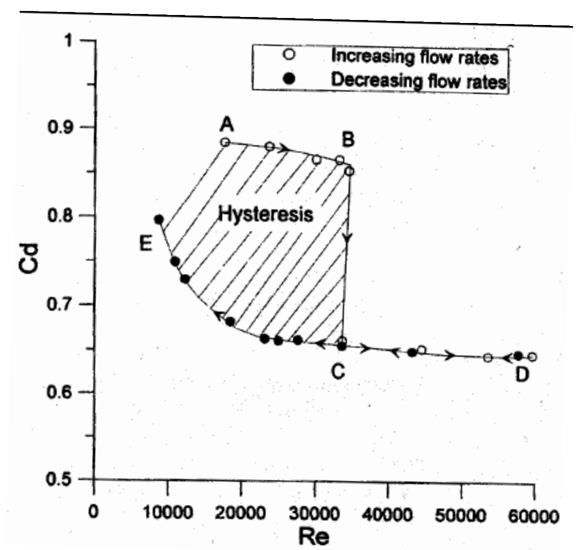


FIG. 4. HYSTERESIS BEHAVIOR OF DISCHARGE COEFFICIENT

had a length to diameter ratio of about 1.7. The initial flow was cavitated (detached flow with lower Cd) and this gave a smaller value of chamber pressure than the design value. The transition to attached flow, which happened about three seconds after ignition gave rise to minor chamber pressure oscillations and a dip in the chamber pressure from the poor spray which is formed in the

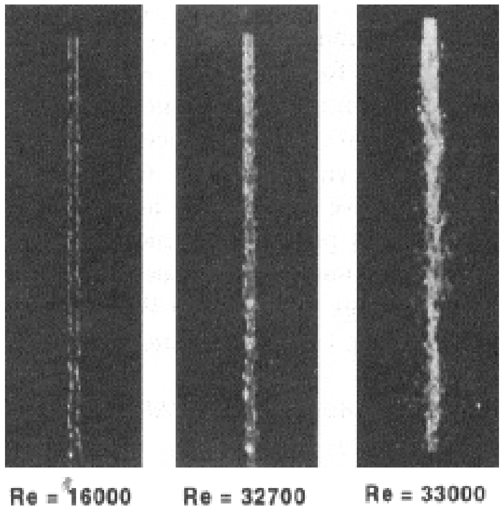


FIG. 5. DISTURBED JET FORMED DURING TRANSITION FROM DETACHED TO ATTACHED FLOWS

disturbed zone of transition before reaching the steady state value. The disturbed nature of a liquid jet formed during such transition process, as determined with a water jet, is shown in Fig. 5.

When the flow reattached in the orifice, the design value of chamber was achieved with a delay. Ground tests done subsequently showed coupling of the disturbances with the tangential mode of oscillations in the chamber during the transition probably due to mixture ratio striations from the poor spray. The problem could be corrected by modifying the length to diameter ratio of the orifice such that the transition between attached and separated flow do not occur. The transition times tallied reasonably with the characteristic time required for flow to relax in the feedline and the regenerative cooling path [3].

4. DAMPING OF OSCILLATIONS

The acoustic waves formed from small fluctuations or perturbations in combustion process get reflected at the walls and other boundaries of the combustion chamber. The interaction of the incident and reflected waves leads to the formation of standing waves in the chamber and these are characterized by nodes and antinodes at fixed locations. Maximum variations occur at the antinodes whereas at the nodes no variations take place. The nature of the standing waves formed is similar to that in an organ pipe except that the standing waves could also be in tangential, radial and mixed modes.

If significant energy release from combustion takes place at locations in the region of antinodes in phase with the perturbations, it is quite possible that the pressure perturbations already present at the anti-node regions can grow further when the combustion responds at the frequency of the standing wave. In liquid propellant rockets and gas turbine combustors, fuel is injected at the end-walls of the chamber which corresponds to the region of pressure anti-node. When spontaneous combustion takes place at the injector-end of the combustor, the energy release from the combustion could augment the amplitude at this anti-node region of the standing wave. This could lead to a growth in the amplitude of the oscillations and hence combustion instability. On the other hand, if the energy release from combustion is in the region of the nodes, there is no significant pressure perturbation present at this location which can interact with combustion. Instability is therefore not to be anticipated. Distributing the combustion heat release away from the pressure antinodes by having larger droplets in the spray or by reducing the reactivity are therefore means of suppressing the combustion instability. The response of the combustion process at the frequencies of the standing wave modes is an important parameter and needs to be negligibly small.

Rather than employ the reduction of the combustion response alone, the possibility of shifting the frequencies of the standing wave modes of the chamber such that they cannot interact with combustion or else increasing the factors contributing to the damping of the amplitude of the oscillations has been used for controlling the instability. These methods comprise of the use of baffles, Helmholtz resonators, quarter wave tubes or ablative chambers and have been discussed adequately in the literature [4]. The ability of small annular slots to suppress combustion oscillations has also been studied in liquid propellant rockets and implemented. The slots are generally provided at the corner of the injector/combustion chamber interface viz., at the pressure antinodes for the different modes of oscillation. These are known as corner wall slots and are either radial or axial. They dissipate the acoustic energy at their characteristic frequencies.

In a large thrust liquid propellant engine Viking, with a thrust level of 700 KN, (Fig. 6) radial annular slots have been provided towards the nozzle end of the chamber to increase the stability margin of the engine. This is done in addition to distributing the combustion zone away from the antinodes. The provision of the slots at the nozzle-end cannot be expected to explicitly influence the combustion response since combustion takes place predominantly near the injector. If the characteristic frequency of the chamber is shifted away from the frequency at which maximum combustion response is obtained, then the slot would give rise to an improved stability. A series of cold-flow tests was done with the acoustic model of this high thrust engine to determine the influence of the slots on the standing waves formed in the chamber. The slots were provided at the nozzle-end portion in the acoustic model as in the Viking engine. The model chamber was excited at different frequencies and also at random frequency using a speaker. The sound pressure level and the characteristic frequencies in the chamber were determined using microphones.

The measured sound pressure spectrum in the chamber, when slots were not provided, is illustrated in Fig. 7. The characteristic frequencies corresponded to the first longitudinal and tangential modes, the combined first longitudinal and radial mode, the second tangential mode and the first radial at 670, 1370, 1525, 2275 and 2855 Hz respectively.

Figure 8 illustrates the influence of the slots on the amplitude of the oscillations in the region near to the characteristic frequencies of 670 and 1370 Hz corresponding to the first longitudinal and first tangential modes. The slots had an axial width between 0.5 and 2 mm and a radial depth between 4 and 12 mm. The Sound Pressure Level (SPL) is given in volts and is proportional to Pascals. The suppression of the amplitudes was hardly seen in the first longitudinal mode for all dimensions of the slot used. A small reduction in amplitude is seen for the first tangential mode. However, for the higher frequencies near to the tangential modes of the cavity a more significant reduction was seen. This is shown in Fig. 9.

FIG. 6. 700 KN ENGINE

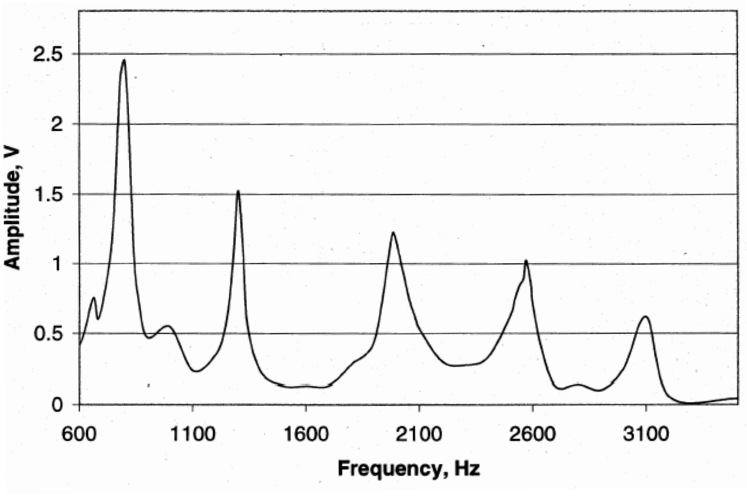


FIG. 7. CHARACTERISTIC FREQUENCIES OF THE COMBUSTION CHAMBER

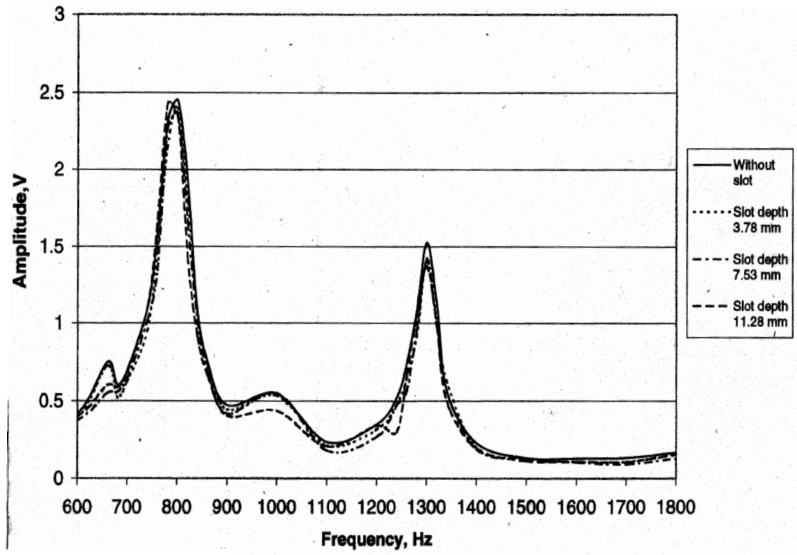


FIG. 8. DAMPING OF AMPLITUDES AT LOW FREQUENCIES

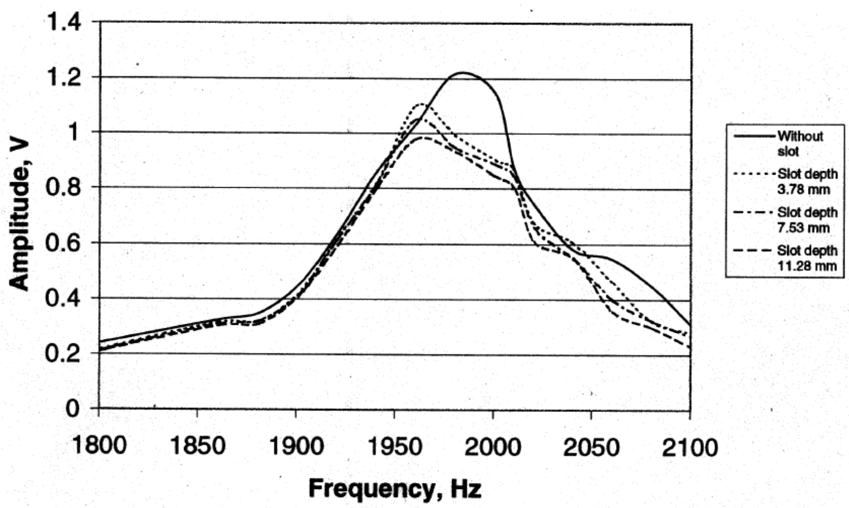


FIG. 9. DAMPING OF AMPLITUDES AT HIGH FREQUENCIES

It is also observed from Fig. 9 that whenever the amplitude of oscillation got significantly reduced due to the provision of the slot, a shift of the frequency towards a lower value occurs. The frequency shift suggests that the annular slot helps to slightly detune the combustion chamber rather than damp the amplitudes alone. Variations of the radial depth of the slot between 4 and 12 mm and the slot widths between 0.5 and 2 mm did not have a significantly influence the reduction

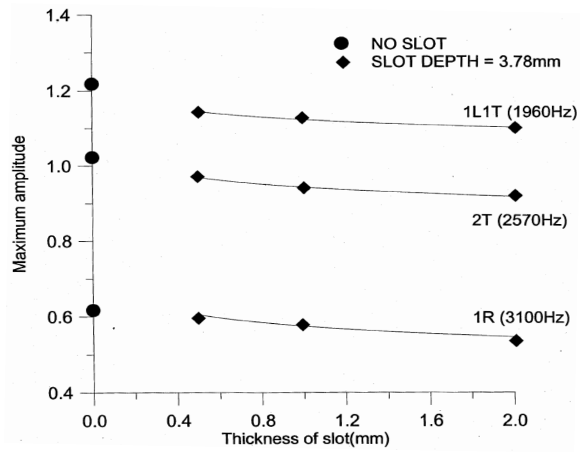


FIG. 10. INFLUENCE OF SLOT DEPTH AND WIDTH ON AMPLITUDE OF THE STANDING WAVES

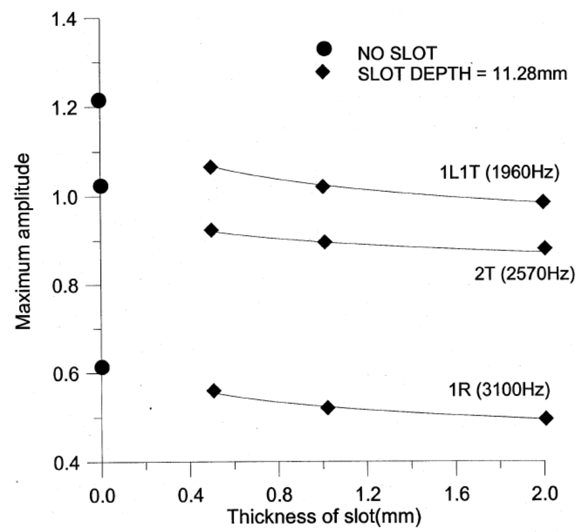


FIG. 11. INFLUENCE OF SLOT DEPTH AND WIDTH ON AMPLITUDE OF THE STANDING WAVES

of the amplitude of the oscillations and these findings are illustrated in Figs. 10 and 11. The insensitivity of the slot width and depth is to be anticipated if the nature of the oscillations in the slot is in the tangential direction. A reduction in the amplitude of some of the higher standing wave modes of the chamber through annular slots was seen to be possible when the tangential frequency associated with the slot is near to these higher frequencies of the chamber. The sizing of the width and depth of the slot does not appear to be important as long as these are very much smaller than the mean radius of the chamber at which the annular slot is provided. The radius at which the slot is provided governs its tangential frequency and is therefore important.

5. TONES IN FLOW AND STABILIZATION

Premixed and diffusion jet flames either anchored to the burner rim or lifted away from the rim are used in several applications. The controlled experiments of Herding et. al [5] using liquid oxygen injected in a co-annular flow of gaseous hydrogen show the flame to be anchored to the injector rim. The nature of flame is illustrated in Fig. 12. The flame begins at the tip of the oxygen orifice and becomes thicker and wrinkled in the downstream region. The co-flowing gas of hydrogen is warmer than the liquid jet and the hydrogen gas cools down adjacent to the liquid jet. The flame gets anchored in this low velocity regions. A decrease in the hydrogen gas temperature could reduce the cooling and decrease the velocity and result in the flame being lifted away from the injector rim.

Experiments with diffusion flames show a lifted flame to be characterized by a significant increase in the noise level unlike the low-level wide band noise of the anchored flame. The additional time during which the fuel and oxygen spend together prior to ignition in the lift-off or stand-off phase provides an incubation time for any incipient disturbances to grow. Lifted flames also have hysteresis. These two factors make the lifted flames more prone to combustion instability. In fact, it is known that a low value of hydrogen gas temperature (less than about 110 K) in certain injector configurations could lead to instability and we noted earlier the possibility of the low hydrogen temperature to favor the formation of a lifted flame.

Studies on interaction of high frequency sound waves with jet flames have demonstrated a significant decrease in the noise generated by the flame [6]. Experiments in which pure tones at high frequencies are imposed in the injector flow also demonstrate a reduction in the lifted jet flame noise [7]. High frequencies can be superimposed in the injector flow by positioning a small cavity upstream of injection port. The dimensions of the cavity could be chosen so as to give standing wave modes at the high frequencies.

Figures 13 and 14 give the overall noise attenuation and the reduction in flame lift-off heights in jet diffusion flames determined by improvising a cavity upstream of the injection port [7]. The reduction is primarily from the better mixing due to the tones from the cavity. The attenuation at the

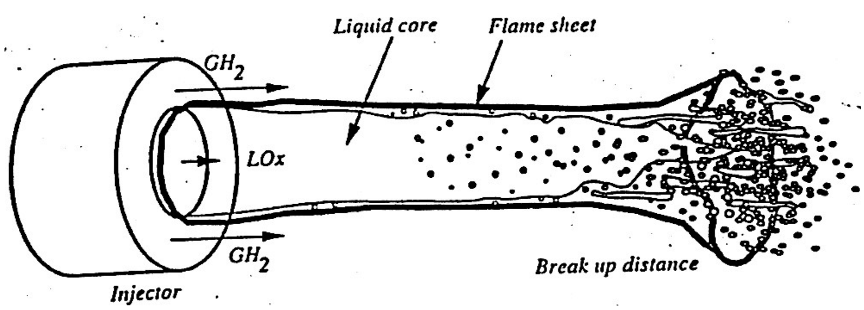


FIG. 12. TUBULAR FLAME ATTACHED TO INJECTOR BASE WITH COAXIAL INJECTION

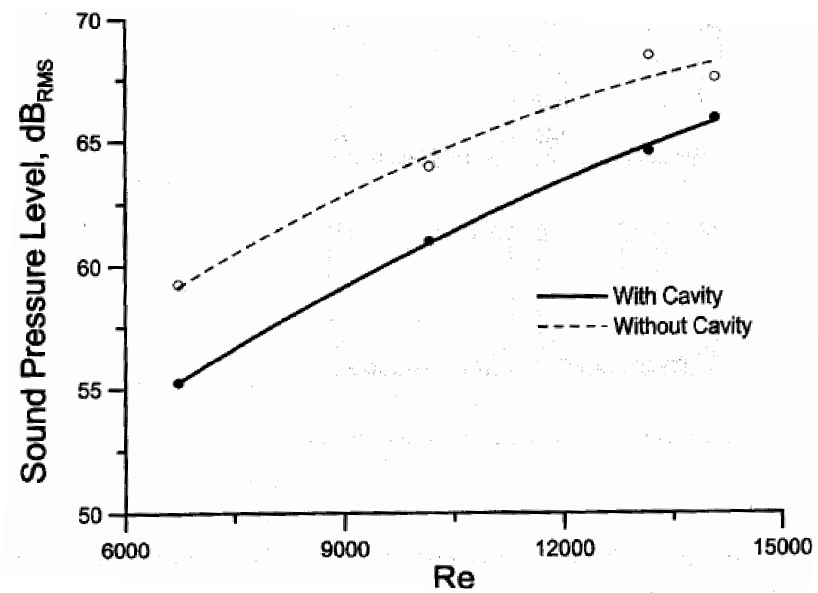


FIG. 13. REDUCTION OF SOUND PRESSURE LEVEL IN FLAMES FROM INDUCING FREQUENCIES IN FLOW BY A CAVITY

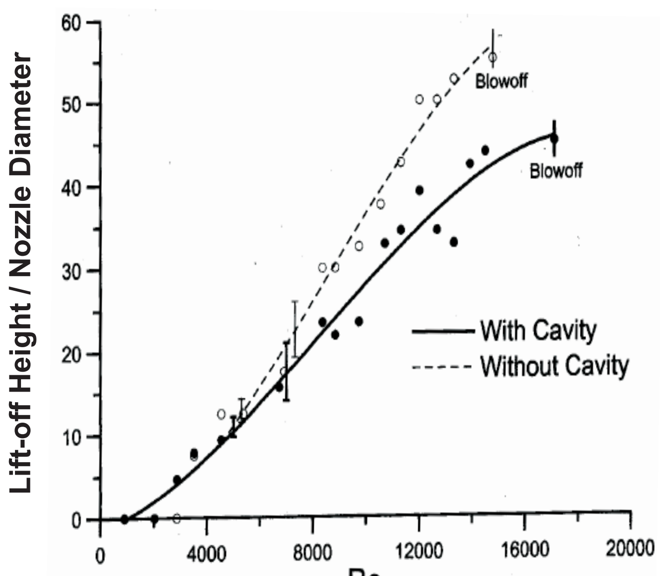


FIG. 14. REDUCTION OF FLAME LIFT-OFF HEIGHTS

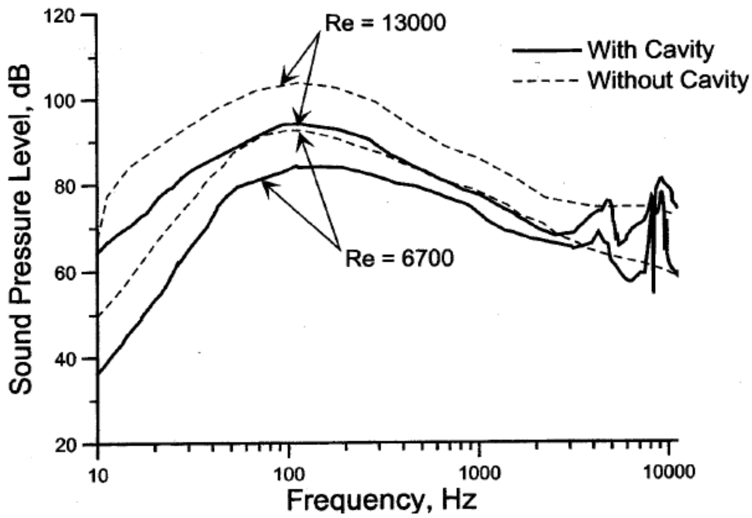


FIG. 15. ATTENUATION AT DIFFERENT FREQUENCIES

different frequencies is shown in Fig. 15. It is seen that significant amplitudes are obtained only at cavity frequencies. If the cavity frequencies can be kept away from the frequencies of the standing waves in the chamber, stabilization of combustion could be possible. This aspect needs further experimentation and demonstration.

Tones induced in the flow system are seen to be effective in reducing the amplitude of the oscillations and could contribute to stabilize the combustion.

6. CONCLUDING REMARKS

A few cases of combustion instability encountered with liquid propellant rockets and methods for suppressing the instabilities are examined. These illustrate some of the factors contributing to the combustion instability and methods for controlling the instability. The acoustics of the chamber and the feed system are examined. Tones induced in the flow system are shown to be important. While a low frequency of a canalization tube caused combustion instability in a cryogenic propellant rocket, high frequency from a cavity could lead to stabilization of the combustion. The large number of processes associated with injection, combustion and flow have different time scales and frequencies and these are influenced by the configuration and dimensions of the combustion system and the feed system. Being a system-level problem, no unique solution for control of combustion instability would be possible for the different combustors. The large variety of processes involved in the problem makes a generalized study difficult even though there are reasons for making judgments to eliminate the problem of instability. Recirculation flows with a plethora of eddies and associated length scales and frequencies offer scope of controlling the mixing and stabilizing combustion and studies for characterizing recirculation fields and applying them for combustion control would be useful.

ACKNOWLEDGEMENTS

R.K. Patnaik, T.Sharma and S. Venkatesh of the Liquid Propulsion Systems Centre, Trivandrum contributed to the different experiments and analysis.

REFERENCES

[1] D.T. Harrje and F.H. Reardon, Liquid Propellant Rocket Engine Combustion Instability, NASA-SP-194(1972)

[2] K. Ramamurthi, K. Nandakumar, S. Shankar and R.K. Patnaik, Hysteresis and Bifurcation of Flow in Fuel Injection Nozzles, J. Aerospace Engg., Proc. Inst. Mech. Engrs. London, Vol.215, 2000, pp. 49-59.

[3] K. Ramamurthi, K. Nandakumar and R.K. Patnaik, Two-step Start Transients with Long Feedlines Discharging Liquid Through Sharp-edged Cylindrical Nozzles, Atomization and Sprays, Vol. 12, 2002, pp. 193-207.

[4] E. Laudien, R. Pongratz, R. Pierro and D. Prelik, Experimental Procedures Aiding the Design of Acoustic cavities, Ch. 14, pp. 377-399, Progress Series in Aerospace Engg, (1993)

[5] G. Herding, R. Snyder, P. Scouflaire, C. Rolon, and S. Candel, Flame Stabilization in Cryogenic Propellant Combustion, 26th Int. Sym Comb., 1996, pp 2041-2047.

[6] F.E.J. Briffa and D.R. Romaine, Effect of Acoustic Fields on Flames, Combustion Science and technology, Vol. 7, 1973, pp. 93-98.

[7] K. Ramamurthi and R.K. Patnaik, Noise Reduction in Non-Premixed Lifted Jet Flames, Flow, Turbulence and Combustion, Vol. 72, 2004, pp. 49-67.

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

ELECTRICAL PROPULSION AND STATIONARY PLASMA THRUSTER

K. RAMAMURTHI*

1. INTRODUCTION

Electrical propulsion systems are being developed in the recent years for satellite propulsion, for attitude and orbit control of satellites and for planetary missions. The use of electrical propulsion for satellites is, however, not new. Electrically heated ammonia was proposed in 1971 for the Indian National Satellite (INSAT). The Application Technology Satellite (ATS-6) of Hughes, which was employed in India for the Satellite Instruction and Television Experiment (SITE) in 1974 to demonstrate television connectivity and health and educational programmes in remote villages, also used electrical propulsion. Here an electrostatic rocket of 4.5 mN thrust and jet velocity of 25000 m/s was used. Mercury was used as the working fluid. The first successful test flight of an electrical propulsion system (electrostatic using mercury) took place way back in August 1964 in Space Electric Rocket Test (SERT-1) wherein the electrical propulsion module was operated for 31 minutes in a sub-orbital flight.

The promise of electrical propulsion systems for space exploration was recognized ever since the scientific basis of rocket propulsion was understood [1]. Higher jet velocities are desirable for space missions since they provide a larger values of rate of change of momentum viz., specific thrust and very high jet velocities are possible by accelerating the gases electro-statically or electromagnetically. The early rocket pioneers Konstatin Tsialkowskii, Hermann Oberth and Robert Goddard considered electrical propulsion for interplanetary missions. Vladimir Glushkov in 1929 put together the first working hardware of an electrical rocket by generating hot vapor from electrically vaporized liquid metals or products of electrically exploded wires and expanding the hot vapor, so formed, through a nozzle.

Compared to chemical rockets for which the available energy is limited depending on the amount of fuel and oxidizer carried, the availability of energy for an electrical rocket is unlimited. This is because electrical power is available from solar energy in an orbiting or planetary spacecraft. Larger quantities of power, when required could also be generated by a nuclear power

*Ex. ISRO, Trivandrum. Presently Mechanical Engineering Dept., Indian Institute of Technology Madras, Chennai

plant in space. The available electric power is transferred to a working medium and the medium accelerated to high velocity using different physical processes. This feature makes the electrical rocket particularly attractive for space missions.

In this paper, we address the different types of electrical propulsion systems, their characteristics and limitations. The stationary plasma thruster (SPT) which uses Hall effect for the generation and movement of ions is gaining in popularity in the field of electrical propulsion. It is to be used in an upcoming mission of the geosynchronous satellite (GSAT) in India. This would therefore be dealt with in some detail.

2. CLASSIFICATION OF ELECTRICAL PROPULSION SYSTEMS

The classification is done into four categories depending on the process by which the electric power is used to generate high velocity of the working fluid.

i) <u>Electro-thermal</u>: Here the electrical power is used to heat the working medium (called as propellant) to form high temperature gases. The hot gases are expelled through a nozzle at high velocities. The system is similar to a chemical rocket except that the chemical heat release is replaced by the electrical heating. The electrical heating could either be resistive heating or arc heating.

ii) <u>Electrostatic</u>: The working medium is ionized and the ions are accelerated by electrostatic forces to high velocities - i.e., by applying a high potential.

iii) <u>Electromagnetic</u>: The ions are accelerated by an electromagnetic field i.e., using electrical and/or magnetic forces.

iv) <u>Pulsed Plasma</u>: Here the ions are generated in pulses from a solid such as Teflon using localized energy release near the surface of teflon(say by an electrical spark or otherwise). The ions so generated are accelerated by an electromagnetic field.

The number of operational satellites for which propulsion is done using the electro-thermal principle is the largest next to the satellites employing chemical propulsion. As on July 2005, about 150 satellites made use of electro-thermal principle [2]. Of these 115 used resistive heating while 35 used arc heating. There are about 40 satellites using electrostatic propulsion [2]. One satellite uses pulsed plasma propulsion. There is a growing trend towards use of Stationary Plasma Thruster (SPT) which is a particular class of electrostatic thruster and which would be discussed subsequently. As on May 2005, 14 operational satellites used the SPT [2] in addition to a large number of Russian satellites.

3. ELECTROTHERMAL THRUSTERS

The working medium is heated using electrical power and the hot gases, so generated, are expanded in a nozzle to generate high values of jet velocity. The heating could be done by electric resistive heating in which case the device is known as a resistojet. A schematic of the resistojet is shown in Fig. 1.

Different working fluids could be used – e.g., ammonia, hydrazine, water, hydrogen, carbon-dioxide. The maximum temperature to which the gases could be raised corresponds to the

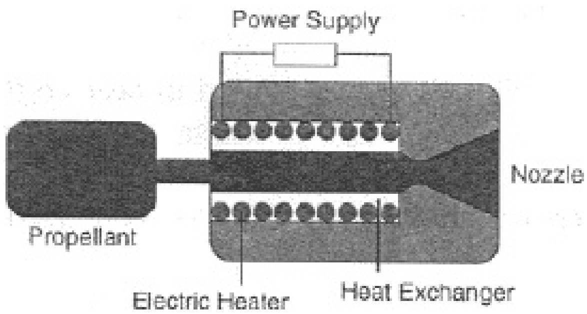


FIG. 1. SCHEMATIC OF A RESISTOJET [4]

temperature which the materials of construction can withstand. The upper limit of temperatures is about 3000 K. Resistojets have efficiencies around 80% and have been extensively used in low-Earth orbiting satellites.

When higher jet velocities are desired, the gas temperature needs to be higher. In this case, the working medium is heated using arc discharge. The discharge is obtained by striking an arc between a rod-shaped cathode in the chamber and an anode placed upstream of the nozzle throat. A typical arc discharge heated thruster, known as arc-jet is shown in Fig. 2. The temperature of the working medium are several tens of thousands K and the working medium gets ionized. The efficiency is lower than that of the resistojet.

A typical system using ammonia with storage tanks is shown in Fig. 3

When hydrazine is used in monopropellant thrusters, the hydrazine decomposes into ammonia, nitrogen and hydrogen:

$$3N_2H_4 \rightarrow 4\,(1 - x)\,NH_3 + (1 + 2x)\,N_2 + 6x\,H_2$$

Here x denotes the fractional dissociation of ammonia. As the dissociation increases, the temperature of the gases drops. Typically the level of dissociation for which the catalytic bed is designed is about 0.3 and the corresponding gas temperature is about 1400 K. It is possible to increase the gas temperature further by resistive or arc heating. Such a system of increasing the temperature of monopropellant combustion by electrical heating is known as Augmented Electro-

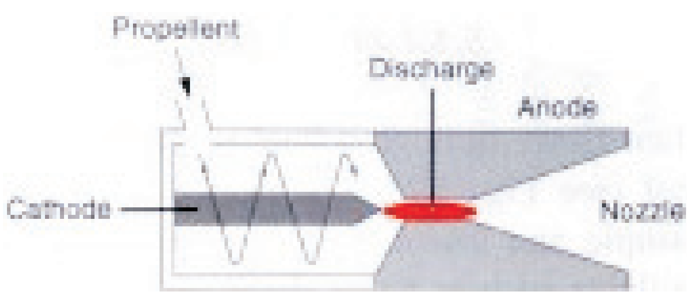


FIG. 2. ARC-JET (FROM [4])

FIG. 3. AMMONIA ARC JET WITH FEED SYSTEM

FIG. 4. AUGMENTED ELECTRO-THERMAL HYDRAZINE THRUSTER

Thermal Hydrazine Thruster (AEHT). These have been flown in several satellites. Figure 4 shows the AEHT hardware.

The viscous boundary layer losses during the nozzle expansion, especially with the low density flows, reduce the performance. Figure 5, taken from Boyd [2], shows typical Mach number contours in a hydrogen resistojet and the loss in the overall velocity. The partitioning of energy into the rotational and vibrational modes at the higher temperatures of the arc-jet and freezing of it during the expansion causes additional losses.

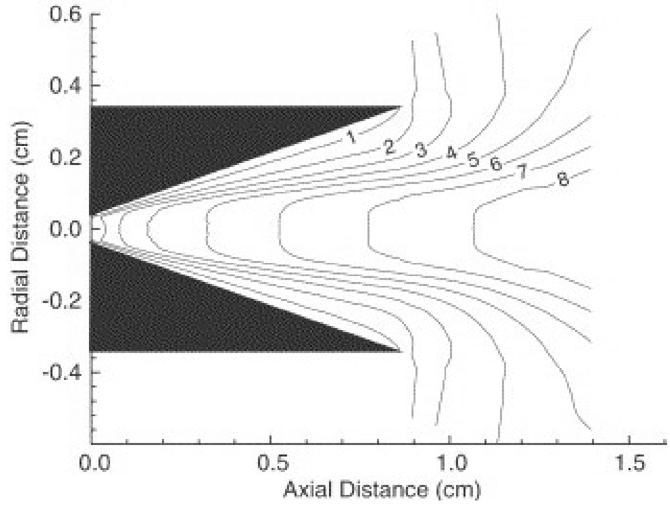


FIG. 5. ILLUSTRATION OF VISCOUS BOUNDARY LAYER LOSSES (FROM [2])

4. ELECTROSTATIC AND ELECTROMAGNETIC THRUSTERS

4.1 Electrostatic and Electromagnetic Forces

a) Electrostatic

Coulomb gave the law between electrically charged particles in a form identical to Newton's law for Gravitation. The law states that the force between a pair of point charge is directly proportional to the product of the charges and inversely proportional to the square of the distance between them.

$$\vec{F} = C\,\frac{q_1 q_2}{\left|\vec{r}^{\,3}\right|}\,\vec{r} \tag{1}$$

Here q_1 and q_2 are in Coulombs, r is in m and the proportionality constant C is

$$C = 9 \times 10^9\ \text{Nm}^2/\text{coulomb}^2$$

C is often written as $\frac{1}{4\pi\varepsilon_0}$ where ε_0 is defined as the electrical permittivity of free space and $\varepsilon_0 = 8.852 \times 10^{-12}$ Farads/m. The unit of Farad as seen from Coulomb's law is seen as follows; C has units of $\text{Nm}^2/\text{Coulomb}^2$ giving the unit of ε_0 as $\text{Coulomb}^2/\text{Nm}^2$. Farad is defined as $\text{Coulomb}^2/\text{Nm}^2$ ($1\ \text{F} = 1\ \text{Coulomb}^2/\text{Nm}^2$). The permittivity is a physical quantity that determines the ability of the material to allow the charge to be stored in it. It is related to the electrical field which is defined in a manner analogous to the gravitational field about a heavy object. A region surrounding a charged particle is said to have an electrostatic field if a charge placed there experiences a force. The electrostatic field from Coulomb's law is:

$$\vec{E} = \frac{Cq_1}{\left|\vec{r}^{\,3}\right|}\,\vec{r} \tag{2}$$

If a charge q_2 is placed in the field

$$\vec{F} = \vec{E}\,q_2 \tag{3}$$

the unit of electric field $\vec{E}$ being N/Coulomb = N-m / Coulomb-m = volt/m. Permittivity depends on the field and the characteristic of the material. A term dielectric constant is defined which gives the ratio of the permittivity in the material (ε) to the permittivity of free space (ε_0).

b) Magnetic

In electro-magnetism, the ability of materials to magnetize in a given magnetic field (equivalent to holding a charge) is defined in terms of permeability μ which is the ratio of the flux to the field. The permeability of free space μ_0 is related to ε_0 through the expression:

$$\mu_0 = \frac{1}{c^2\varepsilon_0} \tag{4}$$

Here c is the velocity of light in m/s. The unit of permeability is Henries. The permeability of free space (vacuum) is 1.257×10^{-6} Henry/m ($4\pi \times 10^{-7}$). The relative permeability of a material is given by $\mu_r = \mu/\mu_0$. The unit of Henry is

$$\frac{Nm^2}{Coulomb^2}\frac{s^2}{m^2} = \frac{N}{A^2}$$

If a charge moving in space experiences a force, but does not experience the force if it is at rest, then the region in space is said to have a magnetic field. The magnetic flux is denoted by symbol B and is a measure of the quantity of magnetism taking into account the strength and the extent of the field. With the field having units of V/m, the unit of magnetic flux is (v/m) (m^2)/ (m/s) = Volt.sec. This unit is called Weber (Wb). The magnetic flux density is Wb/m^2 and is known as Tesla (T). Hence 1 T is given by:

$$1\frac{Wb}{m^2} = 1\frac{Vs}{m^2} = 1\frac{N}{Coulomb} \div \frac{m}{s} = 1\frac{N}{Am}$$

In CGS system, the unit used for magnetic flux density is Gauss with 1T = 10,000 Gauss. The unit Oersted has also been used for magnetic field strength with 1Oersted being equal in vacuum to magnetic flux density of 1 Gauss.

The force on a charge moving in a magnetic field with velocity $\vec{v}$ is given by:

$$\vec{F} = q\vec{v} \times \vec{B} \tag{5}$$

where $\vec{B}$ is the strength of the magnetic field in Weber/m^2(T). The Earth's magnetic field is typically about ½ Gauss i.e., 0.5×10^{-04}T and inadequate to give force unless high values of electromagnetic field is extraneously introduced. The possibilities of magnetic propulsion have therefore not so far been successful using the natural magnetic fields available in interstellar cosmic space. Though the intensity is small, it is extensively available in space and further innovations are required to apply it in practice [3].

c) Combined electrical and magnetic fields

In the presence of both electric and magnetic fields the force experienced by a moving charge is:

$$\vec{F} = q\vec{E} + q\vec{v} \times \vec{B} \tag{6}$$

This is known as the Lorentz force equation where $\vec{F}$ is the force in N, q is the charge in Coulomb, $\vec{E}$ is the electrostatic field in F, $\vec{v}$ is the velocity of the charge in m/s and $\vec{B}$ is the strength of the magnetic field in T. The exit velocity of a charged particle can be determined from the above force equation.

5. ELECTROSTATIC PROPULSION

These are electrostatic propulsive devices. Charged particles are generated and accelerated in a strong electric field generated by high voltage difference between electrodes. Charged particles comprise positively charged ions and electrons. The electrons have a mass of about 10^{-30} kg while a proton is heavier by some 1840 times. The atomic mass of a substance is equal to its atomic weight divided by Avagadro's number. The atomic weight and mass of an atom of some substances are given below:

Table I. *Atomic Mass of Some Substances*

S. No.	Substance	Atomic Weight	Mass of an atom
1	Nitrogen	14	2.32×10^{-26} kg
2	Hydrogen	1	0.167×10^{-26} kg
3	Mercury	200	33.2×10^{-26} kg
4	Cesium	133	22.1×10^{-26} kg
5	Xenon	131	21.8×10^{-26} kg

Heavy substances such as mercury, cesium or xenon are to be preferred as working substances if a significant value of thrust is required. The requirement is therefore to generate positive ions of these substances and accelerate them to high velocities in a high intensity electric field.

Figure 6 gives a schematic of the device. The positive ions from the ionized gases are attracted and accelerated by the accelerating grid as shown. These ions leave at high velocities to produce the thrust.

The departure of the positive ions from the device will cause the satellite to become negatively charged and attract the ions. Electrons are injected into the exhaust of the positive ions to maintain the spacecraft potential and form a neutral plasma. If only the positive ions were sent out at high velocities, the strong positive potential in front of the device would prevent outflow of the ions. The neutralizing of the charge particles by electrons, as shown in Fig. 6, is therefore important.

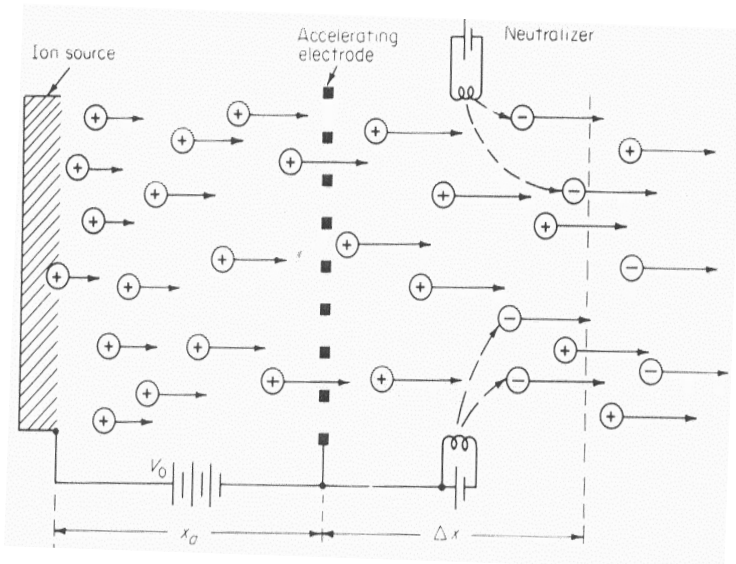


FIG. 6. SIMPLIFIED SCHEMATIC OF ELECTROSTATIC PROPULSION

In practice, the propellant gas (working medium) flows into a hollow cathode where it meets the electrons. Low work function materials (Work function denotes the minimum amount of energy needed to remove an electron from a metal) are used for the cathode. The material could be tungsten (4.5 ev), molybdenum (4.57 ev), silver (4.26 ev) or some alkali metal which has small values of work function. The metal is heated and electrons are released by thermionic emission. The electrons are released in the gas and ionize it. The motion of the electrons from the cathode to the anode is further controlled by providing a magnetic field so as to increase the path length of the electrons. This ensures improved contact between the electrons and the atoms and the bombardment of the neutral atoms of the gas by the electrons and ionizing it. A lower ionization potential for the gas is to be preferred. The first ionization potential in ev for cesium, mercury, xenon, krypton and argon are 3.89, 10.44, 12.13, 14 and 15.80 respectively.

The positive ions, formed during the ionization, drift towards a set of grids- extractor grid and acceleration grid - across which a large potential difference is applied in order to accelerate the ions. A large number of fine holes are provided in the grids and these holes focus the path of the ions. Since grids are used for the extraction and acceleration of the ions, the device is known as gridded ion thruster. Figure 7 shows the constructional features of the thruster. Electrons are subsequently provided for neutralization of the ion beam. The constructional features of the thruster are shown in Fig. 8.

5.1 Performance of Ion Rocket

For ions having mass m kg and charge q Coulomb in an electrical field E volts/m, the force experience by the ion is:

$$F = q\,E = q\,V/L$$

where, V is the voltage difference between electrodes separated by a distance L.

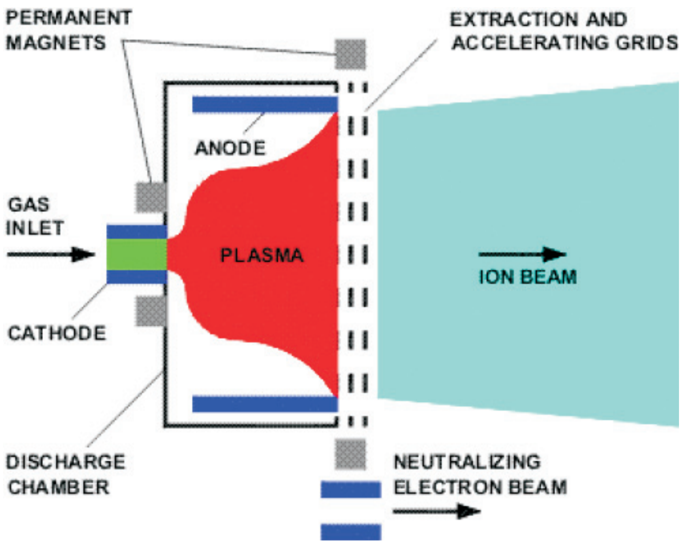


FIG. 7. SCHEMATIC OF GRIDDED ION THRUSTER

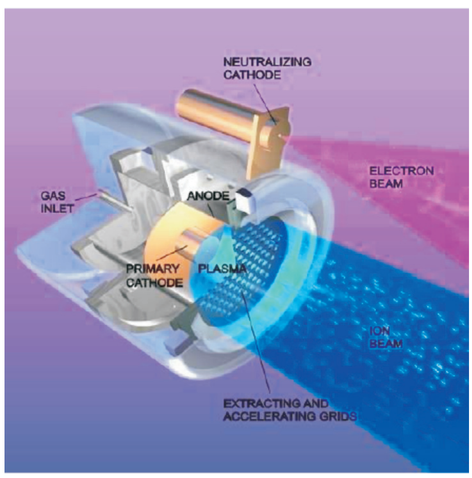


FIG. 8. CONSTRUCTION OF ION ROCKET

The work done by the charge when it moves a distance L is FL = qV J
This will equal the increase in the kinetic energy ½ mV_j^2, giving

$$\frac{mV_j^2}{2} = qV \quad \text{or} \quad V_j = \sqrt{2qV/m} \tag{8}$$

The velocity acquired by the ion therefore scales as $(q/m)^{1/2}$.

The force can be written as rate of change of momentum:

$$F = \dot{m}\,V_j = m\,\frac{I}{q}\,V_j$$

giving

$$F = \mathrm{I}\,\sqrt{2Vm/q} \tag{9}$$

The force or thrust therefore scales as $(m/q)^{1/2}$.

From equations (8) and (9), we see that an ion having a small mass would give higher jet velocities whereas a heavier ion would provide better thrust. We also see the need to have a large beam current.

The power developed is:

$$\frac{\dot{m}V_j^2}{2} = \mu\,P = \mathrm{hP} = \mathrm{hV\,I} \quad \text{giving}$$

$$V_j = \sqrt{2\eta VI/\dot{m}} = \sqrt{2\eta qV/m} \tag{10}$$

As the beam current I increases, the beam current density (j = I/A) would increase for given thruster. The field E which extracts the charge at the charge emitting surface decreases as the charge density increases. This is because the electric field of the ion cloud opposes the electric field of acceleration and revises the potential in the gap. As the current density increases the extraction gradient ultimately becomes zero and no flow of ions is possible. This condition is known as the space-charge limited current for a given field E i.e., potential difference V and length L. The Child Langmuir law gives the space charge limited current density as:

$$j = \frac{4\varepsilon_0}{9}\sqrt{2q/m}\,\frac{V^{3/2}}{L^2} \tag{11}$$

Here ε_0 = permittivity of free space and V = voltage between two electrodes L m away. The thrust per unit area from Equation (9) is:

$$\frac{F}{A} = j\sqrt{2Vm/q}$$

For the limiting (maximum) current density, the value becomes:

$$F/A|_{\max} = \frac{8}{9}\,\varepsilon_0\,\frac{V^2}{L^2} \tag{12}$$

The maximum thrust per unit area depends on the average field intensity. With a beam current of diameter d:

$$F = \frac{2}{9}\,\pi V^2 R^2 \tag{13}$$

where R=d/L and is called the beam aspect ratio.

From Equation (8), $V_j = \sqrt{2qV/m}$, which when substituted in the above gives:

$$F/A|_{\max} = \frac{2}{9}\frac{\varepsilon_0}{L^2} V_j^4/(q/m)^2 \tag{14}$$

The strong dependence of thrust per unit area is seen for the case of the space charge limited condition. It is also seen necessary to reduce the electrode spacing as much as possible. The spacing is about 0.7 mm. The limit is about 0.5 mm and accelerating potentials are about 1.5 to 2.5 kV.

The beam current can be increased for a given V_j by using an acceleration – deceleration system with three electrodes (grids) that allow operation at maximum potential gradient. Typical values of V_j are about 25,000 m/s and thrust between 20 and 200 mN. Molybdenum is used for the grids due to its low sputter. The cathode comprises low work function materials. Considering the contamination problems with mercury and cesium, an inert gas such as xenon is more widely used for the working medium.

The satellite Deep Space which flew by two asteroids and a comet (1998-2001) used the ion thruster. Figure 9 shows the ion thruster of Deep Space.

5.2 Hall Effect and Stationary Plasma Thruster

The space charge limited current density limits the thrust produced per unit area in the ion thruster. The extraction and acceleration grids in the ion thruster considered in the last section are subject to sputtering and lead to a restricted life of the ion thruster. These limitations are overcome using the Hall effect. *Hall effect states that if a current flows in a conductor (or semiconductor) and there is a magnetic field perpendicular to the current flow, then the combination of current and magnetic field will generate a voltage perpendicular to both.* The Hall effect is used to trap the electrons and use them to efficiently ionize the propellant. A schematic of the arrangement is given in Fig. 10.

An annular anode chamber is used in which a radial magnetic field is induced, as shown, especially at towards the end of the chamber. The cathode emits electrons which are attracted towards the anode. The current from the flow of electrons in the radial magnetic field produces an

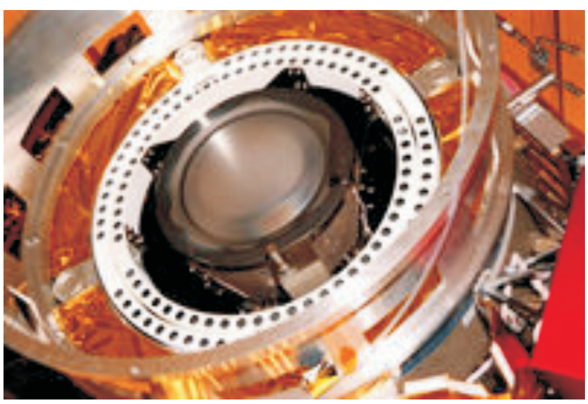


FIG. 9. ION THRUSTER USED IN DEEP SPACE 1

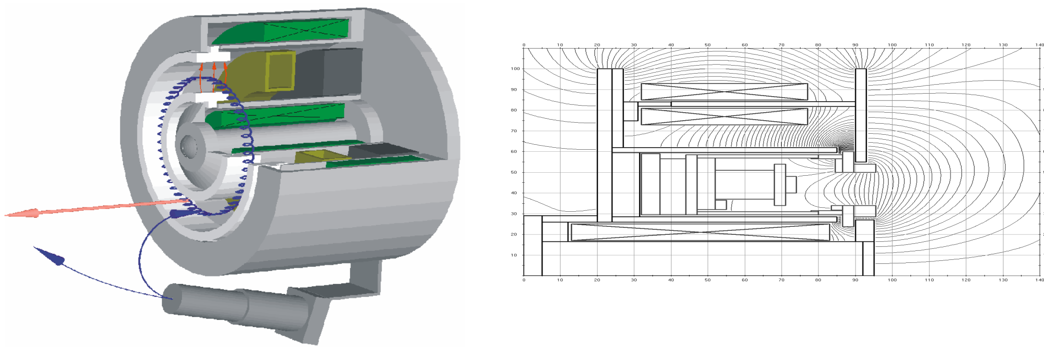


FIG. 10. SCHEMATIC OF HALL EFFECT THRUSTER

azimuthal voltage which causes the low-mass electrons to rotate. A propellant gas, typically xenon, is admitted in the anode chamber. The spirally electrons ionize the gas efficiently, in view of the large residence time available for the interaction from the spiral flow of electrons. While the electrons get deflected by the Hall effect, the heavy positive ions are accelerated by the electrostatic potential to produce thrust. The acceleration of the ions takes place in a neutral plasma and the limitations of space charge limited current density do not exist.

Typically a voltage of 300 Volts is applied between the anode and cathode. About 30% of the discharge current is the electron current that does not produce thrust. The efficiency of the thruster is therefore limited to less than 70%. The exit velocity of the ions is between 12,000 and 18,000 m/s. The magnetic field near the exit of the chamber is about 0.04 T and the discharge current about 4 A. Typical mass flow rates are a few mg/s. The ions are neutralized by electrons from the cathode as shown in Fig. 10.

The Molnya satellites of Russia have traditionally used the Hall Effect Thrusters. They are called as Stationary Plasma Thrusters. More than 200 such thrusters have been flown in over 100 Russian satellites and about 15 American and European satellites. The defect with these thrusters is the wide spread of the plume compared to gridded ion thruster. The impingement of the plume can cause surface heating and contamination. In addition, the collision of the ions with the walls results in a loss of momentum and charge. This is less with anode layer thruster (TAL) which uses a shorter length of the acceleration chamber. The walls of TAL are of molybdenum while that of SPT are of ceramic (boron nitride or silicon carbide). The front end of a SPT is shown if Fig. 11. The ceramic annular chamber and the cathode are seen.

6. ELECTROMAGNETIC PROPULSION

In electromagnetic propulsion, the acceleration of ions is achieved either by an externally applied magnetic field or by a self-induced magnetic field. The Magneto-Plasma-Dynamic (MPD) thruster uses a self-induced magnetic field for the acceleration. As indicated in Fig. 12, a high current discharge ionizes the propellant flow between the electrodes. The current induces a magnetic field

FIG. 11. STATIONARY PLASMA THRUSTER

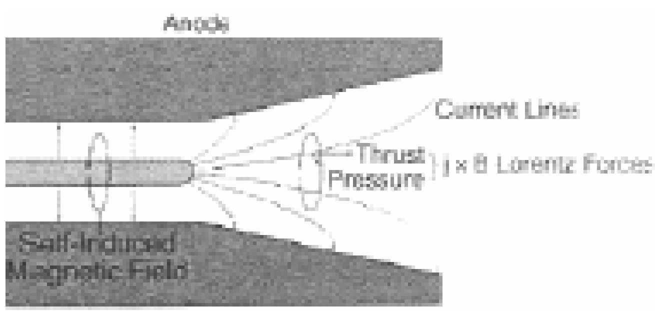


FIG. 12. ELECTROMAGNETIC THRUSTER (FROM [4])

$\vec{B}$. The interaction between the magnetic field and the current generates a force $\vec{I} \times \vec{B}$ perpendicular to the direction of I and B.

When the magnetic field is induced by the current, the device is known as a self-induced MPD thruster. When permanent magnets or coils around the electrode are used to enhance the magnetic field, the device is called as applied field MPD thruster.

Instead of operating in a steady mode, a pulse or puff of plasma could be generated by triggering a rapid electrical discharge across the surface of a solid fuel such as Teflon. Carbon and fluorine atoms and ions in the puff (from Teflon) are accelerated – the ions electromagnetically and the atoms gas dynamically. Figure 13 illustrates the working of a PPT. Very small impulse bits are obtained by such PPT. The construction is simple and PPTs are strong contenders for small spacecrafts.

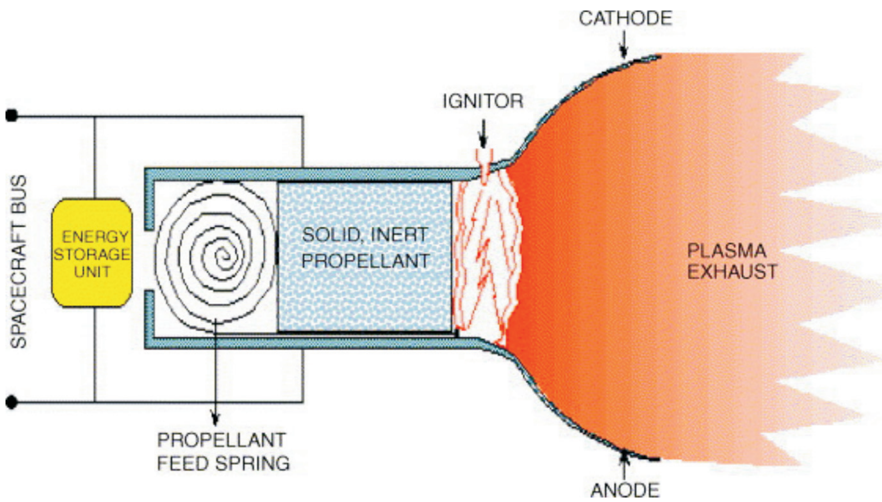


FIG 13. PULSED PLASMA THRUSTER (FROM REF[2])

7. CONCLUDING REMARKS

The different principles of operation of electrical thrusters are reviewed. The high performance and long life of the stationary plasma thruster and the relatively larger thrust achievable with this thruster has focused the new developments in this area. The work in electro-thermal thrusters has come down. The higher performance of the ion thrusters (both V_j and efficiency) compared to the stationary plasma thruster makes it a strong contender; however, the life time of the extraction and acceleration grids and the limitation in the magnitude of thrust are factors that inhibit the growth potential. The pulsed plasma thruster, being simple in construction, has promise for small satellite applications.

Several methods of generating ions other than those discussed in this report have been applied [5]. These include (i.) ionization using radio frequencies of several MHz. (Radio-Ionization Thruster RIT), (ii.) use of colloids with electrical field to generate charged particles of colloid, (iii.) exposing surface of liquids to strong electric field to form charged particles and accelerate them with the same field (Field Emission Thruster – FEEP) and (iv.) laser-assisted plasma propulsion. Magnetic propulsion for interplanetary missions is still to evolve. The power plant of electrical propulsion systems tend to be heavy and optimization of the payload weight is required taking into account the power requirements for the propulsion system.

REFERENCES

[1] E.Y. Choueiri, A critical history of electric propulsion: The first 50 years, *J. Propulsion and Power,* vol. 20, 2004, pp. 193-203.

[2] I.D. Boyd, Numerical modeling of spacecraft electrical propulsion thrusters, *Progress in Aerospace Sciences,* vol. 41, 2005, pp. 669-687.

[3] V. Pulatov, Physics of magnetic propulsion, *Progress in Aerospace Science,* vol. 41, 2005, pp. 64-91.

[4] M. Tajmar, Advanced space propulsion systems, *Springe*r, Wein, 2003.

[5] K. Ramamurthi, Advances in rocket propulsion, *J. Inst. Engrs.,* vol 85, pp. 54-59, 2004.

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

PULSE DETONATION ENGINES – PROBLEMS OF INITIATING DETONATIONS IN HYDROCARBON-AIR MIXTURES

K. RAMAMURTHI*

1. INTRODUCTION

Initiation of detonations in Hydrocarbon-air mixtures remains a major "technological hurdle" [1] in practically implementing a hydrocarbon fuelled Pulse Detonation Engine (PDE). The energy required to initiate a detonation in a typical hydrocarbon-air mixture say a stoichiometric mixture of propane-air at atmospheric pressure is about 100 kJ compared to about 1 mJ required for the initiation of a flame in the mixture [2]. For ethylene-air mixture the minimum energy required is about 55 kJ [3]. Such large energy releases are inherently difficult in the pulsed mode of operation of a PDE. Instead of forming a detonation directly, which is difficult, the general effort has been to form a flame by an ignition source and then get the flame to transit into a detonation. However, this has not been easy and it has not been possible to have the transition in a small volume [1]. A large length scale of transition of the flame to detonation reduces the maximum pressure at the head-end of the PDE and hence the thrust of the PDE.

Schauer et al. [1] showed that a detonation could be formed by transition from a flame in stoichiometric mixtures of propane, JP-8 and JP-10 with air when blockages were provided in the detonation tubes. Blockages have the disadvantage of causing losses and therefore reduce the impulse of a PDE. It is desirable to devise improved methods of spontaneously forming a detonation in hydrocarbon-air mixture. The possibility of initiating a detonation in a hydrocarbon-air mixture using constant volume combustion products from a pre-chamber has not been reported so far and is investigated in this paper. For this purpose a better understanding of the detonation process in general and the formation of detonation in premixed air-fuel mixtures are essential.

The discussions in the paper are organized in six sections. The importance of the shock structure and the role of the transverse waves in forming a detonation are given in section 2. This is followed in section 3 by a short review of the different methods of initiating a detonation in premixed explosive gas mixtures. The theoretical model for the phenomenon is given in section 4 with

*ex. ISRO, Trivandrum 695547. Presently Department of Mechanical Engineering, IIT Madras, Chennai 600036

specific reference to the hydrocarbon-air mixture. The possibilities of forming a detonation using a hot jet from constant volume combustion of the hydrocarbon-air mixture in a pre-chamber in a PDE is given in section 5. The conclusions of the study are given in section 6.

2. ROLE OF TRANSVERSE SHOCKS IN THE INITIATION AND PROPAGATION OF DETONATIONS

The formation of detonation in premixed explosive gas mixtures has been extensively studied over the last fifty years starting with the model of Zel'dovich, Kogarko and Semenov in 1956 [4]. Detailed experiments and computational fluid dynamic modeling have also been carried out to identify the mechanisms involved in the process of initiation, propagation and limits of detonation. Both the experiments and computations have brought out the pivotal role of transverse shock waves in the three dimensional shock structure of the detonation in forming and sustaining the detonation. Figure 1, taken from Lee and Radulescu [5] illustrates the multiple shock structure of a detonation wave. The structure is obtained using schlieren and placing smoked foil on the walls. Transverse waves are seen behind the multi-headed shock front of the detonation. This structure under idealized conditions can be represented at different instants of time by the shocks and their interaction as shown schematically in Fig. 2. The interacting shocks form the cells of detonation.

ABCD in Fig. 2 denotes a cell of detonation. The length of the cell is L and its width is λ as shown. The cell boundaries are formed by the interaction of the transverse shocks with the leading

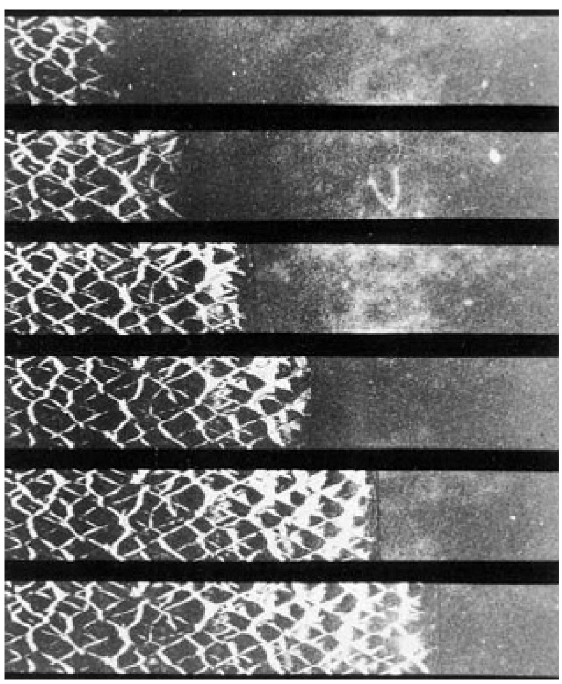


FIG. 1 MULTIPLE SHOCK STRUCTURE OF A DETONATION

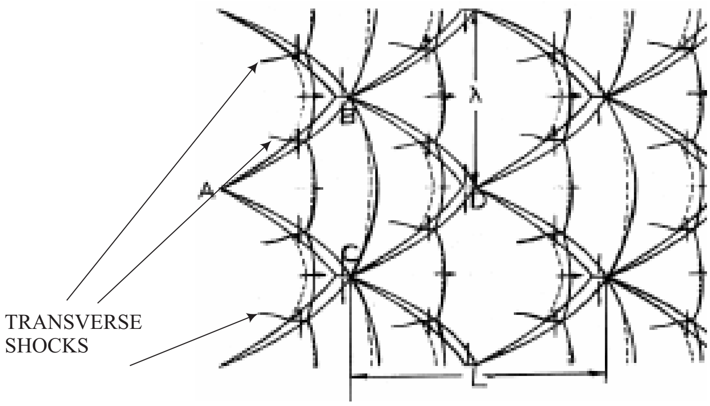


FIG. 2 TRANSVERSE SHOCKS AND THE DETONATION CELL

normal shock and correspond to the trajectory of the triple shock Mach interaction points. The width λ corresponds to the average distance between the triple point trajectories.

The shock front within a cell of detonation is continually decaying in strength (Fig. 3). Its strength increases abruptly at the end of the cell when the transverse waves from adjacent cells collide with each other.

This structure of a detonation and the associated length scales need to be kept in mind while determining a method of initiating a detonation and sustaining it. The presence of the transverse waves helps to create hot spots or local zones of intense chemical reactions at the end of a cell of detonation and it is this phenomenon which sustains a detonation. Experiments done specifically by damping the transverse waves through selective absorption at the walls [6] showed the detonation to decay. The interaction of the transverse waves creates conditions for chemical reactions to take place at a sufficiently fast rate because of the much higher temperatures from the Mach reflection of the interacting shocks than from the individual shocks. The higher rate of chemical energy release couples with the shock front to form a detonation. These aspects are to be considered while determining the requirements of initiating a detonation in a hydrocarbon-air mixture.

3. DIFFERENT METHODS OF INITIATING DETONATIONS IN PREMIXED GASES

Five different methods have been used for generating detonations in a premixed explosive gas mixture. These are:

i. Direct initiation of detonation at the ignition source
ii. Forming a flame or deflagration by the ignition source which transits to a detonation
iii. Using a hot gas jet or using a detonation in a more sensitive mixture to start a detonation in a less sensitive mixture
iv. Using a hypervelocity projectile

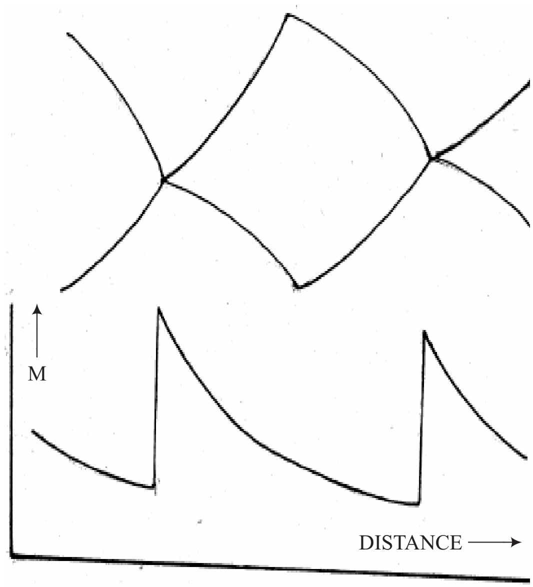


FIG. 3 DECAYING SHOCKS WITHIN DETONATION CELL (Y COORDINATE DENOTES SHOCK STRENGTH)

v. Depositing chain carriers in a certain region of the reactive gas mixture and initiating rapid chemical reactions.

The mechanisms involved in the formation of the detonation in the different methods are given below.

3.1 Direct Initiation at the Ignition Source

Direct initiation of a detonation refers to the instantaneous formation of a detonation by the ignition source. Either a strongly driven detonation known as an "overdriven detonation" is formed by the ignition source or else a strong blast wave (decaying shock wave) of some critical strength is formed such that it becomes a detonation after a certain distance of travel from the ignition source. In the overdriven detonation, the chemical reactions occur adjacent to the shock and drive it at speeds exceeding the Chapman Jouguet velocity. The shock-reaction complex of the overdriven detonation decays down subsequently to the multi-shock detonation wave which travels overall at the Chapman-Jouguet velocity. In the critical blast initiation mode, the blast wave formed by the ignition source is not initially coupled to the heat release from the chemical reaction. It decays in strength to a value of about half the Chapman Jouguet detonation velocity. After a short quasi-steady period of travel at this reduced speed, which provide the shocked gas conditions corresponding to auto-ignition of the gas mixture, localized explosions occur leading to the formation of a detonation.

The requirement that a shock wave of certain threshold value of strength be formed by the ignition source calls for a certain minimum rate of energy to be released by the igniter. The

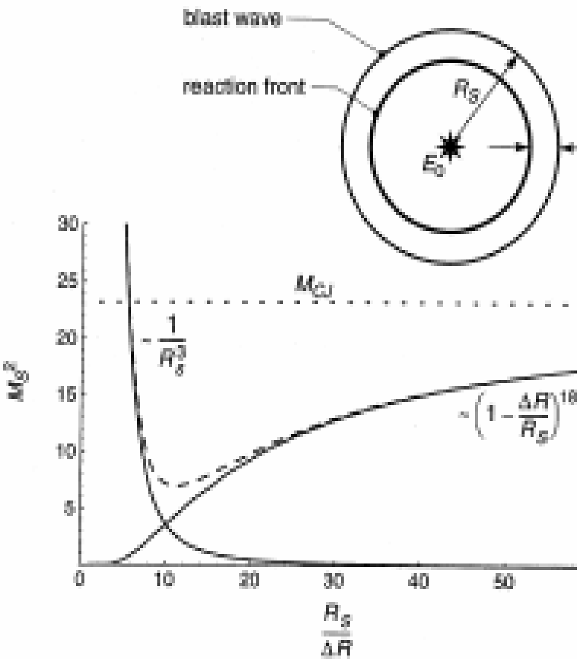


FIG. 4 PHYSICAL MODEL FOR DIRECT INITIATION OF A DETONATION

detonation kernel theory [7], which considers decay of the shocks formed by the energy release and the strengthening of the shocks by heat release from chemical reactions, predicts the requirements of directly initiating a detonation well. It considers for successful initiation of a detonation, the ignition energy to be such that a blast wave is formed whose strength when it has decayed down to the auto-ignition limits can generate sufficient heat release exceeding the applied ignition energy. A schematic of the decay and energisation by heating and chemical reactions which occur at a distance ΔR behind the shock is shown in Fig. 4. Ms in the figure refers to the shock Mach number and Eo is the energy released at the source. The initiation involves re-acceleration of the blast wave from Ms to the detonation Mach number by the dominating influence of the increasing chemical energy contained within it. Since a strong blast wave is formed by the ignition source, this mode of initiation is also spoken of as the blast initiation of detonations.

3.2 Deflagration to Detonation Transition

A flame formed at the ignition source accelerates as it moves away from the source. The acceleration results from the increased heating of the mixture and the enhanced turbulence from the pressure waves propagating ahead into the unburned gas from the propagating flame front. The flame front thus gets wrinkled. If a large eddy at the flame front is formed which engulfs a considerable amount of unburned gas and mixes it with the hot products of combustion, spontaneous energy release takes place leading to a detonation. The eddy is required to be sufficiently large so the shock can be amplified sufficiently.

A gradient in the rate of chemical reactions in the reacting eddy is postulated to give rise to a gradient in the induction times and lead to amplification of the shock [8]. If the amplification process is such that a shock of sufficient strength is created to bring about auto-ignition of the mixture, then a detonation wave can be formed. This mode of forming a detonation is spoken of as Shock Wave Amplification by Coherent Energy Release (SWACER)[8]. Merging of the shock waves ahead of the accelerating deflagration can also form a strong shock and lead to a detonation [9]. The availability of chain carriers is also important and this decides the rate of energy release [10].

Placing of obstacles in the path of the flame helps in accelerating the flame by improving gas mixing and forming new ignition centers from the heterogeneities. An initial bubble of detonation is formed from the ignition centers which sweep through the entire surface of the turbulent mixture. The formation of the bubble is similar to the formation of localized explosions and detonation in the quasi-steady shock reaction complex in the direct initiation of a detonation.

It is also shown by Lee [11] that a self sustained high speed combustion wave can propagate if sufficient turbulence is generated by the obstacles. In this case ignition is achieved by the rapid turbulent mixing between the reacting gases and the products rather than by shock heating. The nature of the propagation is like sonic combustion at speeds corresponding to Chapman Jouguet deflagration. A detonation bubble could get generated during this propagation and lead to the initiation of a detonation. The spacing between obstacles and the type of obstacle becomes important for the deflagration to detonation transition since the transverse waves are formed and intensified here. Resonance could to the intensification. The final stage of forming a detonation is similar to the acceleration observed in the direct initiation of detonations.

3.3 Hot Gas Jet and Diffraction of Detonations

Here a hot gas jet from the products of combustion at constant volume in a pre-chamber mixes with the unburned gas as it expands out rapidly. The starting vortex of the transient jet is conducive to provide good mixing with the unburned gas since the strong density gradients at the interface promote surface instability and mixing. If a sufficiently large kernel of rapid heat release is formed, a detonation can be formed by the SWACER mechanism given earlier.

In the diffraction of detonations, a detonation from a more sensitive mixture enters the less sensitive mixture which is to be detonated. The incident detonation partially quenches. If the energy release in the shock-reaction complex is adequate to mix and form a kernel of detonation as in the case of direct initiation and deflagration to detonation transition, a self sustaining detonation wave is formed.

3.4 Hypervelocity Projectile

When a projectile travels at high speeds in a reactive gas mixture, it does work on the gas at its surfaces due to the aerodynamic drag. The work done is equivalent to depositing energy in the gas in contact with the surface of the projectile [12]. If the projectile is moving at hypersonic speeds, the rapid rate of energy dissipation leads to formation of cylindrical blast wave. The formation of detonation with the hypervelocity projectile is therefore similar to the mechanism in direct initiation of detonation.

3.5 Deposition of Chain Carriers

The deposition of chain carriers will bring about chemical reactions and if the zone of deposition is significant a shock wave can be formed due to focusing of shock from the induction time gradients. Detonation can then be formed through the SWACER mechanism.

4. MODELLING THE INITIATION OF DETONATION: LARGE ENERGY REQUIREMENTS FOR HYDROCARBON-AIR MIXTURES

The detonation was seen to be a three dimensional interacting shock structure with decaying shock fronts and transverse shock waves. In spite of the strong three-dimensional transient shock wave structure, the overall front still propagates as a one dimensional wave at the Chapman Jouguet speed. The propagation velocity is very close to that predicted from a one dimensional model which considers chemical reactions to take place at the shock.

The classical model of a detonation is one of a shock sustained by chemical reactions that take place due to shock heating. The unsteady nature viz., a decaying shock front within a cell of detonation introduces an unsteady flow field behind the shock front. The curvature of the shock front within the detonation cell also leads to expansion of the gases behind the shock. These two effects modify the rate of heating of the gases behind the shock. It must be borne in mind that the shock wave, while passing through the reactive mixture, heats it up to a level wherein spontaneous chemical reactions take place and the energy released during the combustion process goes to maintain the strength of the shock. The coupled complex of the shock and chemical reactions constitutes the detonation. The zone of chemical reactions due to the modulated flow field causes the heat release to be substantially delayed.

The length scale of the chemical reactions in the unsteady flow field is termed as the hydrodynamic thickness of the detonation. This length scale has been linked to the characteristic size of the detonation cell or the critical diameter required for the propagation of a detonation [5]. Once this length scale is estimated, the requirements of initiating a detonation can be readily worked out from an equivalent one dimensional model by ensuring that the shock wave formed by the ignition source is such that spontaneous chemical reactions behind the shock will drive it subsequently.

The energy driving a shock comes from the energy released by the ignition source and the chemical reactions occurring behind the shock at a distance ΔR behind the shock. This was shown in Fig. 4. The shock Mach number is decided by the competing influence of the source energy and chemical reactions. Initially the shock is driven by the energy from the energy source in driving the shock. When the chemical reactions take over the driving process, a detonation is formed. The detonation kernel theory [7] makes use of the experimental observation that by the time the shock decays to a critical limit at which auto-ignition of the shocked gas is possible, the chemical energy release must be sufficient to overcome the decaying influence of the ignition energy and thus maintain its strength. The energy estimated for stoichiometric propane-air mixture works out to be several hundred KJ which is clearly not possible to be generated in the PDE.

The characterization of the ignition source energy in terms of a characteristic length R_0^* known as the explosion length shows it to be typically about 32 times the transverse wave spacing

($R_0^* = 32\ \lambda$) based on experiments in ethylene air mixtures at near-atmospheric pressures [3]. The explosion length $R_0^* = (E_0/P_0)^{1/3}$. If we take the mean value of the transverse wave spacing for a hydrocarbon-air mixture based on different investigations to be about 50 mm [12], the energy for initiating detonation works out to be about 400 kJ, which is near to the value of energy cited by Schauer et al.[1].

In the case deflagration to detonation transition, the delay in forming a detonation leads to smaller values of head-end pressure and hence lower value of impulse. Experimental studies of Schauer et al.[1] have demonstrated that detonation can be achieved by improvising vigorous mixing using Schelkin coils. This is not very desirable considering the losses and it would be of interest to study the possibilities of getting the detonation initiated through some other means near the head end of the detonation tube. For this purpose we look at the feasibility of using hot jets for initiating a detonation in hydrocarbon-air mixtures.

5. PREDICTIONS FOR INITIATION OF DETONATIONS USING HOT JETS FROM A PRE-CHAMBER

Propane-air mixture is considered to be representative of hydrocarbon-air mixtures based on the experimental observations of Schauer et al. [1]. A jet of hot gas from the constant volume combustion of stoichiometric propane-air in a pre-combustor is assumed to expand into the unburned propane-air mixture in a detonation tube through an orifice, entrain and mix with it. The head of the transient gas jet is in the shape of a mushroom and provides good mixing between the products of combustion of the expanding hot gas jet and the unburned gas mixture. Figure 5 shows the transient head of the starting vortex and the mixing process with the unburned gases. A kernel comprising of a mixture of the expanded jet and unburned gas is formed. The possibility of forming a detonation from this kernel is to be determined.

The pressure of constant volume combustion of hydrocarbon mixtures contained at atmospheric temperature and pressure is typically about 0.7 to 1 MPa. Choked flow conditions will therefore prevail in the orifice between the pre-combustor and the detonation tube. The expansion of the hot gases to the atmospheric pressure in the unburned mixture in the detonation tube is rapid and can be assumed to isentropic. The temperature of the expanded gases in the starting vortex can be approximated as:

$$T_{j,\,\exp} = T_{cv}[p_d/p_{cv}]^{(\gamma-1)/\gamma} \tag{1}$$

Here $T_{j,exp}$ denotes the temperature of the hot jet after the expansion, T_{cv} is the temperature of the constant volume combustion, p_{cv} is the constant volume combustion pressure driving the hot jet and p_d is the initial pressure of the hydrocarbon-air mixture in the detonation tube. The value of p_d is 1 atmosphere (~ 0.1 MPa). The chemical composition of the expanded jet can be assumed to be the same as the products of combustion in the constant volume combustion in the pre-chamber. This is reasonable considering the fast expansion process.

The temperature and composition of the kernel of gas formed by the mixing of the frozen expanded jet with the unburned gas (Fig. 5) can be estimated using the conservation of energy and the conservation of the chemical species. For this purpose a volume V_j of the expanded jet is assumed to mix with a volume V_u of the unburned gas. The energy conservation gives

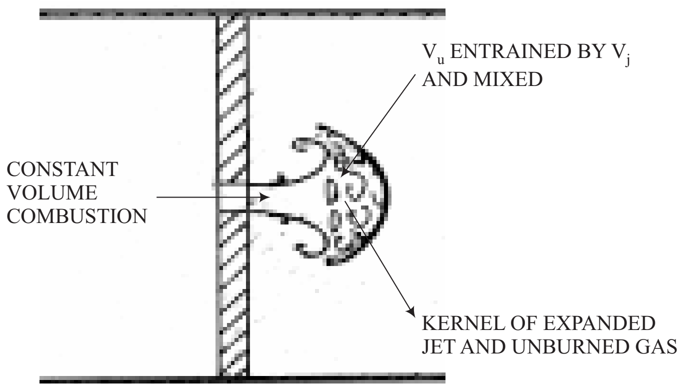


FIG. 5 PHYSICAL MODEL FOR INITIATION USING HOT JET

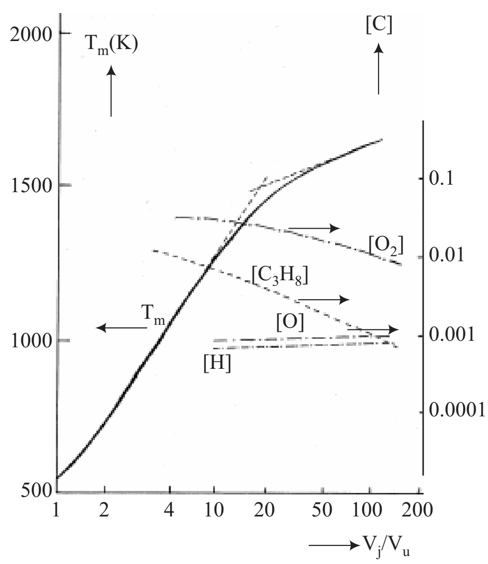


FIG. 6 PROPERTIES OF KERNEL FORMED BY MIXING UNBURNED GAS WITH THE HOT EXPANDING JET

$$\rho_j V_j Cp_j T_{j,\,\mathrm{exp}} + \rho_u V_u Cp_u T_u = (\rho_j V_j Cp_j + \rho_u V_u Cp_u) T_m \tag{2}$$

In the above equation, ρ_j is the density of the expanded jet and Cp_j is its specific heat. ρ_u, Cp_u and T_u denote the unburned gas density, specific heat and temperature. T_m is the temperature of the mixed gases in the kernel.

Equation (2) can be rewritten for the temperature T_m as:

$$T_m = \{1/[1 + (\rho_u Cp_u/\rho_j Cp_j)/R]\}\, T_{j,\,\mathrm{exp}} + \{1/[1 + ((\rho_u Cp_u/\rho_j Cp_j)/R)^{-1}]\}\, T_u \tag{3}$$

Here R denotes the ratio of the volumes of the expanded jet to the unburned gas (V_j/V_u).

When the value of R $\rightarrow$ 0, i.e., when V_j is small, $T_m = T_u$. For R $\rightarrow \infty$, i.e., V_j is large, $T_m = T_{j,exp}$. The temperature of the kernel has therefore the asymptote of the unburned gas temperature when the proportion of the hot gas jet in the kernel is small and the asymptote of the expanded gas jet temperature when the jet gases in the kernel is large. In between the two, the temperature T_m has an inflection point following a rapid increase. Figure 6 shows the variation of the kernel temperature for a propane air mixture for which ρ_u =1.205 Kg/m^3, Cp_u = 0.251 Kcal/Kg K, ρ_j =0.201 Kg/m^3, Cp_j = 0.357 Kcal/Kg K, $T_{j,exp}$ = 1683 K and T_u = 298 K. The inflection point is seen to occur for a volume ratio of about 20. The inflection point is of interest as it denotes the arrest of a sharp increase of temperature and this could well be the region wherein a sharp change in the thermally induced chemical reactions can occur.

The composition of gases in the kernel is determined from the frozen composition of the expanded jet and the volumes of the unburned gas and burnt gas with due weightage being given for their respective volumes. Thus if the mole fraction of a particular specie C is $[C_j]$ in the expanded jet and $[C_u]$ in the unburned gas, the mole fraction in the kernel $[C_k]$ is given by:

$$[C_k] = (V_j[C_j] + V_u[C_u]) / (V_j + V_u) \tag{4}$$

The variation of the concentration of the different species with change in the volume ratio is shown in Fig. 6. The concentration of the chain carriers OH, H and O are not seen to change significantly over a wide range of the volume ratios. The mole fraction of OH is much higher than of O and H. The proportion of the fuel propane and oxygen decreases with increase in the volume of the burnt gases.

The temperature of the kernel from Eq. (3) and the mole fraction from Eq. (4) are used to determine the induction time (τ) and the energy release (ΔE) of the kernel of mixed gases. The Chemkin code was used for the evolution of temperature due to chemical reactions with time for the initial conditions specified by T_m and $[C_k]$. The induction time is assumed to be 1/e of the time taken to reach the maximum temperature T_b (Fig. 7) since strong ignition associated with rapid energy release is important for the detonation. The temperature difference ($T_b - T_m$) is used to calculate the energy release ΔE.

The variation of the induction time and the temperature increase due to combustion is shown in Fig. 8. It is observed that the induction time rapidly falls from several hundred milliseconds to about 50 μs as V_j/Vu increase from 1 to 20. Beyond a volume ratio of 20, the induction time increases again. The increase is not as rapid as the fall at the lower volume ratios. Though the temperature of the gas mixture goes up, the lower concentration of the fuel and oxygen do not permit fast reactions.

The temperature increase ($T_b - T_m$), which is proportional to the energy release from the kernel, is shown by the shaded region in Fig. 8. It is seen to progressively decrease as V_j/Vu increases. A volume ratio of around 20 is seen to be essential to give the smallest induction time and a meaningful value of heat release.

The choice of the above volume ratio of the mixed gases in the kernel is most favorable for a detonation to occur. However, it requires a certain minimum length scale for shock waves to be formed and strengthened either through SWACER or otherwise. The length scale of the kernel should also be such that when the detonation exits from the kernel to the unburned gas, it does not decay. Two values of length scale therefore become relevant: one for formation of shock waves and

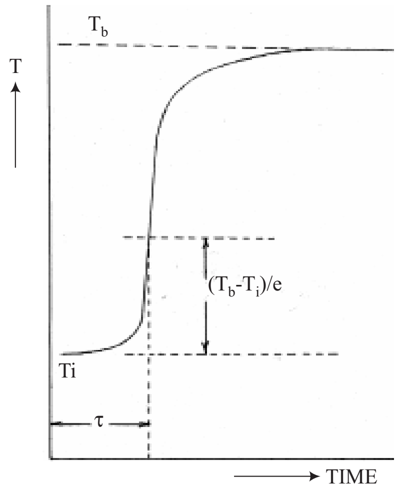


FIG. 7 INDUCTION TIME

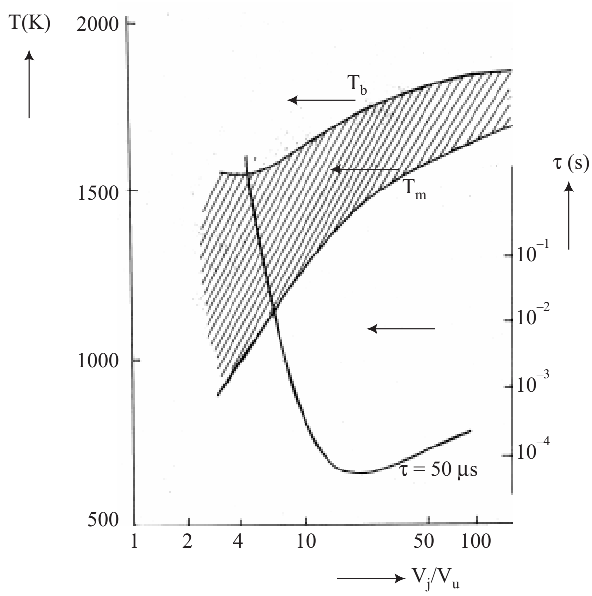


FIG. 8 INDUCTION TIME AND TEMPERATURE INCREASE IN THE KERNEL

its amplification to a detonation and the other to ensure that the kernel volume is sufficient to prevent the decay of the detonation subsequently. Further work is required to be done to model and quantitatively determine these length scales.

An assessment of the order of magnitude of the length scale viz., the size of the jet required for initiating a detonation in Propane-air mixture can be based on experiments done on jet initiation with more sensitive mixtures such as acetylene-oxygen. Experiments in a sub-atmospheric acetylene-oxygen mixture show that an orifice diameter of 73 mm was required for initiating a detonation [14]. The induction time of the kernel formed by this mixture, if calculated using Eq. (3) and (4) and the chemical kinetic data, is 2 μs at a volume ratio R of 5. The sound speed in the kernel of different hydrocarbon-air mixtures would be reasonably similar and the characteristic length should therefore scale as the induction time. With the minimum induction time in the propane-air mixture being 50 μs, the orifice diameter required for propane air would be about $50 \times 73/2 \sim 2$m. This is an unduly large diameter and suggests that the formation of detonation from a hot gas jet for a PDE would be difficult.

An estimate of the orifice size required for a detonation from a more sensitive mixture to form a detonation in the propane-air mixture can be guessed based on the number of transverse waves required so that the detonation does not decay. Studies on diffraction of detonations have shown that a minimum of about 13 cells of detonation are required to sustain a detonation [15]. The characteristic cell size in hydrocarbon-air mixture is typically between 46 and 75 mm [13]. This suggests that an orifice size of about 0.75 m would be required. This is again unusually large to be used for a PDE.

6. CONCLUDING REMARKS

The different modes on initiating a detonation in a hydrocarbon-air mixture are examined with specific reference to the multiple shock structure of a detonation. The direct initiation of detonation demands high ignition energies and is not suited for PDE. The possibility of using a pre-combustor in which constant volume combustion can take place and exhaust the hot gases into the unburned hydrocarbon-air mixture in a detonation tube is also found to be difficult considering the very large length scales involved. Introduction of chain carriers in the initiating kernel would reduce the induction times and hence the length scale. Operation at richer mixtures or preheating of the hydrocarbon or addition of hydrogen to the hydrocarbon would help.

The jet fuels contain alkanes and alkenes and the presence of the alkenes improves the detonability. The hydrocarbon fuel JP-10 which comprises of a single hydrocarbon molecule (exo tetra hydrodi cyclo pentadiene) has been shown to be more difficult to ignite and detonate than the other hydrocarbon fuels in the jet series [16] which have a combination of different hydrocarbon molecules. Suitable blending of fuels would ease the detonability and needs to be explored.

The formation of a flame at the ignition source and making it transit to a detonation by the acceleration of the flame by obstacles such as Schelkin spiral has been demonstrated[1]. A better understanding of the formation of transverse waves by the obstacles which are critical to the formation and sustenance of a detonation is required. This would help in suitably configuring the obstacles, creating transverse shocks and ensuring that lower velocity quasi-steady detonations are not formed.

The narrow range of limits of detonation of hydrocarbon-air mixtures [17] additionally require that the hydrocarbon vapor and air are well pre-mixed at near-stoichiometric proportion to ensure detonability. The possibilities of obtaining intensive mixing in the transient flow field of a jet such as at the head of a mushroom-shaped vortex needs to be studied. Sufficient intense fine scale structure in the overall large scale structure of the vortex head would be required to ensure good mixing.

ACKNOWLEDGEMENTS

Part of the work on jet initiation of detonation reported here was done during 2002 when I was a Visiting Professor at McGill University, Montreal. I wish to record my thanks to my mentor Professor John H. Lee who introduced me to this fascinating subject on detonations first as a graduate student and then through a series of interactions.

REFERENCES

[1] F.R. Schauer, C.L. Miser, K.C. Tucker, R.P. Bradley and J.L. Hoke, 2005, "Detonation Initiation of Hydrocarbon-Air Mixtures in a Pulse Detonation Engine", AIAA 2005-1343, *43rd AIAA Aerospace Sciences Meeting,* Reno, 2005.

[2] V.P. Zhukov and A. Starikovskii, 2006, "Effect of Nanosecond Gas Discharge on Deflagration to Detonation Transition", *Combustion, Explosion and Shock Waves*, vol 42, 2006, pp 195-204.

[3] M.I. Radulescu, A.J. Higgens, J.H. Lee and S.B. Murray, "On the Explosion Length Invariance in Direct Initiation of Detonation", *Proc. Comb. Inst.*, vol 28, 2000, pp 637-644.

[4] I.B. Zeldovich, S.M. Kogarko, and N.N. Siminov, 1956, "An Experimental Investigation of Direct Initiation of Spherical Detonation of Gases", *Soviet Physics Technical Physics*, vol 1, 1956, pp 1689-1698.

[5] J.H.S. Lee and M.I. Radulescu, "On the Hydrodynamic Thickness of Cellular Detonations", *Combustion, Explosion and Shock waves*, vol 41, 2005, pp 745-765.

[6] G. Dupre, O. Peraldi, J.H.S. Lee, and R. Knystautas, "Propagation of Detonation Waves in an Acoustic Absorbing Walled Tube", *Progress in Astronautics and Aeronautics*, vol 114, 1998, pp 248-263.

[7] J.H. Lee, and K. Ramamurthi, "On the Critical size of a Detonation Kernel", *Combustion and Flame*, **27**, 1976, pp 331-340.

[8] R. Knystautas, J.H. Lee, I. Moen and H.G. Wagner, "Direct Initiation of a Spherical Detonation by a Hot Turbulent Gas Jet", *17th Sym.(Int.) Comb.,* 1979, pp 1235-1245.

[9] P.A. Urtiev and A.K. Oppenheim, "Transverse Flame Shock Interactions in an Explosive Gas", *Proc. Roy. Soc.*, vol A, 304, 1968, pp 379-392.

[10] L. He, "A scaling Analysis of Critical Conditions for Deflagration to Detonation Transition", *Combustion Theory and Modeling*, vol 4, 2000, pp 107-122.

[11] J.H. Lee, "Initiation and Propagation Mechanisms of Detonation Waves", *Paper at 3rd Int. High Energy Materials Conference, Trivandrum,* 2000.

[12] J.H. Lee, "Initiation of Detonation by a Hypervelocity Projectile", *Progress Astro. Aero.*, AIAA, vol 173, 1997, pp 293-310.

[13] Detonation Data Base, www.galcit.caltech.edu/detn_db/html/db_120.html/.

[14] M. Inada, J.H. Lee and R. Knystautas, "Photographic Study of the Direct Initiation of Detonation by a Turbulent Jet", *13th ICDERS, Nagoya*, 1991.

[15] R. Knystautas, J.H. Lee, and C.M. Guirao, "Critical tube Diameter for Detonation Failure in Hydrocarbon-Air Mixtures", *Comb. Flame*, vol 48, 1982, pp. 63-83.

[16] S.R. Tieszens, D.W. Stamps, C.K. Westbrook and W.J. Pitz, "Gaseous Hydrocarbon-Air Detonations", *Combustion Flame*, vol 84, 1991, pp 376-390.

[17] A.A. Borisov and S.A. Loban, "Detonation Limits of Hydrocarbon-Air Mixtures in Tubes", *Comb. Explosions and Shock waves*, vol 13, 1978, pp 618-621.

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

TURBOCHARGING OF AN AUTOMOTIVE DIESEL ENGINE UNDER RETROFIT CONDITIONS – A CASE FOR EVALUATION AND ANALYSIS

DR. C.L. DHAMEJANI*

In the past, diesel engines were not accepted as automotive power plants due to their space and onboard load constraints. Higher specific weight and space requirements have always been cited as major shortcomings of diesel engines with reference to mobile applications. However, in the current context, where diesel conservation is of prime importance, the aforesaid shortcomings are far superseded by the single advantage of fuel economy. Adoption of diesel engines as a popular power plant for automotive applications has resulted in immediate need of intense research efforts in the direction of making these machines more compact and light.

Turbocharging is considered as one of the most effective approaches towards reduction of specific weight and size of engines. Engine breathing capacity is enhanced by pre-compressing the air charge using energy of exhaust. The power output of an engine being a function of its breathing capacity, turbocharging implies increased output from the same size of engine. The associated welcome features of turbocharging and for that matter any other kind of supercharging, include improved efficiency and diminished fuel sensitivity of the engine.

In a vast country like ours, the defence vehicles have to operate under varying ambient conditions, from those existing at sea level to the ones which refer to the altitudes of the order of 5000 m. High altitude operation of the vehicles where both ambient pressure and temperature are considerably low has been a serious problem faced by Defence. Power capacity of an engine under these conditions reduces to just about 60 percent of the sea level power. To be able to cope up with these conditions, defence vehicles are rated at a much lower payload capacity.

Supercharging or turbocharging is a promising method through which the engine power could be substantially boosted. Turbocharging holds equal promise for recoupement of loss of power under high altitude operation of the vehicles.

One of the important defence vehicles Shaktiman was powered by an engine rated at 71 kW at 2500 RPM and torque rating of 32.4 mkg at 1500 RPM under naturally aspirated sea level

*Director, Vehicles Research & Dev. Estt, Ahmednagar Defence Res. & Dev. Org., Min. of Defence New Delhi

conditions. It is rated at 3 tonne payload. For obvious reasons interest in uprating the payload capacity of the Shaktiman vehicle existed. The interest was particularly intensified from the point of view of its operation at high altitudes. Increasing power capacity of Shaktiman engine by 50 percent therefore was considered as an important task for past two decades.

Boosting power capacity of Shaktiman engine by turbocharging was essentially a retrofit situation. Retrofitting a turbocharger to an existing engine was a multifacet and case specific problem having several prerequisites. Adequacy of mechanical strength of the machine under possible increased mechanical and thermal loading has to be pre-established in such an attempt.

The gas exchange system has to suit the requirements of turbocharger characteristics. This implies need for basic redesign of the cooling system as well as the exhaust and the suction systems. The fuel injection system has also to be redesigned for increased fuelling requirements. Experimental verification of the redesigned systems for the results aimed forms an important constituent of the retrofitting task.

This work encompasses all the aforesaid requirements of retrofitting. The theoretical content of the work consists of performance prediction using an existent mathematical model, which has been improvised to suit the particular requirements of this work. Experimental investigations have been carried out to prove the redesigned systems as well as to provide validation of the improvised mathematical model.

The work was planned and carried out in five distinct phases, the logistic of which was worked out on the basis of then existing published information. Tackling the problem in a phased manner as described in succeeding paragraphs was particularly chosen to be able to analyse affect of different parameters which is important from the point of view of utilizing the information so generated. As can be seen from the following description, turbocharging of Shaktiman engine consisted of a number of changes in different systems for ensuring brought about faced a constraint due to various limitations and were effected in a phased manner.

The work carried out and the results achieved are discussed in the following paragraphs sequentially.

Engine performance was experimentally determined under naturally aspirated as well as turbocharged conditions for baseline data generation and examination of the firing sequence of engine from point of view of its suitability for turbocharging.

Engine was then modified and the firing sequence was changed. This implied redesigning of the engine camshaft and fuel injection pump camshaft. This phase yielded really encouraging results. The engine power as well as the efficiency improved. Considerable improvement in torsional vibration amplitudes was achieved. (This engine had a problem of premature crankshaft failures, which was solved using redesigned camshafts for changed firing sequence). The firing sequence was changed from 1-2-4-6-5-3 to 1-5-3-6-2-4.

Redesign of exhaust manifold to suit the turbine requirements constituted the next phase of work. Redesigned exhaust manifold worked out to be of a very simple layout. Higher degree of supercharging was achieved with the changed firing sequence and redesigned exhaust manifold. The maximum power output of the engine increased from 92.5 kW to 101.5kW without any change in fuelling which implies a considerable reduction in specific fuel consumption (from 263 to 231g/kWh).

The redesigned exhaust manifold being a much compact unit, it provided adequate space in the engine compartment of incorporation of a charge cooler which was designed and incorporated in the next phase of work. Reduction in exhaust temperature from 914 k to 825 k at the maximum power point i.e. with 40 percent boost of power was achieved with incorporation of the charge cooler. Yet another achievement of the charge cooler incorporation consisted of the increase in the maximum power capacity from 40 percent boost to 50 percent boost, which proved adequacy of the attempted retrofit.

Calculations were made to determine the main fuel injection system parameters referring to 50 percent power boost and changes were accordingly incorporated. Experimental investigations with new fuel injection equipment showed decreased exhaust temperatures (from 829 k to 816 k) and high brake thermal efficiency (from 37.9 to 38.7 percent) at peak power point as anticipated.

An important part of this project consists of improvisation of an existant mathematical model so that the effect of change in the firing sequence on engine performance in naturally aspirated condition could be taken care of. This model predicts the engine performance under turbocharged conditions using certain basic experimental data as input. The parameters computed by this model include power and specific fuel consumption. The model has also been tested for prediction of engine performance under extrapolated high altitude conditions. In this model, balance of turbine and compressor power has been used for computing the compressor pressure ratio. The predictions using this model suggest that the Shaktiman engine under turbocharged conditions developed as much as 95 percent of the sea level power even at 5000 m altitude. It will therefore not be necessary to derate the engine i.e. reduce the fuelling for high altitude operation.

A prototype of retrofitted turbocharged Shaktiman engine was made and its performance was evaluated as the final phase of this work. The prototype yielded 50 percent power increase as aimed. The mechanical and thermal loading of the engine were verified to be within safe range. After exhaustive proving tests which were carried out it was adopted by the Defence.

REFERENCE

1. Dhamejani C.L, 'Turbocharging of an Automotive Diesel Engine Under Retrofit Conditions – A case for Evaluation and Analysis' submitted in partial fulfillment of the requirement of the Ph. D. at I.I.T. Bombay : 1987. Under the guidance of Prof (Mrs.) P.P. Parikh and Prof. S.P. Sharma, Maj. Gen. P.M. Menon (co-guide)

CHEMICAL ENGINEERING AND BIOTECHNOLOGY

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

TECHNIQUES IN SEPARATION: ULTRAFILTRATION, NANOFILTRATION AND REVERSE OSMOSIS

S. MAYADEVI*

1. INTRODUCTION

Separation processes span a diverse range of applications in chemical process industries and contribute to 30 – 70% of the capital cost [1]. Some of the industries that spend a large percentage of their total cost on separation include commodity and specialty chemicals, biotechnology/ pharmaceuticals, microelectronics, food processing, plastics, metals and fertilizer. The emerging trend has been to develop cheaper, faster and efficient separation processes to meet the intense global competition, stringent environmental regulations and quality control. Membrane based separation processes have been either increasingly replacing or used in combination with the traditional separation processes due to their low energy requirement and increased efficiency.

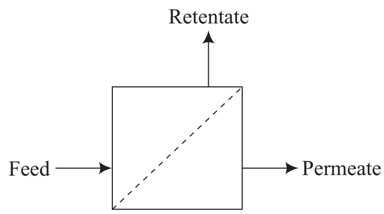


FIG. 1 A SCHEMATIC DIAGRAM OF A MEMBRANE SEPARATOR

A membrane is a selective barrier between two phases. The two phases can be both liquid, both gas or liquid and gas. The membrane selectively permeates one or more components present in the feed. The components passing through the membrane are called permeates and those retained by it are called the retentates. The driving force for separation can be pressure difference (ΔP), potential difference (ΔE), concentration difference (ΔC) or the ratio of temperature to pressure difference ($\Delta T/\Delta P$) across the membrane. Some common applications of membrane based separation are microfiltration (MF), ultrafiltration (UF), nanofiltration (NF), reverse osmosis (RO), dialysis, gas

*National Chemical Laboratory, Pune 411 008

separation (GS), vapor permeation (VP), pervaporation (PV) etc. The driving force for separation for all the above applications except dialysis is the pressure difference across the membrane. In dialysis, the separation is driven by the concentration difference across the membrane. MF, UF, NF, RO and dialysis are liquid-phase operations, GS and VP are gas-phase operations whereas PV has the feed and retentate in the liquid-phase and permeate in the gas/vapor phase.

2. MICROFILTRATION

Microfiltration is generally used for the separation of particles by sieving, based on the particles size. Porous polymer/ceramic membranes of pore size between 10 – 0.05 microns are used for the separation. These membranes retain particles of size between 0.1 and 10 microns and higher. They operate at relatively low pressures (0.1 – 2 bar) and have a high flux (> 50 $l/m^{-2}.h^{-1}.bar^{-1}$).

A variety of membrane materials are used for the preparation of MF membranes. These include polymers, ceramics, metals and glasses. Polymer membranes can be hydrophilic or hydrophobic. Some of the hydrophobic membranes are made of polytetrafluoroethylene (PTFE, Teflon), poly(vinylidene-fluoride) (PVDF), polypropylene (PP) and polyethylene (PE). Examples of hydrophilic membrane materials are: cellulose esters, polysulfone/polyethersulfone (PSf/PES), aliphatic polyamide (PA) and polyetheretherketone (PEEK). Ceramic MF membranes are available in alumina, zirconia, titania and silicon carbide. They are prepared by sintering, stretching, track-etching or phase inversion techniques. Sintering can produce low/medium porosity membranes with narrow/wide pore size distribution. Medium to high porosity membranes with narrow/wide pore size distribution are prepared by stretching. Membranes with low porosity and narrow pore distribution are prepared by track-etching method. Phase inversion method produces highly porous membranes with narrow/wide pore size distribution. These can by symmetric or asymmetric. The thickness of the separating layer in symmetric membranes range from 10 – 150 microns whereas it can be as low as 1 micron in asymmetric membranes.

Normally MF is used for the separation of particulate materials from suspensions and emulsions. An example is the separation of an aqueous solution of salt, sugar, proteins and fats. MF membrane retains the fats and permeates all the other components. Some commercial applications of MF are cold sterilization in food and pharmaceutical industry, cell harvesting, clarification of fruit juice, wine and beer, production of ultra-pure water in semi-conductor industry, metal recovery, separation of oil-water emulsions in waste water treatment and in continuous fermentation process.

3. ULTRAFILTRATION

Ultrafiltration is generally used for the separation of macromolecules (e.g. bacteria, yeast) by sieving, based on the particles size. It can also be used for the separation of high molecular weight compounds from low molecular weight compounds. Porous polymer/ceramic membranes of pore size between 0.01 – 0.1 micron are used. These membranes retain particles of size between 0.01 and 0.1 micron and higher. They operate at 1 – 5 bar pressures and have flux between 10 and 50 $l/m^{-2}.h^{-1}.bar^{-1}$.

Asymmetric polymer based membranes and ceramic membranes are used in UF applications. Polymer membranes are made of PSf/PES/sulfonated PSf, PVDF, polyacrylonitrile (PAN), cellulosics (e.g. cellulose acetate), polyimide/polyetherimide (PI/PEI), aliphatic polyamides and polyether ketone. Alumina and zirconia based ceramic UF membranes are also used in several applications. Asymmetric polymer UF membranes are mostly produced by phase inversion. Different ceramic membrane preparation methods are used in the production of ceramic UF membranes.

UF membrane permeates sugars and salts from a solution containing proteins, sugars and salts and retains the proteins. It is used for the separation of whey from milk and in cheese making (diary industry), for the separation of potato starch and proteins in food industry, for the separation of oil-water emulsions and electro paint recovery (metallurgy and automotive industry), for the separation of enzymes and antibiotics in pharmaceutical industry, and in textile industry for water treatment.

4. NANOFILTRATION

Nanofiltration is used for the separation of low molecular weight solutes from solutions. The separation mechanism is based on solution-diffusion. These membranes have pore size less than 0.002 micron and retain molecules of size 0.002 – 0.0005 microns. They operate at 5 – 25 bar pressure and have flux between 1.4 and 12 $l/m^{-2}.h^{-1}.bar^{-1}$.

Composite membranes are used for NF. These typically have a sub-layer of about 150 microns and top layer of one micron. Polyamide membranes used for NF are prepared by interfacial polymerization. By NF it is possible to separate sugars from a solution containing sugar and salt. Salt solution permeates while the sugar is retained by the membrane. NF can be used for the separation of mono-, di- and multi-valent ions from solution, desalination of brackish water, removal of pollutants etc. Typical applications include desalination of food, diary and beverage products, color reduction of food products, concentration of food, beverage and diary products, concentration of fermentation products and desalination of dyes and optical brighteners.

5. REVERSE OSMOSIS

Reverse osmosis (RO) is used for the separation of very small solute molecules from solution. The separation occurs by solution-diffusion mechanism. These membranes are semi-permeable and have pore size less than 0.002 micron. When a solution and solvent are separated by a semi-permeable membrane, the solvent flows to the solution side till the chemical potential on both sides of the membrane are same. This phenomenon is called osmosis and leads to the development of a pressure difference (osmotic pressure) between the solution and solvent sides of the membrane. The application of pressure higher than the osmotic pressure on the solution side leads to the reverse flow of solvent from the solution side to the solvent side. This phenomenon is called Reverse Osmosis.

Asymmetric or composite membranes (sub layer: 150 microns, top layer: 1 micron) with pore size << 0.002 microns are used in RO. This is also a pressure driven process. The applied pressures are very high – 15 to 25 bar for brackish water and 40 – 100 bar for sea water. The membrane flux of RO membranes are very small (0.05 – 1.4 $l/m^{-2}.h^{-1}.bar^{-1}$). Typical RO membrane materials are cellulose triacetate, poly(ether urea) and aromatic polyamide.

RO is extensively used for the removal of salts from sea water and brackish water (desalination). Other applications include concentration of fermentation products, concentration of food, beverage and diary products, color reduction of food products, production of ultrapure water in electronic industry, solute concentration and solvent purification/recovery.

5.1 Pressure Retarded Osmosis

Pressure retarded osmosis is a process derived from Reverse Osmosis. It enables the generation of energy from a concentration difference. In osmosis, water flows through a semi-permeable to the concentrated solution side till the osmotic pressure is attained. When the pressure applied is less than the osmotic pressure, water keeps flowing from the solution side. This water can be made to flow through a turbine to generate electricity. The main application of pressure retarded osmosis is the generation of electricity.

6. SOME DEFINITIONS

6.1 Membrane Performance or Efficiency

The performance/efficiency of a membrane is determined by two parameters: (1) Membrane flow/ membrane flux/permeation rate and (2) Membrane selectivity.

6.2 Membrane Flux

Membrane flux is the volume flowing through the membrane per unit area per unit time. The units are, $l.m^{-2}.h^{-1}$; $l.m^{-2}.day^{-1}$; $gal.ft^{-2}.day^{-1}$. This is the volume flux. It can also be expressed as mass flux (volume flux × density) or mole flux (volume flux × density/Molecular weight).

6.3 Membrane Selectivity

Membrane selectivity is generally expressed by two parameters – the retention (R) or the separation factor (a).

6.3.1 Retention (R)

For dilute aqueous solutions, the selectivity is normally expressed in terms of the retention of the solute; the solute is either partially or completely retained and the solvent passes through the membrane. The retention of a membrane is defined as

$$R = (c_f - c_p)/c_f = 1 - (c_p/c_f) \qquad (1)$$

Here c_f is the solute concentration in the feed and c_p is the solute concentration in permeate. R varies from 100% to 0%. When R is 100%, there is complete retention of the solute and the membrane is semi-permeable. When R is 0% both the solute and solvent passes freely through the membrane, i.e. there is no separation.

6.3.2 Separation factor (a)

Membrane selectivity towards gas mixtures and mixture of organic liquids is expressed in terms of the separation factor a, defined as

$$\alpha_{A/B} = (y_A/y_B) / (x_A/x_B) \tag{2}$$

A and B are the components of the mixture, y_A and y_B are the concentration of A and B in the mixture, and x_A and x_B are the concentration of A and B in the feed. The selectivity is chosen in such a way that its value is always greater than unity. If the permeability of B is higher than that of A, the selectivity is expressed as $\alpha_{B/A}$. When $\alpha_{A/B} = \alpha_{B/A} = 1$, there is no separation.

7. TRANSPORT THROUGH POROUS MEMBRANES

7.1 Hydraulic Permeability

The volume flux through the membrane is given by the Hagen-Poiseuille equation,

$$J_v = L_p \, \Delta P/\Delta x \tag{3}$$

Here J_v is the liquid flux, ΔP is the pressure difference across the membrane, and Δx is the membrane thickness. The proportionality factor L_p is called the hydraulic permeability and is related to the pore characteristics (pore radius r, porosity ε, tortuosity τ) and liquid viscosity η by the expression:

$$L_p = \varepsilon \, r^2/(8 \, \eta\tau) \tag{4}$$

L_p can be experimentally determined from the slop of the linear plot of J_v versus $\Delta P/\Delta x$. The Hagen-Poiseuille equation assumes that the pores are straight and cylindrical, and is applicable to MF and UF membranes.

7.2 Solute Permeability

Solute permeability w is related to the concentration difference across the membrane Δc, and solvent flux J_v by the equation:

$$(J_s / \Delta c) = \omega + (1 - \sigma) \, J_v \, (\hat{c}/\Delta c) \tag{5}$$

where

$$\hat{c} = (c_f - c_p)/\ln(c_f / c_p) \tag{6}$$

Here c_f and c_p are the solute concentration in the feed and permeate respectively, J_s is the solute flux and σ is the reflection coefficient. A plot of $(J_s/\Delta c)$ versus J_v $(\hat{c}/\Delta c)$ will be linear, and the intercept will give the value of ω**.** From the slope the reflection coefficient σ can be obtained.

The characteristic behavior of the membrane to the solute is reflected in the value of σ. σ equal to unity indicates that there is no solute transport, which means that the membrane is completely semi-permeable. $\sigma < 1$ means that there is some solute transport, the membrane is not completely semi-permeable. $\sigma = 0$ means that the membrane passes both the solute and the solvent with equal ease; there is no selectivity.

8. TRANSPORT THROUGH NON-POROUS MEMBRANES

When the sizes of the molecules to be separated are of the same order of magnitude, (e.g. O_2/N_2, hexane/heptane), porous membranes cannot effect separation. In such cases non-porous membranes are used. In this the pores are present at the molecular level. Non-porous membranes are used for both liquid and gas transport. Transport through non-porous membranes is described by a solution-diffusion mechanism. The permeability P is related to the solubility S and diffusivity D by the expression:

$$P = S \times D \tag{7}$$

Solubility S is a thermodynamic parameter and gives a measure of the amount of penetrant sorbed by the membrane under equilibrium conditions. Solubility of gases in polymer membranes is often described by the Henry's law.

9. SUMMARY

An introduction to membrane based separation methods, MF, UF, NF and RO is presented in this paper. It gives an overview of each separation method, the type of membranes used and typical applications. This is followed by the definition of some of the terms commonly used and an introduction to transport through membranes. (This lecture was delivered to the undergraduate students in the Chemical Engineering Department of Government Engineering College, Trichur, Kerala under the INAE distinguished visiting Professor program).

REFERENCES

[1] M. Mulder, "Basic Principles of Membrane Technology" *Kluwer Academic Press*. Dordrecht, 1998.

[2] R. Agrawal, "Separations: Perspective of a Process Developer / Designer" *AIChE Journal*, Vol. 47, No. 5, May 01, pp 967-971.

[3] H. Strathmann, "Membrane Separation Processes: Current Relevance and Future Opportunities" *AIChE Journal*, Vol. 47, No. 5, May 01, pp 1077-1087.

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

THERMAL ANALYSIS FOR CHARACTERISATION OF POLYMERS

K.N. NINAN*

1. INTRODUCTION

Polymers have played a vital role in the ongoing revolution in all fields of technology, whether it comes domestic (house hold) industry, microelectronics, telecommunications or aerospace. A thorough understanding of structure-property relationships and mastering of the technology for molecular design and industrial production of polymeric materials have helped in the technology advancements in the above mentioned areas. An important area where polymers have made enormous strides is that of aerospace, where advanced polymer materials have become inevitable in high performance solid propellants, composite structures, thermal protection system, coatings, structural adhesive, etc. Polymers are indispensable constituents of many components in satellite and space stations. In composites, polymers constitute either the continuous phase (matrix resin) or the reinforcement or both.

The properties of products made from polymers are greatly dependent on the physico-chemical and thermo-mechanical profile of the latter, which in turn, are determined by their molecular features. Therefore for every application, thorough characterization becomes imperative for the development of the material and for a reliable prediction of its properties. The exponential growth of polymers/composites has led to the development of several techniques for their characterization and no single technique has proved more useful than thermal analysis. This method drives information often unobtainable by other means and has been used in many areas of basic and applied research, production and quality control.

The term thermal analysis refers to a group of techniques in which some physical or chemical property of a system is measured as a function of temperature. All materials, as they experience changes in temperature, undergo changes in their physical and/or chemical properties. These changes can be detected by suitable that are collected and analyzed to give 'thermograms' showing the property change as a function of temperature. The applications in thermal analysis have grown tremendously since the latter half of the 20^{th} century mainly due to the revolution in

*Outstanding Scientist & Deputy Director, Propellants, Chemicals and Materials Entity, Vikram Sarabhai Space Centre, Thiruvanathpuram-695 022, E-mail: kn_ ninan@vssc.gov.in

instrumentation. A host of thermoanalytical techniques have been used in characterization of polymers and composites and the most widely used among them are the following.

1.1 Thermogravimetric Analysis (TG)

TG measures weight changes associated with thermal events. Commonly used for compositional analysis, for determining thermal stability and for the evolution of thermal decomposition kinetics in order to predict long-term as well as short-term thermal stability. Usual working temperature range is from ambient to 1500°C, although equipment with temperatures as high as 2400°C are available for characterizing materials for high temperature applications.

1.2 Differential Scanning Calorimetry (DSC) /Differential Thermal Analysis (DTA)

These techniques measure the temperature and heat flows associated with transitions in materials. Such measurements provide qualitative and quantitative information about physical and chemical that involves endothermic (heat absorbed) or exothermic (heat evolved) processes, or changes in heat capacity. In addition to the determination of heats and temperatures of physical and chemical transitions, DSC has specific use in finding out calorimetric purity and second order transitions (e.g. T_g). In DTA the service temperature ranges from RT to 1500°C (high temperature option up to 2400°C) whereas in DSC the range is generally from-180 to 750°C (high temperature option up to 1600°C).

1.3 Thermomechanical Analysis (TMA)

TMA measures the linear or volumetric changes in the dimensions of materials, provides valuable information on coefficient of thermal expansion, viscosity, gel time, glass transition temperature, modulus and creep/stress relaxation. TMA is normally used in the temperature range-180 to 1000°C (high temperature option up to 2400°C).

Another example is in the evaluation of competitive products like fluoro polymer based O-rings where the TGA decomposition profile furnishes direct information about the quality. Fig. 2 shows two TGA curves for O-rings – one for the O-ring supplied by a new source and the other for the original one. The relatively large volume of residue from O-rings supplied from the new source shows that it contains the larger amount of filler material which is detrimental to the endues properties of the O-ring.

1.3.1 Rapid Determination of Carbon Black Elastomers

The addition of carbon black to elastomers can have a positive effect upon its mechanical, electrical, and physical properties. By heating elastomers in the TGA under nitrogen and then switching to oxygen at 800°C, it is possible to determine the amount of oils, polymers, carbon black, and inorganic filler present. In the example of elastomers chosen here, fillers were not found to be present in the initial TGA scan, and hence the TG procedure for determining carbon black as shown in Fig. 3 was further simplified to heating the elastomers at 100°C/min. to 600°C in nitrogen and holding isothermal for five minutes. The residue was a direct indication of the carbon black content.

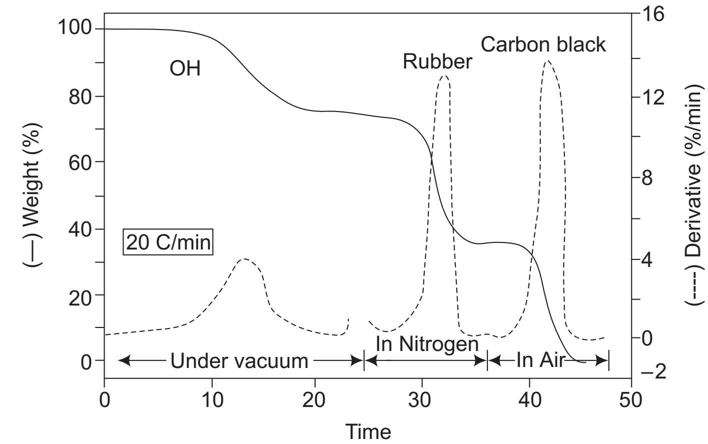


FIG. 1 TG-DTG CURVES OF A RUBBER FORMULATION

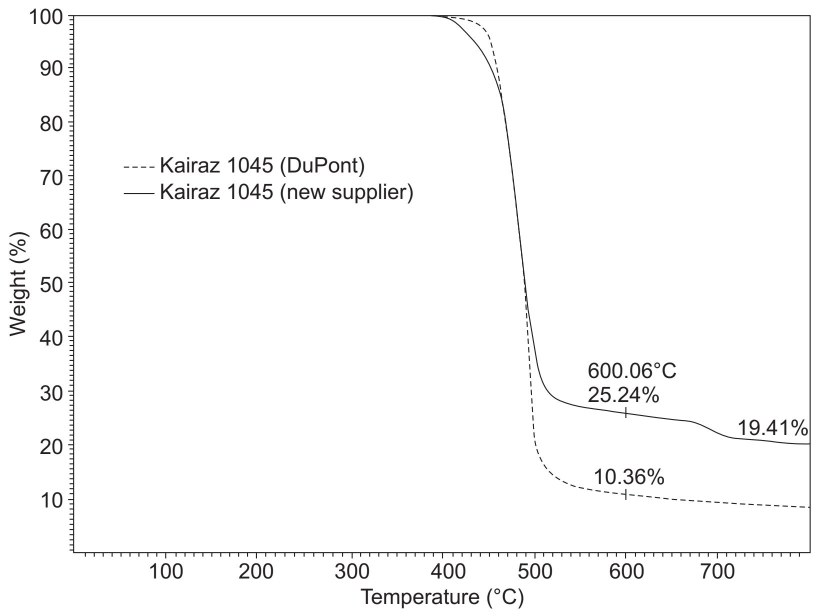


FIG. 2 TG CURVES OF O-RINGS

1.3.2 Compositional Analysis of Composites

TG provides a rapid means of the composition analysis of composite materials used as light weight structures in rockets and spacecrafts. eg., silica phenolic used as nozzle throat inserts. The analysis provides a rapid means of pre-preg composition analysis and hence can be used as a quality control parameter (Fig. 4).

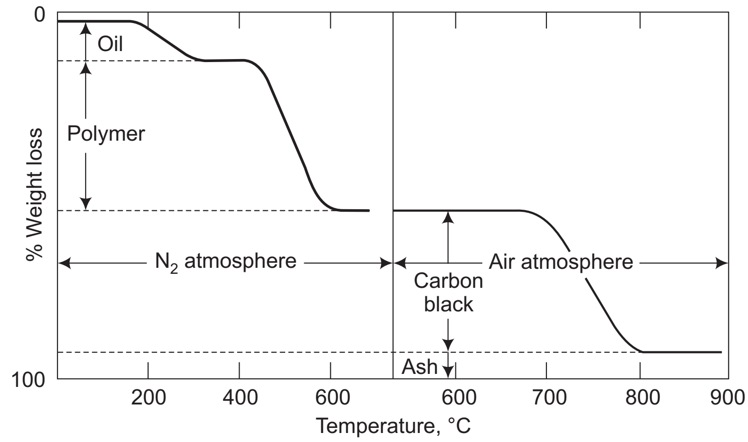


FIG. 3 ESTIMATION OF ELASTOMER COMPOSITION

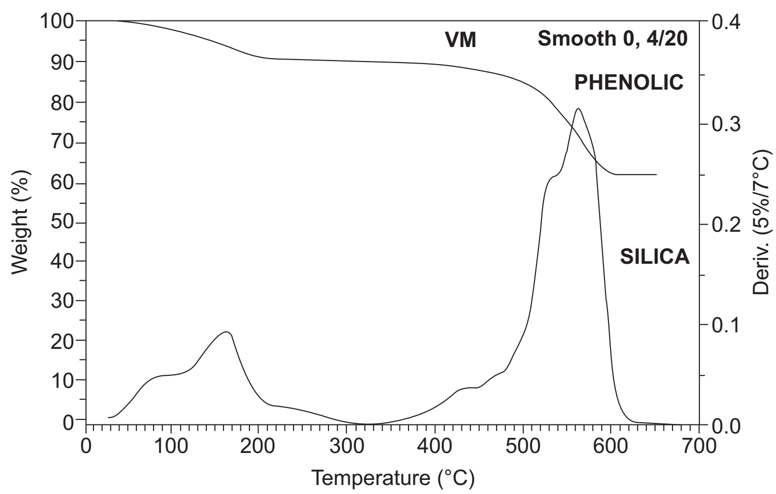


FIG. 4 TG-DTG CURVES OF SILICA-PHENOLIC PREPREG IN NITROGEN

Carbon-phenolics are used as rocket nozzle inserts at regions where the temperature experienced is above 1600°C. The above method is not applicable in this case because carbon fibres also will get oxidized in air atmosphere. On the other hand, a char residue is formed from phenolic resin when the experienced is done in nitrogen atmosphere. A specific method was therefore developed for the compositional analysis of carbon-phenolics by incorporating a correction for the char formed from the neat phenolic resin under identical conditions. The volatile matter (VM) content, resin content, and reinforcement content of bismaleimide based silica-polymide composites have also been evaluated from their TG data. From the TG curve, the density variation of the composite has been evaluated as a function of temperature during the polymer pyrolysis regime.

1.4 Thermal Stability of Polymers

Polymers combining unsurpassed properties of strength, flexibility, low thermal conductivity, easy processability, and viability for tailoring to meet any desired end use are correctly termed as the "wonder materials" of the present century. However, being essentially organic in nature, they suffer from the drawback of poor thermal stability in comparison to metallic materials and ceramics, especially in oxidizing environments. The low thermal diffusivity (arising from low thermal conductivity and high specific heat) makes polymers the candidate materials as thermal protection systems for protecting costly and sensitive components from high heat fluxes. Thus polymers have as important role in aerospace applications-for example a rocket has to be protected from high heat fluxes arising from hot combustion gases emanating from propellant combustion or from aerodynamic friction while ascending through the atmosphere. Therefore, researches world over are on constant endeavour to enhance the thermal stability of polymers for such applications. Every degree Celsius increase in thermal stability is a great boon to polymer scientists.

1.4.1 Relative Thermal Stability

Even though the degradation in mechanical properties precedes mass loss in many cases, the relative stability of different polymer systems can be assessed with fairly good accuracy from the temperature of onset and the nature of their mass-loss curves. This is possible from their TG curves, recorded under identical conditions. Fig. 5 shows a common example.

A number of speciality polymers have been specifically synthesized to meet the high demand of aerospace applications and an example is given in Fig. 6.

1.4.2 Oxidative Stability of Polymers

Certain polymers systems such as unsaturated polymers are prone to undergo rapid degradation in oxygen atmosphere. One widely used technique for predicting the resistance to oxidation involves heating them in an oven and periodically testing for weight gain due to oxygen uptake. TGA allows the same measurements to be made continuously, as shown in Fig. 7. The sample is heated to the temperature of interest in an inert atmosphere, then purge gas is changed to oxygen. The time from oxygen introduction to the onset of weight gain is a measure of the oxidative stability of the polymer system. In fact this method can be extended to edible oils also.

1.4.3 Silicon Resin as Thermal Protection Systems

The low pyrolysis rate coupled with high yield of stable silica residue makes silicones a suitable candidate for the thermal protection of rocket from external aerodynamic heating during the ascend through earth's atmosphere (Fig. 8). Polydimethyl siloxane based thermal protection systems have been successfully used in India's Polar Satellite Launch Vehicle (PSLV). Future reusable launch vehicles are to be designed for reentry mission wherein the temperature rise of the rocket body will be much higher than that in an ascending vehicle. Silicone TPS systems are not applicable here. Ceramic tiles and coating are more effective here. Polymer precursors for ceramics such as polysilastyrene are characterized by TG for their silicone carbide yield.

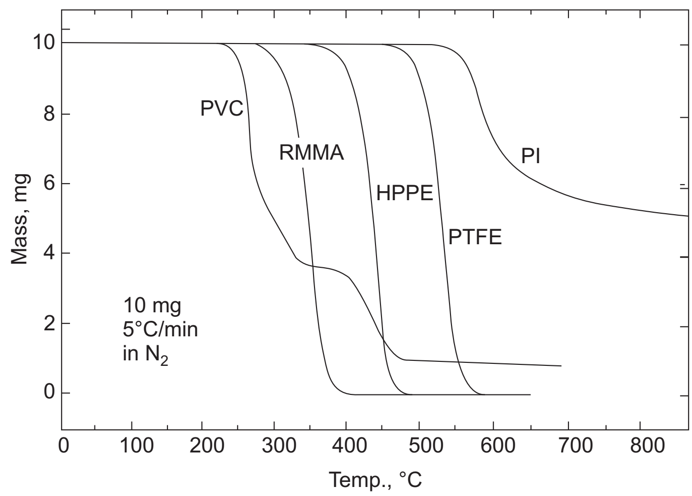


FIG. 5 RELATIVE THERMAL STABILITY OF POLYMERS

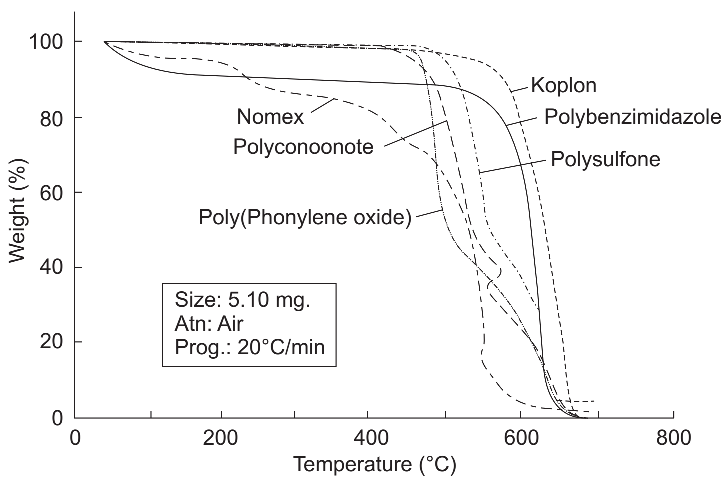


FIG. 6 EVALUATION OF HIGH TEMPERATURE POLYMERS

1.5 Polymer Flammability

TG is extensively used for evaluating the flammability of polymers and composites. The TG curves of the non-flame-retarded and flame retarded polymers differ considerably (Fig. 9). Flame retardancy is believed to result from the presence of one or more key elements in the retardant, such as phosphorous, nitrogen, chlorine, bromine or of a volatile compound. Despite of wide disagreement about how some of these substances suppress burning, there are several accepted theories and their supporting thermoanalytical data. According to condense-phase mechanism the

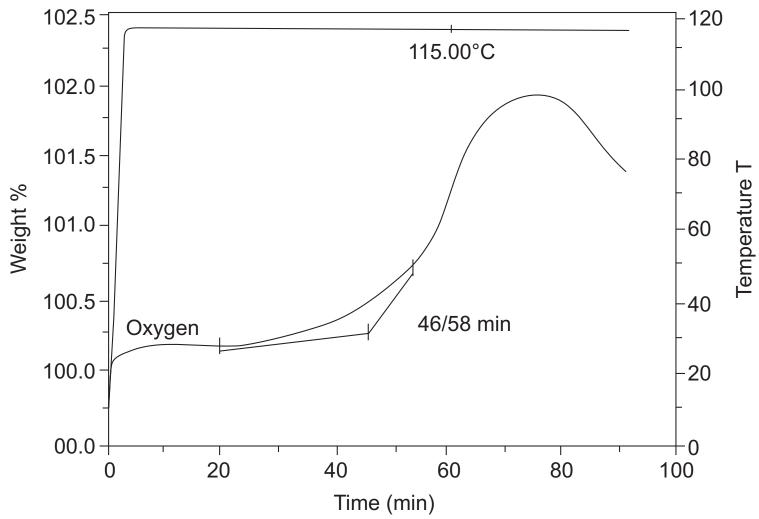


FIG. 7 OXIDATIVE STABILITY OF POLYMERS

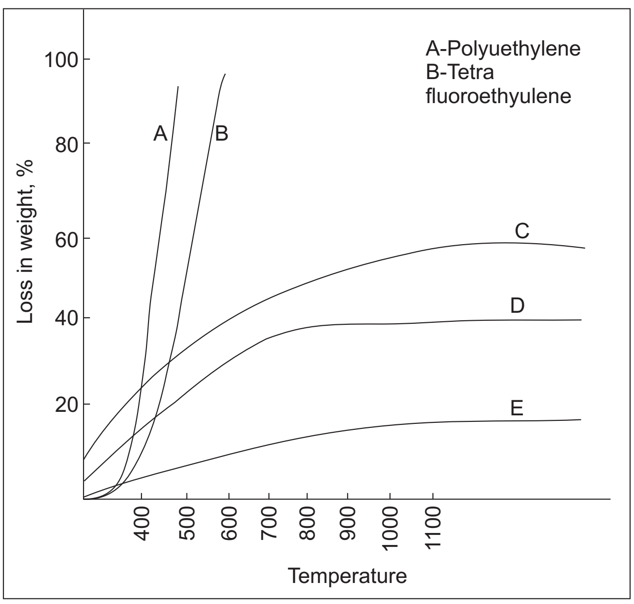


FIG. 8 THERMO-OXIDATIVE STABILITY OF VARIOUS ABLATIVE TYPE POLYMERS

flame retardants have the ability to increase the conversion of polymeric materials to a char reside during pyrolysis and thus decrease the formation of flammable, carbon containing gases. The char helps to protect the substrate by interfering with the access of oxygen.

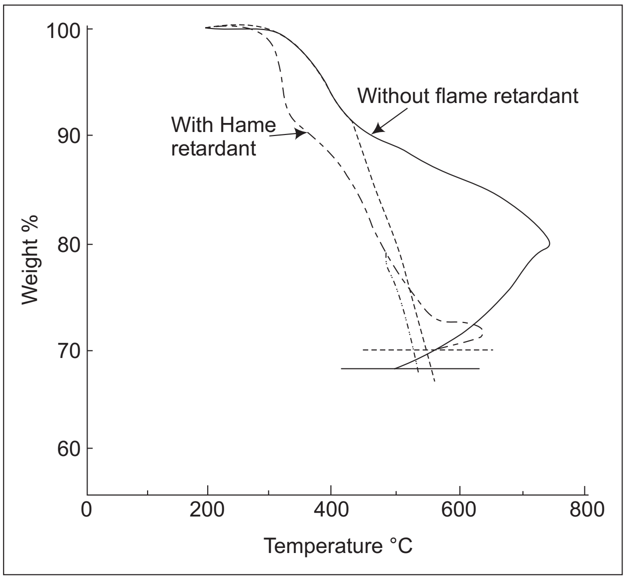


FIG. 9 EFFECT OF FLAME-RETARDANTS ON THERMAL STABILITY

1.6 Char Yielding Ablative Polymers

Carbon and silica composites with ablative polymers, capable of yielding stable char of graphite structure, as matrix resin are used in rocket nozzles through which hot gases (1500-2500K) flow. The major criteria for choice of the resin are (i) stability of the char and (ii) the % char yield. Resole type phenol formaldehyde resin are the conventional ones used. However, recently Phenolic-Triazine (PT) resin which give higher char yield are being considered as better matrix resin for ablatively cooled rocket nozzles.

1.7 Copolymer Analysis

TGA can be effectively used as a tool to differentiate random and block (graft) copolymers if the thermal decomposition pattern is distinctly different for the individual homopolymers. Fig. 10. shows the TG curves of polystyrene, poly α-methyl styrene, poly (styrene-statistical-α-methyl styrene) and the block copolymer of the two. As can be seen the TG curve of the statistical copolymer is distinct from that of the block copolymer. The TGA also furnishes the composition of the latter. The weight loss curve of the random copolymer is smooth and lies in the temperature range between its two homopolymers.

1.8 Life Time Prediction of Polymers

TG provides a method for accelerating the life time testing of polymers so that short term experiments can be used to predict their lifetime. The procedure consists of recording TG curves of the sample at different heating rates as shown in Fig. 11.

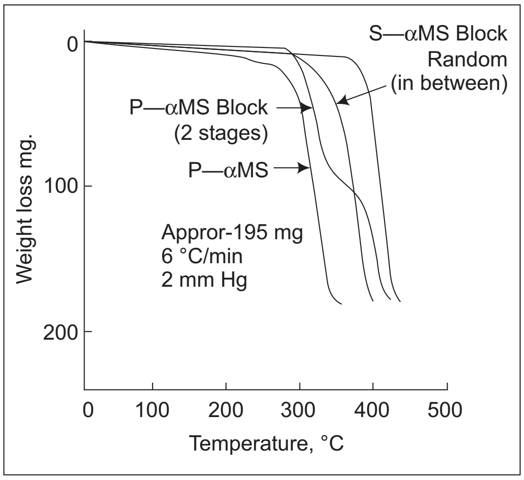


FIG. 10 TG CURVES OF BLOCK VS RANDOM COPOLYMERS

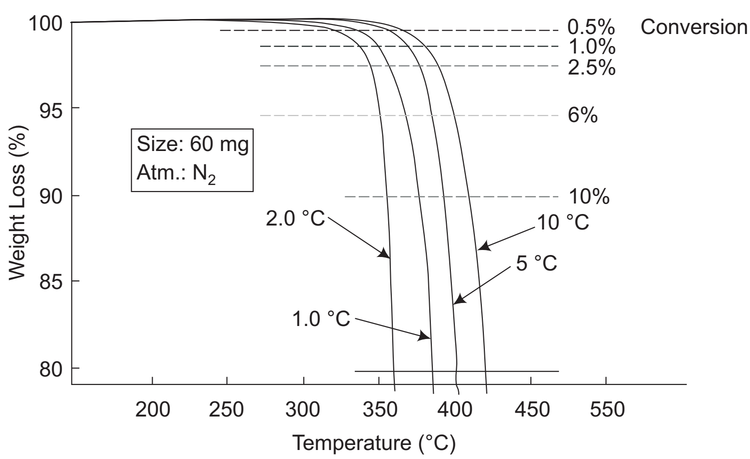


FIG. 11 TG CURVES AT DIFFERENT HEATING RATES

The temperature (T) for a selected value of conversion (say 5%) is noted from the TG curve. From the slope of the plots of heating rate (ϕ) versus 1/T (Fig. 12) the activation energy "E" is evaluated by the method proposed by Flynn and Wall using the equation,

$$E = -R/b[d\log \phi/d(1/T) \tag{1}$$

where E = activation energy (J/mol), R = gas constant (8.314 J/mol K), T = temperature at constant conversion (K), ϕ = heating rate (°C/min), b = constant (0.457) evaluated from the Flynn and Wall method.

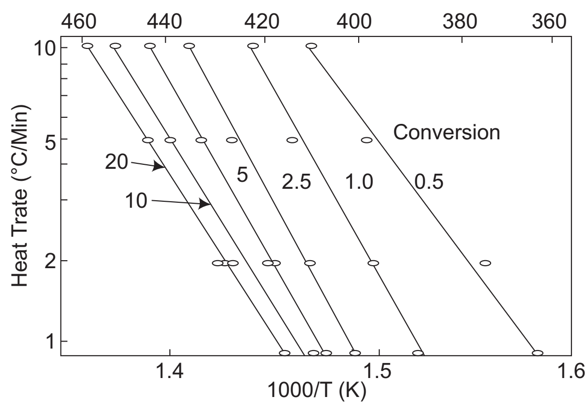


FIG. 12 PLOTS OF HEATING RATE (ϕ) VERSUS 1/T

Estimated Time of Failure (t_f) can be evaluated using the Toop equation,

$$\ln t_f = E/RT_f + \ln [(E/\phi R) . p(X_f)] \tag{2}$$

where t_f = estimated time of failure (min), T_f = Failure temperature (K), $P\ (X_f)$ = a function whose value depends on E at the failure temperature (Fig. 13).

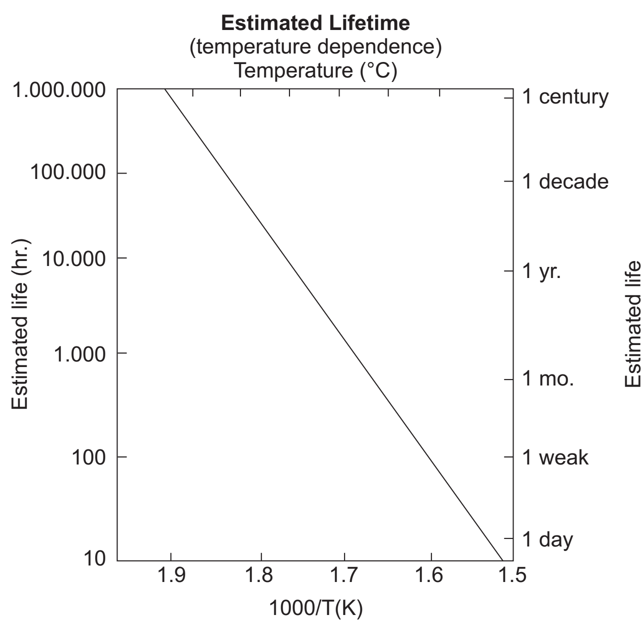


FIG. 13 LIFE TIME PREDICTION CURVE

1.9 Thermal Decomposition Kinetics

TGA has been most extensively used for the evaluation of thermal decomposition kinetics of polymers to evaluate the basic parameters of the reaction viz., rate constants (k), activation energy (E), pre-exponential factor (A) and order of reaction (n). The reaction rate is related to the degree of conversion α obtained from the TG curves (It can also be obtained from DSC curves; however for decomposition reaction α from TGA is more reliable).

Isothermal method is the conventional method of computing kinetic parameters. At constant temperature

$$d\alpha/dt = \mathrm{k}(1-\alpha)^n \tag{3}$$

which on rearranging and integration gives,

$$g(\alpha) = \frac{1-(1-\alpha)^{(1-n)}}{(1-n)} = kt \text{ for } n \neq 1 \text{ and} \tag{4}$$

$$-\ln(1-\alpha) = kt, \text{ for } n = 1 \tag{5}$$

(The LHS of the above equation is called $g(\alpha)$). From the values of k at different temperatures and by applying Arrhenius equation $k = Ae^{-E/RT}$ or $\ln k = \ln A - E/RT$, the kinetic parameters are computed.

Non-Isothermal method offers the simplicity of computing the kinetic parameters from a single measurement. For non-isothermal condition the rate expression is

$$d\alpha/dt = Ae^{-E/RT}(1-\alpha)^n \tag{6}$$

The integral method for solving the above equation yields the best results. Introducing the terms for linear heating rate $\phi = dT/dt$, we get

$$\int_0^{\alpha} \frac{d\alpha}{(1-\alpha)^n} = \frac{A}{\phi}\int e^{\frac{-E}{RT}} dT$$

As seen above, the LHS is g (α), but the RHS cannot be integrated in a closed form. Among the various approaches for solving the exponential integral, the following three give most satisfactory results.

Coats- Redfern Equation

$$\ln\left[\frac{g(\alpha)}{T^2}\right] = \ln\left[\left(\frac{AR}{\phi E}\right)\left(1-\frac{2RT}{E}\right)\right] - \left[\frac{E}{RT}\right] \tag{7}$$

MacCallum-Tanner Equation

$$\log_{10}[g(\alpha)] = \log_{10}\left[\frac{AR}{\phi E}\right] - 0.485\,E^{0.435} - \left[\frac{(0.449+0.217E)\times 10^3}{T}\right] \tag{8}$$

Madhusudhanan- Krishan –Ninan Equation

$$\ln\left[\frac{g(\alpha)}{T^{1.9215}}\right] = \ln\left[\frac{AR}{\phi E}\right] + 3.7721 - 1.9215 \ln E - 0.120309\left[\frac{E}{T}\right] \tag{9}$$

where α is the fractional decomposition, T is the temperature in K, A is the pre-exponential factor, E is the energy of activation, ϕ is the heating rate and R is the gas constant.

Though one school of thermal analysis advocates isothermal as the only correct method of computing kinetic parameters, an unbiased analysis reveals that at infinitesimal time interval the non-isothermal experiments take place isothermally so that a non-isothermal TG curve can be considered as an integral of infinite number of such isothermal runs – an argument substantiated by the close agreement in result obtained by the two methods for a large number of organic and inorganic materials.

Glass-phenolic laminates are used as ablative thermal protection systems for rockets, and their thermal decomposition has been described by Arrhenius type equations. As the heating rates experienced by heat shield in flight environments are different from those employed in usual TG experiments, it has been attempted to evolve correlation between heating rate and kinetic parameters, so that TG results are more useful in evaluating heat shield materials.

$$\log E = C_1 - C_2\Phi + C_3\Phi^2 - C_4\Phi^3 \tag{10}$$

$$A = k_1 + k_2/\Phi + k_3/\Phi^2 + k_4/\Phi^3 \tag{11}$$

where C_1 to C_4 and k_1 to k_4 are empirical constants.

Hydroxyl terminated polybutadiene (HTPB) is termed as “workhorse” polymeric fuel binder for modern solid propellants and it is prepared by either free radical or anionic polymerization of butadiene. A comparison of the overall thermal decomposition kinetics of the two types of HTPB with those of CTPB, (carboxyl terminated polybutadiene – another fuel binder) has shown that the values of E and A are not dependent on the method of polymerization and the nature of the terminal groups. However, when HTPB having a broad molecular weight distribution was fractionated based on molecular weight parameters and correlation can be summarized as

$$E \text{ (or } \log A) = k_1 - k_2 \cdot M_n \ (M_w \text{ or } M_z) \tag{12}$$

Detailed thermal studies in conjunction with analysis of volatile products have shown that HTPB undergoes two stage decomposition, the first stage corresponding to depolymerisation coupled with cyclisation and cross linking and second stage to pyrolysis. And it is the first stage decomposition kinetics, which is profoundly influenced by the molecular weight parameters, whereas the second stage kinetics is bis (3,3) itaconimido phenyl) sulfone the kinetic parameters are not appreciably affected by the degree of polymerization.

2. APPLICATION OF DTA/DSC

The information obtained from DTA and DSC is more or less identical, except that DSC is more sensitive and the data is more quantitative. Hence the two techniques are discussed together in this section. Perhaps the greatest number of applications of DTA and DSC in recent years has been in the area of polymeric materials. These two techniques are routinely used to measure glass transition temperature (T_g), melting point (T_m), degree of cure, degree of crystallinity, heat of fusion and/or crystallization, decomposition temperature, and numerous other parameters. A schematic DTA/DSC curve illustrating the various thermal process of a typical polymer is given in Fig. 14. In reality all these transitions may not be so well defined for all the polymers.

A schematic chart showing the application of DTA and DSC to polymers is given in Fig. 15.

2.1 Identification of Polymer (Blends)

Polymeric materials either alone or in mixture exihibit their characteristic peaks and hence DSC/DTA is an effective tool to identify them. The identification of polymer blends is illustrated in Fig. 16 which shows the DSC curves of a physical mixture of seven commercial polymers, viz., high-pressure polyethylene (HPPE), low-pressure polyethylene (LPPE), polypropylene (PP), polymethylene (POM), Nylon 6, Nylon 66 and polytetrafluoroethylene (PTFE). Each component shows its own characteristic melting endothermic peak at 108, 127, 165, 174, 220, 257 and 340°C respectively. PTFE has an additional peak at about 20°C, corresponding to crystalline transition.

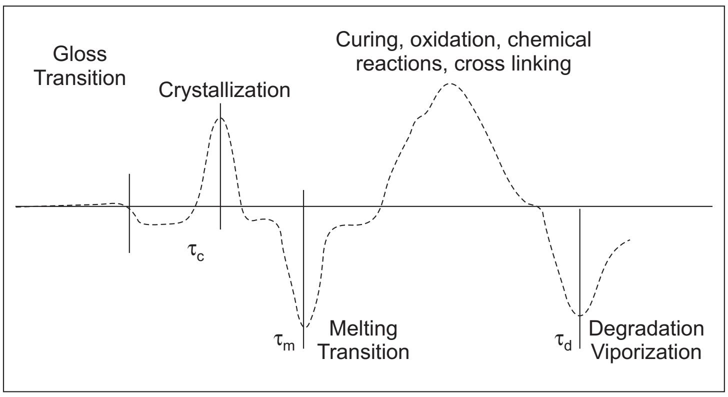


FIG. 14 SCHEMATIC DTA/DSC CURVE OF POLYMER

Polymers			
DTA/DSC			
Qualitative measurements		Quantitative measurements	
Detection of Tg, Tm, etc.	Qualitative Identification	Thermochemical Measurements	Kinetics
Thermal stability	Flammability detection	ΔH_f	Decomposition
Oxidation reaction	Polymerization	Specific heat	Crystallisation
Others	Cure reactions	ΔH Polymerisation	

FIG. 15 APPLICATION OF DTA AND DSC TO POLYMERS

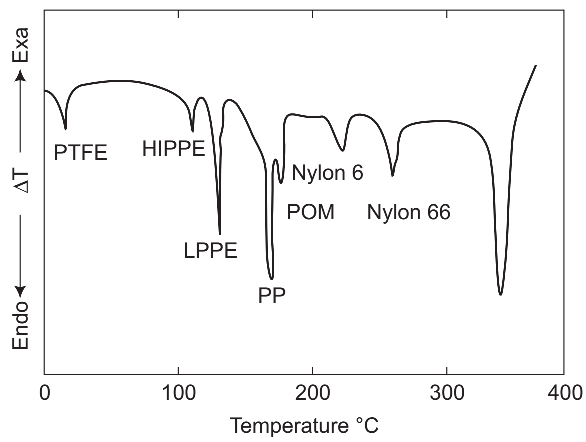


FIG. 16 DSC CURVES OF A PHYSICAL MIXTURE OF SEVEN COMMERCIAL POLYMERS

DSC/DTA curves can be used as a characterization index for polymers and offer unique advantage over other isothermal methods, especially for insoluble and intractable polymer systems.

2.2 Quantitative Analysis

The area under the curve in DTA/DSC is proportional to the heat of transition and the mass of the reactive sample.

$$\Delta Hm = KA \tag{13}$$

where ΔH is a heat transition (or reaction), m is the mass of the reactive sample, K is the calibration constant and A is the area under the DTA/DSC curve. This simplified expression relating peak area with ΔH and m is based on the following assumptions:-

(i) The thermo physical properties of the sample do not change with the temperature and they remain same before and after the transformation – an assumption which is invalid in most of the cases leading to an error of 5–10%, unless identical sample used as calibration standard.

(ii) The changes occur under quasi-stationery conditions achieved in linear and slow heating rates, the latter producing low temperature effects can adversely affect the accuracy.

The calibration coefficient, K is determined using compounds of known heats of transition (usually heat of fusion of pure metals). The calibration coefficient is temperature dependent for DTA whereas it is independent of temperature for DSC sand thus the latter is more suitable for quantitative estimations.

From the area of characteristic peak, sample mass and known heat of transformation it is possible to estimate the amount of polymer in blends. Thus the linear content of polyethylene blends was estimated from DTA curves (Fig. 17). The second peak (ΔT_{max} = 134°C) corresponds to crystals of linear polymer.

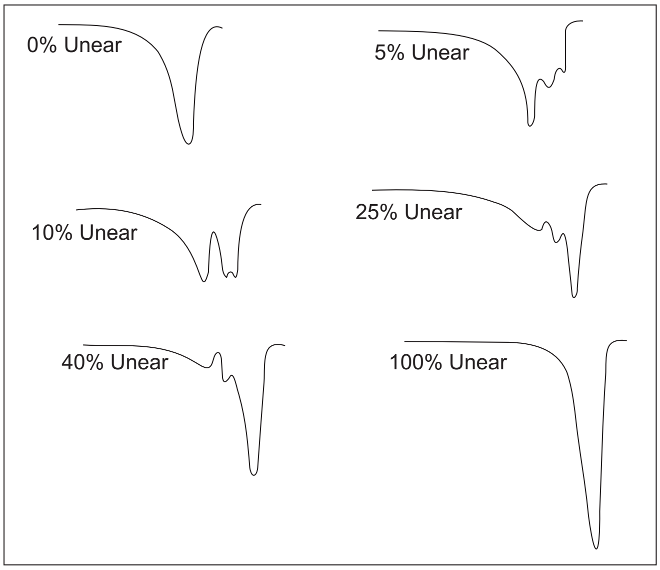


FIG. 17 DTA CURVES OF PE BLENDS SHOWING LINEAR CONTENT

2.3 Oxidative Stability of Polymers

Oxidative stability is an important polymer property to be considered for many applications such as fiber production. DSC is an effective technique for rapidly evaluating the oxidative stability of polymers. Oxidative degradation is seen as an exothermic peak in DSC and extrapolated onset temperature of the peak gives an idea about the relative thermal stability of the polymer (Fig. 18).

2.4 Determination of Polymer Crystallinity

Perhaps no fundamental property affects the physical properties of a polymer in so general a way as the degree of crystallinity.DSC provides a rapid method for determining polymer crystallinity based on the heat required to melt the polymer. The method is based on measurement of polymer sample's heat of fusion (ΔH_f) and comparing it with that of a perfectly crystalline sample, ΔH_f° (degree of crystallinity, $x = \Delta H_f/H_f^\circ$). A 100% crystalline polymer is a hypothetical situation, and hence 'x' is normally obtained by comparison to a standard of known crystallinity. An understanding of the degree of crystallinity for a polymer is important since crystallinity affects physical properties such as storage modules, permeability, density, and melting point. While most of these manifestations of crystallinity can be measured, a direct measure of degree of crystallinity provides a fundamental property from which these other physical properties can be predicted.

Figure 19 shows the melting endotherm for a sample of polytetrafluoroethylene. From differences in the melting endotherm in DSC after variation in thermal history of the samples it is possible to learn a lot optimizing conditions.

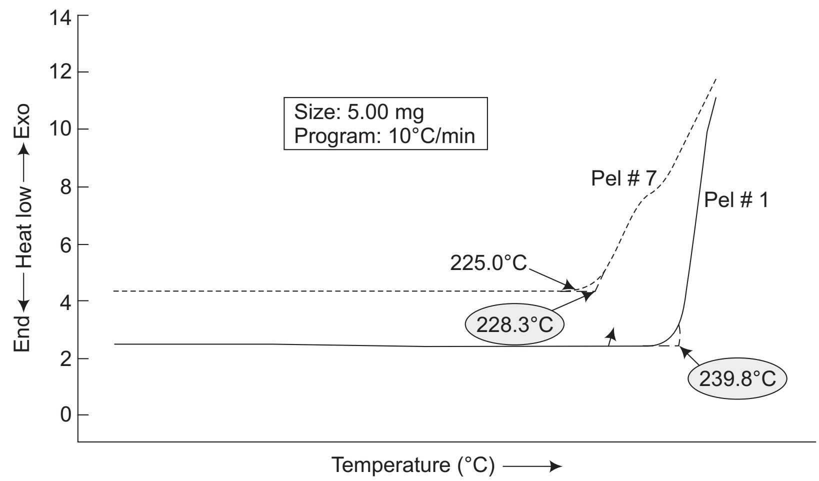


FIG. 18 OXIDATION INDUCTION TEMPERATURES IN POLYETHYLENE TEREPHTHALATE

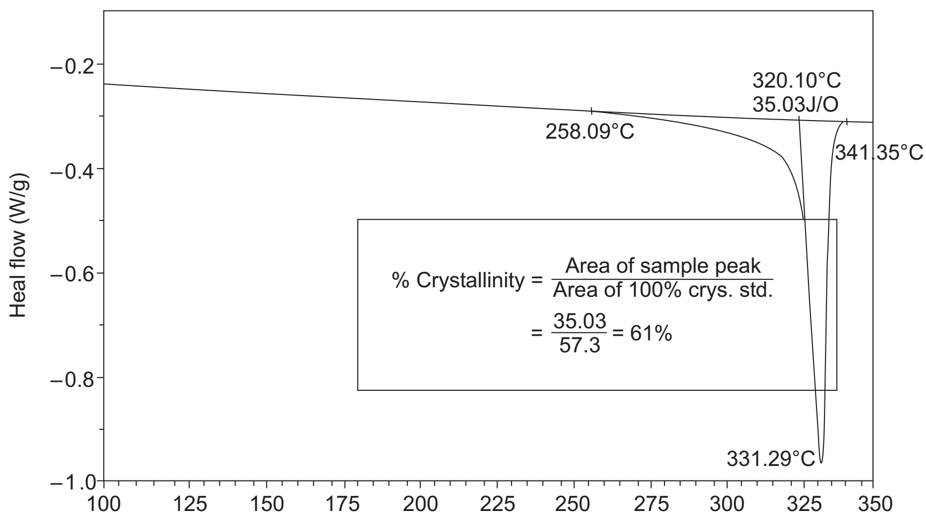


FIG. 19 CRYSTALLINE CONTENT IN PTFE BY DSC

2.5 Detection of Glass Transitions

Glass transition temperature, T_g is the temperature at which the amorphous phase of the polymer is converted between rubbery and glassy states. Since cooling causes the polymer to freeze thermodynamic transition related to the flexibility of a polymer backbone, the secondary interchain forces acting on it, and the free volume of the system. The straightforward determination of T_g and T_m via DSC has been quite helpful for polymer characterization.

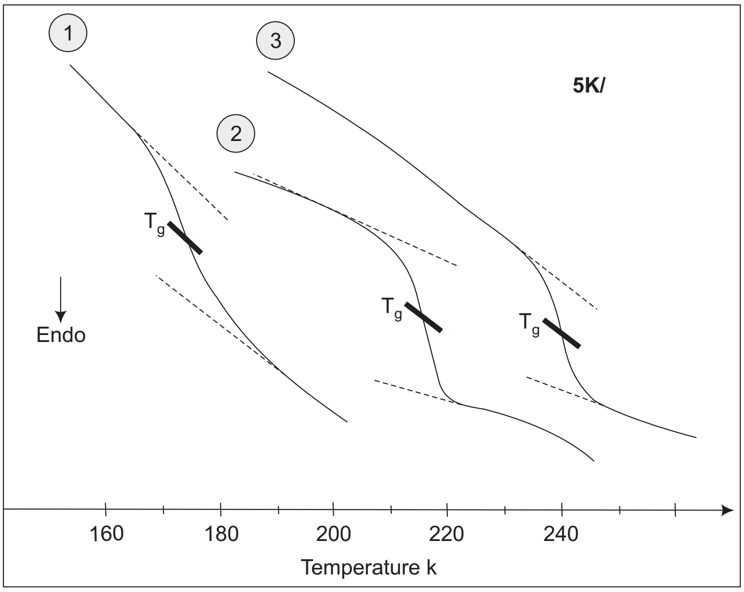


FIG. 20 GLASS TRANSITION TEMPERATURES BY DSC

2.6 Specific Heat

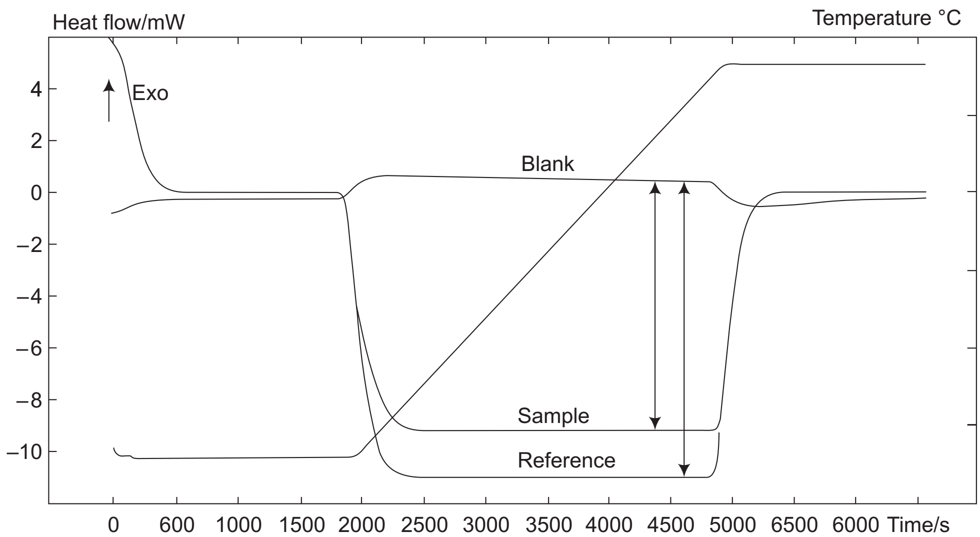
One of the major applications of polymer and composites is as thermal insulation. For the calculation of the optimum thickness of the insulation, it is necessary to know the specific heat of the material as a function of temperature. DSC offers the unique advantage of providing this data quickly. The method is illustrated in the following figure (Fig. 21). From the heat flow values of blank (empty container), reference (inert material, e.g. alumina), and sample mass, *m* and the specific heat of reference, the *Cp* of a sample at a given temperature C_p (T) is calculated as:

2.7 Cure Reaction Studies

The DSC can be used for monitoring the cure reaction. From the DSC curve the extent of cure can be determined. Curing studies can be done using two methods namely (i) isothermal scan or (ii) dynamic scan. The area under the exotherm gives the enthalpy change of the cure reaction. A typical curing of an epoxy resin using dynamic DSC is given below (Fig. 22). DSC is also used to optimize cure conditions and establish 'residual cure' attributed to further polymerization of under-cured resin.

The manufacturing of composites structures is usually a two-step process, and the first step is the impregnation of the reinforcement with the matrix resin to form the 'prepreg', which is cured at the desired temperature and pressure latter on. Since the curators are also added to the resin prepreg stage, the cure reaction often sets in on storage (though at a lower rate).

Such prepregs are to be stored at sub-ambient conditions is the usage is delayed. The advancement of cure reaction beyond a certain limit can adversely affect the flow characteristics of



$$C_P(T) = \frac{HF_{sample} - HF_{blank}}{HF_{ref} - HF_{blank}}, \frac{Mass_{ref}}{Mass_{sample}}.C_{P_{ref}}(T)$$

FIG. 21 SPECIFIC HEAT MEASUREMENT USING DSC

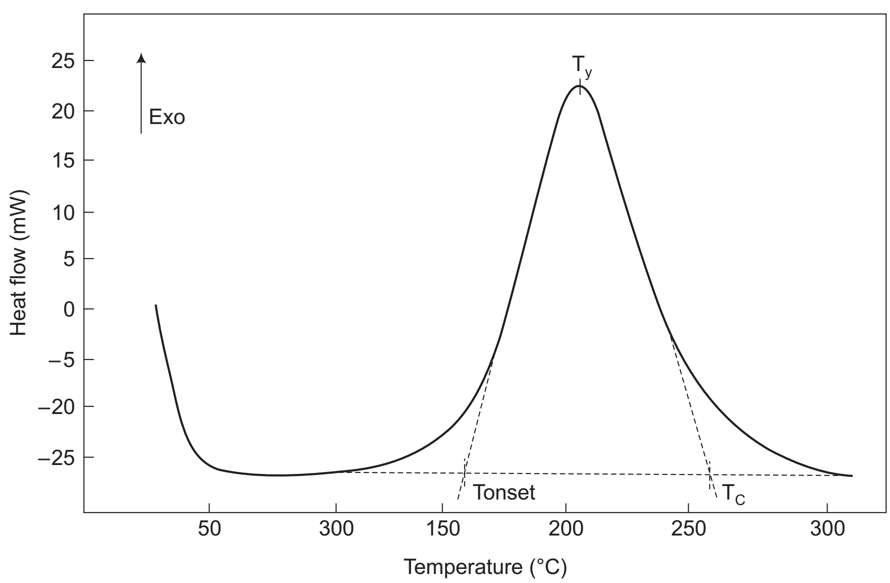


FIG. 22 DSC CURE EXOTHERM OF AN EPOXY RESIN

the resin during further processing, leading to products of inferior properties. Hence it is necessary to establish the degree of cure of the prepreg, which is readily obtained from the DSC cure exothermic peak.

2.8 Evaluation of Purity of Chemicals and Monomers for Polymer Synthesis

The melting endotherm of DSC is made use of to evaluate the purity of high pure organic compounds, based on Van't Hoff's equation. A typical example is shown in Fig. 24. Purity analysis a sample of ammonium dinitramide, an advanced propellant oxidizer, was carried out by DSC and the results are also given in Fig. 23.

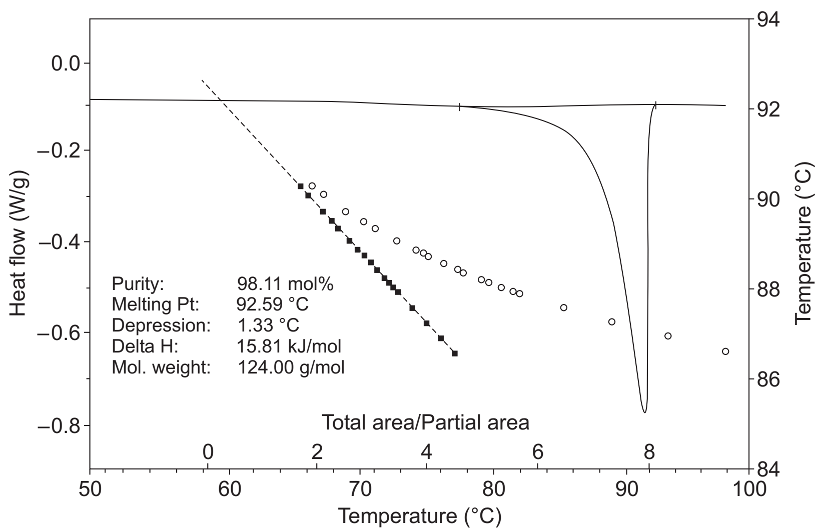


FIG. 23 PURITY ANALYSIS BY DSC OF A SAMPLE OF AMMONIUM DINITRAMIDE

The monomer purity plays a critical role in deciding the degree of polymerization of condensation polymers and DSC is an ideal tool in this regard. It gives reproducible results as seen in example taken from monomers for the synthesis of polypyromelittimide.

Sample	T_{max} *(°C)*	*Purity (%)*
DDE	190.9	99.98, 99.96, 99.97
PMDA	286.3	99.86, 99.84, 99.86

2.9 Kinetics of Cure Reaction

Kinetic parameters can be computed from the DSC curves. The heat of reaction is used to quantify the extent to which the reaction takes place (α). The reaction rate $d\alpha/dt$ is related to the rate of heat generation dH/dt, as represented by the equation,

$$d\alpha/dt = [dH/dt]/\Delta H \text{ and} \tag{15}$$

$$(1 - \alpha) = 1 - [\Delta H_r/\Delta h] \tag{16}$$

The temperature dependence is assumed to reside in the rate constant through an Arrhenius relationship giving activation parameters. All the equations derived for kinetics analysis for TG can be employed for analyzing the DSC data also. There are specific methods like the Borchardt and Daniels methods developed for kinetic analysis of ESC data.

Several interesting findings have emerged from DSC kinetic analysis. Thus, quantitative correlation between heat of polymerization and activation parameters (E and A) for the polymerization reaction has been established in our laboratory for the cure reactions of (i) reactive poly amide (made from dimmer acid) with epoxide and (ii) thermal polymerization of N, N-bismaleimido- p,p'-diphenylmethane, modified by the addition of N,N-bisitaconimido-p-p' diphenyl ether. DSC studies on the cure reaction kinetics of HTPBTDI (toluene diisocyanate) system have shown that the computed kinetic parameters can be used as a means for evaluating Lewis acid (BF_{3-} & $AlCl_3$) as reaction retardants, with a view to use them as a pot-life extenders in solid propellant formulations.

2.10 Modulated DSC (MDSC)

In MDSC the sample is exposed to a linear heating method which has a superimposed sinusoidal oscillation (temperature modulation) resulting in a cyclic heating profile. Deconvolution of the experimental heat flow provides the total heat flow (as from conventional DSC) and also the reversing (heat capacity related) and non-reversing (kinetic) components. This makes it possible to detect Tg in cases where overlapping reaction make it difficult in normal DSC (Fig. 24) analysis.

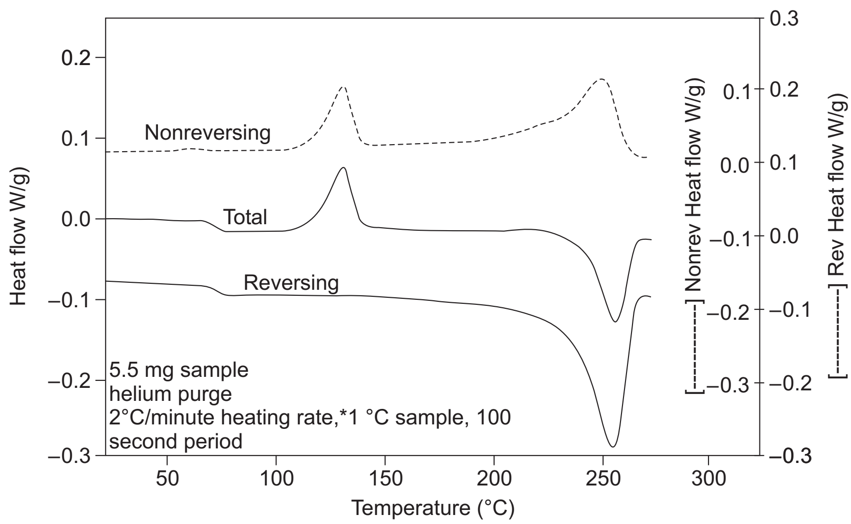


FIG. 24 TOTAL, REVERSING AND NON-REVERSING HEAT FLOW COMPONENTS IN MDSC

2.11 Combined DSC/DTA– TG Studies

The information obtained from DSC/DTA and TG are complementary and complete information can be obtained only when both the studies are carried out on a given sample. Thus our recent TG-DSC studies on HTPB, carried out in air and nitrogen atmosphere have shown three transitions. The product isolated after the first transition (exothermic DSC peak and mass gain in TG at 170–240°C), seen only in air atmosphere shows addition of oxygen to the polybutadiene back-bone (from FTIR). The study has helped in understanding the kinetics and mechanism of the reaction responsible for the viscosity build-up of the resin and thereby increase the useful storage life of the resin.

Combined TG-DSC studied on a new class of low melting, bismaleimide resin on monomers of 2,2'bis [4 – (4–maleimide phenoxy) phenyl] propane and o,o'–diallyl bisphenol have helped in assessing the heat and kinetics of cure reaction and the thermal stability and clear content of the cured product. Similar studies on silica–phenolic prepregs (for making composites for nozzle application) have been carried out to evaluate the effect of phenol and other additives on heat and kinetics of cure reaction and the char yield of the cured product.

3. APPLICATIONS OF TMA

3.1 Thermal Expansion and Glass Transition Temperatures of Polymers

TMA is used to measure the thermal expansion characteristics of polymers and the glass transition temperature at which there is a characteristics increase in the thermal expansion profile. Figure 25 shows the thermal expansion and glass transition temperature of an epoxy printed circuit board under expansion mode. A linear expansion is observed up to 100°C above which a rapid increase in the expansion is observed. This change is due to glass transition phenomenon. The extrapolated

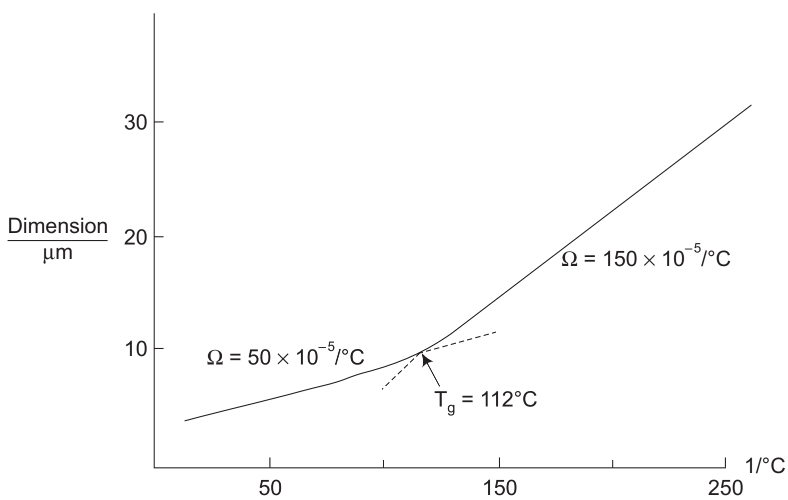


FIG. 25 GLASS TRANSITION BY TMA

expansion curve meets at a point, which represents the T_g of the sample. TMA offers specific advantage of determining T_g of polymers whose DSC transition is not quite sharp.

3.2 Melting/Softening point

TMA experiments can be performed with penetration probe under suitable loading conditions. The glass transition is not sharp as in the case of the expansion mode; however the melting temperature is quite distinctly seen. The softening point values determined by TMA in penetration mode are in close agreement with standard VICAT softening temperatures.

3.3 Characteristics of Multilayer Thin Films

Multilayer thin films are becoming more common for applications such as packing. TMA, because of its ability to measure small dimensional changes, is ideal for evaluating thin films. Displayed in Fig. 26. are the result for a multilayer packing film comprised of a metallic substrate coated on both sides with polymer. The TMA curve (penetration) shows two sharp negative dimensional change at 103°C and 258°C. These two penetrations indicate that the film is probably a three layer film containing two polymer layers and the metal foil layer.

In an attempt to identify the nature of the individual polymer film layers, several homopolymer films were evaluated and from their sharp penetration at 105°C and 255°C, the layers were identified as polyethylene and polyester respectively. TMA thus provides quick alternative to spectral technique especially when multilayers are involved.

TMA with parallel plate rheometer can measure the flow characteristics of polymers, as function of temperature. From these parameters such as gel time, cure time, flow index etc can be evaluated.

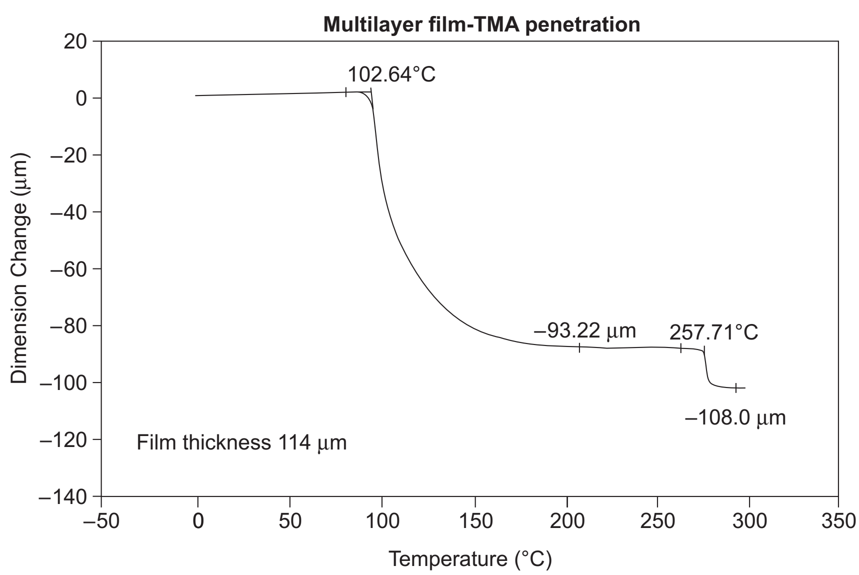


FIG. 26 TMA IN PENETRATION OF A MULTILAYER PACKING FILM

Using approximate probe configurations, the creep characteristics of polymers can be measured. TMA has been used in the optimization of post baking of polymide polymer, prepared from N,N'– bis (itaconamic acid) and p,p'– diphenyl methane. The optimum time was chosen to get the maximum value of T_g for the system. The effect of fiber orientation on the thermal expansion of fibre-reinforced structural composites for space applications is evaluated from TMA studies.

4. APPLICATIONS OF DMA

Dynamic mechanical analysis (DMA) also known as Dynamic Mechanical Thermal Analysis (DMTA) is often able to detect transitions which are difficult to observe by other techniques. It has special applications in pre-pregs and composites, where the DSC sensitivities are low due to the low polymeric content. A host of primary and secondary viscoelastic parameters are derived from the stress–strain frequency and temperature data.

Primary Viscoelastic Functions	*Secondary Viscoelastic Functions*
Flexural Storage Modulus (E′)	Flexural Storage Compliance (S′)
Flexural Loss Modulus (E″)	Flexural Loss Compliance (S″)
Shear Storage Modulus (G′)	Shear Storage Compliance (J′)
Shear Loss Modulus (G″)	Shear Loss Compliance (J″)
Tan Delta (E″/E′ or G″/G′)	Recovery
Flexural Creep Compliance (S)	Flexural Creep Modulus (1/S)
Shear Creep Compliance (J)	Shear Creep Modulus (1/J)
Flexural Stress Relaxation Modulus (E)	Flexural Stress Relaxation Compliance (1/E)
Shear Stress Relaxation Compliance (G)	Shear Relaxation Compliance (1/G)
Relative Creep Complex	Viscosity (η*)
Stress Dynamic	Viscosity (η′)

4.1 Monitoring of Thermoset Cure

Cure time, temperature and onset of gelation are important criteria for optimizing the process parameters of polymer based composites. Since DMA measures stiffness, it is able to monitor thermoset cure as a function of time or temperature. In this example, (Fig. 27) a liquid resin was coated onto a fiber glass cloth and the curing was done at 177°C. Approximately 25 minutes later, the onset of cure is seen in the frequency (stiffness) curve. Both the frequency and damping curves show complete cure at about 215 miuutes after loading the epoxy. Thus DMA is a rapid and inexpensive method for cure-time-temperature prediction and optimization of other process conditions of pre-pregs and composites.

A programmed time-temperature profile of slow heating to 80°C, hold at 80°C, again slow heating and hold at 170°C done by DMA has yielded optimum properties for silica phenolic composites. In the initial phase of heating, softening of the resin occurs leading good flow, which ensures good compaction under pressure. In the second phase of heating, gelatin and cross-linking take place. The optimized time-temperature profile ensures good composite properties.

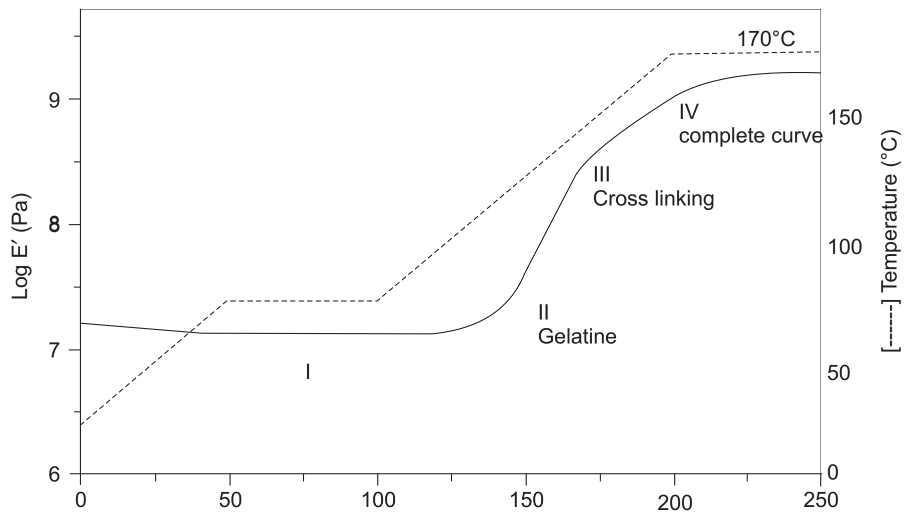


FIG. 27 TIME– TEMPERATURE CURE – OPTIMISATION – DMA

4.2 Characteristics of Thermoplastics and Elastomers

DMA is an extremely sensitive technique, making it the technique of choice for analysis of polymer blends, copolymers and composite structures. It is sensitive enough to accurately measure the glass transition (T_g) of the semicrystalline polypropylene sample (Fig. 28) through a maximum. Tan δ is the ratio of loss modulus and storage modulus.

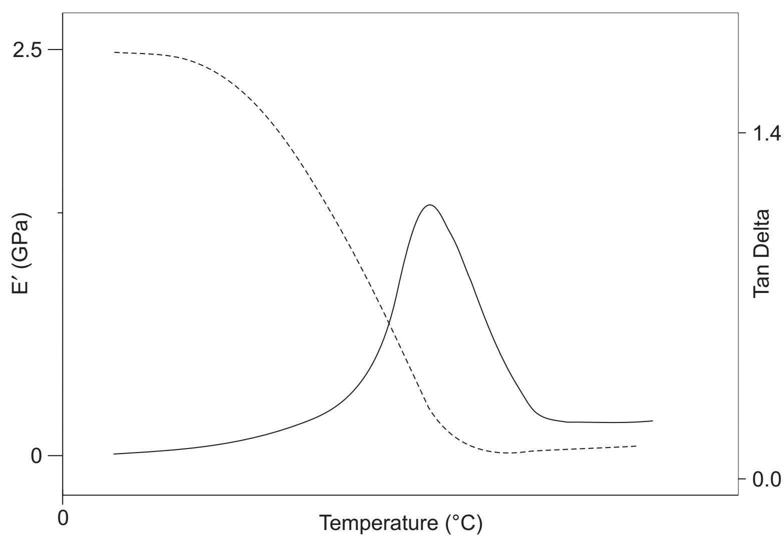


FIG. 28 GLASS TRANSITION OF SEMICRYSTALLINE POLYPROPYLENE SAMPLE

DMA provides a profile of the sample's hardness (stiffness) and energy– dissipation capability versus temperatures, which act as an index for the choice of suitable elastomers as vibration damping materials. Impact resistance is a critical end-use property for many commercial plastic products. DMA provides a convenient method for evaluating impact resistance in materials such as polyethylene-modified polypropylene. The damping information obtained can be correlated with impact properties in materials.

4.3 Failure boundaries

The time-temperature superposition capability of DMA makes it possible to generate the master curve and WLF shift factor (Fig. 29). From the stress-strain data at various strain rates and temperature and the shift factor, it is possible to predict a failure boundary within which it is safe to use polymer or composite (Fig. 30). Such prediction are very important in critical application like aerospace and rocketry.

5. CONCLUDING REMARKS

In this article it is attempted to present typical applications of four important thermoanalytical techniques for the characterization of polymeric materials including composites. A few examples from our laboratory are cited to show the vital role played by TA techniques in development of polymers and composites for space applications. As the future technologies warrant the development of speciality materials needing stringent quality requirements in term of performance and resistance to extreme and hostile environments, thermoanalytical techniques are also evolving

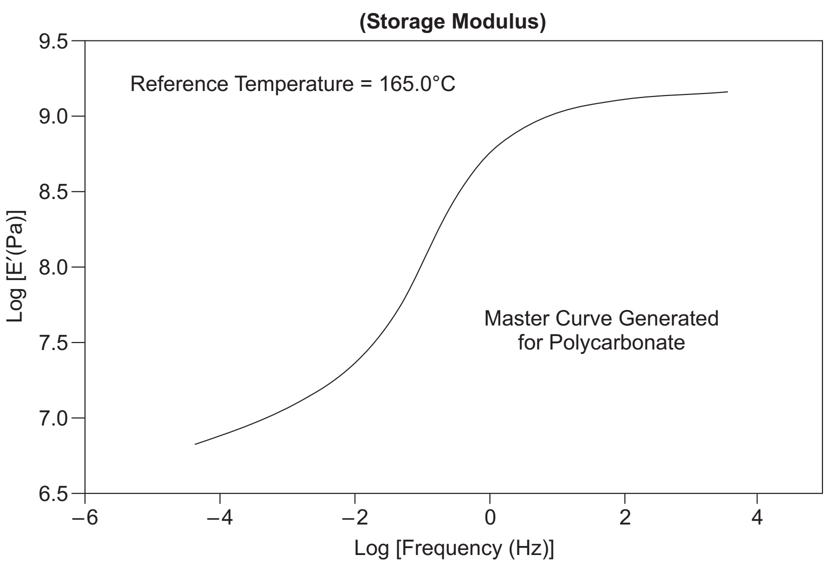


FIG. 30 MASTER CURVE GENERATED FOR THE STORAGE MODULUS OF POLYCARBONATE

parallely. The introduction of novel techniques, Modulated DSC etc., are testimony to these facts. These advanced techniques will provide information, which is normally inaccessible by the conventional thermal means. It is beyond the scope of this article to illustrate them all, as these techniques themselves merit separate discussions.

ACKNOWLEDGEMENTS

The author expresses his gratitude to Dr. Bina Korah and Dr. R. Rajeev for their help in preparing this article.

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

THE NEXT GENERATION BIOLOGICAL ENGINEERING

PURNENDU GHOSH*

There is optimism in the biotech world. The optimism is due to its ability to manipulate and build living things. This ability is becoming as important as the ability to manipulate the flow of electrons and bits. In order to retain this optimism engineering and biology must come closer to each other.

Biology is now so intimately connected with almost all engineering disciplines. Modern biology as basic science course along with physics, chemistry, and mathematics is already being introduced in several undergraduate engineering curricula. Questions such as ‘how much biology is enough for engineering students’, ‘how do we simplify biology for engineers’ are being debated. An engineer’s approach to understand complex biological systems is different. Accordingly biology should be taught differently to engineers. There is a view which says that “function-based approach with the idea of nature as the designer, and evolution as the design tool” should be followed to teach biology to engineers. It means “start with how it works, then talk about the parts”.

BIOPROCESS ENGINEERING

The traditional role of a bioprocess engineer has been cultivation of microbial, plant and animal cells (both in the ‘free’ and ‘immobilized’ state), understand cell’s and enzyme’s kinetic behavior, improvise methods for the transfer of oxygen from the gas-to- liquid-to-cells and to remove metabolic heat, optimize, design and operate bioreactors based on scale-up/scale down concepts, design efficient purification and recovery processes for biological products, etc. With the shift in focus from cellular to molecular processes, product quality rather than its concentration in the broth becomes important. As a consequence of this downstream processing has become as relevant as upstream processing; for some products downstream has become more important than upstream. Accordingly size of the bioreactor and scale of operation are no longer the sole criteria of economic success.

Bioprocess engineers have understood the need for better interaction at the molecular and whole cell level as well as at the whole cell and community level. They want to better understand the

*Birla Institute of Scientific Research, Statue Circle, Jaipur-302001, Rajasthan

behavior of biological systems when perturbed by genetic, chemical, mechanical, or material interventions or subjected to pathogens or toxins. They have realized that there is a need to evaluate, and if needed, to manipulate metabolic networks. Biocatalysts are also expanding their horizon; they are now being designed for specific uses. New bioinformatics based strategies are being developed for protein design. Efforts are being made to tailor fundamental properties of biocatalysts in terms of stability, activity, and surface properties. Newer tool boxes are evolving to develop biocatalysts with new abilities. Development and widespread adoption of bio-sensors and bio-monitors that produce biological data relevant to all levels of biological organization is one of the major reasons of optimism among biological engineers.

BIOMEDICAL ENGINEERING

A revolution of sorts is also taking place in the field of biomedical engineering, where the major concern is to establish an interface between man and machine that are relevant for healthcare and medicine. A biomedical engineer studies biomechanics at the molecular, cell, organ, and whole-body levels. Cell engineering has extended its boundary beyond the cells. It has included the use of cells and part of cells to build structures, such as tissue and ultimately the whole organs. Cell engineering uses, besides cell biology, engineering principles, chemistry, nanotechnology, and material science to control cell behavior. The major application of biomedical engineering is in the areas of biomaterials and tissue engineering. It develops drug delivery platforms, technologies for imaging the body and its components, devices for replacing neurological function, biomaterials for use in the body, organ-replacement systems, artificial blood, etc.

BIOLOGICAL ENGINEERING

Both bioprocess and biomedical engineering are application based engineering disciplines like ceramic, agriculture, and petroleum engineering. In the changing biotech scenario it has become essential to bring bioprocess and biomedical engineering disciplines together on the same platform. Biological engineering, the merged discipline, will have extended territory and knowledge base. This broad based engineering programme should be developed in the pattern of science based engineering disciplines such as electrical or mechanical engineering. Science based engineering disciplines are not specific to any particular industry; electrical engineering is not only restricted to electrical industry but also serves several other industries.

NEXT GENERATION BIOLOGICAL ENGINEERING

Post genome era has raised our hopes and aspirations. The requirement and need of quantitative biology has necessitated effective interfacing between biology and computational sciences. There is growing relevance of system biology in studying the behavior of biological ensemble. At the other end biological systems are being designed to perform a specific function using synthetic biology concepts (1-5).

Molecular and cellular biology have made revolutionary advances in our understanding of biological function and information processing. Systems biology is an attempt to establish a bridge between molecular level information and system level understanding. It emphasizes integrative rather than reductionist approach.

Synthetic biology seeks to construct biological components, such as genetic circuits, metabolic pathways, parts of enzymes as per the required design specifications. Such a designed component can then be assembled into larger integrated systems to resolve specific problems. Synthetic biology is an integration of various sciences as is evident from its many acronyms: NBIC (nanotechnology, biotechnology, information technology and cognitive sciences), GRAIN (Genetics, Robotics, Artificial Intelligence and Nanotechnology), COMBINE (Cogno, Meets Bio, Info, Nanotech), GRIN (Genetics, Robotics, Informatics and Nanotechology), BANG (*Bits, Atoms, Neurons and Genes*).

Synthetic biology promises to build biological systems from the scratch. It includes redesigning existing genome for a new function. Craig Venter's group, for example, is developing a synthetic version of the *Mycoplasma genitalium* genome that has been stripped down to the absolute minimum number of genes required to support independent life. The goal of this "minimum genome project" is to build a simplified microbial platform to which new genes can be added, creating synthetic organisms with known characteristics and functionality. Protein design, systematically altering the genes that code for certain proteins to achieve desired modifications in protein stability and function, is another application of synthetic biology. Biological engineers working at Lawrence Berkeley National Laboratory have expanded the genetic code to specify unnatural amino acids, which can be substituted into proteins to modify their stability as well as their catalytic and binding properties. This technique has made it possible to design protein-based drugs that can resist rapid degradation in the body. Natural product synthesis is another promising application of synthetic biology. Molecular cloning techniques are being successfully used to produce human insulin. Scientists are now engineering microbes to perform complex multi-step syntheses of natural products by assembling "cassettes" of animal or plant genes that code for all of the enzymes in a synthetic pathway. Artemisinin, a natural product, is highly effective in treating malaria. Jay Keasling's group at the University of California, Berkeley is trying to reduce the cost of the drug by engineering a metabolic pathway for the synthesis of its immediate precursor, artemisinic acid in yeast. This technique may lead to produce the compound cheaply and in large quantities by fermentation.

With the growing interface between biology and engineering the biological function is being seen as 'biomolecular machines' and its regulation as 'biomolecular circuits'. How best to address biological engineering problems? Drew Endy (6) of MIT has some suggestions to transform biology in the way integrated-circuit design has transformed computing.

An engineer uses 'nuts and bolts' to assemble a thing. He doesn't do research when he applies these tools. He is confident that the tools he is using are the right ones. This confidence is lacking in biological engineering. One view is that "Genetic engineering techniques are abysmally primitive, akin to swapping random parts between random cars to produce a better car". 'Biological nuts and bolts', and 'cut and paste technology' need better standardization as this will make combination of parts made by different manufacturers easy. Drew Endy thinks that "biological engineering community will have to develop formal, widely used standards for most classes of basic biological functions". It means "promulgation of standards that support the definition, description and characterization of the basic biological parts as well as standard conditions that support the use of parts in combination and overall system operation". The MIT Registry of Standard Biological Parts is a first step toward that end. Decoupling, "to separate a complicated problem into many simpler problems that can be worked on independently, such that the resulting work can eventually be combined to produce a functioning whole", is another idea that could be helpful. The problems of biological system complexity can be managed by allowing individuals to work at any one level of complexity without regard for the details that define other levels, and also allowing the principled exchange of limited information across levels.

"Chemical engineers are good at integrating lots of pieces together to make a large scale chemical plant, and that is what we're doing in modern biological engineering – we're taking lots of little genetic pieces and putting them together to make a whole system," explains Keasling.

The biotech industry has recognized that if bigger is better then small can also be beautiful. Biologists were so far reading the genetic code. Now synthetic biologists are in the early stages of being able to write the code. They are trying to add new letters, rearranging them into new genetic networks, putting them into an artificial 'chassis' to go forth and multiply.

Biological engineering has potential to redesign enzymes, genetic circuits, and cells to their specifications, or even build biological systems from scratch. Though engineering the living world is not easy, in 2008 a bacterial genome has been synthesized from scratch. As gene synthesis becomes cheaper and faster, it is believed that in the near future it will be comparatively easier to synthesize a microbe than to find it in nature or retrieve it from a gene bank.

A biological engineer designs living devices, components and systems. Like any other engineer a biological engineer is expected to work within the constraints provided by technical, economic, business, political, social, and ethical issues. The commercialization of this 'extreme biotech' will not be easy. The developments in this area are quite bothersome to a section of society. It wants proper debate concerning socio-economic, security, health, environmental and human rights implications. Biological engineering, if it wishes to emerge as the new maharaja of engineering sciences, must expand its territory and knowledge base to tackle ethical issues. It must remember that "continued inbreeding with established species only adds to the richness of the gene pool, but does not lead to a new species'. New species require novel breeders.

REFERENCES

1. Roger Brent, A Partnership between Biology and Engineering, *Nature Biotechnology*, Vol 22, No. 10, 2004, p. 1211.
2. ETC Group Report, *Extreme Genetic Engineering – An Introduction to Synthetic Biology*, January 2007.
3. Jonathan Tucker, Raymond Zilinskas, The Promise and Perils of Synthetic Biology, *The New Atlantis, Spring,* 2006.

4. NEST High-level Expert Group Report, *Synthetic Biology – Applying Engineering to Biology*, 2005.
5. Merry Buckley, *Systems Microbiology: Beyond Microbial Genomics*, American Society of Microbiology, 2004.
6. Drew Endy, Foundations for Engineering Biology, *Nature*, Vol 438, 2005, p. 449.

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

LIGNOCELLULOSE BIOETHANOL TECHNOLOGY

PURNENDU GHOSH*

Greater concern about petroleum supplies and environmental consequences of fossil fuels has driven interest in bioenergy programmes all over the world (1-2). The solution is not fossil fuel or biofuel but fossil fuel and biofuel. Rate of oil discoveries peaked in 1960's. Today oil is discovered at $1/3^{rd}$ the peak worldwide rate. We use 4 barrels of oil for every new barrel we discover. Replacement of oil is thus a necessity. Oil companies have acknowledged the need to develop energy sources to replace oil. It is believed that a partial transition of oil to biofuels would stabilize the energy market significantly.

To be a viable alternative, a biofuel should provide a net energy gain, have environmental benefits, be economically competitive and be producible in large quantities without reducing food supplies (3). Bioethanol produced from lignocellulosic biomass is one such viable alternative. It can significantly reduce our dependence on foreign oil, create new jobs, improve rural economies, reduce greenhouse gas emissions, and improve national security.

The ability to produce ethanol from low cost biomass will be the key to making ethanol competitive with gasoline. The R&D in this area thus focuses on raw material availability, cost effective pretreatment technology, designed cellulases, and recombinant organisms capable of co-fermenting whole range of sugars to ethanol (4).

RAW MATERIAL

Development of sustainable technologies to supply biomass is one of the most critical factors for the running of a biorefinery. Three distinct goals are associated with development of biofuel feedstocks: maximizing the total amount of biomass produced per acre per year, maintaining sustainability while minimizing inputs, and maximizing the amount of fuel that can be produced per unit of biomass. Exact values for each of these parameters will vary from one type of energy crop and one growing zone to another. A yield of 20 dry tons per acre per year may be considered a reasonable target in areas of the country with adequate rainfall and good soils, whereas 10 dry tons per acre per year may be acceptable in drier or colder zones.

---

*Birla Institute of Scientific Research, Statue Circle, Jaipur-302001, Rajasthan

Lignocellulosic biomass contains cellulose, hemicellulose, and lignin; typically 35-50% cellulose, 20-40% hemicellulose, and 10-25% lignin. In the lignocellulose bioethanol technology, cellulose and hemicellulose are hydrolyzed to sugars, mainly glucose, xylose, arabinose, galactose, and mannose, using acids or enzymes. Lignin is the major interference in the hydrolysis of native lignocellulose. Metabolic engineering of the lignin biosynthetic pathway has been suggested as a method for modifying lignin content in feedstocks.

PRETREATMENT

In the enzymatic process, the lignocellulosic biomass is pretreated in order to increase the accessibility of cellulolytic enzymes to the substrate. Typically hydrolysis yields in the absence of pretreatment are less than 20% of theoretical yields, whereas yields after pretreatment often exceed 90%. The rationale for pretreatment has thus been to separate individual components of lignocellulose with minimum component losses, concomitant with an increase in surface area and a decrease in crystallinity. Various mechanical, physical, chemical, and biological approaches, either singly or in combination have been attempted to meet these objectives. An ideal technology is expected to produce a reactive fiber which will require little or no size reduction, and can be operated at a high solid/liquid ratio.

Ligninases and hemicellulases are important enzyme systems for biomass preconditioning. Consequently, research is needed to identify, characterize, improve, and economically produce these enzymes. These enzymes would be applied before or after traditional pretreatment to minimize and, eventually, replace thermochemical processes, thus lessening the effects of overall pretreatment severity at the macromolecular level and simplifying processing. The ultimate goal of this research is to produce a recombinant ligninase-hemicellulase microbial system with enhanced catalytic activity and stability.

ENZYME SYSTEM

The important parameters of enzymatic hydrolysis are sugar yield, duration of hydrolysis, and enzyme loading, and of course the characteristics of substrate cellulose and enzyme cellulases. The most desirable attributes of cellulases include the ability to produce a complete cellulase system with high specific activity, high catalytic activity against crystalline cellulose, thermal stability, decreased susceptibility to enzyme inhibition by the products of hydrolysis (glucose, cellobiose), selective adsorption on cellulose and ability to withstand shear forces. Strategies to improve cellulases include discovering new enzymes through bioprospecting, creating new/better mixtures of enzymes, and developing improved expression systems through protein design. Creating a more effective cellulose binding domain in the enzyme molecule is another approach to increase enzyme efficiency.

Industrial enzyme producers Genencor International and Novozymes Biotech achieved up to 30 fold reduction in enzyme cost. Further cost reduction, however, is required to support an economical and robust cellulose biorefinery industry. For example, costs of amylase enzymes for converting corn grain starch to ethanol are about 1 to 2 cents per gallon of ethanol produced, but the most optimistic cost estimates for cellulase preparations now are about tenfold higher than that.

FERMENTATION

Conversion of cellulose and hemicellulose to ethanol comprises hydrolysis followed by fermentation of hexoses and pentoses by ethanologens.

Simultaneous saccharification and co-fermentation (SSCF), in which hydrolysis is integrated with fermentation of both hexose and pentose sugars but with cellulase produced in a separate step. For example, development of thermophilic ethanol-producing organisms for use in SSCF could allow the consolidated process to run at higher temperatures, thus realizing significant savings by reducing cellulase requirements. Combining cellulase production, cellulose hydrolysis, and cofermentation of C-5 and C-6 sugars in a single step is called "consolidated bioprocessing" (CBP). Widely considered the ultimate low-cost configuration for cellulose hydrolysis and fermentation, CBP has been shown to offer large cost benefits relative to other process configurations in both near-term and futuristic contexts.

An ideal ethanologen has the following characteristics: high ethanol tolerance, capacity to withstand high osmotic pressure, high temperature, and low pH, high cell viability, appropriate flocculation and sedimentation characteristics, capability to ferment broad range of sugars mainly to ethanol and possibly negligible levels of by-products (such as acids and glycerol) , resistance to inhibitory compounds present in the pretreatment/hydrolysis stream, etc.

The yeast *Saccharomyces cerevisiae*, due to its large size, thick cell wall, resistance to bacterial and viral infection and its ability to produce high ethanol yield has found greater acceptance as ethanol fermenting organism in the industry. The bacteria *Zymomonas mobilis* produces ethanol at a much faster rate than the yeast and has a higher osmotolerance and alcohol tolerance. Both *S. cerevisiae* and *Z. mobilis* can ferment only glucose, fructose, and sucrose. On the other hand, *Echerichia coli*, a common bacteria, found in the gut, utilizes besides these other sugars (xylose, arabinose, mannose, galactose). But *E. coli* lacks two enzymes required for ethanol fermentation, namely pyruvate decarboxylase and alcohol dehydrogenase. Since all the sugars should be utilized for economical ethanol production, an organism having this capability and also the desired enzymes would be an ideal ethanologen. Recombinant ethanol producing organism having these characteristics is thus the solution. Such recombinant ethanologenic organisms have been created. They, however, produce lower ethanol titers (5 to 6 wt % ethanol). Improvements in ethanol yields and tolerance are needed to increase rates of production (>1.0 g/L/h) from all sugar constituents of lignocellulosic biomass. As titers are increased, rates slow down and eventually cease at ~6 wt % ethanol, the upper limit for wild-type *E. coli*. By comparison, wild-type yeast and *Z. mobilis* can reach titers of >15% ethanol from cornstarch glucose but have failed to achieve these levels on pentose sugars.

A strategy for increasing ethanol tolerance or other traits could use *evolutionary engineering* concepts and methods. This strategy would allow the microbial process to evolve under the proper selective pressure (in this case, higher ethanol concentrations) to increasingly higher ethanol tolerances.

Most methods of biomass pretreatment to produce hydrolysates also produce side products (e.g., acetate, furfural, and lignin) that are inhibitory to microorganisms. These inhibitory side products often significantly reduce the growth of biocatalysts, rates of sugar metabolism, and final ethanol titers. In all cases, the impact of hydrolysates on xylose metabolism is much greater than that of glucose.

The following figures are suggested for a biomass-to ethanol process that will be cost-competitive relative to current cornstarch ethanol operations: Use of both hexoses and pentoses to produce ethanol at a yield greater than 95% of theoretical yield, final ethanol titers in the range of 10 to 15 wt %, overall volumetric productivity of 2 to 5 grams of ethanol per liter per hour.

ETHANOL RECOVERY

Conventionally distillation is used for separating water from ethanol present in the fermentation broth. In order to obtain anhydrous ethanol additional energy is required to break the azeotrope. Various distillation configurations such as vapor recompression, low pressure distillation etc. has been used to make substantial savings in steam consumption for ethanol separation. Even though, in terms of degree of separation, distillation is very effective, new techniques have been proposed to further bring down the energy costs (4). Some non-distillation approaches such as selective sorption of ethanol, preferential adsorption of water, and membrane separation have shown promise.

Distillation is energy intensive, but is the most reliable method of ethanol recovery. Other approaches need further scale-up studies and collection of engineering data to realistically assess their potential. In the immediate future bioethanol technologies will use distillation for ethanol recovery along with much improved energy recycling systems.

CONCLUDING REMARKS

Lignocellulosic biomass is less expensive than sugar or starch-based feedstocks, but its conversion to ethanol at present is more costly. The commercialization of this technology has to overcome various bottlenecks. These include feedstock availability, scale of operation, cheaper and effective pretreatment technologies, efficient hydrolytic agents, availability of recombinant organisms capable of co-fermenting the whole range of sugars at a temperature compatible to optimum hydrolysis, and better co-product value.

Logistics of raw material availability (collection, storage and handling) to meet large future demands is a major issue. For example, rice straw-based ethanol production would require the location of the plant within a reasonable distance from the rice farms. Moreover, seasonal availability of the feedstock would need either large storage facilities or would need plants to operate on multiple feedstock for their continued operation throughout the year.

In India ethanol plants are comparatively small in capacity. This brings to the fore another related issue: scale of operation vis-à-vis feedstock availability. Relatively large investments will be required to install lignocellulose bioethanol plants. Keeping in view the logistics of feedstock procurement, a decision is needed if it is advisable to build very large plants as increased feedstock cost (due to collection and transport of large amounts of feedstock) may offset savings due to economies of scale.

Pretreatment of lignocellulose continues to be a major barrier for the development of a viable technology. We have to ascertain what is more important for enzymatic hydrolysis – the extent of delignification which requires harsher conditions for complete lignin separation or will the loosening of cellulose-hemicellulose-lignin bonds under milder conditions will be sufficient. The benefits of lignin solubilization need to be weighed against the potential for fermentation inhibition by soluble lignin derivatives. Development of energy plants with traits such as increased cellulose and hemicellulose and less lignin not only has the potential to improve ethanol yields but also the application of much simpler pretreatment technologies.

Another critical element for the success of bioethanol technology is the availability of cellulases at a cost that will dictate the ethanol price to be paid by the consumer. Currently, the major market for cellulase enzymes is the textile industry, and the enzymes produced are tailored to meet the requirement of this industry. It is important to recognize that biomass application needs are significantly different from textile applications. Cellulase enzymes are too expensive for bioethanol. There is, however, a good possibility of producing effective cellulases at a much reduced cost. For the hydrolysis of pretreated biomass, extremely complex cellulases may not be required; simpler cellulase systems may serve the purpose.

Companies are engaged in creating synthetic microbes to accelerate the conversion of agricultural waste to ethanol. Microbial diversity collected from the seawater is being used as the raw material to create a new synthetic microbe. The hunt is on for a better microbe that will cheaply and efficiently break down cellulose to sugars and then ferment those sugars into ethanol – without costing energy. That's where synthetic biology comes in. Such designer organisms are yet not available (5).

The projected cost of ethanol from lignocellulose has declined significantly in the last ten years. Further cost reduction is needed. This is possible by employing a cellulose and hemicellulose rich but lignin lean feedstock and using highly efficient cellulase producers, a more efficient fermentation process using better ethanologens. Some indicative figures are as follows: biomass with >50% cellulose, 30% hemicellulose, <15% lignin, enzyme productivity 2000 FPU/l-h, fermentation temperature 55degree C.

Commercialization of ethanol needs the attention of researchers, entrepreneurs, and more importantly, the policy makers. Who will be the major promoter of large scale bioethanol technology – agriculturist, sugar industry, or petroleum industry? Other than science, what kind of support is needed to make this a viable technology? Under what circumstances or conditions refiners might consider participation in the ethanol industry will continue to remain a big question. Only if return on investment for ethanol plants exceeds that of petroleum processing facilities, some major companies might invest in this venture. It is heartening that India is becoming a 'player' instead of a 'spectator' in the bioenergy race.

REFERENCES

1. *Breaking the Chemical and Engineering Barriers to Lignocellulosic Biofuels, Next Generation Hydrocarbon Refinery*, National Science Foundation, 2008.
2. *Breaking the Biological Barriers to Cellulosic Ethanol*, US Department of Energy, 2006.

3. J. Hill, E. Nelson, D. Tilman, S. Polasky & D. Tiffany, Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels, *PNAS,* Vol 103, No. 30, pp 11206-11210.
4. P. Ghosh & T.K. Ghose, Bioethanol in India: Recent Past and Emerging Future, *Advances in Biochemical Engineering/Biotechnology*, Vol 85, 2003, pp 1-27.
5. *Systems Biology for Energy and Environment*, US Department of Energy, 2005.

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

BIOTECHNOLOGY INDUSTRY TRENDS

PURNENDU GHOSH*

Why industry should turn to biotechnology? Industry expects that biotechnology will provide inexpensive and reliable healthcare means. It will give technology to grow more food from cultivable and not-so-cultivable land. It will provide new methods to protect the environment. Industry believes that biotechnology has better means to fulfill the above promises.

INDUSTRIAL BIOTECHNOLOGY

The product portfolio of industrial biotechnology includes enzymes, biofuels, biomaterials, organic acids, amino acids, vitamins, probiotics, specialty chemicals, etc (1-9).

Take the industrial chemical market size, which is presently worth $1000 Billion. The current share of biotech based products is only 6%. In another ten years time the share is projected to jump to 20%. This jump is possible only if biotech based processes are better than chemical synthesis based processes, both economically and environmentally. In order to do this we may have to change the entire process or only a few steps. This will need further R&D to replace a conventional catalyst by a biocatalyst. This will need development of biocatalysts with capabilities to catalyze a broad range of reactions having greater versatility. This will need increase in biocatalyst's thermo stability, activity, and solvent compatibility. This will need, in place of a natural organism, a genetically manipulated organism. This will need reduction in material, energy, and water consumption by over 30%. This will need switch from batch to continuous processes. This will need cleaner processes.

This will need a proper biocatalyst inventory. Such an inventory will keep track of number and type of industrially significant reactions, cofactor use and regeneration, tools that assist in the operation and use of biocatalysts in process systems. Because of these there is a change in human resources trends in conventional chemical industry; chemical industries are now hiring increasing number of life scientists; chemists are embracing biocatalysts as tools for new synthesis.

*Birla Institute of Scientific Research, Statue Circle, Jaipur-302001

MEDICAL BIOTECHNOLOGY

The purpose of medical biotechnology is to enhance the quality of our life. Its product portfolio includes pharmaceuticals, disease diagnostics, therapies, vaccines, neutraceuticals, etc (10-12). Developments in medical biotechnology call for a change in current medical practices. Currently we go to a doctor only when we have some problem. We ask him "Doctor, what is wrong with me?" After few years we will go to a doctor even when we are not sick and we will ask him "Doctor, what is going to be wrong with me in the next ten years?" Current medical practices follow routine examination, diagnosis, prognosis, treatment, and follow-up. The future medical practices will follow personalized therapy based on patient's unique genetic profile. There will be emphasis on prevention, treatment, and more testing. Disease will be understood as a process that evolves over time through interaction of genetic, environmental, and life style factors. This view acknowledges that most diseases can not be tied to a single cause. Tomorrow's medicine will recognize and give due weight to person's 'individuality'. Nearly 30% of patients do not benefit from medicines. This has to change. If patients are different then medicines must be differentiated. Sickness will continue to be the priority of the doctors. But then there will be a large group for whom wellness will also be equally important.

There is big market for obesity, diabetes, Alzheimer's, anti infectives (antibiotic resistance), wellness (preventive/predictive cure), and age related diseases. According to one report (11) greatest market and scientific opportunity currently are in the areas of joint health, weight management, inflammation, glucose management, and cardiovascular diseases.

Health and wellness market will be driven by food companies, agricultural companies, pharmaceutical companies, consumer product companies, and also genomic companies. Consumers will drive this market. Reason: awareness; more and more people want to be updated about health benefits of food products and dietary supplements.

Health informatics will play a big role in deciding the future scenario of medical biotech industry. Health informatics is the rational study of the way we think about patients, and the way that treatments are defined, selected, and evolved. It is the study of how clinical knowledge is created, shaped, shared, and applied. Ultimately it is the study of how we organize ourselves to create and run healthcare organizations. If physiology is the logic of life, pathology the logic of diseases, then health informatics is the logic of health.

Bioinformatics related to healthcare will make significant impact. According to one estimate, three-fourth of the cost of developing a successful drug goes to paying for the failed hypotheses and blind alleys pursued along the way. Bioinformatics promises to bring the costs down by suggesting ways so that shorter lines are followed to reach the destination.

The history of medicine has been interestingly summed up thus: 2000 BC – Eat this root; 1000 BC – Eating root is not good, say prayers; 1800 AD – Prayer is superstition, drink this potion; 1940 AD – That potion is not good, swallow the pills; 1990 AD – Pill is ineffective, take antibiotics; Present – Antibiotics doesn't work, eat this root. We will be willing to eat the root but we will like to know what we are eating. Systematic efforts are being made towards this goal.

AGRICULTURAL BIOTECHNOLOGY

The global population in the year 1900 was 1.6 Billion; today it is 6 billion; in 2030 it will reach 10 Billion. According to the UN estimates world food production will have to double on the existing land over the next few years if it wants to keep pace with anticipated population growth. According to one survey, the growth in agriculture sector in our country during 1990-2007 is mere 1.2% whereas during the same period our population has grown at an annual rate of 1.9%. What is the solution to fill this gap? The solution is to produce more food in cultivable and not-so cultivable land. To improve agricultural productivity it is essential to decrease inputs (water, fertilizer), provide environment friendly pest control methods, develop new diagnostic tools for early detection of plant diseases, etc. Transgenic crops reduce pesticide needs significantly and thus reduce the consumption of harmful agricultural chemicals. The transgenic BT-cotton is a success story (10). In 5 years we have doubled cotton production (3.1 million bales) due to improved biotech based agricultural practices. Presently transgenic crops which are at various developmental stages include cabbage, cauliflower, corn, cotton, eggplant, groundnut, mustard, potato, rice, tomato.

Animal biotechnology holds vast promise for improving our quality of life. Use of animal genomics, an extension of traditional animal breeding, is accepted as safe and is largely unregulated. It provides new tools for improving animal health and increasing livestock and poultry productivity through 1. Enhanced ability to detect, treat and prevent diseases and other problems. 2. Better feed derived from transgenic crops designed to meet the dietary needs of different farm animals. 3. Improved livestock productivity through improved animal breeding and disease resistance.

BIOTECHNOLOGY INDUSTRY

Global biotech industry is worth $60 Billion. The industry spends nearly one-third of its revenues on R&D. The Indian biotech industry scenario is given in Tables 1 and 2.

The industry collectively received revenue of $2.57 Billion during 2007-08; nearly two-third of the revenue came from export. There is slump in the biotech industry during the year 2007-08 (3). It registered 20% growth compared to more than 30% during the previous five years. One reason assigned for this slump is appreciation of Rupee against the US Dollar. Another reason is the slump

Table 1. *Biotech Industry Revenues (Rs. Crore)*

Segment	*2007-08*	*2006-07*	*2005-06*	*2004-05*	*2003-04*
BioPharma	6900	5973	4708	3570	2752
BioServices	1572	1102	720	425	275
BioAgri	1201	926	598	330	130
BioIndustrial	410	395	375	320	238
BioInformatics	190	145	120	100	80
Total	10273	8541	6521	4745	3475

Table 2. *Biotech Industry – Export vis-à-vis Domestic*

Sector	*Export*	*Domestic*
BioPharma	58%	42%
Bioservices	96%	4%
BioAgri	4%	96%
BioIndustrial	7%	93%
BioInformatics	79%	21%

in growth (16%) in the biopharma sector. The Indian biotech sector, according to one estimate, shall generate revenues to the tune of Rs 65000 Crore by the year 2015.

The key success factors and challenges for the biotech industry are reflected in Tables 3 and 4 respectively.

The success of biotech industry in India will depend on the success of formation of bioclusters (Figure 1).

Some well known international bioclusters are USA: San Francisco, San Diego, North Carolina, Boston, Los Angeles, Seattle, New York/New Jersey, Washington DC, Europe: Brussels, Montreal, Toronto, Paris, Strasbourg, Munich, Dublin, Stockholm, Cambridge, Oxford, Edinburgh, Asia-Pacific: Sydney, Brisbane, Adelaide, Perth, Beijing, Shanghai Hong Kong, Tel Aviv, Jerusalem, Kanto, Kansai, Tokyo, Auckland, Dunedin, Singapore, Taipei. The geographical distribution of Indian biotech industries is: South (46%), West (41%), and North (13%). The major bioclusters in India are Bangalore, Hyderabad, Chennai, Mumbai, Pune, and Ahmedabad.

CONCLUDING REMARKS

The biotechnology revolution holds great potential. But it has yet to fulfill the promises it has made to the common man. Some think biotechnology will create a wider wedge between the two worlds.

Table 3. *Key success Factors*

Research Institutions	*Available*
Networking	Needs improvement
Risk Capital	Not satisfactory
Workforce and Talent Pool	Available
Specialized Facilities and Equipment	Needs improvement
Supportive Business, Tax, and Regulatory Policies	Not yet in place

Table 4. *Key Biotech Industry Challenges*

- Growing competition from other regions
- Diffusion of research into commercial innovation
- Establish local biotech industry
- Leveraging the resources and networks of other bioclusters

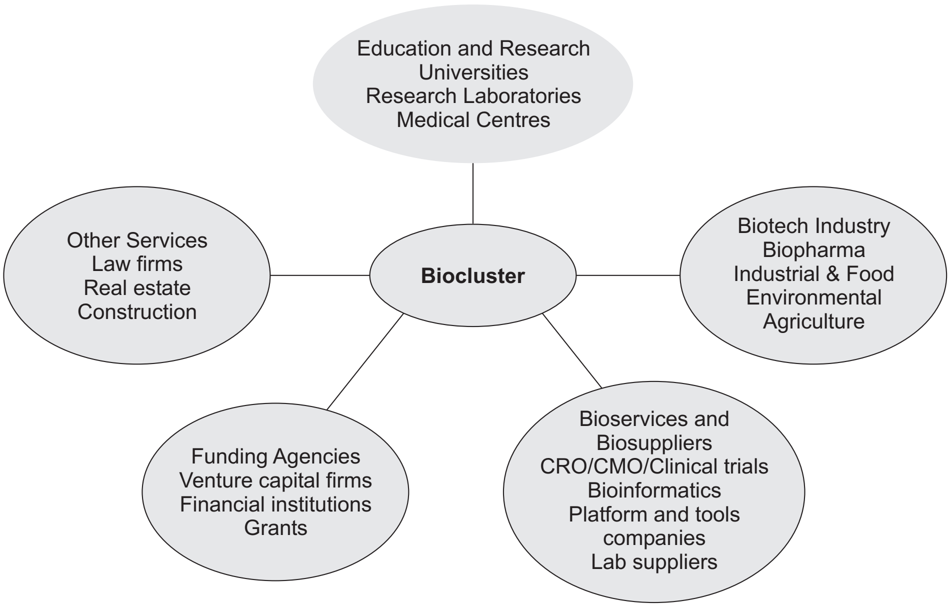


FIG. 1. COMPONENTS OF A BIOCLUSTER

A technology which touches so many sides of human enterprise will surely confront ethical issues. A Biotechnology Industry Organization (BIO) document (13) provides guidelines of the ethical principles. It includes the following: We respect the power of biotechnology and apply it for the benefit of humankind. We listen carefully to those who are concerned about the implications of biotechnology and respond to their concerns. We help educate the public about biotechnology, its benefits and implications. We place our highest priority on health, safety and environmental protection in the use of our products. We respect the animals involved in our research and treat them humanely. We adhere to strict informed consent procedures. We follow the policy of opposing the use of biotechnology to develop weapons of any sort that contain pathogens or toxins aimed at killing or injuring humans, crops or livestock. We continue to support the conservation of biological diversity. In this scenario dissemination on the benefits of biotech through public education and mass media is essential.

Biotechnology shows us another side of human empowerment. Aldous Huxley portrayed it nicely in Brave *New World.* His technologically enhanced inhabitants live cheerfully, without disappointment or regret, "enjoying" flat, empty lives devoid of love and longing, filled with only trivial pursuits and shallow attachments. I do not wish for such 'technologically enhanced people'. I only wish biotechnology promises are not 'pie in the sky' promises.

REFERENCES

1. T.K. Ghose & P. Ghosh, Biotechnology in India, *Advances in Biochemical Engineering/ Biotechnology:* Vol 84&85, 2003.

2. *The Guide to Biotechnology*, Biotechnology Industry Organization, 2007.
3. *Biospectrum*, Vol 6, Issue 7, July 2008.
4. *Technology, Talent and Capital: State Bioscience Initiative*, Report prepared for Biotechnology Industry Organization by Battelle Technology Partnership Practice, 2008.
5. *Microbiology for the Twenty-First Century: Where are we and where are we going?,* American Society of Microbiology, 2004.
6. *Industrial Biotechnology and Sustainable Chemistry*, Royal Belgian Academy Council of Applied Sciences, 2004.
7. S. Herrera, Industrial Biotechnology – A chance at redemption, *Nature Biotechnology*, Vol 22, 2004, p 671.
8. *White Biotechnology: Gateway to a more sustainable future*, EuropaBio, 2004.
9. *Top Value Added chemicals for Biomass Volume 1 – Results of screening for potential candidates from sugars and synthetic gas*, UD Department of Energy, 2004.
10. *Biotech News, Special Issue: BIO 2008*, June 2008
11. *State of Biotechnology Industry Trends, Concerns and Outlooks*, Burrill & Company, 2005.
12. *Bioscience 2015, A Report to Government by Bioscience Innovation and Growth Team.*
13. *Bioethics – Facing the Future Responsibility*, Biotechnology Industry Organization Document.

ELECTRICAL ENGINEERING

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

STUDIES ON HIGHER HARMONICS IN HIGH VOLTAGE NETWORKS

DR. R.P. BHATELE*

ABSTRACT

The paper presents main results of studies in the area of higher harmonics in electric networks of 132kV and higher voltage network, which have been achieved recently in an electric utility. The operating conditions of electric networks at higher harmonics have been analyzed by the distortion power method, which applies the notions of distortion power generation and absorption and the software package is ulitized "High Harmonics", which enables the calculations, analysis and studies of operating conditions at higher harmonics in extended electric networks. The information, was obtained from measurements of operating parameters at higher harmonics in the 132kV-220kV electric networks in the connection points of large consumers, i.e. higher harmonics sources (steel plants, railways). It has been revealed that the lengthy lines favour the increase in higher harmonics voltage levels and greatly influence the absorption of higher harmonics currents. The high levels of higher harmonics are typical of the nodes with the minimum absorption of distortion power. The nodal voltages are formed due to interaction of a large number of non-linear loads. The higher harmonics voltage levels in complex networks can be normalized in a centralized way by installation of several special filters. The higher harmonics modes should be calculated by the full representation of the electric networks of 132kV and higher voltage network.

INTRODUCTION

The 132-220kV networks are extended and have a great number of connected large consumers such as steel plants, railroads. In accordance with the current standards the normal allowable value of the total harmonics distortion for 220kV networks is equal to 2 per cent of the nominal voltage. The maximum voltage is 4 per cent of the allowable level of harmonics voltage components of higher harmonics and accounts for 2 per cent. Measurements of the parameters of higher harmonics modes in 220kV networks of an electric utility and the higher harmonics voltage levels (HHVLs) for many nodes exceeded the allowable levels. The 220kV networks were characterized by prevalence of the voltage harmonics 11 and 13 due to operation of steel plants. The level of these harmonics was

*Executive Director (T&P), MPPTCL, Jabalpur

considerably lowered as a result of change in rectification technology by the appropriate adjustment of the parameter. Adjustment leads to mutual compensation of the current harmonics 11 & 13.

The parameters of higher harmonics modes and the parameters of the basic frequency modes were measured in the 220kV network nodes to evaluate power quality with respect to higher harmonics and to determine the factual values of generation of higher harmonic currents. Measurements taken in electric networks with long transmission lines have revealed that there is no direct relation between HHVLs and the value of higher harmonic current of non-linear loads.

A method of the distortion power balance that reveals interaction of an electric network and a consumer has been devised to analyze the process of HHVL formation. The devised method was applied to study the higher harmonic modes in complex networks with extended transmission lines. It is shown that the levels of higher harmonic voltages in these networks can be normalized by installation of several filters.

METHOD OF DISTORTION POWERS

The method of distortion power balances applies the calculation of higher harmonic conditions in the electric network in the current values that are widely used [1,2]. The condition is calculated for each harmonic separately. Non-linearity of nodes are modeled by the sources of current, the values and phases of which can be adjusted during calculation as function of parameters of the node, to which the non-linear load is connected.

In this method the electric network and consumer with respect to node I on the n-th harmonic are represented by the equivalent two-pole elements consisting of parallel connected sources of current, conductance and susceptance (Fig. 1). The letters used in the indices are: g-generator, *a*- absorption, *s*-networks, *h* – load, *g*-active, *r* – reactive, *i*- number of harmonic.

The current of the equivalent sources of the network I_{sin} is equal to the sum of currents of sources from the non-linear loads of the network

$$I_{\text{sin}} = \sum_{j=1}^{m} I_{\text{gjin}} \tag{1}$$

Where $I_{\text{gjin,}}$ – equivalent source of current in node I reflecting the non-linearity of load in node j

$$I_{\text{gjin,}} = A_{\text{jin}}\, I_{\text{jn,}} \tag{2}$$

I_{jn} – source of current in node j reflecting the non-linearity of load that is connected to node j;

A_{jin} – coefficient of transmission of the current of the n-th harmonic from the source situated in node j to node I is determined by the method of active one-port network from the short-circuit conditions in node I,

m – number of network nodes.

In Fig.1 the balance of current is observed in node I on the n-th harmonic, i.e. generation of currents by their sources to node I is equal to the current absorbed by nodal admittances

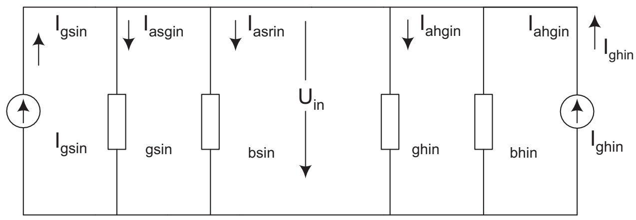


FIG. 1 EQUIVALENT CIRCUIT OF THE NETWORK AND CONSUMER BY THE METHOD OF DISTORTION POWER BALANCE

$$I_{\text{sin}} = \sum_{j=1}^{m} I_{\text{gjin}} + I_{\text{ghin}} = (I_{\text{asgin}} + I_{\text{ahgin}}) + j(I_{\text{asrin}} + I_{\text{ahrin}}) \tag{3}$$

where m – number of nodes.

The balance of distortion powers which is obtained by multiplying the currents of the n-th harmonic by the nodal nominal voltage U, corresponds to the balance of currents of the n-th harmonic of the node in expression (3).

The nodal voltage expressed by the distortion powers is determined as per Eq.(4) to Eq. (10) in ref[3].

The table of distortion powers and higher harmonic voltage levels is constructed to analyze the nodal operating conditions. It contains the following: generation of distortion powers of the network and every non-linear load; three values of the total generation of distortion power of the node: deterministic, expected, maximum; active total absorption of the node and that of each non-linear load connected to the node, reactive total absorption of the node and that of each non-linear load; overall absorption; absorption reserve at the expected generation; contribution of each higher harmonic source to higher harmonic voltage, three estimates of the higher harmonic voltages in the node: deterministic, expected, maximum resonant voltages and resonant capacities of the capacitors.

The use of the distortion power balances for the analysis of operation conditions results in the following advantages:

- generation and absorption power processes are expressed in the same units that allows their comparison;
- transition through the transformer does not change power that allows comparison of operation conditions of the nodes separated by the transformer and estimation of the impact of transformer parameters on the conditions of high harmonics;
- use of the distortion powers makes in possible to formulate the condition for provision of the admissible HHVLs at consumer connection and to introduce the concept of reserve with respect to distortion power absorption;
- distortion powers reflect the economics in respect of inclined of the admissible HHVLs. The costs on the increase in the reserve of distortion power absorption of the node by the resonant filter, that is the main means to normalize HHVLs; are proportional to the absorbed distortion power;

- application of the tables for distortion power balances reveals the significance of the processes of generation of distortion power by the non–linear loads and absorption of distortion power by the network and load for formation of HHVLs.

The devised method was used as a methodological base for creation of the software package “High Harmonics”. The package calculates HHVLs, tables of the distortion power balances, relationship between the nodal operating parameters and the parameters of the network elements, etc. for a complex electric network.

NETWORK PROPERTIES CAUSED BY DISTORTION POWERS

The properties were studied on the scheme with the minimum number of nodes in which properties of bulk power systems are displayed (Fig. 2). The scheme consists of two sections of the single-circuit 220kV transmission lines of the same length with the wire cross-section of 300mm^2 at the current density of 1.1 A/mm. The power transmitted over transmission lines accounts for 135 MVA. The network is connected to the large 400kV node through the 500 MVA autotransformer. In the nodal, HHVLs amount to 1% of the rated voltage of the basic frequency for all the studied harmonics. Consumers are connected to nodes 21, 22 and 23 though the 63 MVA transformers with the secondary voltage of 33kV. The plots presented illustrate the results of studies on the effect of line length change on generation and absorption of distortion power of nodes. The studies were performed on harmonic 11, as ¼ of her wave length for this lies within the economic lengths for transmission lines of the class (100 – 300 km).

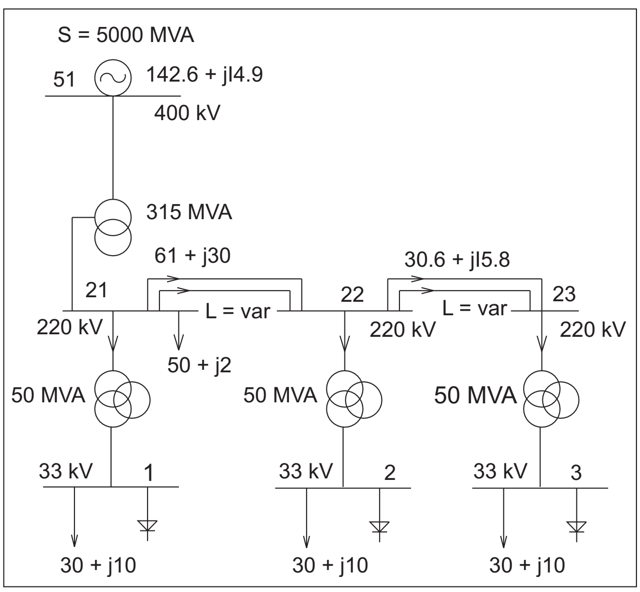


FIG. 2 THE SCHEME FOR STUDYING ELECTRIC NETWORK PROPERTIES

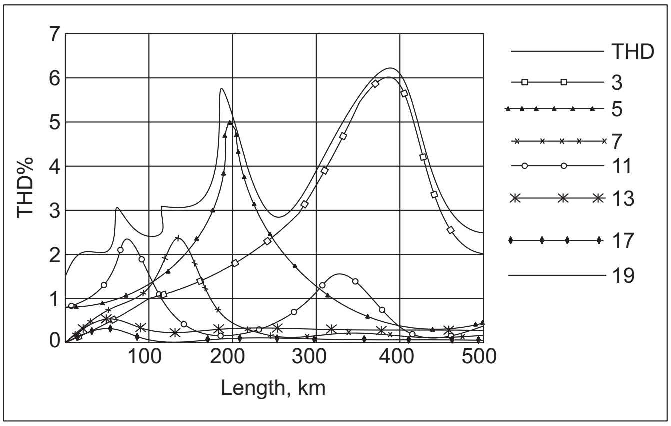


FIG. 3 TOTAL HARMONIC DISTORTION AND HHVLS IN THE NODE 23

Figure 3 illustrates the effect of transmission line length on the levels of THD and the higher harmonic voltage in node 23. The transmission line considerably increases THD due to growth of the voltage level of one or several harmonics. For the 220kV transmission lines 150-230 km. long, the voltage of harmonics 5 has the larges value equal to 6%. Increase in the load power of the node feeding the line leads to decrease in the levels of THD and the higher harmonic voltages.

Traction Load of Railroad

The traction load of railroads is a large non-linear load. The power consumed by a traction substation can reach 15 MW and more at an average load value equal to 5-8 MW. The traction load is single-phase and its change for a short time interval is large. As a rule the traction substations are connected to a 220kV transmission line in 40-60km. in order that the traction load be supplied from different phases and the total load be symmetric. As a result several tens of traction substations appear to be connected to the 220kV transmission line.

The parameters of higher harmonic mode of the traction load of connected to 220kV buses through a 50 MVA transformer are given below. Seven transmission lines transmitting 500-700 MW of power come to this substation.

The active power changes considerably during a short time interval. The phase currents vary similarly. Analysis, of the measurement data shows that the harmonic mix of current varies essentially with time. However, there is no correlation between the current value and the higher harmonic voltage level.

The harmonic mix of currents and voltages (Table 1, Table 2) varies greatly during measurements. The interval of changing the current phase for harmonic 5 makes up from 100 to 200 degrees. The other harmonics are also characterizes by a wide range of variation of the current value and its phase.

Table 1. *Statistical characteristics of HHVLs of traction load (%)*

Value	*Harmonics*					
	3	*5*	*7*	*11*	*13*	*17*
Mean value	0.88	1.50	0.28	0.44	0.17	0.08
Standard deviation	0.41	0.41	0.12	0.14	0.08	0.04
Interval	1.90	1.95	0.61	0.70	0.39	0.21
Minimum	0.07	0.60	0.01	0.11	0.01	0.01
Maximum	1.97	2.56	0.63	0.82	0.41	0.22

Table 2. *Statistical characteristics of higher harmonic currents of traction load (%)*

Value	*Harmonics*						
	3	*5*	*7*	*11*	*13*	*17*	*19*
Mean value	8.25	7.21	4.23	1.83	1.23	0.86	0.92
Standard deviation	4.32	2.20	1.41	0.72	0.50	0.42	0.44
Interval	17.24	14.34	6.79	4.09	2.50	2.22	2.69
Minimum	0.53	0.81	0.35	0.25	0.06	0.12	0.05
Maximum	17.78	15.16	7.15	4.35	2.57	2.35	2.75

NORMALIZATION OF HIGHER HARMONIC MODES IN THE NETWORK

Depending on the specific network and load power the higher harmonic modes can be normalized by the following method:

- decrease of the higher harmonic current generation of rectifier loads by the increase of pulsation and commutation angles;
- installation of filters for higher harmonic currents in the points of connection of large rectifiers;
- installation of filters for higher harmonic currents in one or several nodes of the electric network with extended transmission lines.

As was shown above, it is insufficient to normalize the mode in the connection nodes of large non-linear loads for normalization of the operating conditions in all the network nodes, since the extended transmission lines cause the growth of higher harmonic levels. In order to normalize the higher harmonic mode in the extended lines the following features of the mode are used: high levels of harmonics correspond to nodes with the lower of absorption and generation and change in absorption in one of the nodes of the extended line result in the change of higher harmonic voltage in other nodes. These features allow normalization of operating conditions in a great number of nodes of the extended line by installing relatively small filter in one of the nodes with a higher voltage level. The studies have shown that the capacity and points for installation of devices for the centralized normalization are determined by the network scheme and its conditions, due to the strong interrelation of the higher harmonic modes of nodes, which are located at considerable

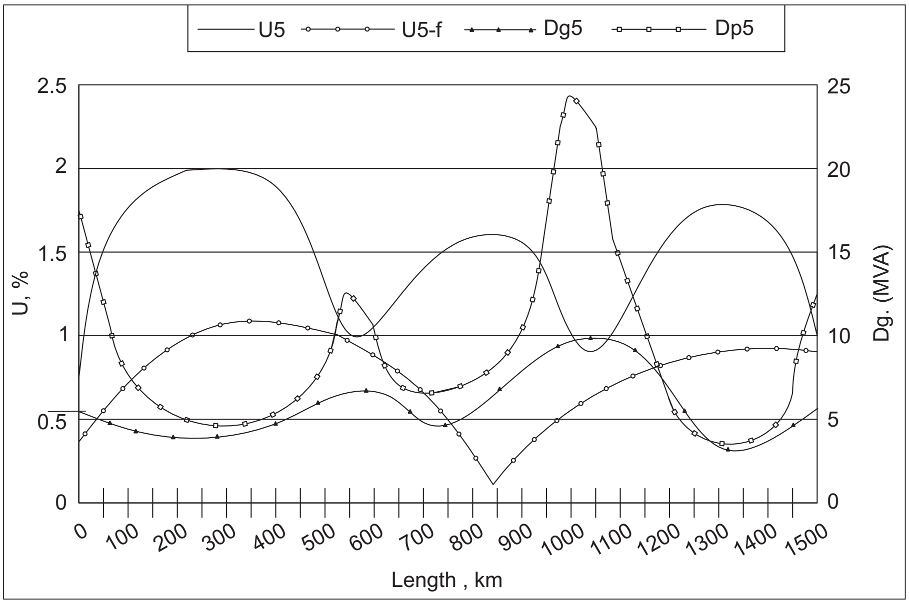


FIG. 4 GENERATION AND ABSORPTION OF DISTORTION POWER, VOLTAGE 5-TH HARMONIC WITH FILTER AND WITHOUT

distances. As a rule, the capacities of devices do not exceed 2-4 MVA. Application of the centralized normalization of the higher harmonic modes allows the solution of their normalization problem in the networks of 132kV and higher.

CONCLUSIONS

1. The method of distortion power balance is suggested in order to analyze nodal operating conditions of an electric network, which are due to high harmonics, and to reveal interaction between the consumer and electric network.
2. Application of the distortion power balances for the analysis of operating conditions makes it possible:
    - To reveal the effect of the network and non-linear loads on HHVLs;
    - to formulate the condition providing admissible HHVLs at consumer connection;
    - to evaluate conditions and danger of resonant phenomena.
3. The distortion power method makes it possible to represent formation of the higher voltage harmonic levels in the network node as a result of interaction of two processes: higher harmonic generation and absorption.
4. The extended transmission lines of 132kV and higher promotes the increase of higher harmonic voltage levels, harmonics 5 and 7 being dominant. The high levels of higher harmonic voltages correspond to the nodes with small values of distortion power generation and absorption.

5. The stable value of higher harmonic current amplitude and the wide range of higher harmonic current variation is typical of the rectified load of steel plants. Besides the current of harmonics 11 and 13 the high current values are generated in harmonics 3,5,7 due to the scheme asymmetry.
6. The amplitude and phase of higher harmonic current of the traction load changes within broad limits. The total harmonic distortion also varies quickly and in a wide range. It can be evaluated by using the probabilistic characteristics.
7. The filters being installed in 132-220kV networks, the higher harmonic voltage levels in the extended networks will be decreased in a radius to 500km and more.

REFERENCES

1. S.S. Smirnov, L.I. Kovernikova, Consumer contribution to the voltage levels of high harmonics in electric network nodes. – Elektrichestvo, 1966, NI, pp58-64. (in Russian)
2. S.S. Smirnov, L.I. Kovernikova, Analysis of operating conditions of electric networks on high harmonics by the distortion powers, Proceedings of the Twelfth Power Systems Computation Conference, Volume II, Dresden, Germany, August 20-23, 1996, pp 1149-1154.
3. S.S. Smirnov & L.I. Kovernikova, “Studies on Higher Harmonics in High Voltage Networks”, presented in CIGRE Regional Meeting on “Power Quality Assessment of Impact”.

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

DEVELOPMENT OF FACTS DEVICES IN BHEL

M. ARUNACHALAM*

ABSTRACT

Flexible AC Transmission System (FACTS) devices are being used more and more in modern Power System to enhance the efficiency and capability of the system. FACTS is the application of Power Electronics to Power System. Bharat Heavy Electricals Limited (BHEL) has been in the business of applying Power Electronics to Power System since 1974.The first Static VAr System (SVC), which is recognized now as one of the FACTS devices, was developed by BHEL in 1970s. In this paper, a number of FACTS devices developed by BHEL since then are discussed.

1. INTRODUCTION

Power Electronics revolutions are taking place in the following areas [1, 2]:

- End use electrical appliance services to increase efficiency
- Power quality and reliability
- System solutions by introducing High Voltage DC Transmission and FACTS

In this paper, the emphasis is on the FACTS devices developed by BHEL since 1990s and the performance of them in commercial operations. Whereas HVDC is used for bulk power transmission in view of better controllability and economics, FACTS is used to enhance the stability and usable transmission capacity of the existing transmission lines avoiding comparatively larger investment and right of way required to construct additional transmission lines.

The various FACTS devices designed and developed by BHEL are discussed in this paper. The commercial performance of these systems and their operational benefits are also discussed.

2. CONSTRAINTS IN UTILISING TRANSMISSION CAPACITY

The following are the constraints in utilizing the existing capacity of the transmission lines:

*Electronics Division, Bharat Heavy Electricals Limited, Bangalore 560026

Dynamic:

- Transient and dynamic stability
- Sub synchronous oscillations
- Dynamic over voltages and under voltages
- Voltage collapse
- Frequency collapse

Steady state:

- Uneven power flow between the lines
- Excessive reactive power flow
- Voltage issues
- Thermal issues
- Stability issues
- Power swing oscillations etc.

In order to overcome these problems and to enhance the capability and efficiency the transmission assets, FACTS and HVDC technologies are being used more and more in the modern power systems.

HVDC has got potential to handle power of the order of 10000MW and FACTS of the order of 10 to 500 MW.

3. A COMPARISION BETWEEN HVDC AND FACTS

A comparison between HVDC and FACTS devices are given below:

FACTS	HVDC
• Potential to handle power of the order of 10000MW	• Potential to handle power of the order 500MW
• Power and voltage control	• Power, voltage and frequency control
• Stability control	• Lower line cost
• Enhancement of usable transmission capacity	• Lower Right of Way requirements

4. LIST OF FACTS DEVICES AND THEIR APPLICATIONS

Following is the list of some of the FACTS devices and their applications:

FACTS DEVICES	APPLICATIONS
1) STATicCOMpensator (STATCOM)	1) PF correction, Harmonics reduction
2) Static VAR Compensator (SVC)	2) PF correction, stability, voltage control
3) Fixed Series Capacitor (FSC)	3) Enhance power transfer capability
4) Thyristor Controlled Series Capacitor (TCSC)	4) Enhance power transfer capability, stability, damping of power swing

5) Controlled Shunt Reactor (CSR)	5) Voltage control, power transfer capability
6) Static Synchronous Series Capacitor (SSSC)	6) Enhance power transfer capability, control, stability
7) Unified Power Flow Controller (UPFC)	7) Power flow and voltage control
8) Solid State Phase Shifting Transformer (SSPST)	8) Power flow control and loop flow control

In the above list, BHEL has involved in the indigenous development of 1, 2, 3, 4, 5 and 6. In section 7, the ratings of these devices developed by BHEL are given.

5. THYRISTOR VALVES AND CONTROLS FOR FACTS DEVICES

Thyristor valves used in FACTS for high power applications like transmission are air insulated, fibre optically triggered and water cooled. They are modular in construction. A number of thyristor modules with a number of thyristors connected in series in each module are arranged in series in a valve structure. Figure 1 shows a typical thyristor module developed by BHEL and is used in their FACTS devices like TCSC, SVC and CSR. Figure 2 shows the block diagram interfacing valve, de-ionized water cooling and control.

6. KANPUR- BALLABHGARH TCSC PROJECT

This is the first indigenously developed TCSC project, funded jointly by BHEL, Dept of IT and Power Grid Corporation of India Ltd (PGCIL). The project was test commissioned in 2004 and is in the process of putting in to continuous operation.

In FACTS devices and in HVDC, high power high blocking capability thyristors are used. Modular constructions are employed as discussed in the previous section to facilitate ease of production and operational maintenance. The other features of this thyristor module are:

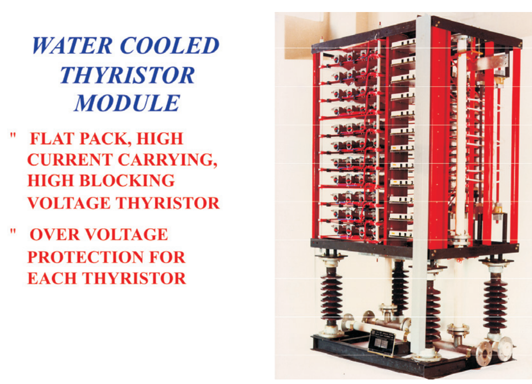


FIG. 1 THYRISTOR MODULE

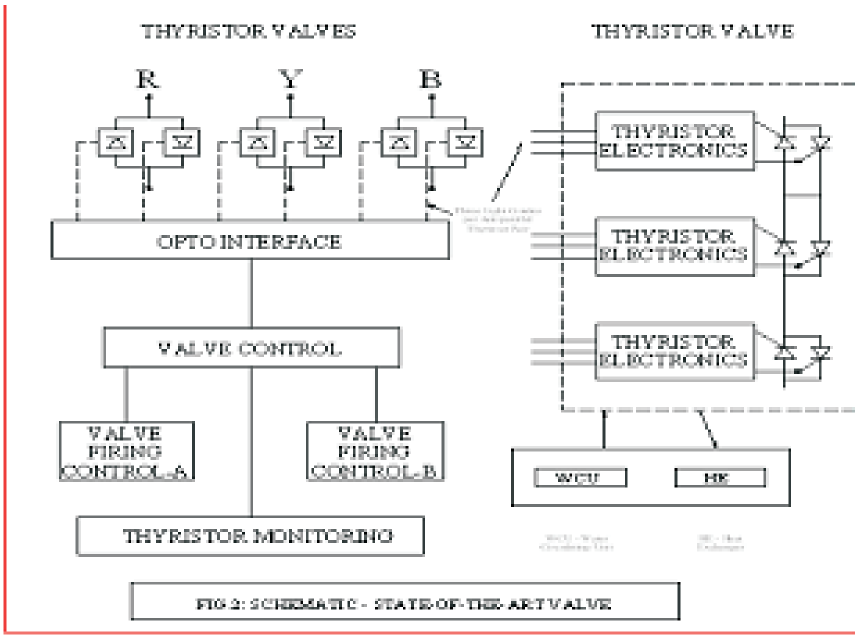


FIG. 2 INTERFACE BETWEEN VALVE, COOLING SYSTEM AND CONTROL

- Flat pack high current and high blocking capability thyristors (5-7 kV)
- Over-voltage protection at each thyristor level
- Monitoring of healthiness of each thyristor level and annunciation in the operator mimic etc.

Figure 3 shows the single line diagram of TCSC. It consists of 27% fixed and a portion of variable compensation. The variable portion consists of 12% fixed is made variable from 12 to 20 % by connecting anti- parallel connected phase controlled thyristor valves in parallel as shown in Figure 3. This 12% fixed and thyristor combination provides smooth variation capacitive reactance from 12% to 20% by varying the firing angle of the thyristor valves one in each phase. The rating of the thyristor valve used in this project is 22 kV, 1500A in each phase. Figure 4 shows the various components of the 400 kV Kanpur - Ballabhgarh TCSC installation.

A digital high speed controller has also been developed for the project indigenously with the following features:

- Master control
    - Current/reactance mode selection
    - Open/closed loop selection
    - Voltage and current measurements
    - Reactance/current level settings
- Sub-segment control
    - Depending on system requirements, thyristor blocked, bypassed or put in vernier mode

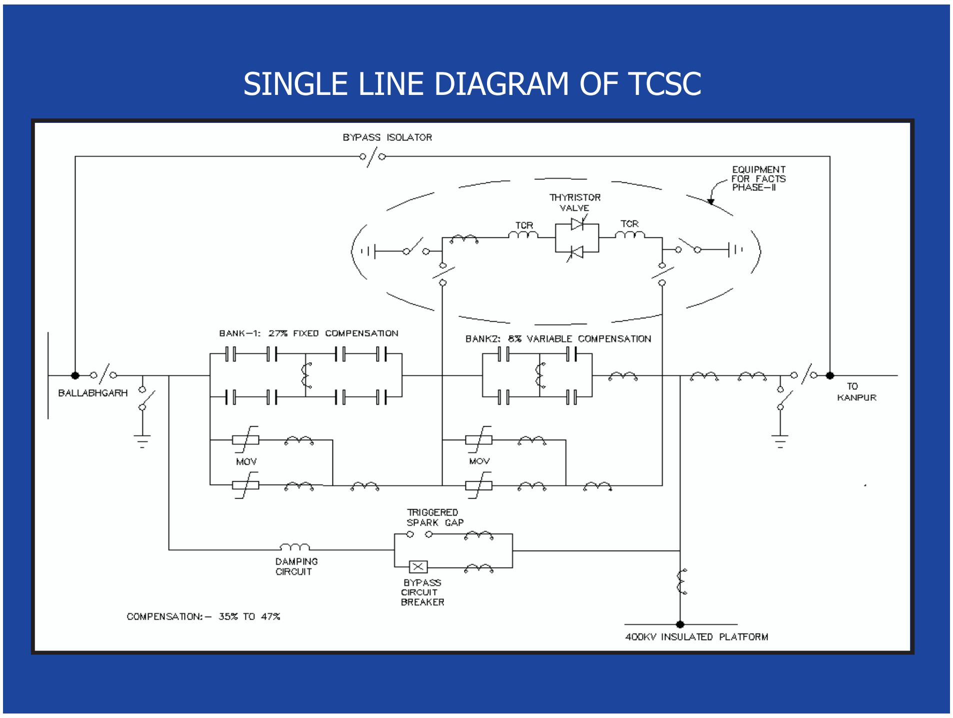


FIG. 3 SINGLE LINE DIAGRAM OF TCSC

- Protection: capacitor over voltage and TCR over current
- Damping control
- Firing control

- Base Electronics and Thyristor Monitoring
  - Interfacing of firing pulses to thyristor valve levels and monitoring healthiness of thyristors.
- Thyristor Control Unit
  - Converts light pulses received from Base Electronics in to electrical pulses to trigger thyristors. Also, sends thyristor healthiness status signals to thyristor monitoring system.
- Control system

Figure 5 gives block diagram of the control system. The detail control functional blocks and the respective transfer functions were developed using Matlab. The control blocks were realized using VME bus based digital hardware to meet the high speed requirements of the system. The control panel was tested using Real Time Digital Simulator (RTDS) at CPRI, Bangalore before sending to site.

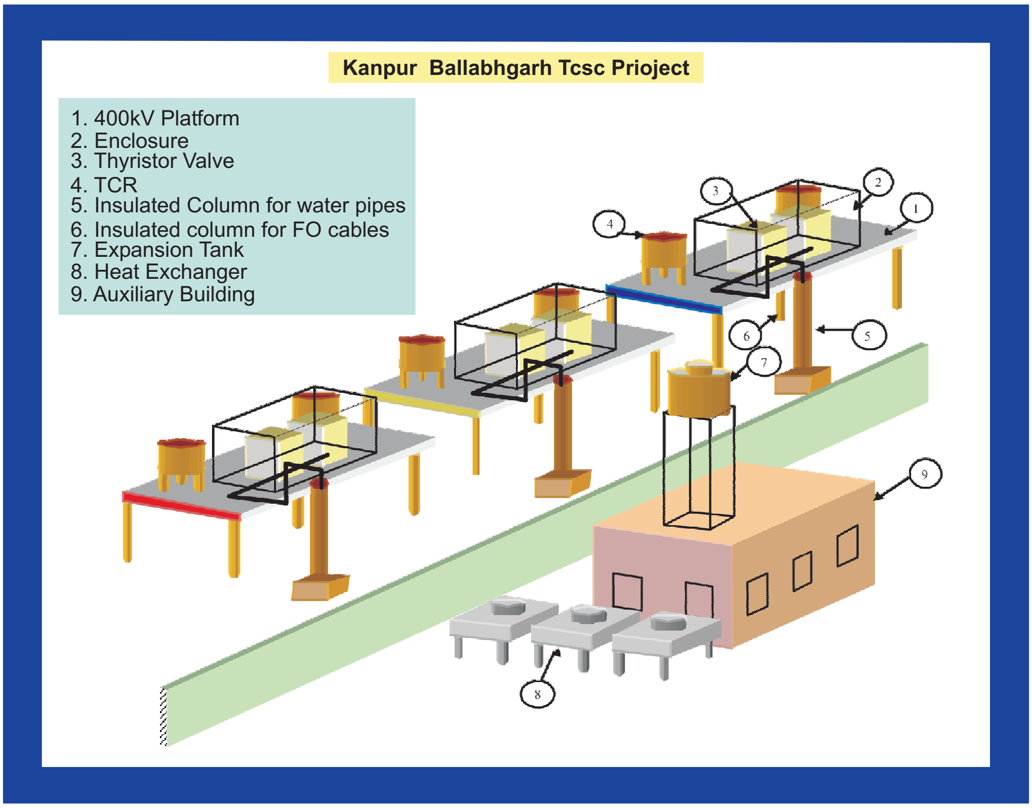


FIG. 4 KANPUR -BALLABHGARH TCSC INSTALLATION

In RTDS, the Kanpur - Ballabhgarh 400kV line, the dynamic equivalent of the rest of the system including the Delhi- Rihand HVDC system and Kanpur SVC were represented. In addition, Thyristor monitoring and Base Electronics were also represented in RTDS. The controller under test was hooked up to RTDS as hardware in the loop.

The following tests were conducted using RTDS to verify the performance of the control equipment [3]:

- Open loop and manual mode checking
- Impedance control mode: Figure 6 give the response of the control system for step change in impedance settings from 12.5 ohms to 26 ohms and back from 26 to 12.5 ohms.
- Current control mode: The responses of the controller with respect to step change in current settings 700A to 800A and back from 800A to 700A are shown in Figure 7.
- Frequency compensation: Frequency changed from 47Hz to 53Hz and found satisfactory performance.
- Capacitor over voltage and TCR over current protection
- Internal line faults

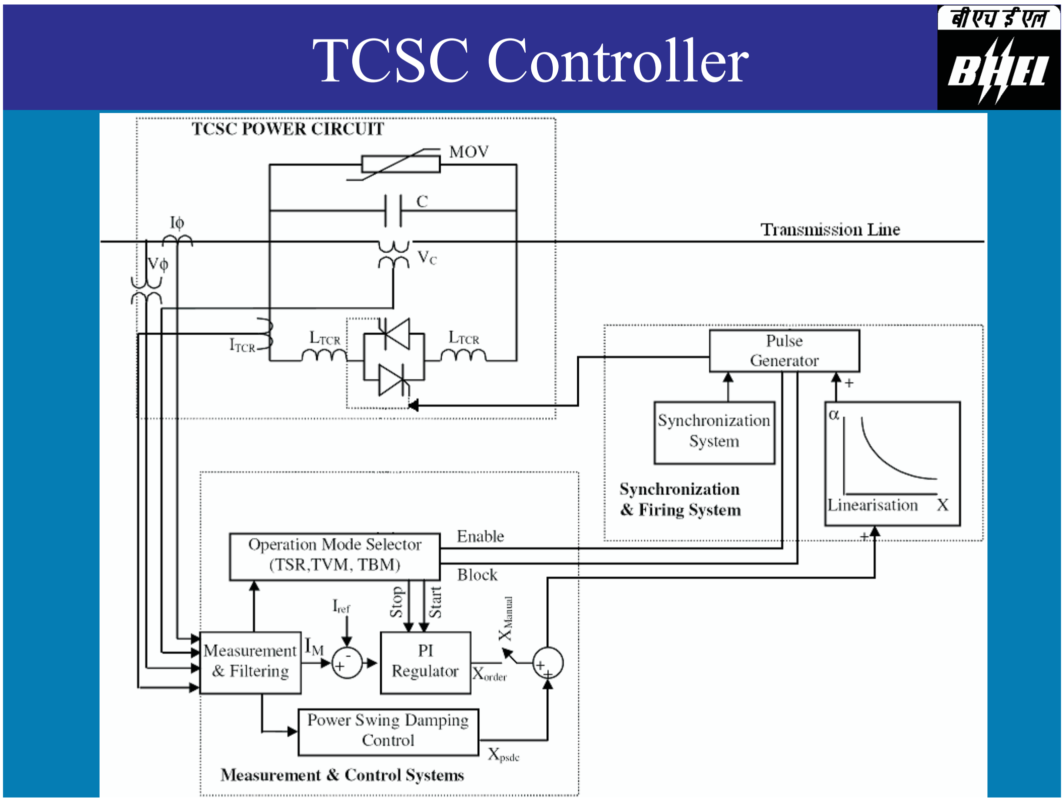


FIG. 5 BLOCK DIAGRAM OF THE TCSC CONTROLLER

- External line fault: TSR was enabled successfully as expected, helping to reduce the stress on MOV as shown in Figure 8.
- Power swing oscillations: Figure 9 shows the damping on outage of HVDC line.

7. CAPABILITIES OF BHEL IN FACTS AND CUSTOM POWER DEVICES

In section 4, list of some of the FACTS devices developed by BHEL has been discussed. The capability of BHEL in terms of the maximum ratings of these devices both for industrial and transmission system applications are given below:

7.1 Industrial and Distribution Applications

- Static VAr Compensators (SVCs)
- Thyristor based: Few KVAr to 105 MVAr, up to 33Kv
- STATCOM
  IGBT based: +/–500KVAr, 415V

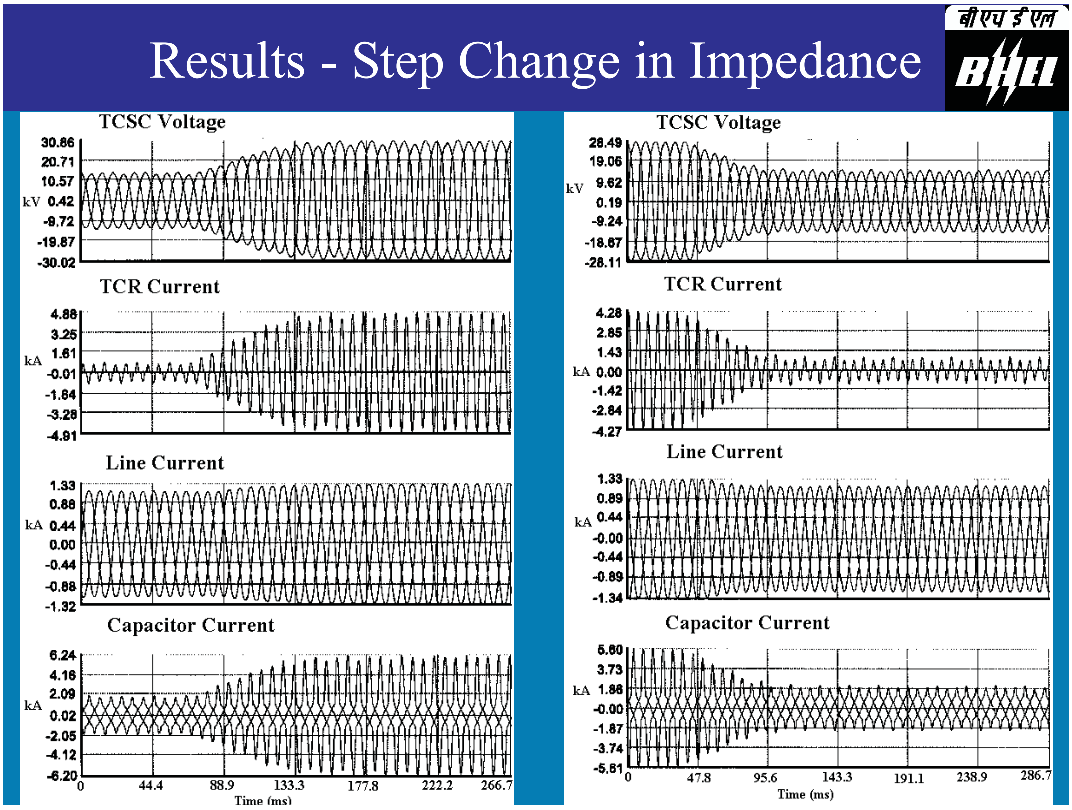


FIG. 6 STEP CHANGE IN IMPEDANCE

- Thyristor Switched Capacitors
  Different ratings up to11Kv, 8 MVAr
- Capswitches
  Up to few hundred KVAr, 415V

7.2 Transmission Applications

- Controlled Shunt Reactor (CSR)
  Thyristor based: For 400kV, 50/63/100 MVAr
- Thyristor Controlled Series Capacitor (TCSC)
  Up to 400kV, 200 MVAr
- Static Compensator (SVC)
  Up to 33Kv, 250 MVAr

The list of such schemes, supplied and commissioned by BHEL, is given in Tables 2 and 3.

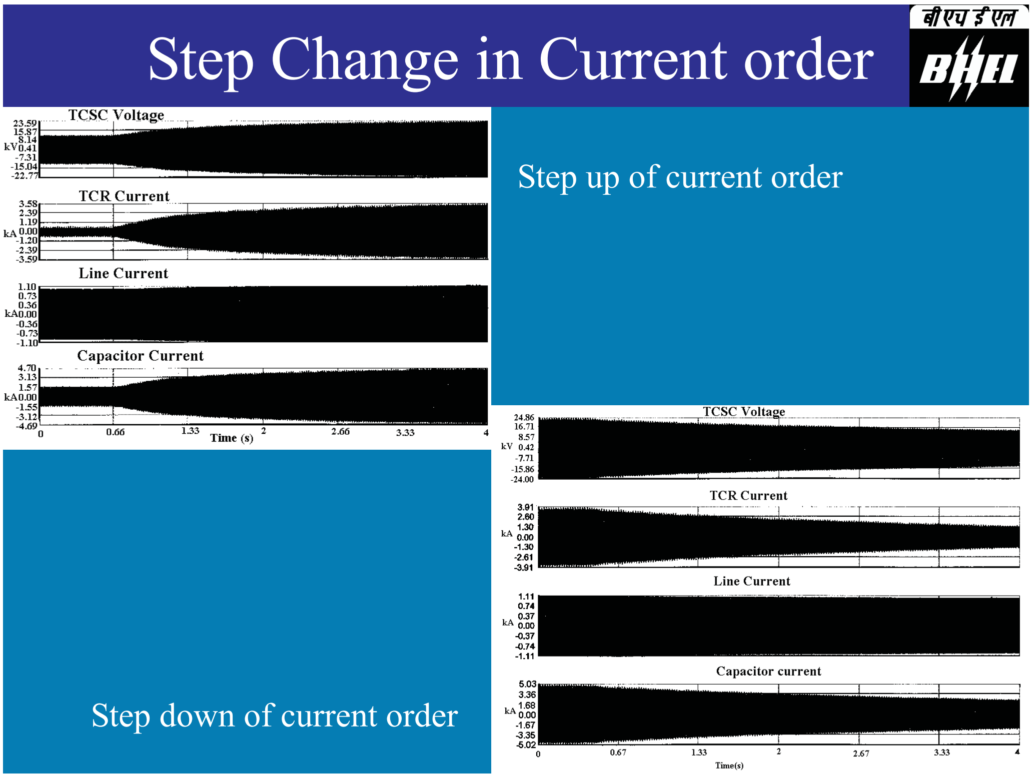


FIG. 7. STEP CHANGE IN CURRENT

8. OTHER FACTS DEVICES

The performance of other FACTS devices developed and supplied by BHEL is discussed in this section.

8.1 Statcom

BHEL has developed +/– 500 KVAr, 415V Voltage Source Converter using IGBT and built a STATCOM of 0 to 1 MVAr STATCOM and tested extensively as discussed in [4]. The system was commissioned at MIDHANI, Hyderabad in 2004. It is continuous operation since then correcting power factor close to unity.

8.2 Static VAr Compensator (SVC)

Table 1 gives the SVC systems supplied by BHEL, ratings varying from few KVAr to 106 MVAr, mostly for steel mill applications. The Thyristor valves designed are employing air insulation, water cooling and fiber optical triggering. The basic thyristor module construction looks as in Fig 1. This

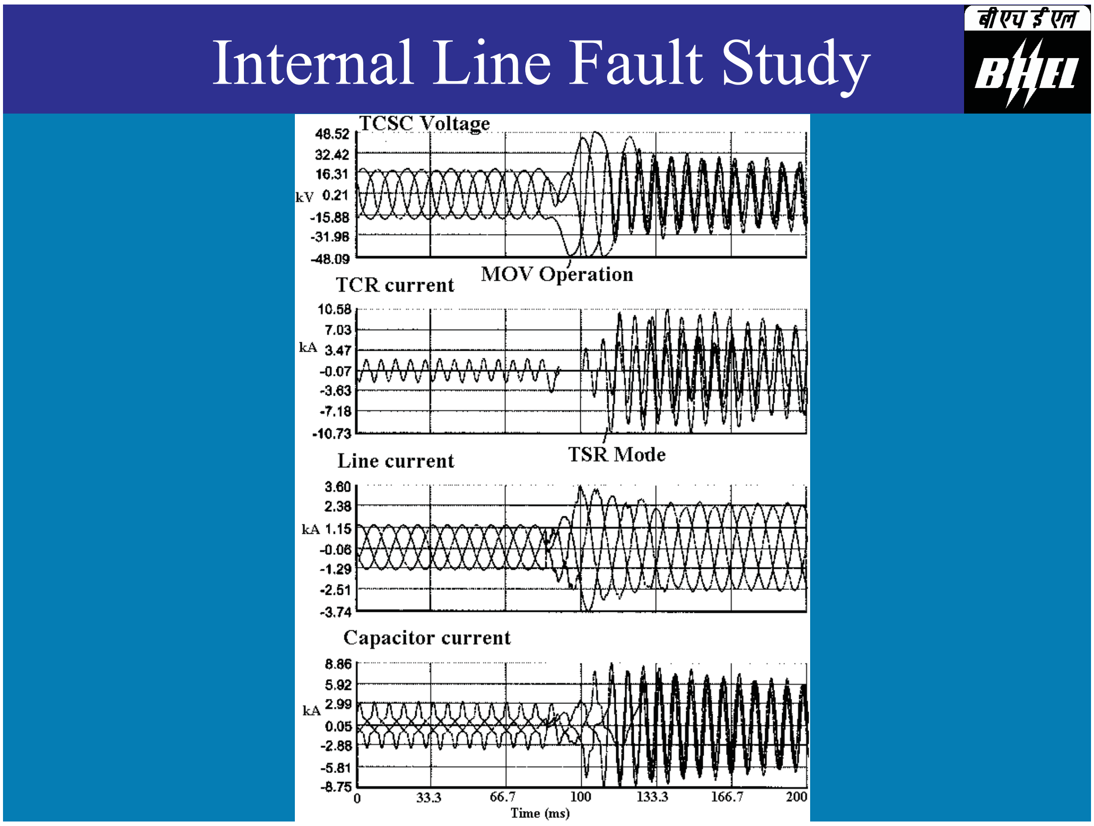


FIG. 8 RESPONSE TO LINE FAULT

system helps to mitigate voltage dips due to rapid change in reactive power in rolling mills. The power factor has also improved considerably resulting in saving in the electricity bill of the order of rupees 1 to 5 crores in a year.

8.3 Fixed Series Capacitor (FSC)

List of FSCs supplied by BHEL for both 220 kV and 400 kV transmission systems are given in Table 2. These systems are used to enhance the power transfer capabilities of the existing system and there by increasing the utilization of the assets and avoiding new investments required otherwise constructing new lines. The Right of Way (RoW) requirements are also avoided. The benefits realized in these systems are also indicated in Table 2.

8.4 Controlled Shunt Reactor (CSR)

In [5], a 420Kv, 50 MVAr developed, designed and supplied by BHEL has been presented. The system was commissioned at 400 kV Itarsi substation of PGCIL in the year 2002 and is in

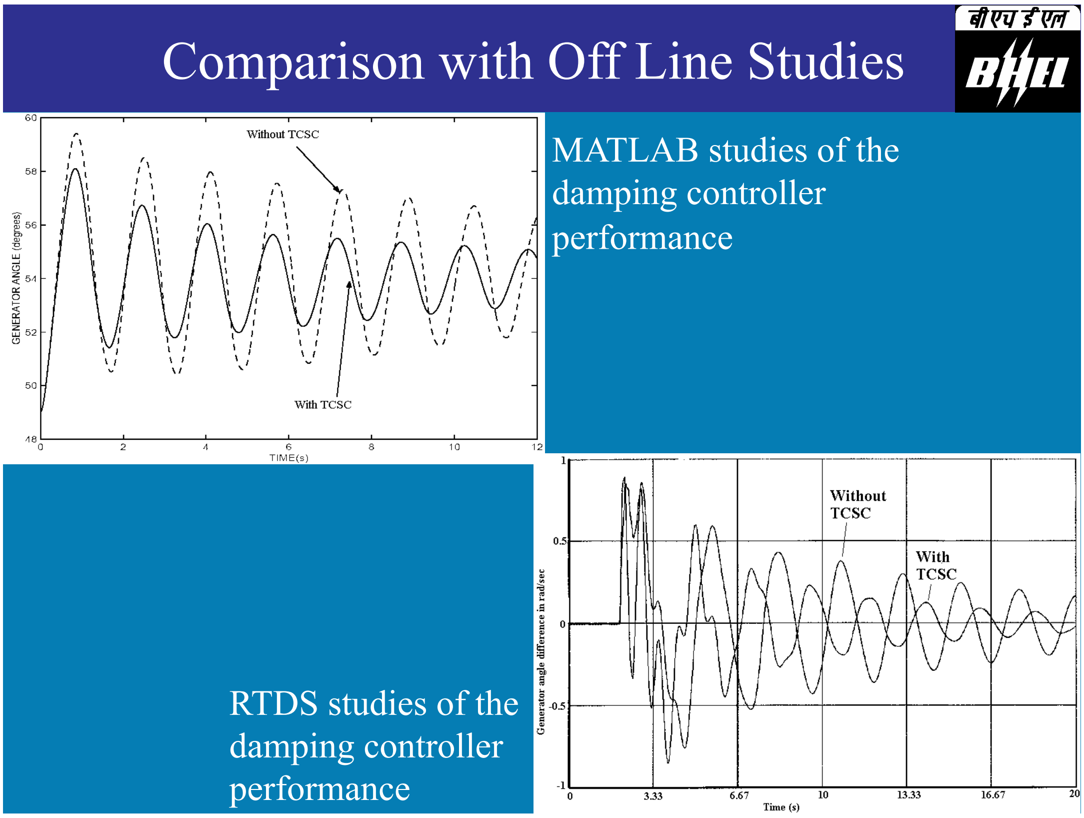


FIG. 9 DAMPING OF POWER SWING OSCILLATIONS

continuous operation since then. The thyristor valve is used on the secondary side of the 420/12 kV, 100% impedance transformer. The controller brings the thyristor valve in to ON state whenever under voltage is experienced. Thyristor switches OFF at natural current zero subsequent to blocking of the trigger pulses to thyristors under system over voltage situations. The point on wave switching is incorporated in order to ensure transient free switching of the CSR.

In [6], a compact thyristor valve has been evolved. The thyristor valve is kept ON only for a short duration of about 100 ms during switching ON and within this period the by-pass breaker which is connected parallel to the thyristor valve closes. This approach helps to reduce the duty on thyristor valve and there by elaborate cooling requirement of thyristor valve is avoided [6]. Some of the advantages of CSR are improvement of system voltage profile, reduction of losses as CSR is brought in to the system only when the system voltage deviates from the pre determined band avoiding continuous presence in the system and there by losses.

Table 1. *SVCs commissioned*

Sl. No.	*Customer/education*	*Rating*	*Sets*	*Remarks*
1.	Bhel, Bhopal GPN	11 KV +/– 3.75 MVAR	1 No.	Reactive power control
2.	MPEB. Barsoor	12 KV +/– 30 MVAR	1 No.	Voltage control of 220LV line
3.	TNEB	11 KV +15/– 30 MVAR	3 Nos.	Voltage control of 220LV line
4.	Sail, alloy steel plant, Durgapur	33 KV 0 – 35 MVAR	1 No.	PF improvement of sms with arc furnace
5.	Sail, BSP Bhalai	1 KV 0 –12.5 MVAR	1 No.	Reactive power control for rail and structure mill
6.	Tisco Jameshpur	33 KV 0 – 50 MVAR	1 No.	Reactive power control for hot strip mill
7.	Sail, RSP Rourkela	33 KV 0 – 50 MVAR	1 No.	Power factor voltage control of tandem mill
8.	Sail, VSP Vizag	1.65 KV +/– 5 MVAR	2 Nos.	Reactive power control of medium merchant structural mill
9.	HDC, Malanpur	33 KV ±100 MVAR	1 No.	Flicker control of arc furnace load

8.5 Solid State Phase Shifting Transformer (SSPST)

Phase Shifting Transformers are used to control the power flow in a line using mechanical tap-changer. In order to meet the fast responses required, thyristor based Solid State Tap-changers have been evolved in [7]. The challenge involved in this development is the commutation failure free switching of thyristor valves during up and down movements of taps. Fast response has been achieved as reported in [7].

9. CONCLUSION

The FACTS and Custom Power devices developed and supplied by BHEL to utilities and industries are SVC, STATCOM, TSC, TCSC, and CSR. The ratings of these devices supplied by BHEL are given in Tables 1 and 2.Out of these devices, CSR is the device introduced by BHEL to the FACTS family. A number of them in commercial operation and are instrumental in saving electricity bills of the order of about 1 to 5 crores of rupees per annum.

It is expected that more and more of such devices will find applications both in industries and utilities and contribute to enhance the effectiveness and efficiency of the system.

Table 2. *References for series compensation*

Bharat heavy electric limited references for series compensation

Sl. No.	*Customer location*	*Voltage*	*Description of the feeder*	*Capacitor bank rating*	*Improvements achieved anticipated*
01	KSEB Kozhikode	220 KV	221 km long Ldukki-Kozhikode	36 Ohm, 650A 23.4 KV, 45.75 MVAR 40% compensation	• increase in power transfer capability from 170 KW to 230 MW Voltage from 0.8 PU to 0.9 PU • loss reduction by 38 MW
02	PDD J&K Wanpoh	132 KV	152 kms long Udhampur-Pampore	43.3 Ohms, 400A 17.32 KV, 2 X 20.28 MVAR, 60% compensation	• Increase in power transfer capability by 70 MW • Voltage from 0.85 PU to 0.92 PU at Pampore • Loss reduction by 1.4 MW
03	MPEB Pichhore	220 KV Double circuit line	240 km long Gwalior-Bina	34.8 Ohms, 600A 20.88 KV, 2 X 36.6 MVAR, 45% compensation	• Increase in power transfer capability by 60 MW • Voltage from 0.84 PU to 0.9 PU at Gwalior • Loss education
04	MSEB Miraj	220 KV	84.6 km long Karad-Miraj	24.2 Ohms, 650A 14.8 KV, 30.66 MVAR, 70% compensation	• Increase in power transfer capability by 30 MW • Voltage from 0.82 PU to 0.9 PU at Miraj • Loss education by 2.5 MW
05	MPEB Handia	220 KV Double circuit line	Itarsi-Barwaha	34.8 Ohms, 600A 20.88 KV, 2 X 36.6 MVAR, 45% compensation	• Increase in power • Voltage and transient stability improvement
06	MPEB Bijawar	132 KV	210 km long Damoh-Bijawar-Prithvipur	50 Ohms, 350A 17.5 KV, 18.4 MVAR, 60% compensation	• Increase in power transfer capability by 6 MW • Voltage from 0.82 PU to 0.9 PU at Bijawar • Loss at Bijawar by 2.0 MW
07	MSEB Aurangabad	220 KV Double circuit line	89 km long Bableshwar-Aurangabad	22.3 Ohms, 650A 14.5 KV, 2 X 28.2 MVAR, 60% compensation	• Voltage from 0.85 PU to 0.92 PU at Aurangabad • Loss education by 1.85 MW

(Contd.)

08	Powergrid Kishenpur-J&K	220 KV Double circuit line	174 km long Kishenpur-Pampore-J&K	30 Ohms, 800A 24 KV, 3 X 19.2 MVAR, 42.86% compensation	• Increase in power transfer
09	Power grid Ballabhgarh	400 KV	390 km long Kanpur-Ballabhgarh line	35 Ohms fixed + 10.3 Ohm variable 1200A, 42 KV 151.2 MVAR fixed + 45 MVAR variable 27% fixed + 8% variable	Increase of power transfer capability by 120 MW
10	Power grid Ballabhgarh	400 KV Thyristor controlled series compensation	390 km long Kanpur-Ballabhgarh line	35 Ohms fixed + 10.3 Ohm variable 1200A, 42 KV 151.2 MVAR fixed + 45 MVAR variable 27% fixed + 8% variable	Outdoor thyristor value & controls being developed
11	Power grid Muradnagar	400 KV	395 km long Panki-Muradnagar line	53 Ohms 1200A, 76.32 MVAR/ phase 40%	Under execution

REFERENCES

[1] Narain G. Hingorani, National Workshop on Custom Power Devices, Key note Speech, organized by ER&DC, Dept of IT, GoI, June, 2003.

[2] Narain G. Hingorani, Laszlo Gyugyi, Understanding FACTS, IEEE Press, New York, 2000.

[3] Arunachalam M., Ghamandilal, Rajiv C.G., Babu Narayanan M.M., "Performance Verification of TCSC Control and Protection Equipment Using RTDS," Power System Computational Conference, Liege, 22-26, August 2005, session 32, paper 6.

[4] Bishnu P. Muni, Eswararao S., etc Arunachalam M., "Design and Development of Distribution STATCOM for Power Factor Improvement, National Power Electronics Conference, October 2003.

[5] Jithin Sunder SVN, Bhageria SC, etc, Arunachalam M, "Design,Testing and Commissioning of First 420 kV, 50 MVAr Controlled Shunt Reactor", CIGRE, Session 2002.

[6] Arunachalam M, Dipak Dutta, Khoday C.D., "An Improved Controlled Shunt Reactor", Patent filed in 2005.

[7] Arunachalam M, Dipak Dutta, Khoday C.D., S VN Jithin Sunder, "A Solid State Tap-changer for Phase Shifting Transformer," Patent filed in 2005

ELECTRONICS AND COMMUNICATION ENGINEERING

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

NEW ADVANCES IN SENSORS RESEARCH

V.R. SINGH*

ABSTRACT

Sensors are heart of an instrumentation system. With the advancement of materials technology, new types of sensors are being developed now, day by day. New advances in sensors research are given here. Trends in sensor technology, sensor materials and new sensors with biomedical and engineering applications are given. Recent research in sensors at this Laboratory is also mentioned in brief. The new sensors would explore new avenues of research in science and technology for newer applications.

1. INTRODUCTION

1.1 The Sensor

Sensors and transducers are integral parts of instrumentation in science, engineering, medical and other fields. Sensor technology is a rapidly growing field that has significant potential to improve the operation, reliability, serviceability, and utility of many engineering systems. Advances in materials science and engineering have paved the way for the development of new and more capable sensors. Also, with the advancement of technology, nanostructures and nanomaterials are now used to develop new types of sensors. New nano-sensors are being developed and used for various industrial, biomedical and other applications [1-53]. Nanobiotechnology tools and processes of nano/microfabrication are used to build new micro/nano scale devices and biosystems [7-21].

Sensors can be categorized in a number of different ways, most notably by either their chemical composition or their principle of operation. The basic terms associated with sensor technology are important to be understood. A *sensing element* is the fundamental transduction mechanism (e.g., a material) that converts one form of energy into another. A *sensor* supplies a usable output in response to a specific measurand (input) in a predictable way; a sensor's physical configuration includes the sensing element together with its physical packaging and external connections (e.g.,

---

*National Physical Laboratory, New Delhi-110012, India, E-mail: vrsingh@ieee.org, and Electrical and Instrumentation Department, Thapar University, Patiala-147001 1ndia

[1]Talk delivered at TIET – Thapar University, Patiala on December 21, 2006

electrical or optical). A *sensor system* is a sensor and its assorted signal processing hardware (analog or digital) with the processing either in or on the same package or discrete from the sensor itself.

The diversity of sensor technologies and applications and the resulting diversity of materials needs is to lead to conclude that the concept of an "ideal" sensor material is inadvisable. It is frequently possible to fulfill a given sensing need by more than one type of sensor. Thus, identification of the "best" sensor material should only be done within the context of a specific application.

Two categories of sensors and sensor materials development can be distinguished: high-volume, low-cost sensors and low-volume, high-cost sensors. Further identification of opportunities for sensor materials development and application areas are to be explored.

1.2 Trends in Sensor Technology

The recent research [7-8, 19-25, 50-53] in the following areas is important:

(i) *Current sensor development is trending toward increased complexity in sensor systems.*

The greater flexibility and lower cost of advanced, integrated electronic technology allows signal processors to be reduced to a microelectronic chip; however, from the perspective of the end user, the sensor system appears simpler.

(ii) *The principal technical drivers for sensor development are becoming enabling/supporting technologiesother than materials technology.*

Most recent advances in sensors have not originated from the synthesis of new transduction materials; they have been due to sensor technologies not heavily dependent on transduction materials and to innovations that have significantly decreased the cost of a sensor system.

(iii) *Networking of large sensor systems can provide improved spatial and temporal sampling in low-cost, lowmaintenance systems.*

A network of sensors can provide data to a central processor that monitors performance or helps characterize defects. Also, the individual outputs from an array of sensitive but moderately selective transducers can provide a composite indication that is both sensitive and selective.

(iv) *Sensor R&D lends itself to dual use and commercialization efforts.*

Sensors are an enabling technology with a wide spectrum of applications.

1.3 R&D Strategy

(i) *Development of sensors solely for the sake of advancing sensor technology.*

Historically, sensor research and development efforts [1-53] have been pursued for for large applications that required sensors. A new developmental approach will require the implementation of a research planning process that addresses the needs of a broad set of users and applications.

(ii) *A generally accepted framework to describe both sensor application requirements and sensor performance capabilities is needed.*

A common set of descriptors for use by sensor users, suppliers, and researchers is to be identified as the most important step in facilitating the identification of sensor materials R&D opportunities and in accelerating the development and use of advanced sensor technologies.

(iii) *Sensor development centres may provide useful guidelines for improving sensor R&D strategy.*

Essential characteristics include: a multidisciplinary approach with emphasis on teamwork; capabilities ranging from an initial proof of concept through engineering prototypes; focus on selected sensor technologies for a broadly defined range of applications; and strong linkage to industry to guide the general relevance of the research.

1.4 Objective

As new types of sensors are being developed now, day by day, new advances in sensors research are given here. Trends in sensor technology, sensor materials and new sensors with biomedical and engineering applications are mentioned. Recent research in sensors at this Laboratory is also given in brief. The new sensors would explore new avenues of research in science and technology for newer applications.

2. SENSORS MATERIALS RESEARCH

2.1 Sensor Materials

Different types of materials are used for the development of various types of sensors [44-45]. As said above, advances in materials science and engineering have paved the way for the development of new and more capable sensors. Drawing upon case studies from manufacturing and structural monitoring and involving chemical and long wave-length infrared sensors, this

is to expedite the consideration of new and novel sensor materials.

The trend is toward the development of materials tailored to specific, or targeted, applications rather than toward fundamental R&D without a specific application-focus (i.e., toward materials development driven by “market pull” as opposed to “technology push”). However, technology-driven, leading edge research is vitally important, since the results from these efforts have the potential to create entirely new products and markets [44, 45, 51].

2.2 Materials R&D

(i) *Sensor materials include all materials required by the sensing system.*

These materials encompass those required by the transducer, package, and leads [47, 52].

(ii) *Sensor materials R&D can be divided into two main categories: the development of new materials and materials engineering for particular applications.*

These two categories frequently require very different approaches to materials R&D [47, 49].

(iii) *High impact for new sensor materials in the near term will come primarily from R&D on existing materials rather than synthesis of new compositions of matter.*

The fundamental research on new compositions of matter is the highest risk element of sensor materials R&D. Exploiting materials developed for purposes other than sensing can lead to rapid sensor technology advancements at relatively low cost and risk [42, 47].

(iv) *Materials processing science will be the foundation for developing affordable sensor materials.*

Materials synthesis and processing will facilitate the transfer of innovations in materials science to commercially viable products [45, 48].

(v) *Universities and federal research laboratories play a critical role in conducting frontier research.*

Universities may conduct frontier research and to use such programs as vehicles to educate students. Federal research laboratories generally sponsor frontier research and conduct a portion of the research in-house to keep abreast of the leading-edge technologies. Long-term commitment to such research is essential to remain technically competitive internationally [48, 50].

2.3 Manufacturing and Processing

Intelligent processing can reduce product variability, optimize use of processing facilities, and decrease costs. As defined by the committee, intelligent processing involves event-based control of process variables [45-52].

Sensors are currently the weak link in intelligent processing. They must perform reliably in hostile manufacturing environments and provide data that permit accurate representation in time and space of the changes occurring as the material is processed.

Intelligent (self-directed) processing of aerospace structural polymeric composites (e.g., thermosetting resin systems) is an illustrative example. In this case, the state of cure of the polymer is determined, and then the process variables are adjusted to achieve the most efficient cure cycle. Sensor materials and technologies are improved for this application, notably by directly measuring the molecular structure of the polymer to more precisely determine the degree of cure. Laser-fiberoptic sensor technology.

The fabrication and processing of complex semiconductor materials is an important and expanding industrial endeavor. There is a significant need for in situ diagnostics to permit precise on-line process control during epitaxial growth of electro-optical thin films. The manufacture of low-cost, reproducible, uniform, and tailorable structures needs noninvasive sensors to measure film thickness, composition, interlayer sharpness, and other properties. Optical sensing technologies (e.g., ellipsometry, laser-induced fluorescence, and fiberoptic probes) can significantly improve the processing of complex semiconductor devices.

3. SENSOR TECHNOLOGIES

3.1 Structural Monitoring and Control: Smart Technology

"Smart" structures that incorporate active materials are emerging as an important broad-based discipline. A step toward the development of such "smart" structures is the incorporation of sensor systems that provide accurate information describing the state of a structure at any given time throughout its life cycle. This life-cycle management [29, 44-48]

The availability of advanced sensors and actuators, together with developments in signal processing, communication, and control technologies, has led to a surge of interest in smart structural materials systems that can adapt to an ill-defined, changing environment. Evaluating of the long-term performance of sensors used for in-service monitoring, understanding the sensor/host interactions for embedded sensor applications, and focusing on improved long-term reliability of sensors for in-service environments, are studied.

3.2 Long-Wavelength Infrared Sensors

The sensing of electromagnetic radiation is essential for a wide variety of activities [13, 15, 20, 45]. Sensing radiation in the long-wavelength infrared (LWIR) window (spectral region with a wavelength of 8 mm to 14 mm) allows detection of unilluminated objects that are approximately at room temperature. Infrared sensors are attractive because they are non-contracting and can quickly sense a temperature change over an area. This capability is important for applications such as measuring part temperature, detecting process defects, enabling night vision, and identifying chemical species. There are three materials strategies—at different levels of maturity—for obtaining high-efficiency photodetector LWIR sensors.

- ***Mercury cadmium telluride (MCT) compounds:*** The quality of LWIR MCT has improved over the last decade, and continued incremental improvements may eventually yield temporally stable, uniform detector arrays for LWIR applications. However, materials instabilities still result in major challenges.
- ***Artificially structured materials with tailored band gaps based on multiple-quantum-well devices (e.g., GaAs/AlGaAs system):*** The theoretical sensitivity of these sensors is lower than that of MCT; however, because of superior response uniformity, arrays of these structures already exhibit performance exceeding that of MCT detector arrays for selected applications. These materials can be developed using real-time sensor-based process controls developed for GaAs.
- ***Artificially structured materials with tailored band gaps based on strained-layer superlattice structures (e.g., those that exploit In [As, Sb] alloys):*** The fundamental detectivity limit of these materials is higher than that of MCT. It is the least mature of the technologies, and substantial improvements in performance appear to be possible, since no fundamental limitations have emerged as obstacles.

3.3 Chemical Sensors

Chemical sensors are devices or instruments that determine the detectable presence, concentration, or quantity of specific chemical substances (analytes). Arguably the most severe limitation on current chemical sensor technology is the inability to obtain a selective response to a target analyte, given the millions of known molecular species, the variations in environmental conditions (presence of water, etc.), and the variations in analyte amount or concentration by factors of 1023 or greater. Applications of chemical sensors include monitoring manufacturing processes, environmental sensing, and health monitoring [10-12, 19-25].

Direct-reading selective sensors, such as electrochemical sensors, detect species in the gas or liquid phase. They achieve molecular selectivity through interaction at the sensor-sample interface. This selectivity depends on recognition of the size, shape, or dipolar properties of the analyte by molecular films, phases, or "receptor sites," with resultant selective binding, absorption, or permeation of the analyte. Selectivity of direct-reading chemical sensors can be enhanced through the development of analyte-specific films, membranes, and coatings.

Miniaturization of these sensor systems could lead to compact, lightweight, portable monitoring systems. As an alternative, analyte selectivity can be addressed by using sensors that incorporate preliminary sample separation steps, such as chromatography and electrophoresis. Conventional analytical chemistry methods, such as mass spectroscopy, may then be used if the analysis can be performed fast enough and the equipment is significantly compact and inexpensive. Miniaturized and integrated platforms that incorporate both separation and detection systems may be developed further.

3.4 Intelligent Sensors: Nano Technology-Based Sensors

Current nanotechnology permits the operation on the scale of atoms and molecules. This promises to have a dramatic impact on sensor design and capabilities. Nanotechnology has become a key technology in sensor development. Sensors can now exploit novel properties of materials at the nano-scale. Chemical and biological materials operate at the nano-scale, hence nanotechnology is well suited to design of chemical and biological sensors [8-22, 25, 29].

Initial research in nanotechnology involved ministration of the macro techniques. The small size of these sensors leads to reduced weight, low power requirements and greater sensitivity. Main industries befitting from nanotechnology include transportation, communications, building, medicine, safety and security.

3.5 Wireless Sensors

Wireless sensors are being developed at a rapid pace, with many different standards and communications protocols in play. The management and integration of different hardware products and wireless protocols is growing faster. A single platform for application development and integration of wireless sensors may be used easily [45, 52, 53].

Reusable software components that allow for the rapid deployment of 'end-to-end solutions' by connecting to back-end systems or custom applications are possible.

Any number of sensor points may be supported for different types of wireless protocols including mesh networks (Zigbee or proprietary platforms), active RFID tags, point-to-point RF sensors, and other types of smart sensors.

A universal system utilises components called application profiles to handle specific types of wireless sensor data. This allows a developer to use an existing profile from the software library. This is especially useful for developers of Smart Sensors, providing an immediately available software management and integration framework for wireless sensor systems.

3.6 Health Care Sensors

For better healthcare facilities to better manage, archive and process images and information related to mammography screening, the systems may be Information Management Solutions (IMS) as these solutions–driven by the powerful software platform–combine software, hardware and professional services to enable the consolidation, control, continuity and access to all fixed archived content within a healthcare facility or among multiple healthcare facilities within a geographic region. With IMS, the right medical images and information are available to medical practitioners in the right place and at the right time. These data comprise analytical tools required for quality assurance and compliance, as well as visualization tools dedicated to streamlining the often-complicated workflow processes involved with mammography [19-25, 29-31, 42-44, 50-53]. The solutions that evolve from integration will enable healthcare facilities to collect, retrieve, store and distribute digital images and information on demand, thus helping enhance workflow efficiencies.

Patients with personal health record management tools are provided to securely upload and the digital medical images and electronic health records into a private web site. The program allows patients to identify a service bureau to digitize mammography films; store prior mammograms from previous healthcare providers; locate digital imaging facilities; and link to healthcare provider records stored.

Health data may enable women, who are undergoing screening mammography exams, may be linked to the main centre. A physician and patient together may be tied up for the patient's entire life, including self-populated family history and self-breast examination logs [45, 46, 50].

A data management and communications infrastructure enables on-demand access, visualization and distribution of diagnostic quality images and related clinical data [47].

3.7 Nano Mechanical Oscillators

Nano quartz crystals in electronic circuits can be used to detect and identify bacteria and viruses.

An AFM uses a tiny probe that moves slowly just above a surface. Electrostatic attraction or repulsion between the atoms in the tip of the probe and those in the surface causes the probe to move up and down, creating an image of the surface so detailed that individual atoms show up as bumps. Alternatively, the AFM can be used in "tapping mode," literally bouncing off the surface.

MEMS (microelectromechanical systems) are machines with moving parts measured in microns, or millionths of a meter; NEMS (nanoelectromechanical systems) are measured in nanometers, or billionths of a meter. A nanometer is about the length of three atoms in a row. When the NEMS oscillator is too small to be observed by laser light, it could still be coupled to a MEMS probe that in turn would be large enough for a laser readout.

To measure the vibration of a nanomechanical oscillator, the AFM probe moves along the length of the oscillating rod. The result is a complex bouncing interaction between the probe and the oscillator — imagine shaking one end of a spring and watching the vibrations at the other end — from which the frequency of vibration of the oscillator can be determined mathematically.

Nanomechanical oscillators are often cited as potential tools for detecting bacteria, viruses or other organic molecules. An array of tiny cantilevers might be created with antibodies to many different pathogens attached to them. An experimental solution could then be washed over the array, allowing microbes to bind to the cantilevers with matching antibodies. Since the cantilevers are so tiny, an attached bacterium or virus represents a significant change in mass, which changes the frequency at which the oscillator will vibrate.

In a practical device, a MEMS probe could be mounted above each NEMS oscillator to read out which oscillators in the array show a change in frequency — and thus identify which pathogens are present [48].

4. NEW SENSORS

MEMS Device Controls Human Cells

To get liver cells to express specific liver functions, they needed to manipulate supporting stromal cells for 18 hours. The device would allow biologists to keep the stromal, or connective tissue, cells, close together, but would avoid physical contact over the entire 18 hours, which would harm the cells.

Compact, Noninvasive Nano Flow Sensor

MEMS-based thermal anemometry technology, the instrument directly measures the mass-flow rate of liquids in an isolated flow channel and displays it on an LCD readout.

Nanoporous Silicon-based Biosensor Detect Selected Bacteria

The system is based on the nanoporous silicon-based biosensor technology to detect selected bacteria. One of the bacteria covered in the field of use is E-coli.

Industrial Use of MEMS and Nanotechnology

MEMS and nanotechnology-based solutions are available for application-specific developments.

Light Power Sensors

Especially in cell and molecular biology, nanotechnology based instruments and sensors enable better understanding of life at the molecular scale. High-end optical techniques play an important role and have a huge potential for the life sciences.

Sensors for Cancer Nanomedicine

Emerging areas in Cancer Nanomedicine with nanotechnologies for treating oncology are developed.

E. Coli Sensor

The sensor features a vibrating cantilever, supported at one end and coated with antibodies. The antibodies are specific to the desired target such as E. coli, anthrax or proteins that are biomarkers for diseases such as prostate cancer. When the target is present in a sample flowing past the sensor, it binds to the cantilever and changes the frequency of vibration so it can be read electronically.

Sensor to Detect Presence of Water

Metal complexes are embedded in the film steal electrons from the water molecules. When the number of electrons in the metal complexes changes, so does their color, and this change can be read optically. Devices based on optical readout do not need to be wired directly to larger-scale electronics

New Environmental Sensor

The Cloud and Land Surface Interaction Campaign and a coordinated experiment known as the Cumulus Humilis Aerosol Processing Study involve a combination of aircraft and ground-based measurements to enable researchers to capture the full breadth of fluctuations in carbon dioxide, moisture, aerosol particles, cloud properties and radiative energy from within the Earth's surface to the top of the atmospheric boundary layer.

Nano Pressure Sensor

A new implantable intravascular biosensor to monitor health and disease progression in humans is developed.

New Nanotube Sensor

A new type of nanotube built of gold, silver and other nanoparticles has been developed [50]. The tubes exhibit unique electrical, optical and other properties, depending on their components, and as such, may form the basis for future nanosensors, catalysts and chemistry-on-a-chip systems.

Nanotubes are tiny cylinder-shaped structures (a nanometer is one millionth of a millimeter). Discovered in 1991, the first nanotubes were made of carbon and captured the attention of scientists worldwide when they proved to be the strongest material ever made (100 times stronger than steel), as well as being excellent conductors of electricity and heat. Here, the properties can be altered by choosing different types of nanoparticles or even a mixture, thus creating composite tubes. Moreover, the nanoparticle building blocks can serve as a scaffold for various add-ons, such as metallic, semiconducting or polymeric materials.

The tubes are produced at room temperature. First, a nanoporous aluminum oxide template is made which is modified chemically to make it bind readily to gold or silver nanoparticles. When a solution containing the nanoparticles (each only 14 nanometers in diameter) is poured through, and bound to the aluminum oxide membrane, creating multi-layered nanotubes in the membrane pores. In step three, the aluminum oxide membrane is dissolved, leaving an assembly of free-standing, solid nanotubes.

The resulting tube is porous and has a high surface area, distinct optical properties and electrical conductivity. Collectively, the tube‘s unusual properties may enable the design of future sensors and catalysts (both requiring high surface area), as well as microfluidic, chemistry-on-a-chip systems applied in biotechnology, such as DNA chips (used to detect genetic mutations and evaluate drug performance). Various metal and composite nanotubes, including gold, silver, gold/palladium and copper-coated gold tubes have been developed. A nanotube sensor is shown in Fig. 1.

Gene-based Nano-Sensors

Figure 2 depicts bacteria laden with “smart nanoparticles,” which could carry genes, drugs, nanosensors or other cargo into the interior of cells. The nanoparticles are used to precisely position cargo inside cells for the early diagnosis and treatment of cancer and other diseases. The green and red spheres represent drug-carrying polystyrene nanoparticles of varying sizes [49].

FIG. 1

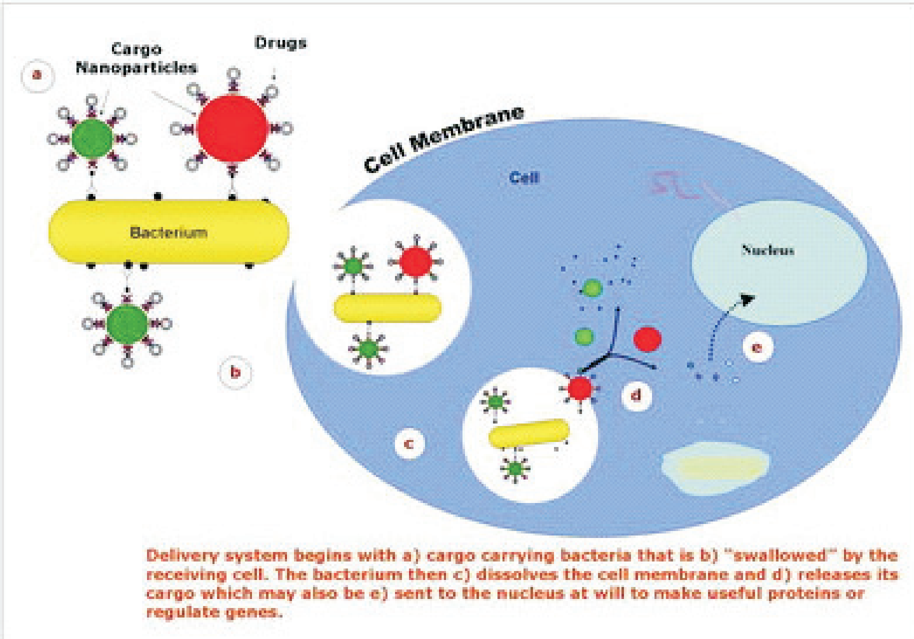


FIG. 2

New Nano-Space Sensors

A chemical nanosensor has been developed for space research recently (Fig.3). The nanosensor module (middle image) is roughly 12 centimeters by 12 centimeters by 4 centimeters. It contains a data-acquisition board, a sampling system, a tank of nitrogen dioxide, and the sensor chip (bottom image). The nanosensor is a one-centimeter-by-one-centimeter silicon wafer that has 32 channels for detection and uses different nanostructure materials for sensing [19,51].

5. NEW APPLICATIONS

Physical sensors, electro-sensors, chemical sensors and bio-sensors can all benefit from nano-technology [24-25, 30-53].

Walter de Heer has devised a *Balance*. This balance is the smallest in the world. A particle to be weighed is placed on a nanotube. The mass of the particle is calculated from changes in the vibrational resonance frequency with and without the particle. The balance can be used to weigh signal molecules.

Measurement of electricity is important and the bases for a large number of sensors. A *submicron mechanical electrometer* demonstrate charge sensitivity below a single electron charge.

The use of sensors based on nanotubes can be used to *trap molecules* . Results might be sensitive enough to detect single chemical and biological molecules.

FIG. 3

Nanotechnology will also enable the very selective, sensitive detection of a broad range of biomolecules. For example cover the surface of a chip with millions of vertically mounted nanotubes 30–50 nm in diameter. When the DNA molecules attached to the ends of the nanotubes are placed in a liquid containing DNA molecules of interest, the DNA on the chip attaches to the target and increases its electrical conductivity. These nanotubes are connected to an integrated electric circuit and the results analyzed.

A lightweight, portable chemical detection sensor is developed to play an important role in the detection of biological and chemical weapons.

The new possibilities now available are endless. The ability for atomic bricklaying lets the designer be precise when designing new sensors. This ability will reduce the amount of defects in new devises. At the atomic level the materials have new properties which can be exploited like surface and quantum effects. Nanotubes have been shown to have a number of uses in sensor technologies. They are extremely narrow hollow cylinders made of carbon atoms. The orientation of the carbon atoms can affect the conducting and semi-conducting properties. These can be used to integrate electrical circuits for the design of sensors. These nanotubes can be grown on existing structures.

The existing IC technologies can be used to integrate these nanosensors into integrated electronic circuits. The sensor chips can be used as building blocks to build new more complex sensors.

Nanotechnology has a vast number of applications, say in data storage proximal probes. These can make and read nanometer-scale indentations in polymers. The current densities are around 1x1012 bits per square inch which is greater then the current magnetic based recording devices.

Advances in nano-manufacturing have been from the top-down approach and bottom-up approach. Conventional microelectronics (lithography, etching, and deposition) has approached the nanometer scale. The current line widths in chips are near the 100 nm. Manufacturing from the bottom-up is also possible using individual atoms and molecules to build useful structures. Designers can also combine micro and nanotechnologies to develop new sensor systems.

Using computers for the design of new nanotechnology is important. Understanding the interactions of atoms and molecules is required when simulating using powerful computers and algorithms.

Working at such a small level also has it's own problems. The new sensors are very sensitive. These sensors are prone to degradation from the effects of foreign substances, heat, and cold. At such a small scale the micro effects become more significant. This problem can be partially overcome by installing hundreds of sensors in a small space. This allows malfunctioning devices to be ignored in favor of good ones.

Lab-on-chip sensors have been developed for different applications, with the design aspects and technological developments.Current lab-on-a-chip products automate only two or three analytical steps, but further advantages can be realised when multiple steps are fully automated. There is also a need for automation to increase throughput and reproducibly. Existing technologies, such as gel electrophoresis, and to address novel analytical problems that cannot be solved today, can be solved. This may well come about because of the convergences of the "micro" lab-on-a-chip world with the increasing miniaturization of the "macro" HPLC world. Indeed, "nano" LC systems are rapidly becoming a reality.

6. SENSORS RESEARCH AT NPL

Sensors research has been pursued at NPL-India extensively in different types of sensors and transducers for various applications in science, engineering and biomedicine. Piezoresistive semiconductor and piezoelectric ceramic materials for sensor applications have been developed indigeneously. Semiconductor, ultrasonic, polymers, biological and biochemical types of sensors have been designed and fabricated [5, 39, 41].

Research on biochip-based and piezo-composite based sensors is also going on. Figs 1 and 2 show respectively The conductance value (Fig. 4) and SEM micrograph Fig.5) of a newly developed piezo-electric composite material are found comparable to piezoelectric ceramics for better use in sensor development.

Piezoelectric composites and polymers have been used in ultrasonic transducers for various applications due to good acoustic matching. Piezocomposite with 0-3 connectivity made with a piezoceramic powder dispersed in a polymer or piezoelectric polymer and copolymer has been used in large areas of underwater acoustics or in thin layers of high frequency biomedical imaging applications or NDT (Nondestructive applications). These soft materials are molded in different shapes, say as focusing bowls, to develop focused ultrasonic sensors. The properties and behaviour of the piezocomposites depend on properties of constituent materials and the local arrangement of different phases e.g., 1-3, 2-2, 3-3 and 0-3, in a particular case. Smart micro and nano chips are also being studied.

CONDUCTANCE VALUES

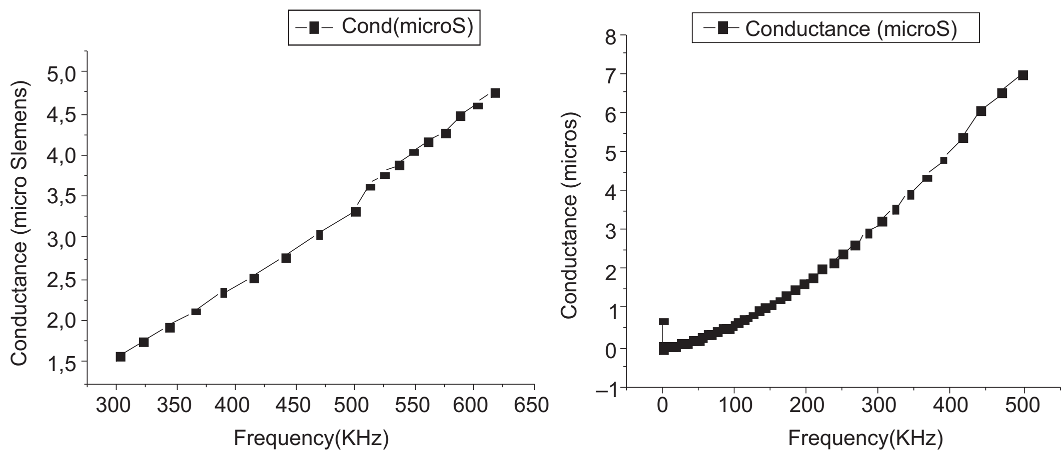


FIG. 4 CONDUCTANCE VALUES OF NEW PIEZO-ELECTRIC COMPOSITE FOR SENSOR APPLICATIONS

SEM OF NANO-PIEZO-COMPOSTE SAMPLE

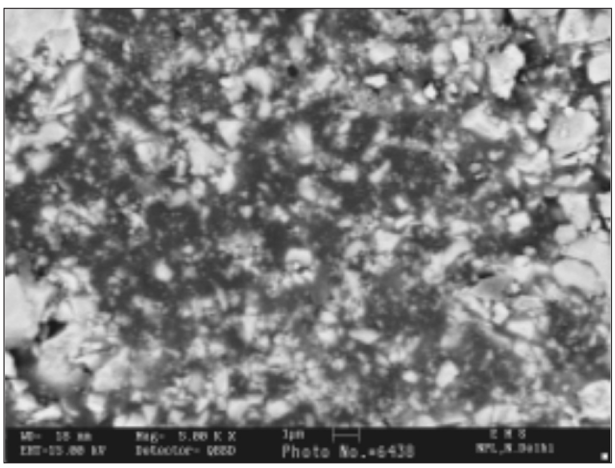


FIG. 5 SEM MICROGRAPH OF NEW PIEZO-ELECTRIC COMPOSITE

7. CONCLUSION

New advances in sensors research have been discussed. Trends in sensor technology, sensor materials and new sensors with biomedical and engineering applications are highlighted. Recent research in sensors at NPL-India has also been mentioned in brief. The research will significantly accelerate the development and use of advanced sensors. The R&D approach aimed at satisfying high payoff opportunities in sensor technology should emphasize the multidisciplinary integration of existing technologies for specific or generic applications. R&D programs that develop sensor materials should focus on selected classes of materials. In view of the diverse range of sensor materials and the high costs of fabrication facilities for many advanced sensor materials, the specific R&D programs set priorities for selected classes of materials rather than attempting to encompass a very broad range of endeavor.

Sensor materials R&D may be pursued further in three main areas, viz., development of processing techniques for existing sensor materials; assessment and development of sensing capabilities in existing materials that have properties not yet exploited for sensor applications; and fundamental investigation of novel sensing approaches, such as using multiple physical responses to a sensing phenomenon and new compositions of matter.

Thus, new sensors will assist in the exploration of new research for new applications [49-51].

REFERENCES

[1] P.R. Hauptman, "Selected Examples of Intelligent (Micro) Sensor Systems: State-of- the-art and Tendencies", *Meas. Sci Technol.* Vol 17, 2006, pp 459-466.

[2] W.H. Ko, "Solid-state Physical Transducers for Biomedical Research", *IEEE Trans. Biomed. Eng.*, Vol. BME-33, No.3, 1986, pp 153-162.

[3] V.R. Singh, "Smart Sensors: Physics, Technology and Applications", *Ind. J. Pure & Appl. Phys.*, Vol. 43, 2005, pp 7-16.

[4] G.L. Cote, "Emerging Biomedical Sensing Technologies and Their Applications", *IEEE Sensors Journal*, Vol. 3, No.3, 2003, pp 251-266.

[5] V.R. Singh, "New Nano-Biomedical Lab-on-Chip Sensors in Nano-Medicine", *Proc. Int Conf on Nanotecnoly in Medicine*, Mumbai, India, Oct 12-13, 2007.

[6] V.R. Singh, "IT in Biomedical Instrumentation", Proc. National Seminar on IT in Emerging Technologies, Patiala, India, April, 2007, pp 1-5.

[7] en.wikipedia.org/wiki/**Lab**-on-a-**chip** - 33k – (lab on a chip)

[8] Geschke, Klank & Telleman, eds.: Microsystem Engineering of Lab-on-a-chip Devices, 1st ed, John Wiley & Sons, 2004.

[9] Stefan Haeberle & Roland Zengerle, "Microfluidic Platforms for Lab-on-a-chip Applications", *Lab Chip*, Vol. 7, 2007, p 1094.

[10] Pampa Dutta, Kasey Hill, Panos G. Datskos & Michael J. Sepaniak, "Development of a Nanomechanical Biosensor for Analysis of Endocrine Disrupting Chemicals", *Lab Chip*, Vol. 7, 2007. p 1184.

[11] Ali Asgar S. Bhagat, Preetha Jothimuthu & Ian Papautsky, "Photodefinable Polydimethylsiloxane (PDMS) for Rapid Lab-on-a-chip Prototyping", *Lab Chip*, Vol. 7, 2007, p 1192.

[12] Jian Gu, Ravi Gupta, Chia-Fu Chou, Qihuo Wei and Frederic Zenhausern, "A Simple Polysilsesquioxane Sealing of Nanofluidic Channels below 10 nm at Room Temperature", *Lab Chip*, Vol. 7, 2007, p 1198.

[13] Larry J. Millet, Matthew E. Stewart, Jonathan V. Sweedler, Ralph G. Nuzzo & Martha U. Gillette, "Microfluidic Devices for Culturing Primary Mammalian Neurons at Low Densities", *Lab Chip*, Vol. 7, 2007, p 987.

[14] Curtis D. Chin, Vincent Linder & Samuel K. Sia, "Lab-on-a-chip Devices for Global Health: Past Studies and Future Opportunities", *Lab Chip*, Vol. 7, 2007, p 41.

[15] http://www.science.org.au/sats2004/images/ryan13.jpg (lab on a chip: DNA and PCR)

[16] M.T. Bohr, "Nanotechnology Goals and Challenges for Electronic Applications", *IEEE Trans on Nanotechnology,*, Vol. 1, No. 1, March, 2002, pp 56-62.

[17] A.Cavalcanti, "Assembly Automation with Evolutionary Nanorobots and Sensor-based Control Applied to Nanomedicine", *IEEE Trans on Nanotechnology*, Vol. 2, No. 2, June, 2003, pp 82 -87.

[18] M. Menon, M., Kyeongjae Cho & D. Srivastava, "Computational Nanotechnology with Carbon Nanotubes and Fullerenes", *Computing in Science & Engineering* [see also *IEEE Computational Science and Engineering]* , Vol. 3, No.4, July-Aug. 2001, pp 42 -55.

[19] C.A.C. Laurvick & B. Singaraju,. "Nanotechnology in Aerospace Systems", IEEE Aerospace and Electronic Systems Magazine, Vol. 18, No.9 , Sept. 2003, pp 18 -22.

[20] G.C. Giakos, K. Meehan & M. Tuma, "Exploitation of Enhanced Fluorescence via Cross-Coupling Principles Toward the Design of an Optical Integrated Thin-film Sensor for Nanotechnology and Biomedical Applications", *IEEE Trans on Instrum. and Meas.*, Vol. 51 No. 5 , Oct. 2002, pp 970 -975.

[21] R.H. Austin, J.O. Tegenfeldt, H., Cao, S.Y Chou & E.C. Cox, "Scanning the Controls: Genomics and Nanotechnology", *IEEE Trans on Nanotechnology*, Vol. 1, No. 1, March, 2002, pp 12-18.

[22] R.C. Merkle, "Nanotechnology: What Will It Mean?", *IEEE Spectrum*, Vol. 38, No. 1, January, 2001, pp 19-21.

[23] D.R. Forrest, "Molecular Nanotechnology", *IEEE Instrum & Meas Magazine,* Vol. 4, No. 3, September 2001, pp 11-20.

[24] R.C. Merkle, "Are We Prepared for the Nnanotechnology Revolution?", *Proceedings of the IEEE*, Vol. 88, No. 1, January 2000, pp 107-108.

[25] K. Pourrezaei, "Application of Nanotechnology to Medicine", *Proc. 23rd Annual Intl IEEE-EMBS Conf. on Engg. in Med & Biology,* Vol. 4, October 25-28, 2001, pp 4087-4088.

[26] T. Kawai, "Nanotechnology Toward DNA Electronics", *Proc. 2001 Int. Conf. on Microprocessors and Nanotechnology*, 31 October-2 Nov. 2001, p 74.

[27] V. Weil, "Zeroing in on Ethical Issues in Nanotechnology", *Proceedings of IEEE*, Vol. 91 No. 11, Nov, 2003, pp 1976 -1979.

[28] U-M Scientists to Develop Nanosensors for Astronauts: http://www.med.umich.edu/opm/newspage/2002/nanosensors.htm.

[29] Biosensors: http://www.u-bourgogne.fr/LPUB/opsub/Nanosensors/capteur.html; http://www.directionsmag.com/article.php?article_id=375.

[30] Molecular Nanotechnology and Medicine http://www.bbriefings.com/businessbriefing/pdf/lifescience2002/publication/phoenix.pdf.

[31] Kanika Singh, "Microfractal Electrodes for EEG Sensing", *Proc. 2nd ASM - IEEE EMBS Conference on Bio, Micro and Nanosystems, S*an Francisco, (USA), January 15-16, 2006.

[32] Kanika Singh, H.H.Kim & K.C.Kim, "Biomems for Osteoproteogerin Detection with Gold NanoparticleS", Proc. MicroTAS, Paris (France),October 7 – 11, 2007.

[33] Kanika Singh & K.C. Kim, "Investigating BioMEMs Techniques for Early Detection of Osteoporosis", Proc. 29th Annual Int. Conf. of the IEEE Engg in Med. and Biology Soc., Lyon, (France), August 23-26, 2007).

[34] Kanika Singh, "A Bone Material Based Sensor", Proc. 26th Ann. Intl. Conf. of the IEEE-EMBS, San Francisco (USA), September 1-5, 2004.

[35] J.W. Choi, K. W. Oh, A. Han, N. Okulan, C. A. Vjayawardhana, C. Lannes, S. Bhansali, K. T. Schlueter, W. R. Heineman, H. B. Halsall, J. H. Nevin, A. J. Helmicki, H. T. Henderson & C. H. Ahn, "Development and Characterization of Microfluidic Devices and Systems for Magnetic Bead-Based Biochemical Detection", Biomedical Microdevices, Vol. 3, No. 3, 2001, pp 191-200.

[36] Fei Su, Krishnendu Chakrabarty & Richard B. Fair, "Microfluidics-Based Biochips: Technology Issues, Implementation Platforms, and Design-Automation Challenges", IEEE Tans on Computer-aided Design of Integrated Circuits and Systems, Vol. 25, No. 2, February, 2006.

[37] J.W. Choi, K.W. Oh, J.H. Thomas, W.R. Heineman, H.B. Halsall, J.H. Nevin, A.J. Helmicki, H.T. Henderson, and C.H. Ahn, "An Integrated Microfluidic Biochemical Detection System for Protein Analysis with Magnetic Bead-Based Sampling Capabilities," Lab on a Chip, Vol. 2, No. 1, 2002, pp 27-30.

[38] K. Singh, S.H. Lee, and K.C. Kim, "Osteoporosis: New Biomedical Engineering Aspects", J. of Mechanical Science & Technology (KSME Int. J), Vol. 20, No. 12, 2006, pp. 2265-2283.

[39] V.R. Singh, "Recent Advances and Future Directions in Ultrasonic Diagnosis of Tissues", in book, 'Advances in Medical Diagnostic Techniques and Procedures' (eds. M.Singh and S. Radhakrishnan), Anamaya Publishers, New Delhi, 2004, pp 7-11

[40] M. Singh, H.R. Singh, A. Kumar, V. R. Singh, K. Pant & P. Sethi, "Development of Tissue Equivalent Phantoms for Biomedical Ultrasonic Applications", IJBET (Intl. J. Biomedical Engineering and Technology), 2007 (in press).

[41] V.R. Singh, "A New Semiconductor Lithotripsy Sensor", Proc. 26th Ann. Int. Conf. IEEE-EMBS (EMBC-2004), San Francisco, California, USA, September 1-5, 2004, p 568.

[42] Fundamentals of BioMEMS and Medical Microdevices: *SciTech Book News,* June, 2006.

[43] www.ucsf.edu/desailab/papers/2004/Neurotransmitter%20analog%20tethered%20to%20a%20silicon%20platform%20fo.

[44] Expanding the Vision of Sensor Materials: http://www.nap.edu/catalog/4782.html.

[45] www.octavetech.com 2006.

[46] www.kodak.com.

[47] www.ndma.us.
[48] http://www.cornell.edu.
[49] www.physorg.com/news100970000.html.
[50] www.bioisrael.com/index.asp?NewsPaperId=41.
[51] Billy Smith, Space Sensor, www.technologyreview.com.
[52] Silicon Nanowire-Based Biochemical Sensors: www-bsac.eecs.berkeley.edu.
[53] findarticles.com/p/articles/mi_m0QLT/is_2006_June/ai_n16675739 - 26k.

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

DESIGN OF EMBEDDED SYSTEM FOR SAFETY APPLICATION

P. SWAMINATHAN*

ABSTRACT

Physically and functionally distributed embedded systems are increasingly used for on-line supervision and control of process plants. The information is transmitted to control room through plant area network. The design of hardware and application software should take into account that any random single failure should ensure safe shut-down of the process plant. For ensuring availability, fault tolerant architecture is recommended. While developing application software, waterfall model is recommended. MISRA-C guidelines need to be followed during the coding stage. The system shall be robust enough to withstand normal electromagnetic interference ,which is expected at site.

1. INTRODUCTION

Microprocessor or Micro-controller based instrumentation systems, which are also called embedded systems, are increasingly used for supervision and control of process plants such as Thermal power plant, Nuclear power plant, Chemical plant, Cement plant, Steel plant etc. In every process plant, large number of process signals are processed by physically and functionally distributed embedded systems. The processed information is transmitted to control room through local area network. The information is managed by expert system and displayed to Plant operator. For safety application, the specified reliability is achieved through fault tolerant architecture. For reliable operation of embedded system, it is very important to concentrate on hardware design and software design. Safety analysis of embedded systems need to be carried out to ensure that hardware or software failure should be detected and shall ensure safe shut down state of process plant.

2. EVOLUTION OF MICROPROCESSORS

Microprocessors made possible the advent of the microcomputer. Before this, electronic CPUs were typically made from bulky discrete switching devices (and later small-scale integrated

*Outstanding Scientist, Indira Gandhi Centre for Atomic Research Kalpakkam, India 603 102;E-Mail: swamy@igcar.gov.in

circuits) containing the equivalent of only a few transistors. By integrating the processor onto one or a very few large-scale integrated circuit packages (containing the equivalent of thousands or millions of discrete transistors), the cost of processor power was greatly reduced. Since the advent of the IC in the mid-1970s, microprocessor has become the most prevalent implementation of the CPU, nearly completely replacing all other forms. One or more microprocessors typically serve as central processing unit (CPU) in a computer system or handheld device. The evolution of microprocessors has been known to follow Moore's Law which suggests that *complexity of an integrated circuit, with respect to minimum component cost, doubles every 24 months*. This dictum has generally proven true since the early 1970s. From their humble beginnings as the drivers for calculators, the continued increase in power has led to the dominance of microprocessors over every other form of computer; every system from the largest mainframes to the smallest handheld computers now uses microprocessor at its core. The microprocessor is a central processing unit (CPU) on single chip. To create a viable computer requires memory components, interface components, timing and control circuits, a power supply, and a cabinet or other enclosures. The CPU card contains the microprocessor and its associated control circuitry. A block diagram of a typical computer system is shown in Fig.1.

2.1 Choice of Back Plane or Bus

Back plane or bus is a set of communication system through which CPU dialogues with memory and Input/output systems. Normally CPU is made of standard Intel microprocessors (8085, 8086) or Motorola micro processors (68000, 68020), or Intel micro controllers (8051, 80251) or Motorola micro controllers (683XX). The software is normally stored in Read only memory (ROM). Necessary dynamic data is stored Random Access Read/Write Memory (RAM). Microprocessor reads one by one the instruction from ROM and executes them. In this process, the necessary data is stored in RAM. The calculated results are written back in RAM. For reading the instruction or data from memory, first CPU will put the required address information in the address bus. The required service, namely, read command is also put in the command lines. CPU also puts Master Sync signal in the bus in the case of asynchronous bus. Memory Unit will put the addressed data in the data lines. In the case of Asynchronous bus, memory unit will also put "Ack" signal. On receiving "Ack" signal, CPU will read the data from the data lines. The cycle is completed.

In the case of write cycle, CPU will put the required address in the address lines. Data to be written is put in the data lines. CPU then asserts MSYN signal. Memory will take the data from the data lines and write it in the required location. Memory Unit will assert slave sync signal. CPU, will drop MSYN signal, thus completing the bus cycle. Similar Read/Write operation takes place between CPU and Input/Output system. Motorola microprocessors use Asynchronous bus. For Intel microprocessors, synchronous bus is used. Here Read or Write cycle is completed within the specified clock cycles. For safety application, asynchronous bus is recommended.

2.2 Design of CPU board

Normally CPU board consists of the following:

- Microprocessor or micro controller

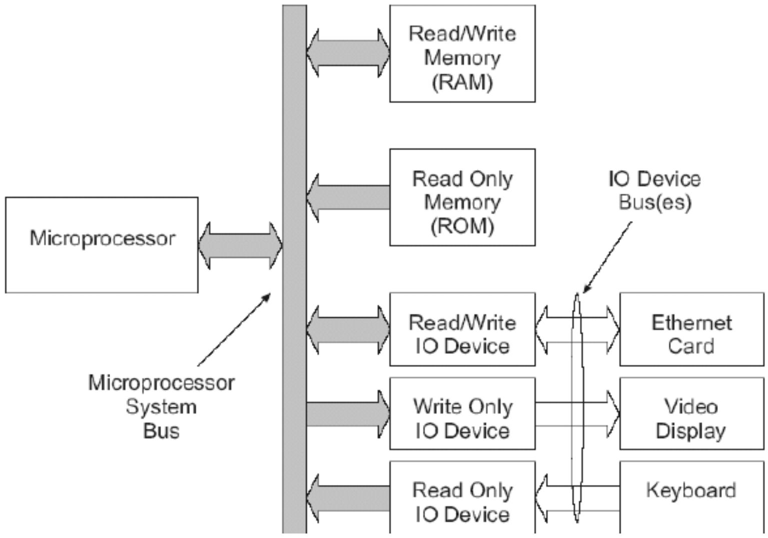


FIG. 1 THE BASIC COMPUTER SYSTEM HARDWARE STRUCTURE

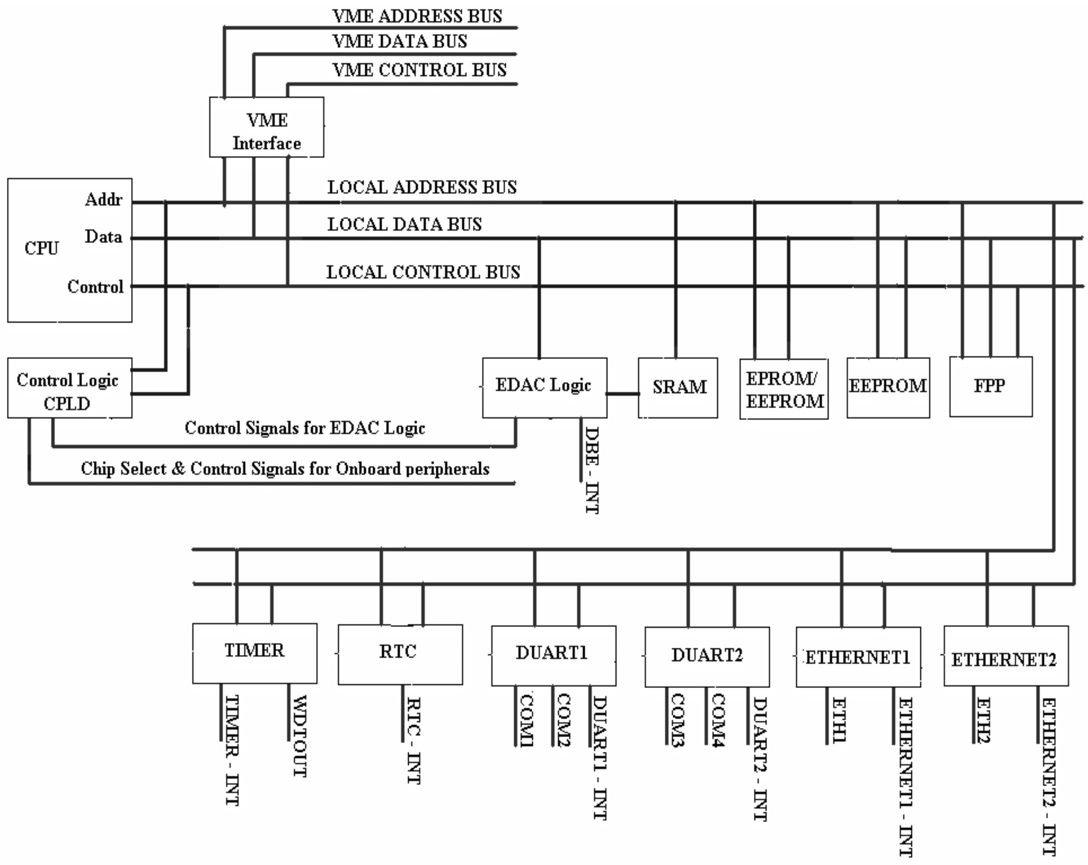
- ROM & RAM
- Interconnection bus between CPU and memory
- Bus interface logic
- Watch dog timer
- Clock circuit

Typical block diagram of 68020 based CPU card is given below:

RAM memory is prone to failure. It is necessary to detect single bit memory failure and correct the same. At the same time two bit memory failure shall be detected and CPU shall be informed through interrupt. Standard Error detection and correction (EDAC) chip is available in the market. This is integrated in the CPU card. Watchdog timer shall be refreshed periodically by software. Otherwise it will be decremented by clock. When watchdog timer reaches “zero” then, on-board mounted relay can be made to deenergise. The change of state of relay contact can be used to take necessary remedial action. Normally whenever double bit memory error occurs or if slave-Ack is not received in the back plane (bus) or if the microprocessor hangs, then the watchdog will time-out.

2.3 Design of Analog Input Card

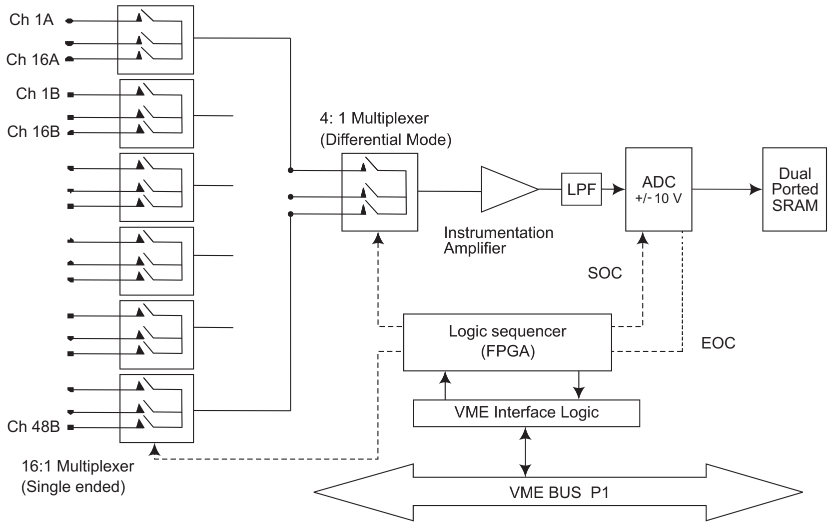
Signals from process sensors like thermocouple, RTD, flow meter, pressure transducer, level sensor, etc. are first signal conditioned (amplified, isolated and filtered) and then received by Analog Input Card. If the process sensor is located at a long distance, then current signal (4-20 mA) is used. Current signal is less sensitive to electrostatic/electromagnetic noises. It is always preferable to use isolation amplifier between the process sensors and Analog to Digital Converter. This will eliminate circulating ground loop currents.



VME Bus Based Motorola 68020 CPU card with EDAC

Analog input card consists of Multiplexer, Analog to Digital converter, on-board memory and control logic. The block diagram of typical analog input card is given below:

CPU initiates the scanning by issuing the necessary command to the sequencer. The address input to input multiplexer is incremented in steps by the sequencer. The multiplexed input signal is analog to digital converted and stored on the on-board memory. Normally a 12 bit or 16 bit, successive approximation type Analog to Digital Converter (ADC) is used. In situations where 50Hz pick up from nearby power lines is dominant, integrating type ADCs may be used for reducing the effects of this noise. Each Analog input card is provided with on-board calibration sources, which are, in turn, connected to the input multiplexer. Diagnostic software will analyze the signal level from calibration source. This will enable to detect drift in amplifier or error in ADC. Normally scanning rate shall be greater than double the frequency of process signals. To minimize the effect of noise, each sample will be compared with previous sample. If the difference is greater than the allowed limit, then the present sample is discarded. Similarly, to overcome fluctuating noise, average of ten or fifteen samples is used instead of the sample itself.



BLOCK DIAGRAM OF ANALOG INPUT CARD

2.4 Design of Digital Input Card

Digital signals from the process plant are received either as electrical signal (OV or (5V/12V/24V/32V/48V) or as voltage free relay contact.

To eliminate ground loop problem, opto coupler is used for every digital input signal. CPU periodically reads the status of the digital inputs and analyses them. Some opto-couplers may fail in conducting or non-conducting state. State-of-the art digital input cards are provided with force ‘O’ and force ‘1’ option. This is periodically carried out by on-line diagnostics to detect the failed opto-coupler. Each digital input card houses 8 or 16 or 32 or 48 input channels.

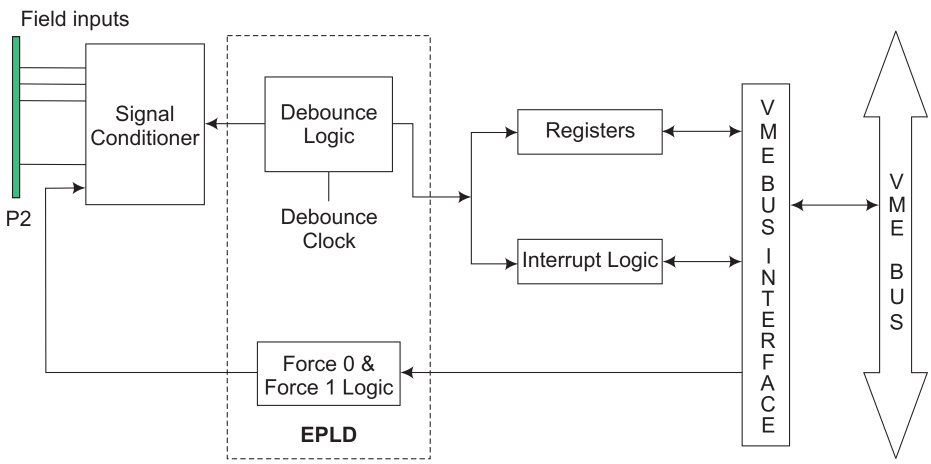
2.5 Design of Analog & Digital Output Card

Decision taken by the embedded system is communicated to the plant/equipment through digital output card and Analog output card. A typical block diagram of Digital output card is given below:

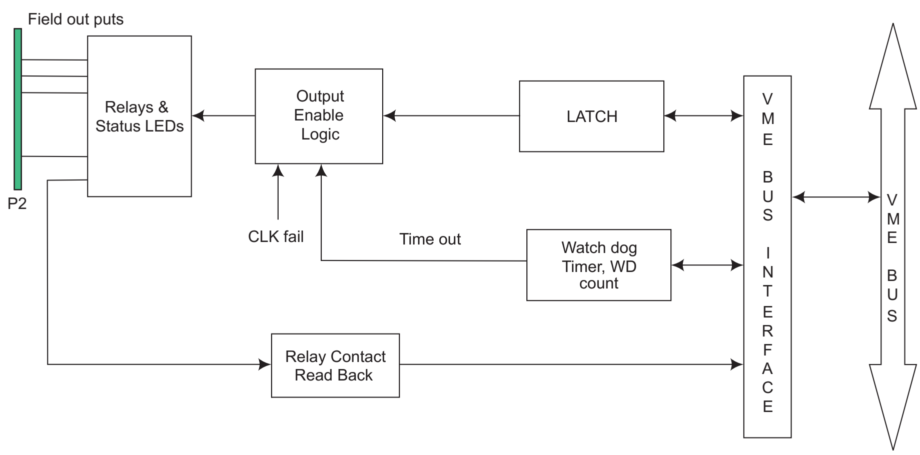
Digital signals are communicated to the plant as voltage free relay contacts or open collector transistor outputs.

In state-of-the art digital output card, there is provision to read back the status of the output relay. Each relay is provided with two contacts. One contact is wired to the plant white the other contact is read back by the CPU. Each digital output card will house 8 or 16 or 32 output channels. The status of each digital output is available through LED lamp .For safety application, the card is designed such that software periodically loads the output value in the on-board latch. If microprocessor hangs or if software enters endless loop due to memory fault, then on-board watch dog timer will time out. This in turn will reset the on-board latch. The digital outputs from latch is wired such that process safe state is ensured when latch is reset by watchdog timer.

Block diagram of Analog output card is given below:



BLOCK DIAGRAM OF DIGITAL INPUT

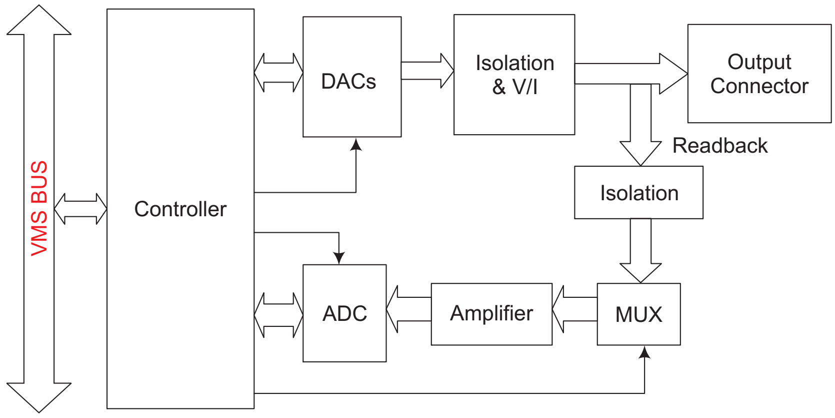


BLOCK DIAGRAM OF RELAY OUTPUT

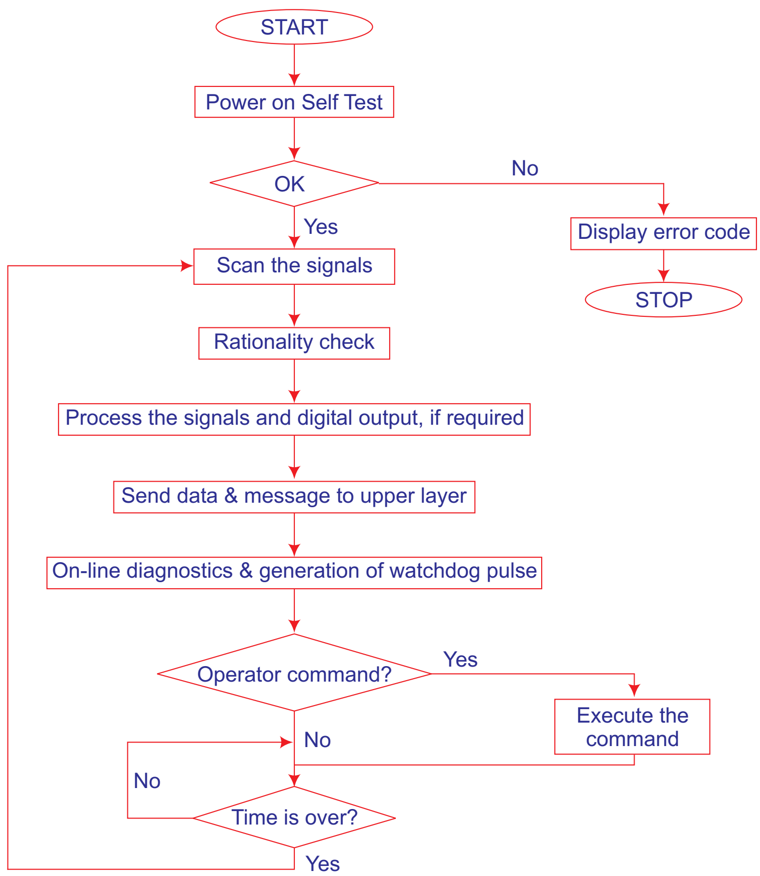
Analog output signal is available as 4 to 20 mA or as 0 to 5 or 10 V. For transmitting analog signal over long distance, current mode is selected. In analog output card, 12 bit DAC is normally used to convert digital signal to analog signal. Normally each analog output card will house four analog output channels. If the microprocessor hangs, there is provision to hold on to the recently sent analog output value,such that safe condition of the plant is ensured. There is also provision to read-back the output values for diagnostic purposes.

3. SOFTWARE ARCHITECTURE OF EMBEDDED SYSTEM

Commercially available operating systems consists of scheduler, memory management, I/O management etc. In embedded application same task is executed at fixed time interval.The listing of commercially available operating system is also not made available for verification. Hence for



BLOCK DIAGRAM OF ANALOG OUTPUT CARD



safety application, usage of commercially available operating system is not recommended. The application software normally will consist of power on diagnostics, scanning software, signal processing software, communication software and diagnostics .The arrangement is shown below:

During powering the system, power-on reset is generated. This in turn gives control to power on self test. During this phase all parts of hardware will be checked. If any error is detected then corresponding error code is displayed and system stops. Otherwise control is given to scanning software. During rationality check, the process values will be compared with absolute low and high of process conditions. If process signal value is not within specified validation limits, the sample is rejected. To minimize the 50Hz noise, average value of the scanned process samples is taken for further processing. After carrying out the required processing, necessary analog/digital outputs are delivered to the plant. The information about the value of the process signal and generated messages are transmitted to upper layer for display to plant operator. On-line diagnostics periodically checks all parts of the hardware. If any error is detected, corresponding error code is displayed in the front panel and system stops. The value of analog or digital output is forced to fail-safe state with respect to the process plant. Provision is also made in the software such that plant operator will be able to edit software threshold through Dump terminal.

After the specified time interval, control is given back to scanning software once again.

3.1 Process Models

Waterfall model

The waterfall model is a sequential software development model (a process for the creation of software) in which development is seen as flowing steadily downwards (like a waterfall) through the phases of requirements analysis, design, implementation, testing (validation), integration, and maintenance.

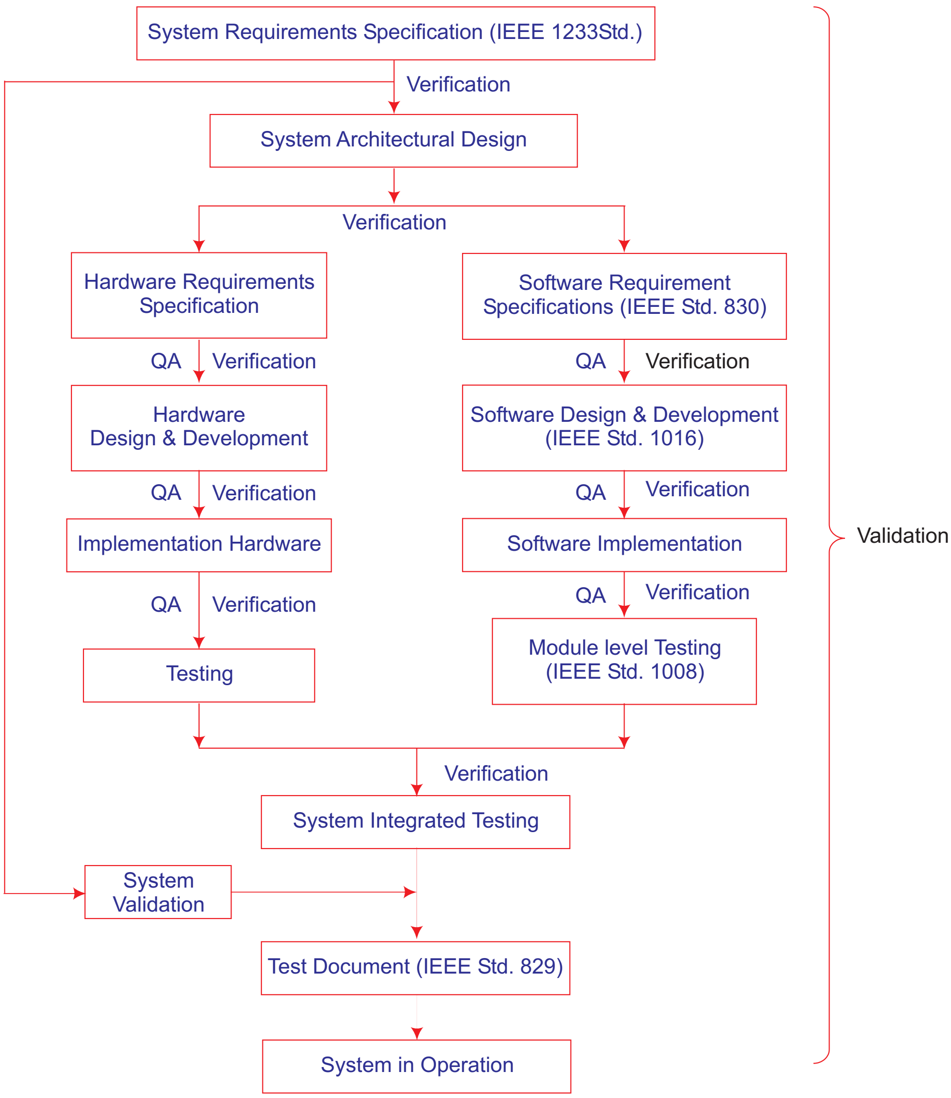
Waterfall model is used in the development of embedded system for safety application, where requirement is well understood. Relevant IEEE standards are to be followed at every life cycle stage of development of embedded system as shown below:

Quality Assurance (Q.A.) : QA process at every life cycle involves cheking the conformance of the product to specified standards.

Verification : Verification involves checking the conformance of product at every life cycle stage to requirement specification.

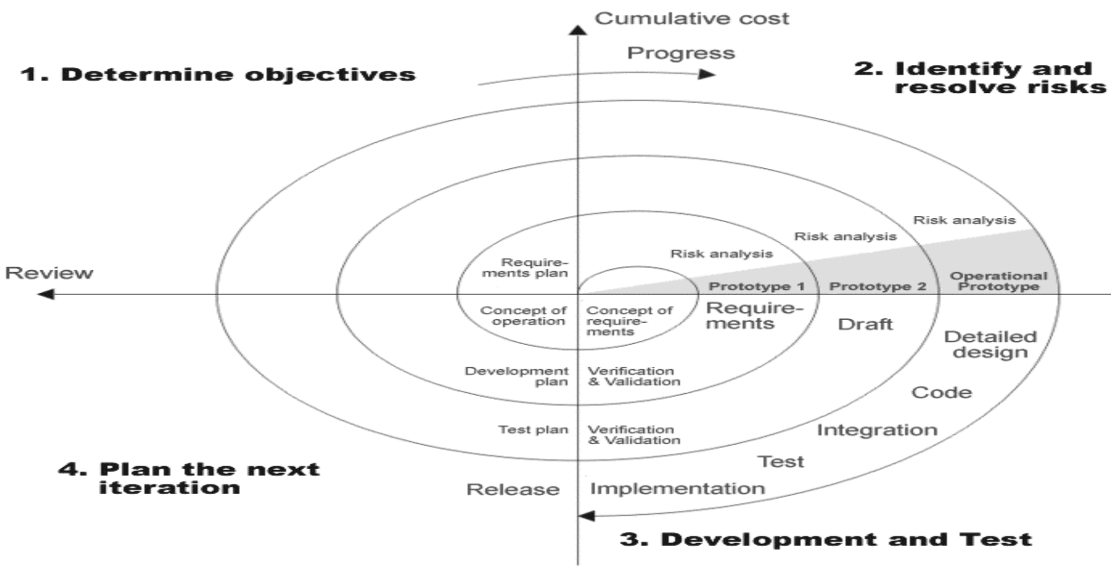
Validation : Validation involves checking the final system for compliance to requirement specification of the end-user. There is need to carryout independent verification and validation at every life cycle stage of development of embedded system. FORMAL method is also recommended in modelling the requirement specification of embedded system. Either Z or B language is used in modelling the specification. It is very important to acquire necessary domain knowledge of the process for finalising the requirement specification. Any error in the requirement will sail through the final stage and it will be very costly to rectify the error.

Prototype model: The prototyping model is a software development process that begins with requirements collection, followed by prototyping and user evaluation. Often the end users may not be able to provide a complete set of application objectives, detailed input, processing, or output



requirements in the initial stage. After the user evaluation, another prototype will be built based on feedback from users, and again the cycle returns to customer evaluation. The cycle starts by listening to the user, followed by building or revising a mock-up, and letting the user test the mock-up, then back. There is now a new generation of tools called Application Simulation Software which help quickly simulate application before their development. Normally prototype model is used for developing Human Machine Interface System.

Spiral model: The spiral model, also known as the spiral lifecycle model, is a systems development method (SDM). This model of development combines the features of the prototyping model and the waterfall model. The spiral model is intended for large, expensive, and complicated projects like Distributed Digital Control System for large process Plant.



The steps in the spiral model can be generalized as follows:

1. The new system requirements are defined in as much detail as possible. This usually involves interviewing a number of users representing all the external or internal users and other aspects of the existing system.
2. A preliminary design is created for the new system.
3. A first prototype of the new system is constructed from the preliminary design. This is usually a scaled-down system, and represents an approximation of the characteristics of the final product.
4. A second prototype is evolved by a fourfold procedure:
    1. evaluating the first prototype in terms of its strengths, weaknesses, and risks;
    2. defining the requirements of the second prototype;
    3. planning and designing the second prototype;
    4. constructing and testing the second prototype.
5. At the customer's option, the entire project can be aborted if the risk is deemed too great. Risk factors might involve development cost overruns, operating-cost miscalculation, or any other factor that could, in the customer's judgment, result in a less-than-satisfactory final product.
6. The existing prototype is evaluated in the same manner as was the previous prototype, and, if necessary, another prototype is developed from it according to the fourfold procedure outlined above.
7. The preceding steps are iterated until the customer is satisfied that the refined prototype represents the final product desired.
8. The final system is constructed, based on the refined prototype.

9. The final system is thoroughly evaluated and tested. Routine maintenance is carried out on a continuing basis to prevent large-scale failures and to minimize downtime.

3.2 System Requirements Specification (SyRS)(IEEE Std.1233)

A System Requirements Specification is a document that communicates the requirements of the customer to the technical community who will specify and build the system. The collection of requirements that constitutes the specification and its representation acts as the bridge between the two groups and must be understandable by both the customer and the technical community. The purpose of the SyRS is to provide a black-box description of what the system should do, in terms of the systems interactions or interfaces with its external environment. The SyRS should completely describe all inputs, outputs, and required relationships between inputs and outputs.

The recommended uses of the SyRS, which vary as the development cycle progresses, are as follows:

a) During systems design, requirements are allocated to subsystems, hardware, software, operations, and other major components of the system.
b) The SyRS is utilized in constructing the resulting system. The SyRS is also used to write appropriate system verification plans. If the system contains hardware and software, then the hardware test plan and software test plan are also generated from the system requirements.
c) During the implementation phase, test procedures will be defined from the SyRS.
d) During the validation process, validation procedures based on the SyRS are used to provide the customer with a basis for acceptance of the system

One possible outline for SyRS is shown below:

1.0 Introduction

1.1 System purpose

1.2 System scope

1.3 Definitions, acronyms, and abbreviations

1.4 References

1.5 System overview

2.0 General system description

2.1 System context

2.2 System modes and states

2.3 Major system capabilities

2.4 Major system conditions

2.5 Major system constraints

2.6 User characteristics

2.7 Assumptions and dependencies

2.8 Operational scenarios

3.0 System capabilities, conditions, and constraints

NOTE- System behavior, exception handling, manufacturability, and deployment should be covered under each capability, condition, and constraint.

3.1 Physical

3.1.1 Construction

3.1.2 Durability

3.1.3 Adaptability

3.1.4 Environmental conditions

3.2 System performance characteristics

3.3 System security

3.4 Information management

3.5 System operations

3.5.1 System human factors

3.5.2 System maintainability

3.5.3 System reliability

3.6 Policy and regulation

3.7 System life cycle sustainment

4.0 System interfaces

4.1 Interfaces with Field Inputs

4.2 Interfaces with Actuators

4.3 Interfaces with other systems

3.3 System Architecture Design (SAD):

This document consists of details of functional requirements, definition of a collection of subsystems, hardware as well as software that will form the system and their interfaces with each other as well as with the external world. It also covers allocation of functions to both hardware and software. This follows the System Requirements Specification (SyRS) document. This document mainly explains how the main system will be decomposed into several subsystems physically & functionally and the rationality behind it.

One possible outline for SAD is shown below.

1.0 Overview

2.0 Introduction

2.1 System purpose

2.2 System scope

2.3 System Context

3.0 Functional requirements specifications

3.1 Requirement specification of function 1

3.2 Requirement specification of function 2

….

3.n Requirement specification of function n

4.0 System overall architecture

4.1 Architecture of partition 1

4.1.1 Functional requirements

4.1.2 Non-functional requirements

4.1.3. Hardware Architecture

4.1.4 Software architecture

4.2 Architecture of partition 2

4.2.1 Functional requirements

4.2.2 Non-functional requirements

4.2.3 Hardware Architecture

4.2.4 Software architecture

……..

4.n Architecture of partition n

4.n.1 Functional requirements

4.n.2 Non-functional requirements

4.n.3 Hardware architecture

4.n.4 Software architecture

3.4 Software Requirements Specification (SRS):

The SRS is a specification for a particular software product, program, or set of programs that performs certain functions in a specific environment.

The basic issues that the SRS shall address are the following:

a) ***Functionality***.

What is the software supposed to do?

b) ***External interfaces***.

How does the software interact with people, the systems hardware, other hardware, and other software?

c) ***Performance***.

What is the speed, availability, response time, recovery time of various software functions, etc.?

d) ***Attributes***.

What are the portability, correctness, maintainability, security, etc. considerations?

e) ***Design constraints imposed on an implementation***.

Are there any required standards in effect, implementation language, policies for database integrity, resource limits, operating environment(s) etc.?

Note : The SRS should avoid placing either design or project requirements in the SRS.

An SRS should be

a) Correct
b) Unambiguous
c) Complete
d) Consistent
e) Ranked for importance and/or stability
f) Verifiable
g) Modifiable
h) Traceable

One possible outline for SRS is shown below.

1. Introduction
- *1.1 Purpose*
- *1.2 Scope*
- *1.3 Definitions, acronyms, and abbreviations*
- *1.4 References*
- *1.5 Overview*

2. Overall description
- *2.1 Product perspective*
- *2.2 Product functions*
- *2.3 User characteristics*
- *2.4 Constraints*

3. Specific requirements
- *3.1 External interface requirements*
  - *3.1.1 User interfaces*
  - *3.1.2 Hardware interfaces*
  - *3.1.3 Software interfaces*
  - *3.1.4 Communications interfaces*
- *3.2 System features*
  - *3.2.1 System Feature 1*
    - *3.2.1.1 Introduction/Purpose of feature*
    - *3.2.1.2 Stimulus/Response sequence*
    - *3.2.1.3 Associated functional requirements*
      - *3.2.1.3.1 Functional requirement 1*
      - *3.2.1.3.n Functional requirement n*
  - *3.2.2 System feature 2*
    - ………
  - *3.2.m System feature m*

3.3 Performance requirements

3.4 Design constraints

3.5 Software system attributes

Other requirements

Appendixes and Index

3.5 Software Design Description (SDD)

An SDD is a representation or model of the software system to be created. The model should provide the precise design information needed for planning, analysis, and implementation of the software system. It should represent a partitioning of the system into design entities and describe the important properties and relationships among those entities.

The design description model used to represent a software system can be expressed as a collection of design entities, each possessing properties and relationships. To simplify the model, the properties and relationships of each design entity are described by a standard set of attributes. The design information needs of project members are satisfied through identification of the entities and their associated attributes. A design description is complete when the attributes have been specified for all the entities.

A *design entity* is an element (component) of a design that is structurally and functionally distinct from other elements and that is separately named and referenced. A *design entity attribute* is a named characteristic or property of a design entity. It provides a statement of fact about the entity.

Purpose of an SDD

The SDD shows how the software system will be structured to satisfy the requirements identified in the software requirements specification . It is a translation of requirements into a description of the software structure, software components, interfaces, and data necessary for the implementation phase. In essence, the SDD becomes a detailed blueprint for the implementation activity. In a complete SDD, each requirement must be traceable to one or more design entities.

One possible outline for SDD is shown below.

1. Introduction

1.1 Purpose

1.2 Scope

1.3 Definitions and acronyms

2. References

3. Decomposition description

3.1 Module decomposition

3.1.1. Module 1 decomposition

3.1.3 Module 2 decomposition

3.1.n Module n decomposition

3.2 Concurrent process decomposition

3.2.1 Process 1 description

3.2.2 Process 2 description

…….

3.2.n Process n description

3.3 Data decomposition

3.3.1 Data entity 1 description

3.3.2 Data entity 2 description

3.3.n Data entity n description

4. Dependency description

4.1 Intermodule dependencies

4.2 Interprocess dependencies

4.3 Data dependencies

5. Interface description

5.1 Module interface

5.1.1 Module 1 description

5.1.2 Module 2 description

……..

5.1.n Module n description

5.2 Process interface

5.2.1. Process 1 description

5.2.2 Process 2 description

……..

5.2.n Process n description

6. Detailed design

6.1 Module detail design

6.1.1 Module1 detail

6.1.2 Module 2 detail

…..

6.1.n Module n detail

6.2 Data detail design

6.2.1 Data entity 1 detail

6.2.3 Data entity 2 detail

……..

6.2.n Data entity n detail

3.6 Software Test Planning Documentation

The test plan prescribes the scope, approach, resources, and schedule of the testing activities. It identifies the items to be tested, the features to be tested, the testing tasks to be performed, the personnel responsible for each task, and the risks associated with the plan.

A test plan shall have the following structure:

a) *Test plan identifier*
b) *Introduction*
c) *Test items*
d) *Features to be tested*
e) *Features not to be tested*
f) *Approach*
g) *Item pass/fail criteria*
h) *Suspension criteria and resumption requirements*
i) *Test deliverables*
j) *Testing tasks*
k) *Environmental needs*
l) *Responsibilities*
m) *Staffing and training needs*
n) *Schedule*
o) *Risks and contingencies*
p) *Approvals.*

3.7 Software Configuration Management (SCM)

Software configuration management (SCM) is a "set of activities designed to control change by identifying the work products that are likely to change, establishing relationships among them, defining mechanisms for managing different versions of these work products, controlling the changes imposed, and auditing and reporting on the changes made." In other words, SCM is a methodology to control and manage a software development project.

The goals of SCM are generally:

Configuration Identification- What code are we working with?

Configuration Control- Controlling the release of a product and its changes.

Status Accounting- Recording and reporting the status of components.

Review- Ensuring completeness and consistency among components.

Build Management- Managing the process and tools used for builds.

Process Management- Ensuring adherence to the organization's development process.

Environment Management- Managing the software and hardware that host our system.

Teamwork- Facilitate team interactions related to the process.

Defect Tracking- making sure every defect has traceability back to the source

3.8 Coding Guidelines

Most of the applications software is developed in 'C' language. For embedded systems, it is highly recommended that it is MISRA-C guidelines(Motor Industry Software Reliability Association) are followed:

Important guidelines are

- The goto statement shall not be used
- Pointer arithmetic should not be used
- The right hand operator of a && or ïï operator shall not contain side effects
- The basic types of char, int, short, long, float and double should not be used, but specific-length equivalents should be typedefined for the specific compiler and these type names used in the code
- Dynamic heap memory allocation shall not be used
- The time handling functions of library TIME.H shall not be used
- Floating point variables shall not be tested for exact equality or inequality
- All automatic variables shall have been assigned a value before being used
- Assembly language code shall not be embedded within normal C code
- Functions shall not call themselves directly or indirectly
- A function should have a single point of exit

4.0 BURNING IN ROM/EPROM

Software development for embedded systems involves a set of sequential activities using the following tools. Compiler, Linker and Locator are normally integrated into single development tool.

Compiler:

Translates C source files to machine instructions, produces object files. Normally the development platform will be different from that of the target embedded system's platform hence cross compiler is used to compile the source code to object code corresponding to the target platform.

Linker:

Resolves addresses of different object files in terms of offsets from base address of program. It integrates multiple object files of the same project into a single executable file.

Locator:

Determines final memory image of program. Locator includes some mechanism for programmer to determine placement in memory – Some parts need to be in RAM, others in ROM.

Programmer:

This binary file shall be fused/burnt into an EPROM/PROM chip using the EPROM programmer with the supporting programming software tool. If EPROM is split into odd and even address banks in the hardware, the binary file shall also be split into odd and even and then shall be programmed into the respective EPROM chips accordingly. The start address of the program shall also be supplied to the EPROM programming software.

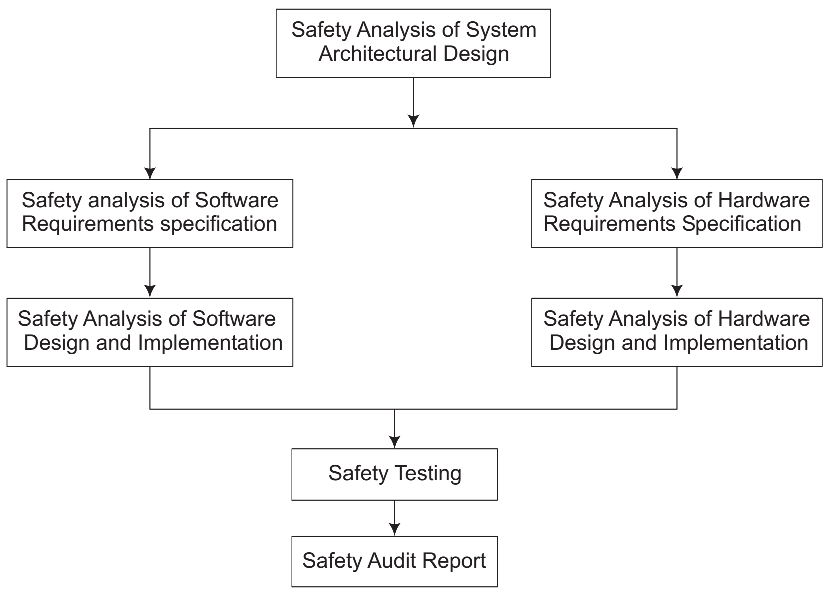
Guidelines during Operation and Maintenance (O&M) phase

During development phase, development team, QA team, verification & validation team etc will work together in realising reliable embedded systems for safety application. During O&M phase, the following guidelines shall be followed:

a) All the hardware and software documents, generated during development phase, shall be made available during O&M phase
b) Separate personnel shall be identified, trained and qualified for O&M phase
c) All the proposed hardware and software changes shall be analysed for impact on the safety of the Plant. The proposed changes shall be introduced in representative system. After Verification and Validation, changes shall be ported to target system.
d) All the relevant documents shall be updated to reflect the changes carried out.
e) A Log file shall be maintained at the site consisting of history of changes carried out.
f) The network of embedded systems shall be physically disconnected from the Internet. This will protect the embedded systems from virus and worms.

5.0 SAFETY ANALYSIS OF EMBEDDED SYSTEMS

For safety application safety analysis need to be carried out at every development life cycle stage of embedded system as shown below:



Safety Analysis of System Architectural Design

System architectural design shall be analysed in detail to establish that all system level safety requirements are carried into the system design and allocated to software or hardware or a combination of them. The system level hazards shall be traced through the system architecture to show that hazardous states cannot occur. The design shall be shown to be fail-safe taking into account the various failure modes of hardware and software.

Safety Analysis of Software Requirements Specification

Analysis of software requirements specification shall be carried out to establish that it incorporates all system level safety requirements allocated to software and they are clearly described, and are testable. These should include the on-line (in service) safety test requirements, mandated by the technical specifications of the plant and to be implemented in software.

Safety Analysis of Hardware Requirements Specification

Analysis of hardware requirements specification shall be carried out to establish that it incorporates all system level safety requirements allocated to hardware and they are clearly described, and are testable. These should include the on-line (in service) safety test requirements, mandated by the technical specifications of the plant and to be implemented in hardware.

Safety Analysis of Software Design and Implementation

Software design and implementation shall be analysed in detail to establish that software design and implementation incorporates all safety requirements given in Software Requirements Specifications. Analysis should establish that software satisfies all safety requirements, does not cause any unsafe action under any operating condition and allows on-line tests to be carried out without compromising the performance of safety functions. The design of the software shall be shown to handle hardware failures gracefully without causing unsafe conditions in the plant. Catastrophic failure of the software (i.e. When it is not able to perform the intended function) should be shown to lead to fail safe outputs from the Computer-based System(i.e safe conditions in the plant).

Safety Analysis of Hardware Design

Hardware design shall be analysed in detail to establish that hardware incorporates all safety requirements given in Hardware Requirements Specifications. Analysis should establish that hardware satisfies all safety requirements, does not cause any unsafe action under any operating condition and allows on-line tests to be carried out without compromising the performance of safety functions. Failure of the hardware should be shown to lead to fail safe outputs from the Computer-based System(i.e. safe conditions in the plant).

Safety Testing

The system shall be subjected to tests that will confirm its overall safe behavior. This is the final demonstration safety. The testing shall be done to check that

1. All safety requirements are correctly implemented
2. System behavior is failsafe.
3. All on-line tests can be conducted without compromising the performance of safety functions.

Safety Audit

The Safety Audit shall be carried out to verify the safety analysis and establish that safety requirements have been implemented. The Safety Audit shall cover the following phases of safety life cycle:

- System Architectural Design

- Software Requirements
- Hardware Requirements
- Software Design and Implementation
- Hardware Design
- Safety Testing

The safety analysis of overall architecture shall address the following failure of subsystems.

- Non availability of power supply
- Sensor fault
- Sensor over range
- Noise in input signal
- Process signal fluctuation
- Failure of microprocessor
- Failure of memory
- Failure of acknowledgement signal in the bus
- Failure of multiplexer, amplifier, analog to digital converter and sequencer in analog input card
- Failure of optical isolator in digital input card
- Failure of latch and relay in digital output card
- Endless loop in application software
- Irrational data entry for changing software threshold
- Failure of data server and message sensor and graphic user terminals

A general fault tree shall be constructed . The design shall ensure that any postulated fault will result in ordering digital output, which in turn ensures safe state of the nuclear reactor.

6.0 RELIABILITY ANALYSIS OF EMBEDDED SYSTEM

Faults in embedded systems can be classified as safe fault and unsafe fault. If the fault results in ordering analog or digital outputs for placing the process in safe state, then the fault is classified as safe faults. The failure of power supply of the embedded system is example of safe fault. On the other hand, if there is demand for shut down of the plant, and if shut down order is not delivered, then the fault is defined as unsafe fault. Again the unsafe fault is further classified as on-line detectable unsafe faults and on-line undetectable unsafe faults. In embedded system, on-line diagnostics will detect unsafe fault such as drift in signal amplifier, ADC fault, memory fault, failure of opto coupler in digital input/output cards, failure of ACK signal etc.

If any fault is detected, on-line diagnostics will not refresh watch dog timer. This will result in time out of watch dog timer thus resulting delivery of shutdown order to the process. There are still unsafe faults which can not be detected such as failure in watch dog circuit, welding of relay contacts in digital output card etc. The safe fault or failure rate is represented as $\lambda_{s.}$ The failure rate of unsafe faults which can be detected by on-line diagnostics is represented as $\lambda_{u1.}$ The failure rate of unsafe faults which can not be detected by on-line diagnostics is represented as $\lambda_{u2.}$

6.1 Safe Failures & Unsafe Failures

The total failure rate in the system can be divided into Safe and unsafe (dangerous) failures.

$$\lambda = \lambda_S + \lambda u1 + \lambda u2$$

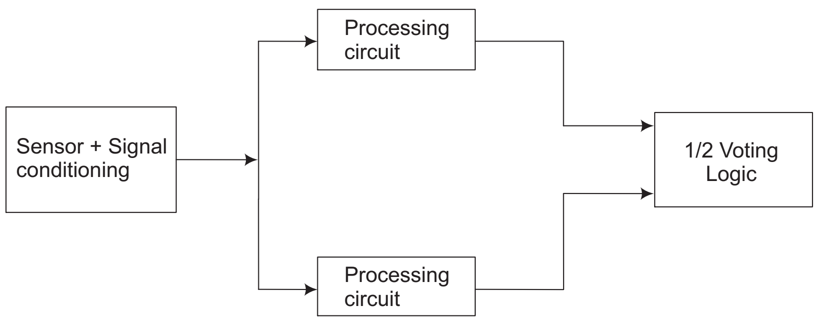
6.2 System Configurations

Generally embedded systems used in process applications will follow one of the configuration discussed below.

1/2 Configuration: In this model two identical systems are operational as shown below**.**

Overall Unsafe failure rate = $\lambda_{u2} * \lambda_{u2}$

Overall Safe failure rate = $\lambda_s + \lambda_s + \lambda_{u1} + \lambda_{u1}$



1/2 Voting Logic

Thus 1/2 configuration ensures safety but causes high spurious trips.

2/2 Configuration: In 2/2 model, two identified systems will be processing the input signals but outputs will be routed through 2/2 logic as shown below.

Overall Unsafe failure rate = $\lambda_{u2} + \lambda_{u2} = 2\lambda_{u2}$

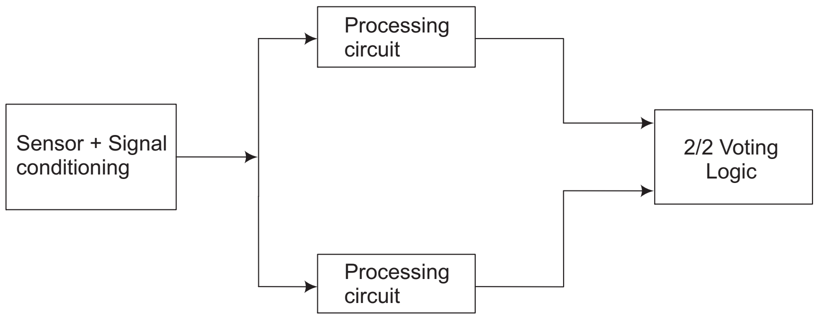
Overall Safe failure rate = $(\lambda_s + \lambda_{u1}) * (\lambda_s + \lambda_{u1}) = (\lambda_s + \lambda_{u1})^2$

In this configuration safe failure rate is satisfactory but unsafe failure rate may not be acceptable.

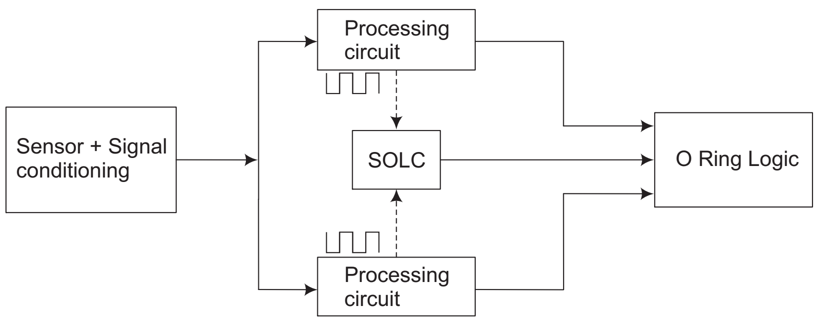
Hot standby Logic: In fault tolerant model, two identical systems are operational. One will be acting as main system while the other will be acting as hot standby. If main system fails, automatic switchover will take place to connect active standby system. The architecture is shown below:

Unsafe failure rate (assuming Reliability of switch over logic is unity) = λ_{u2}

Overall Safe failure rate assuming that Reliability of switch over logic is unity = $(\lambda_s + \lambda_{u1})^2$



2/2 Voting Logic



Hotstandby Logic

Disadvantage of this configuration is that unsafe faults which are not detected by online diagnostics, will not cause switch over. Switch over logic system and ORing logic may fail in unsafe mode thus affecting the safety of the Process Plant.

2/3 Configuration: ***In this model, three identical signal-processing systems are used as shown below . Trip outputs are routed through 2/3 voting logic.***

Overall Unsafe failure rate = $3\lambda_{u2}^{2}$

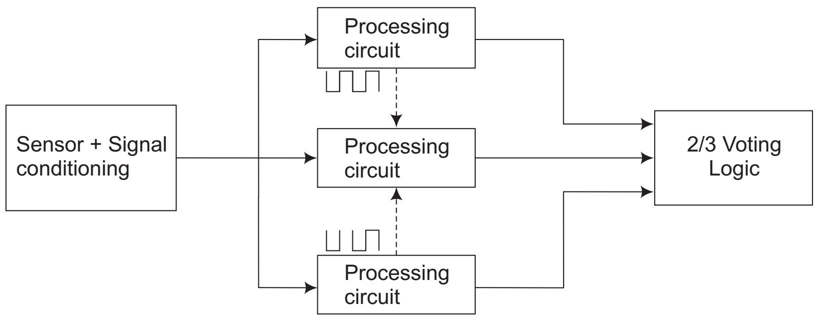
Overall Safe failure rate = $3\,(\lambda_{s} + \lambda_{u1})^{2}$

This model balances between safety and availability with minimum cost. Normally 2/3 architecture is used for safety critical instrumentation system as shown below.

If the same hardware and application software in used in fault tolerant architecture, common mode problems can not be avoided. To avoid common mode problem, hardware and software systems shall be developed by three diverse teams. However, maintenance of diverse systems is not easy during operation and maintenance phase.

7. ENVIRONMENTAL QUALIFICATION OF EMBEDDED SYSTEM

Normally electronic systems work satisfactorily under laboratory conditions where temperature and humidity are controlled. But when the same systems are deployed in a process plant, where

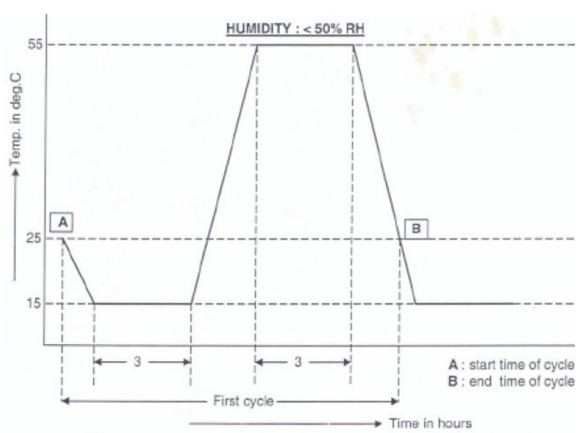


temperature and humidity are not always controlled, the systems may not function as per the requirement. To ensure the proper functioning of the systems during the harsh environmental conditions, the systems shall be qualified environmentally. The following environmental tests shall be carried out as per IS 9000, 1997 or latest:

a) Temperature cycling test:	IS 9000 Part XIV
b) Vibration test:	IS 9000 Part VIII
c) Dry heat test:	IS 9000 Part III, Section 5
d) Damp heat test:	IS 9000 Part V, Section 1
e) Dry cold test:	IS 9000 Part II, Section 4

7.1 Temperature Cycling Test

The objective of this test is to determine the ability of components, equipment or other items to withstand and function satisfactorily when the ambient temperature changes in the process plant. The Equipment under test (EUT) is kept in a chamber where the temperature can be controlled as shown in the following figure. EUT shall be under power off condition during the test.



Temperature Cycle

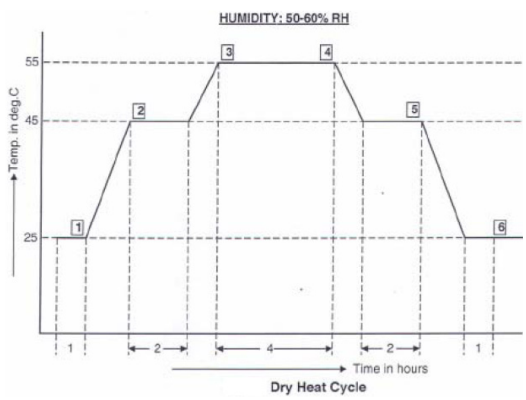
At the end of the test, the system shall remain in normal atmospheric conditions for half an hour. After this, the performance of the equipment shall be satisfactory.

7.2 Vibration Test

The vibration test is applicable to electronic systems which, during transportation or in service, may be subjected to conditions involving vibration of a harmonic pattern, generated primarily by rotating, pulsating or oscillating forces when the equipment is transported in ships, aircraft, land vehicles etc. Vibrations are also caused by seismic events. Equipment under Test in power-on condition, is subjected to sinusoidal vibration over a given frequency range(10 Hz to 55 Hz) with amplitude of 0.75mm. The equipment shall work satisfactorily after the test.

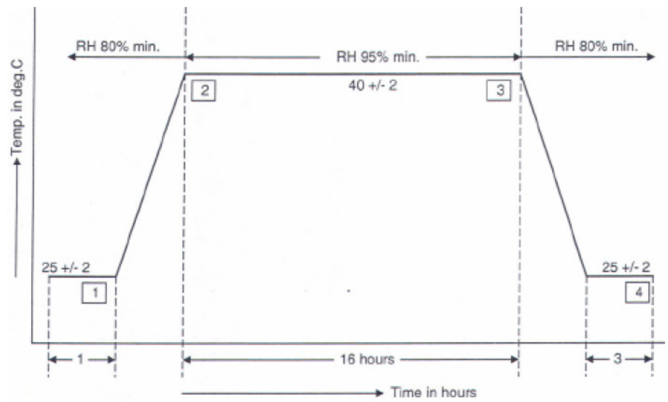
7.3 Dry Heat Test

The objective of this test is to determine the suitability of the system for use under conditions of high temperature when humidity is not high(less than 60%). The equipment is kept in a chamber and powered ON. The humidity is maintained at 55 to 60 %.The temperature of the chamber is increased to 45 degree from 25 degree. The performance is again checked. The temperature is again raised to 55 degree and performance is checked again. After maintaining the temperature at 55 degree for twelve hours, the performance is checked again. The temperature is lowered in steps as shown below and performance of equipment will be checked.



7.4 Damp Heat Test

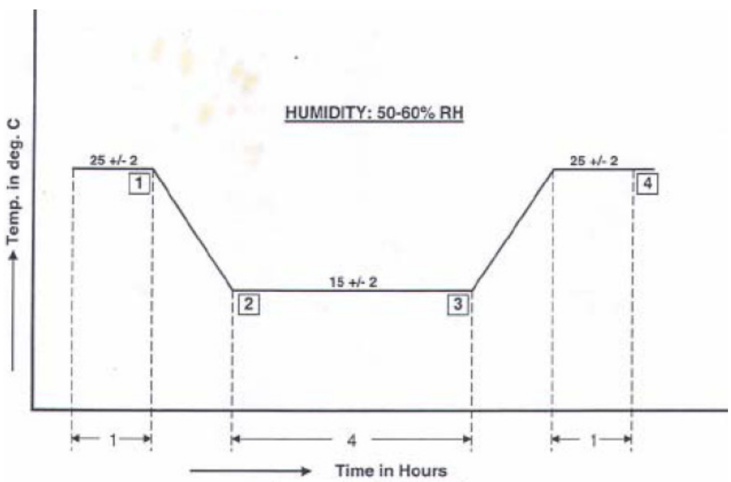
The objective of this test is to determine the suitability of the system / equipment for use and storage under conditions of high relative humidity when combined with cyclic temperature change. The equipment is kept in a chamber in power OFF condition. The temperature is raised from 22 degrees to 38 degrees. The relative humidity is also increased from 45% to 90%. At this time ,the equipment is switched ON and performance is checked. After this the equipment is switched OFF and both temperature and relative humidity is lowered in steps as shown below and performance is again checked.



Damp Heat Cycle

7.5 Dry Cold Test

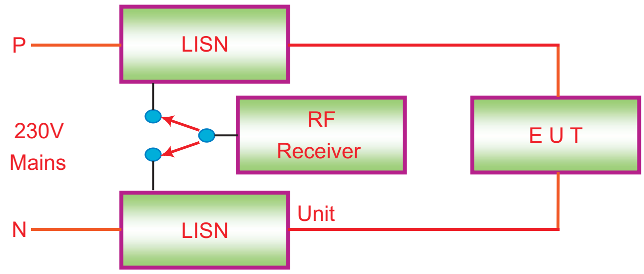
The objective of this test is to determine the suitability of the system for use under conditions of low temperature and humidity less than 60%.The Equipment under test is kept in chamber in power OFF condition. The temperature is maintained at 25 degrees and relative humidity is maintained at 50 to 60 %. The temperature is lowered to 15 degrees as shown below and performance of the equipment is checked. After four hours the temperature is increased to 25 degrees and performance is checked again.



Dry Cold Cycle

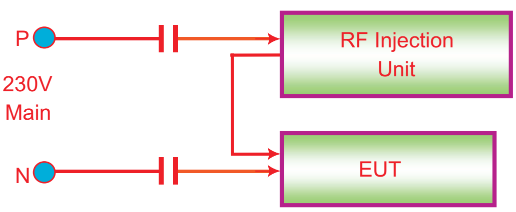
7.6 Conducted Emission Test

In Conducted Emission Test, the noise transmitted in the power supply line by the equipment is measured as shown below. Line input Stabilization Network (LISN) is used to isolate the input mains from EUT or vice versa such that noise does not propagate in either direction. The average value of noise in the range up to 30 MHz shall be less than 66 dBmV. If the noise is larger than acceptable level, the design of the equipment should be modified.



7.7 Conducted Susceptibility Test

The 50Hz AC power supply is not purely sinusoidal. Whenever inductive or capacitive loads are switched ON or OFF, the wave shape is distorted. The spikes in the power supply system will affect the smooth operation of embedded systems. Hence it is preferable to design proper filter system in the input stage of power supply of embedded system. To measure the robustness of embedded system, spikes of value 0.5kV, 1kV, 2kV, 4kV and of 20 msec duration is injected in the power supply line as shown below. The embedded system shall continue to work satisfactorily.



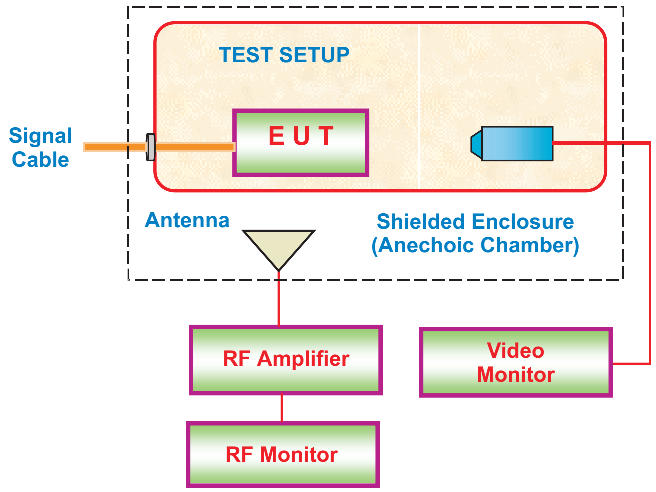
7.8 Radiated Emission Test

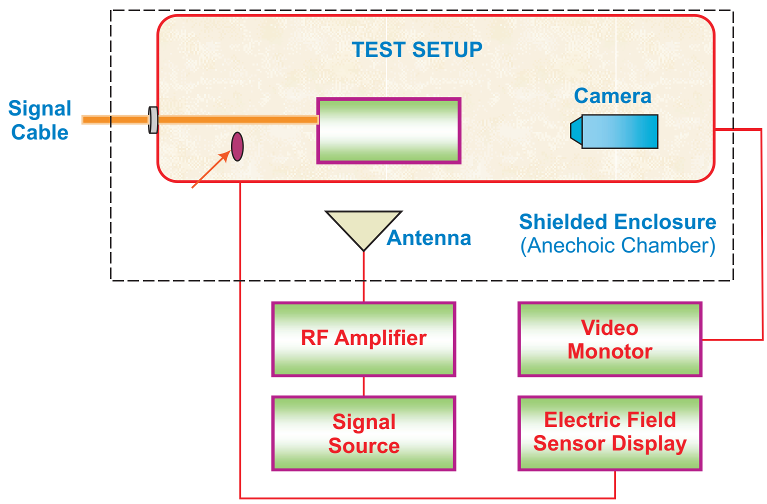
The embedded system shall be properly shielded and grounded to minimize the radiated noise. To measure the intensity of radiated noise, the equipment under test is kept in anechoic chamber as shown below. The radiated noise is received by antenna, kept at three meters from the equipment. After amplification the noise is measured by RF monitor. The measured noise level shall be less than 50.45 dBmV/meter in the range 30MHz to 230 MHz and it shall be less than 57.45 dBmV/meter in the range 230 MHz to 1GH.

7.9 Radiation Susceptibility Test

The shielding and grounding shall be properly designed such that the embedded system shall withstand radiated noise from electrical/electronic equipments in process plants.

In the Radiation Susceptibility test, the performance of the equipment under radiated noise will be measured as shown below. Signal source is tuned to radiate noise in the range 26MHz to 1 GHZ with field strength 1V/meter. The system shall function satisfactorily. The field strength can be increased to 10V/meter if needed.





8. CONCLUSION

Physically and functionally distributed systems are interconnected by optical fibre. The safety actions are directly delivered to the plant by the embedded systems. However the processed data and alarm messages are transmitted to the control room through plant area network. The information is normally stored in fault tolerant servers. Plant operator can query the information from the servers regarding messages, history & trend of process variable etc. Large video display screens are used in control room to display plant mimic diagrams. The value of process variables are dynamically updated by servers. The operational status of embedded systems is also made available to plant operator through servers.

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

EDUSAT A SATELLITE FOR DISTANCE EDUCATION

D. JOHN

1. INTRODUCTION

The various challenges facing the educational sector of the country need to be seen in the context of the several technological possibilities that are available. One such possibility is the space technology, and space technology in the country has indeed evolved to meet the pressing developmental needs in several fields. It is well recognized that Education is an important factor for development. In many respects, primary education makes a positive contribution towards combating the problems of poverty, degradation of environment and improvement of health. Education increases the capacity of the people to transform the vision of the society into operational realities. It therefore becomes the primary agent of transformation towards sustainable development.

The challenge before education in India is two-fold – the challenge of number and the challenge of quality. Indeed, the challenge is, educating meaningfully a one billion strong nation on a continuing basis. Considering the importance of Education sector in the development of the nation, the Indian Space Research Organisation launched a thematic satellite, EDUSAT, to exclusively meet the needs of education in the country. Before the advent of EDUSAT pilot projects were conducted to experiment new trends and techniques in communication system, networking, learning management system etc. At present EDUSAT is being utilized with full participation of the State and Central Governments and the Academic Institutions in the country, to realize its goal of Distance Education.

2. SPACE APPLICATIONS FOR DEVELOPMENT

The Indian Space efforts started in the early sixties with the establishment of Thumba Equatorial Rocket Launching Station near Thiruvananthapuram for the investigation of ionosphere using sounding rockets. Dr. Vikram Sarabhai, the father of Indian Space Programme, had a vision to make India second to none in the development of space technology and in particular its application to accelerate the process of development. He saw new opportunities in using the global vision of planet Earth from space for solving the developmental problems of a large country like India and the use of satellites for spreading education to every nook and corner of the country.

Let us briefly look at the activities in ISRO and the applications of space, for development in communication area, before going in to Distance Education. The three major systems developed by ISRO, are, the INSAT System consisting of communication and meteorological satellites, the Earth Observation System consisting of the Indian Remote Sensing satellites and the Launch Vehicle system comprising of PSLV and GSLV launchers that launch the above two classes of satellites. Providing leadership for space applications, supporting the large user base of the space technology, industrial participation, space commerce, international cooperation in space related activities, maintenance of the space assets and continued development of the state of the art technology related to space systems through R & D programmes are the other related activities of ISRO.

The Indian Space Programme has come a long way since the first launch of a sounding rocket on November 21, 1963 for the study of the upper atmosphere. As of today ISRO has accomplished 50 satellite missions for India, sixteen satellite missions for other countries and twenty six launch vehicle missions. Today India has eleven satellites in the geo-stationary orbit of which ten are communication satellites and one an exclusive meteorological satellite. At present, there are eight remote sensing satellites in the sun-synchronous orbits which are operational. Four decades of space research in the country has resulted in self reliance and today we are one among the six nations having the end to end national capability covering indigenous satellites and launch vehicles and a number of meaningful developmental applications for societal needs based on our space infrastructure.

The ten communication satellites have a total of 210 communication transponders, operating in the various frequency bands viz, Ku band, C band, Ext C band and S band. They meet the requirements of the various users both Government and private. The INSAT applications are, telecommunication i.e. speech circuits on trunk routes, TV broadcasting, business communication, radio networking, VSAT connectivity, meteorological data collection and imaging, disaster warning, tele-education and telemedicine.

The services provided by INSAT include telecommunication circuits reaching remote inaccessible areas with digital technology. In recent times there has been a dramatic increase in corporate communications with VSATs. Introduction of digital TV has resulted in a more efficient way of using the communication satellite transponders. The unique services provided by INSAT are: 1. Training and developmental Education TV services for rural development, 2. Networking of all AIR stations, 3. Data collection for forecasting floods in the various flood-prone rivers of the country, 4. Earth imaging, cyclone warning and monitoring cloud motion vectors for meteorological services on regular basis.

The cyclone warning system provides advance information on impending disaster from approaching cyclones. Specially designed receivers that work with INSAT have been installed in the vulnerable coastal areas of Andhra Pradesh, Tamilnadu, Orissa, West Bengal and Gujarat. IMD has installed 350 receiver stations under INSAT system. Special warning bulletins are transmitted in local languages to the affected areas.

For the satellite aided search and rescue, INSAT system complements the international COSPAS SARSAT system, for providing service in the Indian Ocean region. Signals picked up from distress beacons, from sea, air or land are transmitted through the INSAT transponder to the Mission Control Centre and from there to the Rescue Coordination Centre for search and rescue operation. This is an important service resulting in the saving of human lives and goods.

An indigenous satellite based regional GPS augmentation system is being planned, for navigation of aircraft over the Indian airspace. This system will seamlessly bridge the gap between the European EGNOUS system and the Japanese MSAS system. The system will provide an accuracy of 5 metres for aircraft navigation and landing.

3. SATELLITE INSTRUCTIONAL TELEVISION

There are therefore many areas in which space communication has played a very major role in improving the quality of life. Distance education is one such area, which has generated tremendous excitement in both formal and informal realms of education. The objective of ISRO is to enable the relevant users and Institutions to take education to all parts of the country, more particularly to the difficult and inaccessible areas. First experiment in this field was the Satellite Instructional Television Experiment (SITE) using NASA's ATS-6, conducted in 1975-76. Using the spacecraft and with the help of Direct Reception TV sets, education programmes were beamed to 2,400 community reception centres in remote villages scattered over six clusters of rural villages in central India. Programmes were generated in local languages meeting the specific needs of the illiterate population. It was proven that the tele-education programme can bring significant gains to rural population in the areas of health, hygiene, nutrition, animal husbandry, social values etc., thus enriching the quality of life. The Ministry of Education and National Council for Education Research and Training (NCERT) were active partners in this effort, which also contributed to training of about 50,000 teachers and supplementing the curriculum at primary and secondary schools in those areas.

Later in 1996, under Jhabua Development Communication Project (JDCP) 150 receive terminals at village level and one talk back terminal in each of the twelve block headquarters were installed in a tribal area as part of a pilot project. Now this network has grown to 1062 villages of three districts of Madhya Pradesh, namely Jhabua, Dhar and Barwani. A large scale survey of JDCP programme revealed that substantial gains accrued to tribal community in all subject areas and health awareness increased to a very great extent. Success of JDCP led to establishment of similar network in many states of the country. Through Training and Development Communication Channel (TDCC), programmes are broadcast for educating or training the target audience in several states of the country.

4. EDUCATION SCENARIO IN INDIA

The efforts by various agencies, States and Central Government Establishments, Open Universities and ISRO, over the years, have succeeded in improving the quality of education to a great extent. But the problem continues to be formidable given the size of the country. India, even today, has the largest illiterate population in the world; up to 30% of the Indians above the age of six do not know to read and write. To appreciate the challenges posed, let us look at the requirements. In case of secondary education, at the end of the 10th Five Year Plan there was an addition of about 6.9 million students requiring 0.13 million new teachers and about 35,000 new schools. Studies indicate that participation in higher education is only 6% of the eligible young people in the relevant age group, where as, the figure stands at 50% in the developed countries. At school level 87% of the students drop out by the time they reach standard XII mostly due to shortage of infrastructure. Lack of good qualified teachers has affected the performance and the failure rates are very high in Mathematics, Science and English at school level. Higher education is also faced with dearth of facilities such as scientific equipment, laboratories, libraries etc apart from qualified and good faculty.

The Indian education system is very vast and full of complicated peculiarities. There are a lot of wide gaps that prevent the universalisation of education. Higher education has its own share of problems. Some of the critical issues that come to the mind in the context of the India scenario are – reaching the unreached, continuing education, resource crunches and the need for lifelong enrichment.

For a great majority of people living in the countryside and in small towns, it is often difficult to gain access to the benefits of quality education, latest scientific equipment, the use of up-to-date libraries and research facilities. Nor it is physically possible to take all facilities to small towns and villages. Even within big cities, there are institutions, which are not equipped with well-qualified teachers, laboratories and libraries etc. In the case of training and skill up-gradation courses, experts cannot travel to every location of training and so the opportunity to learn from experts is reduced. Thus in the education system there is a serious crunch both in terms of infrastructure and human resources. This is where the intelligent use of technology helps. Satellite communication can help in establishing the backbone required to spread education to remote regions and achieve universal access to knowledge.

5. TECHNOLOGY FOR EDUCATION

The use of technology for education has its own dimensions and can increase the efficiency of the education sector. Use of suitable technology helps in the following ways.

- Facilitates reaching to students in a cost effective way
- Helps weaker students to learn with timing flexibility
- Offers equal quality contents to all
- Tackles the pluralistic character of the country

Information and Communication Technology (ICT) has power to store, retrieve, sort, filter, distribute and share information seamlessly and this can lead to substantial efficiency in production, distribution and market educational contents. The increase in efficiency and subsequent reduction of costs brought about by ICT is leading to the creation of new products, services and distribution channels within traditional, as well as innovative business models. Intangible assets like intellectual capital are increasingly becoming the key source of value.

Most of the developed nations have laid the foundation in using the ICT for education and several other developmental works. Though not a cure for all the ills of education, it can certainly be the best compliment for the existing conventional education system. It has also the potential of reaching the un-reached at a faster pace.

One of the most clearly demonstrated applications is distance education. Distance education has been a particularly successful model in developing countries where affordability and geography have been real barriers to access. ICT-mediated learning is an excellent tool in the delivery of higher secondary school lessons and vocational training. Through the use of ICT data repositories and curricula can be developed collaboratively, educational materials can be procured more cost effectively, staff and student time can be scheduled more efficiently, and individual student performance can be monitored more closely.

With a judicious mix of ICT, satellite communication and high quality content most of the problems facing the present school education could be mitigated. The major advantages of satellite based distance education include:

- Simultaneous delivery of classroom lessons to a large number of geographically spread-out people in short time.
- Uniformity of the classroom content.
- Access to experts in the subject and his/her lessons.
- Repeatability of delivery of the lessons from the archives.
- Capability to share the same network by different groups of users such as students, students requiring coaching for public exams and teachers.
- Significant savings in expenditure due to economies in travel, replication of lessons and logistics.

The distance education has come a long way in India starting with postal correspondence courses as the mode of delivery. Postal correspondence courses are still doing well; however today they are supplemented and enhanced by the inclusion of electronic media and satellite based communication. One-way television and radio communication for distance learning has been improved by the introduction of two-way interactive audio and video communication. Digital connectivity to remote area is an important element for this, and satellite communication provides for it. Five transponders of INSAT system are already carrying education programmes, Gyandarshan, TDCC programmes, APNET-MANA TV, ERNET for academic institutions, and also EDUSAT pilot projects that were established before the launch of EDUSAT.

6. EDUSAT PROGRAMME

Considering the growth in demand for satellite connectivity for education, ISRO conceived the EDUSAT project through which the full capacity of a geo-stationary spacecraft would be committed to meet the demand of distance education. It was envisaged that it would meet the challenge of number and quality in digital connectivity facilitating effective teachers training and curriculum based teaching in several regional languages, and also by providing access to quality resource persons especially in higher and professional education. It was expected to strengthen the distance education efforts initiated by various agencies and take education to all parts of the country and provide access to new technologies.

EDUSAT (GSAT-3), the first satellite of the Education Satellite System, was launched by the third flight of the indigenous launch vehicle GSLV on September 20, 2004. The communication transponder specifications of EDUSAT are given in Table 1. The satellite was specially configured to have multiple spot beams covering different regions of India, keeping the diversity in mind. The satellite is meeting substantially, the requirements of different parts of the country. This is a

Table 1. *Communication transponder specifications of EDUSAT*

COMMUNICATION PAYLOAD	EIRP	G/T
5 Ku Band Regional Beams	55 dBW	7 dB/K
1 Ku Band National Beam	50 dBW	2 dB/K
6 Ext C Band National Beams	37 dBW	1 dB/K

powerful, state-of-art, satellite operating in Ku band frequency facilitating the use of easy-to-handle small ground terminals. The satellite provides communication coverage through five regional beams and a national beam. It also provides full Indian Coverage including the Indian islands in Extended-C band.

The concepts of multicasting and interactive multimedia mode of delivery are effectively used. The satellite provides a sustainable distance learning service in India using convergence of advanced technologies. This system is primarily for school, college and higher level of education. However it also supports non-formal education and developmental communication. One of the primary objectives of EDUSAT project is to take virtual classroom to all regions of the country; thus making available high quality education programmes by eminent teachers and scientists to less-privileged classrooms at remote areas. It also provides interactive mode of teaching by which effectiveness of delivery through electronic media can be improved.

Another important area is science education. The scarcity of qualified science teachers can be overcome by beaming the lectures and transmitting the notes on the tele-link to far-flung schools. Lessons and experiments from science labs also could be transmitted through this media. The tele-education terminal at remote areas can be used for adult literacy classes as well as imparting training to farmers and artisans for improving their knowledge base and skills by time sharing the system with regular education programmes.

Thus the potentials of the satellite for tele-education are high. Integrating the hardware and development of software for its operation were being validated during the pilot project phase of EDUSAT. But the major challenge is the content generation. One of the primary modes is to capture the class by eminent teachers as they are delivered. Addition of multimedia and access to database can improve the quality of service. Generating contents on various programmes, range and modes do demand the attention of many experts and agencies.

In the first phase, in the pre GSAT-3 period, three pilot projects using the Ku band transponder of INSAT-3A/3B were evaluated in Maharashtra, Karnataka and Madhya Pradesh. These projects were subsequently improved and expanded in their reach. In Maharashtra the Y.B. Chavan State Open University, Nashik, uses it for distance education. In Karnataka the Visweshwaraiah Technical University (VTU) is using it for transmitting virtual-classroom lessons to all the engineering colleges of the university. In Madhya Pradesh, the Rajiv Gandhi Technical University uses it for teaching support. Each location caters up to about 100 classrooms. The pilot phase provided valuable inputs which were implemented in the subsequent phases. Some of these inputs are: The Schools/ colleges with inadequate teaching need to be given preference over well established institutions for EDUSAT connectivity. Quality should be improved emphasising professionalism and media experts need to be consulted to provide good Video and Audio effects.

Transmission rate of teaching end video and audio to be increased to 1 Mbps. Provision should exist to record the programme at college end. Distributed viewing of lectures at colleges should be facilitated through LAN connection. Proper training needs to be given to coordinators and students to use the interactive system.

In the second phase, EDUSAT was being used in a semi-operational mode with uplinks in each beam and about 100 to more than 1000 classrooms per beam using the EDUSAT. Today there is no state in India without EDUSAT terminals. The types of connectivity provided by EDUSAT are shown in Table 2. More users came forward to fund the ground receive network during the semi-operational phase.

Table 2. *Types of connectivity provided by EDUSAT*

CONNECTIVITY	SYSTEM	USER
Broadcast	Receive only terminal	Primary Education
Interactive Teaching	Satellite Interactive Terminal	Secondary and Higher Secondary Education
Video Interactive Teaching Education	Satellite Interactive Terminal with	Higher and Professional Return Channel Video

Based on pilot phase experience the hardware and system parameters were optimized during semi-operational phase. The receive only terminals (ROT) were configured to work with 90 cm antennas for Ku band operation. The reception was found undisturbed even under rainy conditions. The satellite interactive terminals (SIT) provide good reception under all weather conditions in Ku band with 1.2 m antennas. The configuration and design of the terminals both ROT and SIT, took in to consideration the cost aspects so that they are affordable for the various education departments/ institutions in the country.

In the third phase, which is the current phase, the EDUSAT network is being made fully operational. All major universities, IITs and Medical colleges are being provided with multimedia, multicasting facilities through which the teaching by experts can be made accessible to the needy, in any part of the country in real time or offline basis. This will also enable interlinking of database and libraries and effective dissemination of information.

7. CURRENT STATUS OF EDUSAT

At present 30,500 receive only terminals are operating on EDUSAT. These terminals are essentially used for primary school education. The very first network under this category was established in Karnataka with terminals in 1000 remote primary schools in the State. High quality, curriculum based programmes are being transmitted regularly from Bangalore, the State capital, in the local language. This network facilitates distance learning for about 50,000 primary school students simultaneously and the content, though in local language, is of the same quality as that imparted in leading schools in Bangalore. Other networks under this category are functioning equally remarkably. In addition there are 3,500 satellite interactive terminals operating on EDUSAT, mostly meeting the requirements of higher education. One network which is note worthy under this category is the INDO-US network in which at present 50 higher technical institutions are connected, one of them being SONA College of Technology, Salem, Tamil Nadu and IIT Bombay in responsible for the content. Under an agreement, Professors visiting India from leading institutions in US use this network for reaching out to students in these 50 institutions. There are at present three SIT networks operating in the North-Eastern States, where such facilities are very much required considering the remoteness and hilly nature of the terrain.

8. CONCLUSION

For a one billion strong nation like India, human resource itself could play a very significant role in the development of the country. Demographic studies and projections tell us that by 2020, productive youth will constitute substantial portion of the population. This may be today considered as a liability for the nation. But through meaningful education, the same uneducated and uninformed human resource can be converted into an economic asset. So in the years to come the economic implications of being an educated nation could be tremendous.

The need of the hour is therefore to prepare learners for the knowledge-based economy and society, to champion the cause of life-long learning and to stimulate the creation of word-class knowledge resources and universal access to them through Information Technology. Thus in today's context, higher education, distance and open learning systems and ICT go hand in hand and essentially become more of a package deal. In the era of globalization and the information age, quality higher education becomes a prerequisite. With ICT and space applications, the benefits of distance and open learning can be leveraged. This is where EDUSAT can play a substantial role.

EDUSAT will help in taking education to the 'doorsteps' of the students and will also make it interactive. "Learning when you want and at the speed you want" will become a reality, as EDUSAT uses not only broadcasting but also interaction. Besides, it can also have the advantage of video on-demand, data broadcasting, easy storage and retrieval, examination through Personal Computers, etc. It is envisaged that distance learning could be the biggest beneficiary of EDUSAT, apart from teachers' training, primary education, secondary education and higher education. Depending on the enthusiasm and demand ISRO will add more satellites to the EDUSAT fleet if needed.

REFERENCES

[1] A. Bhaskaranarayana & D. John, "Edusat A Satellite for Distant Education" BES Review, October-December 2004.

[2] D. John, "EDUSAT Bridging the Digital Divide in Education" Proc. Conf. 4th VSAT India 2004. New Delhi, India, 3 December 2004.

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

SATELLITE COMMUNICATION SYSTEM

D. JOHN

1. INTRODUCTION

The potential of satellites for communications was first realized by Arthur C. Clarke who published a paper in 1945 proposing a three satellite system to provide worldwide relay communications from the geostationary orbit. The geostationary orbit is unique since its orbit period of 24 hours precisely matches the earth's rotation allowing one satellite to provide fixed coverage over a large region. The first satellite to attain geostationary orbit after two failures was SYNCOM III launched on August 19, 1964. The first Indian geostationary satellite was an experimental communications satellite, APPLE, launched on June 19, 1981.

2. PRINCIPLES OF ORBITAL MOTION

The motion of satellites around the earth is governed by the same laws that govern the motion of planets around the sun. In 1609 Johannes Kepler (1571–1630) published his first two laws of planetary motion. Many years later he published the third law after much painstaking work.

2.1 Kepler's Laws

First Law - The orbit of each planet is an ellipse, with sun at one of the foci.

Second Law - The line joining the planet to the sun sweeps out equal area in equal time.

Third Law - The Square of the period of a planet around the sun is proportional to the cube of its mean distance from the sun.

Later in 1665, Sir Isaac Newton (1642–1727) was able to derive Kepler's Laws from his own laws of mechanics and developed the Theory of Gravitation. Kepler's Laws apply generally to any two bodies in space which interact through gravitation.

Satellites remain in orbit because of a delicate balance of forces. A satellite is said to be in orbital motion when two principal forces governing its motion are exactly in balance. These two forces are gravitational attraction of the earth and the centrifugal force on the satellite in circular motion around it.

The force on an earth satellite due to gravity is given by:

$$F_g = \frac{GMm}{R^2} \quad \text{...(1)}$$

Where m is the mass of the satellite, R is the distance between the satellite and the centre of the Earth, M is the mass of Earth and G is the gravitational constant.

The centrifugal force is given by

$$F_c = \frac{mv^2}{R} \quad \text{...(2)}$$

Where v is the velocity of the satellite normal to the radius vector

Equating these two forces we have

$$v = \sqrt{\frac{GM}{R}} \quad \text{...(3)}$$

For a geostationary satellite at a height of 35786 km from earth the orbit velocity would be 3.08 Km/s.

For a LEO satellite at a height of 600 km above earth the orbit velocity would be 7.56 km/s.

From Kepler's third law

$$a^3 \propto T^2 \quad \text{...(4)}$$

$$a^3 = kT^2 \quad \text{...(5)}$$

Where 'a' is the semi major axis of the ellipse, the mean distance from the primary body in space; equal to the radius R in the case of circular orbit. 'T' is the orbit period.

For an artificial satellite orbiting earth

$$k = \frac{GM}{4\pi^2} \quad \text{...(6)}$$

G is universal gravitational constant = 6.672×10^{-11} m^3 /kg s^2

The mass M of earth is 5.9742×10^{24} kg

For a circular orbit with a period of one day (sidereal = 23h, 56 min, 4.09 sec)

$$a^3 = \frac{GM}{4\pi^2} \times 86164^2$$

$$a = R = 3\sqrt{\frac{6.672 \times 10^{-11} \times 5.9742 \times 10^{24} \times 7424 \times 10^6}{4\pi^2}}$$

$$= 42164 \text{ km}$$

Radius of the earth $r = 6378$ km

Hence height of Geo arc above the Earth = 35,786 km

It may be noted that for a satellite to be stationary for an observer on earth, the orbit period should be equal to the rotational period of earth. In one solar day the earth moves in its orbit by 0.986° with respect to sun. Hence earth rotates by 360.986° in 24 hours (solar day) and it takes 23h, 56 min, 4.09 sec (sidereal day) to rotate 360°.

3. ORBITAL ELEMENTS

It maybe noted that the above case of circular orbit is when the eccentricity '*e*' of the elliptical orbit becomes equal to zero. To specify completely the properties of an elliptical orbit, five parameters are needed. Furthermore to specify the position of a satellite in its orbit, the sixth parameter, the time at a given position (such as perigee) is needed. These six parameters representing the classical orbital elements are as given below:

a – Semi-major axis

e – Eccentricity

i – Inclination, the angle that the orbital plane makes with the equatorial plane

Ω – Angle of the right ascension of the ascending node, the angle between the direction of vernal equinox and the ascending node.

ω – Argument of perigee, the angle between ascending node and perigee with respect to centre of the earth.

t – Time of perigee crossing

Some other terms used in describing a satellite orbit are:

Apse line – the line joining apogee and perigee

Line of nodes – the line joining ascending node and descending node (i.e. equatorial crossings)

These orbital elements are defined in the earth centered fixed inertial *ijk* coordinate system with '*k*' pointing upwards through the earth's north pole, '*i*' in the earth's equatorial plane and pointing in the direction of vernal equinox (i.e. where the sun crosses the equatorial plane in spring) and '*j*' completes the orthogonal triad.

4. SATELLITE ORBITS

There are several satellite orbits in use but not all of them are useful for satellite communication. The application of a satellite mostly dictates its orbit. The orbits are as follows.

- Low earth orbits (LEO) have altitudes between 200 and 1000 km. They are of low inclination to the equatorial plane with moderate eccentricity. This may be used for store and forward communication.
- Medium –altitude earth orbits (MEO) with altitudes of around 10000 km. and inclinations in the range of 45° to 60°.
- Polar orbits with high inclination. In this orbit the satellites cross over the poles of the earth.

- Sun synchronous polar orbit. Earth resource remote sensing satellites are placed in this orbit.
- Highly elliptical orbits (HEO) are used for satellite communications at higher latitudes.
- Geostationary earth orbit (GEO) has zero inclination and orbital period equal to the rotational period of earth.
- Geostationary transfer orbit (GTO) is an eccentric orbit designed to position a satellite so that it can be efficiently placed in to GEO.

For continuous communications between two points on the ground, a satellite should be continuously visible from both points. In addition if the satellite remains stationary with respect to the ground locations the antennas can be fixed and do not require tracking of the satellite. Satellites in GEO fulfill this requirement.

The GEO stationary earth orbit has the following characteristics:

- Period – 1 day [23h, 56 min, 4.09 sec (sidereal day), Earth to rotate 360°].
- Shape – circular with earth at the centre.
- Orbit plane – coincident with earth's equatorial plane.
- Orbit height – 35786 Km. (5.61 earth radii) above the earth's surface.

Most commercial satellites such as INSAT, INTELSAT, INMARSAT employ GEO for communication because of the following advantages:

1. The satellites remain stationary with respect to a point on earth; therefore the earth station antenna is not required to track the satellite. This reduces the earth station cost considerably.
2. The satellite covers almost 38% of the surface of the earth with minimum elevation angle of 5°.
3. Three geostationary satellites 120° apart can cover the entire surface of the earth with some overlapping except for the Polar Regions, above 76°N and 76°S latitudes with a minimum elevation angle of 5°.
4. Doppler Shift caused by the satellite drift in orbit is small for all earth stations within the coverage area. This is desirable for many synchronous digital systems. – The small drift is caused by the gravitational attraction of moon and sun.
5. Signal strength is stable owing to constant ground-satellite range.
6. Interference effects are easy to predict owing to stable geometric relationship.
7. Relatively lower number of sun eclipses and hence battery charge/discharge cycles, giving long battery lifetime.
8. Time between launch and deployment/operation is relatively small, of the order of weeks.

However there are certain disadvantages such as coverage not available beyond ± 76° latitude, relatively large propagation delay and path loss and high launch cost. Nevertheless the advantages certainly outweigh the disadvantages and this orbit is the most popular one for satellite communication. Figure-1 shows the orbit and attitude of GEO satellite.

5. SATELLITE COMMUNICATIONS SYSTEM

Satellite communications system can be broadly divided into two segments, (1) Space segment and (2) Ground segment. The space segment includes the satellite and also the ground facility needed to

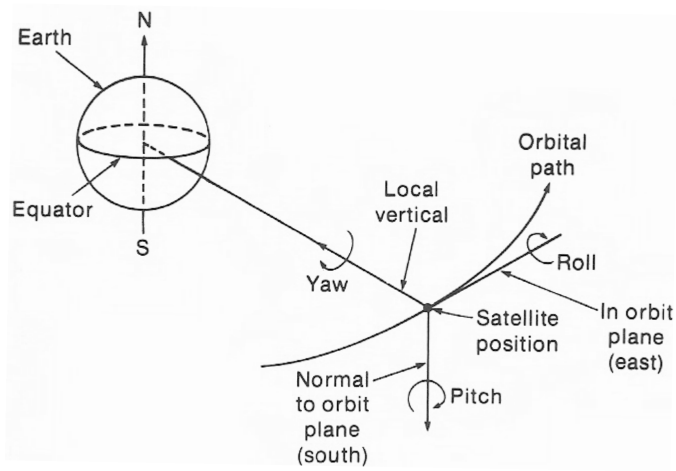


FIG. 1 ORBIT AND ATTITUDE OF GEO SATELLITE

keep the satellite operational. This is referred to as the telemetry, tracking and command (TT&C) system. In many cases a dedicated ground station is employed for the purpose of TT&C.

The space segment can be further classified according to their function; *bus* system and the *payload*. The bus system refers to the various subsystems which provide various support functions required to service the payload. The payload refers to the equipment used to provide the service for which the satellite has been launched.

In a communication satellite the equipment which provides connecting link between the transmitting and receiving antennas of the satellite is referred to as the *transponder*. The transponder forms the main section of the payload, the other being antenna sub-systems.

5.1 Satellite Bus System

Satellite bus system is designed to support the payload operations reliably throughout life of the communication satellite. The bus consists of several subsystems to meet the various specific requirements in the satellite. The bus system has to provide the following:

1. Maintain the position and orientation of the satellite at any specified orbit location and keep the antennas pointed correctly towards the service area. The east-west and north-south drift of the satellite in orbit must be constantly corrected. The pitch, roll and yaw rotations of the satellite also require corrections. This function is performed by the attitude and orbit control subsystem (AOCS).
2. Providing the required torque needed, for maintaining the attitude of the spacecraft and assisting station-keeping in association with the AOCS. This is performed by the propulsion subsystem.
3. Maintain the temperature of the various spacecraft subsystems within the specified limits. This is done by the thermal subsystem.

4. Providing the mechanical and structural support and rigidity for the spacecraft. This is done by the spacecraft structure.
5. Providing DC power to all active components of the spacecraft. This is carried out by the power subsystem that includes the solar arrays and batteries.
6. Providing TT&C support to the spacecraft through onboard TT&C RF and Base band systems that work together with the ground facility. The support includes:
   a) Providing data to ground control centre for monitoring the performance of the various spacecraft subsystems.
   b) Supporting ground station tracking requirements.
   c) Accepting commands from the ground control centre for altering spacecraft configurations and performing vital manoeuvres.

5.2 Communication Payload

The communication payload consists of the transponders and the communication antennas. They together, receive signals from the up-linking ground stations, amplify the same and then transmit to the desired service area on the surface of the Earth.

5.2.1 Transponder

A transponder, some times called a *repeater*, is a series of interconnected units which forms a single communications channel between receive and transmit antennas. Some of the units utilized by a transponder in a given channel may be common to a number of transponders. Hence a specific transponder is an equipment channel rather than a single item of equipment.

Before going into the details of the various units of a transponder, let us look at the overall frequency arrangement of a typical C-band communications satellite. The bandwidth allotted for C-band service is 500 MHz and this is divided into subbands, one for each transponder. A typical transponder bandwidth is 36 MHz and allowing for a 4 MHz guardband between transponders, 12 such transponders can be accommodated in 500 MHz bandwidth. By making use of polarization isolation, this number can be doubled. The block of frequency allotted in Ku-band is also 500 MHz and this is also divided in to12 numbers of 36 MHz bandwidths for the transponders as in the case of C-band.

Polarization isolation refers to the fact that carriers, which may be on the same frequency but with opposite senses of polarization can be isolated from one another by receiving antennas matched to the incoming polarization. With linear polarization, vertically and horizontally polarized carriers can be separated in this way. With circular polarization, left hand circular and right hand circular polarizations can be separated. Because carriers with opposite senses of polarization may overlap in frequency, this technique is referred to as frequency reuse. Frequency reuse may also be achieved with spot-beam antennas; these combined with polarization reuse can increase the effective use of a bandwidth by a factor of four.

The function of a transponder is to receive the uplink RF signals and convert these signals into appropriate downlink frequency and power, for transmission towards the service area. Two types of transponder architectures are possible: (1) transparent transponder and (2) regenerative transponder.

Transparent transponder (sometimes called *bent-pipe transponder*) only translates the uplink frequency to a suitable downlink frequency and power, without in any way processing the baseband signal. Regenerative transponder in addition to frequency translation and signal amplification, demodulates the baseband signal, processes the same and again modulates on the RF carrier. A major advantage is that the uplink noise is eliminated from the signal by recovering the baseband signal onboard. Also, if needed, the received digital signals can be multiplexed as required. However a regenerative transponder is more complex than a transparent transponder.

5.2.1.1 Wideband Receiver

The gain of a transparent transponder is typically around 110 dB. With such high end-to-end gain requirement a transponder tends to become unstable unless the isolation between receive and the transmit section is made high. Isolation is achieved by using filters with high isolation between transmit and receive frequency bands and careful RF shielding. The signals from the receive antenna and feed system are fed through a band-pass filter (BPF) to the wideband receiver. The BPF attenuates all out-of-band signals such as transmission from ground stations of adjacent satellite systems. A duplicate receiver is provided so that if one fails the other is automatically switched in. The combination is referred to as a redundant receiver, meaning that although two are provided, only one is used at a given time.

The first stage of the receiver is a low noise amplifier. The spacecraft antenna is pointed towards a relatively warm earth having noise temperature of about 300° K. Therefore there is no advantage in reducing the noise temperature of LNA much below this level. Nevertheless this amplifier adds little noise and at the same time provides sufficient amplification. The equivalent noise temperature of a satellite receiver may be of the order of few hundred Kelvin. The LNA feeds into a mixer stage which also has a local oscillator signal, for frequency conversion. The second amplification follows the mixer to provide an overall receiver gain of about 60 dB. Splitting the gain between the low noise amplifier at 6 GHz and second amplifier at 4 GHz prevents oscillation (instability) which might occur if the gain were to be provided at the same frequency. The wideband receiver utilizes solid-state active devices. FET /HEMT amplifiers, diode mixers and bi-polar junction transistors are employed for the low noise amplifier, the mixer and the second amplifier respectively. Typically a linear dynamic range of 17 dB (minimum) between high and low gain settings of the transponder channels and an overdrive capability of 20 dB above the low gain setting of the channels is ensured for the wideband receiver in its operation.

5.2.1.2 Input Demultiplexer

The input de-multiplexer separates the broadband input (500MHz) into the transponder frequency channels (1 through 12). The channels are usually arranged in even numbered and odd numbered groups. This provides greater frequency separation between adjacent channels in a group, which reduces adjacent channel interference. The output from the receiver is fed to the power splitter, which in turn feeds the two separate chains of circulators. The full broadband signal is transmitted along each chain and the channelizing is achieved by means of channel filters connected to each circulator. Each filter has a bandwidth of 36 MHz and is tuned to the appropriate centre frequency. Although there are considerable losses in the demultiplexer, these are easily made up in the overall gain for the transponder channels.

5.2.1.3 Power Amplifier

A separate power amplifier, consisting of a driver amplifier followed by a TWTA (traveling wave tube amplifier), provides the output power for each transponder channel. The driver amplifiers are incorporated with telecommandable step attenuators for gain setting of the transponder channels. The amplifiers operate in linear range and have capability to withstand RF overdrive in the uplink.

Switch matrices are provided at the input and the output of the above power amplifier to provide the required redundancy in the event of power amplifier failure.

Traveling wave tube amplifiers are widely used in transponders to provide the final output power required to the transmit antenna. The term TWT (traveling wave tube) is normally used to denote the tube without its power supplies. In the TWT, an electron-beam gun assembly consisting of a heater, a cathode and focusing electrodes is used to form an electron gun. A magnetic field is required to confine the beam to travel along the inside of a wire helix. The space TWT employs permanent-magnet for focusing. The RF signal to be amplified is coupled into the helix at the end nearest to the cathode and it sets up a traveling wave along the helix. The electric field of the wave will have a component along the axis of the helix. In some regions this field will decelerate the electrons in the beam and in others it will accelerate them, so that electron bunching occurs along the beam. The average beam velocity which is determined by the DC potential on the collector of the tube is kept slightly greater than the phase velocity of the wave along the helix. Under these conditions the kinetic energy of the beam is converted to potential energy in the wave. The wave actually travels around the helical path at close to the speed of light, but it is the axial component of the wave velocity which interacts with the electron beam. Because of this effective reduction in phase velocity the helix is referred to as a slow wave structure. The amplified microwave power output is obtained at the collector end of the helix.

At low input powers the output-input power relationship is linear. At higher power inputs the output power saturates, the point of maximum power output being known as the saturation point. The linear region of the TWT is defined as the region bound by the thermal noise limit at the low end and by what is termed the 1-dB compression point at the upper end. This is the point where the actual transfer curve drops 1 dB below the extrapolated straight line of the linear region. Near saturation phase modulation of the signal occurs and this is termed AM/PM conversion. This effect causes intermodulation distortion. To avoid this, a TWT is operated in the linear portion by providing input backoff.

5.2.1.4 Output Multiplexer

The outputs from the transponders are combined into a composite signal in the output multiplexer. The configuration of the multiplexer consists of a bank of channel band pass filters and a common, short circuited rectangular wave-guide manifold. The manifold configuration has a distinct advantage of low insertion loss. The signals combine at the output of the filters, each filter providing the necessary isolation from the adjacent channels. In addition the filters reject the harmonics generated by the nonlinearity of the high power amplifiers. The insertion loss of the filters must be minimized since power loss at this stage necessitates an increase in the size of the power amplifier resulting in higher DC power requirement from the spacecraft. Finally the output of the multiplexer is delivered to the antenna for transmission.

5.2.2 Antennas

Antennas are carried onboard a satellite to provide the dual functions of receiving the uplink and transmitting the downlink signals. There are various types of antennas that range from the one having omni-directional characteristic to highly directional antennas. Omni-directional antennas are used for TT&C. A combination of main quadrifilar helix antenna and an axial mode null filling antenna provides omni-directional pattern. Directional beams which are required for communications are produced by parabolic reflector antennas.

The gain of such an antenna relative to an isotropic radiator is given by:

$$G = h\left(\frac{\pi D}{\lambda}\right)^2 \qquad ...(7)$$

Where λ is the wavelength, D is the reflector diameter and η is the aperture efficiency.

A typical value for η is 0.55.

The 3 dB beam width of the antenna is given approximately by:

$$\theta_{3dB} \cong 70\frac{\lambda}{D} \quad \text{Degrees} \qquad ...(8)$$

The ratio D/λ is seen to be the key factor in these equations, the gain being directly proportional to its square and beam width inversely proportional to it. Hence the gain can be increased and beam width made narrower by increasing the reflector size or decreasing the wave length. The largest reflectors are those for 6/4 GHz band. Comparable performance can be obtained with considerably smaller reflectors in the 14/12 GHz band.

Wide beams for global coverage can be produced by simple horn antennas at 6/4 GHz. To obtain the required shaped beam, the reflector can be illuminated with an array of feed horns suitably arranged. 'Shaped reflectors' can also be used to produce any desired coverage on the surface of the earth. Here the reflector surface is computer modeled as a series of mathematical functions which are changed until the model produces the desired coverage on the surface of the earth through mathematical simulations. Shaped reflector antennas are much lighter in weight and have less loss compared to the ones illuminated by the array of feed horns to obtain the desired coverage. Dual girded reflector antenna, with orthogonal linear polarization can be used for transmit and receive functions, as in INSAT 4A/4B.

Same feed horn may be used to transmit and receive carriers with the same polarization. Transmit and receive signals are separated in a device known as diplexer, and the separation is further aided by means of frequency filtering. Polarization discrimination may also be used to separate transmit and receive signals using the same feed horn. For example, the horn may be used to transmit horizontally polarized waves in the downlink frequency band, while simultaneously receiving vertically polarized wave in the uplink frequency band. The polarization separation takes place in a device known as orthocoupler or orthogonal mode transducer (OMT). Separate horn may also be used for transmit and receive functions, with both horns using the same reflectors.

6. EARTH STATIONS

Earth stations are a vital element in any satellite communication network. Depending on the application, an earth station may have both transmit and receive capabilities, or may have either transmission or reception capability. The simplest earth station is the TV receive-only (TVRO) system for home TV reception and the most complex is the terminal station used for international communications network. Other types of earth stations are those on ships at sea, commercial/military land stations and aeronautical mobile stations.

A fundamental parameter in describing an earth station is its G/T, the figure of merit which represents the sensitivity of the earth station. The other important parameter is its EIRP, which is decided by the power amplifier used and the size of the antenna, and this has an impact on the earth station cost.

Signals from terrestrial network or directly from the user in some applications, are fed to an earth station through a suitable interface. The baseband signals are then processed, modulated and up-converted to the desired frequency, amplified to the required level, combined with other carriers (if necessary) and transmitted via the antenna. The feed system provides the necessary aperture illumination, introduces the required polarization and provides isolation between the transmitted and received signals. The antenna transmits the signals in the uplink.

Signals received by the earth station through the antenna are amplified in low noise amplifier, down-converted to an intermediate frequency (IF), demodulated and transferred to the terrestrial network via the interface (or directly to the user in some applications). Other subsystems such as tracking, antenna control and power supply provide necessary support.

6.1 Earth Station Antenna

Most of the earth stations use reflector antennas since such antennas can readily provide high gain and the desired side lobe characteristics. Cartesian plot of antenna radiation pattern is shown in Figure-2

The earth station can have either axi-symmetric or asymmetric (also called *offset*) antenna configurations based on their geometry. In an axi-symmetric configuration the antenna axes are symmetrical with respect to the reflector, which results in a simple mechanical structure and antenna mount. Two of the most commonly used arrangements are:

1. Prime focus feed
2. The Cassegrain and Gregorian systems

Prime focus system consists of a parabolic reflector with a feed located at the focus of the parabolic reflector. Although simple to implement, such a feed arrangement results in a larger antenna noise temperature for earth station because the feed horn is pointed towards earth, which is relatively hot.

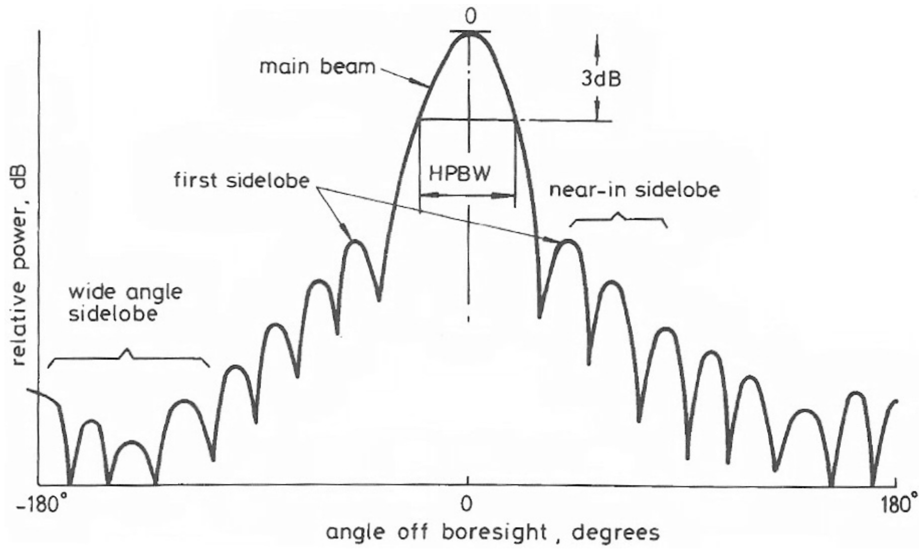


FIG. 2 CARTESIAN PLOT OF ANTENNA RADIATION PATTERN

A Cassegrain antenna system consists of a parabolic reflector and a *hyperbolic sub-reflector* sharing the same focal point. The primary feed which is generally a horn is located at the second focal point of the hyperbolic sub-reflector. This system provides more uniform illumination of the reflector without the spilling over, associated with the prime focus system thus resulting in better efficiency. Also it has lower cross-polarization. Since the primary feed is pointing away from the hot earth the antenna noise is inherently less. Further, because of the easy accessibility of the feed, LNA can be mounted close to the feed, reducing noise contribution from the feeder line loss. The Cassegrain system is widely used in large earth station installations.

The Gregorian antenna system consists of the main paraboloid and an *ellipsoidal sub-reflector*. The sub-reflector once again has two focal points, one of which is made to coincide with that of the main reflector and the other with the phase centre of the feed horn. The performance of the Gregorian system is similar in many respects to the Cassegrain. However it is less commonly used.

The performance of the Axi-symmetric configuration is affected by the blockage of the aperture by the feed and the sub-reflector assembly. The result is a reduction in the antenna efficiency and an increase in the side lobe levels. The *asymmetric or the offset configuration* can remove this limitation. Hence the trend is to use this configuration. The geometry of this configuration is however more complicated and hence difficult to implement in large earth stations. Cassegrain and Gregorian systems can be configured with offset feed arrangements.

6.1.1 Feed System

The primary feed system used in earth stations performs many functions. Depending on the type of earth station, these functions may be:

1. Illuminating the main reflector
2. Separating the transmit and receive bands

3. In dual polarized system separating/combining polarizations
4. Providing error signals for satellite tracking in system employing auto-track

A horn is antenna is commonly used as primary feed in microwave frequencies. A horn antenna is a wave guide which flares so that the free space impedance at the open end matches the impedance of the wave guide. The aperture can be rectangular or circular. Circular aperture horns, known as conical horns, are widely used as primary feeds in earth stations.

6.2 Tracking System

Tracking is essential when the satellite drifts by a significant fraction of the beam-width, as seen by the earth station. Before communication can be established it is necessary to 'acquire' a satellite. One method is to program the antenna to perform a scan around the predicted position of the satellite. The automatic tracking system is switched on when the received signal strength is sufficient to lock the tracking receiver to the satellite beacon. After acquisition, a satellite needs to be tracked continuously. This function is performed by an auto-track closed-loop system. The three main types of auto-track system which are commonly used for satellite tracking are: (1) Conical scan (2) Mono-pulse and (3) Step-track.

6.3 Low Noise Amplifier

With the advancements in Gallium Arsenide field-effect (GaAs FET) transistors the low noise amplifiers with GaAs FET can provide low noise temperature at low cost and are commonly used in modern earth stations. The typical noise temperatures are 75 K in C-band and 170 K in Ku-band for uncooled LNA.

6.4 High Power Amplifier

The two most commonly used high power amplifiers in large earth stations are the TWT amplifier and Klystron. TWTA can offer bandwidth of the order of 500 MHz and are capable of providing power up to 10 kW. Sometimes a linearizer is used to improve the linearity of the amplifier, especially if a multi-carrier operation is desired.

Klystrons are narrow band devices, typically offering bandwidths of the order of 40 MHz, tunable over the entire 500 MHz bandwidth. Maximum power is of the order of 3 kW. A multiple amplifier configuration becomes essential when a larger bandwidth is required. However Klystrons have a higher efficiency, longer tube life, lower cost and are simpler to maintain and operate than TWT amplifiers. TT&C uplink typically uses Klystron power amplifier.

7. TV RECEIVE ONLY EARTH STATION

The direct-to-home digital TV which has replaced the analogue system in recent times uses an offset feed parabolic reflector of 60 cm. in Ku-band. The system consists of an outdoor unit (ODU), an indoor unit (IDU) and an inter facility cable. The outdoor unit consists of the offset fed parabolic antenna and the low noise block converter (LNBC). The standard frequency range of the output is 950 to 1450 MHz. A low loss coaxial cable carries this signal to the indoor unit. The same coaxial cable is used to carry the DC power from indoor unit to the outdoor unit. The indoor unit receives the 950 to 1450 MHz signal and after demodulation feeds the audio-video signals to the TV set.

8. TRANSMIT-RECEIVE STATION WITH REDUNDANCY

Transmit-receive stations are required for telecommunication traffic including network TV. These stations are provided with redundancy. A duplicate or redundant unit is automatically switched into service to replace a corresponding unit that has failed. The terrestrial network is interfaced with the earth station system at the baseband level.

Most large earth stations need to track the satellite accurately because of their use of narrow beam width antenna. Typically the tracking accuracy of less than one-tenth of half power beam width is essential. The mono-pulse technique is therefore commonly used. The step-track technique is commonly used in medium-sized earth stations. The typical maximum slew speed of a large earth station is 2°/sec.

TT&C ground stations are usually dedicated stations and are located separately from the communications ground stations. In C Band 14m or 11m ground station antennas are used for TT&C functions such as spacecraft health data reception, acquiring the satellite at launch, tracking and ranging during orbit raising and during regular operation, maintenance of attitude and orbit and sending command signals to the satellite. In Ku Band 7m ground station antenna is employed for TT&C operations.

For both TT&C and communications earth stations, power supply system must provide power, ensuring that communication outages are not encountered. Battery banks and standby generators are used. When power failure occurs, the system switches to battery banks at the same time starting the generators. When the generators speedup and reach the required power level, the generators take over from the battery banks.

Air conditioning is essential for reliable operation of all large earth stations. The operation of the entire station is controlled from a control console. However many earth stations now have a capability of unattended operation.

REFERENCES

[1] M. Richharia, “Satellite Communication Systems” Macmillan Press Ltd. London, 1999
[2] Dennis Roddy, “Satellite Communications” McGraw-Hill, 2001
[3] B.G. Evans, “Satellite Communication Systems” 3rd Edition, IEE, London, UK, 1999

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

DATA COMMUNICATION AND NETWORK PROTOCOL

D. JOHN

1. INTRODUCTION

In the data communications world, data generally are defined as information that is stored in the digital form. The word 'data' is plural; a single unit of data is called a datum. Data communications is the process of transferring digital information (usually in binary form) between two or more points through the network. The data communications network has several components that need to be understood in order to appreciate its working. The components are data communications code, error control (error detection and correction), character synchronization and finally fundamental hardware that include computer-networking equipment such as line control unit, serial interfaces and data communications modem.

2. DATA COMMUNICATIONS CODES

Data communications codes are prescribed bit sequences used for encoding characters and symbols. Therefore data communications codes are called character codes, character sets, symbol codes or character languages. In essence, there are only three types of characters used in data communications codes: data link control characters, which are used to facilitate the orderly flow of data from a source to a destination; graphic control characters, which involves syntax or presentation of the data at the receive terminal; and alpha/numeric characters, which are used to represent the various symbols used for letters, numbers and punctuation in the English language.

The first data communications code that was widely used was *Morse code*. The Morse code uses three unequal-length symbols (dot, dash and space) to encode alpha/numeric characters, punctuation marks and an interrogation word.

The Morse code is inadequate for use in modern digital computer equipment because all characters do not have the same number of symbols or take the same length of time to send, and each Morse code operator transmits code at a different rate. Morse code also has an insufficient selection of graphic and data link control characters to facilitate the transmission and presentation of the data typically used in contemporary computer applications.

The three most common character sets presently used for character encoding are the Baudot code, the American standard code for information exchange (ASCII) and the extended binary-coded decimal interchange code (EBCDIC).

The Baudot code (pronounced baw-dough) sometimes called telex code was the first fixed-length character code developed for machines. This code is a fixed-length source code sometimes called a fixed-length block code. The Baudot code is a five-bit character code that was used primarily for low-speed teletype equipment, such as telex system and radio teletype (RTTY). The latest version of Baudot code is recommended by the CCITT as the International Alphabet No. 2.

The ASCII (pronounced as-key) has passed through several versions and the 1977 version has been recommended by the ITU as International Alphabet No. 5. ASCII is a standard character set for source coding the alpha/numeric character set that humans understand but computers do not(computers only understand 1s and 0s). ASCII is a seven-bit fixed-length character set. The eighth-bit is not part of the ASCII code but is generally reserved for error detection and is called the parity bit. ASCII is probably the code most often used in data communications networks today.

The Extended Binary-Coded Decimal Interchange Code (EBCDIC) is an eight-bit fixed-length character set developed in 1962 by IBM. With eight bits, 256 codes are possible, although only 139 of the 256 codes are actually assigned characters. The unspecified codes can be assigned to specialized characters and functions if need be. The name binary-coded decimal was selected because the second hex character for all letter and digit codes contain only the hex values from 0 to 9, which have the same binary sequence as BCD (binary-coded decimal) code.

Bar codes are the black-and-white striped stickers that appear on every consumer item these days. A bar code is a series of vertical black bars separated by vertical white bars (called spaces). The width of the bars and spaces represent binary 1s and 0s and combinations of bits identify specific items. There are several standard bar code formats. The bar codes are generally classified as being discrete, continuous or two dimensional.

3. ERROR CONTROL

A data communications circuit can be as short as a fraction of a meter or as long as several thousands of kilometers. Therefore errors will occur and it is necessary to implement error-control procedures. Transmission errors are caused by electrical interference from natural sources, such as lightning as well as from man-made sources such as motors, generators, power lines and fluorescent lights. Noise added in the channel also causes errors.

Data communications errors can be generally classified as single-bit, multiple-bits or burst. Single-bit errors affect only one character within a message. A multiple-bit error is when two or more non consecutive bits within a given data string are in error. A burst error is when two or more consecutive bits within a given data string are in error. Both multiple-bit errors and burst errors can affect one or more characters within a message. Error control can be divided into two categories – error detection and error control.

4. ERROR DETECTION

The most common error detection techniques are redundancy checking, which includes vertical redundancy checking, checksum, longitudinal redundancy checking and cyclic redundancy checking.

Vertical redundancy checking (VRC) is the simplest error detection scheme and is referred to as parity. With parity each character has its own error detection bit, called the parity bit. Since the parity bit is not part of the character, it is considered a redundant bit. An 'n' character message would have 'n' redundant parity bits. Therefore, the number of error detection bits is directly proportional to the length of the message. With character parity, a single parity bit is added to each character to make the total number of logic 1s in the character, including the parity bit, to be either an odd number (odd parity) or an even number (even parity). If one bit is in error the condition changes and it can be detected and the character is discarded as erroneous. The advantage of parity is its simplicity. The disadvantage is that if even number of bits are in error, the parity checker will not detect it as an error.

Checksum is another relatively simple form of error checking where each character has a numerical value assigned to it. The characters within a message are combined together to produce an error-checking character (checksum), which can be as simple as the arithmetic sum of the numerical values of all the characters in the message. The checksum is appended to the end of the message. The receiver replicates the combining operation and determines its own checksum. The receiver's checksum is compared to the checksum appended to the message to detect the transmission error.

Longitudinal redundancy checking (LRC) is a redundancy error detection scheme that also uses parity to determine if a transmission error has occurred within a message. It is therefore, sometimes called message parity. With LRC, each bit position has a parity bit. In other words, b_0 from each character in the message is XORed with b_0 from all other characters in the message. Similarly b_1, b_2 and so on are XORed with their respective bits from all the characters in the message. Essentially, LRC is the result of XORing the "character codes" that make up the message, whereas VRC is the XORing of the bits with in a single character. With LRC even parity is generally used, whereas with VRC odd parity is generally used. The LRC bits are computed in the transmitter and appended to the end of the message as a redundant character. In the receiver, the LRC is recomputed from the data and compared with the LRC appended to the message. If these two LRC characters are the same, most likely no transmission errors have occurred.

The group of characters that comprise a message is often called a block or frame of data. Therefore, the bit sequence for LRC is often called a block check sequence (BCS) or frame check sequence (FCS). LRC detects between 95% and 98% of all transmission errors. However, LRC will not detect transmission errors when even number of characters have an error each, in the same bit position.

Cyclic redundancy checking (CRC) is one of the most reliable error detection schemes, and can detect up to 95.5% of all errors. The commonly used code is CRC-16-CCITT. The basic idea of a CRC is that the transmitter and the receiver both agree that the numerical value sent by the transmitter will always be divisible by a predetermined number, e.g. 9. If the receiver gets a value which is not divisible by 9 then it knows that there has been an error. For example, if a value of 32 is to be transmitted then this value could be changed to 320 so that the transmitter can add to the least significant digit to make it divisible by 9. In this case the transmitter would add 4, making 324. If this transmitted value was to be corrupted in transmission then there would only be a 10% chance that an error would not be detected.

In the CRC-16-CCITT, the error correction code is 16 bit long and is a remainder of the data message polynomial $G(x)$ divided by the generator polynomial P(x) [$x^{16} + x^{12} + x^5 + 1$, that is, 10001000000100001] . The quotient is discarded and the remainder is truncated to 16 bits (frame check sequence-FCS). This is then appended to the message as a coded word.

The division does not use standard arithmetic division process. Instead of subtraction operation an XOR operation is used. This is a great advantage as a CRC only requires a shift register and a few XOR gates to perform the division.

The receiver and the transmitter both use the same generating function $P(x)$. If there are no transmission errors then the remainder will be zero.

The method uses the following steps:

1. Let $P(x)$ be the generator polynomial and $M(x)$ the message polynomial.
2. Let 'n' be equal to the highest exponent of the generator polynomial, $P(x)$.
3. Append 'n' zero bits on to the right hand side of the message so that it contains $m + n$ bits.
4. Using modulo–2 division, divide the modified bit pattern by $P(x)$. [Modulo-2 arithmetic involves XOR operations, i.e. 0 – 1 = 1, 1 – 1 = 0, 1 – 0 = 1 and 0 – 0 = 0]
5. The final remainder is added to the modified bit pattern and this is transmitted as the message. [The remainder consisting of 'n' bits is the frame check sequence (FCS), sometimes called block check sequence (BCS)]

There are four versions of generator polynomials $P(x)$ widely used. CRC-16 [$x^{16} + x^{15} + x^2 + 1$] and CRC-16-CCITT are popular for 8-bit characters in the US and Europe respectively. The CRC-16-CCITT detects all single and double bit errors, all odd number of bit errors, all burst errors of 16 or less, 99.997% of 17 bit error bursts and 99.998% of 18 bit and longer burst errors. For randomly distributed errors, it is estimated that the likelihood of this scheme not detecting an error is 10^{-14}, which equates to one undetected error every two years of continuous data transmission at a rate of 1.544 Mbps.

5. ERROR CORRECTION

Data communications network designers have developed two basic strategies for handling transmission errors: errors-detecting codes and error-correcting codes. Thus for we have discussed errors-detecting codes. Error-correcting codes include sufficient extraneous information along with each message to enable the receiver to determine when an error has occurred and which bit is in error. In data communications, there are two primary methods for error correction: retransmission and forward error correction (FEC). In retransmission method, when the received message is in error, the receiver requests for retransmission and the message or a portion of the message is retransmitted.

Forward error correction (FEC) schemes detect and correct bit error. A popular FEC code is Hamming code. The error correction bits are known as Hamming bits, the number that requires to be added to a character is determined by the expression:

$$2^n \geq m + n + 1$$

Where, m is equal to the number of bits in each data character and n equal to the number of Hamming bits.

Hamming bits insert into the message character in a desired way. Typically, they are added in positions of powers of 2, i.e. the 1st, 2nd, 4th, 8th, 16th, and so on, bit positions. For example to code the character 00111001, starting from the right-hand side, the Hamming bit would be inserted in to 1st, 2nd, 4th and 8th bit positions. Then each position where '1' appears in the coded data-character is represented as a binary value. Next each position value is XORed with each other. The result is Hamming code. In this example:

Table 1. *Transmitted block*

Bit Position	12	11	10	9	8	7	6	5	4	3	2 1
Position value1100	1011	1010	1001	1000	0111	0110	0101	0100	0011	0010	0001
Data bit	D8	D7	D6	D5		D4	D3	D2		D1	
Hamming bit H1					H8				H4		H2
Transmitted block	0	0	1	1	0	1	0	0	1	1	1 1
Codes			1010	1001		0111				0011	

Table 2. *Hamming bit calculation prior to transmission*

Position	Code
10	1010
9	1001
7	0111
3	0011
XOR = H8H4H2H1 Hamming bits	0111

The entire block that is transmitted is 001101001111. Suppose now that data bit 3, in bit position 6, sustains an error and is changed from '0' to '1', the received block is 001101101111. The received Hamming code is still 0111. The receiver performs an XOR of the Hamming code and all of the bit position values for non-zero data bits, with a result of 0110. The non-zero result detects an error and indicates that the error is in bit position 6. These are shown in the following tables. (If the result of the XOR operation at the receiver is zero then no error has occurred).

The only stipulation on the placement of the Hamming bits is that both the sender and the receiver must agree on where they are placed. The code just described is known as single-error-correcting (SEC) code. A variation of this is a single-error-correcting, double-error-detecting

Table 3. *Received block*

Bit Position	12	11	10	9	8	7	6	5	4	3	2	1
Position value	1100	1011	1010	1001	1000	0111	0110	0101	0100	0011	0010	0001
Data bit	D8	D7	D6	D5		D4	D3	D2		D1		
Hamming bit					H8				H4		H2	H1
Received block	0	0	1	1	0	1	1	0	1	1	1	1
Codes			1010	1001		0111	0110			0011		

Table 4. *Hamming bit calculation after reception*

Position	Code
Hamming	0111
10	1010
9	1001
7	0111
6	0110
3	0011
XOR = Syndrome	0110

(SEC-DED) code. Such codes require one additional bit compared with SEC codes. The extra bit is a parity bit over the entire code block. It may be noted that FEC improves the BER performance of the data link allowing a reduction in E_b/N_0. This reduction is referred to as coding gain.

6. CHARACTER SYNCHRONIZATION

Character synchronization involves identifying the beginning and the end of a character within a message. When continuous string of data is received, it is necessary to identify which bit belongs to which character and which bits are the MSBs and LSBs of the character. In data communications circuits, there are two formats commonly used to achieve character synchronization: asynchronous and synchronous.

With asynchronous serial data for each character there is a start bit which is usually logic 0 and stop bits which are two numbers of logic 1s. The start bit is placed ahead of the LSB and the stop bits appear after MSB and parity bit. The character synchronization is achieved using this mechanism.

Synchronous data generally involves transporting serial data at relatively high speeds in groups of characters called blocks of frames. Therefore, synchronous data are not sent in real time. Instead, a message is composed and then the entire message is transmitted as a single entity with no time lapses between the characters. With synchronous data, rather than framing each character independently with start and stop bits, a unique sequence of bits called synchronizing (SYN) characters,(two characters i.e., 16 bits) is transmitted at the beginning of each message.

7. SERIAL AND PARALLEL INTERFACES

To ensure an orderly flow of data between the data terminal equipment (DTE) and the modem (DCE - data communications equipment), a serial interface is placed between them. This interface coordinates the flow of data, control signal and timing information between the DTE and the DCE. In 1962 the Electronic Industries Association (EIA) in an effort to standardize interface between DTE and DCE agreed on a set of standards called the RS-232 specification. The latest version of this standard being RS-232 D introduced in 1987.

The RS-232 specifications identify the mechanical, electrical, functional and the procedural description for the interface between the DTE and DCE. The RS-232 interface uses a 25-pin connector. There is also a 9-pin version of RS-232 interface designed to transport asynchronous data between DTE and a DCE or between two DTEs. On the other hand 25-pin version is designed for transporting synchronous or asynchronous data between a DTE and a DCE. The RS-232 interface is similar to the combined CCITT standards V.28 (electrical specifications) and V.24 (functional description) and is designed for serial transmission of data up to 20 kbps for a distance of approximately 50 feet.

Contemporary data rate have exceeded the capabilities of the RS-232 interface. Therefore, in 1977 EIA introduced the RS-449 specification with the intention of replacing RS-232 interface. The RS-449 interface uses a 37-pin connector that provides more functions, faster data transmission rates and greater distance capabilities. However, the industry never embraced the RS-449 standard. Consequently, in 1987, the EIA introduced the RS-530 standard which was intended to operate at data rate from 20 kbps to 2Mbps using the same 25-pin connector used by the RS-232 interface.

The RS-530 interface does not include electrical specifications. The electrical specifications used with the RS-530 are specified by either the RS-422A or RS-423A standard. The RS-422A standard specifies a balanced interface cable that will operate at bit rate up to 10 Mbps and span distance up to 1200 m. This does not mean however that 10 Mbps can be transmitted 1200 m. At 10 Mbps the maximum distance is approximately 15 m., and 90 kbps is the maximum bit rate that can be transmitted 1200 m. The RS-423A standard specifies an unbalanced interface cable that will operate at a maximum line speed of 100 kbps and span a maximum distance of 90 m.

Parallel interfaces transfer data between two devices 8 or more bits at a time. That is, one entire data word is transmitted (received) at a time as opposed to one bit at a time as with serial interfaces. Parallel transmission is sometimes referred to as serial-by-word transmission. One obvious advantage of parallel transmission is data are transmitted much faster than with serial transmission. Here there is a transmission port for each bit of the word. For example, a system utilizing 8-bit words would have 8 separate communications channels between the transmitter and receiver. Another advantage of parallel transmission is most computer terminals and peripheral equipments process data internally in parallel. A disadvantage of parallel transmission is higher costs for transmission line, especially when there are long distances between transmitter and receiver. Therefore, parallel interfaces are commonly used to transfer data between two or more devices that are located in close proximity to each other, such as, a computer and a printer. The connector used is a 36-pin Amphenol connector.

7.1 IEEE 488 Bus

The IEEE 488 bus uses eight bidirectional data lines connected in parallel to interface up to 15 remote devices (usually computer-controlled test equipment). Because the various pieces of equipment interfaced with the IEEE 488 are located in proximity with each other, the interface was designed for a maximum distance between adjacent devices of 7 feet and the maximum length of the entire bus is about 65 feet. Devices are connected to the bus by a 24-pin ribbon connector. With this bus, the controller (the device which is in charge of the bus) determines, which device can transmit data and which devices are receivers.

8. DATA COMMUNICATIONS HARDWARE

Data terminal equipment (DTE) can be any binary digital device that generates, transmits, receives or interprets data messages. In fact the DTE is where information originates or terminates. DTEs contain hardware and software necessary to establish and control communication between endpoints; however, DTEs seldom communicate directly with other DTEs. Examples of DTEs include video display terminals, printers and personal computers.

Data communications equipment (DCE) is a general term used to describe equipment that interfaces data terminal equipment to a transmission channel, such as a digital T1 carrier or an analog telephone circuit.

Line control unit (LCU) is a DTE. The LCU performs parallel-to-serial and serial-to-parallel conversion. It also houses the circuitry that performs error detection and correction. In addition, data-link characters are inserted and deleted in the LCUs. With LCUs a single special purpose integrated circuit performs many of the fundamental data communications functions. This integrated circuit is called a universal asynchronous receiver/transmitter (UART), if it is designed for asynchronous data transmission, a universal synchronous receiver/transmitter (USRT), if it is designed for synchronous data transmission and a universal synchronous/asynchronous receiver/transmitter (USART), if it is designed for either asynchronous or synchronous data transmission. All three types of circuits specify general purpose integrated circuit chips located in an LCU that allows DTE to interface with DCE.

9. THE TELEPHONE NETWORK

In its simplest form data communication is the transmission of digital information between two data terminal equipment (DTEs). The DTEs may be separated by a few feet or several thousand miles. One convenient method of transmitting digital information is by using the existing public telephone network (PTN). The PTN was designed and most of it constructed long before the advent of large scale data communication. The PTN was intended to be used for transferring voice telephone signal, not digital data. Therefore, to use PTN for data, the data from DTE must be converted to a form suitable for transmission over analog carrier systems. This is done by data communications equipment (DCE) which is a modem.

The telephone companies that operate PTN offer two categories of service: direct distance dialing (DDD) and private line. The DDD (a dial-up circuit) originally included only long distance telephone call without the assistance of an operator. Now it includes the entire public switched network; it includes any service associated with a telephone number. Private line services are dedicated to a single user.

The telephone circuit in general, may be cable pairs or carrier systems, and the information maybe transferred on a coaxial, metallic, microwave, optic fibre or satellite communication system. The telephone companies offer a wide assortment of digital channels ranging from basic 4-kHz voice-band circuit to wideband (30 MHz) microwave channels that are capable of transferring high resolution video signals. However, here we are primarily concerned with only voice-band circuits.

Digital information sources, such as personal computers, communicate with each other using the POTS (plain old telephone system). At the transmitting end the telephone interface converts the digital data from the transceiver to analog electrical energy which is transported through the telephone line. At the receiving end the same telephone interface converts the analog electrical energy received from the telephone line to digital data. For two way communication the source and the destination are interchangeable. Therefore it is more appropriate to describe a data communications system as connecting two endpoints (sometimes called nodes) through a common communications channel. All end points must have three functional components: data terminal equipment (DTE), data communication equipment (DCE) and a serial interface.

10. DIGITAL SUBSCRIBER LINES

With standard voice only telephone lines, the "last-mile" connection between the telephone customer and the central office consists of a twisted-wire pair that feeds to a plain old telephone (POT) line card located at the central office. The POT card interfaces the voice line to a 64 kbps encoded voice signal, connected to the rest of the telephone network. While the twisted pair of wires maybe capable of carrying signal up to 30 MHz, the POT card is designed for voice signal and thus the bandwidth is limited to 3.2 kHz. If the computer interface is made via a modem the data rate is limited to 30 kbps or, in some cases, up to 56 kbps. For interactive video and other high speed services we want to consider a digital subscriber line (DSL) that can handle increased data rates. DSL is primarily a set of standards that define the central office interface. There are several DSL standards, they are ADSL (asymmetric digital subscriber line), HDSL (high bit rate digital subscriber line), ISDN (integrated services digital network) etc.

ADSL is an FDM system where the existing twisted pair cable supports three services: (a) POTs (b) 640 kbps digital data from subscriber to central office (upstream) and (c) 6.144 mbps digital data from central office to subscriber (downstream).

ISDN is another service that supports end to end digital connectivity. It consists of a single twisted cable pair that allows for data rate of at least 128 kbps and enables integration of voice, video and other data sources.

11. DATA MODEM

Telephone circuits were designed for transporting analog voice signals within a bandwidth of approximately 300 Hz to 3000 Hz. In addition, telephone circuits often included amplifiers and other analog devices that could not propagate digital signals. Therefore, voice-band data modems were designed to communicate with each other using analog signals that occupied the same bandwidth used for standard voice telephone communications. Data communications modems designed to operate over the limited bandwidth of the public telephone network are called voice-band modems. In the transmitter (modulator) section of the modem, digital signals are encoded into an analog carrier. The digital signals modulate the carrier, producing digitally modulated analog signals that are capable of being transported through the analog communications medium. Therefore the output of the modem is an analog signal that is carrying digital information. In the receiver section of the modem, digitally modulated analog signals are demodulated. Therefore, modem receivers (demodulators) simply extract digital information from digitally modulated analog carriers.

The most common modems available are the ones intended to be used to interface DTEs through a serial interface to standard voice-band telephone line and provide reliable data transmission rates from 300 bps to 56 kbps. The most sophisticated modems (sometimes called broadband modems) that are capable of transporting data at much higher bit rates over wideband communications channels are also available. Broadband modems operate using different set of standards and protocols than voice- band modems.

Modems are generally classified as either asynchronous or synchronous and use one of the following modulation techniques: ASK, FSK, PSK or QAM. With synchronous modem clock information is recovered in the receiver but with asynchronous modem, it is not. Asynchronous modems generally use ASK or FSK and are restricted to relatively low speed applications, generally below 2.4 kbps. Synchronous modems use PSK and QAM and are used for medium and high speed applications (up to 57.6 kbps). High data rate transmission up to around 56 kbps can be achieved over standard telephone channels using an encoding technique called trellis code modulation (TCM).

It may be noted that these modems are an alternative to DSL modems. The following table gives a selected list of Bell and ITU modem standards along with their respective rates and modulation methods. Shannon limit for standard voice grade telephone lines is 37 kbps. Improvements in modulation and coding have enabled modems so that they can get relatively close to this limit. Most commercial modems have an option so that at the initial connection and during the session the modem will test the telephone line's S/N to adjust the modem's data rate.

Table 5. *Selected telephone line Data Modems*

Model	Bit rate	Modulation	Transmitter frequency (Hz)
Bell 103 A	300 bps	FSK	1070-1270/2025-2225
Bell 212 A	300/1200 bps	DPSK	1200/2400
V.32	9600 bps	16-QAM or 32-QAM with TCM	1800
V.32 bis	14.4 kbps	128-QAM with TCM	1800
V.34	28.8 kbps	960-QAM with TCM	1800
V.34 bis	33.6 kbps	1664-QAM with TCM	1800
V.90	56/33.6 kbps	PCM	-

Other types of modems for computer communication include cable modems for communication via the cable TV network, LAN modems, wireless modems and cellular telephone modems. Cable TV systems with their bandwidths of 300 MHz promise data rates in the Gbps range.

12. ISDN

The internationally agreed definition of ISDN is a 'network evolved from the telephony integrated digital network (IDN) that provides end to end digital connectivity to support a wide range of services, including voice and non-voice services, to which the users have access by a limited set of standard multi-purpose customer interfaces'.

The above definition of ITU-T makes a number of points:

1. ISDN requires a digital network.
2. This digital network is not one between exchanges, but it extends to customers.
3. It provides not only telephony, but a variety of services.
4. A customer does not require a separate interface for each service provided by ISDN, but can use all services via one access point or via a limited set of standard multi-purpose customer interfaces.

ISDN will remove the need for telecommunication customers to have separate links to telephone and low speed networks. Under ISDN a single physical connection is provided to a customer's premises from the digital exchange and a range of services can be made available from it. Furthermore, if required the services may be used simultaneously, because action is not restricted to each individual service in turn.

ISDN is a TDM digital telephone network that integrates voice, video, computer and other data sources. There are two ISDN structures: Basic Rate Interface (BRI) and Primary Rate Interface (PRI).

The BRI consists of two *B* (bearer) 64-kbps channels and a 16-kbps *D* (delta) channel. It is also referred to as $2B + D$. BRI is used primarily by residential and small business customers. In North America, Japan and Korea, PRI has twenty three 64-kbps *B* channels and one 64-kbps *D* channel, or it is $23B + D$. The *B* and *D* channels are full duplex (both directions) and carry voice, computer data and so on, while the *D* channel is generally used to transmit control, signaling, telemetry, or other connection information. With BRI, the $2B$ channels could carry two 64-kbps PCM voice messages or one 64-kbps voice message and up to six multiplexed 9600 bps data streams. Because of faster data rate and increased data channels ISDN is an alternative to the modem for interfacing home computers to telephone lines.

The European and rest of the world standards for PRI consist of $30B + D$ 64-kbps channels, where each frame is 256 bits long. With overheads added the output rate is 2.048 Mbps.

ISDN also has high capacity *H* channels for information in excess of 64-kbps. These include video, high resolution graphic, high fidelity audio, HDTV and so on. $H11$ (1.536-Mbps) and $H12$ (1.92-Mbps) are the high capacity ISDN channels used in North America and Europe respectively. Still higher capacity *H* channels are also available.

13. LOCAL AREA NETWORK (LAN)

LANs are typically privately owned data communications network in which 10 to 40 computer users share data resources with one or more file servers. LANs use a Network Operating System (NOS) to provide two-way communication at bit rates typically in the range of 10 Mbps to 100 Mbps and higher between a large variety of data communications equipment within a relatively small geographical area, such as in the same room, building or building complex. Most LANs link equipments that are within few miles of each other or closer. Because the size of most LANs is limited, the longest transmission time is bounded and known by everyone using the network. Therefore LANs can utilize configurations that are otherwise not possible.

LANs are designed for sharing resources between a wide range of digital equipments, including PCs, workstations, FAX machines, CD-ROM drive, scanners and printers. The resources shared can be software as well as hardware. When PCs are in the LAN there is no need to have dedicated individual printers for the PCs. To print a document, a printer simply sends the information over the network to the server. The server organizes and prioritizes the document and then sends them, one document at a time to the common printer. Meanwhile the PCs are free to continue performing other useful tasks.

LANs allow people to send and receive messages and documents through the network much quicker than they could send through a paper mail system. E-mail is a communications system that allows users to send messages to each other through their computers. To send an e-mail, the PC sends its address and message along with destination address to the server. The server effectively relays the message to the destination PC if they are subscribers to the same e-mail system/network. If the destination PC is busy or not available, for whatever reason, the server stores the message sends it later. The server is the only computer that has to keep track of the location and address of all other PCs on the network. To send e-mail to subscribers of other network, the server relays the message to the server on the destination user’s network, which in turn relays the mail to the destination PC.

14. WIDE AREA NETWORK (WAN)

WAN is the oldest type of data communications network that provides relatively slow speed, long distance transmission of data, voice and video information over relatively large and widely dispersed geographical areas, such as a country or the entire continent. WANs typically interconnect cities and states. They typically operate bit rates from 1.5 Mbps to 2.4 Gbps and cover a distance of 100 to 1000 miles. WAN may utilize both public and private communications systems to provide service over an area that is virtually unlimited. WANs are generally through service providers and normally come in the form of leased-line or circuit-switched technology. Often WANs interconnect routers in different locations.

15. ISO-OSI SEVEN LAYER ARCHITECTURE

Open System **I**nterconnection (OSI) was developed in the late 1970s by the International Standard Organizations (ISO) in association with International Telecommunications Union (ITU) to be a model for the standardization of computer networks. 'Open' means that any two systems can communicate if they adhere to the OSI model. The OSI model consists of seven layers. Each layer consists of specific protocols for communicating. Layers 1 to 3 are part of network connection while layers 4 to 7 are part of data originator source or recipients (destination). This architecture allows any one layer to be modified or improved without affecting other layers. The layers are described as below.

15.1 Physical Layer

This is where the bit stream information is inserted into the network. The physical layer includes electrical and mechanical standards (i.e., electrical signals, connectors etc.). The physical layer specifies the transmission medium and the transmission mode (simplex, half duplex or full duplex). The physical layer includes carrier system used to propagate the data signals between points in the network. Examples are metallic, optical fibre cable, wireless arrangement such as microwave, satellites etc.

15.2 Data Link Layer

Data link structures the data into frames so that each one includes bits for synchronization, error detection/correction and the beginning and end of the frame plus the source's and destination's addresses. It also sends and receives acknowledgement if the frame is received correctly. For example, if a burst error occurred and a frame was corrupted the software in the data link layer would cause the source to retransmit the frame. If an acknowledgement signal was not received by the source, the source would then attempt to send a duplicate frame. Therefore, data link layer would minimize the problem of duplicate frames. The data link layer along with other layers manages the data flow, so that a source with high data rate does not overwhelm a slow destination. A data link layer example is CCITT 1.441/Q.921 ISDN Data Link Protocol (LAPD-Link Access Protocol -D).

15.3 Network Layer

Network layer decides how the data is to be routed from node to node, and is used to manage data flow if there is network congestion. This layer may also be involved in sending and receiving acknowledgements. Network layer provides the upper layer of the hierarchy independence from the data transmission and switching technology used to interconnect systems. It accomplishes this by defining mechanisms in which messages are broken into smaller data packets and routed from a sending node to a receiving node within the data communications network. The network layer also typically provides the source and destination network addresses, subnet information and source and destination node addresses.

15.4 Transport Layer

The transport layer either breaks the message data from the session layer into packets (or some other unit) or takes the packet from the network layer and connects them to form a message.

15.5 Session Layer

Session layer establishes connections between computers or reestablishes connection in the event of a failure. It also links applications such as, audio and video services and authenticates the network users.

15.6 Presentation Layer

Presentation layer performs data formatting function such as code conversion (e.g. ASTII to EBCDIC), encryption/decryption and data compression/decompression.

15.7 Application Layer

Application layer is a user interface to the OSI system and includes such services as file transfer programs, airline reservation data, and e-mail or web browsers.

In practical systems some or all of these layers can be integrated at various degrees into a single hardware device. For example, some PCs have a plug-in modem card (i.e. stand-alone modem) that performs the function of layers 1 to 3, while rest of the computer performs the functions of layers 4 to 7. On the other hand, some PCs have a built-in modem that share the computer's CPU at a lower priority to perform functions of layers 1 to 3, using a process known as 'cycle stealing'. Of course, with this system if there is a great deal of interaction with the network, the system slows down. For connections for the most popular LANs, the standards/ Protocols are Ethernet standards for physical and data link layers and IP (routing data packets) for network layer and TCP or UDP for transport layer.

16. CONCLUSION

In this article, under data communication, we had an over view of data communication codes, error detection and error correction techniques, character synchronization, serial parallel/ interfaces, telephone network as a medium for data communication, DSL, ISDN, LAN and WAN. The functions of the seven layers of the OSI protocol for exchange of information between computers and networks were explained.

REFERENCES

[1] M.P. Clark, "Networks and Telecommunications – Design and Operations" John Wiley Sons, 1997

[2] W. Buchanan, "Applied Data Communications and Networks" Chapman & Hall, London, 1996

[3] Wayne Tomasi, "Advanced Electronic Communications Systems" Pearson Education, Inc, 2004

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

DIGITAL COMMUNICATION

D. JOHN

1. INTRODUCTION

In the world around, the information signals that we encounter are mostly in analog form. Analog signals have continuously varying amplitude with respect to time; the amplitude can take any possible value. However digital signals have discrete values; coded numbers are used for transmitting information. The term digital communications covers a broad area of communication techniques and could mean entirely different things in different contexts. Here we are concerned with systems where relatively high frequency analog carriers are modulated by relatively low frequency digital information signals and systems involving the transmission of digital pulses. In recent times, traditional electronic communications systems that use conventional analog modulation techniques such as, AM, FM and PM are being, rapidly, replaced with more modern digital communications systems which offer several outstanding advantages.

Some of the advantages of digital communication over analog communication are as below:

1. Digital communication is more rugged than analog communication because it can withstand channel noise and distortion much better.
2. The greatest advantage of digital communication over analog communication, however, is the viability of regenerative repeaters in the former.
3. Digital hardware implementation is flexible and permits the use of microprocessors, mini-processors, digital switching and large scale integrated circuits.
4. Digital signals can be coded to yield extremely low error rates and high fidelity as well as privacy.
5. It is easier and more efficient to multiplex several digital signals.
6. Digital communication is inherently more efficient than analog in realizing the exchange of SNR for bandwidth.
7. Digital signal storage is relatively easy and inexpensive. It also has the ability to search and select information, from distant electronic storehouses.
8. Reproduction with digital messages is extremely reliable without deterioration. Analog messages, such as photocopies and films, lose quality at each successive stage of reproduction.

9. The cost of digital hardware continues to halve every two or three years, while performance or capacity doubles, over the same period. Analog technologies for messages, sound, films do not come down in cost.

In decimal numbering system, that we are familiar with, we have ten digits. Using the ten digits we represent any value or number. On the contrary, in binary system only two digits, '1' and '0' are used to represent any value or number. The first mathematician, who recognized that any number can be represented merely by the use of 1 and 0, was Gottfried Wilhelm Leibniz (1646-1716). For digital communication, binary numbering system is used. The advantage being only two voltage levels which are easily distinguishable are needed for carrying all the information.

2. PULSE-CODE MODULATION (PCM)

PCM is the most useful and widely used of all the pulse modulations. Before pulse-code modulation of an analog signal, the signal is sampled and quantized.

2.1 Sampling

Samples of the signal are taken at least at Nyquist rate. If the highest frequency component of a signal is B Hz, the signal can be completely recovered from the samples of the signal taken at a minimum rate of 2B. This rate is called Nyquist sampling rate. The sampling interval, T_s=1/2B is called the Nyquist interval. The original signal is recovered by passing the samples through an LPF with cut off at B Hz.

2.2 Quantization

Analog signal has a continuous range of amplitude. Therefore the samples of the signal have continuous range of amplitudes. This means that within the finite limit of amplitude range, infinite-number of amplitude levels are possible. However it is not necessary to transmit the exact levels of the samples. Discrete levels can be used for successful transmission of the signal. If we assign discrete amplitude levels, with sufficiently close spacing, we can obtain the approximated signal, reconstructed from the discrete samples, practically indistinguishable from the original signal. The process of transforming the 'sample amplitude' of a message signal, at the sampling instant, into a discrete amplitude-level, taken from a finite set of possible amplitude-levels, is defined as the amplitude quantization.

Consider the analog signal $m(t)$ in the range ($-m_p$, m_p), which is partitioned into L sub-intervals, each of magnitude *Äv* = $2m_p/L$. Each sample amplitude is approximated by the midpoint value of the sub-interval in which the sample falls. Each sample is now approximated to one of the L numbers. We can represent the L values of the signal into binary signal by using pulse coding. As an example, if we take L to be equal to 16 then a code formed by binary representation of 16 decimal digits from 0 to 15 is known as natural binary code. Each of the samples in this example is encoded by 4 bits to obtain the PCM signal.

The telephone speech signal has a bandwidth of 3.4 kHz, and this signal is sampled at a rate of 8000 samples per second (8 kHz), which is higher than the Nyquist rate to avoid unrealizable filter requirements for reconstruction. Each sample is finally quantized into 256 levels (L = 256) which requires a group of 8 binary pulses to encode each sample ($2^8 = 256$). Thus the telephone signal requires $8 \times 8000 = 64000$ binary pulses per second.

The compact disc (CD) is a recent application of PCM. This is a high-fidelity situation, requiring audio signal bandwidth to be 15 kHz. Although the Nyquist sampling rate is only 30 kHz, the actual sampling rate of 44.1 kHz is used for reason mentioned above. The signal is quantized into a rather large number of levels (L = 65,536) to reduce the quantizing error. The binary coded samples are then recorded on the CD. This scheme requires a group of 16 binary pulses.

Since the sample value is approximated by the midpoint of the sub interval (of height Δv) in which the sample falls, the maximum quantization error is $\pm \Delta v/2$. The mean square quantizing error can be shown equal to $(\Delta v)^2 / 12$.

In essence, in PCM, a message signal is represented by a sequence of coded pulses, which is accomplished by representing a signal in discrete form in both time and amplitude. The basic operations performed in the transmitter of a PCM system are sampling, quantizing and encoding. The low pass filters prior to sampling is included to prevent aliasing of the message signal. The quantizing and encoding operations are usually performed in the same circuit, which is called an analog-to-digital converter (ADC). The basic operations in the receiver are regeneration of the impaired signals, decoding and reconstruction of the train of quantized samples. Regeneration also occurs at intermediate points along the transmission path as necessary. When Time Division Multiplexing (TDM) is used, it becomes necessary to synchronize the receiver to the transmitter for the overall system to operate satisfactorily.

If the 256 quantization levels of the telephone speech signal mentioned above are equally spaced over the amplitude range of the analog signal, then the low amplitude signals would incur far greater percentage of quantization errors (and thus distortion) than higher amplitude signals. For this reason, quantization levels are not linearly spaced but instead are more densely packed around zero amplitude level. This gives better signal quality in the low amplitude range and a more consistently clear signal across the whole amplitude range. Two particular sets of quantization levels are in common use for speech signal quantization. They are called the A-Law code and Mu-Law code. These two laws differ in the actual amplitude values chosen as their respective quantum levels; in all other aspects they are equal. The A-Law code is a European standard for speech quantization and the Mu-Law code is used in North America.

3. TIME DIVISION MULTIPLEXING

One advantage of using pulse modulation is that it permits the simultaneous transmission of several signals on a time sharing basis, using TDM. Because a pulse-modulated signal occupies only a part of the channel time, we can transmit several pulse modulated signals, on the same channel by interweaving them. Figure 1 shows TDM of two pulse amplitude modulated (PAM) signals. In this manner we can multiplex several signals on the same channel by reducing pulse widths. Another method of transmitting several base band signals simultaneously is frequency division multiplexing (FDM). In FDM various signals are multiplexed by sharing the channel bandwidth. The spectrum of each message is shifted to a specific band not occupied by any other signal. In a way, TDM and FDM are the duals of each other.

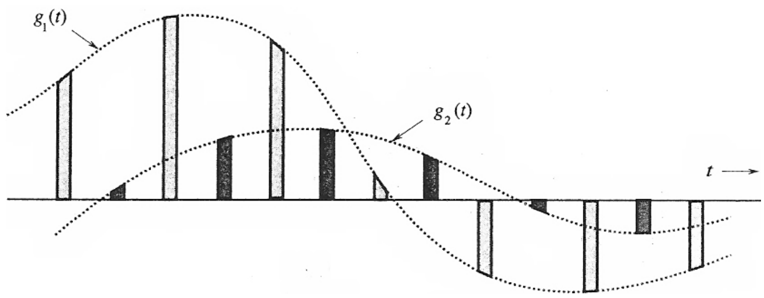


FIG. 1 TIME-DIVISION MULTIPLEXING OF TWO SIGNALS

In the TDM system, each input message signal, of several sources is first restricted in bandwidth by a low pass anti-aliasing filter to remove the frequencies that are nonessential to an adequate signal representation. The low pass filter outputs of various sources are then applied to a commutator, which is usually implemented using electronic switching circuitry. The commutator takes a narrow sample of each of the several input messages at the rate of f_s that is slightly higher than *2B*, where *B* is the cut off frequency of the anti-aliasing filter. The commutator also sequentially interleaves the samples from the various sources inside the sampling interval T_s. The multiplexed signal is then applied to a pulse modulator, the purpose of which is to quantize the sample and encode. At the receiving end, pulse demodulation, which is the reverse of the above process, is done to obtain narrow samples. These pulses are passed through the decommutator to sort out the samples of the various channels. The individual channel is passed through the respective low pass, reconstruction filter to obtain the output message. It is clear that the use of TDM introduces a bandwidth expansion by a factor equal to the number of message signals (N), because the scheme must squeeze N samples derived from N independent message sources into a time slot equal to one sampling interval.

3.1 T- Carrier System

T- Carriers are used for transmission of PCM encoded time division multiplexed digital signals. In addition, T-Carriers utilize special line-coded signals and metallic cables that have been conditioned to meet the relatively high bandwidths required for high-speed digital transmissions. Digital signals deteriorate as they propagate along a cable due to power loss in the metallic conductor and the low power filtering inherent in parallel-wire transmission lines. Consequently regenerative repeaters are placed at periodic intervals. The distance between the repeaters depend on the transmission bit rate and line-coding techniques. The regenerators clean up the digital bit stream. The signal to noise ratio at the output of a regenerator is exactly what it was at the output of the transmit terminal. That is the SNR does not deteriorate as the digital signal propagates through a regenerator; in fact, a regenerator reconstructs the original pulses with the original SNR.

T- Carrier system is a North American, AT&T system used for voice channel time-division multiplexing after Mu-Law encoding. The details of T carrier *hierarchy* are as follows:

The T1 carrier system utilizes PCM and TDM techniques to provide short-haul transmission of 24 voice band signals. The lengths of T1 carrier systems range from about 5 miles to 50 miles.

T1 carriers use regenerative repeaters placed every 6000 feet. The transmission medium for T1 carriers is either a 19 gauge or 22 gauge wire pair. The output data rate is 1.544 Mbps.

It may be noted that simply time-division multiplexing 24 voice-band channels does not in itself constitute a T1 carrier system. The output of the multiplexer is a multiplexed first-level digital signal. The system does not become a T1 carrier until it is line encoded and placed on special conditioned cable called T1 line. Line coding or transmission coding is a process by which the output of a multiplexer consisting of logic '1's and '0's is coded into electrical pulses for the purpose of transmission over the channel. There are many possible ways of assigning pulses to the digital data. Some examples of the line codes are on-off, polar, bipolar, etc. The T2 carrier system utilizes PCM to time-division multiplex 96 voice-band channels into a single 6.312 Mbps data signal for transmission up to 500 miles over a special low capacitance cable. This system is equivalent to multiplexing four numbers of T1 inputs.

The T3 carrier time-division multiplexes 672 PCM encoded voice-band channels for transmission over a single metallic cable. The transmission rate for T3 signal is 44.736 Mbps. This system is equivalent to multiplexing seven numbers of T2 inputs. T4 carrier time-division multiplexes 4032 PCM coded voice-band channels for transmission over a single coaxial cable up to 500 miles. This handles transmission data rate up to 278.176 Mbps. This system is equivalent to multiplexing six numbers of T3 inputs. There is also the T5 carrier system that is in use where in the system time-division multiplexes 8064 PCM coded voice-band channels and transmits them at a 560.6 Mbps rate over a single coaxial cable.

There are in fact, three basic hierarchies of transmission rates which have been standardized for international use; the other two are European and Japanese TDM systems. The European TDM system standardized by CCITT is E-carrier system, where 30 analog channels are A-Law encoded in to 64 kbps digital format and multiplexed in to single 2.048 Mbps E1 digital line system. In this hierarchy, groups of E1 circuits may be bundled onto higher capacity E3 links between telephone exchanges or countries. In practice, only E1 (30 circuit) and E3 (480 circuit) versions are used. The third TDM system used in Japan has Mu-Law encoding applied. Table-1 compares all these digital multiplexing hierarchies. The various hierarchies are incompatible and therefore inter working equipment is required for international links.

The high speed video channels in use are H0 (384 kbps), H11 (1536 kbps), and H12 (1920kbps); these correspond to 6×64 kbps, 24×64 kbps and 30×64 kbps TDM Channels

Table 1. *Digital multiplexing hierarchies*

Level	*AT&T*		*CCITT*		*Japan*	
	No. of voice-band I/P	*O/P rate Mbps*	*No. of voice-band I/P*	*O/P rate Mbps*	*No. of voice-band I/P*	*O/P rate Mbps*
I	24	T1 = 1.544	30	E1 = 2.048	24	J1 = 1.544
II	$24 \times 4 = 96$	T2 = 6.312	$30 \times 4 = 120$	E2 = 8.448	$24 \times 4 = 96$	J2 = 6.312
III	$96 \times 7 = 672$	T3 = 44.736	$120 \times 4 = 480$	E3 = 34.368	$96 \times 5 = 480$	J3 = 32.064
IV	$672 \times 6 = 4032$	T4 = 278.176	$480 \times 4 = 1920$	E4 = 139.264	$480 \times 3 = 1440$	J4 = 97.728

respectively. All the three bit rates may be supported by an E1 system; however T1 or J1 system can support only the first two high speed video channels.

4. DIGITAL RADIO SYSTEM

The property that distinguishes the digital radio system from a conventional AM, FM or PM radio system is the nature of the modulating signals. Both digital and analog radio systems use analog carriers; however, with analog modulation the modulating signal is analog and with the digital modulation the modulating signal is digital. It may be noted that with both analog and digital modulation, the original source information could have been either analog or digital.

In the digital radio system on the transmitter side, the first step is pre-coding. The pre-coder performs level conversion and then encodes or groups the incoming data into controlled words that modulate the analog carrier. The modulated carrier is filtered, amplified and then transmitted through the transmission medium to the receiver. In the receiver the incoming signal is filtered, amplified, and then applied to the demodulator circuit which produces the original source information. Clock and carrier-recovery circuits extract carrier and clock timing information from the incoming modulated signal. The modulation used could be digital amplitude modulation (on-off keying modulation) or frequency-shift keying or phase-shift keying or a variation of any of these types of modulations.

5. DIGITAL MODULATION

The three major classes of digital modulation techniques used for transmission of digitally represented data are:

- Amplitude Shift Keying (ASK)
- Frequency Shift Keying (FSK)
- Phase Shift Keying (PSK)

5.1 Amplitude-Shift Keying

The simplest digital modulation technique is digital amplitude modulation, which is double sideband full carrier AM where the input modulating signal is a binary waveform. When the modulation is 100% the carrier is 'on' or 'off'. This type of modulation is commonly referred to on-off keying (OOK). In its simplest form, the presence of a carrier for a specific duration represents a binary one, while its absence for the same duration represents a binary zero. On-off keying is most commonly used to transmit Morse code over radio frequencies (referred to as continuous wave operation); although in principle any digital encoding scheme may be used. OOK has been used in the ISM (Industrial, scientific and medical) bands to transfer data between computers, for example. OOK is not very spectrally efficient due to the abrupt changes in amplitude of the carrier wave. At low to medium signaling speeds, this can be mitigated by adjusting the rise and fall rates of the carrier's amplitude. At high speeds, more efficient modulation modes (such as frequency-shift keying) are normally used instead. In addition to RF carrier waves, OOK is also used in optical communication systems.

5.2 Frequency-Shift Keying

Frequency-shift keying (FSK) is a form of frequency modulation in which the modulating signal shifts the output frequency between predetermined values.

Usually, the instantaneous frequency is shifted between two discrete values termed the mark frequency f_m and the space frequency f_s. Continuous phase forms of FSK exist, in which, there is no phase discontinuity in the modulated signal. Minimum frequency-shift keying (MSK) is a spectrally efficient form of coherent frequency-shift keying. In MSK the difference between the higher and lower frequency is identical to half the bit rate. As a result, the waveforms used to represent a 0 and a 1 bit, differ by exactly half a carrier period. This is the smallest FSK modulation index that can be chosen such that the waveforms for 0 and 1 are orthogonal.

With binary FSK, there is a change in the output frequency of the modulated signal each time the logic condition of the input signal changes. Consequently, the output rate of change is equal to the input rate of change. In digital modulation the rate of change at the input to the modulator is called the bit rate (f_b) and has the unit of bits per second (bps). The rate of change at the output of the modulator is called baud. Baud should not be confused with bit rate. Baud is a rate of change and equal to the reciprocal of the time of one output signaling element. With FSK the time of an output signaling element is the time of either a mark frequency or a space frequency outputted, which is equal to the time of a single bit (t_b). Therefore with binary FSK, the time of signaling element and the time of a bit are equal; thus, the input and output rates of change are equal and the bit rate and baud are also equal.

5.2.1 Bandwidth of FSK

The peak frequency deviation is given by $\Delta f = \dfrac{|f_m - f_s|}{2}$ ...(1)

The Bandwidth of FSK modulated signal $B = 2(\Delta_f + f_b)$ ...(2)

5.2.2 FSK modulator

FSK modulator is very often a voltage controlled oscillator (VCO). The carrier frequency is chosen such that it falls half way between mark and space frequencies. Logic 1 input shifts the VCO output to mark frequency and logic 0 input shifts the VCO output to space frequency. Consequently the binary input signal changes back and forth between logic 1 and logic 0 condition, the VCO output deviates back and forth between mark and space frequencies. The FSK modulated signal is obtained at the output of the VCO.

5.2.3 FSK demodulator

FSK demodulation is quite simple. The FSK input is split into two signals through a power splitter and the outputs are applied to band pass filters one tuned to mark frequency and the other tuned to space frequency. The respective filter passes only the mark or only the space frequency on to its respective envelop detector. The envelop detectors; in turn, indicate the total power in each pass band and the comparator responds to largest of the two powers. This type of demodulation is referred to as non-coherent detection.

The most common circuit used for demodulating binary FSK signal is the phase-locked loop (PLL). A PLL-FSK demodulator works similar to a PLL-FM demodulator. As the input of the PLL shifts between mark and space frequencies, the DC error voltage at the output of the phase comparator follows the frequency shift. Binary FSK has a poorer error performance than PSK; consequently it is seldom used for high performance digital radio systems. Its use is restricted to low performance, low cost, asynchronous data modems that are used for data communications over analog voice-band telephone lines.

5.3 Phase-Shift Keying (PSK)

Phase-shift keying (PSK) is a digital modulation scheme that conveys data by changing, or modulating, the phase of a reference signal (the carrier wave). Any digital modulation scheme uses a finite number of distinct signals to represent digital data. PSK uses a finite number of phases; each assigned a unique pattern of binary bits. Usually, each phase encodes an equal number of bits. Each pattern of bits forms the symbol that is represented by the particular phase. The demodulator, which is designed specifically for the symbol-set used by the modulator, determines the phase of the received signal and maps it back to the symbol it represents, thus recovering the original data. This requires the receiver to be able to compare the phase of the received signal to a reference signal. Such a system is termed coherent.

Alternatively, instead of using the bit patterns to set the phase of the wave, it can instead be used to change it by a specified amount. The demodulator then determines the changes in the phase of the received signal rather than the phase itself. Since this scheme depends on the difference between successive phases, it is termed differential phase-shift keying (DPSK). DPSK can be significantly simpler to implement than ordinary PSK since there is no need for the demodulator to have a copy of the reference signal to determine the exact phase of the received signal (it is a non-coherent scheme). In exchange, it produces more erroneous demodulations. The exact requirements of the particular scenario under consideration, determine which scheme is to be used.

Two common examples of PSK are "binary phase-shift keying" (BPSK) which uses two phases and "quadrature phase-shift keying" (QPSK) which uses four phases. However any number of phases may be used in PSK scheme. Since the data to be conveyed are usually binary, the PSK scheme is usually designed with the number of phases being equal to a power of 2.

5.3.1 Bandwidth of BPSK

The bandwidth of the BPSK signal $B = f_b$ ...(3)

5.3.2 BPSK modulator

BPSK modulation is realized using a balanced modulator that acts as a phase reversing switch depending on the logic condition of the digital input. The carrier is transferred to the output either in phase or 180° out of phase with the reference carrier oscillator. The balance modulator is shown in Figure 2

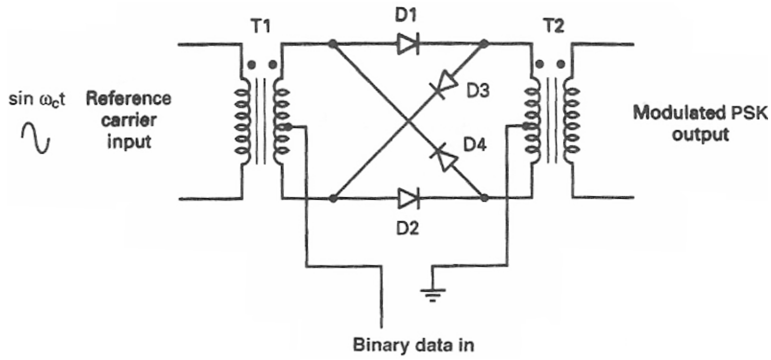


FIG. 2 BALANCE MODULATOR

5.3.3 BPSK demodulation

In the BPSK demodulator, coherent carrier recovery circuit detects and regenerates from the incoming signal, the carrier that is frequency and phase coherent with the original transmit carrier. The incoming signal and the recovered carrier are then applied to the product detector which is nothing but a balance modulator circuit. The output of the product detector is filtered by an LPF to obtain the binary data output.

5.3.4 QPSK modulator

QPSK uses four phases of a single carrier frequency. Because there are four different output phases there must be four different input conditions. The digital input to a QPSK modulator is a binary signal and this has to produce four different input conditions. It takes more than a single input bit to produce four different input conditions. With two bits there are four possible conditions: 00, 01, 10 and 11. Therefore with QPSK the binary input data are combined into groups of two bits called dibits. Each dibit code generates one of the four possible output phases. In this case for each two-bit sent into the modulator, a single output change occurs. Therefore the rate of change at the output (baud rate) is one half of the input bit rate.

After two bits have been serially inputted to the modulator, they are simultaneously parallel outputted. One bit is directed to what is called the 'I' channel and the other to the 'Q' channel. The I-bit modulates the carrier that is in phase with the reference oscillator and the Q-bit modulates a carrier that is 90° out of phase or in quadrature with the reference carrier. It may be noted that when once the bit stream is bifurcated into I and Q channels the subsequent operation is the same as BPSK modulator. When the linear summer combines the two quadrature (90° out of phase) signals there are four resultant phases and this output is QPSK modulated.

5.3.5 Bandwidth of QPSK

In QPSK system, because the input data rate (f_b) are divided into two channels the bit rate in either the 'I' or the 'Q' channel is equal to one half of the input data rate. Thus a bandwidth compression is realized i.e. minimum bandwidth is less than the incoming rate. Also because the QPSK output signal does not change phase until two bits have been inputted the fastest output rate of change (baud) is also equal to one half of the input bit rate. As with BPSK the minimum bandwidth and the baud are equal. The bandwidth of the QPSK is given by the relation $B = \frac{f_b}{2}$ ...(4)

5.3.6 QPSK demodulator

For QPSK demodulation, the input is passed through a power splitter that provides three outputs. The power splitter outputs are then fed into I and Q product detectors and the carrier recovery circuit. The carrier recovery circuit produces original transmit carrier oscillated signals. The required carrier must be frequency and phase coherent with the transmit carrier. The in-phase carrier is given to the I channel product detector while the Q channel product detector is given 90° phase shifted carrier. The I and Q data bits are generated by the respective product detectors. The outputs of the product detectors are fed to the bit combining circuit where they are converted from parallel I and Q data channels to a single binary output data stream. Thus the transmitted data stream is recovered at the output of the QPSK demodulator.

6. BIT ERROR RATE CALCULATIONS FOR DIGITAL CHANNEL

The performance of the digital channel is evaluated based on the probability of error in the channel or bit error rate (BER) encountered. Probability of error P_b and bit error rate (BER) are often used interchangeably although in practice they have slightly different meanings. P_b is a theoretical expectation of the BER for a given system. BER is an empirical record of the system's actual error performance.

Probability of error for a digital communication system is a function of the carrier-to-noise power ratio and it is more specifically the average energy per bit-to-noise power density ratio. Carrier power C is estimated in Watts. The noise power N is expressed mathematically equal to KTB in Watts. Here N equal to noise power in the band, K is Boltzmann's constant (1.38×10^{-23} Joules / Kelvin), T is temperature in Kelvin and B is bandwidth in cycles per second (Hertz).

$$C/N\,(dB) = 10 \log C/N \qquad ...(5)$$

Energy per bit is the energy of a single bit of information. Mathematically, energy per bit is $E_b = C/f_b$, where f_b is the bit rate.

Noise power density $\quad N_0 = N/B$ ...(6)

Therefore, $\quad E_b/N_0 = C/N \times B/f_b$ ...(7)

The bit error performance for the various digital modulation systems is directly related to E_b/N_0 and the number of possible encoding conditions used (*M-ary*).

For BPSK $\quad P_b = \frac{1}{2} erfc \sqrt{\frac{E_b}{N_o}}$ ...(8)

The error performance of QPSK is the same as BPSK. This is because the 3dB reduction in error distance for QPSK is offset by the 3dB decrease in bandwidth.

The error performance for FSK system is as given by the following expressions

For noncoherent FSK $$P_b = \frac{1}{2} exp\left(-\frac{E_b}{2N_o}\right) \qquad ...(9)$$

For coherent FSK $$P_b = erfc\sqrt{\frac{E_b}{N_o}} \qquad ...(10)$$

For the desired BER, the digital communication link is designed to ensure the minimum E_b/N_0 as given by the above equations of P_b. In practice, in the system design, a comfortable margin is built-in to guarantee the desired BER under all conditions. The value of E_b/N_0 required ensuring a BER of 1×10^{-6} is equal to 10.6 dB for BPSK and QPSK. Satellite links are usually designed with 1×10^{-6} BER. The plots of the above equations are shown below in Figure 3. The error performance of PSK (BPSK&QPSK), as can be seen, is the best when compared with all other digital modulations discussed above.

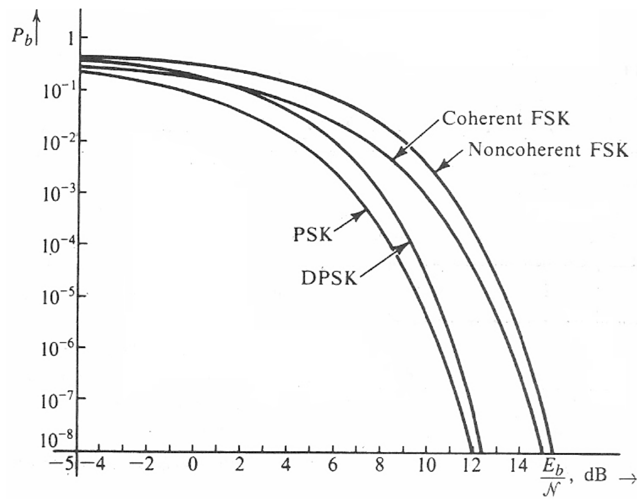


FIG. 3 ERROR PROBABILITY OF PSK, DPSK, COHERENT FSK AND NON-COHERENT FSK

7. CONCLUSION

We have briefly seen the basic techniques of digital communication and the various types of digital modulations. Digital communication has revolutionized the electronic communications scenario over the years in many ways. Digital techniques have also led to the efficient use of communication channels. Developments in digital systems are phenomenal and today capacities closer to Shannon's limits have been realised.

REFERENCES

[1] Wayne Tomasi, “Electronic Communication Systems” Pearson Education, 2001
[2] Simon Haykin, “Communication Systems” John Wiley & Sons, 2001

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

OPTOELECTRONIC DEVICES

D. JOHN

1. INTRODUCTION

Optoelectronics is based on the combination of optics and electronics. The advantage of such a combination results from the large information carrying capacity of light together with the easy controllability of electrons. The large information carrying capacity of light comes from the high frequency of its oscillation. The period of oscillation is of the order of around 10^{-14} second and therefore optical processes can absorb and convey information on the same kind of time scale leading to information rates of the order of optical frequency.

Electrons, unlike photons carry electrical charge. As a consequence they respond to electric and magnetic fields which can be used to control and manipulate them. This is a significant advantage, since modern electronics provide finely adjustable fields at a sophisticated level of flexibility. There are well understood processes by which electrons can give rise to photons and photons to electrons when certain conditions are satisfied. To day a wide range of applications are based on optoelectronics: optical communications, compact audio-discs, laser printers, liquid crystal displays on watches and calculators, supermarket laser check-outs, laser surgery, to name just a few.

2. PHYSICS OF LIGHT

During the early days of the study on light, two emission theories were developed, the corpuscular theory of Newton and the wave theory of Hooke and Huygen**s**. Newton proposed that light consisted of small particles, with which he could easily explain the phenomenon of reflection. With considerably more difficulty, he could also explain refraction through a lens, and the splitting of sunlight into a rainbow by a prism. Newton’s particle theory however, fails to explain other effects, such as interference and diffraction. Wave theory of light, was later understood as part of electromagnetism.

In 1864 James Clark Maxwell combined the equations of electromagnetism in a general form and showed that they suggested the existence of transverse electromagnetic waves. The speed of propagation in free space of these waves was given by

$$C = \sqrt{\frac{1}{\mu_0\, \varepsilon_0}} \qquad ...(1)$$

Where, μ_0 and ε_0 respectively, are the permeability and permittivity of free space.

Substitution of the experimentally determined values of μ_0 and ε_0 yielded a value for C in very close agreement, with the value of speed of light in vacuum, measured independently. Maxwell therefore proposed that light was an electromagnetic wave having a speed C of 3×10^8 ms^{-1} , a frequency of some 5×10^{14} Hz and a wavelength of about 500 nm. Maxwell's theory suggested the possibility of producing electromagnetic waves with a wide range of frequencies (or wavelengths). In 1887 Hertz succeeded in generating electromagnetic waves with a wavelength of the order of 10 m by discharging an induction coil across a spark gap thereby setting up oscillating electric and magnetic fields. Visible light and Hertzian waves are parts of the electromagnetic spectrum. The wave theory became an acceptable theory of light and provides an explanation for optical phenomena such as interference and diffraction. However it fails completely in situations where energy is exchanged, such as the emission and absorption of light and the photoelectric effect.

2.1 Photoelectric Effect

The photoelectric effect, which is emission of electrons from the surfaces of solids when irradiated, was explained by Einstein in 1905. He suggested that energy of light beam is not spread evenly but is concentrated in certain regions which propagate like particles. He called these particles photons. Einstein was led to the concept of photons by the work of Max Planck on the emission of light from hot bodies. Max Planck found that the observation indicated that light energy was emitted in multiples of a certain minimum energy unit. The size of the unit which is called a quantum, depends on the wavelength *ë* of the radiation and is given by,

$$E = \frac{hC}{\lambda} \qquad \text{...(2)}$$

Where h is Planck's constant.

Planck's hypothesis did not require that the energy should be emitted in localized bundles and it could, with difficulty, be reconciled with the electromagnetic wave theory. When Einstein showed, however, that it seemed necessary to assume the concentration of energy traveling through space as particles, a wave solution was excluded. Thus we have a particle theory also; apparently light has a dual nature! The two theories of light are not in conflict but rater they are complementary. Wherever energy exchange is involved the particle nature of light dominates the wave nature of light. On the other hand in dealing with interference and diffraction the wave nature of light dominates.

2.2 Wave Nature of Light

Light as an electromagnetic wave is characterized by a combination of time varying electric and magnetic fields propagating through space. Maxwell showed that both these fields satisfy the same partial differential equation, the wave equation.

$$\nabla^2 \vec{E} = \frac{1}{C^2} \frac{\delta^2}{\delta t^2} \vec{E} \qquad \text{...(3)}$$

$$\nabla^2 \vec{H} = \frac{1}{C^2} \frac{\delta^2}{\delta t^2} \vec{H} \qquad \text{...(4)}$$

The implication of the above equation is that changes in the fields propagate through space with speed C, the speed of light. The frequency of oscillation of the fields, v and their wavelength in vacuum λ_0 are related by:

$$C = v\lambda_0 \qquad ...(5)$$

In any other medium the speed of propagation is given by:

$$\text{v} = \frac{C}{n} = v\lambda \qquad ...(6)$$

Where n is the refractive index of the medium and λ is the wave length in the medium. The refractive index n is given by:

$$n = \sqrt{\mu_r \varepsilon_r} \qquad ...(7)$$

Where μ_r and ε_r respectively, are the permeability and permittivity of the medium.

The electric and magnetic fields are perpendicular to one another and perpendicular to the direction of propagation. That is, the light waves are transverse EM waves. In describing the optical phenomena we often omit the magnetic field vector. However we should always remember that there is a magnetic field component which behaves in a similar way as the electric wave component.

3. POLARIZATION

If the electric vector vibrates in a specific plane the wave is said to be plane polarized and the plane of polarization is the plane containing the electric vector. Light can be polarized by reflection and by absorption. We find that light with its polarization perpendicular to its plane of incidence is perfectly reflected in comparison to the light polarized parallel to the plane of incidence. We can resolve any electric field vector into components parallel and perpendicular to any convenient direction. Here we choose plane of incidence. Thus we may think of unpolarized light as comprising of two equal plane polarized components with orthogonal orientations.

We further find that the reflectance of the surface for the perpendicular and parallel component varies as the function of the angle of incidence and at a particular angle of incidence called *Brewster angle* (θ_B), the entire parallel component is transmitted. It is also seen that for incidence at Brewster angle the reflected and refracted rays are perpendicular to one another. It can be shown that

$$\tan \theta_B = \frac{n_2}{n_1} \qquad ...(8)$$

Where n_1 and n_2 are the refractive indices of the medium 1 from where light passes into medium 2, respectively. The above equation is known as Brewster's Law. One way of polarizing light is simply to pass it through a series of glass plates oriented at the Brewster angle.

In many cases of interest a beam of light may consist of two plane polarized wave trains with their plane of polarization right angle to each other and which may be out of phase. Let us consider a special case where the amplitudes of the two wave trains are equal and having a phase difference of $\pi/2$ between them. In this case the *E* vector will be rotating in a circle perpendicular to the direction of propagation as it propagates; such a wave is said to be *circularly* polarized. In the most general case when the amplitudes are different and having 'any' phase angle between them, the wave is *elliptically* polarized.

4. BIREFRINGENCE

Ordinary glass is isotropic in its properties, but many important crystalline optical materials such as Calcite, Quartz and KDP (Potassium dihydrogen Phosphate) are anisotropic. This anisotropy is due to the arrangement of the atoms being different in different directions through the crystal. In such a crystal the speed of propagation of light wave depends on the direction of propagation and the polarization of the light. That is, the refractive index of the crystal varies with the direction in the crystal. These crystals are said to be birefringent or doubly refracting; the names refer to the fact that there are in general two different directions of propagation that a given incident ray may take, depending on the direction of polarization. The rays corresponding to these directions travel with different speeds and have mutually orthogonal planes of polarization.

In general crystals exhibit three different principal refractive indices and two optic axes. The optic axes are directions in the crystal along which the velocities of the two orthogonally polarized waves are the same. In many important crystals, for example Calcite, two principal indices are the same and there is only one optic axis. Such crystals are called *uniaxial* whereas other doubly refracting crystals, for example Mica, are *biaxial*. In cubic crystals which are isotropic the principal indices are all the same.

When an un-polarized narrow beam of light falls normally on a parallel-sided Calcite plate, the beam is found to divide into two parts. One is called ordinary ray or O ray and it passes straight through the crystal. The other ray called extraordinary ray or E ray diverges as it passes through the crystal and then emerges parallel to its original direction. This is found to be the case *unless* the direction of incidence of the original beam is parallel or perpendicular to the optic axis. The ordinary and extraordinary rays are found to have orthogonal directions of polarizations.

In the case of normal incidence on a crystal plate cut with its surface normal to the optic axis, there is no double refraction. The state of polarization of the emergent light is the same as that of the incident light.

However, for a crystal which is cut such that, the optic axis is parallel to the crystal surface but normal to the plane of incidence, there is no divergence of E and O rays. The rays become increasingly out of phase as they propagate through the crystal. A crystal plate cut in this fashion that introduces a phase difference of $\lambda/2$ (path difference of $\lambda/4$) between the O and E rays is called a *quarter-wave plate*. For quartz, for example, a thickness of 0.0164 mm for Sodium light introduces a phase difference of $\lambda/2$. When plane polarized light is incident on a quarter-wave plate, the emergent light is in general elliptically polarized. If the plane of polarization of the incident beam is inclined at 45° to the privileged directions, the emergent light is circularly polarized. [The two directions of vibrations of the incident light which give only a single emergent beam are called *privileged directions*] In similar way one can produce half-wave and full-wave plates. Such plates are used in light modulation systems.

5. OPTICAL ACTIVITY

Certain crystals have the ability to rotate the plane of polarization of light passing through them. They are said to be optically active. For example, when a beam of plane polarized light is incident normally on a crystal plate of quartz cut perpendicular to the optic axis, it is found that the emergent beam is also plane polarized but its plane of polarization is rotated with respect to the incident light. The plane of polarization may be rotated clockwise looking against the oncoming light (right-handed or dextrorotary crystals) or rotated anticlockwise (left-handed or levorotary). Quartz exists in both forms. It is found that the rotation depends on the thickness of the crystal plate and the wavelength. The rotation produced by a 1 mm thick quartz plate for sodium light is 21.7° while it is 3.67° for 1 mm of sodium chlorate.

The optical activity (*circular birefringence*) can be explained by assuming that in active crystals the velocity of propagation of circularly polarized light is different for different directions of rotation, i.e. the crystal has different refractive indices, n_r and n_l respectively, for *right* and *left circularly polarized* light. A *plane polarized light* can be considered as a combination of two circularly polarized waves with opposite directions of rotation if these components travel through the crystal at different speeds, a phase difference will be introduced between them at different distances through the crystal. This corresponds to a rotation of the plane of the plane polarized wave which results from the combination of the two circularly polarized waves.

6. OPTOELECTRONIC MODULATORS

Use of lasers in a wide variety of applications has led to a demand for devices which can modulate a beam of light. Applications of light modulators include wide-band analog optical communication systems, switching for digital information recording, information storage and processing, pulse shaping, to name a few. A modulator is a device which changes the irradiance (or direction) of the light passing through it. There are several general types of modulators. Here we will discuss electro-optic, magneto-optic and acousto-optic modulators.

6.1 Electro-optic Effect

When an electric field is applied to an optical medium the electrons suffer restricted motion in the direction of the field, when compared with that which is orthogonal to it. Thus the material becomes linearly birefringent in response to the field. This is known as the electro-optic effect. Under an applied electric field naturally isotropic crystals, for example gallium arsenide (GaAs), becomes doubly refracting or naturally doubly refracting crystals such as KDP gets a new optic axis introduce. Consider the incident light which is linearly polarized at 45° to an electric field that acts on the medium transversely to the propagation direction of the light. The electro-optic effect of the medium will cause a phase displacement between components of the incident light which lie, respectively, parallel and orthogonal to the field. The emerging light will therefore be elliptically polarized. This is shown in Figure 1.

A perfect polarizer (polarization analyzer) placed on the output side of the medium with its acceptance direction parallel to the input polarization direction will pass all light in the absence of

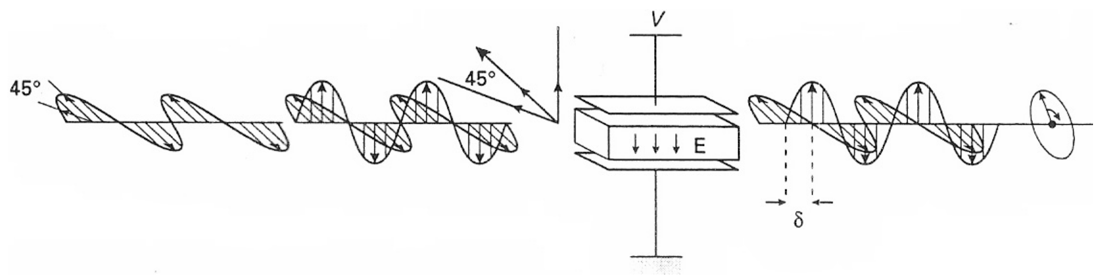


FIG. 1 LINEAR POLARIZATION BECOMES ELLIPTICAL WITH APPLIED E FIELD

the field. When the field is applied, the fraction of light power which is passed will depend upon the form of the ellipse, which in turn depends upon the phase delay introduced by the field. Consequently, the field can be used to modulate the intensity of light. The electro-optic effect is therefore very useful for modulation of light.

The phase delay introduced may be proportional either to the field (*Pockels effect*) or to the square of the field (*Kerr effect*). All materials manifest a transverse Kerr effect. Only crystalline materials can manifest any kind of Pockels effect, or longitudinal (E field parallel with propagation direction) Kerr effect.

In addition to the modulation of light (phase or intensity) it is clear that the electro-optic effect could be used to measure an electric field and/or the voltage which gives rise to it. Modulators and sensors based on this idea are built and used.

6.1.1 Electro-optic modulator

When building a modulator certain adjustments are to be made to make it perform satisfactorily. In the scheme explained above, suppose the acceptance direction of the polarization analyzer is placed at 90° to the original input polarization direction, in the absence of applied field no light passes through the analyzer to the photo-detector placed beyond. On application of the electric field the polarization becomes elliptical and some light passes through the analyzer to the detector. In this arrangement the sensitivity of light striking the detector is not very satisfactory in relationship with the electric field. To improve the sensitivity a bias phase delay of ð/2 is inserted between the polarization components before they reach the analyzer. In this case the light in the absence of the applied electric field reaches the analyzer circularly polarized. When the electric field is applied the photo detector sees a change in intensity linearly proportional to the phase change occurred in the electro-optic medium of the modulator. As shown in Figure 2, in the vicinity of 50% of light transmission by the polarization analyzer the irradiance variation is linear with the applied voltage.

For Pockels effect change in intensity with this biasing is linearly proportional to E field and thus the voltage applied. For Kerr effect these changes are proportional to the square of E (and thus the voltage). However we can use the same kind of biasing to make it a linear variation (if necessary). This time the bias is applied to the external electric field, which is a steady value and the modulation is applied over and above this steady value.

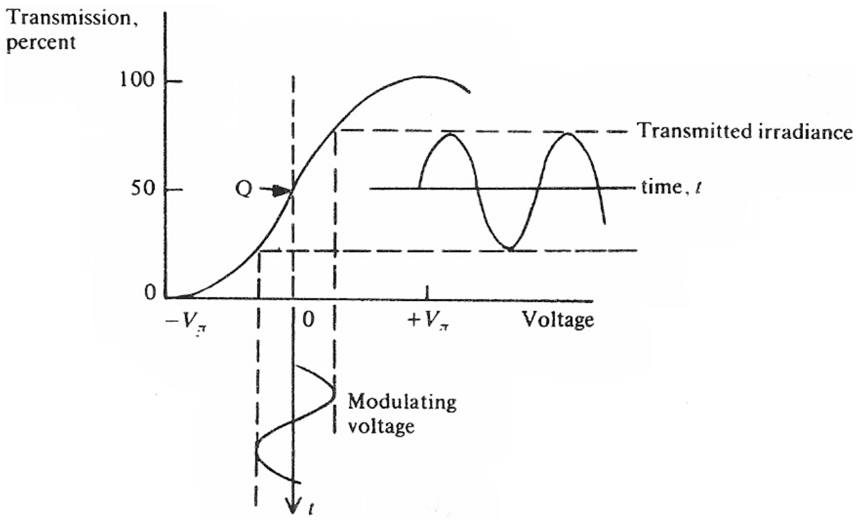


FIG. 2 TRANSMITTED IRRADIANCE VS. APPLIED VOLTAGE

It is possible to devise an electro-optic modulator with either Pockels effect of Kerr effect materials. By and large it is most convenient to use materials which have a strong linear (Pockels) electro-optic effect. The examples of such materials are ammonium dihydrogen phosphate (ADP), KDP, lithium Niobate ($LiNbO_3$), lithium tantalite ($LiTaO_3$), zinc sulphide (ZnS) and gallium arsenide (GaAs). Only non-centre-symmetric crystals exhibit Pockels effect, while all materials exhibit Kerr effect. Therefore in practice the latter is often more accessible.

6.2 Magneto-optic Devices

The presence of magnetic field also affects optical properties of some substances thereby giving rise to some useful magneto-optic devices. *Faraday Effect* is the simplest magneto-optic effect and is the only one of real interest for optical modulators. It concerns the change in the refractive index of a material subjected to a steady magnetic field. Faraday found that when a beam of plane polarized light passes through a substance subjected to a magnetic field its plane of polarization is observed to rotate by an amount proportional to the magnetic field component, parallel to the direction of propagation. This is very similar to optical activity which results from certain materials having different refractive indices n_r and n_l for right and left circularly polarized light (circular birefringence). There is one important difference in the two effects. In Faraday Effect the sense of rotation of the plane of polarization is independent of the direction of propagation. This is in contrast to optical activity where the sense of rotation is related to the direction of propagation. The difference between Faraday Effect and optical activity is shown in Figure 3. Thus in the case of Faraday Effect the rotation can be doubled by reflecting the light back to the (magneto) optic device. The rotation of plane of polarization of the light beam passing through the magnetic field is given by:

$$\theta = VBL \qquad ...(9)$$

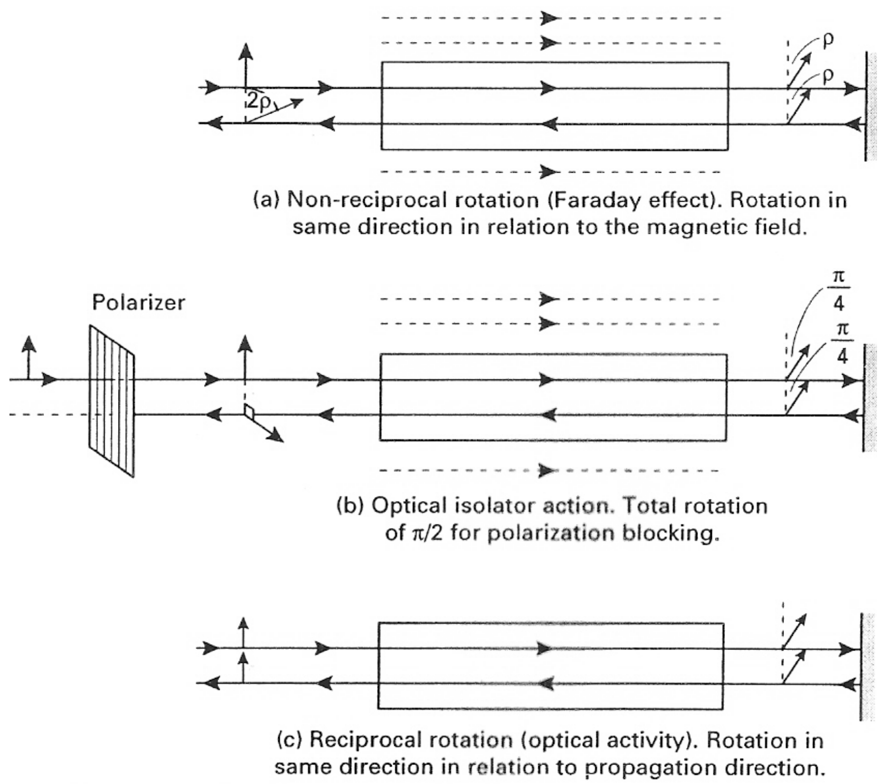


FIG. 3 FARADAY (NON-RECIPROCAL) ROTATION AND OPTICAL ACTIVITY (RECIPROCAL) ROTATION

Where V is known as the Verdet constant, B is the magnetic flux parallel to the direction of propagation and L is the path length in the material.

The Faraday Effect is small and wavelength dependent. The rotation for dense flint glass is $\theta \approx 1.6°$ per mm thickness, per Tesla magnetic flux density at $\lambda = 589.3$ nm.

For optical activity the rotation θ, in terms of refractive indices n_r and n_l is given by the expression:

$$\theta = \frac{2\pi}{\lambda}(n_r - n_l)L \qquad ...(10)$$

6.2.1 Optical isolator

A Faraday rotator used in conjunction with a pair of polarizers acts as an optical isolator which allows a light beam to travel through it in one direction but not in the opposite one. It may therefore be used in laser amplifying chains to eliminate reflected, backward traveling rays, which are potentially damaging.

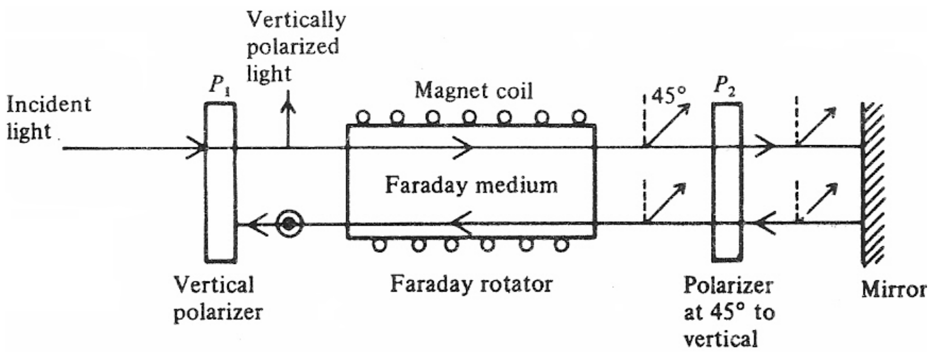


FIG. 4 OPTICAL ISOLATOR BASED ON THE FARADAY EFFECT

Light passing from left to right is polarized in a vertical plane by polarizer P_1, as shown in Figure 4. The Faraday rotator is adjusted to produce a rotation of 45° in the clockwise sense. The polarizer P_2 is set at 45° to P_1 so that it will transmit light emerging from the rotator. However, when a beam entering from the right after reflection by a mirror it will be plane polarized at 45° to the vertical and then have its plane rotated by 45° in the clockwise sense by the rotator. It will therefore be incident on the first polarizer with its plane of polarization at right angles to the plane of transmission and be eliminated. The device thus isolates the component on its left from light incident from the right.

6.2.2 Magneto-optic modulator

Magneto-optic modulator is constructed in a similar manner as the electro-optic modulator. Linearly polarized light enters a magneto-optic medium and emerges to pass first through a linear polarization analyzer and then on to a photo-detector. The acceptance direction of the analyzer is set at 45° to the emerging light's polarization direction in the absence of the magnetic field. The small variation of intensity received by the detector is the proportional to the induced rotation of polarization by the magnetic field.

The magnetic field is produced by the electric current, so for this modulator we require the signal information in the form of a current rather than a voltage. Since the magnetic field is generated using solenoids with large number of turns, which means large electric inductance. This makes the circuit sluggish to fast responses limiting the frequency response of the modulator.

6.2.3 Magneto-optic computer memory

One potential of magneto-optics is large capacity computer memories. Magneto-optic memories developed so far are read via the 'Faraday effect' or the 'magnetic Kerr effect' which relates to the rotation of a beam of plane polarized light reflected from the surface of a material subjected to a magnetic field. In either case a magnetized ferro or ferrimagnetic material rotates the plane of polarization of laser light incident on it. Writing may be achieved by heating the memory elements on the storage medium to a temperature above the Curie point using a laser beam. The element is then allowed to cool down in the presence of an external magnetic field thus acquiring a magnetization in a given direction. Magnetization of the elements in one direction may represent '1' and in the opposite direction '0'. To read the information the irradiance of the laser beam is reduced and then directed to the memory elements. The direction of the change in polarization of the laser beam on passing through or being reflected from the memory elements depends on the direction of magnetization; therefore you can decide if a given element is storing a '1' or '0'.

Systems have been built for example with 50 mW helium-neon laser with a Pockels modulator and manganese bismuth thin-film storage elements enabling information storage and reading and erasing at rates greater than 1 Mbps.

6.3 Acousto-optic Effect

The acousto-optic effect is the change in the refractive index of a medium caused by the mechanical strains accompanying the passage of an acoustic (strain) wave through the medium. The strain and hence the refractive index varies periodically with a wavelength Λ equal to that of the acoustic wave. The refractive index changes are caused by the photo elastic effect which occurs in all materials on the application of a mechanical stress. The change in the refractive index is proportional to the square root of the total acoustic power.

Let us consider the case of a monochromatic light wave, wavelength ë, incident upon a medium, as shown in Figure 5, in which an acoustic wave has produced sinusoidal variations of wavelength Λ, in the refractive index. [The light enters the medium orthogonal to the direction of sinusoidal

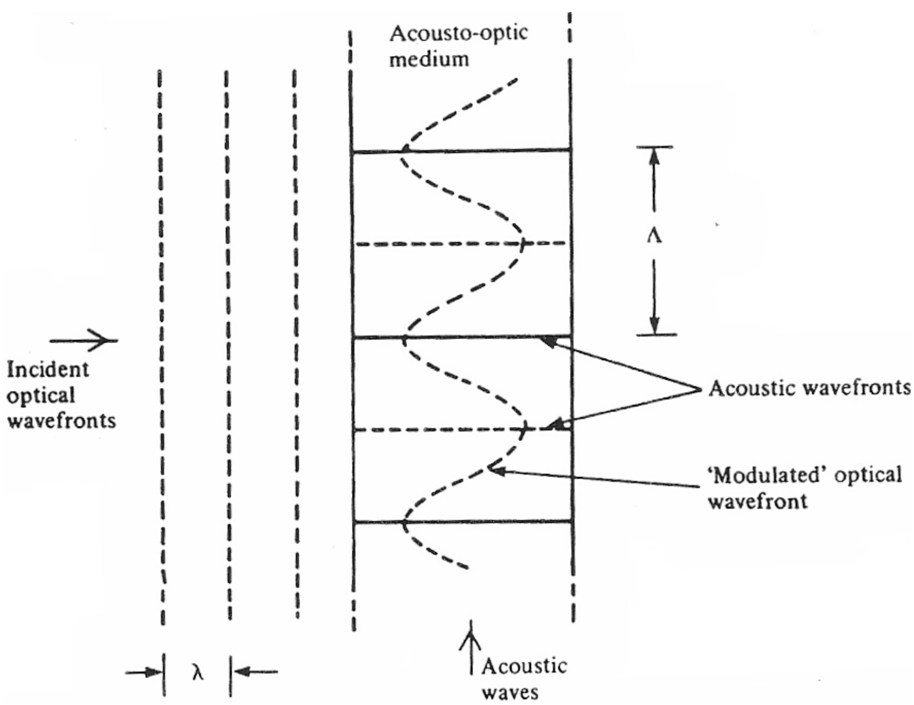


FIG. 5 ACOUSTO-OPTIC MODULATION - SCHEMATIC

index variation caused by impinging acoustic wave.] The portion of the optical wavefront near the pressure peak caused by the acoustic wave will encounter a high refractive index and therefore advance with lower velocity than those portions of the wavefront which encounter pressure minima. The optical wave front in the medium therefore soon acquires the wavy appearance. The acoustic wave velocity is much less than the velocity of light wave. Therefore we can consider the variation in refractive index to be stationary in the medium.

As elements of the light wave propagate in a direction normal to the local wavefront, almost all the wave elements will suffer a change in direction leading to a redistribution of the light flux, which tends to concentrate near regions of compression. In effect the acoustic wave sets up a diffraction grating within the medium so that optical energy is diffracted out of the incident beam into various orders. There are two main cases of interest, namely, Raman-Nath regime and Bragg regime.

In *Raman-Nath regime* the acoustic diffraction *grating is so thin* that the diffracted light suffers no further redistribution before leaving the medium. The light is diffracted as from a simple plane grating. The radiance I of the light, in various diffraction fringes depends on the 'ruling depth', which is related to amplitude of the acoustic grating. This, in turn, is related to the amplitude of the acoustic modulating wave. The fraction of light removed from the zero order beam (direct Beam) is $\eta = (I_0 - I)/I_0$, where I_0 is the transmitted irradiance in the absence of acoustic wave. Thus amplitude variations of the acoustic wave are transformed into irradiance variations of the optical beam.

Physical basis of the *Bragg regime* is that light diffracted from the incident beam is extensively re-diffracted before leaving the acoustic field. Under these conditions, the acoustic field acts very much like a thick diffraction grating, that is, a *grating made up of planes rather than lines*. Consider a plane wavefront incident on the grating planes at an angle of incidence *è*$_i$. Significant amount of light will emerge only in those conditions in which constructive interference occurs. The conditions to be satisfied are the light scattered from a given grating plane must arrive in phase at the new wavefront and light scattered from successive grating planes must also arrive in phase at the new wavefront, implying that the path difference must be an integral number of wavelengths.

That is, $\theta_d = \theta_i$ where θ_d is the angle of diffraction.

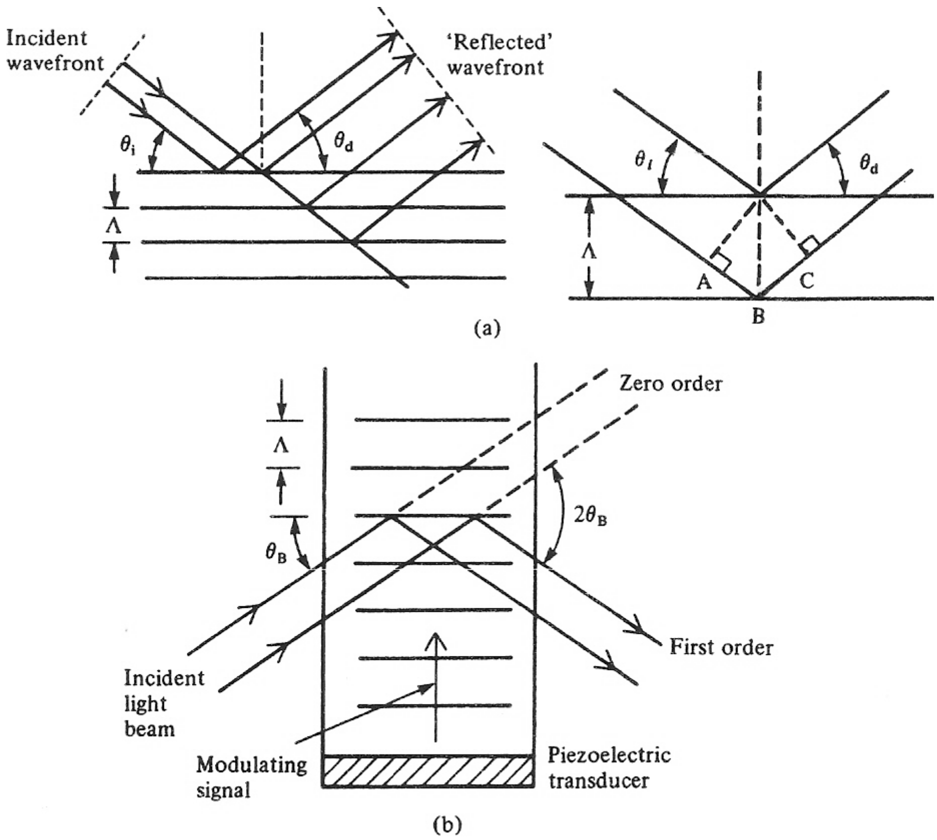
Also $$\sin \theta_i + \sin \theta_d = \frac{m\lambda}{2\Lambda} \qquad \text{...(11)}$$

With $m = 0, 1, 2\ldots$ The two conditions are satisfied simultaneously when

$$\sin \theta_i = \sin \theta_d = \frac{m\lambda}{2\Lambda} \qquad \text{...(12)}$$

The diffraction is similar to that obtained with a plane grating, but only for special angles on incidence; the angle of incidence must be equal to the angle of diffraction.

Although in the simplified theory strong scattering can take place when m is equal to any positive integer, taking into account the fact that the scattering is not from discreet planes but from a continuous medium, it can be deduced that scattering only takes place when $m = 1$. The equation for the so called Bragg angle θ_B becomes $\sin \theta_B = \lambda/2\Lambda$. The modulation depth (or diffraction efficiency) is $(I_0 - I)/I_0$. In this case it can be theoretically equal to 100% in contrast to about 34% for the Raman Nath case.



BRAGG ACOUSTO-OPTIC DIFFRACTION (a) SCATTERING BY SUCCESSIVE LAYERS (b) REFLECTION IN TO FIRST ORDER

The acoustic waves which create the diffraction grating are moving through the medium. The diffracted wave behaves as if it has been reflected from a mirror moving with the same velocity as the grating. Therefore the diffracted wave appears to originate from a source moving at twice the velocity of the mirror (grating). Thus the frequency of the reflected beam is changed by the Doppler Effect. It can be shown that the shift in frequency is equal to the acoustic wave frequency. This change in frequency can be used as a basis of a frequency modulator.

Acousto-optic modulator can in general be used for similar applications to electro-optic modulators, though they are not so fast. On the other hand, because the electro-optic effect usually requires voltages in kilovolt range the drive circuit becomes complex and expensive than that for acousto-optic modulators which operate with few volts.

7. OPTICAL DETECTION DEVICES

Optical detectors may be classified as either thermal or photon devices. In thermal detectors, the absorption of light raises the temperature of the device and this in turn results in changes in some temperature dependent parameters such as electrical conductivity. The output of a thermal detector is usually proportional to the amount of energy absorbed per unit time by the detectors provided the absorption efficiency is the same at all wavelengths of light. In the photon detector on the other hand the absorption process results directly in some specific quantum event (such as photo-electric emission of electrons from a surface) which is then countered by the detection system. Thus the output of a photon detector is governed by the rate of absorption of light quanta and not directly their energy. Furthermore, all photon processors require certain minimum photon energy to initiate them. Since the energy of a single photon is given by $E = h\nu = hC/\lambda$, photon detectors have a long wavelength cut-off, that is, a maximum wavelength beyond which they do not operate.

A photon detector operated in the infrared region encounters photon energies comparable with the average thermal energies ($\approx kT$) of atoms in the detector itself. A relatively large number of quantum events may then be generated by thermal excitation rather than by light absorption and will constitute a source of noise. The obvious way to reduce this noise is to reduce the temperature of the detector. In fact, most photon detectors operating above a wavelength of 3000 nm or so must be cooled to liquid nitrogen temperatures (77 K) or below.

7.1 Thermal Detector

As stated above thermal detectors work on the principle that light falling on the detection device, raises its temperature proportional to the illumination. This temperature is measured accurately using one of the several methods available.

7.1.1 Thermoelectric detectors

Thermoelectric detectors use the principle of thermocouple (Seebeck Effect) whereby the heating of one junction between two dissimilar metals relative to the other causes a current to flow round the circuit which is proportional to the temperature between the junctions. In thermoelectric detectors of light, one junction is used to sense the temperature rise of the sensing element, while the other is maintained at ambient temperature. Although metals are most often used for the junction materials, certain heavily doped semiconductors can offer improved sensitivity but are generally less robust and give rise to constructional problems. The usefulness of the thermoelectric detector lies in the simplicity and rugged construction.

7.1.2 Bolometer

In the bolometer the incident radiation heats a fine wire or metallic strip causing a change in its electrical resistance. The change in resistance may be detected in several ways, for example, inserting into one arm of a Wheatstone bridge. Care must be taken to ensure that any current flowing through the element is sufficiently small not to raise its temperature by a significant amount. The resistivity of metals increases with increase in temperature that is, the temperature coefficient of resistance is positive. Platinum and nickel are most commonly used. Greater sensitivity may be achieved by using semiconductor elements which are sometimes called Thermistors. These consist of oxides of manganese, cobalt or nickel. Carbon resistance bolometers cooled to liquid helium temperature (4.2 K) have proved successful in far-infrared astronomy where very sensitive detectors are required. In the INSAT VHRR payload the ‘infrared detector’ is bolometer type and made of Indium-tin-cadmium-telluride.

7.2 Photon Devices

When radiation with a wavelength less than the critical value is incident upon a metal surface, electrons are found to be emitted. This is called photo-emissive or *photoelectric effect*. When a *photon* of energy *hí* enters the metal its energy may be absorbed and the energy is given to an *electron*. Provided the electron is able to reach the surface and has enough energy to overcome the surface potential barrier (given by $e\varphi$, where φ is the surface work function), it may escape the metal and photoelectric emission takes place. If the electron is initially at the Fermi level, its kinetic energy E on emission is given by:

$$E = h\nu - e\varphi \qquad \text{...(13)}$$

However, the electron may be initially below the Fermi level and may also suffer inelastic collisions before emission. Therefore E above represents the maximum energy available to the emitted electrons. No electrons at all will be emitted when $h\nu < e\varphi$. If the probability of inelastic collisions of the excited electrons is high then, only a fraction of them may be able to escape. The ratio of the number of emitted electrons to the number of absorbed photons is called the quantum yield or *quantum efficiency*.

Pure metals are rarely used as photo cathodes since they have low quantum efficiencies (~ 0.1 %) and high work functions. (Caesium - *Cs* has the lowest value for work function which is 2.1 eV.) Older type photo emissive surfaces consist of thin evaporated layer containing compounds of alkali metals (usually including *Cs*) and one or more metallic elements from Group *V* of the Periodic Table (e.g. Antimony – *Sb*). They are most often designated by an '*S*' number. The other type, called negative electron affinity type of photo emissive surfaces, is formed by evaporating very thin layers of Caesium or Caesium oxide onto semiconductor surface. Photo cathodes using GaAs have been used successfully and operate with quite high quantum efficiency.

7.2.1 Vacuum photo diodes

In the vacuum photo diode a photo emissive surface (cathode) is placed inside a vacuum tube with another electrode (anode) biased positively with respect to it. When photo cathode is illuminated the electrons will be collected by the anode and current flows in the external circuit. If the tube is filled with a gas such as argon under low pressure (~ 1 torr or less), photo electrons on the way from cathode to anode will collide with gas atoms and provided they are sufficiently energetic may ionize them and so generate further electrons. The overall current gained may be typically of the order of 10. The other method of obtaining higher internal gain is by using *photo-multiplier*.

7.2.2 Photon counting technique

A useful technique for dealing with very low level signals is photon counting. If the cathode quantum efficiency is unity, then each photon striking the photon cathode gives rise to a single current pulse at the anode. It has been found that for certain photo- multiplier designs the magnitude of the current pulses lie within closely defined limits, whereas the noise pulses have a much wider amplitude distribution. The photon counting system consists of fast pulse counter coupled to a pulse height discriminator. Pulses are ignored if they do not fall within the preset limits, thus improving the signal to noise ratio. Additionally, if the light signal being detected is chopped on and off then a further reduction in the background noise level may be obtained by subtracting the count rate when the signal is off from the count rate when both signal and noise counts are present.

7.2.3 Photo conductive detectors

We know that the conductivity of a semiconductor depends upon the number of charge carriers in the conduction band. A photon when absorbed by a semiconductor atom is capable of raising an electron from valence band to the conduction band (provided that its energy is greater than the band-gap energy) and thus increasing the conductivity. In doing so and electron hole pair is created; this will eventually be annihilated, since it represents a pair of chargers in excess of thermal equilibrium value. However, while the extra charge exits it can be used to contribute to a current whose value will depend upon the light power incident on the semiconductor. Thus, a small voltage across the photo conductive material layer will give rise to the required measurable current in an external circuit.

For the photon to yield an electron-hole pair, its energy must satisfy $hv > E_{g,}$ where E_g is the band-gap energy of the material. In terms of wavelength, $\lambda < hC/ E_g$

We define the band-gap wavelength $\ddot{e}_g$ as the largest value of wavelength that can cause this transition. That is, $\lambda_g = hC/ E_g$ ...(14)

It is seen that the current produced by a photo conductive detector is directly proportional to the optical power. There are two disadvantages possessed by photo conductive devices when compared to other types of detectors. First is that the response is too slow for most optoelectronic applications. Second is that the electron-hole recombination process is a random one and therefore introduces noise. This 'recombination noise', as it is called, is troublesome in many applications. An advantage, however is that they can be used at long wavelengths up to ~ 20,000 nm. Another advantage of these devices is that they are very cheap.

7.2.4 Photo diodes

Photo diodes and photo transistors convert light energy into electrical energy. Their operation is based on the fact that the number of free electrons generated in a semiconductor material is proportional to the intensity of incident light.

When a *p-n* junction is formed in a semiconductor material a region of depletion of mobile charge carriers is formed with a high internal electric field which is known as the depletion region. If an electron-hole pair is generated by a photon absorbed within this region, the internal field will cause the electron and hole to separate such that the electron moves to the *n* region and hole to the *p* region. We may detect this charge separation in two ways. If the device is left on open circuit, the externally measurably potential will appear between the *p* and *n* regions. This is known as *photovoltaic mode* of operation. The solar cell is basically a *p-n* junction operated under this condition; it can deliver power to an external load. In INSAT VHRR payload, the 'visible light detector' is of photovoltaic type.

On the other hand, we may short circuit the device externally (or more usually operate it under reverse bias) in which case an external current flows between *p* and *n* regions. This is known as *photoconductive mode* of operation.

There are two devices commonly used, to detect light energy in fiber optic communication receivers. They are PIN diode (*p-i-n* structure) and APDs (avalanche photo-diodes).

The device that provides a good long wavelength response with only relatively modest bias levels has the *p-i-n* structure. In this an intrinsic semiconductor region is inserted between the *p* type semiconductor and *n* type semiconductor materials of the earlier discussed *p-n* junction. (Intrinsic semiconductor is nothing but a highly pure semiconductor). The intrinsic region has a high resistivity, so that only a few volts of reverse bias are needed to cause the depletion region to extend all the way through to *n* region and thus provide a large sensitive volume. In practice, the bias is maintained considerably higher than the minimum value and the intrinsic region remains fully depleted of carriers, even at high light levels.

An APD is a *p-i-p-n* structure. APDs are more sensitive than PIN diodes. The disadvantages of APDs are relatively long transit times and additional internally generated noise due to the avalanche multiplication factor.

Photo diodes are reverse biased; the reverse biased current varies, as the amount of light on the diode junctions. The amount of current is normally extremely small, possibly just a few hundred μA.

7.3 Optical Detector Performance

Light detector performance is dependent on its most important characteristics that are given below.

1. **Responsivity** – It is the measure of conversion efficiency of a photo detector. It is the ratio of the output current of a photo detector to the input optical power and has the unit of amperes per watt. Responsivity is generally given for a particular wavelength or frequency.
2. **Spectral response** – The range of wavelength values that a given photo diode will respond. Generally, relative spectral response is given as a function of wavelength or frequency in the form of a graph.
3. **Transit time** – The time it takes a light-induced carrier to travel across the depletion region of a semiconductor. This parameter decides the maximum bit rate possible with a particular photo diode.
4. **Dark current** – The leakage current that flows through a photo diode with no light inputs. Thermally generated carriers in the diode cause dark currents.
5. **Light sensitivity** – The minimum optical power a light detector can receive and still produce a usable electrical output signal. Light sensitivity is generally given for a particular wavelength in dBm or *dB*μ.

8. CONCLUSION

In this article we had an over view of the nature and characteristics of light. Different light modulation techniques were discussed. The concept and application of optical switching were briefly seen. The principle of optical detection mechanism in different detection devices was outlined. The field of optoelectronic devices is indeed vast; here the topic was restricted to cover devices pertaining to communication systems.

REFERENCES

[1] Alan Roger, "Essentials of Optoelectronics with Applications", Chapman & Hall, 1997

[2] J. Wilson & J. F. B. Hawkes, "Optoelectronics - An Introduction", Prentice-Hall of India Pvt. Ltd., 2001

[3] Ray Tricker, "Optoelectronic and Fiber Optic Technology", Newnes, 2002

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

FUNDAMENTAL PRINCIPLES OF COMMUNICATIONS

D. JOHN

1. INTRODUCTION

Electronics communications are ubiquitous and form essential part of our living. It is encountered in our day to day life as well as in highly sophisticated technical and scientific systems. We come across electronics communication in the working of our mobiles, television and FM receptions, working of ATMs, remote control devices, public internet, office internet, fiber-optic networks, terrestrial microwave systems, satellite communications etc.; in fact the list is endless. The fundamental principles of communications behind all these systems are the same. In all communication systems, essentially information is transferred and or exchanged between users of these systems. The information generated is suitably transformed in to electronic signals and conveyed in turn, using higher frequency signals, which provide higher bandwidth. There are various techniques available to accomplish this; however each system uses the technique that is most suited for its functioning based on appropriate engineering decision by the designer.

2. INFORMATION AND COMMUNICATION

The world around us is filled with information. All the time our senses sense the environment around us for information, continuously increasing our 'intelligence' and our 'instructive knowledge'. Conveyance or transfer of such information is communication. This can be done by means of sound, words, images, moving pictures and actions. In all cases information represents some change from previous state. A flow of information arises from a source in which changes are more or less continuous.

Here we are concerned with transfer of information electromagnetically. Telecommunication is transfer of information at a distance; –'tele' in Greek means distance. Synthesizing and designing systems for transfer of information at a distance is Communication Engineering.

In most cases the information originally generated is not in a form that can be readily transported to the receiver. It goes through a transformation, or coding process, often in more than one step. A communication system will contain the following five elements:

1. The source of information
2. Transducer or/and an encoder, for transforming the information into a form suitable for transportation over the transmission system

3. The transmission system
4. Decoder and/or transducer for transforming the electrical signal in a form suitable for interpretation by the receiver
5. The receiver of information

In general the transmission system is used by several signals simultaneously. This is accomplished by proper design of the coder. Each coder must modify the signal from the associated source so that there is a distinctive difference between the several coded signals. At the same time, the coder must retain the identity of the information. This process is called “modulation”. When modulation is involved in the coding process, the decoder must also be modified to perform two additional functions:

1. Filtering out the desired signal for its receiver and rejecting all other signals.
2. Transforming the signal back into a form in which it can be handled by the normal decoding process. This is called “demodulation”.

The complete system from information source to information receiver is called a communication channel, whether a portion of a system is common to other channels or not.

It should be evident that error may creep in, i.e., the information may be distorted and corrupted in the process of transportation. Error arises from two major causes:

1. Distortion of information may occur as it passes through successive links.
2. Extraneous signals or noise may be introduced which will be interpreted by the receiver as part of the signal originating at the source. This noise may intrude at all links in the channels. There are various sources of noise in the communication links which are to be considered in the design of a communication system.

All electrical communication signals show a fundamental characteristic of rapid change at the transmitting end and unpredictability at the receiver. To the extent that the signal can be predicted at the receiving end, the amount of information is reduced, as, for example, when a musical note repeats its wave-form for a long period.

In order that the electrical system may be designed to handle time-varying signals, more detailed knowledge of their nature is required. The time-varying signals, no matter how complicated may be analyzed in terms of its frequency spectrum, i.e., in terms of frequency components with specified relative amplitudes.

It can be shown by Fourier analysis that single-valued repetitive waveform consists of sine waves and/or cosine waves. The frequency of the lowest-frequency or fundamental is equal to the repetition rate of the waveform and all others are the harmonics of the fundamental. There are infinite number of such harmonics. As a general rule, it may be added, that higher the harmonic, lower is the amplitude and so higher harmonics can be suitably ignored. The time domain and frequency domain representation of a square wave is shown in Figure 1.

It may be noted that communications signals are inherently transient in character. Much of the analysis and synthesis of communication circuits are carried out on a steady-state sinusoidal basis, since transient characteristics of a network is uniquely determined by the steady-state network characteristics. This is evident from the existence of the Fourier integral.

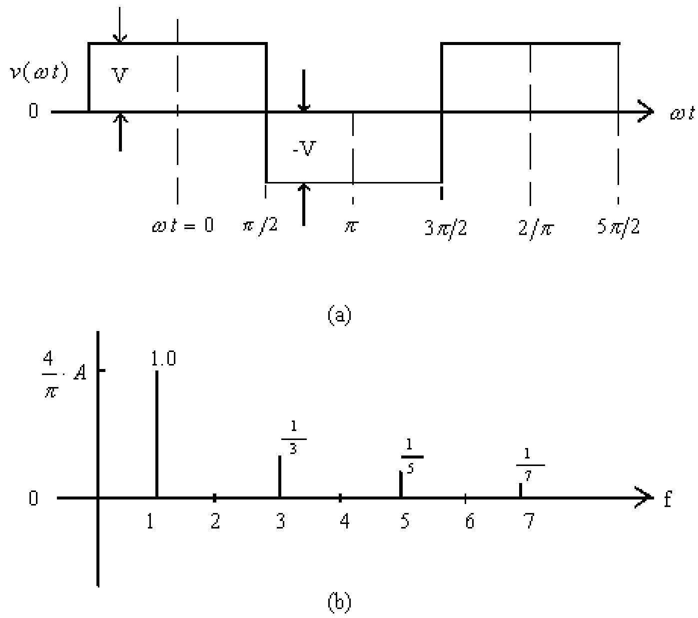


FIG. 1 TIME DOMAIN (A) AND FREQUENCY DOMAIN (B) REPRESENTATION OF A SQUARE WAVE

3. FREQUENCY TRANSLATION

A fundamental principle of communication is that a complex wave, represented by a band of frequencies, can be translated to any other band of at least equal width, if it is so desired, and later retranslated in to its original band. This may be desirable because of a greater effectiveness of the transmitting medium at the new band or in order to convey several messages in the same medium. This translation can be accomplished by a nonlinear process involving multiplication of the signal by a suitable high frequency. There are two reasons for this frequency translation:

1. To have a reasonable size for the antenna. The length of one quarter wavelength antenna for transmitting 1 kHz signal will be 75 km.
2. If all the sources transmit their respective signals, as such, in their original frequency band, they will all be mixed-up making it impossible for the receivers to receive their respective signals.

Modulation of a wave is the process by which a characteristic of a so-called "carrier wave" of a higher frequency is varied in accordance with a time variation of the signal. A general alternating wave may be represented by the equation

$$V = A.\text{Sin}\,(\omega_C t + \theta) \qquad ...(1)$$

Three groups of modulation methods are recognized:

1. Amplitude modulation, where A is varied by the signal.
2. Angle modulation, where *è* is varied by the signal.

3. Pulse modulation, where the signal is turned on or off in pulses, and where the pulses themselves are modulated in amplitude or time of occurrence or width in accordance with the instantaneous values of the signal.

4. AMPLITUDE MODULATION

The amplitude of the carrier is varied in accordance with instantaneous amplitude of the modulating signal. Equation 1 can be written as:

$$V = V_C(1 + m \cdot \text{Sin } \omega_m t) \cdot \text{Sin } \omega_C t \qquad ...(2)$$

Where $m = \dfrac{V_m}{V_C}$ the modulation index

Expanding equation 2, we get

$$V = V_C \cdot \text{Sin } \omega_C t + \frac{mV_C}{2} \text{Cos } (\omega_C - \omega_m)t - \frac{mV_C}{2} \text{Cos } (\omega_C + \omega_m)t \qquad ...(3)$$

Equation 3 illustrates the frequency spectrum of AM: consisting of three discreet frequencies. Of these, central frequency, i.e., the carrier, has the highest amplitude and the other two are placed symmetrically about it, having amplitudes which are equal to each other, but which can never exceed half the carrier amplitude.

Power relation in AM wave is as follows:

The total power in the modulated wave will be

$$P_t = \frac{V_{carr}^2}{R} + \frac{V_{LSB}^2}{R} + \frac{V_{USB}^2}{R} \qquad ...(4)$$

Where all the voltages are root mean square (rms) values and R is the resistance (e.g. antenna resistance where the power is dissipated. It can then be shown:

$$P_C = \frac{V_C^2}{2R} \qquad ...(5)$$

$$P_{LSB} = P_{USB} = \frac{m^2}{4} \times \frac{V_C^2}{2R} \qquad ...(6)$$

$$P_t = P_C\left(1 + \frac{m^2}{2}\right) \qquad ...(7)$$

There are various types of Amplitude Modulation which are in use. Single Side Band modulation (SSB) and Vestigial Side Band modulation are the other popular types of AM used because of their unique advantages. In vestigial side band transmission, a portion of the unwanted lower-side band is also transmitted to eliminate the harmful effects on the received video signals, in TV transmission, caused by the phase response of filters near the edges of the flat pass-band.

4.1 Modulation and Demodulation of AM

There are two possibilities; one, high-level modulation and the other low-level modulation followed by amplification using a linear Class B power amplifier. For high level modulation it is possible to apply the modulating signal in series with the DC supply for the final class C amplifier. This is called plate modulation (or collector modulation in the case of low power AM transmitter). The AM receiver has a simple diode rectifier for demodulation of the signal.

5. FREQUENCY MODULATION

Frequency modulation is a system in which the amplitude of the modulated carrier is kept constant, while its frequency is varied by the modulating signal. Phase modulation is a similar system in which the phase of the carrier is varied instead of its frequency; as in FM, the amplitude of the carrier remains constant.

The instantaneous frequency of the carrier, modulated by the signal $V_m \text{Cos } \omega_m t$, is given by:

$$f = f_C(1 + V_m \text{ Cos } \omega_m t) \qquad \text{...(8)}$$

The maximum deviation for this signal will occur when the Cosine term takes the maximum value, that is, ±1. Under these conditions the instantaneous frequency will be

$$f = f_C(1 \pm V_m) \qquad \text{...(9)}$$

So the Maximum frequency deviation $\Delta f = f_C V_m$

Consider the equation of the carrier:

$$V = A \cdot \text{Sin } \theta \qquad \text{...(10)}$$

From equation 8 it can be written:

$$\omega = \omega_C(1 + V_m \text{ Cos } \omega_m t) \qquad \text{...(11)}$$

We know $\theta = \int \omega \cdot dt$

Therefore $\theta = \omega_C t + \omega_C \dfrac{V_m}{\omega_m} \text{Sin } \omega_m t$

The above can be rewritten as

$$\theta = \omega_C t + \frac{f_C V_m}{f_m} \text{Sin } \omega_m t \qquad \text{...(12)}$$

$$\theta = \omega_C t + \frac{\Delta f}{f_m} \sin \omega_m t \qquad \text{...(13)}$$

Substituting the above in equation 10 we get the equation for FM modulated signal.

$$V = A \cdot \text{Sin } (\omega_C t + \beta \cdot \text{Sin } \omega_m t) \qquad \text{...(14)}$$

Where β is the modulation index of the FM and is defined as $\beta = \dfrac{\Delta f}{f_m}$

It should be noted that as the modulation frequency decreases and the modulating voltage amplitude (i.e., V_m) remains constant, the modulation Index increases. However in PM modulation the modulation index β remains the same as long as the modulating voltage amplitude (i.e., V_m) remains constant. The modulation index β is measured in radians.

Equation 14 is mathematically more complex than the equation for AM. It is an equation of sine of a sine. The solution involves the use of Bessel functions of the *first kind*. Using Bessel functions, equation 14 can be expanded. From the resulting terms it can be seen there is a carrier component and infinite number of pairs of sidebands.

Note:

1. Unlike in AM, FM has an infinite number of sidebands.
2. Sidebands at equal distances from the carrier on either side have equal amplitude.
3. The sideband level gradually diminishes on either side of the carrier but not in any simple manner. The modulation Index determines how many sideband components have significant amplitudes.
4. In FM, unlike in AM, the amplitude of the *carrier component* does not remain constant. It depends on the modulation index *â*. For certain values of â (modulation index), called *eigenvalues,* the carrier component of the FM signal disappears. They are 2.4, 5.5, 8.6, 11.8 and so on.

The practical band width required for FM, as rule of thumb, given by Carson's rule, states that it is twice the sum of maximum deviation and the highest modulating frequency, which is a good approximation.

$$BW_{FM} = 2(\Delta f + f_m) \qquad ...(15)$$

6. PHASE MODULATION

Phase modulation is a system where the phase of the carrier is varied by the modulating signal; however, the amplitude of the carrier remains constant.

The equation for the PM is:

$$V = A \cdot \text{Sin}\ (\omega_C t + \beta_{PM} \cdot \text{Sin}\ \omega_m t) \qquad ...(16)$$

β_{PM}, is the maximum phase deviation

In terms of frequencies $\beta_{PM} = \dfrac{\Delta f_{Max}}{f_m}$ radians ...(17)

In Phase Modulation the maximum phase deviation β_{PM} is proportional to the maximum amplitude of the modulating signal while in FM the maximum frequency deviation, Δf_{Max} is proportional to the maximum amplitude of the modulating signal.

In *space telemetry* Phase Modulation is use due to the following advantages:

1. There is no *threshold* for PM demodulation. Therefore communication can be effective even at low signal levels.

2. Since the carrier frequency is stable unlike in FM carrier can be used for 'Doppler Tracking', which is a requirement in determining the location of the satellite.

Comparison of Eqs.14 and 16 shows them to be identical, except for different definitions of the modulation index. It is obvious therefore that these two forms of angle modulation are indeed similar.

7. PRE-EMPHASIS AND DE-EMPHASIS IN FM

Pre-emphasis is done at the FM transmitter to boost the higher frequencies of the modulating signals. This is to give advantage to higher frequencies over the noise present. De-emphasis is done at FM receiver to compensate and bring down the levels of the higher frequencies at the receiving end.

The network used for pre-emphasis is a differentiator [High Pass Filter (6 dB/octave)] and the network used for de-emphasis is an integrator [Low Pass Filter (-6 dB/octave)].

From the derivation of the equation for the FM, one can deduce that if the signal is passed through an integrator (inverse frequency network) and subsequently phase modulated, FM modulated output is obtained. Conversely phase modulation is frequency modulation with pre-emphasis.

8. PULSE MODULATION

Both amplitude modulation and angle modulation permit the separation of the signal by means of frequency selective networks. This method of coding and decoding for transmission of signals through a common transmission medium or network is generally called 'frequency-division' system.

The use of pulse modulation introduces the possibility of another method of identifying signals of a particular channel on a 'time-division' basis.

Pulse modulation is based on the concept of sampling the signal in accordance with the sampling theorem stated below:

"A signal of duration T whose highest frequency of importance is f_h can be completely specified by $2Tf_h$ [or $2TB$, where B is the band width of the low-pass (LP) network necessary to transmit the signal] samples taken at equal time intervals of $1/2f_h$ sec."

The theorem can be proved using Fourier series. For a continuous signal (rather than one of duration T) the theorem may be restated as follows:

"A continuous signal, whose highest frequency of importance is B can be completely specified by samples taken at a rate of $2B$ per second."

9. HARTLEY-SHANNON LAW

From the above statements it follows that the amount of information in a signal of duration T and the maximum frequency B can be represented by $2BT$ pulses.

Consider the information in a series of pulses describing a signal, each pulse having a voltage or current amplitude proportional to $\sqrt{S+N}$ and an uncertainty in value proportional to $\sqrt{N}$. Then the number of possible recognizable values of signal in the presence of noise is

$$n = \log_2 \sqrt{\frac{S+N}{N}} = \frac{1}{2}\log_2\left(1+\frac{S}{N}\right) \qquad ...(18)$$

Since the information in the original signal can be represented by $2BT$ pulses, the total information *'I'* in the signal will be

$$I = TB \log_2\left(1+\frac{S}{N}\right) \text{ Bits} \qquad ...(19)$$

Here, the *bit* (binary digit) is defined as the quantity of information required to permit the correct selection of one out of a pair of equiprobable values.

The rate of transmission of information, or the capacity, of the channel in bits per second will be

$$C = \frac{I}{T} = B \log_2\left(1+\frac{S}{N}\right) \text{ Bits/sec} \qquad ...(20)$$

Equation 20 is known as Hartley-Shannon law. Hartley provided the concept of the effect of bandwidth, and Shannon added the important effect of *S/N* ratio. The law shows that an increase in information rate can be obtained by increasing either the bandwidth or the signal power or, alternatively by keeping down the noise power.

However it should be recognized that every communication system does not achieve the full capabilities inherent in the power and the bandwidth used. Design of coding and decoding systems has much to do with the degree to which the limit of channel capacity is approached. Frequency modulation is just one method of obtaining improved *S/N* ratio using wider bandwidth.

10. HETRODYNE ACTION AND SUPER HETERODYNE RADIO RECEIVER

Heterodyne action results by super imposing two waves of different frequencies. When two waves are super imposed the envelope of the combined wave fluctuates in amplitude and the frequency of the envelope is the difference between the frequencies. This is called the beat-frequency. Here the two signals are added and no multiplication is involved. If one of them is a pure carrier and the other amplitude modulated, the beat- frequency possesses the amplitude modulation. Also if on the other hand, the signal is frequency modulated, the beat-frequency also possesses the frequency modulation.

Super heterodyne receiver has a mixer that translates the received carrier frequency to an intermediate frequency (IF). The mixer is a multiplier and it mixes the incoming carrier frequency with a local oscillator frequency. The sum and difference frequencies are produced. The difference frequency is selected rejecting the sum frequency using filter and amplified in the IF amplifier. This frequency is then sent to the demodulator which is centered at the IF.

The receiver characteristics that are important are sensitivity, selectivity and fidelity. And all these characteristics depend on the IF amplifier. The RF Amplifier which is the first stage of a receiver couples the receiver to the antenna. It also provides discrimination or selectivity against the *image frequency* and intermediate frequency. Noise figure of the receiver is also decided by the RF amplifier. The receiver is an important element in the communication system; hence the reader may undertake further study on the receiver characteristics and system parameters.

11. FREQUENCY DIVISION MULTIPLEXING (FDM)

FDM involves each individual channel being shifted in frequency and the frequency spectra placed contiguously. The original channels may be recovered at the receiving end by de-multiplexing with filtering and frequency translation back to base band. Essentially each user is single-sideband modulated on to a different carrier frequency and they are summed together. Small guard bands are inserted between each channel to prevent frequency overlap and to facilitate de-multiplexing. As an example one can look at the FDM employed in the satellite communication to share the allotted frequency band among the users.

12. CONCLUSION

The function of transmitting information by electronic communications system is accomplished by the techniques outlined above. As the demand for communication increases, with more users seeking to share the common communication channel in all the frequency bands, an appropriate mix of frequency-division and time-division techniques is employed and new frequency bands are brought in to use. Over the years digital systems (time-division) are gradually replacing analogue systems (frequency-division), in view of the obvious advantage of the former. Today a 36 MHz channel can carry eleven TV channels. Comparing this with the early days of satellite communication when 36 MHz was required for one TV channel, it is evident that the growth is indeed phenomenal and this can be clearly attributed to advances in digital communications and data compression techniques.

REFERENCES

[1] George Kennedy, "Electronic Communication Systems" McGraw Hill, 2002
[2] B.P. Lathi, "Modern Digital and Analog Communication Systems" Oxford University Press, 1995

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

WIRELESS COMMUNICATION[1]

P.S. DHEKNE*

1. INTRODUCTION

Mobile communication is the fastest growing field in the ICT industry. Recent advances in wireless communication are rapidly gaining deployment all over the globe because of its ease of deployment, ad hoc connectivity and cost-effectiveness. The key opportunity that stands out in this emerging technology is its reach and affordable cost. Upcoming wireless technologies starting from ultra short-range RF tags, Bluetooth, Wi-fi to long haul cell phones have opened up variety of new and innovative applications in many areas. Increasingly, manufacturers are turning to wireless technologies to reduce theft and improve delivery efficiency. While satellite-based global positioning systems (GPS) have drawn considerable interest, a mobile data system that uses global system for mobile communications (GSM) and General Packet Radio Service (GPRS) cellular networks to track merchandise shipments and individual goods via the Web have opened many new applications where distributed information gathering is an integral part. In particular Wireless Sensor Networks (WSN), which are low power devices equipped with programmable computing and multiple sensing capabilities networked by wireless radio are deployed in number of remote monitoring applications. This paper provides an overview of different types of commercially available wireless communication systems and future challenges are also discussed.

2. RFID TECHNOLOGY

The Radio Frequency Identification Cards (RFID) or tags that emits radio signals, which can be picked up by devices called readers, has enormous potential to track anything from livestock to products. Unlike barcodes, RFID tags / cards transmit information without contact and line of sight and can be read through rain, snow, fog, ice and paint. It is estimated that RFID technology could help eliminate hundreds of millions of dollars by reducing inventory by detecting consumption of materials via RF tags. Read/write RFID tags can be used as bill of material for transport, or process data, smart cards for e-commerce, quality-control data and personnel security data. RFID could even facilitate real-time manufacturing.

---

*Raja Ramanna Fellow, BARC, Tromaby, Mumbai 400085
[1]This paper is based on lectures delivered at TIFAC-CORE, SASTRA College, Tanjavur

The prices for RF-readers, middleware and system integration are expected to be very competitive, as expected volumes are going to be enormous. Low frequency 30-500 KHz. systems are used for security access, asset tracking and animal identification applications. They have shorter range of approx. few centimeters. High-frequency 850-950 MHz. and 2.4 – 2.5 GHz. systems, offer ranges of over 90 feet and high reading speeds. They are used in railroad, car tracking and automated toll collection. The RFID cards are available around $ 2 – 3 to over $ 6-7 in quantities depending upon data storage capacity and processing capability.

3. BLUETOOTH TECHNOLOGY

Bluetooth, a short-range wireless technology was unveiled in 1998. This technology is named after Danish King Harold Bluetooth, who reported to have united Denmark and Norway. Bluetooth is supposed to unite the world of consumer electronics and computers and make all devices in the living room as well as in the office talk to each other.

Bluetooth is widely promoted as ‘cable replacement’ solution, which works within 10 to 100 meters at 720 kbps. We are going to have lots of new generation cars in year 2004, which will be Bluetooth integrated and let you use all mobile functionality from dash-board and steering wheel. Latest generation of keyboards and mouse devices use Bluetooth USB to connect to a PC. Also Palmtops, mobile devices are increasingly using Bluetooth technology.

With latest Bluetooth 1.2 standard lot of issues related to older generation technology 1.1 of frequency interference with TV & microwave oven etc. have been resolved and have improved sound and data transmission capability.

4. WI-FI TECHNOLOGY

High bandwidth Wireless Local Area Networks (LAN with IEEE 802.11a, b & g standard) popularly known as ‘Wi-Fi’ are rapidly gaining deployment all over the globe.

With the ability to access the net from anywhere within the office or on the road at today’s fractional cost that too at much higher bandwidths, Wi-Fi have the potential to radically enable and enhance many new online services & applications. Coupled with hand held (personnel Digital Assistants) consumer-friendly portable devices many innovative applications are emerging in the market.

The Wi-Fi networks operating in the 2.4 GHz. Band, has several advantages, both technical and commercial, and thus we cannot ignore its emergence. Wi-Fi’s current bandwidth of 11 Mbps (IEEE 802.11b) has increased to 54 Mbps (IEEE 802.11 a & g). It supports multimedia communication including digital video & audio. Major computer vendors such as Intel and CISCO are investing billions of dollars into its development and are proposing it as the mainstream-networking infrastructure of the future. Many enterprises go for complete wireless connected network and the technology is penetrating every field like roadways, railroads, chemical, medical and geological, etc.

5. WI-MAX TECHNOLOGY

The name "WiMAX" was created by the WiMax Forum, which was formed in June 2001 to promote conformance and interoperability of the standard. The forum describes WiMAX as a standards-based technology enabling the delivery of last mile wireless broadband access as an alternative to cable and DSL.

WiMAX technology is intended for wireless "metropolitan area networks" and can provide broadband wireless access (BWA) up to 30 miles (50 km) for fixed stations, and 3 - 10 miles (5 - 15 km) for mobile stations. WiMAX is a second-generation protocol that allows for more efficient bandwidth use, interference avoidance, and is intended to allow higher data rates over longer distances.

WiMAX can be used for wireless networking in much the same way as the more common Wi-Fi protocol and provides wireless data in a variety of ways, from point to point fashion to full mobile cellular type access.

The IEEE 802.16 standard defines the technical features of the communications protocol. WiMax operates in 2.4 GHz ISM band with data rates supporting Speeds up to 40-70 Mbps. WiMAX operates on both licensed and non-licensed frequencies, providing a regulated environment and viable economic model for wireless carriers.

6. CELLULAR TECHNOLOGY

Most cell phones worldwide have the ability to send and receive short text messages, and an increasing number now incorporate more advanced Internet capabilities. Almost all mobile phones except point-to-point radio are cellular: they divide their coverage area into cells, at the center of which sits a Base Transceiver Station (BTS). The simplest cellular systems use Frequency Division Multiplexing (FDM), which gives each conversation its own waveband, just like a radio station. In contrast, second generation (2G) systems are digital. In the USA, two standards are used for second-generation systems: IS-95 (CDMA) and IS-136 (D-AMPS). Europe consolidated to one system called the global system for mobile communications (GSM). Japan uses a system called personal digital cellular (PDC).

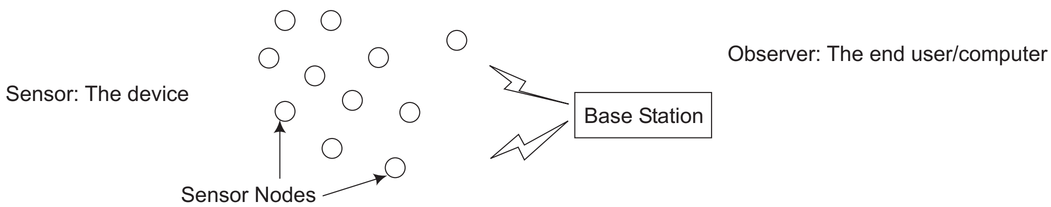
Current cell phone systems built for voice, offer low bit rates and poor interoperability. Next generation mobile phones using 3G and 4G protocols are expected to be still smaller in size, cheaper in cost and much faster in terms of data rates. These aim to provide packet-switched data to a handheld terminal with throughputs measured in hundreds of Mbits/sec. (2 Mbps and 20 Mbps) which, not only encompasses the delivery of voice services to people regardless of location, but also that of data services regardless of location, network, or terminal used. Industry is working on speeding up the development of mobile communication technologies. The ideal communication system where both voice and data services can be delivered regardless of location, network, or terminal may be possible in the near future.

7. WIRELESS SENSORY NETWORKS

Wireless sensory network consists, of a large number of densely deployed sensor nodes. These nodes incorporate wireless transceivers so that communication and networking are enabled.

Additionally, the network possesses self-organizing capability so that little or no network setup is required. Ideally, individual nodes should be battery powered with a long lifetime and should cost very little.

Over the last decade, the cost of computational, sensing and display systems has dropped dramatically, but the cost of electronic connectors and labor has remained relatively constant. At the same time, semiconductor technology has evolved to the point where it is now possible to manufacture low-power radio frequency subsystems that rival the cost of physical electronic connectors. The advent of nano-technology has made it technologically feasible and economically viable to develop low-power devices that integrate general-purpose computing with multi-purpose sensing and wireless communications capabilities.



Phenomenon: The entity of interest to the observer

Aggregating sensor nodes into sophisticated computational and communication infrastructures, called sensor networks, will have a significant impact on a wide array of applications ranging from military, to scientific, to industrial, to health-care, to domestic. Self-organizing, scalable data network for device-to-device and device-to-enterprise applications will become ubiquitous. This paradigm shift will bring smart systems into our everyday lives by overcoming the real barriers to wireless device networking. The typical commercial applications include: inventory management, product quality monitoring, factory process monitoring, disaster area monitoring, biometrics monitoring, and surveillance [Fig.3 & Fig.4].

The fundamental goal of a sensor network is to produce globally meaningful information from raw local data obtained by individual sensor nodes. Intel has taken the initiative to invest in R & D of sensor networks, recognizing this technology as crucial to addressing the pending global age wave and public health crisis.

Wired sensor networks have been around for decades, with an array of gauges measuring temperature, fluid levels, humidity and other attributes on pipelines, pumps, generators and manufacturing lines. Also wired in are actuators, which let the control panel slow down a pump or turn on a heater or a fan in response to the sensor data.

Now advances in silicon radio chips, coupled with cleverly crafted routing algorithms and network software are promising to eliminate those wires, and their installation and maintenance costs [Fig.1 & Fig.2]. Mesh network topologies will let these wireless networks route around nodes that fail or whose radio signal is hammered by interference from heavy equipment. A gateway will create a two-way link with legacy control systems, hosts, wired LANs or the Internet.

These networks can use several different wireless technologies, including IEEE 802.11 wireless LANs, and radio frequency identification. But low-power radios that have a range of about 30 to

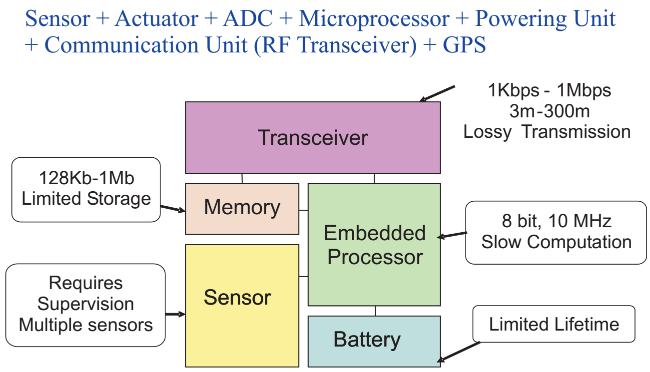


FIG. 1

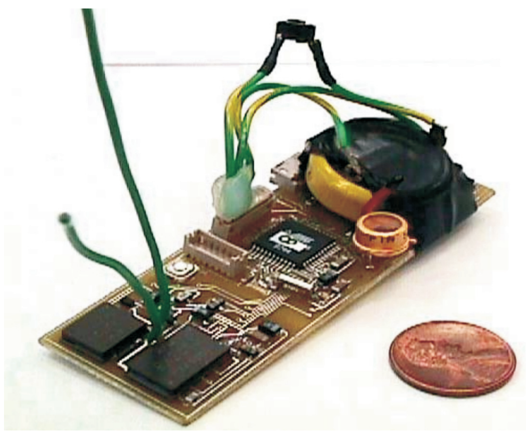


FIG. 2

200 feet and data rates of up to around 300K bit/sec is used commonly. Most of these, with their accompanying network software and APIs, are proprietary products. But the IEEE last year approved the low-rate standard for a simple, short-range wireless network whose radio components could run several years on a single battery. The ZigBee Association, a group of vendors, has finalized an industry specification for the network software that will run on the 802.15.4 radio chips [3, 4, 5].

Crossbow's wireless sensor platform gives you the flexibility to create powerful, tether-less, and automated data collection and monitoring systems. Crossbow's supports a wide range of hardware and sensors for various customer requirements. Most of the hardware can plug-and-play and it all runs TinyOS / nesC from UC Berkeley. The platform consists of Processor Radio boards (MPR) commonly referred to as MOTES. These battery-powered devices run TinyOS and support two-way mesh radio networks. Sensor and data acquisition cards (MTS and MDA) plug into the Mote Processor Radio Boards. Sensor support includes both direct sensing as well interfaces for external

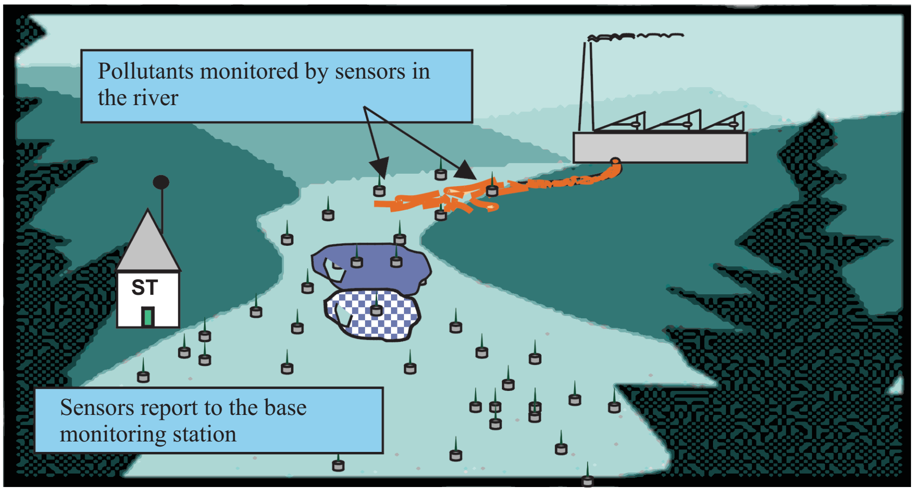


FIG. 3

FIG. 4 SENSORS EQUIPPED WITH BLUETOOTH

sensors. The TinyOS operating system is open-source, extendable, and scalable. Code modules are wired together allowing fluent-C programmers to custom design systems quickly. Accessory products include antennae, cables, and packaging.

7.1 Challenges and Opportunities

It is obvious that the electronics hardware market will grow. The entertainment, communication and computer market is driving the technology with mobile, cell phones and high-end computers playing the key role. At present, the hardware industry across the world is of the order of US $200 billion annually. The automobile market is another segment in which the role of electronics is increasing day by day. Many other services are also becoming more and more computer hardware oriented. A very interesting new field is emerging in the area of MEMS and biosensors, which integrate the capability of MEMS and biological cells. This is an emerging field even in the advanced countries and it is possible to produce high end products at relatively inexpensive facilities with coarse geometries. Another device, which will be useful, is a tiny semiconductor digital camera for remote monitoring and control.

The concept of **fabless industries** is catching fast. These industries sub-contract every activity connected with realization of VLSI chips starting from design, silicon processing, packaging, testing to marketing. Also there are few vendors who offer wireless sensory components.

Sensory's ICs and embedded software are used in consumer electronics, cell phones, PDA's, Internet appliances, interactive toys, automobiles, and other applications where low cost and high qualities are essential.

We need to be innovative in envisaging the applications and put these into use. There are numerous applications, where effective use of next generation wireless sensor technologies can be thought of.

Indoors and Outdoors Environmental Monitoring
Monitoring of Gamma Radiation levels near nuclear plants
Seismic and Structural Monitoring
Power Monitoring
Factory and Process Automation
Classroom / Lab Training on Wireless Networking

The sensors were often power-hungry, so data polling and transmission had to be fine-tuned not to exhaust their batteries. Some data aggregation will be needed at the level of the sensor network. Interfacing with networks complicates the picture even more, and little work has been done on that.

7.2 Research Issues in Sensory Network

Even though sensory networks are being used for number of real-life applications, there are number of unexplored areas in this field such as selection of optimum size sensor network, fault tolerant configuration, redundancy, best routing algorithms, effective use of network bandwidth, error recovery & control, securing data during transmission and many others, which is a subject of advanced research [1].

These issues have been studied using network-based simulators, but they lack practical approach and results produced through simulations may not be applicable directly. Thus there is a need to carry out experiments to verify and validate the actual values with simulated results.

8. SECURITY ASPECTS

When the wireless communications is coming to the offices and the homes, there are new security issues to be taken care of. Today ICT industry is facing greater challenge in providing secure wireless communication. Enhancing the wireless networks with the best security encryption techniques and protocols is a prerequisite. Lot of active research is happening in this area.

Therefore deployment of communication device in motion – battlefield co-ordination, vehicular application, tracking, etc. where wireless is the only alternative, we must use encryption techniques and robust protocols for ensuring secure communication. Before deploying wireless technology, detailed risk analysis of various types of threats [2] must be carried out and we must have counter measure plan to stop & defeat undesired use of information/data while in transmission.

9. CONCLUSION

Wireless technology can provide unconstrained and undeterred secure information transmission. At consumer level, the sudden surge and unabated success of wireless can be attributed to three features of convenience: high bandwidth & reliability, low cost and local reach.

With the advancement of variety of cost effective wireless technologies and enormous capabilities wireless products are rapidly gaining user acceptance and would replace wired devices sooner than you may think.

Wireless device coupled with processing capability, net connectivity and suitable sensors have tremendous potential for deployment in many areas of Control & Instrumentation. We can envisions a proactive computing world in which a multitude of unseen, connected computing nodes automatically acquire and act on real-time data about a physical environment, helping to improve lives, promoting a better understanding of the world and enabling people to become more productive.

In most cases, the information provided by the sensor may be incomplete, inconsistent or imprecise. In many cases some ambiguities will be caused when we use only one kind of sensor to perceive the real world. Additional resources may provide complementary data in addition to the redundant information contents. Merging of redundant data can reduce imprecision and fusion of complementary data can crease a more consistent interpretation. Based on overcoming the ambiguity problem, fusion can also bring some benefits to our perception of the real world. Also lot of work needs to be done in software and data management.

10. ACKNOWLEDGEMENTS

This article is based on the series of lectures delivered by the author at TIFAC-CORE, SASTRA, Engineering College, Tanjavur (TN), under Distinguished Visiting Professorship award scheme with AICTE support and implemented by INAE. Author wishes to thank Prof K. Ushadevi, staff members & students for their interest & encouragement. Thanks are also due to INAE for the sponsorship and pursuance to convert lecture material into an article. Author also wishes to thank DAE authorities for the encouragement and support to carryout this work.

REFERENCES

1. Jayesh Kataria, P.S. Dhekne and Sugata Sanyal, "ACRR:On Demand Distance Vector Routing with Controlled Route Requests" International Journal Computers, Information

Technology and Engineering (IJCITAE) Vol. 1, No. 1, pp 9-15, June 2007, Serial Publications

2. Jayesh Kataria, P.S. Dhekne, and Sugata Sanyal, “A Scheme to Control Flooding of Fake Route Requests in Ad-hoc Networks”, International Conference on Computers and Devices for Communications, CODEC-06, December 18-20, 2006, Kolkata, India
3. P.S. Dhekne, “Wireless Sensor Networks for Health Care”, an invited paper presented during “National Workshop On Digital Signal Processing and its Applications in Communication, Control and Biomedical Engg.” at RAIT, Navi Mumbai held between July 3-6, 2006
4. P.S. Dhekne, “Wireless Sensor Networks” an invited talk given at K.K. Wagh College of Engineering, Nasik during March 2006 and at University of Cincinnati, Cincinnati, USA, Axes Technologies, Texas, USA during October, 2004
5. Prof. Anup Kumar, CECS, University of Louisville, Louisville, KY 40292, USA, Mr. P.S. Dhekne, BARC, Mumbai, India and Prof. Dharma P. Agrawal, ECECS, University of Cincinnati, Cincinnati, OH 45221-0030, USA, “Challenges in utilizing Aggregate Logical Sensor Units for Wireless Sensory Information Networks” an invited paper during 16th International Symposium on Personal Indoor and Mobile Radio Communications held at Berlin, Germany during September 11-14, 2005

AEROSPACE ENGINEERING

III-rd Proof – \\server\e\book work
27-05-10

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

ADVANCED MATERIALS FOR AIRCRAFT AND HELICOPTERS

C. G. KRISHNADAS NAIR*

1. INTRODUCTION

Development of high speed aircraft with ever increasing demands for reducing structural weight requires the use of materials close to the limit of safety regarding their strength, fatigue resistance, elastic modulus, corrosion resistance etc. Consequently, the demands of safety are very stringent and established reliability of performance is a great asset of conventional aircraft materials when compared to new materials. As a result airframes are still to a large extent made of duralumin type alloys introduced in 1911. However, the development of high speed aircraft and engine with higher operating temperatures has catalyzed development of a series of new alloys including stronger and lighter aluminum, magnesium and titanium alloys, nickel, cobalt and iron based super alloys and various high strength steels. Also, there has been development in composite materials such as fibre reinforced plastic, metal and ceramic composites, metal to metal bonded and foam/honeycomb sandwich structures.

2. REQUIREMENTS OF MATERIALS

The objective of the designer is to achieve complete structural integrity at minimum structural weight and at minimum cost. A number of factors are to be considered in the selection of materials for airframes. These include specific gravity, static strength and stiffness, resistance to impact loads, temperature effects, corrosion resistance fatigue strength, rate of crack propagation, etc. and the cost of raw materials and fabrication, as well as fabricability. In addition to these, thermal fatigue, creep strength and hot corrosion/erosion resistance are important properties for materials of high temperature aeroengine components.

2.1 Strength and Rigidity

Static structural efficiency which takes into account both strength (or rigidity) and density is an important parameter on which the designer bases his choice of material. It is measured by 0.2%

*Former Chairman, Hindustan Aeronautics Limited (HAL), President, Society of Indian Aerospace Technologies & Industries (SIATI), Aeronautical Society Building, Suranjandas Road, Off. Old Madras Road, Bangalore 560 075

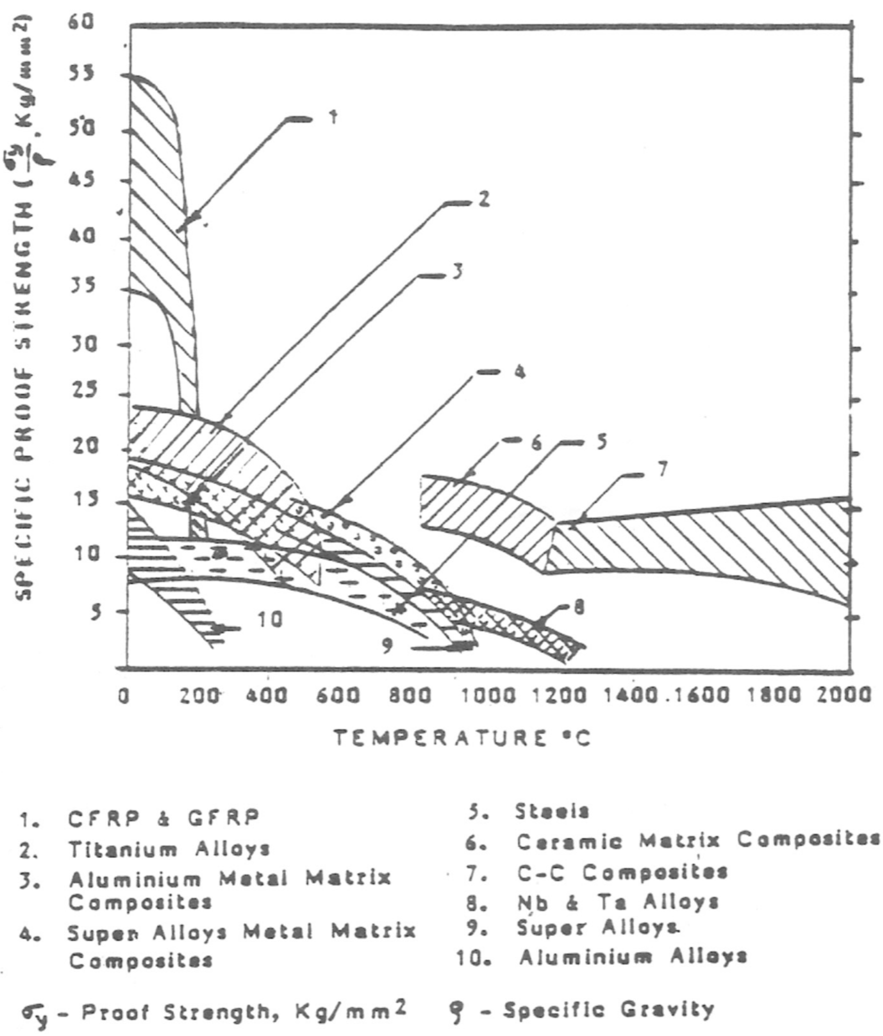
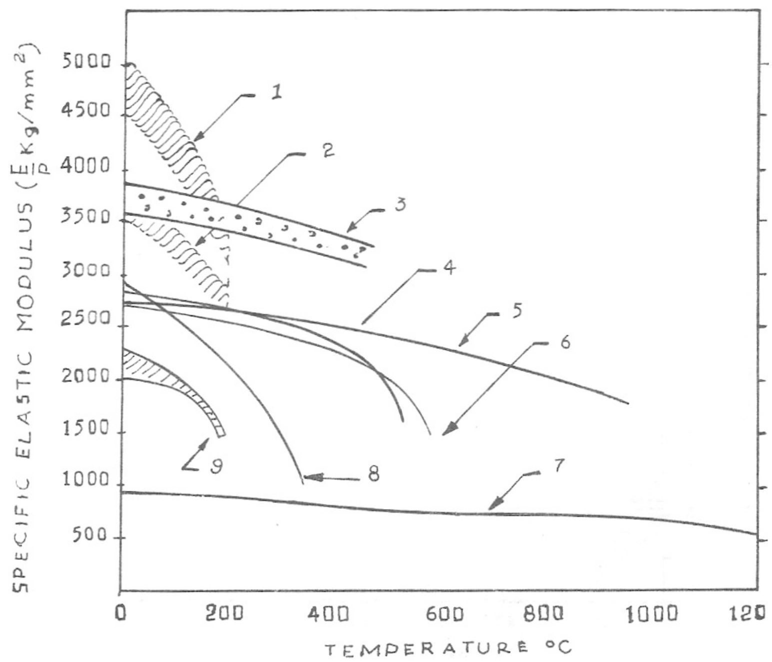


FIG. 1 RELATIONSHIPS BETWEEN SPECIFIC PROOF STRENGTH AND OPERATING TEMPERATURES FOR AIRCRAFT MATERIALS

proof strength's E/ρ or G/ρ ratios where ρ is the density and E&G are tensile and shear moduli respectively. Figure 1 is a plot of specific proof strength versus temperature for a number of common and potential aircraft materials. Figure 2 is a similar plot for specific modulus versus temperature. Aluminum alloys lie at the lower left hand side. Concorde operates at a surface temperature of about 130°C and this is just about the limit of aluminum alloys.

Titanium offers a potential airframe materials upto 500°C and for aeroengine compressor blades/ discs and other components. The refractory metals, ceramics and metallo-ceramic composites have potential for high temperature aeroengine components beyond the capacity of super alloys. While



1 Unidirectional CFRP
2. Woven CFRP
3. Al. MMC
4. Titanium Alloys
5. Nickel Base Alloys
6. Steels
7. Tantalium
8. Aluminium Alloys
9. GFRP

E - Elastic Modules. kg/mm^2 ρ - Specific Gravity

FIG. 2 RELATIONSHIP BETWEEN SPECIFIC MODULUS AND OPERATING TEMPERATURES FOR AIRCRAFT MATERIALS

CFRP and GFRP offer the highest specific proof strength, the GFRP have much lower rigidity than that of the aluminum alloys and will require special stiffening by design. The fibre reinforced plastics although, offer higher strength to weight ratio, have lower temperatures capability than aluminum based alloys. Metal matrix composites based on aluminum alloys and Ni, Co etc., extend the temperature capability, especially with reference to modulus of elasticity (rigidity).

2.2 Fatigue Strength

Failure of components due to fatigue is well known to aircraft designers. It is essential that aircraft materials should have sufficiently high fatigue strength to take full advantage of their ultimate strength, this is not always true. Large improvements in strength of metals obtained by alloying and heat treatment does not produce corresponding improvement in fatigue strength. Alloys used in materials are therefore, not in the highest strength condition, but in a tempered condition which has better resistance to fatigue crack propagation. Metal to metal bonded strictures as compared to riveted structures are superior in fatigue strength. The fiber reinforced composites also offer-better fatigue properties as compared to metals.

2.3 Toughness and Rate of Crack Propagation

Resistance to crack propagation is an important factor in the selection of aircraft materials. The materials should be tough to resist formation of cracks. Once a crack is formed, the rate of propagation should be slow enough to give sufficient time for the detection of crack and for remedial action before catastrophic failure occurs. In the case of metals generally, the stronger the materials the faster the rate of crack propagation in them. Fiber reinforced composites and bonded structures have a distinctive advantage over metals, as the interface in bonded structures and fibres act as crack arrestors. The property is assessed by a variety techniques, including the conventional charpy and izod impact test, and the more sophisticated fracture toughness test and direct measurement of crack propagation by instrumented impact and fatigue test.

2.4 Stress Corrosion

Another source of failure of aircraft materials is stress corrosion which is failure by cracking due to the combined action residual stresses and corrosive environment. Hence, aircraft materials should have adequate resistance or should be protected against it. Modification of alloy composition (e.g.; small additions of Ag, Zr and Cr in Al- -Zn alloys) and special heat treatments (stepped quenching and double ageing treatments for Al- -Zn alloys) are used to minimize the susceptibility to stress corrosion. The occurrence of stress corrosion can also be reduced by keeping down the residual stress level while machining, quenching and assembling.

2.5 Thermal Shock, Creep and Hot Corrosion/Erosion Resistance

High temperature components of aeroengines, for example, turbine blades may develop surface cracks due to repeated thermally induced stresses and strains resulting from thermal gradients and repeated heating and cooling. This is also known as thermal fatigue. Creep is the property of undergoing permanent strain under constant stress (below the yield strength) at elevated temperature. Such permanent strain can alter the dimensions of precision components and affect the performance of aeroengines. Hence high temperature materials for aeroengine components are also evaluated on the basis of thermal fatigue and creep properties. Hot end components of the engines are also subjected to serve corrosive atmosphere and erosion by hot gases/combustion products. Coating materials based on ceramics, intermetallic and metallio-ceramic composites and special process have been developed to coat engine components for enhancing their hot corrosion/erosion resistance

2.6 Simulated Service Tests

The aircraft and the materials used for its construction are subjected to rapid and gradual changes in climatic conditions such as temperature and humidity. These may affect the physical and mechanical properties of the materials used. For example, the rigid cellular plastics used for making sandwich panels may undergo changes in dimension and strength due to change in temperature and humidity. Glass and plexiglass used in the construction of canopies and wind-shield may be cooled from ambient temperature to 0°C about 3- -5 minutes and then heated to nearly 40°C in about the same time when an aircraft climbs to an altitude where temperature is colder and then descends to warmer temperatures. Reinforced plastics and adhesives used in modern aircraft may undergo undesirable changes when exposed to variations in temperature, humidity and ultraviolet radiation. Hence, materials are tested under simulated service conditions before they are selected for specific applications.

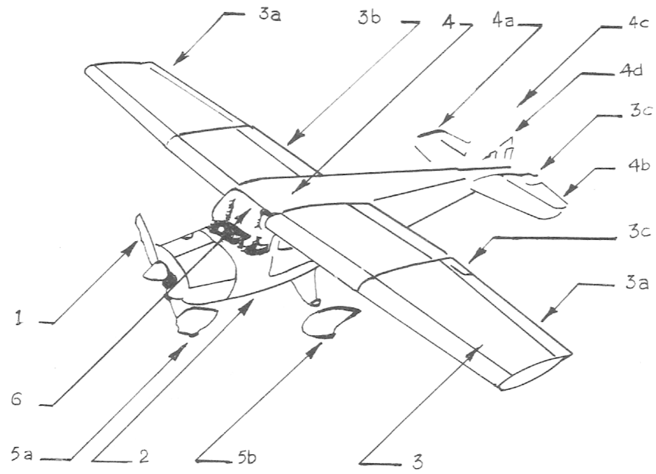
2.7 Cost of Materials and Fabrication

Cost is another factor influencing the selection of materials. Although, titanium alloys are superior to aluminum alloys in strength to weight ratio, they cost too much compared to aluminum alloys and hence can replace the same only beyond the temperature capability of aluminum alloys. The carbon fiber and Kevlar fiber reinforced plastics, metal matrix composites and the high strength ultra light aluminum – lithium alloys are orders greater in cost as compared to conventional materials. The reduced structural weight and the resulting higher performance is to be matched against the cost for selection of appropriate materials. The cost of materials should also be considered together with the cost of fabrication. Castability forgeability, machineablity, formability (as in the case of bending, stretch forming deep drawing, press forming for sheet metals) and weldability are important criteria to be considered for specific shape of components and their applications.

3. MATERIALS AND APPLICATIONS

Figure 3 and 4 illustrate major structures of aircraft and materials used. Figure 5 and 6 illustrate major structures of typical helicopter and materials used. Figure 7 and 8 illustrate construction of a typical aircraft wing and a helicopter rotor blade respectively.

Aluminum alloys are still in the lead of all materials used in airframe. This is due to their good all round properties, namely low cost, low density, relatively high strength/density and modulus/density ratios. The duraluminium type alloys (alloy 2024 and 2014) are largely used for aircraft fittings and brackets and structures such as fuselage and wings. Fittings and brackets are either casting, forging or machined from extruded bars. Fuselage and wing structures are made from sheet metal either by reverting or bonding of skin and stringers. Fuselage frames/bulk heads are made from thick sheets by integrally milling of thick plates. Extrusions are used for longerons and wings and spars and helicopter blade spars. Propellers are made from forging, under carriage composts including wheels are either cast or forged. In some of these structures, fittings and forging, high strength Al - - Zn alloys of the type 7075 was used, but not preferred due to poor resistance to stress corrosion. These are now replaced in new design by 7010 and 7050 type of Al – Zn alloys free from stress corrosion. The lighter aluminum - - lithium alloys offering higher strength to weight ratio have great potential for future application, if the cost can be brought down. Heat and creep resistant Al- -Cu- -Mg- -Fe- - Ni- -Ti alloys are used in the form of forgings for aeroengine compressor blades and discs and as integrally cast compressor rotors and stators for small engines and as sand

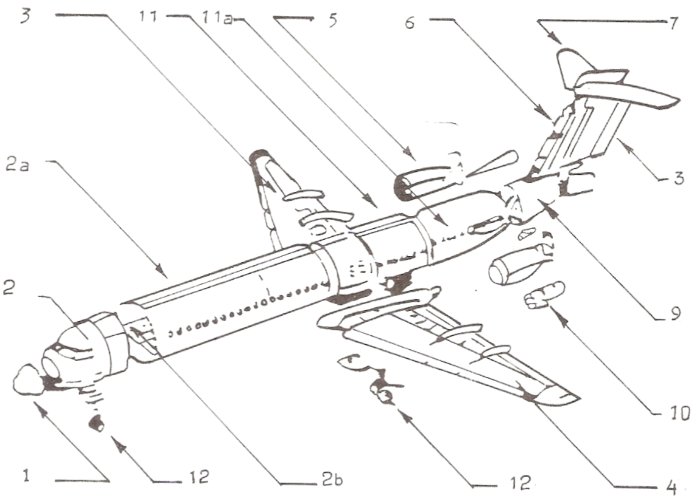


No	Nomenclature	Material
1.	Propellor	—Wood, composite,Forged Al. alloy .
2.	Fuselage	Al. alloy sheet skins, Formed stringers, Extruded components, Machined bulkheads and, Frames of Al. alloys, Riveted/bonded structure.
3.	Wings	Al. alloy sheets/composite skin, Al. alloy extruded soars.
3.(a)	Allerons	Al alloy/composite structure.
3.(b)	Landing Flaps	
3 (c)	Trim Tap	
4.	Empennage	Al. alloy composite structure
4 (a)	Horizontal Stabilizer	
4 (b)	Elevato	
4 (c)	Rudder	
4 (d)	Vertical Stabilizer	
5 (a)	Nose Landing Gear	Al. & Mg alloy forgings/castings/ steel &Ti forgings, rubber tyres
5 (b)	Main Landing Gear	
6.	Wind Shield	Glass/cast & Stretched Acrylic/polycarbonate.

FIG. 3 AIRFRAME STRUCTURES & MATERIALS FOR A TYPICAL LIGHT AIRCRAFT

castings for engine casings. Good corrosion resistance and weldablity of Al- -Mg alloys such as 5054 and 5056 find application in aircraft fuel tanks, fuel and oils lines etc. With the Anglo-French venture of building Concorde with aluminum alloy, the problem of creep appeared. With a speed of Mach 2.2 skin temperature of the order of 150°c reached and it became necessary to use an alloy with good resistance to creep. Creep resistant aluminum alloy sheets such as Hinduminium RR58 (Al- -Cu- -Fe- - Ni- -Ti) originally developed for engine applications found use in the airframe. Research on unidirec-tional reinforcement of aluminium by fine wire and other fibres has been undertaken by some investigators. Such aluminium composites have not yet been used as aircraft materials. However, aluminium alloy-SiC particulate and fibre reinforced composites are emerging as potential materials requiring high rigidity and minimum thermal expansion.

Magnesium is the lightest (Sp. Gr. 1.74) of common engineering metals and is a potential weight saver in airframes. Strength to weight ratio of magnesium alloy castings is superior to aluminium

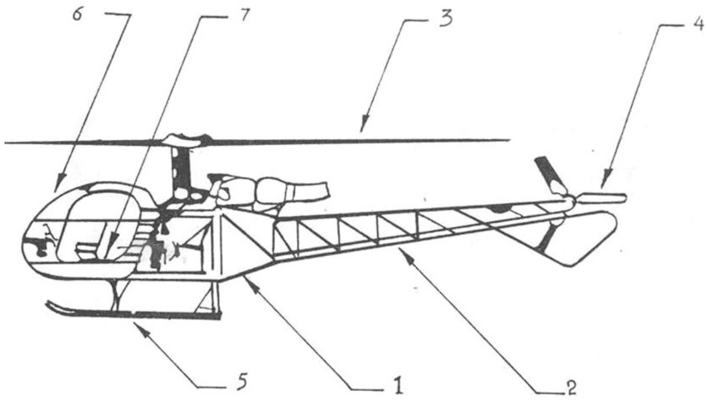


No.	Nomenclature	Material
1.	Nose Cone	— Composite
2.	Nose Fuselage	Al. alloy, Glass/acrylic windscreens
2.(a)	Front Fuselage	Thermoplastic, Composite interiors
2.(b)	Floor Boards	Honeycomb sandwich structure
3.	Wing-Port	Al. alloy/composite
4.	Wing-starboard	
5.	Engine Nacelle	– Al. alloy/Ti & composite
6.	Fin	
7.	Stabilizer	Al. alloy/composite
8.	Rudder	
9.	Rear Fuselage	→ Al. alloy/
10.	Power Plant	– Ti & super alloys
11.		
11 (a)	Centre Fuselage	→ Al. alloy, Ti alloy & steel
12.	Landing Gears	— steel, Ti alloy, Al. alloy & rubber tyres

FIG. 4 AIRFRAME STRUCTURES & MATERIALS FOR A TYPICAL LARGE AIRCRAFT

alloy castings. A variety of Mg—Al, Mg—Zn—Zr, Mg—Zn Rare Earth alloys machined parts, forgings and mainly castings are used for brackets, fittings, gear cases, landing gear wheels, engine compressor housings etc. Magnesium alloys have poor corrosion resistance and are protected by chemical treatments such as anodising and chromating followed by painting.

Titanium alloys have excellent strength and resist oxidation upto 500—550°C and offer the best materials in the temperature range of 200°C—500°C. Considerable weight saving is achieved by using titanium alloys for undercarriages, hydraulic jacks, firewalls, helicopter transmission drives and fasteners. Ti—6Al—4V alloy is the work horse alloy for aircraft applications used for aircraft gas turbine compressor blades and forgings. Titanium alloy forgings are also used for undercarriage compo-nents, fasteners, landing wheels and other airframe structures. Titanium alloy sheets are



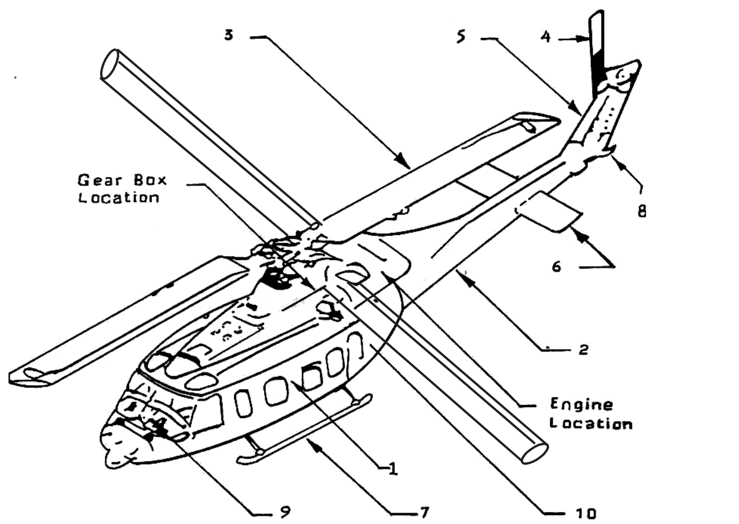
No. Nomenclature	Materials
1. Fuselage	Steel Tubular Structure
2. Tail Boom	
3. Main Rotor	Metal & Composite Combination
4. Tail Rotor	
5. Main Skid	Welded Low Alloy Steel
6. Canopy/Wind Screen	Transparent Plastic
7. Floor Board	Honey comb Sandwich Construction

FIG. 5 STRUCTURES, MATERIALS, SYSTEMS AND EQUIPMENTS FOR LIGHT HELICOPTER

used for fire walls, engine bay doors, aircraft fuselage and wing skins operating at 200–300°C high speed Mach 3 aircraft and welded combustion chamber outer casings, engine inlet ducts, nose cones, engine casings, exhaust shrouds etc. Titanium alloy tubes are used for fuel and hydraulic lines. Titanium alloy honeycomb sandwich construction are used for large fan blades of modern high bypass engines of passenger aircraft.

Steels are used generally for parts which require maximum load carrying capacity in the smallest volume, such as undercarriage components, hydraulic jacks, wing attachment and other structural fittings, fasteners, etc. and also for transmission items such as gears and splines. High strength low alloy steels and special steels such as maraging steels are used for these purposes. Corrosion and heat resistant steels are used for parts such as exhaust collectors, engine casings and pipe bellows, manifolds of fire walls and rear fuselage structures in the proximity of engine. Precipitation hardenable stainless steels are used for valves, gears, splines, springs and other structural members. Low alloy steel tubes are used for welded helicopter body structures, landing struts, skids. A variety of stainless steel tubing is also used for airconditioning and high pressure hydraulic systems.

Copper alloy forged/cast and machined fittings and tubes are used in the electrical and hydraulic system. Super alloy sheets and tubes are used for combustion chambers/flame tubes, reheat and exhaust manifolds, jet pipes and hot gas ducts in aero engine. Turbine blades, discs and nozzle guide vanes are made of superalloy forgings and castings.

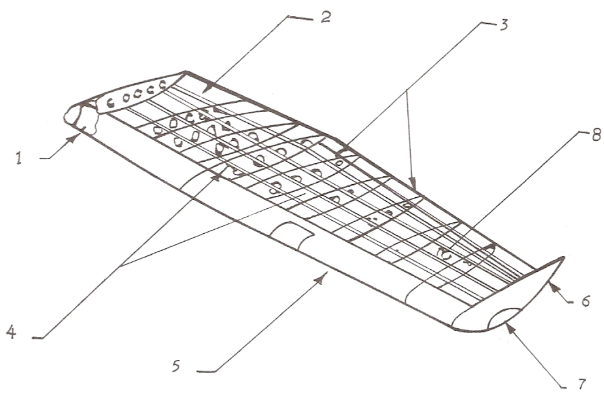


No. Nomenclature	Materials
1. Fuselage 2. Tail Boom	Al alloy Sheet & Stringer construction, riveted/bonded or honeycomb sandwich with CFRP/GFRP/Kevlar/hybrid composite
3. Main Rotor Blade 4. Tail Rotor Blade 5. Vertical Stabiliser 6. Horizontal Stabiliser	Metal & Composite Combination
7. Main Skid* 8. Tail Skid	Welded Low Alloy Steel tube
9. Canopy/Wind Screen	Transparent Plastic
10. Air Intake	Aluminium alloy structure

* In some Helicopters Main Skid is replaced by conventional Landing Gear as in Light Aircraft

FIG. 6 STRUCTURES, MATERIALS, SYSTEMS AND EQUIPMENTS FOR LARGE HELICOPTER

Wood was used in earlier aircraft for light weight structures including fuselage, rudder flaps, propellers. Wood is now-a-days used in the construction of models and gliders. A variety of fabrics was used as fuselage and wing skins in earlier light aircraft. Glass and plastics are used for wind shields, side windows, canopy, etc., for aircraft and helicopters. A variety of structural adhesives is used to make metal to metal, metal to plastics and honeycomb/foam cored bonded sandwich structures. Aluminium alloy and plastic honeycomb sandwich structures with aluminium alloy or FRP skins are used for helicopter and aircraft floor boards. Such sandwich construction is also used for helicopter rotor blades. Glass fibre reinforced plastics are used for making fuselage fairings/ side panels and light weight ducting for aircraft and helicopters. Glass, aramid, carbon reinforced plastics, and their hybrids are used for structural parts including helicopter fuselage structures, rotor



No.	Nomenclature	Material
1.	Front Spar	aluminium alloy extrusion
2.	Rear Spar	
3.	Inter Spar Ribs	aluminium alloy sheets/machined plates
4.	Stringers	aluminium alloy extrusion.
5.	Leading Edge	Aluminium alloy sheet
6.	Wing Tip	
7.	Navigation Light	electric bulb with transparent plastic cover green or red
8.	Closing Panel	aluminium alloy sheet

FIG. 7 CONSTRUCTIONAL FEATURES OF A TYPICAL AIRCRAFT WING

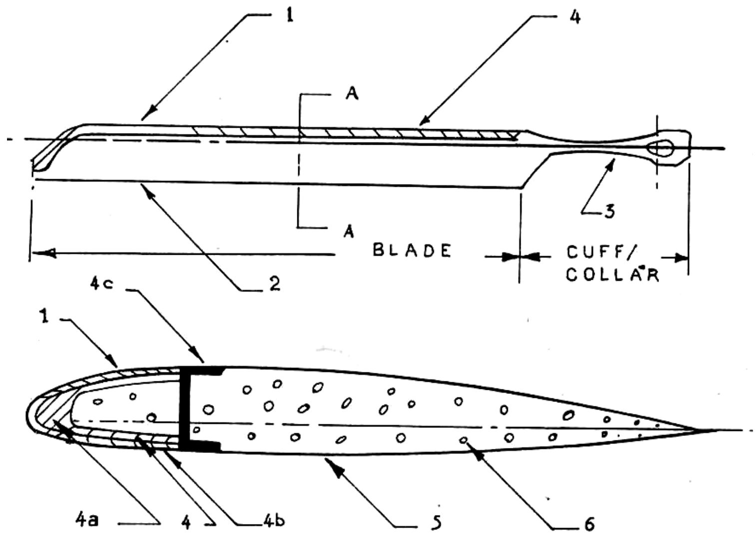
blades, aero-engine fan blades, aircraft wings, rudder, horizontal and vertical stabilizers, trim panels etc.

4. ENGINEERING THE REQUIREMENTS

Now let us look at the aircraft materials requirements Table 1. We know that aircraft materials require high strength to weight ratio, high stiffness, higher fatigue strength, low rate of crack

Table 1. *Aircraft materials requirement*

- High Stiffness
- Higher Strength to Weight Ratio
- Higher fatigue strength
- Lower rate of crack propagation
- Higher fracture toughness
- Environmental tolerance, temperature extremes and Corrosion resistance
- Stealth Characteristics
- Smart/Intelligence characteristics



No. Nomenclature	Materials	
1. Erosion Protection	Stainless steel	
2. Blade	Metal & composite combination	
3. Cuff/Collar	Steel Ti alloy forging or CFRP/GFRP structure	
4. Spar		CFRP/GFRP structure Al-extrusion or taper milled Al-extrusion
a. Core of Spar	Unidirectional CFRP/GFRP structure	
b. Torsion Wrap c. Channel	CFRP/GFRP or Hybrid	
5. Aerofoil Skin 6. Aerofoil Core	Al. alloy sheet or GFRP/CFRP foam core/honeycomb core	

FIG. 8 HELICOPTER MAIN ROTOR BLADE AND ITS SECTION

Table 2. *New & ever growing requirements are met by*

- Advancement in Material
- Engineering of new materials
- New Process Technologies
- Combination of the above

propagation, higher fracture toughness, environmental tolerance and corrosion resistance, stealth characteristics and smart and intelligent characteristics. Table 2 lists some of the routes through which these are achieved.

These can be met by development of new materials, engineering of new materials & process technologies, can combinations of the above. Just to give an example to meet the higher strength to weight ratio for airframe structures you can either increase strength or reduce density or do both and also use innovative design. Now we had in the 1950s aluminium copper alloys & aluminium magnesium alloys for various skins and stringers, and, steel forgings predominantly for fittings to transfer heavy load. We have now higher strength Al alloys and lighter Al & Mg alloys. Similarly there are titanium alloys and titanium forgings which can replace many steels. Also we have a series of high strength low alloy steels, micro alloyed steels and maraging steels all of which have very high strength to density ratio. Then we have a new series of materials coming up viz., intermetallics. We can talk about aluminium intermetallics, titanium aluminides and so on. Then we have got foils and foils technology.

On the structure side, earlier it was all riveted structures and later on we moved onto integrally machines structures reducing the number of rivets used. Then came the era of adhensive bonding and sandwich structures. This was possible because of the development of a low-density core materials, called honeycomb, made out of thin foils of both metallic and non-metallic materials. It was an engineering ingenuity and is finding wide applications in aircraft primary and secondary structures offering quite an amount of weight savings. Now we are onto Super Plastically Formed and Diffusion Bonded structures, composites and hybrid structures. Stealth is achieved by use of materials such as composites with lower radar cross section, external coatings with radar absorbant materials and by design, for example by facetted structures.

So on one side there is development in materials and on the other side there are developments in structural design and how these materials are effectively engineered into structures. Thus in effect there are 3 ways; first one is development in materials, the second one is engineering of materials and the third one is engineering of structures.

5. TRENDS IN AIRCRAFT MATERIALS

Trends in usage of airframe materials over the years is given in Figure 9. There is increasing use of Titanium alloys alloys composites. A comparison of materials used in Jaguar of 1980's and LCA for the 2000 and beyond is shown in Figure 10 and a similar comparison for our older generation and new generation advanced helicopter is shown in Figure 11. In the Jaguar it is seen that nearly 55% of the structural weight is made of aluminium alloys whereas in LCA it is around 40%.

The use of composite materials in Jaguar is around 7% whereas it has moved up to around 33% in the LCA. In the case of Advanced Light Helicopter almost 60% of its wetted area is made out of composites. In terms of weight, we can see, the Chetak Helicopter, which is of the older generation, has utilized around 5% of composites, about 40% of aluminium and 2% of titanium alloys. Whereas in ALH the usage by weight is 35% of aluminium alloys, 30% of composites and about 8% of titanium alloys.

6. SMART MATERIALS AND STRUCTURES

We made a brief mention about intelligent materials and their applications earlier. We have a variety of intelligent materials developed and under development which could be used in future (Table 3).

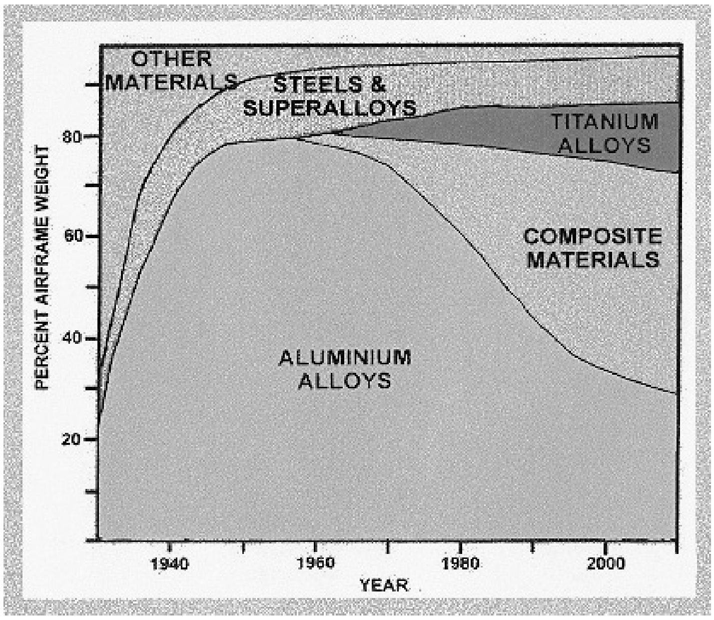


FIG. 9 TRENDS IN AIRCRAFT MATERIALS

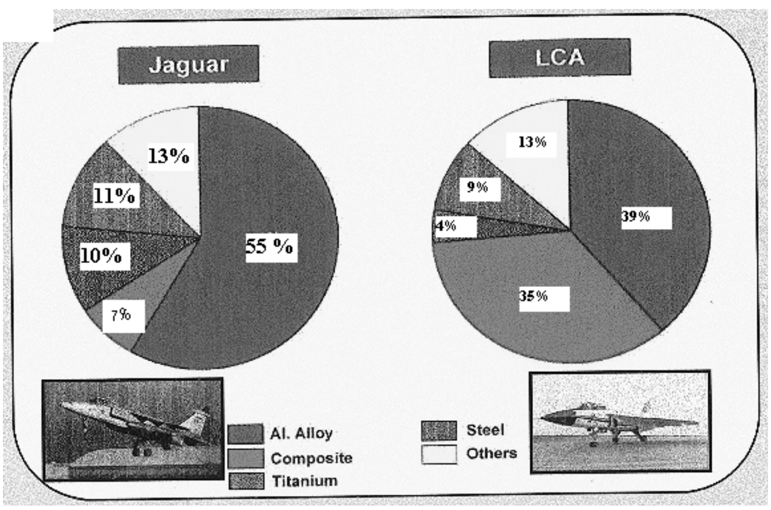


FIG. 10 MATERIALS USED IN JAGUAR AND LCA

To quote a few examples; Embedded sensors to identify and inform online computers for corrective and safety measures, which can be used on control surfaces of the aircraft; Piezo ceramics which dampen acoustically excited vibrations which will be useful for aircraft structures; Shape memory

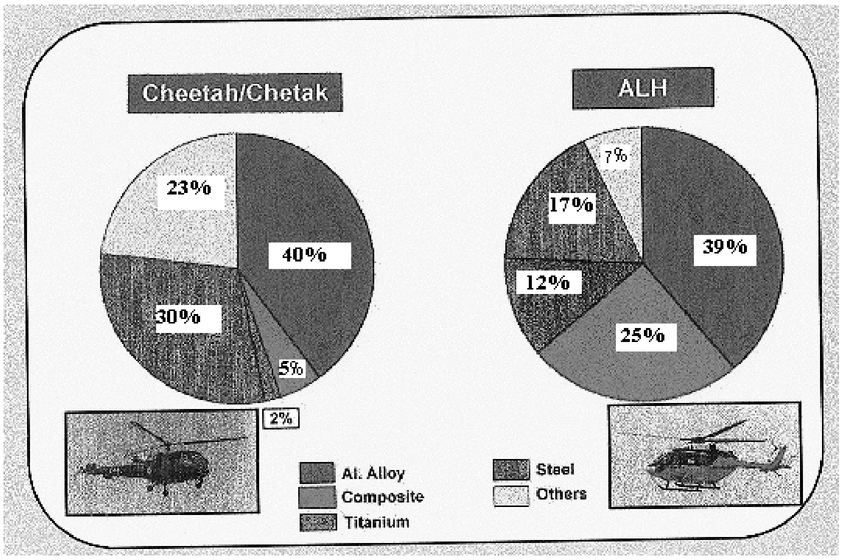


FIG. 11 MATERIALS USED IN CHEETAH/CHETAK AND ALH

Table 3. *Intelligent Materials*

Embedded Sensors	Identifies & informs on-line computer for corrective/safety measures
Piezoceramics	Dampens acoustically excited vibrations
Shape Memory Alloys	Acts as thermal sensors and controllers
Electro-rheological fluids	When electrically excited, provides extra stiffness in specific directions to assist in damping of sudden loads
Optical Fibres	Acts as nerves; Sensing heat, stress & damage

alloys which can change shapes used in aerospace quite a lot now-a-days and which can act also as thermal sensors and controllers; Electro rheological fluids, on being electrically excited providing extra stiffness in specific directions to assist in damping of sudden loads; Optical fibers which act as nerves, sensing heat, stress and damage. Optical fibres are also being used for the active controls of fly-by-light Aircraft. Figure 12 a, b, c, d show the principles of building Intelligent Structures by the confluence of four different disciplines.

Figure 13 shows the applications of Smart Materials in different roles in different zones of a future transport aircraft.

7. CONCLUSION

Development of advanced high strength light weight structural materials, variety of engineered composite materials and structures, shape memory and intelligent materials will play a major role in

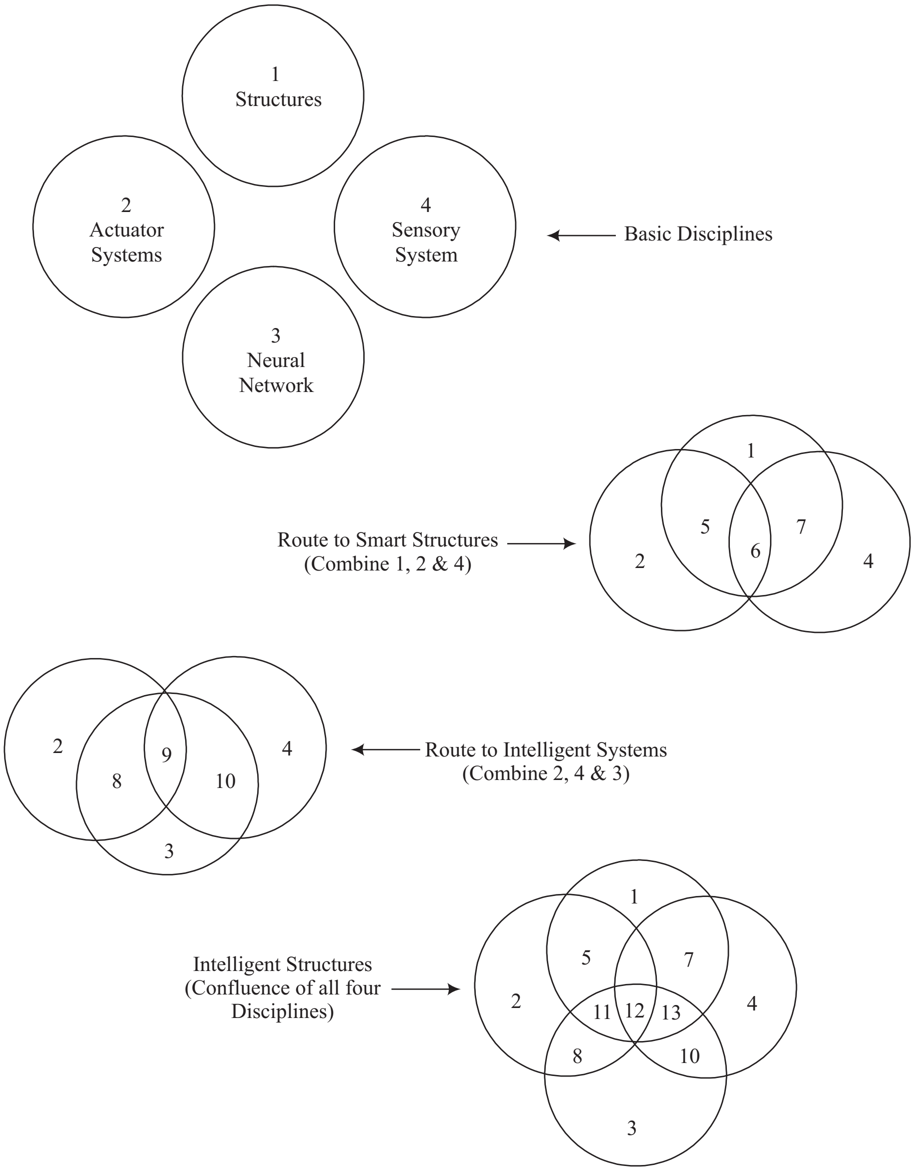


FIG. 12 PRINCIPLES OF BUILDING INTELLIGENT MATERIALS/STRUCTURES 5. CONTROLLED STRUCTURES, 6. SMART STRUCTURES, 7. SENSORY STRUCTURES, 8. LEARNED RESPONSE SYSTEMS, 9. INTELLIGENT SYSTEMS, 10. EXPERIENCED BASED SYSTEMS, 11. LEARNED REACTIVE STRUCTURES, 12. ADAPTIVE LEARNING SENSORY STRUCTURES, 13. INTELLIGENT STRUCTURES

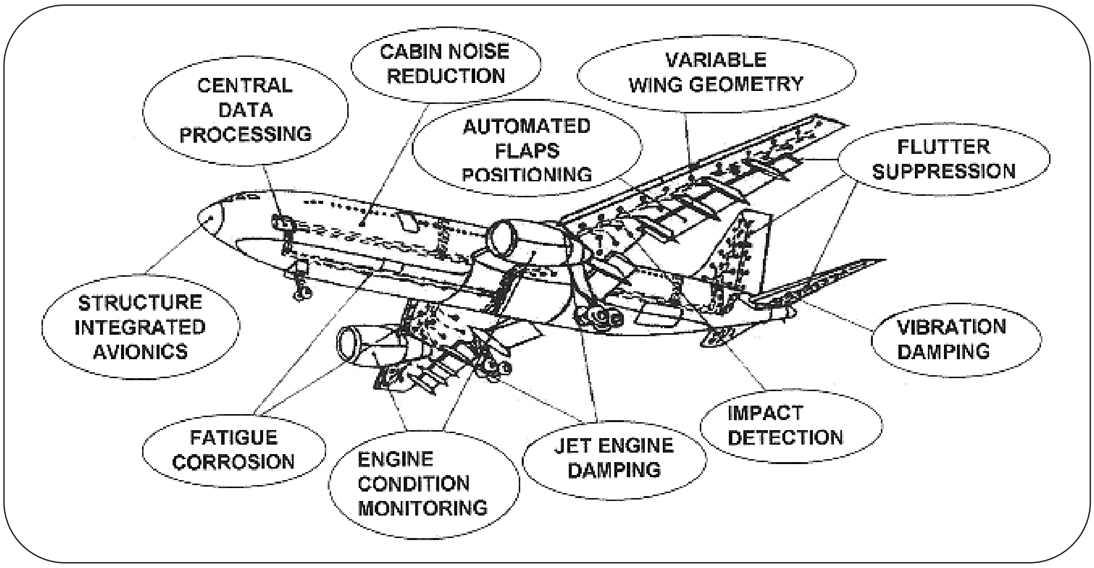


FIG. 13 SMART STRUCTURES & MATERIALS IN AIRCRAFT

the development of tommorow's aircrafts, the Jumbos, the micro and minis and unmanned intelligent air vehicles. Man will continue to observe and learn from nature and mimic with his own creative intelligence to soar higher in aerospace technology and applications.

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

INDIAN AIRCRAFT INDUSTRY (NETWORKING OF INDUSTRY, R&D AND ACADEMIA)

C. G. KRISHNADAS NAIR*

1. INTRODUCTION

Aviation has come a long way from the early dreams and adventures. It is today one of the most important technological influences with major role in national security, domestic and international transport, business, industry, tourism and economic growth. Hindustan Aeronautics Limited, which is synonymous with the Indian Aircraft Industry, was founded in December 1940 by a visionary industrialist Lt. Shri Walchand Hirachand. Over the years it grew into the largest integrated aeronautical complex in South Asia manufacturing not only aircraft but also engines, systems, accessories and equipment and varieties of sophisticated components and structures. The industry in fact took the initiative right from the beginning in aeronautical education and research and development. It was Dr. V.M. Ghatge, HAL's Chief Designer who set up the Department of Aeronautics, the first of its kind in India at the Indian Institute of Science in Bangalore. It was under his leadership that India designed and developed its first glider in 1941 and the first Trainer HT-32 in 1951. HAL has designed and developed a number of light aircrafts, trainers, fighters and helicopters. It has also grown by collaborating with overseas industries and manufacturing transport and fighter aircrafts and helicopters, engines, equipment, systems and accessories under license and thereby absorbing technology from others.

In recent times, there has been systematic growth in research and development both inhouse and in collaboration with DRDO, National Laboratories and academia enhancing self-reliance. There have been considerable efforts to involve private sector for indigenous development and manufacture of materials, equipment and systems. With the liberalization of economic policies and the need to be more competitive, the industry has redefined its mission:

"To become a globally competitive aerospace industry while working as an instrument for achieving self-reliance in design, manufacture and maintenance of aerospace defence equipment and diversifying to related areas, managing the business on commercial lines in a climate of growing professional competence."

---

*Former Chairman, Hindustan Aeronautics Limited (HAL), President – Society of Indian Aerospace Technologies & Industries (SIATI), Aeronautical Society of India Building, Suranjandas Road, Off. Old Madras Road, Bangalore – 560075

New R&D Centres have been created in order to strengthen existing ones. These R&D centres are co-located with production divisions for better interaction, close coordination, concurrent engineering and achieving greater and speedier performance. With the reorganization of the R&D centres co-located with manufacturing divisions and the setting up of joint venture companies, new divisions & business development units, the present HAL group is as follows:

Today HAL has about 30,000 highly skilled employees including about 2000 Design Engineers and Technology Managers. HAL puts in 20% of its profit or 2% of its sales turn over which ever is higher, to the R & D reserve each year. This together with customer funding makes HAL, one of the top industries in terms of investment in R & D, amounting to almost 8 – 10% of its annual budgeted expenditure of R & D. There has been substantial improvement upgradation of technology and equipment both for design and manufacture by conscious efforts and it has led to substantial improvement in productivity and profitability.

THE HAL GROUP

Divisions	*R&D Centres*	*New Divisions Business Development Groups*
• Aircraft Division Bangalore	• Aircraft R & D Centre, Bangalore	• Aerospace Division Bangalore
• Aircraft Overhaul Division Bangalore & Barrackpore Branch	• Aircraft Upgrade R & D Centre, Nasik	• Industrial and Marine Gas Turbine Division, Bangalore
• Aircraft Division Nasik	• Transport Aircraft R & D Centre, Kanpur	• Airport Services Business Group
• Transport Aircraft Divison, Kanpur	• Rotary Wing R&D Centre, Bangalore	• Simulator Business Group
• Helicopter Division Bangalore	• Engine Test Bed R & D Centre, Bangalore	*JVCs*
• Engine Division Bangalore	• Gas Turbine R & D Centre, Koraput	• Indo Russian Aviation Limited
• Engine Division Koraput	• Aerospace Systems & Equipment R & D Centre, Lucknow	• BAe-HAL Software Private Limited
• Accessories Division, Lucknow	• Strategic Electronic R&D Centre, Hyderabad	
• Avionics Division Hyderabad	• Central Materials & Processes Lab & NDT Centre, Bangalore	
• Avionics Division Korwa		
• Foundry & Forge Division, Bangalore		

The technology base of HAL is also used to make a significant contribution to other National projects. A variety of complex castings, forgings and machined components and systems are supplied to naval frigates, earthmovers and armoured vehicles. HAL's association with ISRO dates back to the Seventies when the structures for first Indian satellite ***Aryabhatta*** was manufactured at HAL. This was followed by structures of ***Bhaskara, Rohini, Apple*** and Stretched ***Rohini Satellite*** series and the first operational satellite ***Indian Remote Sensing Satellite.*** Likewise the light alloy structures for SLV and ASLV were manufactured in HAL. A dedicated facility is set up for the manufacture of Light alloy structures and tankages for PSLV, GSLV and Communication Satellites. The production and supply of Light Alloy structures for PSLV, GSLV and INSATS have greatly enhanced self-reliance in the field and made HAL emerge as a major industry partner for ISRO.

Networking of HAL's management executives and designers with scientists from DRDO and other National Laboratories and Academic Institutions have made some of the recent Design initiatives highly successful. I would like to highlight a few of these, which will have a bearing on the future growth of the industry and the integrated growth of R & D, Education and Industry in the country.

2. ADVANCED LIGHT HELICOPTER (ALH)

ALH is a state-of-the-art helicopter incorporating several advanced technologies such as Carbon-Kevlar composite structures, Hingeless composite Main Rotor with an elastomeric bearing, Bearingless composite Tail Rotor, Integrated dynamic transmission systems, Full authority digital engine control and Crash worthy airframe. The helicopter is developed as a multirole helicopter for military as well as civil applications in the 4-5 tonnes category. In addition to various component testing and structural testing and full scale fatigue testing, it underwent 1,200 hrs. of flight-testing to certify the helicopter for airworthiness. The helicopter in multirole capability is now in production and is named **'DHRUV'** by Hon'ble Minister for Defence. The ALH military version compares very well in performance with other armed helicopters. The ALH civil variant with its wide body, spacious and quite cabin, hinged doors, emergency exits, smooth and vibration free flights make it ideal for passenger transport and have a seating capacity of 12 – 14 passengers. It compares extremely well in performance and cost with world's top class passenger helicopter. The Air ambulance-ALH Variant is virtually a Flying hospital. It is configured with 2 lightweight litters and for medical attendants or 4 litters with 2 medical attendants and it is equipped with First Aid, on board oxygen and emergency treatment kits, patient monitoring data link equipment and with day & night all weather navigational aids.

3. LANCER

Lancer is a light attack helicopter developed by Hindustan Aeronautics Limited as an effective airmobile area weapon system. Basic structure of the Lancer is derived from reliable and proven Cheetah Helicopter. The Lancer is optimized for

- Anti Insurgency Operations
- Close Air support
- Intimidation and Harassment of enemy troops and insurgents
- Anti Armour Role

Lancer has two jettisonable combination gun-cum-rocket pods on either side. Each pod carries one 12.7 mm gun and three 70mm rockets. A gun sight is provided for accurate aiming and firing. HAL produced and supplied twelve Lancers to Indian Army and there is great potential for export and domestic market for police and paramilitary forces.

4. LIGHT OBSERVATION HELICOPTER (LOH)

HAL has taken up conceptual design studies and development of key technologies for a "Light Observation Helicopter", which will replace existing fleet of Cheetah, and Chetak helicopter. Major roles will be reccee and surveillance, border patrol, commuter, rescue, and other – missions at altitudes in the range of six kilometers. It will be a 3 tonne class helicopter with 1500 kg useful load. LOH is being designed with ALH technologies, and a ducted fan bail rotor and maximum commonality with ALH systems thus reducing the development lead time and cost. India will have the world record for having designed and built the advanced technology helicopter for operations in the highest altitude – a remarkable world record.

5. LIGHT ATTACK HELICOPTER (LAH)

HAL has plans to take up development of Light Attack Helicopter as a follow on project for ALH. The LAH will have improved features over ALH in the areas of weight, lower visual signature and reduced Radar cross-section, and a new fuselage structure with armour.

6. LIGHT COMBAT AIRCRAFT (LCA)

The successful maiden flight of the Light Combat Aircraft (LCA), (India's most prestigious aircraft design project) on 4th January 2001, was a high point in HAL's Mission realization. LCA is a versatile niche product for airborne defense and tactical air superiority, supersonic speed at all altitudes, and with superior agility. HAL and Aeronautical Development Agency are working on the new generation state-of-the-art aircraft for the 21st century. LCA integrates modern design concepts such as Tail-less Compound Delta Platform, with Relaxed Static Stability and Fly-by-Wire Flight Control System. Advanced Digital Cockpit, Multi Mode Radar, Integrated Avionics System, Advanced Composite Material for structures and a Flat Rated Engine, short take off and landing, high manoeuvrability with excellent maintainability and a wide range of weapon fit are some of its salient features.

HAL has been the main partner with DRDO for design and manufacture of airframe, systems/ equipments as well as system integration. Chairman, HAL was co-chairman of the technical committee headed by the Scientific Advisor to the Hon'ble Minister for Defence. LCA was not just an aircraft design project, it was a total R&D project to develop various technologies, equipments, and systems and build up greater self-reliance in design and development of advanced military aircraft, involving in addition to HAL, and DRDO a large number of research labs, private industries and academic institutions thus building an integrated National capability. The second technology demonstrator flew on 6th June 2002. It was again an incredible sight seeing the two LCA prototypes flying during the Aero India international exhibition and flying displays 2003. Prime Minister of India christened LCA as TEJUS on 4th May 2003. HAL has been given clearance for setting up facilities for limited series production and the aircraft is scheduled to be inducted into service from 2008-09 onwards.

7. INTERMEDIATE JET TRAINER (IJT)

Considering the need for a modern Intermediate Jet Trainer, HAL took up the design and development of HJT-36, this aircraft will replace the famed Kiran Aircraft currently used for Stage – II Training. The aircraft made its maiden flight on 7th March 2003 in a record time of 42 months from the start of the design. HJT-36 compares favorably in price and performance with similar trainer aircrafts. The HJT-36 is superior with respect to cost and performance when compared with some class of trainers. The development cost and time were reduced by networking with Indian and overseas equipment suppliers as partners in the project, and using the CFD and other techniques, and various equipments already developed for the LCA Project.

8. SARAS LIGHT AIRCRAFT

The design & development of the 14 seater multi purpose Light Transport Aircraft - SARAS is an initiative by the National Aerospace Laboratories. It is designed to have a maximum cruise speed of 625 km/hr and a maximum range of 2800 kms. HAL has networked with NAL for design and development of wings and landing gear and provide all other support as required. A number of vendors already developed by HAL for component and equipments for other projects are utilized by NAL for the project.

9. MULTIROLE TRANSPORT AIRCRAFT (MTA)

HAL has signed an agreement with Ilyushin Aviation Complex and Irkutsk Aviation Production Organisation of Russia for joint design and development and co-production of a Tactical Transport Aircraft of 15 to 20 tonne capacity / 100 seater to meet the services as well as the civil airlines requirements. The aircraft will have a range of 2500 to 6000 kilometers depending upon the load to be carried with the cruise speed of around 850 kilometers per hour. This aircraft is planned to be developed with the involvement of all the major aerospace labs and academic institutions (of the country) who were networked and participated for the LCA and ALH projects. Also participation of private sector for development and production is planned.

10. OTHER PROJECTS

In addition to the above HAL has launched a number of aircrafts' life extension and upgrade projects and development of remotely piloted vehicles, helicopters and aircraft engines, various aircraft systems, accessories and equipments in collaboration with other agencies.

11. HAL TODAY

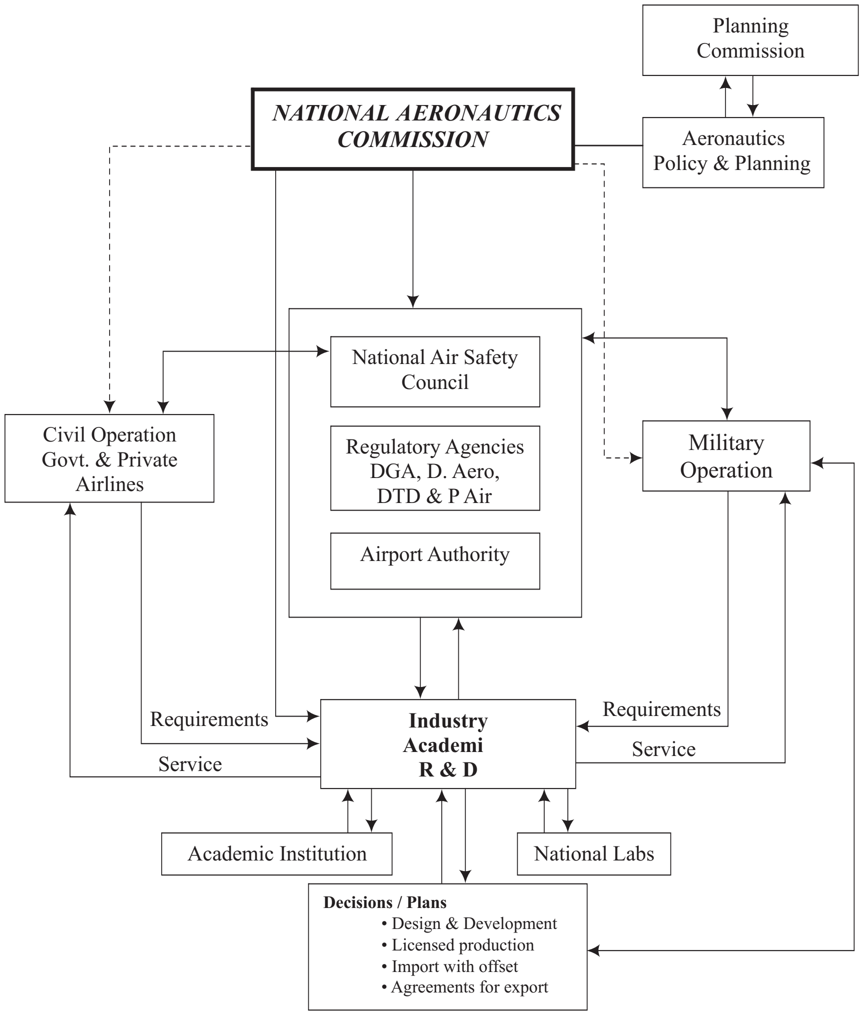
With greater emphasis on indigenous design in collaboration with R & D organisations in the country and with global co-operation both in co-development and co-production, diversification, HAL has emerged as a globally competitive aerospace industry. Its financial performance is as good as its technological excellence. Various financial parameters such as profit to sales, profit to capital employed, sales to net assets, earnings per share etc., compared well with the leading aerospace industries of the world. Among the Indian Public Sector Companies, HAL had the distinction of getting award for excellent performance in successive years and was judged as the best managed company in 2001. HAL has drawn up an impressive long term plan to increase the sales turnover to about Rs.100,000 million by the year 2007 and achieve an export sales of Rs.5,000 million and achieve about 25 – 30% of its total sales to the civil customers.

12. THE FUTURE

From the aircraft of Wright Brothers built with mostly fabric, tubes and wires to the modern Jumbo Jet and the advanced supersonic fighter aircraft, there has been tremendous development in technology during this century. What could be the shape of things to come in future? It is not an easy task to look into the crystal ball and predict future, but we know that technology will play a major role. The emphasis is on larger transport aircraft, and higher speed and maneuverability for fighter aircraft, Research & Development will be needed to meet the challenges and opportunities, especially in fields such as newer lightweight and strong materials including metal and ceramic components, new aerodynamic configurations such as blended body, mission adaptive wings and aircraft skin, fly-by-wire to active controls, micro and mini aircrafts, electronic counter measure, advanced data link, greater noise and vibration control intelligent materials, smart and intelligent structures and information technology. We need greater interaction between industries, R & D laboratories, civil and military aviation, government bodies, academia and collective action from the aeronautics community to continue our march towards greater excellence. In future R & D investment will be so high that success can be achieved only through partnership, by sharing expertise as well as resources and risks/gains. Growth through National and International collaborations in research, design and manufacture and marketing should be our goal.

Looking at the near future, the world aerospace market for the next 10 years is valued at approximately US $850 billion comprising of military and commercial aircraft. The largest demand will be in the commercial Jet Transport estimated at about 550 billion inclusive of regional commuters. The demand for military fighters and military transport is valued about US $ 150 billion and the requirement for civil and military helicopter is estimated to be about 70 billion. India's defence budget has been recently raised and it is about US $ 10 billion annually. The large and emerging demand for military aircraft in India are being met by indigenous design and developments as well as license production and co-production. As per survey conducted by different agencies, the demand for civil transport aircraft will be of the order of US $20 billion in the next 10 years. Airport infrastructure including air traffic and control equipments is estimated to be around US $ 4 billion during this period. Spares requirement for maintenance of both military and civil aircrafts is estimated to be around US $ 100 Million annually.

While we should be open to share our market, we should also plan and utilise our legitimate right in sharing the world market through mutually rewarding co-operations. A large industrial infrastructure in the private sector consisting of small, medium and large scale organizations is required in addition to HAL, other defence PSUs, DRDO and other R&D labs to do this. SIATI (Society of Indian Aerospace Technologies & Industries) was an initiative in this direction.



ORGANISATIONAL STRUCTURE FOR MANAGEMENT OF SYNERGETIC GROWTH OF AERONAUTICS IN INDIA - *VISION*

Founded in 1991, it has now over 330 small, medium and large scale private industries, in addition to major aerospace players who are engaged in design, analysis, software and software solutions, manufacturers of precision machined parts, sheet metal components, composites and composites parts, various types of rubber and plastic components, batteries, connectors, PCB's and embedded systems, electrical and electronics equipments and materials and consumables. The Society of Indian Aerospace Technologies & Industries have made pioneering efforts in bringing industry R&D collaborations within India and with overseas organisations and enhance self-reliance in aerospace technology and manufacturing.

SIATI, an important 'Gateway' to Indian Aerospace business and Global co-operation for tie-up in R&D, technology, joint ventures, collaborations, co-productions and information-exchange, is a professionally managed non-profit, non-commercial organisation.

It is also required to have a National Aeronautics Policy and an Apex body to integrate and stimulate aeronautical activities in the country. I had proposed this idea at the annual conference of Aeronautical Society of India at Mumbai in 1992, when I delivered the *Dr. Ghatge Memorial Lecture*. Subsequently a team consisting of Prof. A K Rao - Indian Institute of Science, Mr. K Srinivasa – Director of Aeronautics, Dr. Kota Harinarayana- Aeronautical Development Agency, Dr. B R Somashekar – National Aerospace Laboratory, Capt. K M Balasubramaniam – Air India, Dr. G K Agarwal – Indian Airlines, Mr. K N Ardhanareeswaran – National Airport Authority, Dr. K V L Rao – Aeronautical Development Agency, and myself Dr. C G Krishnadas Nair prepared a draft Aeronautical policy under the guidance of Prof. APJ Abdul Kalam, then President of Aeronautical Society of India and submitted to PM of India. More recently the present and a few past president of ASI have taken up the case again. Proposed organizational structure for management of synergetic growth of aeronautics in India is given below.

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

DEVELOPMENT AND APPLICATIONS OF CFD TECHNIQUES IN AERO-PROPULSIVE CHARACTERIZATION OF MISSILES

DEBASIS CHAKRABORTY*

ABSTRACT

CFD applications for complex aerodynamic and propulsive flow problems pertaining to DRDO missiles are presented. Number of grid generators, 3D Euler and Navier Stokes solvers are developed indigenously using state of art numerical techniques and physical models. These indigenously developed softwares are used extensively for aerodynamic characterization of missiles over a wide range of Mach number, angle of attack and control surface deflections etc. leading to significant reduction of wind tunnel testing. Six degrees of freedom trajectory equations are integrated with the flow solver for separation dynamics studies. Significant contributions are made in the design of high speed propulsion systems of various ongoing and future missiles through numerical analysis of reacting and nonreacting flow field inside the propulsion system. Key performance parameters were predicted and propulsion systems were optimized through the analysis of various thermochemical parameters obtained from numerous numerical simulations. Understanding of complex flow phenomena inside the propulsion system has reduced the developmental cost and time of the system significantly. Capabilities are being developed for many advanced topics including computational aeroelasticity, coupled Euler –Boltzmann solver, conjugate heat transfer studies etc.

1. INTRODUCTION

The current maturity of CFD techniques, the general availability of powerful computers and the complexity of the problems make the use of CFD an integral part of aerodynamic and propulsion design of missile system. Presently, CFD methods are employed routinely for the estimation of various complex aerodynamic and propulsion flow parameters where experimental data cannot be obtained economically or feasibly. CFD has emerged as one of the important design tool alongwith the wind tunnels and other experimental testing and contributing significantly in reducing developmental cost and time of the missile system.

---

*Defence Research and Development Laboratory, Hyderabad - 500058, INDIA
E-mail: debasis_cfd@drdl.drdo.in

DRDL has developed a host of indigenous grid generators, surface modeler and industry standard three dimensional Euler and Navier Stokes solver using advanced numerical techniques and physical models for accurate prediction of complex aerodynamic flow phenomena. Systematic validations were carried out through comparisons against reliable experimental results before applying in the design exercises. These softwares are continuously upgraded through inclusion of new numerical techniques and physical modeling to increase the accuracy. Core competencies are being developed in numerical simulation of turbulent reacting flows. Important contributions are made in the design and analysis of propulsion systems of various ongoing and future missile projects using commercial softwares. These softwares were validated extensively against experimental results to find its error band and range of applications. Careful examination of various thermochemical parameters obtained from numerical simulations has enabled the making of important design changes leading to improvement of performance parameters. Understanding of complex flow phenomena inside the propulsion system has reduced the developmental cost and time of the system significantly. The role of CFD in the missile design and analysis in DRDL is presented in this paper. The code development activities and few examples of aerodynamic and propulsive characterization of missile systems are highlighted.

2. CFD CODE DEVELOPMENT ACTIVITIES

An integrated CFD code consists of three major parts, namely grid generator, flow solver and post processor. A full 3-D elliptic grid generator AUTOELGRID [1] is indigenously developed to generate body fitted structured volume grid for DRDO missiles. The grid generation is based on Sorensen - Steger procedure for controlling grid line orientation and spacing. The grid generator is highly automated and is being used routinely for generating high quality grids around the missile configuration.

A 3-D upwind finite volume Euler Solver KFSG based on Deshpande’s Kinetic Flux Splitting (KFVS) [2] has been developed with characteristics based boundary conditions such as Kinetics Characteristics Boundary Condition (KCBC) and Kinetic Outer Boundary Condition (KOBC) applied in the computational boundaries. This software has been validated for test cases ranging for subsonic to hypersonic speeds [3] and is used extensively for aerodynamics characterization of number of DRDO missiles and projectiles. The surface flow parameters obtained from the code is used as the input for aerodynamic heating analysis. This solver has been parallelized using both MPI and ANUPAM library. For aeroelastic analysis, KFVS formulation has been extended for moving Grid (KFMG)[4] and an unsteady Euler Solver has been developed and integrated with 2 DOF structural dynamics model. The developed aeroelastic code has been applied to study pitch and plunge motion [5] and ‘Transonic Dip’ phenomena [6] of NACA 64A006 aerofoil. A three dimensional unsteady Euler Solver based on KFMG has been developed for structured grid using Spring analogy and Transfinite Interpolation method for grid movement due to deflection and deformation of the oscillating body. The developed code has been validated for ogive-cylinder-ogive configuration in transonic flow [7] and ONERA M6 wing configuration [8]. The development of 3D unsteady Euler Solver in unstructured grid is under progress.

Grid generation around complex flight vehicle is a difficult and highly time consuming task. Recently Grid free methods have emerged as efficient tools for solving many aerospace problems.

These methods do not require a grid to solve the partial differential equations but operates on cloud of points distributed over the domain. Least square Kinetic Upwind Method (LSKUM) is popular among grid free methods [9]. The extension of LSKUM to second order accuracy using entropy variable (q) known a q-LSKUM [10] is proved to be more efficient. A 3-D grid free Euler code has been developed at DRDL using q-LSKUM to simulate flow past flight vehicles with control surface deflection. A modified LU-SGS method has been implemented [11] to get good speed-up without any extra memory and the implicit code is parallelized [12] using MPI to reduce the turn around time. The complex flight vehicle is decomposed into simple component and simple grid generation methods are used to generate cloud of points around each component and points are overlapped to generate distribution of point in the computational domain [13]. Efficient blanking and search algorithm have been developed to generate connections from each point. Recently 6-DOF trajectory equation solver has been coupled with q-LSKUM code and validated for store separation problem [14]. The parallel q-LSKUM is routinely used in DRDL to solve many complex aerodynamic problems of DRDO missiles. A grid free viscous flow has also been developed and the validation of the code is currently in progress.

DRDL has developed a general purpose 3D RANS solver CERANS [15,16] to solve the flow equations in cell centered finite volume framework on sequential and parallel computers. The flow gradients can be evaluated by cell-centered weighted least squares method or the diamond path reconstruction approach. Higher order accuracy can be obtained using the method of reconstruction or Polyhedral MUSCL approach. The convective fluxes are modeled using several state-of-the-art numerical flux formulae including the KFVS & van Leer flux vector splitting schemes, AUSM family, Roe family, Modified Steger-Warming and HLL family of schemes. The turbulence closure is effectuated by the one-equation Spalart-Allmaras turbulence model and SST turbulence model. Implicit time integration is carried out using either the Point Jacobi method or Lower Upper Symmetric Gauss Siedel Method. A robust reliable blended universal law of the wall with compressibility and heat transfer correction has been used for modeling the near wall flow. Incorporation of real gas effects and transition model are under development. This code is routinely used at DRDL for simulation of complex high speed flows [17,18].

During the reentry phase, three principle types of flows; namely, free molecular, transitional and continuum are encountered. For transitional and free molecular flows, Boltzmann equations are applicable and are generally solved using Direct Simulation Monte Carlo (DSMC) Method [19]. With the decrease in altitude when mean free path and mean time between collisions decrease, these methods demands excessive computing time and memory. In recent time, there is a special interest in combining continuum and non-continuum solver [20] to reduce the simulation cost without sacrificing the accuracy of the results. In this approach, continuum and non-continuum regions are identified and respective solvers are applied to get the flow field. Euler equations are solved using a particle based method [21] to circumvent the problem of interface between the two regions. The deviation of local velocity distribution function from the Maxwellian distribution is considered as a measurement of equilibrium and is used for the identification of continuum and non-continuum region. A 2D and 3D Euler-Boltzmann coupled solver has been developed and validated for standard test cases.

3. APPLICATION OF CFD METHODS TO EXTERNAL FLOWS

The indigenously developed CFD codes are routinely used in DRDL to predict various aerodynamic parameters pertaining to DRDO missile systems. Aerodynamic characterization of missiles in complete M-α flight regime, control surface deflection studies, heat shield separation of hypersonic air breathing missiles, store separation from the aircraft, study of plume impingement on jet deflector, low speed (incompressible limit) characterization of vehicles etc., are some of the important applications of the codes. Few important results are presented in the next subsections.

3.1 Aerodynamic Characterization of Missiles with Control Surface Deflection

To increase the performance of a surface-to-surface missile, aerodynamic surfaces (wings and fins), were redesigned significantly. Typical configuration of the missile is shown in Figure 1. The inhouse developed grid free method (q-LSKUM) was used to determine the aerodynamic characteristics of the vehicle including the effect of the fin deflection. The grid generation in the defected fin condition is critical because of the presence of gaps between the corebody and fins. A separate block needs to be created in the gap region and new gird to be generated for the new defected position. This problem has been overcome in the grid free methods. Structured grids around each component such as core-body, wings, fins, etc, are generated independently and the grids are overlapped to get the distribution of points within the domain. The redistribution of points required for fin deflection is accomplished by appropriately rotating the grid blocks around the hinge line. Zoomed view of the gird in the wing and fin region is shown in Figure 2. The inviscid flow field around the configuration is computed for critical flow conditions (different Mach number, angle of attack and fin defection) obtained from the trajectory. Figure 3 presents the Mach

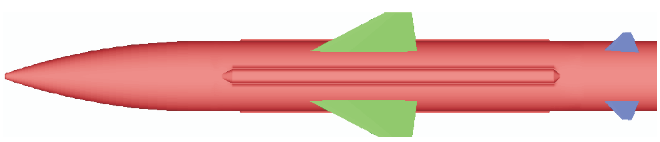


FIG. 1 TYPICAL CONFIGURATION OF THE FLIGHT VEHICLE

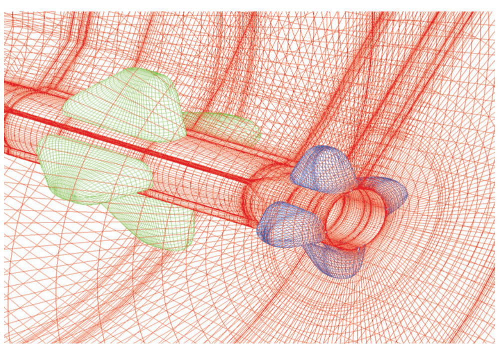


FIG. 2 ZOOMED VIEW OF THE GRID NEAR WING AND FIN REGION

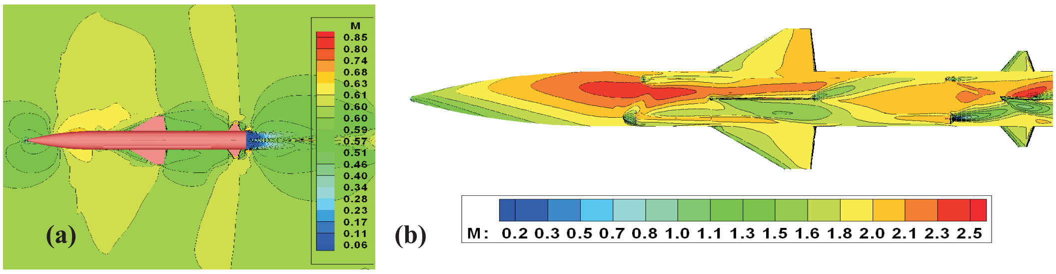


FIG. 3 FLOW PARAMETERS FOR M = 2, A = 14 AND WITH DIFFERENT FIN DEFLECTIONS (a) MACH CONTOUR IN PITCH PLANE (b) SURFACE MACH NUMBER

contour in the pitch plane and surface Mach number distribution for M = 2 and $\alpha = 14°$ with 5° fin deflection. The computed aerodynamic forces and moments were used in the mission analysis. Later on, limited wind tunnel tests were carried out for the configuration. The comparisons of the computed and measured aerodynamic forces and moments presented in Table 1 for few typical Mach numbers and angle of attacks show a good match between the two. Further details of the work are available in Ref. 22.

3.2 Nose Petal Separation Studies of Hypersonic Airbreathing Vehicle

To demonstrate the autonomous functioning of a scramjet engine, an airframe integrated scramjet engine (cruise vehicle) carried to the operating attitude of 32.5 km at Mach 6.5 with the help of a launch vehicle. The cruise vehicle is put above the first stage of a rocket and is covered with a heat shield to protect it from aerodynamic heating during its ascent phase. At the required altitude and Mach number, the heat shield need to be jettisoned at an optimum angle and the safe separation of heat shield without touching the cruise vehicle/launch vehicle must be ensured for the success of the mission. Typical configuration is shown in Figure 4. CFD studies [23] are carried out to estimate the forces and moments acting on the panel at different opening angles. Separate grids have been generated for different panels and they are overlapped and the cloud of points around the geometry is generated after blanking the solid node. The cloud of points in the computational domain is shown in Figure 5. 3D grid free Euler Solver q-LSKUM is used to solve the flow field for different panel opening conditions. The grid blocks are only rotated about the hinge lines to get the grids for

Table 1. *Comparison of computed and measured aerodynamic forces and moments*

M_∞	α	C_N		C_m	
		CFD	*Experiment*	*CFD*	*Experiment*
0.6	2.5	0.494	0.490	–2.308	–2.170
0.6	5	0.984	1.046	–4.723	–4.712
3.0	2.5	0.335	0.375	–1.475	1.653
3.0	5	0.705	0.793	–3.175	–3.535

Table 2. *Comparison of computed and measured jet vane parameters*

Angle of attack	C_N		X_{cp}/c	
	CFD	*Experiment*	*CFD*	*Experiment*
6°	0.157	0.15	0.57	0.59
12°	0.326	0.32	0.54	0.54
618	0.515	0.52	0.53	0.53

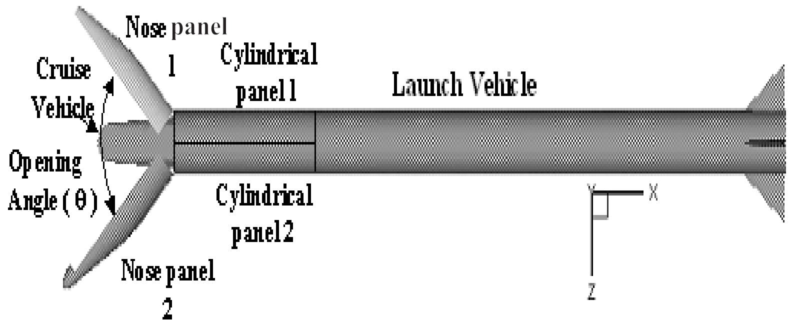


FIG. 4 HEAT SHIELD SEPARATION SYSTEM FOR HYPER-SONIC VEHICLE

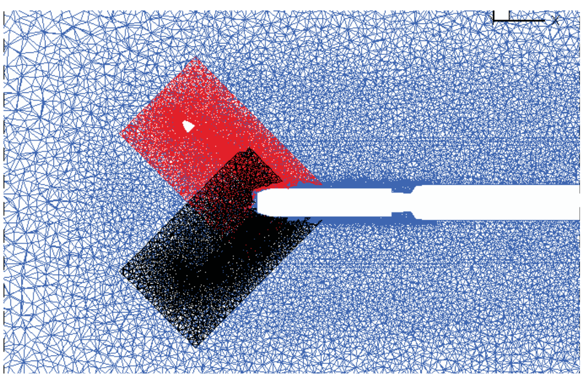


FIG. 5 CLOUD OF POINT FOR HEAT SHIELD SEPARATION SIMULATION

the different opening angles and the flow field for the previous conditions has been used in the initial guess for the next opening condition. This has reduced the grid generation time and the flow convergence time drastically. The Mach number contour for the different opening angle 50° and 120° are shown in Figure 6. For small opening angle, an attached shock is visible for panel 1 and a bow shock for panel 2 (for nose bluntness) is seen to interact with the cruise vehicle bow shock. For higher opening angle, the triple shock structure is replaced by a strong bow shock. The forces and moments on the panels for different opening angles are shown in Figure 7. There results are being used for the separation dynamics studies of the nose and cylindrical panels of the hypersonic technology demonstrator vehicle.

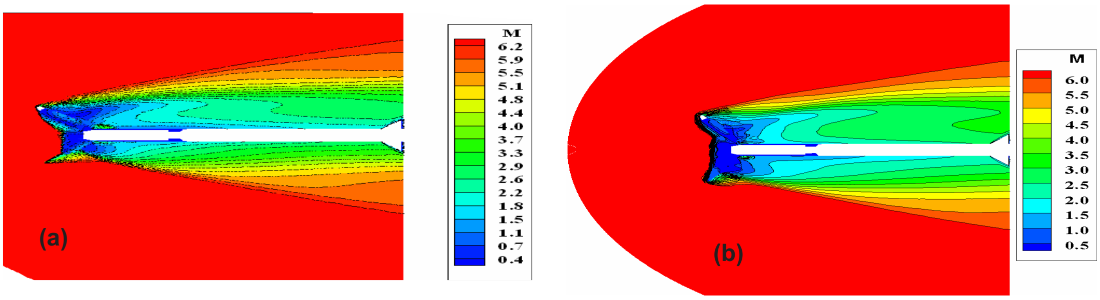


FIG. 6 MACH CONTOURS AT SYMMETRY PLANE FOR DIFFERENT OPENING ANGLES (a) 90° (b) 120°

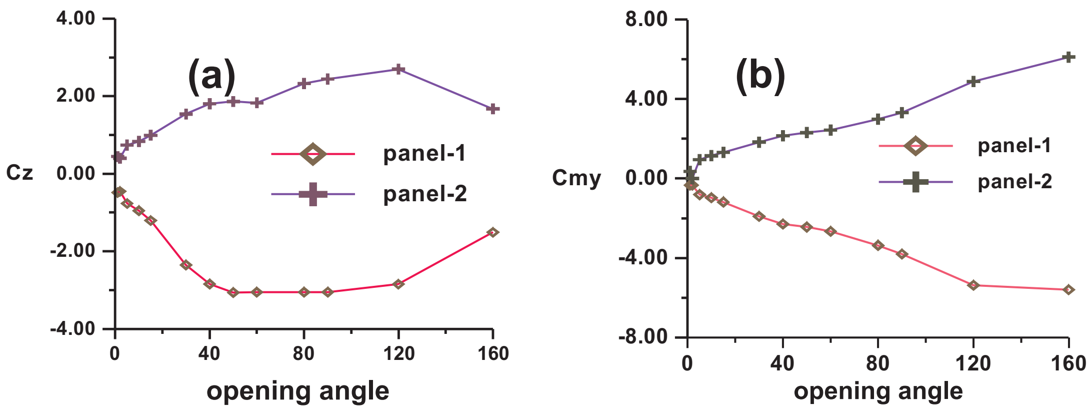


FIG. 7 AERODYNAMIC (a) FORCES AND (b) MOMENTS OF THE PANELS FOR DIFFERENT OPENING ANGLES

3.3 Missile Separation from Aircraft Wings

Safe separation of missile from aircraft is mission critical. Hence, accurate estimation of aerodynamic load on the missiles and aircraft is very much necessary to design the separation sequence. Flight testing and CTS wind tunnel tests were generally carried out for this design of separation system. In recent years, CFD approaches in simulating store separation become very successful. In DRDL, 3D Euler Grid free method (q-LSKUM) has been integrated with 6 DOF trajectory program [14] to carryout the separation dynamic study of the missile from the aircraft wing/full aircraft. The preprocessor overlap the grid around wing and the missiles to from a chimera cloud of points and generate connectively using grid information and gradient search method. The grid cloud around the missile is moved relative to the grid of the wing and the connectively is updated at every time step. The aerodynamic forces and moments obtained from CFD solver are used to solve the 6 DOF equation of motion for obtaining the new position of missile [24].

The chimera cloud of points around the missile and the wing is shown in Figure 8 and typical surface Mach number distribution of the wing and missile at various instant of time is shown in

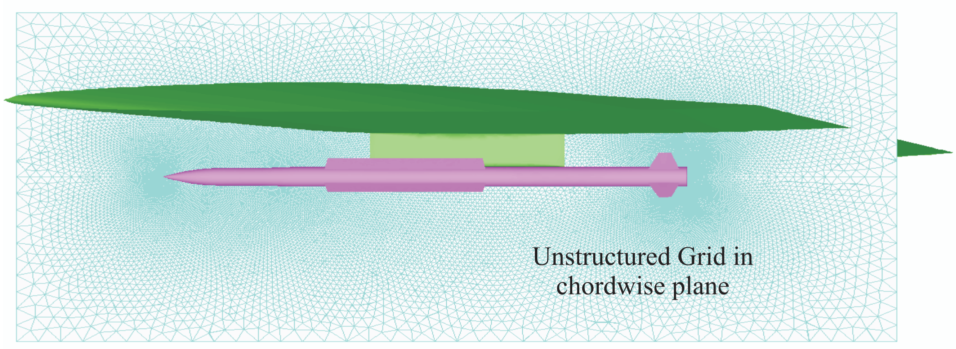


FIG. 8 CHIMERA CLOUD OF POINTS AROUND THE MISSILES

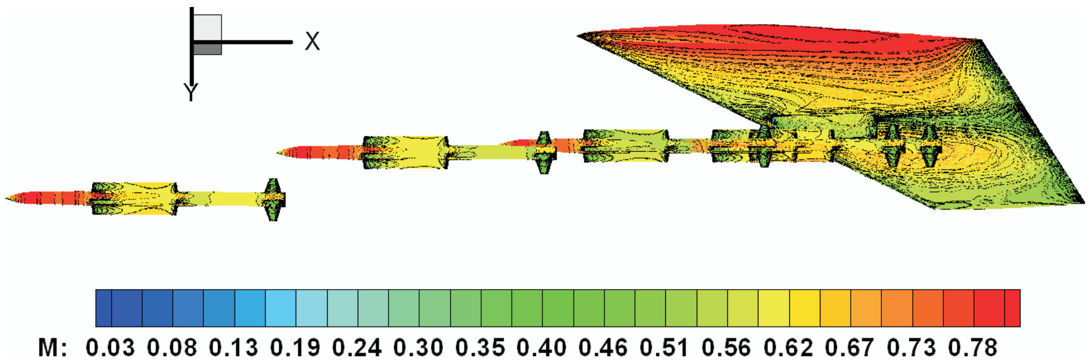


FIG. 9 SURFACE MACH NUMBER DISTRIBUTION OF WING AND MISSILE AT VARIOUS INSTANT OF TIME

Figure 9. The motion is restricted to 1DOF till the vehicle leaves the pylon as it is constrained by rail motion. After it leaves pylon, it has all 6DOF motion and the missile experience pitch down motion.

4. APPLICATION TO INTERNAL FLOWS

Core competence is also being developed in numerical simulation of turbulent reacting flows. Propulsion systems of various ongoing and future missile projects are being designed and analyzed. Three dimensional RANS equations are solved alongwith κ-ε / κ-ω / SST turbulence model and Eddy dissipation based combustion models using commercial software [25]. Liquid fuel is considered as dispersed phase fluid and is modeled through Lagrangian tracking method. The software was validated extensively against reliable data pertaining to reacting and nonreacting propulsion tests. Notable internal flow studies include scramjet combustor flow field simulation, Jet vane flow field simulations, Air intake characteristics estimation, exhaust plume-free stream interaction at high altitude at base region, conjugate heat transfer studies, starting process in the nozzles and diffusers etc. Few of these internal flow simulations are presented in this paper.

4.1 Simulation of Installed Air Intake Flow Field

Intake performance is a critical point in the design of ramjet and other supersonic airbreathing mission. The intake of a supersonic airbreathing engine is required to capture and efficiently compress the air so that after heat addition the flow can be expanded in the nozzle to produce thrust. The intake is very sensitive to the interaction with the upstream external flow and downstream combustion process and hence exhibits complex flow phenomena in its range of operation. The performance of the individual intake gets modified when it is installed to the core body at downstream location. The variation of the performance may be very significant when the vehicle is at angle of incidence. The flowfield of the installed air intake mounted on the rear portion of a missile is analysed. The methodology has been validated against the literature results of isolated mixed compression air intake flow field [26] and was used to predict the installed intake performance at an angle of incidence [27].

Four intakes are attached to the rear portion of ogive-cylinder core body of the ramjet missile. The intakes are connected to a dump chamber and back pressure is simulated through a plug in the dump chamber. The computational domain of the problem is shown in Figure 10 which include the forebody and the internal flow path in the intake and the dump chamber. As the interest of the study is to estimate the intake characteristics in installed mode, the external flow domain is terminated at certain downstream distance of the cowl lip. Simulations are carried out with different back pressure to find out the pressure recovery (π) vs mass flow ratio (η) characteristics of the intake at M = 2 and 6° angle of attack. The computed values are compared with the experimental values [28] in Figure 11. Good overall agreement has been obtained. The cause of the slight difference in the performance of left and right intakes in supercritical range of operation found in the experimental results is not very clear.

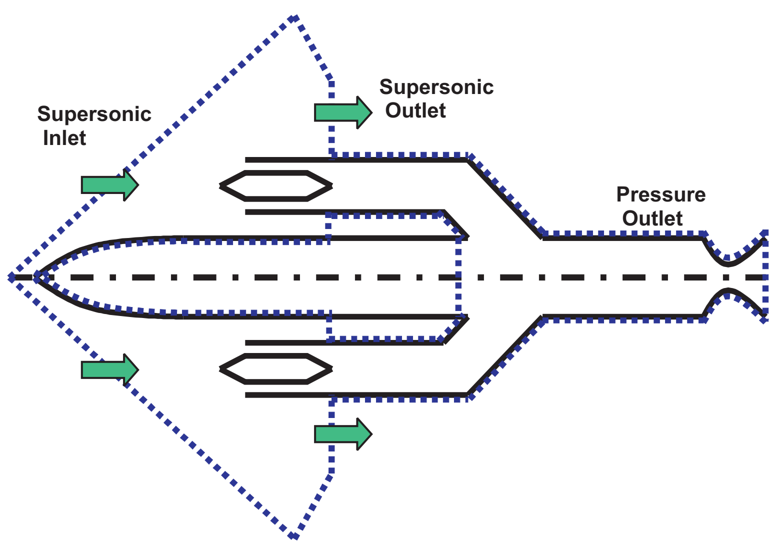


FIG. 10 SCHEMATIC OF COMPUTATIONAL DOMAIN FOR INSTALLED AIR INTAKE FLOW FIELD SIMULATION

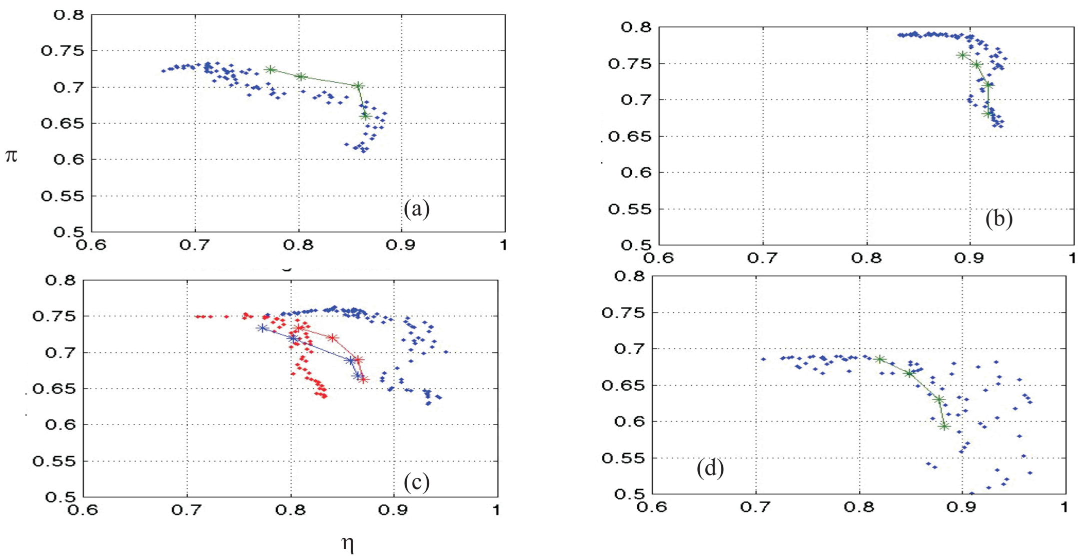


FIG. 11 COMPARISON OF COMPUTED AND MEASURED INTAKE CHARACTERISTICS AT M = 2 AND A = 6° (A) TOP INTAKE (B) BOTTOM INTAKE (C) LEFT AND RIGHT INTAKE (D) DUMP END

4.2 Simulation of Jet Vane Flow Field

At the initial stage of missile launching, when the speed is not large, the control and stability requirements of the missiles are met by Thrust Vector Control (TVC) system. Jet Vane is one of the TVC system and is employed in many missiles. The system is operated with small torque, installed in small space and control pitch, roll and yaw simultaneously. A typical Jet Vane configuration mounted in the rear section of nozzle is shown in Figure 12. Numerical simulations are carried out for the aerodynamic characterization of Jet Vane exposed to hot exhaust gas of the rocket nozzle with temperature as high as 3000 K and velocities up to Mach 3.5 that exerts extreme mechanical and thermal load on the vane. The computed normal force coefficient (C_N) and the center of pressure (X_{cp}/c), where C is the root chord length, is compared with the experimental results for cold flow conditions for different angle of attack up to 18° and a good match is obtained [29].

Numbers of simulations are carried out with different deflections of opposite vanes in the presence of hot rocket exhaust and the computed pressure field and jet vane characteristic was used in the structural design of the vanes and control design of the missile respectively. Typical Mach number distribution in the jet vane flow field is shown in Figure 13. The computational characteristics of the jet vane match with the static test results till the erosion of the vane surface is limited. Efforts are continuing to couple erosion modeling with CFD solver.

4.3 Simulation of Starting Process in Supersonic Nozzle

Short duration facilities like shock tunnel and expansion tubes are very important to study hypersonic flow. They can generate the required combination of temperature and pressure for

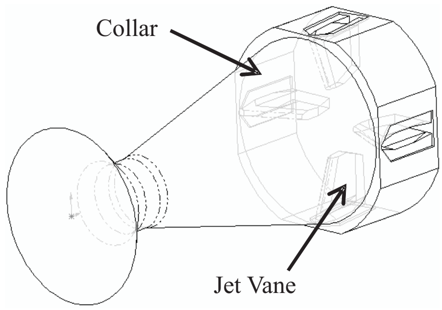
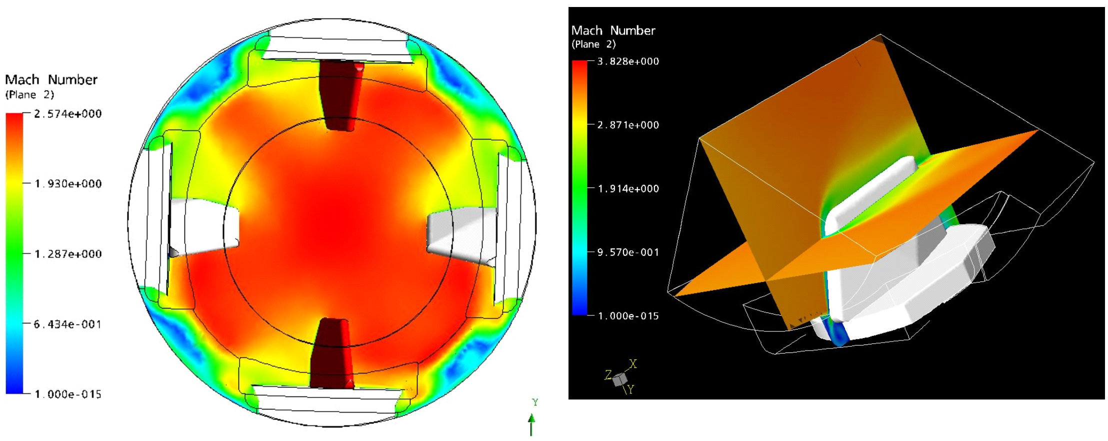


FIG. 12 TYPICAL JET VANE CONFIGURATION



(a) (b)

FIG. 13 MACH NUMBER DISTRIBUTION IN THE JET VANE FLOW FIELD (a) CROSS SECTIONAL VIEW IN THE NOZZLE END (b) AROUND THE JET VANE

simulating the high Mach number flight condition. The prediction of starting process of the shock tunnel is very important to determine the run time of the facility. This phenomena is induced by unsteady shock / boundary layer interaction. The presence of reflected waves in the nozzle also affects the starting up process. The mechanism of shock wave propagation during the nozzle start up process is quite complex and a fundamental knowledge of flow physics is still needed to understand the starting process. The transient flow field in a nozzle driven by a shock tube is explored numerically [30]. The experimental condition of nozzle starting process [31] is taken as the validation case and the flow parameters are compared with the experimental and other numerical simulation results. The numerical schlieren in the flow field at $t^* = 4$ is presented in Figure 14 to depict the primary shock, secondary shock, contact discontinuity and the reflected normal shock. The time is normalized on $t^* = t\, a_0/h_c$, a_0 is the speed of sound of gas initially at rest and h_c is the throat diameter. The primary shock is seen propagating into medium at rest, while the secondary shock is moving relative to the gas. The contact discontinuity is also clearly seen in the schlieren picture; it is getting distorted owing to Ritchmeyer Meshkov instability. The numerical

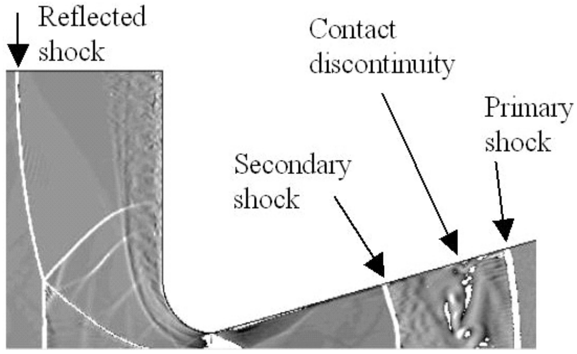


FIG. 14 NUMERICAL SCHLIEREN OF THE NOZZLE TRANSIENT FLOW FIELD AT t* = 4

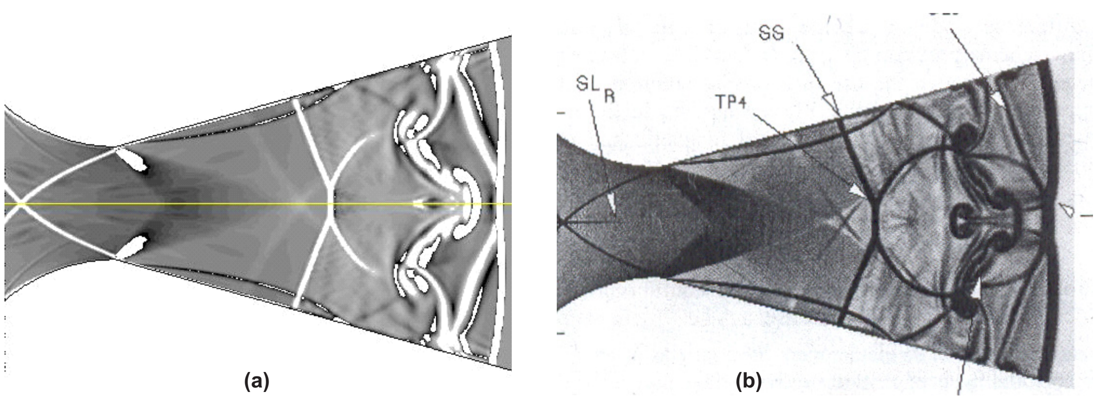


FIG. 15 COMPARISON OF NUMERICAL SCHLIEREN OF STARTING UP PROCESS IN THE NOZZLE (a) PRESENT COMPUTATION (b) COMPUTATION OF REF [32]

schlieren of the flow field near the nozzle throat at early stage of start up process at t^*=1.14 is compared with that of Mouronval and Hajjadi [32] in Figure 15. The complex flow structure between the primary shock and contact discontinuity is captured in the simulation.

The comparison of experimental and the numerical values of the primary and the secondary shocks trajectories are presented in Figure 16. The axial distance is normalized by the nozzle throat diameter and the time is normalized by t^* as defined earlier. The position of primary and secondary shock at any instant of time has been determined from the location at which the wall pressure is discontinuous. The excellent agreement between the experimental and numerical values demonstrate that the simulation has captured all the features of the flow accurately. Viscosity does not alter the flow feature significantly as seen from the comparison of viscous and inviscid results in the figure.

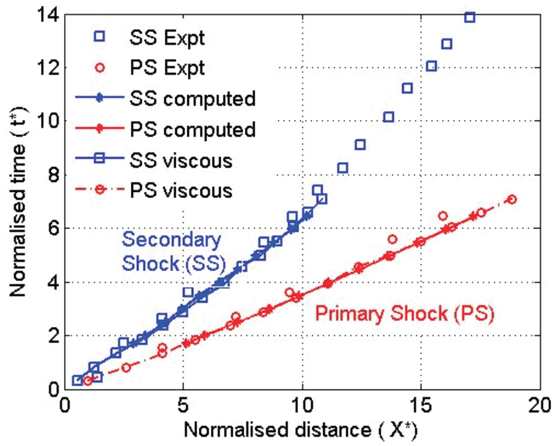


FIG. 16 COMPARISON OF PRIMARY AND SECONDARY SHOCKS TRAJECTORIES BETWEEN EXPERIMENT AND COMPUTATION

5. CONCLUSIONS

CFD is playing a very important role in the aerodynamics and propulsion design of various DRDO missiles. Number of grid generator (AUTOELGRID), 3D Euler (KFSG, q-LSKUM) and Navier Stokes (CERANS) solvers have been developed inhouse and being used routinely in aerodynamic characterization of missiles in complete M- á regime. State of art numerical techniques, namely; kinetic schemes, grid free methods and robust turbulence closures like Spalart Allmaras, SST turbulence models etc are used in these softwares. The development of entropy based (q) Least Square Kinetic Upwind grid free Method (q-LSKUM) and its integration with 6 degrees of freedom trajectory equations has enabled us to study the control surface deflection and store separation problems with ease. A new unsteady CFD solver (KFMG) and Euler Boltzmann coupled solver have been developed for application to computational aeroelasticity and low density reentry flows respectively. These developed solvers are continuously upgraded with the inclusion of new numerical techniques and better physical modeling to increase its accuracy and range of application. Significant contributions were made in the numerical simulation of reacting and nonreacting internal flow field using commercial software. The software was validated extensively against reliable propulsion test results and used to optimize various high performance propulsion systems. Important inputs are provided in the design of scramjet combustor, jet vane, divert thruster motor, base shroud etc. Various advanced topics like computational aeroservoelasticity, Large Eddy Simulation, Ignition modeling, atomization modeling etc. are being pursued in-house and also with collaboration with national and international academic institutes.

ACKNOWLEDGEMENTS

The work carried out by the scientists of the Directorate of Computational Dynamics (DOCD), DRDL, Hyderabad are compiled and presented here. Author is thankful to all the members of DOCD for their help in preparing this paper.

REFERENCES

[1] R. Krishnamurthy, K. Anandhanarayanan and M. Nagarathinam, “Elliptic Grid Generation Around Flight Vehicle Configurations”, Proc. 4th ECCOMAS Conference, Athens, Greece, 7-11 September. 1998.

[2] J.C. Mandal and S.M. Deshpande, (1994), “Kinetic Flux Vector Splitting for Euler Equations”, *Computers & Fluids*, vol. 23, pp. 447-478

[3] S. Mondal, K. Anandhanarayanan and R. Krishnamurthy, “CFD based aerodynamic characterization for a flight vehicle configuration”, published in the proceedings of Recent Advances in Computational Methods in Aerospace Engg., MIT, Chennai, India, February 2006

[4] R. Krishnamurthy, B.S. Sarma and S.M. Deshpande, “Kinetic Flux vector splitting scheme for unsteady Euler equations on Moving Grids (KFMG)”, presented in 3rd Asian Computational Fluid Dynamics conference (ACFD3), held at Bangalore, India and published in proceedings of the conference (December 1998).

[5] R. Krishnamurthy, B.S. Sarma and S.M. Deshpande, “KFMG based Euler solver for Computational Aeroelasticity”, presented in first International Conference on Computational Fluid Dynamics (ICCFD1), held at Kyoto, Japan (June 2000) and published in Lecture notes in physics.

[6] R. Krishnamurthy, B.S. Sarma and S.M. Deshpande, “Kinetic scheme for computational aeroelastic analysis of 2-D airfoils in transonic flows” published in the proceedings of 34th AIAA Fluid dynamics conference, Portland, USA, Paper no. : 2004 – 2236.

[7] R. Krishnamurthy, B.S. Sarma and S.M. Deshpande, “3-D KFMG Euler Computations for Unsteady Flows Around Oscillating Geometries”, presented in second International Conference on Computational Fluid Dynamics (ICCFD2), held at Sydney, Australia (July 2002).

[8] R. Krishnamurthy, B.S. Sarma and S.M. Deshpande, “3-D unsteady flow computations around oscillating wing using Kinetic upwind Euler solver”, published in the proceedings of 24th AIAA Applied Aerodynamics Conference, San Francisco, Paper no.: 2006 – 3379.

[9] A.K. Ghosh and S.M. Deshpande, “Least Squares Kinetic Upwind Method for Inviscid Compressible Flows”, AIAA Paper No. 95-1735, 1995.

[10] S.M. Deshpande, K. Anandhanarayanan, C. Praveen and V. Ramesh, “Theory and Applications of 3-D LSKUM Based on Entropy Variables”, Int. Jl. of Numerical Methods in Fluids, Vol. 40, 2002, pp. 47-62.

[11] K. Anandhanarayanan, M. Nagarathinam and S.M. Deshpande, “An Entropy Variable-Based Grid-free Euler Solver with LU-SGS Accelerator”, Proceedings of 24th ICAS Congress, Yokohama, JAPAN, August 29–September 3, 2004.

[12] K. Anandhanayanan, M. Nagarathinam and S.M. Deshpande, “Parallelisation of a Grid-free Kinetic Upwind Solver”, AIAA paper No. AIAA 2005-4846, 2005.

[13] K. Anandhanayanan, M. Nagarathinam and S.M. Deshpande, “Development and applications of a Grid-free kinetic upwind solver to multi-body configurations”, AIAA paper No. AIAA 2005-4846, 2005.

[14] G. Harish, M. Pavanakumar and K. Anandhanarayanan, "Store Separation Dynamics using Grid-free Euler Solver", AIAA Paper 2006-3650, June 2006.

[15] R. Balasubraminan and N. Balakrishnan, "Development of a 3D Laminar Compressible Navier-Stokes Solver", 2nd Internal Conference on Computational Fluid Dynamics, Sydney, Australia, July 2002.

[16] R. Balasubraminan, K. Anandhanarayanan and N. Balakrishnan, "Development of a 3D Compressible Reynolds Averaged Navier-Stokes solver", ECCOMAS-2004, Jyvaskyla, Finland, July 2004.

[17] R. Balasubraminan, K. Anandhanarayanan and J. Umakant, "Computational and Experimental investigations of Flight Vehicle Aerodynamics", EWHSFF-2005, Beijing, China, October 2005.

[18] R. Balasubraminan, K. Anandhanarayanan, "Viscous Computations for Complex Flight Vehicles using CERANS with Wall Function", Proc. of the SAROD-2005, Hyderabad, India, December – 2005.

[19] G.A. Bird, "Molecular Gas Dynamics and the Direct Simulation of Gas Flows", Clarendon press, Oxford, 1998.

[20] P.Le. Tallec and F. Mallinger, "Coupling Boltzmannn and Navier-Stokes Equations by Half Fluxes", J.Comp. Phy. 136, 51-67, 1997.

[21] S. Tiwari, "Coupling of the Boltzmann and Euler Equations with Automatic Domain Decomposition", J. Comp. Phy. 144, 710-726, 1998.

[22] K. Anandhanarayanan, Vaibhav Shah, R. Krishnamurthy and Debasis Chakraborty "Numerical Exploration of an Aerospace Vehicle with deflected fins using a grid-free solver", Proceedings of SAROD-2007 held at Thiruvananthapuram November 22-23, 2007, pp 536-544.

[23] Rohit R. More, S. Srinivas Raju, K. Anandhanarayanan and R. Krishnamurthy " Nose panel opening study of a hypersonic launch vehicle" Proceedings of International Conference on high speed Transatmospheric Air and Space Transportation, June 2007, Hyderabad, pp 171-177.

[24] P.K. Sinha, R. Krishnamurthy and Debasis Chakraborty: "CFD in Aero propulsive design of missiles", Proceedings of 7th Asian Computational Fluid Dynamics Confernce (ACFD-7) held at Bangalore November 26-30, 2007, pp 68-83.

[25] CFX-TASC Flow Computational Fluid Dynamics Software, Version *2.11.1*, AEA Technology Engineering Software Ltd. 2001.

[26] Oh, J. Y., Fuhaha Ma, Shih–Yang Hseih and Vigor Yang (2005): "Interactions between shock and acoustic waves in a supersonic inlet diffuser", Jl. of Propulsion and Power, Vol 21, No. 3, pp 486-494.

[27] Soumyajit Saha, P.K. Sinha and Debasis Chakraborty: "CFD Prediction of Ramjet Intake Characteristics at Angle of Attack", Proccedings of 8th National Conference on Air breathing engines and Aerospace Propulsion, held on December 12-14, at DIAT, Pune pp 97-107.

[28] Theerthamalai P., and Mani Sekharan S., Umakant J. and Selvarajan S., (2002): “Experimental investigation of an Installed air intake model”, National Conference on Air Breathing Engine-2002, Bangalore, January 2003.

[29] S. Saha, P.K. Sinha, Sampath Kumar, A.Satyanarayan and Debasis Chakraborty : “Jet vane flowfield characteristics using CFD”, Recent trends in Aerospace Design and Optimization, Proceedings of SAROD-2005. pp 335-346.

[30] Soumyajit Saha and Debasis Chakraborty: “Numerical Exploration of Starting Process in Supersonic Nozzle” Aeronautical Journal Vol. pp 51-58.

[31] Amann H. O., Experimental study of the starting process in a reflection nozzle, Physics of Fluids, Vol 12, No 5, 1969, pp 150-153.

[32] Mouronval, A. S. and Hadjadi A., Numerical study of the starting process in a supersonic nozzle, Jl of Propulsion and Power, Vol 21, No. 2, 2005, pp 374-378.

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

NUMERICAL SIMULATION OF LIQUID FUELED SCRAMJET COMBUSTOR FLOW FIELD

DEBASIS CHAKRABORTY*

ABSTRACT

Numerical simulations are carried out for various reacting flow fields related to scramjet propulsion system using commercial CFD Software. Three-dimensional Navier stokes equations are solved along with K-ε turbulence model. Modeling of turbulence chemistry interaction is done through infinitely fast rate chemical kinetics. The software was validated extensively by comparing different experimental condition for scramjet combustor to find its error band and range of application. Good agreement between the experiment and computation are obtained for the scramjet combustor flow field with sturt, pylon and cavity injection system with both hydrogen and hydrocarbon fuel. The CFD tools are applied in the design exercise of flight sized kerosene fuel scramjet combustor of a hypersonic airbreathing mission. Significant improvement of combustion efficiency and thrust could be achieved by relocating the fuel injection system through the analysis of various thermochemical variables in the scramjet combustor.

1. INTRODUCTION

The success of efficient design of a hypersonic airbreathing cruise vehicle largely depends on the proper choice of propulsion system. This type of vehicle, according to current proposal will use Scramjet propulsion system. Both hydrogen and hydrocarbon fuels are considered depending on applications and speed range. Although, hydrogen is having attractive features in terms of specific impulse, ignition characteristics etc., liquid hydrocarbon is required for volume limited applications in lower hypersonic region ($M < 8$). Starting from the pioneering work of Ferri [1] significant advances are made in the design of Scramjet engines. Over the last few decades, there has been much effort on the research pertaining to analytical, experimental and CFD areas to understand the mixing and combustion process in the Scramjet combustor. In a recent review,

*Defence Research and Development Laboratory, Hyderabad - 500058, INDIA
E-mail: debasis_cfd@drdl.drdo.in

Curran [2] has identified two emerging Scramjet applications namely (1) hydrogen fueled engine to access space and (2) hydrocarbon-fueled engines for air-launched missiles.

Considerable efforts have been focused on different injection schemes like cavity, sturt, pylon for different geometrical configurations and flow conditions in the past two decades. Selected methods that have been used to enhance the mixing process in the Scramjet engines are summarized and reported in Ref 3. Atomization, vaporization, mixing and slow chemical reactions are some of the major barrier in the realization of liquid hydrocarbon fuel based scramjet. The problem of slow lateral fuel transport in the air stream can be circumvented by injecting the fuel in the core region of the flow by means of struts and or pylons. The oblique shocks generated from the struts also augment the mixing which is very much needed in high speed propulsion devices. A good number of experimental and numerical studies [4-11] were reported in the literature to focus on various aspects of flow phenomena including drag losses, mixing, combustion, intake combustor interaction etc., in strut based scramjet combustors with hydrogen fuel. The reported experimental and numerical studies on kerosene fueled supersonic combustion mostly address the issues of cavity based flame holder and injection system [12-18] in laboratory scaled combustor.

The studies on strut-based scramjet combustor with kerosene fuel are highly limited. Vinogradov et al [19] conducts experimental investigation to determine the ignition, piloting, and flame holding characteristics in a strut based scramjet combustor operating on kerosene. In order to improve the fuel distribution and mixing, kerosene was injected from the strut located in the middle of the duct. Stable combustion of kerosene was achieved even after turning off pilot hydrogen. Bouchez, Dufour and Montazeal [20] carried out experimental investigation of hydrocarbon fueled scramjet combustor. Two identical metallic water-cooled and liquid kerosene cooled struts were used for the fuel injection in the combustor. To ensure ignition, pilot flames with gaseous hydrogen was used at the base of the struts. Kerosene equivalence ratio was varied from 0 to 1.0. Various flow parameters (wall pressure, wall heat flux, total temperature at combustor exit, thrust etc.) were measured. Optical methods including passive spectroscopy were also used to characterize the flow.

With the advent of powerful computer, robust numerical algorithm, CFD is complementing ‘difficult to perform’ experiment and thus playing a major role in developing a comprehensive understanding of the key phenomenon that dominate performances. Only very few numerical studies were reported on strut based liquid fueled scramjet combustor. Dufour and Bounchez [21] have numerically simulated the scramjet experiment [20] using a three dimensional Navier Stokes solver and single step chemical kinetics. A reasonably good match is obtained between the computational and experimentally measured wall static pressure. Recently, Manna, Behera and Chakraborty [22] presented a CFD based design and analysis for a flight scale scramjet combustor with kerosene fuel injected from struts placed in the combustor flow path and emphasized that higher combustor entry mach number and distributed fuel injection system is required to avoid thermal chocking. But before the CFD software is used for the design exercise, it is necessary to make a thorough validation checks to find its range of application and error band. This paper present the validation studies for a few hydrogen and kerosene fuelled scramjet combustor and the use of CFD techniques for the development of full scale scramjet combustor for an hypersonic airbreathing mission.

2. METHODOLOGY

The software, used in the present study, is a three dimensional Navier Stokes code – CFX-TASCflow [23] which is an integrated software system capable of solving diverse and complex multidimensional fluid flow problems. The code is fully implicit, finite volume method with finite element based discretisation of geometry. The method retains much of the geometric flexibility of finite element methods as well as the important conservation properties of the finite volume method. It utilizes numerical upwind schemes to ensure global convergence of mass, momentum, energy and species. It implements a general non-orthogonal, structured, boundary fitted grids. In the present study, to circumvent the initial numerical transient, the descretisation of the convective terms are done by first order upwind difference scheme till few time steps and subsequently, the convective terms are discretized through 2nd order scheme to capture the flow features more accurately. The turbulence model used was K - ε model with wall functions

The chemistry of Hydrogen-air combustion reaction is represented on a molar basis by $H_2 + 0.5O_2 = H_2O$; whereas kerosene – air combustion reaction is represented by: $C_{12}H_{23} + 17.75O_2 = 12CO_2 + 11.5H_2O$. The mixing rate determined from the Eddy Dissipation Model (EDM) is given as.

$$R_{k,edm} = -A_{ebu}\,\rho\,\frac{\varepsilon}{K}\,\min\left\{Y_f\,,\frac{Y_o}{r_k},\,B_{ebu}\frac{Y_p}{1+r_k}\right\}$$

where ρ, Y_f, Y_o and Y_p are the density and mass fractions of fuel, oxidizer and products respectively, A_{ebu} and B_{ebu} are the model constants and r_k is the stoichiometric ratio.

3. RESULTS AND DISCUSSIONS

3.1 Validation Studies

Four different experimental conditions are selected for comparisons with numerical simulations so that most of the issues related to Scramjet propulsion could be addressed.

- Staged supersonic combustor with hydrogen fuel from strut and wall injector[24]
- Hydrogen fuelled Scramjet combustor with pylon injector[25]
- Kerosene fuelled scramjet combustor with cavity injector[26]
- Kerosene fueled scramjet combustor with ramp-cavity combustor[27]

3.1.1 Staged supersonic combustor with strut injection

Strut based Scramjet combustors with hydrogen fuel are extensively investigated experimentally at NAL, Japan [4-8] focusing on various aspects of drag losses, mixing, combustion, intake-combustor interactions etc. Injection of excessive fuel form strut can cause inlet-combustion interaction that may lead to engine unstart condition. To reduce such inlet-combustor interaction, a staged supersonic combustor with strut for the first stage injection and second stage wall injection at divergent section was experimentally studied by Tomioka et al. [8, 9] in a direct-connect wind tunnel facility. The combustor test were carried out for the combustor entry Mach number of 2.5, total temperature of 1500k and total pressure of 1.0 MPa which are almost the combustor entrance condition under Mach 6 flight condition. Sonic hydrogen was injected from the strut and the wall with various equivalence ratios. The schematic diagram of the combustor for which the computations are carried out is shown in Figure 1. A strut with blunt leading edge (1mm in radius) and a compression part (43.8 m) with half wedge angle of 6°, followed by 28 mm straight portion is installed in the constant area section. Downstream of the contrast area section, the sidewall diverged at an angle of 3.1° for 600 mm. Taking the advantage of two plane of symmetry, only one fourth of the geometry is simulated [24]. The computed axial pressure distribution for the non-reacting case is compared with the experimental values in Figure 2. Pressure has been nondimensionlized by total pressure and the origin (X = 0) is taken at the location of step. It can be seen that by changing the grids from 1,71,105 to 2,95,071 has not changed the results appreciably thus demonstrating the grid independence of the results. A very good agreement between the experiment and the computation has been obtained. All the shock interaction in the combustor has been captured very nicely in the computations.

Reacting flow computations are carried out with various fuel injections from the strut and the wall. The temperature and water mass fraction distribution in the combustor for the strut injection case are shown in Figure 3. It is clear from the figure that the reaction occurs only in the central zone of the combustor. The flow field in the rest of the combustor almost remains unreacted. The surface pressure comparison between the experiment and the computation is presented in Figure 4. A very good match is observed except near the fuel injection location, where the computation overpredicts the surface pressure. This difference is due to the use of fast chemistry, which cause instantaneous heat release in the modeling results in prediction of higher surface pressure.

The surface pressures for various reacting and non-reacting cases are compared in Figure 5. With the second stage injection from the wall, there is considerable increase in pressure in the divergent section, which is very necessary to increase the thrust. For the case of injection from strut and wall

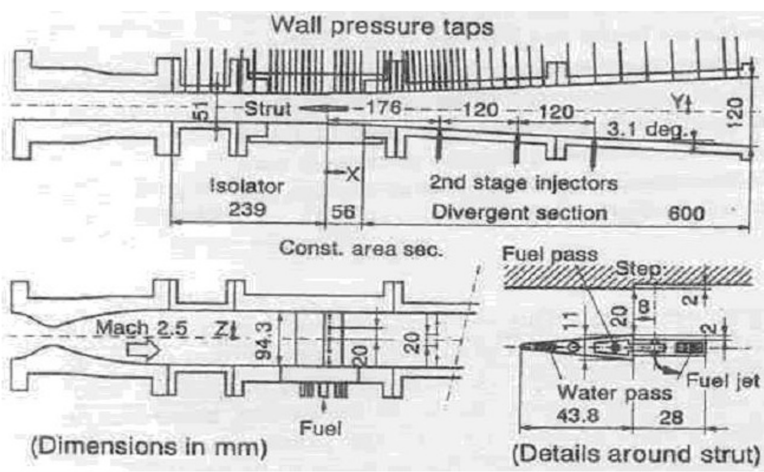


FIG. 1 SCHEMATIC OF THE SCRAMJET COMBUSTOR WITH THE STRUT

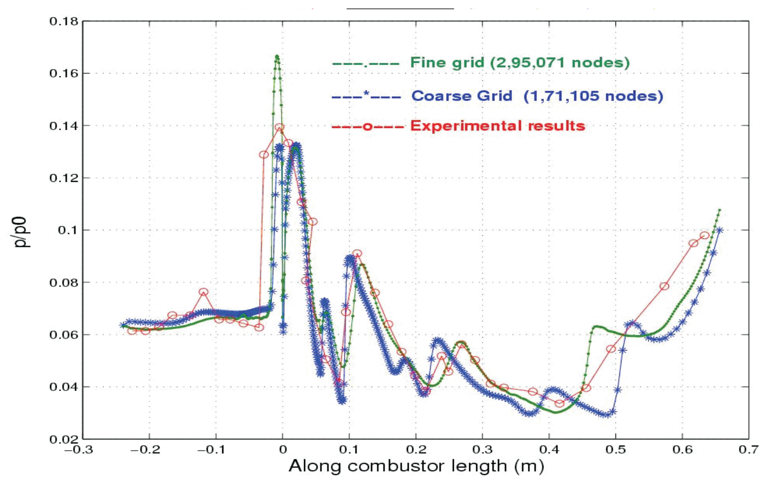


FIG. 2 COMPARISON OF WALL PRESSURE DISTRIBUTION FOR NON REACTING CASE

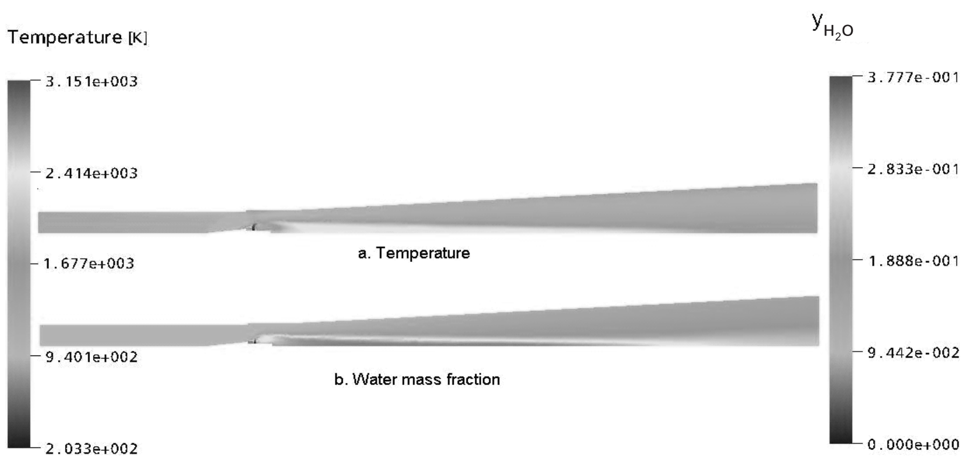


FIG. 3 FLOW VARIABLES IN THE SYMMETRY PLANE FOR REACTING CASE WITH STRUT INJECTION (φ_1 = 0. 34) (a) TEMPERATURE (b) WATER MASS FRACTION

(Φ_1=0.35 + Φ_2=0.44), there is no influence of second stage injection in the first stage, while for the fuel rich case (Φ_1=0.35 + Φ_2=0.90), the influence of wall injection has felt slightly upstream but did not effect the pressure peak caused due to strut injection.

3.1.2 Supersonic combustion with pylon injection

Selecting proper injector geometry is one of the most important considerations for the Scramjet engine development. Fuel injection from the wall will result in reaction zones that occupy only a

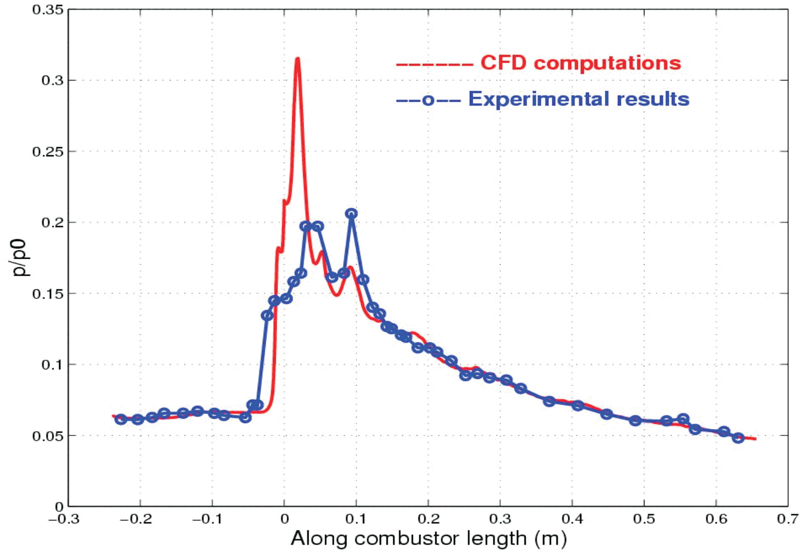


FIG. 4 WALL PRESSURE COMPARISON FOR REACTING CASE (FUEL INJECTION FROM STRUT (φ_1 = 0. 34))

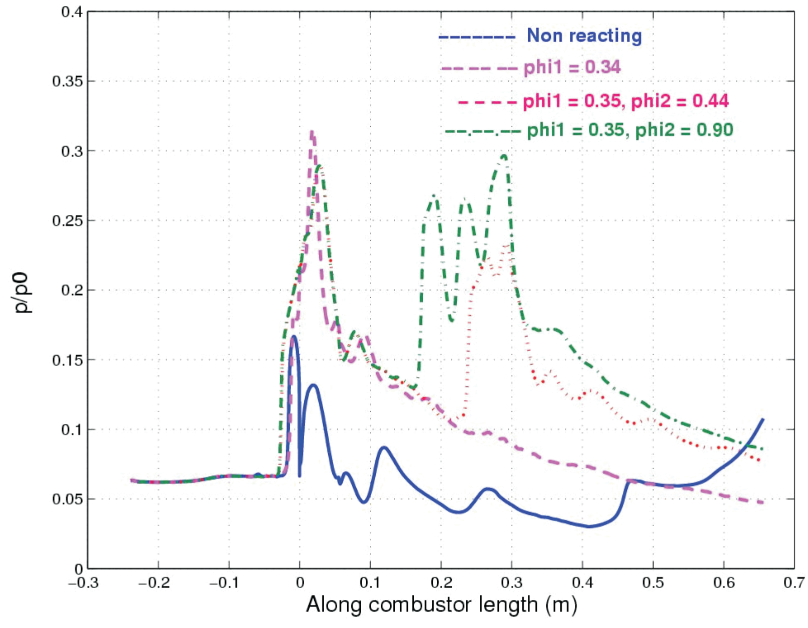


FIG. 5 COMPARISON OF WALL PRESSURE DISTRIBUTION FOR ALL REACTING AND NON REACTING CASE

small fraction of the flow field. Therefore, not all of the oxygen supplied by the air stream entering the combustor can participate in the chemical reaction. This problem of slow lateral fuel transport in the airstreams can be circumvented by injecting the fuel in the center of the flow by mean of pylons. Although, few pylon designs have been investigated experimentally [29, 30] to study their effectiveness in fuel mixing and combustion in the Scramjet combustion, numerical simulations of the reacting flow field for pylon injected scramjet combustor is very limited. Numerical simulation of experimental conditions of Grunneig et al [28] has been presented here. The combustor is having 645 mm long, 25 mm width and its height is 27.5 mm at the entry and 40.5 mm at the exit. The pylon is placed at a distance of 45 mm from the combustor entry. An expansion angle of 5° is provided on the lower surface of the combustor for 150 mm just after the pylon. The combustor along with the grid distribution with blown up view near the pylon is shown in Figure 6. Sonic hydrogen with 3.9 bar total pressure and 288 K total temperature is injected at 120° from slot shaped orifice to the vitiated air of Mach 2.15, P_o=7.8 bar & T_o=1350 K. Water mass fraction distribution in the plane of injection is presented in Figure 7, which clearly depicts the reaction zone occupying a significant portion of the combustor flow field. The comparison of surface pressure distribution for the top wall between the experimental and computational values are shown in Figure 8. A reasonably good match is obtained.

3.1.3 Kerosene Fuelled Scramjet Combustor With Cavity Injector

The schematic of cavity based scramjet combustor experiment [13] for which the computations are carried out is shown in Figure 9. The combustor has rectangular cross section with an entry of 51×70 mm^2. The length of the combustors is 1070 mm and consists of four sections. Different types of integrated wall injector cavity configuration were designed and tested at various stagnation conditions with liquid kerosene fuel. Numerical investigations [25] are carried out with three

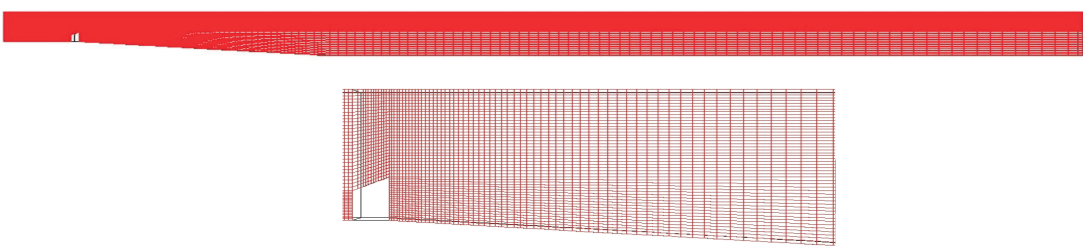


FIG. 6 GRID DISTRIBUTIONS IN THE INJECTION PLANE WITH BLOWN UP VIEW NEAR PYLON

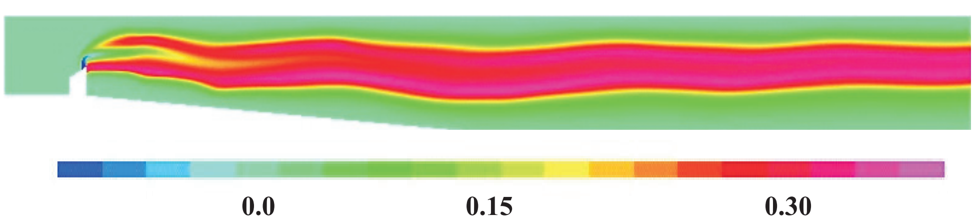


FIG. 7 WATER MASS FRACTION DISTRIBUTIONS IN THE INJECTION PLANE

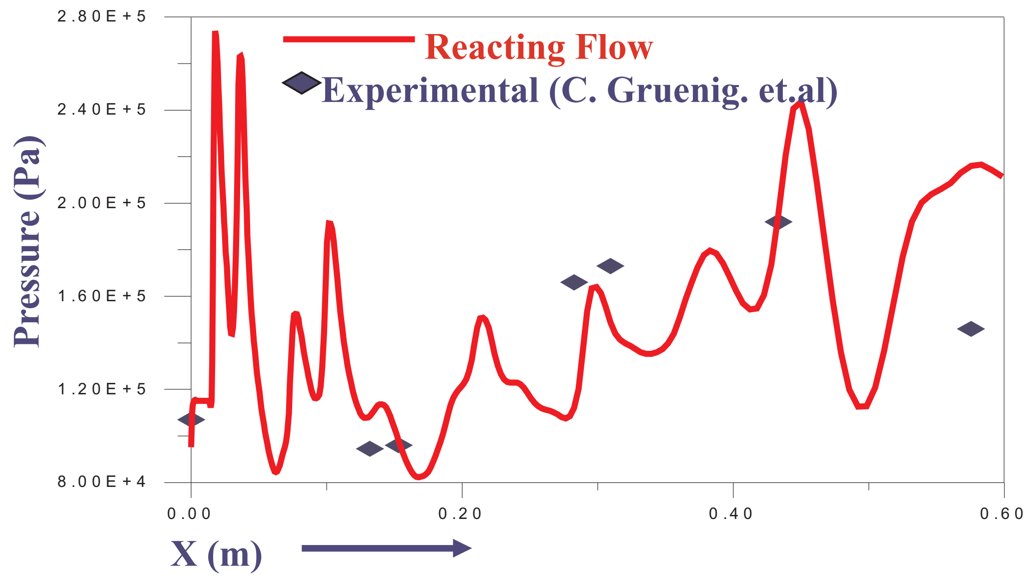


FIG. 8 COMPARISON OF AXIAL DISTRIBUTION OF SURFACE PRESSURE FOR PYLON BASED SCRAMJET COMBUSTOR

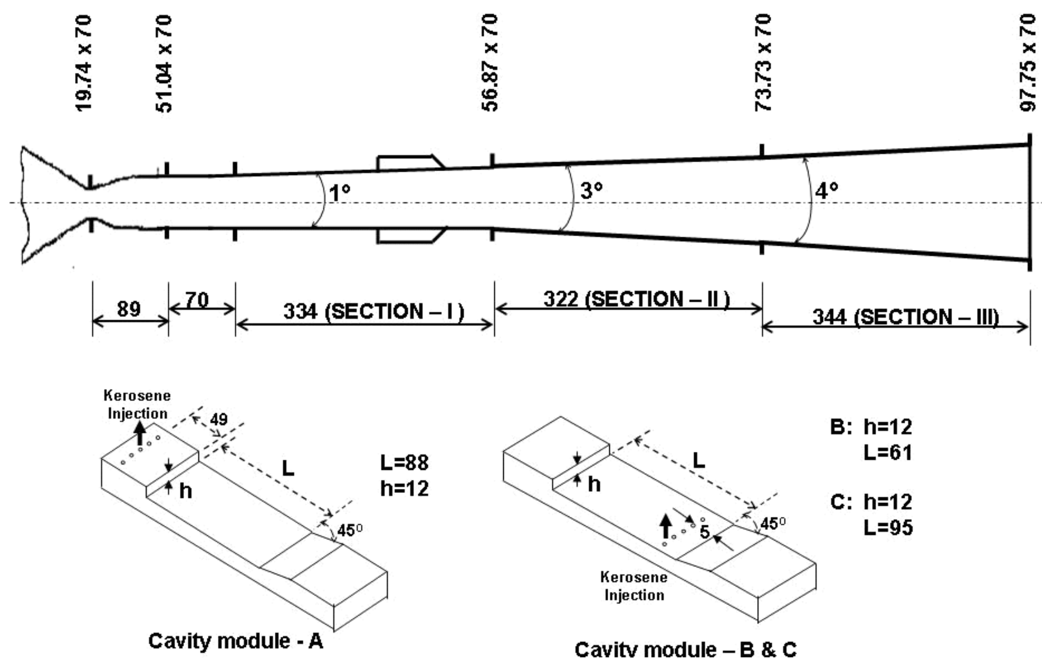


FIG. 9 CAVITY BASED SCRAMJET COMBUSTOR [16] FOR WHICH COMPUTATIONS ARE CARRIED OUT

different cavity configurations (cavity module A, B &C). The depth of the cavities is 12 mm whereas the lengths of the cavities are 88, 61 and 95 mm given the L/h ratio of 7.33, 5.08 and 7.92 respectively. Kerosene was injected normally to the vitiated air stream via five orifices of 0.6 mm diameter. For cavity module A, fuel is injected upstream of cavity at an equivalence ration of 0.45, while for the cavity B and C, the fuel injected at the floor of the cavity at the equivalence ratio of 0.45 and 0.78 respectively. The total temperature, total pressure and mach number of the vitiated air is 1840 K, 10.44 bar and 2.5 respectively. More detailed descriptions of the experiments are available in Ref. [13]

The cross sectional views of the mass fraction of CO_2 – the reaction product are presented in Figure 10 to depict the zone covered by reaction. Although, the kerosene is injected at 49 mm upstream of the cavity, the presence of CO_2 is seen at 136 mm upstream of the injection point as the fuel has diffused through the recirculation separation bubble. Although, CO_2 has covered complete width of the cross section, reaction is not very intensive. This is mainly because of relatively low equivalence ratio of 0.45. The computed side wall surface pressure for the three cavity configurations (A, B, C) are shown in Figure 11(a) to 11 (c). As mentioned earlier, the equivalence ratios of cavity A and cavity B configurations are 0.45, whereas the equivalence ratio for cavity C is 0.78. The surface pressure for the non-reacting case for the cavity configuration is also shown in Figure 11(a). The increase in static pressure starts much upstream of injection location showing significant upstream interaction due to heat release. The static pressure was seen to reach an approximately isobaric plateau in the nearly constant area section X = 250 mm and decrease continuously till the combustor exit because of flow expansion in the divergent section of the combustor. A good comparison with experiment and computational values are obtained except in the region of fuel injection where the computations have shown a higher value. The higher heat

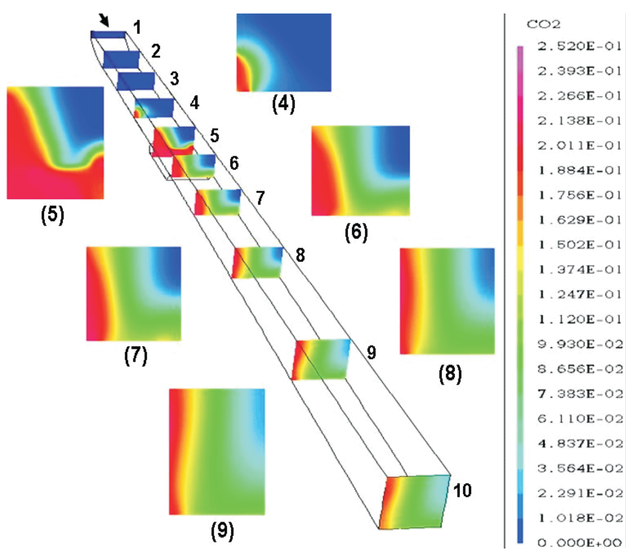


FIG. 10 CO_2 MASS FRACTION AT VARIOUS AXIAL STATIONS FROM FACILITY NOZZLE THROAT TO EXIT

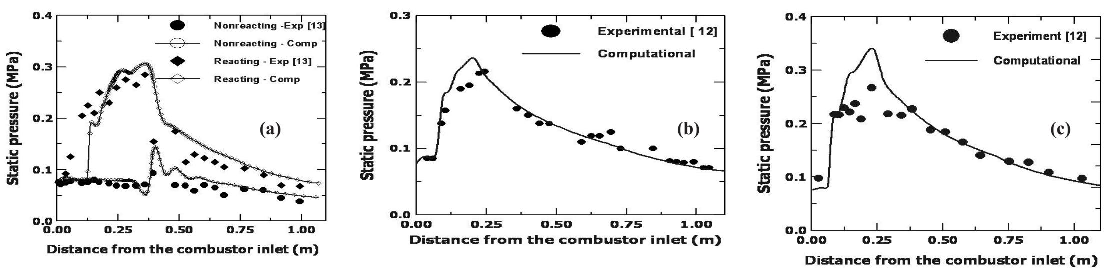


FIG. 11 SURFACE PRESSURES COMPARISON (a) CAVITY MODULE A (b) CAVITY MODULE B (c) CAVITY MODULE

release caused due to fast chemistry assumption in the simulation is conjectured to be the cause of higher surface pressure in the injection zone. In the divergent portion, the principle thrust producing element of the combustor, the agreement between the two is very good. The difference between the surface pressure for the non-reacting and reacting cases presented for the cavity module A in Fig.11(a) quantifies the effect of heat release in the surface pressure of the combustor. The effect of droplet diameter on the surface pressure was determined by carrying out the simulation with the different particle diameters of 1, 5, 10 and 20 mm for the cavity module A configuration with equivalence ratio 0.45. The computed surface pressures with the different droplet diameters are compared with experimental values in Figure 12. It can be observed that with lesser droplet diameter, the evaporation is faster and heat release is intensive. This has lead to surface pressure to rise near the injection zone. The higher heat release is also responsible for more upstream interaction for the lesser droplet diameter case. The effect of the droplet diameter on the surface pressure is insignificant in the divergent portion of the combustor.

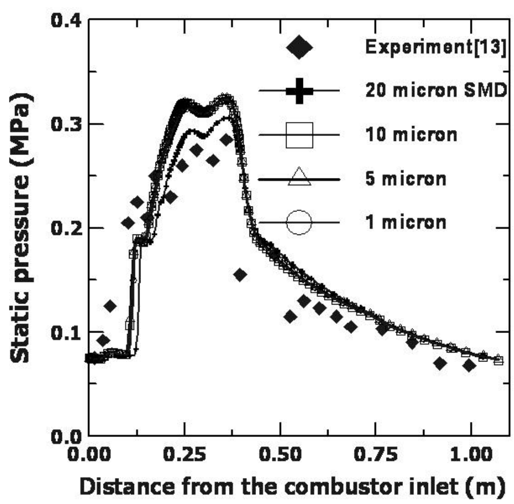


FIG. 12 EFFECT OF DROPLET DIAMETER ON THE SURFACE PRESSURE FOR CAVITY MODULE A

3.1.4 *Kerosene fueled scramjet combustor with ramp-cavity combustor*

The combustor configurations for which the computations [26] are carried out are taken from Ref. 30. The combustor configuration is presented in Figure 13. The combustor consists of three parts namely, the facility nozzle, the constant area section of length and the divergent section with divergence 3.2^0. Three and two distributed ramps of length are provided on the bottom and top walls respectively in the constant area portion of the combustor as shown in Figs 13(b) and (c). One cavity each of length to depth ratio of 7.25 is placed in both top and bottom wall at the end of the ramps for flame holding purpose. Kerosene is injected in the combustor through 10 numbers of injectors of 0.4mm diameter. The vitiated air from the burner accelerated through a two dimensional convergent-divergent nozzle of Mach 2.0 into the combustion chamber. The total temperature and total pressure of the vitiated air is 0.9 MPa and 1645 K respectively. The equivalence ratio is 0.21.

Mach number distribution for the reacting and non-reacting cases in the plane of symmetry is compared in Figure 14. The flow structure is different between the two cases. The terminal shock for the reacting case is positioned in the upstream location (compared to the nonreacting case) because of heat release due to reaction. The flow accelerates again in the divergent section of the combustor. The axial distribution of the computed non-dimensional surface pressure (P/P_0), (P_0 being the stagnation pressure) at the top surface of the plane of symmetry is compared with the experimental values in Figure 15. The axial length has been normalized with the height of the throat of the facility nozzle. A very good match has been obtained within experiment and numerical value except near the injection location, where the computation predicts higher value because of fast chemistry assumption.

3.2 Application of CFD Technique for Scramjet Combustor Design

The validation exercises in the previous section reveal that although predicted pressure rise is more in the fuel injection zone because of fast chemistry assumption, the computed pressure match

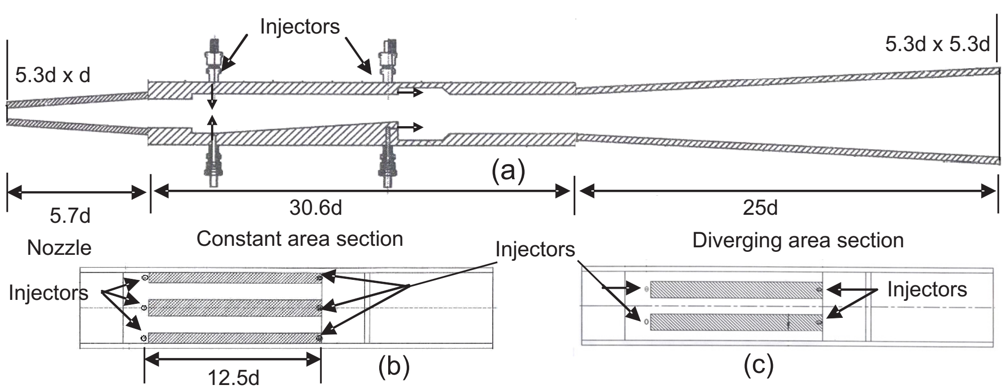


FIG. 13 RAMP CAVITY SCRAMJET COMBUSTOR CONFIGURATION: (A) FULL COMBUSTOR, (B) BOTTOM PLATE OF CONSTANT AREA SECTION, AND (C) TOP PLATE OF CONSTANT AREA SECTION

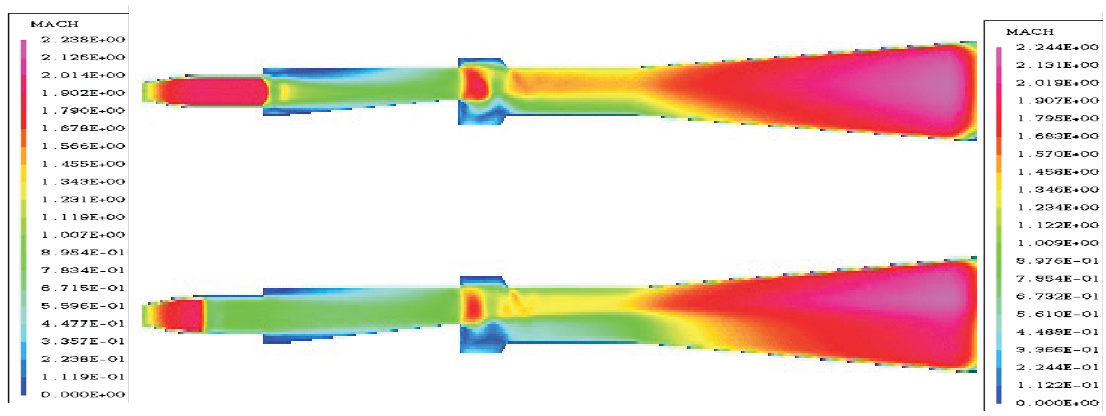


FIG. 14 MACH NUMBER DISTRIBUTION IN THE SYMMETRY PLANE: (A) NON-REACTING AND (B) REACTING

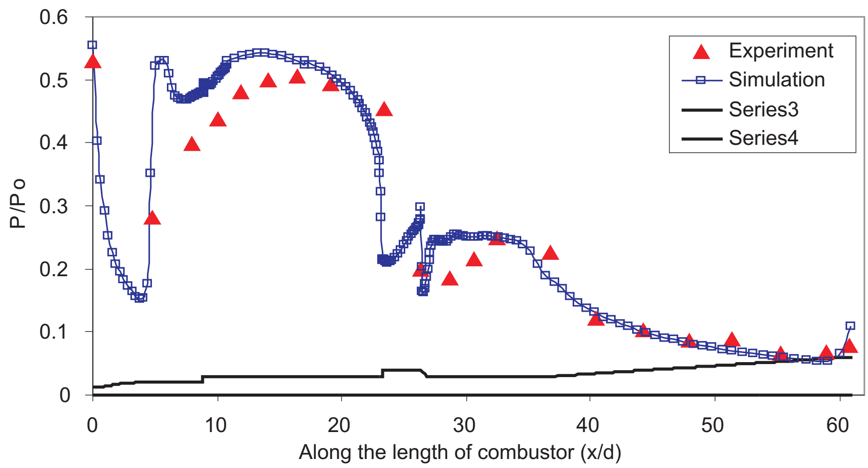


FIG. 15 COMPARISON OF SURFACE PRESSURE DISTRIBUTION AT SYMMETRY FOR REACTING CASE

reasonably well in the divergent portion of the combustor where the major portion of thrust is produced.

A number of reacting and nonreacting simulations [31] were carried out for half width and full width scramjet combustor for hypersonic airbreathing mission. A typical geometry of the combustor is shown in Figure 16. Kerosene fuel is injected from a V shaped strut with obtuse total angle. Mcquardth Corportion, USA investigated this type of struts to increase the three dimensionality of the flow field. Starting from an initial design, the performance of the combustor was improved progressively by relocating the struts and injection locations through analysis of various thermochemical parameters in the combustor obtained from the numerical simulation.

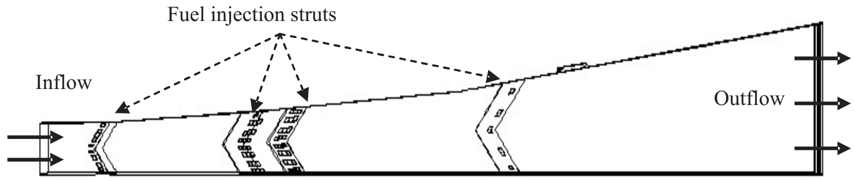


FIG. 16 TYPICAL GEOMETRY OF THE SCRAMJET COMBUSTOR

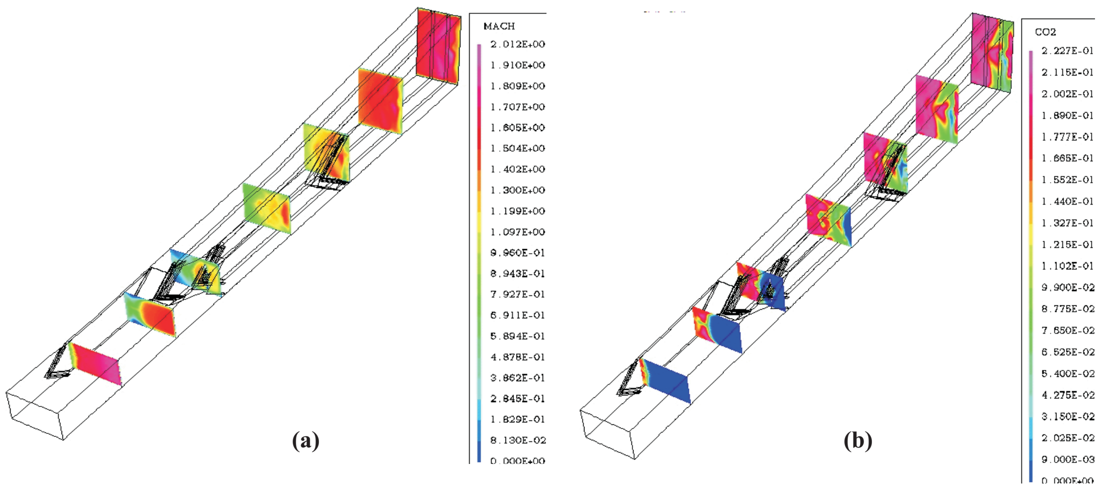


FIG. 17 FLOW PARAMETERS AT VARIOUS AXIAL LOCATION (a) MACH NUMBER (b) CO_2 MASS FRACTION

Typical Mach number and carbon dioxide distributions at different axial locations are presented in Figure 17. The combustor so designed was tested in DRDL test facility and the comparison of top wall surface pressure for reacting and non-reacting cases are presented in Figure 18. The simulation results for supersonic outflow boundary condition are also presented in the figure to show the point of flow separation in the combustor. Although a very good match is obtained for the nonreacting case, the computed pressure is lower than the experimental value in mid-region of the combustor. The achieved thrust from the full width combustor was found to be adequate for the mission.

4. CONCLUSIONS

Numerical simulations are carried out for the design and analysis of strut based kerosene fueled scramjet combustor. Commercial CFD software CFX-TASCflow is used to solve three dimensional Navier Strokes Equations along with K-ε turbulence model. Combustion is modeled by Eddy Dissipation Concept based on infinitely fast rate kinetics. The evaporation and mixing of liquid kerosene droplet is studied by employing a Lagrangian Dispersed Phase Analysis. Detailed validation of the software has been carried out by comparing reliable experimental results of

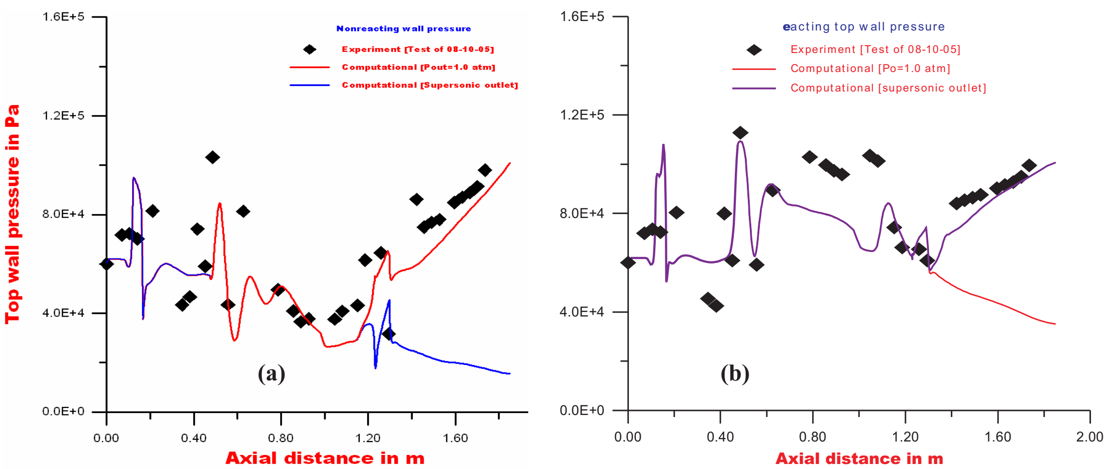


FIG. 18 COMPUTATIONAL AND EXPERIMENTAL WALL PRESSURE COMPARISON (a) NONREACTING (b) REACTING

various reacting flow cases pertaining to scramjet combustor flow field with both kerosene and hydrogen fuel with different fuel injection systems. It has been found that although predicted pressure rise is more in the fuel injection zone because of fast chemistry assumption, the computed pressure match reasonably well in the divergent portion of the combustor where the major portion of thrust is produced. This properly validated numerical tool is used for the design and analysis of flight sized scramjet combustor for an airbreathing mission. Computed flow parameters were analysed and design modificationa are carried out by relocating the struts and fuel injection system to achive better performance of the combustor in terms of thrust and combustion efficiency. CFD tools are playing a very significant role in developing the kerosene fuelled scramjet combustor for hypersonic airbreathing mission.

ACKNOWLEDGEMENTS

The work carried out by the scientists of the Computational Combustion Dynamics (CCD) Division of Directorate of Computational Dynamics (DOCD), DRDL, Hyderabad are compiled and presented here. Author is thankful to all the members of CCD Division of DOCD for their help in preparing this paper. The author expresses his sincere thanks to Dr. V. Ramanujachari, scientist DRDL for providing the experimental results for comparisons with CFD results.

REFERENCES

[1] A. Ferri (1964): "Review of problems in application of Supersonic Combustion" Journal of the Royal Aeronautical Society, Vol. 68, No. 645, pp 575-597.

[2] E.T. Curran (2001): "Scramjet Engines: The first forty years" Journal of propulsion and power, Vol. 17, No.6, pp 1138-1148.

[3] M. and Kenzakowski, D.C. (2001), "Historical survey on enhanced mixing in Scramjet engines", Jr. of Propulsion and Power, Vol. 17, No. 6, pp 1273-1286.

[4] Masuya, G., Komuro, T., Murakami A., Shinozaki, N., Nakamura, A., Murayama, M. and Ohwaki, K. (1995): "Ignition and combustion performance of Scramjet combustor with fuel injection struts", Journal of propulsion and power, Vol. 11, No. 2, pp 301-307.

[5] Mitani, T., Kanda T., Hiraiwa, T., Igarashi, Y. and Nakahashi, T. (1999): "Drags in Scramjet engine testing - Experimental and Computational Fluid Dynamics studies", Journal of propulsion and power, Vol. 15(4), pp 578-583.

[6] Mitani T., Chinzei, N., Kanda T. (2001): "Reaction and Mixing-Controlled Combustion in Scramjet engines", Journal of propulsion and power, Vol. 17(2), pp 308-313.

[7] Gerlinger. P., Kasal, P., Stoll, P. and Bruggemann D. (2001): "Experimental and theoretical investigation on 2D and 3D parallel Hydrogen/air mixing in a supersonic flow", ISABE paper No 2001-1019.

[8] Tomioka, S., Murakami, A., Kudo, K. and Mitani, T. (2001): "Combustion tests of a staged supersonic combustor with a strut", Journal of propulsion and power, Vol. 17(2), pp 293-300.

[9] Tomioka, S., Kobayashi, K., Kudo, K., Murakam A. and Mitani T. (2003): "Effects of injection configuration on performance of a stage supersonic combustor" Journal of propulsion and power, Vol. 19, No. 5, pp 876-884.

[10] Glawe, D.D., Saminy, M., Nejad A.S. and Cheng, T.H. (1995): "Effects of nozzle geometry on parallel injection from base of an extended strut into a supersonic flow, AIAA paper 95-0522.

[11] Wepler Ulrich & Kaschel W. (2002): "Numerical investigation of turbulent reacting flows in a Scramjet combustor Model" AIAA paper No. 2002-3572.

[12] Mathur, T., Gruber, M., Jackson, K., Donbar, J., Donaldson, W., Jackson, T., Billig, F., (2001): Supersonic combustion experiments with a cavity-based fuel injector. Journal of Propulsion and Power. 17 (6), 1305-1312.

[13] Yu, G., Li, J.G., Chang, X.Y., Chen, L.H., and Sung, C.J., "Fuel Injection and Flame Stabilization in Liquid – Kerosene Fueled Supersonic Combustor," *Journal of Propulsion and Power*, Vol.19, No.5, 2003, pp.885-893.

[14] Burnes, R., Parr, T.P., Wilson, K.J., Yu, K., 2000. Investigation of supersonic mixing control using cavities - Effects of fuel injection location. AIAA paper 2000-3618.

[15] Yu, K H., Wilson, K.J., Schadow, K.C., (2001): Effect of flame holding on supersonic – combustion performance.", Journal of Propulsion and Power. Vol. 17(6), pp 1287-1295.

[16] Hsu, K.Y., Carter, C., Crafton, J., Gruber, M., Donbar, J., Mathur, T., Schommer, D., Terry, W., 2000. Fuel distribution about a cavity flame-holder in supersonic flow. AIAA paper 2000-3585.

[17] Gruber, M.R., Donbar, J.M., Carter, C.D., Hsu, K.Y., (2004): "Mixing and combustion studies using cavity-based flame holders in a supersonic flow". Journal of Propulsion and Power., Vol. 20(5), pp 769-778.

[18] Yu, G., Li, J.G., Chang, X.Y., Chen, L.H., Sung, C.J., (2003): “Fuel injection and flame stabilization in liquid – kerosene fueled supersonic combustor”. Journal of Propulsion and Power, Vol. 19(5), pp 885-893.

[19] Vinogradov, V.A., Kobigsky, S.A. and Petrov, M.D. (1995): “Experimental investigation of kerosene fuel combustion in supersonic flow”, Journal. of Propulsion and Power, Vol. 11(1), pp 130 -134.

[20] Bouchez, M., Dufour, E., Montazel, X., 1998. Hydrocarbon fueled scramjet for hypersonic vehicles. AIAA paper 1998-1589.

[21] Dufour, E., Bouchez, M.(2001). “Computational analysis of a kerosene fueled scramjet” AIAA paper 2001-1817.

[22] P. Manna, Ramesh Behera, and Debasis Chakraborty **“**Design and Analysis of Liquid Fueled Strut Based Scramjet Combustor – A CFD Approach” *AIAA Journal of Propulsion and Power* Vol 24, No. 2. March – April, 2008, pp 274-281.

[23] CFX-TASCflow Computational Fluid Dynamics Software. Version 2.11.1. AEA Technology Engineering Software Ltd. 2001.

[24] Soumyajit Saha and Debasis Chakraborty “Reacting flow computation of staged supersonic combustor with strut injection”, AIAA Paper No. 2006-3895. 36th AIAA *Fluid Dynamics Conference, San Francisco, USA, June 2006*

[25] Afroz Javed and Debasis Chakraborty **:** “Numerical simulation of supersonic combustion of pylon injected hydrogen fuel in scramjet combustor” *Journal of Institute of Engineers* Vol. 87, May, 2006, pp 1-6.

[26] P. Manna, Ramesh Behera and Debasis Chakraborty: “Thermochemical exploration of a cavity based supersonic combustor with liquid kerosene fuel” *Journal of “Aerospace Sciences and Technologies”* Vol. 59, No. 4 pp 246-258, Nov, 2007.

[27] Ramesh Behera and Debasis Chakraborty: “Numerical simulation of Kerosene Fueled Ramp Cavity Based Scramjet Combustor” *Journal of Aerospace Sciences and Technologies,* Vol. 58, No. 2, May 2006, pp 104-111.

[28] Gruenig C., Avrashkov, V., and Mayinger, F. (2000): “Fuel injection into a supersonic airflow by means of pylons”, Journal of propulsion and power, Vol. 16, No. 1 pp 29-34.

[29] Gurenig C., Avrashkov, V. and Mayinger, F. (2000): “Self-ignition and supersonic reaction of pylon injected hydrogen fuel”, Journal of propulsion and power, Vol. 16, No. 1, pp 35-41.

[30] Satish Kumar, Charyulu, B.V.N., Moorthy, J.V.S. and Chandrasekhar, C. (2005), “Scramjet Combustor development”, Newsletter, The Combustion Institution (Indian Section), Vol. 25, No. 4, April 2005, pp 16-31.

[31] Debasis Chakraborty: “Role of CFD in Aero-propulsive Characterization of missiles in DRDL”, Proceedings of SAROD-2007 held at Thiruvananthapuram November 22-23, 2007, pp 23-40.

MINING, METALLURGICAL AND MATERIALS ENGINEERING

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

PERSPECTIVES IN ARMOUR MATERIALS AND DESIGNS

T. BALAKRISHNA BHAT[1], V. MADHU[2] AND N.K. GUPTA[3]

ABSTRACT

Increased protection leads to multiplied psychological stamina for moving free of fear in the battlefield and to a shift in the internal paradigm of the soldier. Increased protection is possible only through use of advanced armour technology. This paper describes the advances in metallic, non-metallic and composite armour materials and concepts for protection. It also describes newer approaches to design of protective systems through reactive armour, bulging armour, electric armour and energy beams. An effective armour for man is his family. The family is an unparalleled armour against all forms of attack like hatred, anger, depression, addictive, jealousy, loss of faith, divorce, violence, ill health, and all forms of physical, mental and emotional attacks. With this armour firmly around, the state of happiness and composure expands many fold. Yet, the most effective active armour for any human being is himself and the infinite consciousness around him. A few comments on total defence are also made.

Keywords: Metallic armour, Ceramic armour, Bulging armours, Electric armour, Dynamic armour, Energetic armour, intelligent armour, Total protection.

1. INTRODUCTION

Life is precious. It is also being increasingly recognized as precious. For example, the second world war ended only after crores of people sacrificed their lives. But Vietnam war stopped when a few lakhs of people lost their lives. Afghanistan war, IndoPak war and the Iraq war subsided when casualities ran into thousands. Clearly, this trend indicates that even in war, large number of casualities are no longer acceptable. No longer can people be treated as cannon fodder. Protection has thus become paramount in defence.

[1]INAE Distinguished Visiting Professor, IIT/Delhi, Defence Metallurgical Research Laboratory, Kanchanbagh, Hyderabad – 500 058
[2]Defence Metallurgical Research Laboratory, Kanchanbagh, Hyderabad – 500 058
[3]Department of Applied Mechanics, IIT/Delhi, New Delhi – 110016

In the words of General Shergill (1997), “Those days have gone when human life had scant respect and men were considered as cannon fodder. Today, casualities are a cause of alarm and human life can no longer be taken for granted and no longer the casualities will be accepted for want of planning. Human factors such as crew comfort, crew fatigue and survivability assume utmost importance. We do not wish to witness lives of crews being squandered and at the very worst see crews abandoning their tasks because of lack of survivability. This has happened in the past in many a country and that country has naturally paid its price in terms of humiliation and may be defeat. Please do not allow this to happen”. Human beings and the life in general are also faced with other challenges from nature and those of our own making. Deaths and damage in these far exceed the toll in war. Some thoughts on armouring against these attacks are also very important for the mation’s survival and prosperity.

Because armour technology is one of the most classified areas of research no country, which develops the protection systems, reveals its armour technology. As a result, it becomes imperative to have one’s own indigenous integrated armour research activity to keep pace with the developments of anti-armour threats around the world. Also today, no single material is capable of effectively defeating all the wide range of threats and hence a wide variety of armour materials and systems need to be developed to suit various applications.

A clever way for such a defence is through camouflage. Here one makes oneself practically invisible, by reducing the accoustic, magnetic, infrared, visible and radar signatures. Shape and size of the equipment, choice of materials of construction and application of well designed coatings play a crucial role in successful camouflaging. If done well camouflage is a wonderful means of protection. However, this aspect will not be discussed in the present article.

To provide protection by directly resisting attacks, one needs a solid armour. Such an armour is sometimes called passive armour. This is a misnomer because it is not really passive; instead it reacts vigorously and absorbs the energy often breaking and deforming the attacking projectile within a few microseconds. Design of armours which enables one to directly withstand the attack depends on the type of threat and severity of threat. These armours are usually made of high strength materials such as steel, in the strength range of 1000 to 2000 MPa, aluminium alloys in the strength range of 300 to 500 MPa, titanium alloys in the strength range of 800 to 1400 MPa, high strength polymeric materials such as aramids, polyethylene, and hard ceramic materials such as alumina, boron carbide, silicon carbide, and tough composites made of glass fiber reinforced composites. Because on impact, the materials undergo deformation or fracture at very high strain rates, the design principles involved in optimizing them are quite different from those involved in optimizing typical engineering materials. This is the challenging area of materials research related to armour materials.

Protection can also be achieved through reactive armour. In this, the armour reacts, often exploding locally, when hit upon by a missile or a projectile. This dissipates or damages the jet coming out of the missile, or in the case of projectiles, it breaks the projectiles, thereby making the attack ineffective. By mounting pressure and velocity sensors and decision chips in the armour modules, we can make this armour intelligent and develop intelligent dynamic armours which respond only when the threat is serious enough. Even better is active armour. In this, one neutralizes the threat midway even before it reaches the target. One way of doing this is to disturb the sensors and guidance system of the attacking missile with the help of properly designed screens or jamming devices. Another way is to detect the incoming threat and direct a burst of counter attacks so as to neutralize it. This can also be achieved more elegantly by electromagnetic armour which causes a yaw to the attacking projectile. High intensity electromagnetic beams can also be used to break up the projectile. This is another possible manifestation of the exciting electromagnetic armour.

In the recent past there have been major advances in the area of armour materials, concepts and designs some aspects of which are described here. Also some thoughts relevant to armouring ourselves against various natural and man made causes are shared.

2. MECHANISMS OF FAILURE

The efficacy of any material or design depends on the mechanisms of failure of the armour and also on the mechanisms of damage they inflict on the attacking projectile. An effective mechanism for many types of KE threats is to shatter or deform the carrier of the KE and make the shot absorb the KE on itself or to make it lose its integrity. For example, to stop a 7.62mm armour piercing projectile, we may need armour steel weighing about 130 kg/m^2. However, if the projectile is made to shatter and deform and absorb energy for its own deformation, through the use of a facing high hardness steel or ceramics, then the weight to stop the same projectile could be brought down to about 50 kg/m^2. Deflecting the shot by adjusting the armour parameters is another method especially effective when used in combination with the method of shot degradation. As much as 30-40% weight can be saved by this approach. Absorbing the energy is the conventional safe method and it needs maximization of the flow stress, volume of deformation, strain in the deformed zone and strain to fracture. This in turn is related to optimization of a variety of metallurgical or structural parameters and to delay the operation of failure of the armour and to induce severe failure processes in the projectile, which is the essence of the science of armour materials.

Various mechanisms of failure of the target during high velocity impact events are shown in Figure 1.

A well designed tough metallic armour deforms and fails in the form of petals at the front (Fig. 1a) and ductile bulge at the back (Fig.1b) When the toughness is a bit lower than required, it begins to scab (Fig. 1c). When the toughness is still lower, plugging and cracking becomes visible (Fig. 1d). Delamination is another mode of failure which can occur even in homogenous materials loaded with shock waves (Fig. 1e). Adiabatic shear band formation is a particularly damaging microscopic failure mode in high strength metallic materials (Fig. 1f). Fiber pull out, fiber fracture and delamination are the common modes in fiber reinforced materials (Fig. 1g). Tougher fibres and bonded fabrics fail by stretching and breaking (Fig. 1h). Ceramics fail by fracture into microscopic particles close to the impact (Fig. 1i) and into larger fragments in surrounding regions (Fig. 1j). By increasing the toughness of the ceramic, it is possible to minimize the area of damage to regions close to the point of impact. A special mode of failure (bulging of plates) occurs in the case of a sandwich plate consisting of a polymeric layer in between two metallic plates when impacted by high velocity jets coming from a hollow charge war head. Here, the polymeric material vapourises to cause the metal plates to bulge as shown in the (Fig. 1k). The bulging plates, as they move across the path of the jet, consumes it in a highly efficient manner.

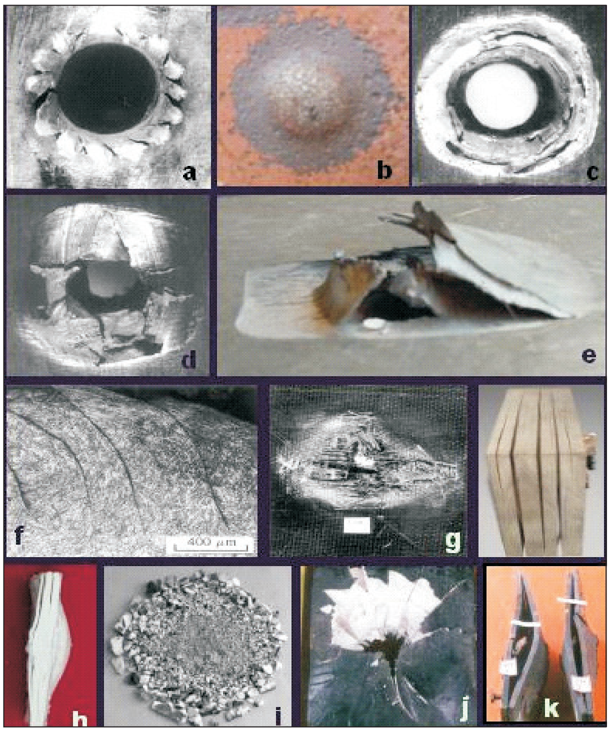


FIG. 1 DAMAGE MECHANISMS IN TARGET DURING HIGH VELOCITY IMPACT

In metals, the strength of material, and its toughness could be improved by refining the grain size, by micro-alloying and by heat treatment. Similarly, in composites, performance could be improved by the use of high performance fibers having high specific strength and modulus, high fiber content in the composite, higher strain to failure and tailored weave pattern. Figure 2 shows energy absorbed by various target materials against 7.62 armour piercing small arms projectile as a function of the target thickness [Madhu, Sivakumar and Balakrishna Bhat (1999)]. From the figure it is seen that harder materials like ceramics and steel need much lower thickness than softer materials like Aluminum. It also shows that some materials exhibit a threshold thickness below which, little energy is absorbed while some others show a non-linear variation of energy absorption with thickness.

Even as armour undergoes changes during impact, often projectiles too suffer severe damage. For example, hard projectiles when hitting a harder ceramic may partly break up, as shown in Figure 3a, into micron sized particles. Under less severe conditions, it may develop adiabatic shear band or break up into larger size pieces. In a laminated or layered structure shots can be made to turn (Fig. 3b). Shot may also turn and ricochet when we place the armour plate at an appropriate angle (Fig. 3b). Softer projectiles deform into a splash or mushroom as shown in the Fig. 3c. Longer rods break up into a large number of fragments (Fig. 3d). As in armour, formation of adiabatic shear bands is an important microscopic failure mechanism in the projectiles too (Fig. 3e).

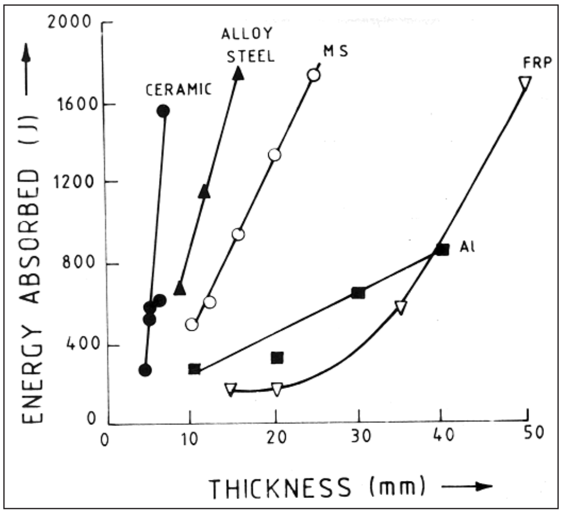


FIG. 2 ENERGY ABSORPTION BY A FEW MATERIALS AS A FUNCTION OF THEIR THICKNESS

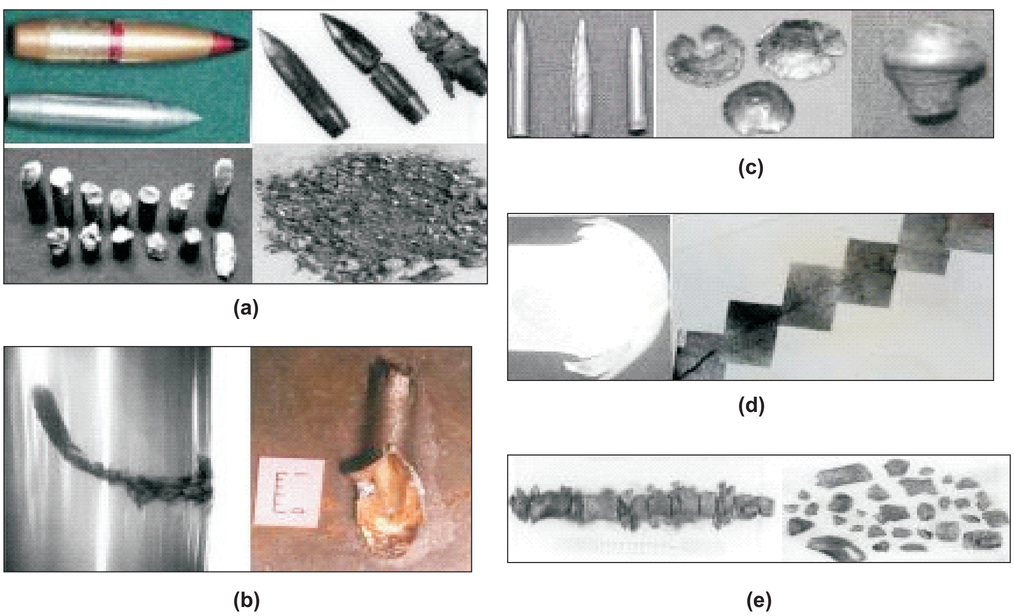


FIG. 3 TYPICAL DEFORMATION AND FRACTURE MODES OF VARIOUS TYPES OF PROJECTILES WHEN IMPACTED ON ARMOUR MATERIALS

3. ARMOUR MATERIALS

3.1 Armour Steels

Metallic armour materials are used in battle tanks, ICVs, light combat vehicles, helicopters and even as body armour. Excepting selective use of ceramics or composites in some locations for providing protection against HEAT and for damaging the shape of kinetic energy penetrators, on the whole more than 90% of the armour in the world is steel. This is largely due to its low cost, and wonderful ability to be heat-treated to a wide range of excellent properties.

The steel used for structure of battle vehicles consists of 30% of total weight of armour steel. Any improvement in ballistic properties of these steels will reduce the total weight of the tank significantly. Increasing the hardness of the steel is one direct approach Figure 4 shows the trend in improvement in ballistic limit of a 6mm thick armour steel plate when subjected to impact of 7.62mm ball ammunition. At present the hardness of steel being used for the structural purpose is around 320 VHN. Increasing hardness to very high levels may also pose problems related to welding or cracking on impact. It is expected that the problem may not arise at medium hardness levels of about 400 VHN. Some high hardness steels having more than 550 BHN have shown a higher performance and can be used as components of armour modules, fitted to the tanks or can be used as add-on armour on ICVs. Such high hardness steels are therefore going to play an important role in the design of fighting vehicles for giving required protection against the kinetic energy threats with improved performance. However, inhomogeneous deformation mechanisms limit the advantage beyond certain hardness values.

3.2 Aluminium Armour

Aluminium alloys, with excellent properties, such as high strength to weight ratio, fatigue and fracture resistance, corrosion resistance, sub-zero mechanical properties, weldability,

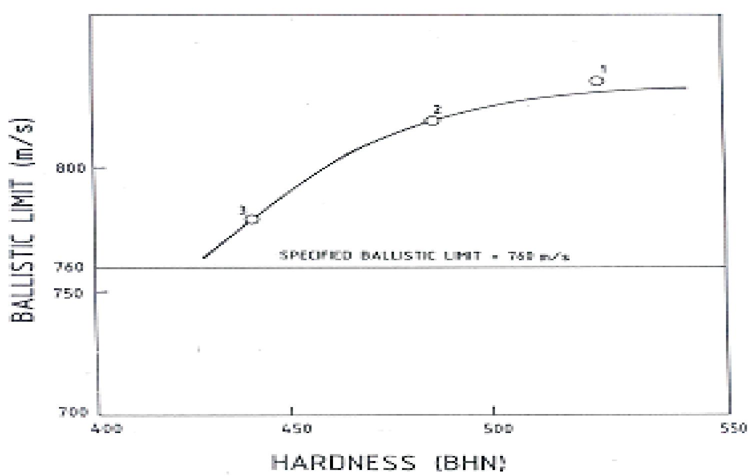


FIG. 4 VARIATION OF BALLISTIC LIMIT WITH HARDNESS

machinability and the ability to be formed into intricate shapes, are attractive materials for armour applications. Alloy designated as 5083 shows about 16% weight saving when compared to steel against 155mm artillery shell fragments. But against 0.30AP and 14.5mm AP the Aluminium alloy is inferior. The areal density of 7039 is lower, particularly when subjected to attacks by AP bullets. It is also superior to steel except at intermediate angles of obliquity of 30-50 degrees. It is noteworthy that the relative improvement shown by the 7039 alloy over 5083 and steel increases as the caliber of the attacking projectile increases.

Studies on the ballistic behaviour of Al-7017 against low caliber projectiles at room temperature and at cyrogenic temperatures at DMRL have shown interesting resuslts. The strength of the alloy (Figure 5) increased at cryogenic temperature (–120°C). Also a significant improvement in ballistic resistance and absence of adiabatic shear bands has been observed (Figure 6) at cryogenic temperature, which gives new direction of research on high strength alloys [Balakrishna Bhat, Madhu and Pappu (2005)].

3.3 Titanium Armour

Advantages of titanium arise from its high strength-to-weight ratio, excellent corrosion resistance and excellent ballistic performance compared with steel and aluminium. Ti alloys are readily fabricated in existing production facilities and are easily recycled. One disadvantage with Titanium is adiabatic shear band formation, which may result in fragmentation or spalling. Another disadvantage of titanium is its high cost.

Although titanium alloys have been used successfully in aircraft for many years, the high cost of titanium coupled with the sparse information on its ballistic properties have prevented widespread

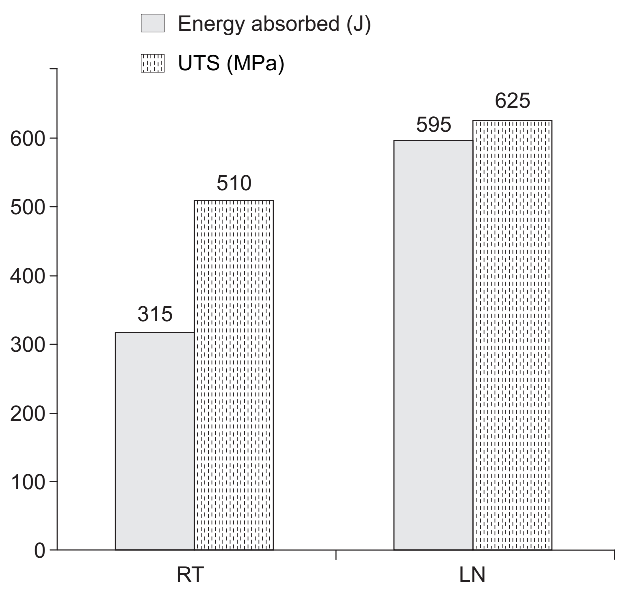


FIG. 5 ENERGY ABSORBED AND STRENGTH OF AL-7017 AT ROOM TEMPERATURE AND LIQUID NITROGEN TEMPERATURE.

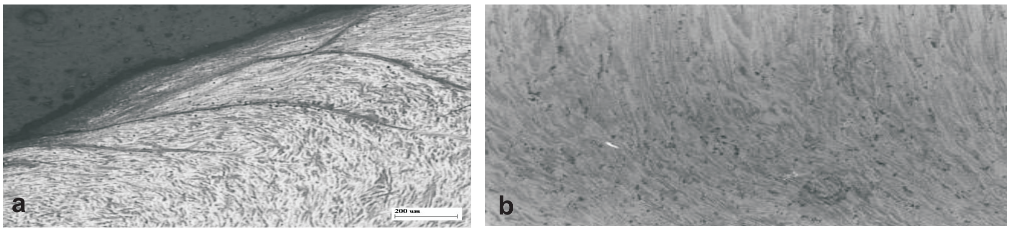


FIG. 6 MICROSTRUCTURE OF BALLISTICALLY IMPACTED PLATE OF AL-7017 (a) AT RT SHOWING ADIABATIC SHEAR BANDS (b) AT LNT SHOWING NO BANDS.

use of titanium in ground vehicles. They are two military specifications applicable for Ti-6Al-4V alloy: MIL-A-46077 and MIL-T-9046. Ballistic testing indicated that reductions in interstitial elements such as carbon, oxygen, nitrogen, and hydrogen improved the ductility and, hence, the spall resistance and ballistic protection of the plate. Consequently, the MIL-A-46077 armor specification was developed for ELI grade Ti-6AI-4V.

The need to reduce the possibility of injury to law enforcement and military personnel has generated considerable interest in enhancing the protection level provided by vests, helmets, shields, vehicles and other armors while minimizing weight. Flexible armor materials (including aramids, nylons, and other polymeric materials) typically provide very efficient protection against most small arms, particularly lead-core ammunition. However, many types of ball ammunition contain steel components. Although most of this ball ammunition is not intentionally “armor piercing”, the steel components tend to enhance the ability of the round to penetrate flexible body armor. It is against these types of threats that combinations of titanium with flexible armor can be especially attractive.

To reduce the weight of the battle tanks, a number of components on the tanks that are currently made from RHA steel are being replaced with titanium components such as turret blow-off panels, gunner’s primary sight cover, NBC cover, engine top deck, turret pivot rack, commander’s Hatch, commander’s independent thermal viewer (CTTV) cap, which would result in weight savings of more than 30% in these areas.

The ballistic efficiency of the titanium armour against 7.62 AP and 120 FSAPDS has been assessed at DMRL and are given in Table 1.

Performance of titanium armour also depends on its heat treatment, though not as strongly as in the case of steel. Some results are shown in Figure 7.

Table 1. *Mass efficiency of Ti-alloy armour*

Material	*Mass efficiency (Em)*	
	7.62AP	*120FSAPDS*
RHA (300 BHN)	1.0	1.0
Ti-6Al-4V	1.5	1.5

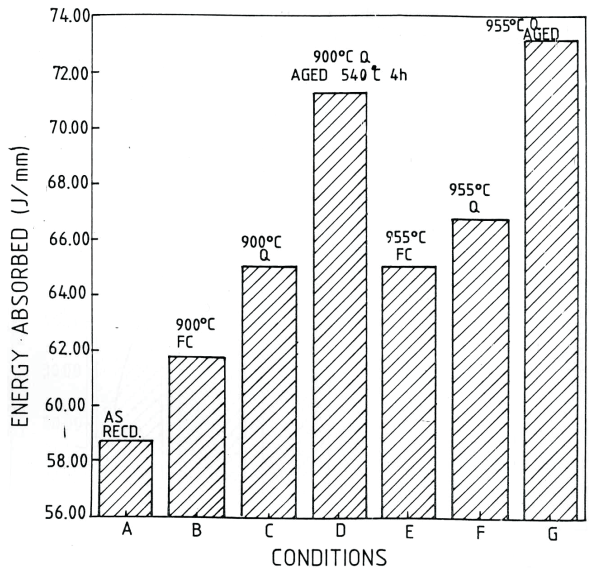


FIG. 7 ENERGY ABSORPTION IN TI-6AL-4V AS A FUNCTION OF HEAT TREATMENT

Thus, our own results confirm the attractiveness of titanium armour against non- deformable small caliber shots as well as deformable large caliber penetrators. As a result of this work Ti-6A1-4V grade material has been recently developed as an armour for ICV. This armour proved itself as highly weight efficient armour in comparison to hard steel or other material combinations against a combination of threats. The first Titanium armour plate used in India is shown in Fig. 8.

3.4 Ceramics

Ballistic performance of ceramics depends on a number of properties. These properties include density and porosity, hardness, fracture toughness, Young's modulus, sonic velocity and compressive strength.

Considering dense armor ceramics, such as alumina ceramics, one can only say that porosity should be minimized (zero or practically zero; i.e., water absorption should not exceed 0.02%). Density should be low. However, obtaining a decrease in density by underfiring (i.e., by porosity increase) is unacceptable, because it results in the deterioration of other properties important for ballistic performance. Hardness of ceramics should be high; it should at least be higher than the projectile hardness. Because hardness data in the literature are performed using measurements under various conditions (using varied techniques and under various loads), comparative evaluation of ceramics is sometimes difficult. To achieve consistent ballistic test results, hardness for alumina ceramics should exceed HV 1220–1250. For comparison, a standard NATO 7.62 mm armor-piercing steel core has a hardness value of HV 800–870, and a 14.5 mm armor-piercing tungsten carbide core used by machine guns has a hardness value of HV1150. The term EH/ñ is sometimes used as an indicator of the ballistic figure of merit for ceramic. In this term 'E' is the

FIG. 8 TITANIUM ARMOUR PLATE FOR USE IN INFANTRY COMBAT VEHICLES

Table 2. *Comparison of ceramic armour materials*

Materials	*Physical Properties*				
	Density(ρ) g/cm^3	*Hardness(H) kg/mm^2*	*Strength(MOR at 20°C)MPa*	*Stiffness (E) GPa*	*Ballistic merit EH/rkg/S^2 $\times 10^{15}$*
Hot Pressed B_4C	2.5	3000	300	400	480
Sintered Al_2O_3	3.8	1600	350	340	143
Hot Pressed TiB_2	4.5	3300	350	570	418
Al_2O_3 + 30 wt% TiB_2	3.99	2059	–	402	207

Elastic modulus 'H' is the hardness and 'ρ' is the density of the material. Properties of a few ceramic materials are given in Table 2.

Also some balance between levels of hardness and fracture toughness must be maintained. For instance, some carbide based ceramics with elevated *Kic* and metal-infiltrated ceramics, such as Lanxide™ SiC/Al, which have high fracture toughness values, exhibit a high level of ballistic performance. As a result, many of these materials have been successfully used as armor materials. Sonic velocity (which indicates the ability of hard ceramics to dissipate energy from the impact area) should be high. High sonic velocity also indirectly indicate the achievement of a high level of densification of ceramics and low closed porosity.

3.5 Titanium Diboride

Properties of TiB_2 are listed in Table 3. Despite many attractive properties and potential applications, like many other covalently bonded materials, the production of high density TiB_2 is very difficult because the mass transport for densification is restricted. In addition, the thin oxygen rich layer existing on the surface is known to be very detrimental to densification. Therefore various strategies such as hot pressing are to be adopted to consolidate TiB_2 based materials.

Table 3. *Properties of titanium di-boride*

Property	*Value*
Density, g/cc	4.5
Crystal structure	Hexagonal
Melting point, °C	3225
Microhardness, Kg/mm^2	3300
Thermal conductivity, W/m.K	65 – 122
Young's modulus, GPa	574
Flexural strength, MPa	756
Fracture toughness, MPa/m	6.7
Electrical resistivity, Ωm	9 – 15

TiB_2 is a potential candidate for protection against heavy threats due to its high modulus, high hardness and moderate density. Ballistic figure of merit for TiB_2 and Al_2O_3 + TiB_2 composite have been compared with those of the conventionally used Al_2O_3 and boron carbide in Table 2.

This clearly shows the advantage of using B_4C and TiB_2 for armor application. Al_2O_3 + TiB_2 composite shows better figure of merit compared to pure Al_2O_3. With this background, experimental work was carried out to make Al_2O_3 + TiB_2 composite by self propagating high temperature synthesis. Though the composite was heavily cracked due to thermal stresses, Al_2O_3 + TiB_2 composite showed a performance comparable with Al_2O_3 [Subramaniam (2006)].

3.6 Boron Carbide

Boron carbide is an attractive ceramic armour material for protection against small arms projectiles. With its low density, high hardness and high compressive strength, B_4C has already found extensive applications. It is reported to be currently used extensively in body armours as well as to a limited extent on infantry combat vehicles and helicopters.

Boron carbide in hot-pressed form is a lightweight armour material. Tiles of B_4C for personnel vests are hot pressed in large hot-press furnaces. Sintered B_4C with metallic additives has been studied for years but no suitable armour material has evolved from these studies. Manufacturers also use a sinter-HIP process to make complicated shapes of B_4C like in the case of the helmets.

Boron carbide is produced by reacting and fusing a mixture of boric oxide and carbon in an electric arc furnace [Riedel (2000)]. The product essentially consists of partially fused slush having the composition near that of B_4C. The liquid in contact with solid boron carbide does not have the same composition as the solid. The material left in the molten state becomes increasingly richer in carbon than in solid. When the last liquid solidifies its composition consists of eutectic of mixture of boron carbide and graphite. Hence, in practice, all arc furnace products have free graphite. This mixture is then sent through hammer mills and jet mills to reduce the particle size to about 1.0 to 1.5 μm. The typical run time is about one hour for 8-10 kg powder. The powder is then acid washed and sedimentation is carried out. It is then dried in vacuum.

Generally boron carbide is hot pressed and in a few cases sintered followed by HIP. Typical hot pressing conditions of powder (< 10 μm) are 2100°C at 35 MPa for 30 minutes. Typical sintering and HIP of powder (< 1 μm) consists of sintering at 2200 to 2250°C for 30 minutes at low pressure (~ 10 Pa) followed by HIP at 2000°C at 200 MPa in argon for 2 hours. Optimum conditions depend on many parameters such as particle size, impurities and additives. Spark Plasma pressing is an emerging attractive route for making boron carbide.

3.7 Alumina Ceramic

Because of its low cost and easy availability alumina is the most widely used ceramic for all armours. Its hardness, 1600 kg/mm^2, exceeds the hardness of all typical projectile materials. The way it breaks up an armour projectile into fine dust is shown in Fig. 3a With refinement of grain size and additions of zirconia, the fracture toughness of the alumina has been steadily improving to make alumina, the preferred choice of designers of armour systems.

Response of ceramics to projectile impact has been the subject of many investigations in the past two decades. These studies have described various aspects of ceramic fracture both in compression and tension and have made their ballistic performance assessments. Evidence of ceramic comminution during projectile penetration has been provided by [Wlkins (1978)] and [James (1998)]. Studies have also shown that when backing is sufficiently thick, this phenomenon is essentially due to induced stress levels exceeding the compressive strength of the material.

Evaluation of ballistic performance of ceramics has been a difficult task due to the number of variables like the type of threat, projectile velocity, projectile geometry, nature of ceramics, target configuration in terms of front and backing material and their thicknesses, angle of impact and support conditions affecting the results. However, the depth of penetration (DOP) method has been found to be quite satisfactory for the purpose [Rosenberg and Yeshurun (1988)]]. An attempt has been made to study the ballistic efficiency of two grades of alumina ceramics, backed by thick metal plates, against impact of 12.7mm armour piercing projectiles [Vemuri, Ramanjaneyulu, Balakrishna Bhat and Gupta (2005)]. The thickness of the ceramic tile and the impact velocity of projectile were varied and the influence thereof on the ballistic efficiency studied. Comparative studies with 7.62mm AP projectiles were also made keeping the diameter to thickness ratio identical. The ceramic layer consisted of nine tiles of size of 50mm x 50mm each bonded close together over aluminum armour backing plate without any other additional confinement. This rather small size was chosen in order to provide multi hit capability, which is a common requirement in modern armour.

In the ballistic tests shown schematically in Figure 9, the alumina ceramic tiles backed by metal plate targets were impacted by the 12.7mm armour piercing projectiles and the depth of penetration was measured in each case.

Ballistic performance is assessed by a dimensionless factor, which combines the mass efficiency and thickness efficiency of the material as defined in Eqs (1) to (3) [Gooch and Burkins (2001)].

At different velocities, the ballistic performance of ceramic is calculated by comparing the depth of penetration [P_{REF}] of the projectile into a semi-infinite thick aluminium alloy (Al-7017) target plate to the residual penetration [P_{RES}] in the plate when backing against ceramic armour tiles of

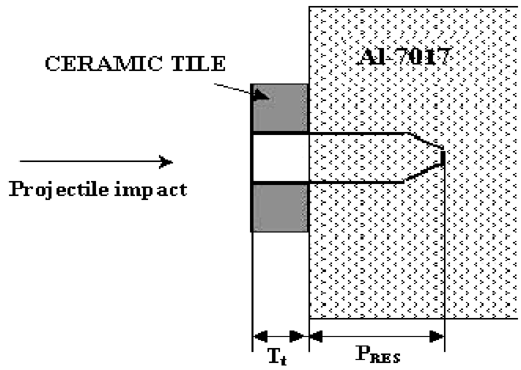


FIG. 9 TARGET CONFIGURATION IN THE DOP TEST

various thicknesses. Ballistic performance of the ceramic is calculated in terms of the thickness efficiency [E_t], the mass efficiency [E_m] and the ballistic efficiency factor q^2 as given below:

$$E_m = E_t \cdot \frac{\rho_{REF}}{\rho_t} \tag{1}$$

$$E_t = \frac{P_{REF} - P_{RES}}{T_t} \tag{2}$$

$$q^2 = E_m \times E_t \tag{3}$$

The ballistic efficiency factor has a significance for armour designers as this factor connects both the mass and the thickness efficiency factors.

Alumina ceramic tiles of two grades backed by plates of 7017 aluminium alloy, a commonly used grade of armour in the light armoured vehicles, were used in the present experiments. The mechanical properties of the above alloy plates are given in Table 4. Physical and mechanical properties of the two grades of alumina ceramics, used in the current experiments, are given in Table 5. The ceramic tiles used were 50mm × 50mm in size with thicknesses of 10mm, 12mm and 14mm. The size of the ceramic tiles have been chosen in such a way as to provide multi-hit protection, which is a common requirement in any armouring. These tiles were bonded to a 40mm thick 7017 aluminium alloy backing plate of size 200mm × 200mm using epoxy resin as shown in Figure 10. Only one shot was fired close to the center of each of the panels.

For every thickness of the ceramic tile bonded to aluminium alloy plate, projectiles were fired at velocities ranging from 500 ms^{-1} to 830 ms^{-1} at intervals of about 100 ms^{-1}. The residual depth of penetration in the backing plate was measured by dipping digital dial calipers in to the craters to an accuracy of ± 0.05 mm. The penetration of projectile obtained in the reference backing plate for each velocity was also measured. In this case, two 40mm thick aluminium alloy plates were bonded together and projectiles were fired on the combined target to measure the reference penetration. Using the data of residual penetration and the reference penetration calculations of the thickness efficiency (E_t), mass efficiency (E_m) and the ballistic efficiency (q^2), as defined in Eqs. 1 to 3, were made.

Table 4. *Mechanical properties of aluminium alloy A1-7017*

Density (g/cc)	*Yeild Strength (MPa)*	*UTS (MPa)*	*Elongation*	*Hardness (VPN)*
2.9	420-435	480-490	10%	135

Table 5. *Properties of the two grades Alumina ceramic*

Property	*Unit*	*D95 ceramic*	*C99.5 ceramic*
Density	Gm/cc	3.68	3.85
Hardness	HV	1200	1510

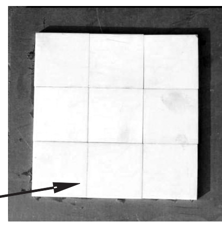


FIG. 10 PHOTOGRAPH OF THE TARGET SHOWING CERAMIC TILES BONDED TO AL-PLATE

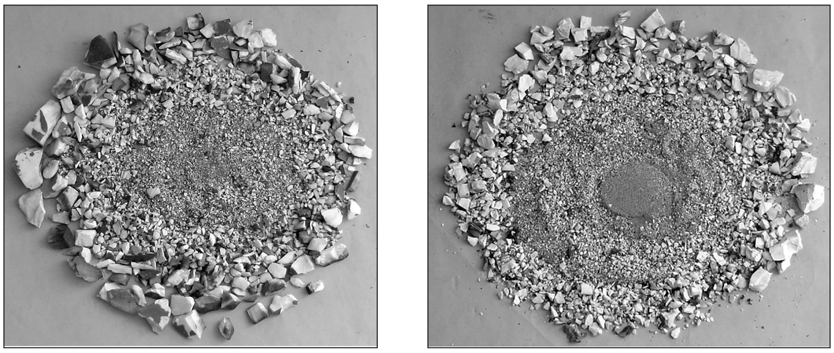
In order to study the nature of fracture and fragmentation, the target and projectile debris were collected during a few experiments. The debris collection was done by placing a closed box in front of the target and covering the front of the box with layers of polymeric cloth. The projectile debris was then separated from the total debris through magnetic separation. Figure 11 shows the target debris collected from the two grades of 14mm thick ceramic tiles fired at about 800 ms^{-1}. It is seen that a larger amount of the 95% grade ceramic debris was in the form of fine powder and very small pieces compared to the 99.5% grade ceramic.

To study the possible difference in their failure mechanism, fractographic study was carried out on the fractured ceramic tile collected from the target debris using a Scanning electron microscope. Figure 12 shows the SEM images of the fractured pieces of the two grades of ceramic. It is observed that in the case of 99.5% grade ceramic, the fracture is planar and appears across the grains, which can therefore offer greater resistance to penetration. However, in the case of 95% grade, the fracture surface is less defined along the paths of least resistance.

It is found that as the velocity of the projectile is increased, both the projectile and the target break into a larger number of finer pieces, leading to more efficient energy absorption. Figure 13 shows debris of the projectile when fired at 614 ms^{-1} and 696 ms^{-1}.

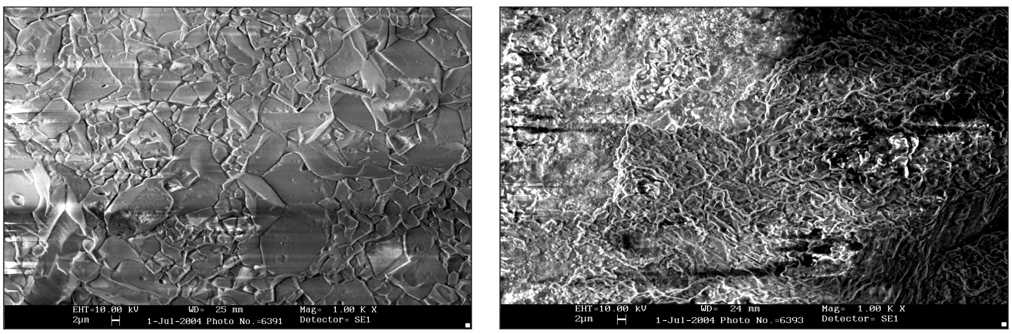
The typical finer dust particles of the projectile debris is seen to be of the order of 10 microns. Figure 14 shows the variation of the normalized ballistic efficiency against the projectile velocity for different thicknesses of ceramic.

From the data, it is seen, that for both types of ceramic grades, the efficiency increases linearly with the increase in velocity for the range of tests conducted in the present study. The increase in efficiency is found to be greater in 99.5% ceramic than in the 95% grade. From the results, it is



(a) (b)

FIG. 11 TARGET DEBRIS OF (a) C99.5 GRADE AND (b) D95 GRADE CERAMIC TILES OF 14MM THICKNESS



(a) (b)

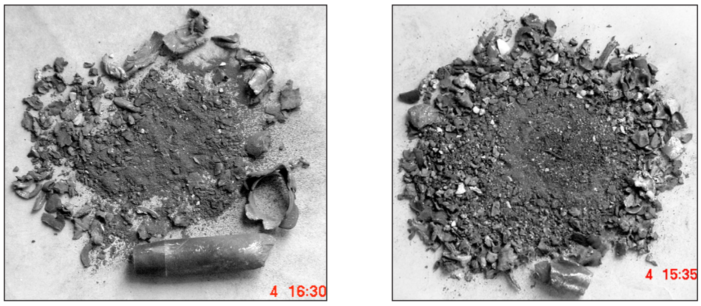
FIG. 12 SEM IMAGES OF (a) FRACTURED C99.5 CERAMIC TILE AND (b) FRACTURED D95 CERAMIC TILE

observed that the mass efficiency increases by a factor of about 1.9 when velocity changes from 500 ms^{-1} to 800 ms^{-1} in case of 99.5% grade alumina. But the increase in 95% grade is only by a factor of about 1.6.

From Fig. 10 it can be seen that the efficiency of 99.5% grade is significantly higher than that of the 95% grade at any given thickness or impact velocity. This can be partly attributed to the higher hardness and corresponding dynamic strength of 99.5% grade of alumina ceramic than the 95% grade and partly to the difference in microstructures of the fractured surfaces.

Figure 15 shows the variation of the normalized ballistic efficiency against the tile thickness for different velocities of the projectile. From the data, it is observed that with the increase in thickness of the ceramic tile, the ballistic efficiency factor for a given velocity decreases by a factor of about 1.3 in the case of 99.5% grade for all the velocities, whereas it increases by nearly the same amount in the case of 95% grade ceramic.

This ‘thickness effect’ is actually a result of two distinct effects: (a) the free surface effect and (b) the velocity effect. The free surface at the top of the ceramic generates release waves, which cause cracking and reduced stresses, thereby decreasing the resistance. The effect of this on the overall



(a) (b)

FIG. 13 DEBRIS OF THE PROJECTILE COLLECTED WHEN FIRED AGAINST 99.5% GRADE CERAMIC AT VELOCITIES OF (a) 614 MS^{-1} AND (b) 696 MS^{-1}

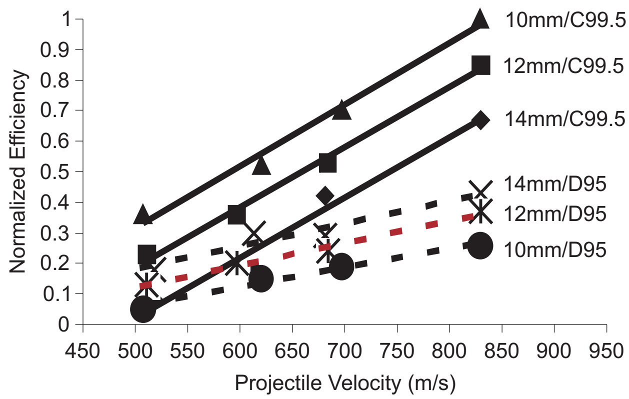


FIG. 14 NORMALIZED BALLISTIC EFFICIENCY VS PROJECTILE VELOCITY FOR CERAMIC TILES

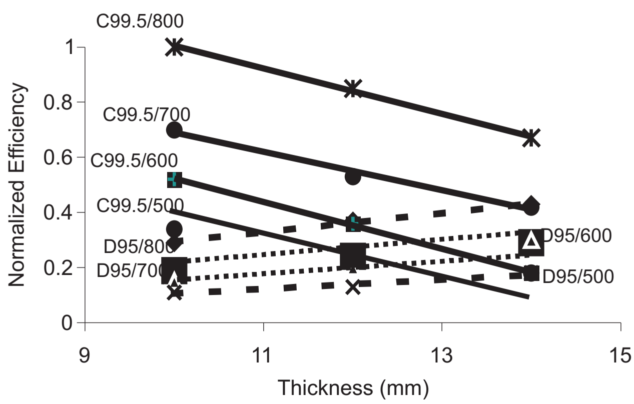


FIG. 15 NORMALIZED BALLISTIC EFFICIENCY VS THICKNESS OF CERAMIC TILES

performance diminishes with the increase of thickness of the ceramic giving rise to an apparent increase in the overall performance. The "velocity effect" is a consequence of the decreased performance with decrease in velocity. For any given incident velocity, there is a definite and strong gradient of velocity as the projectile moves from the top of the ceramic to the bottom.

Essentially then, a thick ceramic tile can be considered to be made of elemental thicknesses, wherein, each succeeding layer encounters the projectile at a lower velocity than at the previous layer and hence exhibits a corresponding lower efficiency. The average efficiency, being an average over the whole thickness, therefore decreases. The net effect due to the "free surface effect" and the "velocity effect" depends on which one among the two effects is dominant for any given ceramic. In the case of 99.5% grade, the velocity effect which is higher, apparently dominates and leads to a decrease in efficiency with an increase in tile thickness. In the case of 95% grade for which the velocity effect is only about $1/3^{rd}$ as strong, the surface effect dominates, to create a mild increase in efficiency with increasing thickness.

3.8 Polymers and Composites for Armour

High performance fibres are currently being used extensively to produce light weight PMCs for vehicle armours. The availability of different high performance fibres and the ability to tailor these fibres, allows versatility in designing fibre reinforced PMC armours. Woven fibre mats and fibre-reinforced PMCs absorb energy in different ways. The amount of energy absorbed by fibres is largely dependent upon their strain to failure [Hogg (2003)]. A fibre cloth with high strength and high elongation to failure is thus expected to absorb energy through plastic deformation and stretching of the fibres. The strain in a fibre is equated to the impact velocity divided by the sonic velocity of the fibre as given in Eq. 4 [Yang (1993)]:

$$\varepsilon = V / c \tag{4}$$

where,

ε – strain,

V – impact velocity and

c – sonic velocity of the fibre

The sonic velocity is proportional to the fibre's elastic modulus as shown in Eq. 5. Thus a higher elastic modulus results in the impact energy wave traveling longer length of the fibre due to a higher sonic velocity, and thus a greater volume of fibre absorbs the projectile energy.

$$c = \sqrt{E/\rho} \tag{5}$$

where,

E – elastic modulus

ρ – density of the fibre

Once fibres are fabricated into a laminate with a resin matrix, their ability to deform and absorb energy changes. In PMCs, the fracture process takes place in two phases. High velocity impact causes localized compression of the composite, and subsequently shearing of fibres and spalling of

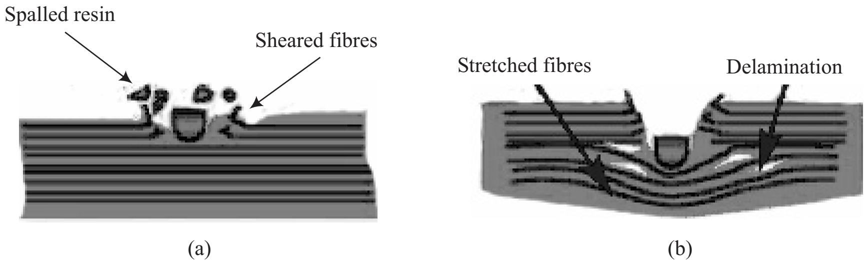


FIG. 16 FIBRE REINFORCED PMC ENERGY ABSORPTION MECHANISMS

resin, as depicted in Figure 16a. As the projectile travels through the laminate, it slows down in velocity and the composite deforms causing fibre stretching, pullout, and delamination of composite layers as shown in Figure 16b.

Three dimensional fibre weaving could reduce delamination and confine damage to a small area [Fink, Monib and Gillespi (2001)]. However, this may result in an increase in fibre damage during stitching, leading to a decrease in compressive strength after ballistic impact, and thus lower load carrying ability. Fibres can be woven together into a number of configurations, some of which are illustrated in Figure 17, to provide varying degrees of performance and flexibility.

A few advanced fibres currently being used world over for ballistic applications are S-2 glass, Aramid, HPPE, PBO and PIPD. Table 6 gives properties of the various ballistic fibres.

S-Glass is composed of silica (SiO_2), alumina (Al_2O_3), and magnesia (MgO), and has a strength about 40% higher than that of E-glass [Walling (1985)]. Its cost is significantly higher than that of E-glass, but its strength advantage, and consequently performance per unit weight advantage, gives it an edge for use in ballistic applications.

Aramid fibres were developed during the 1960s and first introduced commercially by DuPont in the 1970s under the trade name Kevlar.

There are other commercially available aramid fibres with trade names of Twaron and Technora. Various aramid fibres available today are Kevlar 29, Kevlar 49, Kevlar 129 and Kevlar KM2 from DuPont and CT709, T750 etc. from Twaron depending on the denier. Aramid fibres show a decrease in tensile strength when exposed to heat or moisture. At temperatures up to 355^{o}F, a strength loss of $\leq$ 20% occurs [Chang (2001)]. Aramid fibres are also vulnerable to damage from ultraviolet light,

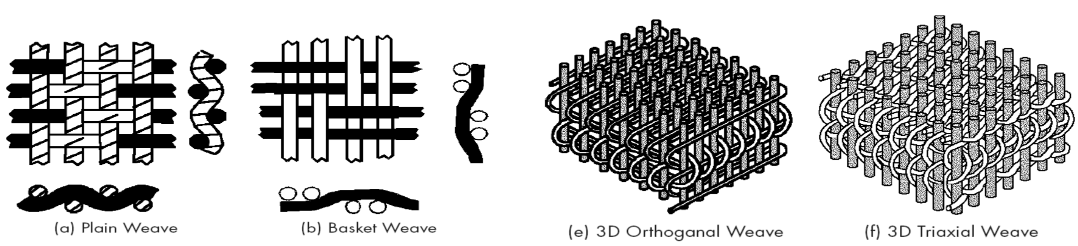


FIG. 17 WOVEN FIBRE STRUCTURES

Table 6. *Properties of ballistic fibres*

Fibre Material	*Fibre Density (g/cm^3)*	*Elastic Modulus (GPa)*	*Y.S (MPa)*	*Failure Strain (%)*
S-glass	2.48	90	4400	5.7
Technora	1.39	70	3000	4.4
Twaron	1.45	121	3100	2.0
Kevlar 29	1.44	70	2965	4.2
Kevlar 129	1.44	96	3390	3.5
Kevlar 49	1.44	113	2965	2.6
Kelvar KM2	1.44	70	3300	4.0
Spectra 900	0.97	73	2400	2.8
Spectra 1000	0.97	103	2830	2.8
Spectra 2000	0.97	124	3340	3.0
Dyneema	0.97	87	2600	3.5
Zylon AS	1.54	180	5800	3.5
Zylon HM	1.56	270	5800	2.5
M5	1.70	271	3960	1.4
M5 (goal)	–	450	9500	2.5

with as much as 49% loss in strength [Bunsel (1988)]. Strong acid and alkaline environments also reduce their performance.

High Molecular Weight Poly Ethylene (HMWPE) has a simple structure consisting of a repeating ethylene unit $[CH_2\text{-}CH_2]_n$. Two commercially produced HMWPE fibres are Dyneema [DSM (2005)] and Spectra [Honeywell (2005)]. HMWPE fibres which have the lowest density of all fibres currently used for armour applications involving low velocities and soft shots.

Polybenzobisoxazole (PBO) fibres were developed as a result of the US Air Force's research during the 1980s [Sikkema, Northolt and Pourdeyhimi (2003)]. PBO fibres have very high tensile strength and better penetration resistance than HMWPE fibres, but suffer from low compressive strength like HMWPE. A PBO fibre is currently available commercially under the trade name Zylon. Tests have shown that this fibre undergoes tensile strength degradation in hot and moist conditions, and when exposed to ultraviolet and visible light [NIJ Report (2004)]. A 40% loss in strength may occur at a temperature of 176°F and 80% relative humidity.

A new high performance fibre – Polylyridobisimidazole (PIPD), denoted by M5 – has been developed by Akzo Nobel and shows promising results. Similar to PBO, it has a rigid rod structure. Due to strong inter-molecular hydrogen bonding, its compressive strength is significantly higher than that of PBO fibres. Its decomposition temperature is about 985°F, [Sikkema, Northolt and Pourdeyhimi (2003)]. The fabrication technologies and ballistic evaluation for M5 fibres are still in developmental phases as some properties of the fibres fall short of their theoretical potential.

The ballistic performance of fibre reinforced PMC armours is largely attributed to the fibres. Figure 18 shows typical ballistic V_{50} data of some fibre-reinforced PMC armour materials when impacted against 0.30 caliber FSP [Richard (2005)].

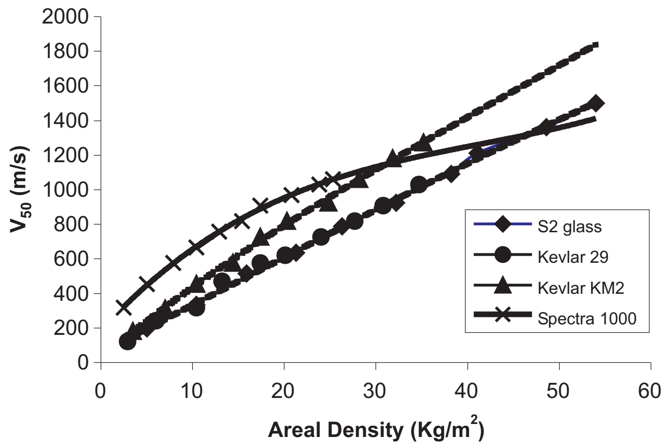


FIG. 18 BALLISTIC PERFORMANCE OF PMCS

Maximizing fibre volume in a composite using the best weave structure is necessary to optimize the ballistic performance of composites.

4. ARMOUR SYSTEMS

The increasing capability of modern anti-armour threats and the need to field lighter combat vehicles, capable of engaging any opponent with little preparation have increased the need for highly effective passive armour systems. In this context, because of their high hardness, compressive strength and low density, ceramics offer significant advantages for meeting future protection requirements. Even more promising are the systems of explosive reactive armour, bulging armour, electric armour and active armour as well as protection systems involving energy beams.

4.1 Designing of Armour Systems with Passive Materials

The design of a system using metallic armour essentially calls for studying and optimizing the angle of projectile impact and thickness of the armour and solving the problems of integration. In polymeric materials joint design can become a serious problem. Inspection, quality control and post damage acceptance is another problem. Optimisation of weaves, fabrics, rasins and their procuring is also an important problem.

The investigation and application of ceramics against small arms threats has a long history dating back to early 1960s. However, achieving similar performance against large caliber KE penetrators has been a difficult challenge. It is believed that ceramics, when used in appropriate configurations and encased macrostructures, can inflict even “interface defeat” to the projectile [Lundberg, Renstrom and Lundberg (2000)]. The key step for high efficiency is the design of a suitable system for holding the ceramic in place. Thus it is a design problem rather than a purely materials problem.

Until these problems are solved, ceramics will perform below their full potential for ballistic protection. Intense work is believed to be going on throughout the world in this area. Recent studies have tried to address the problems by providing support at the sources of damage and by introducing a weak interface layer to manage the flow of failed projectile material at the surface of the ceramic. When this approach was used, high quality ceramics have sustained minimal damage while defeating long rod projectiles launched at velocities ranging from 1600 to 2000 m/s. Yet, further studies have detected a weakness in support at the front of the ceramic and it appears that the dream of creating a practical and usable armour configurations has not yet been fully realized.

The small arms projectiles are more commonly made of hard steel (HV 700-800). These projectiles typically have a length to diameter (L/D) ratio of 3:1 to 5:1 with muzzle velocities below 1000 m/s. These projectiles tend to produce a total KE of the order of 10^3 to 10^4 J. The perforation of ceramic armour systems against these projectiles occurs in three distinct phases; shattering, erosion and catching. During the shattering, the penetrator fractures and breaks on the surface of the ceramic plate; the high compressive strength of the ceramic overmatches the loading produced by the penetrator impact, and the penetrator material flows or shatters. During the second stage, the ceramic material cracks. However, it still contributes to the defeat of the penetrator core through erosion mechanisms. During the catching phase, the ceramic loses considerable strength, but the ceramic and the backing together reduce the velocity of projectile further through momentum transfer mechanisms.

The mechanism of defeat of long rod penetrators is more complex. These penetrators are commonly made of high strength, high density materials like tungsten carbide, tungsten sintered alloy and depleted Uranium. Their L/D ratios are typically 18:1 to 25:1 and are fielded at muzzle velocities of about 1.6 to 2.0 Km/s yielding Kinetic Energy in excess of 6-8 MJ. In such a case even if the frontal portion of the penetrator can be effectively damaged, a major portion of the rod remains intact to continue the armour penetration process.

These long rods demand continuous longer duration resistance from the armour. At DMRL, the first experiments were carried out by confining ceramic tiles in large boxes of armour steel. Combining various metallic, ceramic and polymeric materials in optimum ways to maximize performance is a major challenge in armour research and development.

4.2 Explosive Reactive Armor (ERA)

Explosive reactive armour consists of layer of explosive sandwiched between two metallic plates. When the jet from hollow change hits the explosive, the explosive explodes and the high pressure zones disturb the alignment of the jet particles. The moving plates further enhance the performance of the armour.

To be effective against long-rod penetrators, plates of integrated ERA sandwiches must still be relatively thick to impart sufficient transverse momentum to the rods to deflect and to break them. However, a fine balance has to be struck between making the front sandwich plates heavy enough to be effective against the penetrators and not too heavy to be contained by the outer layer of armor. The impact of the rear plate on the inner layer of armor is less of a problem and can be reduced by making the sandwiches asymmetric, so that the rear plate is significantly thinner than the front plate. Making the front plate thicker also helps to prevent premature detonation of the explosive interlayer by minor threats. Given optimized designs, integrated ERA offers future tanks highly effective protection both against the penetrators of APFSDS projectiles and the jets of shaped charge weapons (even those with tandem warheads). [Ogorkiewicz (1997)]. Figure 19 shows a diagrammatic representation of integrated explosive reactive armor.

Use of explosives as in explosive reactive armour (ERA) – though an attractive option – has some limitation such as significant edge effects, needs for protection of the armour against KE shots and high sensitivity to angle of attack (as shown in Figure 20), fire and explosion hazard, possibility of collateral damage and limited shelf life.

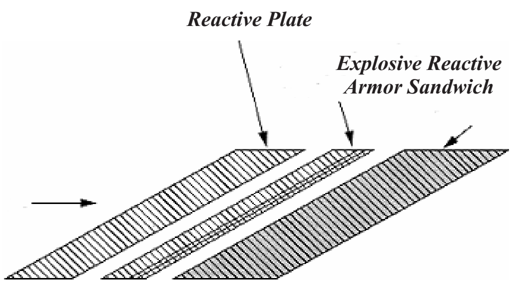


FIG. 19 INTEGRATED EXPLOSIVE REACTIVE ARMOR

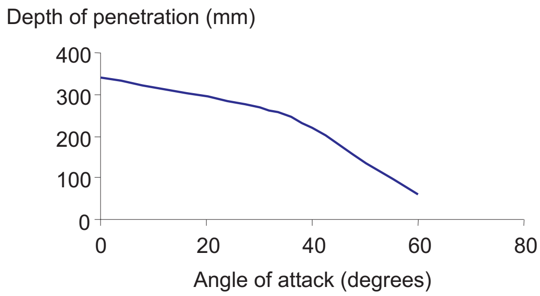


FIG. 20 EFFECT OF ANGLE OF ATTACK ON PENETRATION DEPTH IN RHA BEHIND AN ERA AGAINST A HEAT

4.3 Electric Armor

The electric armor typically has two widely spaced plates, one of which is connected to a high-voltage capacitor bank while the other is grounded. In an attack, a shaped-charge jet penetrates the plates and acts as a short circuit between them causing a heavy current to surge through the jet [Smith (2002)]. Magnetomechanical instabilities in the jet lead to its break-up. In the case of the penetrators too, electrical currents can cause instabilities and disruption of the penetrator. This armour can be made intelligent by coupling to a detector system based on radar signatures during flight or pressure and velocity signatures during initial stages of impact.

The physical cause that underlies the damage to the jet, is the necking and fracture due to the magneto hydrodynamic instability. The magnetic pressure that acts on the jet surface can be written as

$$P = \mu_\circ/2\ [I/2\pi R]^2 \tag{6}$$

where,

$\mu_\circ = 4\pi \times 10^{-7}$ H/m is the permeability of vacuum

I = Current flowing through the jet

R = Radius of the jet

It can be seen that a decrease in R causes a rapid increase in pressure. So any perturbation shrinkage in radius causes accelerated shrinkage thus leading to necking. Also as the jet leaves the interelectrode gap there is a sudden release of this pressure causing the jet to expand and break up. The electrical field disperses the particles into a harmless cloud over the surface of the vehicle. This way, it is possible to withstand grenade and HEAT attacks with an armour weighting one to two tons in place of 10–20 tons of conventional armour over typical fighting vehicles.

When the pair of plates are separated by other layers of insulating material, the system becomes an electrothermal armour. When this layer is penetrated, the shorting generates heat and the layer evaporates and expands rapidly to throw the plates apart. To be effective against long rod penetrators, armour plates have to be heavy enough to affect the projectile. This may need about 1 MJ of energy. Fortunately, the velocity of the plate does not have a decisive influence on its effectiveness, with velocities as low as 50 m/s being adequate.

4.4 Active Protection

Active protection relies on detecting the approaching threat and then launching some type of counter-measure. The counter measure may be decoys or jammers or fragmentation cassettes. These countermeasures fall into two categories. These are the soft-kill systems and hard-kill systems The soft kill make the attacking munitions miss their targets without damaging them. They confuse the incoming missile, by using decoys, smoke and electro-optical signals, infrared or laser jamming Threat warning or detection systems are a prerequisite to hit avoidance. The hard kill is designed to intercept and destroy the incoming projectile or missile before it hits the vehicle. Countermeasures include fragmentation charges, steel bars, high pressure shock waves that destroy the incoming threat, destabilize or disrupt it flight path, or divert it from its course. These systems use millimeter-wave radar to detect and track approaching missiles. Shtora and Arena of Russia; Galix, KBCM and Spatem of France; FSAP, FCLAS and IAAPS of USA; MIDAS of UK; AwiSS and Muss of Germany; and Pomals and Trophy of Israel are some of the known Active Protection Systems around the world. But reliable and quick sensor systems are a prerequisite to hit avoidance. Also the sensors must discriminate between true and false attacks [Meyer (1998)].

4.5 Protection through Directed Energy Weapons

Future battlefields may see a revolutionary change as the directed energy weapons begin to get displayed. Moving with the speed of light, they can shoot down missiles, shells, bombs, or aircrafts and directly serve the purpose of protection. Lasers can be produced in a range of wavelengths from infrared to X-rays. They are produced using chemically doped glass, oxygen, duetorium fluoride, energetic reactions, semiconductors or magnetically confined free electrons. The power output can range from a few Kilowatts to Megawatts.

One problem that affects laser beams is "blooming" which occurs as the air on the path turns into a plasma. Dust, smoke, rain of cloud can also absorb the laser! Another problem is that lasers travel only in a straight line and would need appropriate mirrors to redirect it. There further arises a challenge of protection against these laser beams. One rather immediate threat which needs attention is that of laser dazzlers that temporarily blind sensors, optics and personnel. One possibility is use of appropriately designed mirrors. Another is the development of materials aggregates which locally absorb the energy without generating global damages or shock waves. High power microwaves – in the megawatt range – form another set of directed energy tools. It induces a current in unshielded electronic components and cause them to malfunction. At higher energy levels it can cause severe burnout of semi-conductors and electronic devices and can be used as a means of protection. Pulsed power devices or devices which explosively compress an electrically charged coil are used to generate the required bursts of power. On the other hand when the need is protection from microwaves we need to use properly designed conductive metal cages or equipments or dress materials for individuals. Yet another future possibility is particle beam weapons, which poses tremendous Kinetic energy too, making them nearly unstoppable especially for deriving protection.

5. APPLICATIONS

Developments in armour research around the world have delivered many advanced materials and systems. These are : metallic materials like medium & high hardness steels, aluminium alloys, titanium alloys, ceramics like alumina, silicon carbide, boron carbide and titanium di–boride, polymers like FRP, aramids like Kevlar & twaron high performance polyethylene like dyneema & spectra, polycarbonates, silk., cellular materials, sandwich structures, compound armours, explosive reactive armour. Newer materials and systems including dynamic and active armours are now being developed.

At DMRL, we have been able to develop a variety of effective armour materials which may be used singly or in combination. These range from porous ceramics, fiber glass laminates, advanced ceramics, various types of armour steels like spade M1 (medium hardness), Jackal & DMR1700 (high hardness), titanium alloys, fiber reinforced composites and compound armour named “Kanchan”. Using them, it is possible to tailor exciting armours for withstanding any threat or combination of threats. These materials and designs have been used extensively for various applications as shown in Figure 21.

6. GENERAL COMMENTS ON TOTAL DEFENCE AND PROTECTION

Injury deaths and losses due to war is actually a miniscule fraction of the total injuries, unnatural deaths and losses in the world. Many other important resources also need protection [Bhat (2006)].

There are many challenges in this area. Precious resources such as minerals, plants, animals, insects, sea life, water, sunlight, health, time, state of education, population, institutions, past glory and even more precious emotional roots such as culture, tradition, religion, economy and intellectual properties or the very spirit of nationhood, may be subjected to attack and may suffer continuous slow bleeding or instantaneous fracture. These need to be protected by appropriate armours. Any nation or group which ignores this requirement will either perish or soon resort to aggressive offensive means such as war or terrorism for its own defence. Obviously the technology of armour - the technology of protecting ourself without having to attack others - in each of the areas are quite different from what we have discussed until now. Yet they are too important to be ignored. Therefore a few armouring concepts relevant for these areas are indicated. Some comments are also made on armouring concepts relevant to various facets of nation’s survival and growth.

FIG. 21 PHOTOGRAPHS OF SUCCESSFUL APPLICATIONS OF ARMOUR 1. VIJAYANTHA TANK 2. MBT – ARJUN 3. ICV-ABHAY 4. MI-17 HELICOPTER 5. LANCER HELICOPTER 6. BODY ARMOUR 7. VIP CARS

Attacks on emotional capitals such as spirit of nation, culture, tradition and religion is extremely damaging. They must be resisted by suitably devised armours of thoughts. These thoughts must be constantly updated to match and block the virus that tries to attack. It is not to say that good thoughts from outside should not be accepted. Just as we accept food from outside - plants, animals - and digest and make and strengthen our own body using them, we may accept good thoughts from outside, digest and convert them and use them and constantly update and rejuvenate ourselves. If we do not do this and also we do not ban or resist the attacks and be merely dumb sheep, we would be eventually converted and lead to the slaughter house. Neither our base, nor our envelop and the spirit of nationhood will survive if we leave organized thought attacks unattended in the name of freedom, leaving everything to the hapless individuals and small groups to handle by themselves subjected to attack by crafty organized outsiders.

Another precious commodity is time. Time that man has with him for doing useful things and for attaining higher levels, for himself and for the surroundings, is limited. The amount of time taken away from men through worthless TV serials, broadcasts, advertisements, junk reading materials, commentaries, news papers, and excessive disabling schooling, disputes, quarrels and long duration legal battles, habit forming behaviors, addictions, excessive medication, poor health maintenance practices is colossal. More than half of the time is taken away in these activities in many countries. In effect it means that half the country, generation after generation, is invisibly killed and vaporized by these rather inconspicuous attackers. Suitable armour for this is simply the envelope of knowledge - an acute awareness of the colossal damage they are causing to the elements of the nation. Time should utilized in honing and developing our practical skills in as many areas as possible along with the theoretical knowledge rather than in dumbly watching the shows or reading about them. Right from the childhood one should work along with elders in all possible ways to achieve the stamina to work.

Water is life. Rivers are the arteries and veins. A litre of pure distilled water would cost about 5 rupees to make. So, 40" of rain means a gift of Rs. 5, 000 on every square metre of the country's land. To let it go from the pure form in our terrace, then slightly impure form from our garden, then little more impure form from our land and finally to let it join the salty ocean even as we are thirsty, is an unimaginable foolishness. It can be shown that roof water alone is ten fold more than adequate to satisfy all the drinking needs in most parts of the world. The armour against loss of water - this injury and bleeding - is various forms of bunds, dams, lakes and collections. Differential pressure dam at the sea coast [Bhat (2006)] is another wonderful way to protect water. This dam can also play the role of armour against attacks such as enemy ships or Tsunami.

Soil, the thin nutritious cover of earth is the "real country". All below is Volcano. Therefore, erosion of soil due to attack of rain water etc., must be resisted. Cover of grass is the best known armour for this. Grass can very well protect the soil against impact of water, wind, etc.

There is an intense energy attack on earth occurring all the time. The solar and other radiations amount to a kilowatt on every square metre of earth. It is like a gas stove all the time burning on every square metre and burning us down. We need an armour to protect us from this. If not, global warming may make all life extinct. Viewed another way sunlight is a highly valuable gift from the sun. Therefore, in this case, the armour should be ideally of that type which captures the whole thing intact, transforms it and stores it. Very little should be reflected or dissipated. Here, too, a thick layer of Chlorophyl – grass and plants – is the best form of armour.

The plants, animals, insects, sea life and other forms of life also need protection against attacks from insecticides, pesticides, herbicides, terminators etc. For this ignorant short term greedy attack by humans, proper education is the best armour. Still better would be a law making all forms of life "citizens"! In any case, weapons of "mass destruction" must be banned from being used against any form of life. A law, a sheet of energised paper as it were, is the best armour here!

Biological attack is another very serious threat. An example is AIDS. Best form of armour for protection against this is detection. Every few years, the whole population with no exception, should be screened. Following detections, simple stand off – a spaced armour - is enough. Keep a distance and avoid intimate contacts with those carrying the terrorist virus! In a few years the virus would find no shelter, exhaust itself.

Even air needs protection not only from excessive poisonous chemicals but also from more deadly forms such as those which punch holes in it - the ozone holes - as if attacked by a bomb. Early detection and high level of taxation to dissuade such activities and use of damaging chemicals is the best form of armour : again a sheet of energetic paper!

Usually armour is for protecting against threats - hits - from others. A very common but peculiar killer situation is when we hit others. All traffic accidents are of this class. If we look at the statistics, the number of disablings and deaths due to this reverse impact in one year alone exceeds the number of deaths due to war and such situations during the entire past fifty years of independence in India! What are the problems? Why are they not addressed? What do we need to do? Actually the problem is technically quite simple. The speeds involved are not even one tenths of the speeds involved in the case of bullets and shots. A few millimeter thick light weight composite fibre glass, kevlar or even nylon or cotton, worn in the form of helmets, and pads on the legs - somewhat like what the cricketers do is enough to drastically reduce the casualties on two wheelers. In automotives such as cars or buses shifting to light weight crash resistant structures made of the very same materials described in earlier chapters will be able to save nearly all the lives being lost on the roads. Put in some belts around and the protection becomes nearly complete. Solutions exist. Again what is needed is to activate and implement a brave paper of legislation!

Then there is an extreme situation of satellites or spacecrafts. They collide against the fine dust like particles on their path. The collision velocities - here, mind boggling 30-100 km/sec, - are one thousand times higher than what they are in our road accidents. The energy density involved becomes a million times higher. Under these conditions of impact, materials get shocked to such temperatures and pressures that they vapourise and even ionise instantly. Nearly all the concepts of armouring involving increased strength, toughness, hardness, elongation, energy absorption etc., described in the earlier chapters are thrown to the wind under such situations. Fortunately, it turns out that the most efficient armour here is paper thin bumpers made of materials such as cadmium or aluminum which causes the impacting materials to vapourise and dissipate in the form of a bubble. All that is needed is to provide enough space or gap for the bubble to expand and loose its punch. Once again a well designed thin 'paper' like armour functions as an effective armour.

Finally, an effective armour for man is his family. The family is an unparalleled armour against all forms of attack – attacks of hatred, anger, depression, addictive agents, lust, greed, delusion, intoxication, jealousy, loss of faith, divorce, violence, ill health, and all forms of physical, mental and emotional attacks. With this armour firmly around, the state of happiness and composure expands manyfold. Yet, the most effective active armour for any human being is himself and the Infinite Consciousness around him. One way to achieve this is the process of Yoga and Pranayam. While at work and at home, when everyone is controlled and protected with this best armour – the active, dynamic, intelligent armour – designed by god – all guns will automatically fall silent and total protection will be at place for everyone. With individuals being at deep peace nations too will be at peace with one another.

7. REFERENCES

Balakrishna Bhat, T. (Jan 2006) Total Defence, *Rachana* DMRL, Vol. 2, pp 1-5.

Balakrishna Bhat, T; Madhu, V; Pappu, **S.** (2005) Approaches to design of materials resistant to high strain rates, *Proceedings of the 13th International Conference on Computational and Experimental Engineering and Sciences, ICCES'05*, (Chennai, India), pp 2851-2856.

Bunsel, A.R. (1988) Ed, Fibre Reinforcements for Composite Materials, *Elsevier Science Publishers*.

Chang, K.K. (2001) Aramid Fibers, *ASM Handbook* – Vol. 21: Composites, ASM International.

DSM. (2005) Manufacturer data, http://dsm.com

Fink, B.K; Monib, A.M; Gillespie Jr., J.W. (2001) Damage Tolerance of Thick-Section Composites Subjected to Ballistic Impact, *Army Research Laboratory*, (USA) ARL-TR-2477.

Gooch, W.A; Burkins, M.S. (2001) Ballistic development of U.S. high density tungsten carbide ceramics. *Pac Rim IV conference : Advanced Ceramics and Glass*, (Hawaii) , pp 53-61.

Hogg, P.J. (2003) Composites for Ballistic Applications, *Journal of Composites Processing*, CPA, Bromsgrove (U.K).

Honeywell International, Inc., (2005) http://spectrafiber.com

James, B. (1998) Direct evidence for ceramic comminution ahead of penetrator, *17th Int. Sym. On Ballistics*, Midrand, (South Africa), Vol. 3, pp 9-16.

Lundberg, P; Renstrom, R; Lundberg, B (2000) Impact of metallic projectiles on ceramic targets : transition between interface defeat and penetration, *Int J Impact Engng*., Vol. 24, pp 259-275.

Madhu, V; Sivakumar, K; Balakrishna Bhat, T. (1999) Force multiplication through enhanced protection using advanced armour materials, *International seminar on 'Force Multiplier Technologies for Naval and Land Warfare'*, Vigyan Bhavan, (New Delhi, India), pp 416-425.

Meyer, T.J. (June 1998). Active Protection Systems : Impregnable armor or simply enhanced survivability?. *Armor*, pp 7-11.

Michael Smith. (2002) http://www.dstl.gov.uk/.

Ogorkiewicz, R.M. (1997) Future tank armor revealed : Development in electric and explosive reactive armour. *Janes International Defense Review*, Vol. 30, pp 50-51.

Ralf Riedel. (2000) *Handbook of ceramic hard materials*, Wiley-VCH, (Germany), Vol. 2, pp 838-856.

Richard Lane, **A.** (2005) High performance fibers for personnel and vehicle armor systems, *AMPTIAC*, Vol. 9, No. 2, pp 3-9.

Rosenberg, Z; Yeshurun, Y. (1988) The relation between ballistic efficiency and compressive strength of ceramic tiles, *Int. J. Impact Engng*, Vol. 7, No. 3, pp 357-362.

Shergill, M.S. (1997) Battle field milieu of the 21st century, *Proceedings of the seminar battle field milieu of the 21st century – Challenges,* editor: Balakrishna Bhat T, Rao, S.S, (DMRL, Hyderabad), pp 1-6.

Sikkema, D.J; Northolt, M.G; Pourdeyhimi, B. (2003) Assessment of New High-Performance Fibers for Advanced Applications, *MRS Bulletin*, Vol. 28, No. 8, pp 579-584.

Status Report to the Attorney General on Body Armor Safety Initiative Testing and Activities, *National Institute of Justice*, (March 2004).

Subramaniam, J (2006) DMRL : Personal Communication.

Vemuri Madhu; Ramanjaneyulu, K; Balakrishna Bhat, T; Gupta, N.K. (2005) An experimental study of penetration resistance of ceramic armour subjected to projectile impact, *Int J Impact Engng*, Vol. 32, pp 337-350.

Walling, S.J. (1985) S-2 Glass Fiber: Its Role in Military Applications, *International Conference on Composite Materials, Metallurgical Society of AIME*, pp 443-456.

Wilkins, M.L. (1978) Mechanics of penetration and perforation, *Int. J. Engng. Sci*,, Vol. 16, pp 793-807.

Yang, H.H (1993) *Kevlar Aramid Fiber*, John Wiley & Sons.

8. FURTHER READING

Anderson, C.E. Jr.; Morris, B.L. (1992) The ballistic performance of confined Al_2O_3 ceramic tiles, *Int. J. Impact Engng*, Vol.12, pp 167-187.

Anderson, C.E. Jr.; Royal Timmons S.A. (1997). Ballistic performance of confined 99.5% Al_2O_3 ceramic tiles, *Int. J. Impact Engng*, Vol.19, pp 703-713.

Anderson, C.E. Jr.; Walker, J.D; Houver, G. (1992) Target resistance for long rod penetration in to semi-infinite targets, *Nuclear Engineering Design*, Vol.138, pp 93-104.

Angers, R; Beauvy, M. (1984) Hot pressing of boron carbide, *Ceramic International*, Vol.10, pp 49-55.

Backman, E.M. (1986) Stagnation cap formation on blunt projectiles penetrating metallic and brittle targets, *Metallurgical application of shock wave and high strain rate phenomena*. (NewYork), Dekker publications, pp 823.

Backman, E.M. (1986) Scaling rules for adiabatic shear, *Metallurgical application of shock wave and high strain rate phenomena*, (New York), Dekker publications, pp 675.

Backman, M.E; Goldsmith, W. The mechanics of penetration of projectiles into targets, *Int. J. Engng. Sci*., Vol.16, pp 1-99.

Bonsignore; Ezio; Ian Bustin. (1993) Towards Electronic Armour, *Military Technology*, p 10.

Bose, D.K; Nair, K.U; Gupta, C.K. (1986) Production of high purity boron carbide, *High temperature materials and processes*, Vol. 7, pp 133-140.

Bougoin, M; Thevenot, F; Dubois, J; Fantozzi, G. (1985) Synthesis et caracterisation de ceramiques denses en carbure de bore, *J Less Common Metals*, Vol. 114, pp 257-271.

Chen, S; Wang, D.Z; Huang, J.Y; Ren, Z.F. (2004) Synthesis and characterization of boron carbide nano particles, *Applied Physics A*, Vol. 79, pp 1757-1759.

Chung, S.K. (1990) Fracture Characterization of Armor Ceramics, *American Ceramic Soc. Bulletin*, Vol. 69, No. 3, pp 358–366.

Corbett, G.G; Ried, S.R; Johnson, W. (1996) Impact loading of plates and shells by free flying projectiles : a review, *Int. J. Impact Engng*, Vol.18, pp 141-230.

Curran, D.R; Seamen, L; Cooper, T; Shockey, D.A. (1993) Micromechanical model for communition and granular flow of brittle material under high strain rate application to penetration of ceramic targets, *Int. J. Impact Engng,* Vol.13, pp 53-83.

Donald, A. Shockey. (1986) Material aspects of the adiabatic shear phenomena, *Metallurgical application of shock wave and high strain rate phenomena,* (New York), Dekker publications, pp 633.

Edward, J. F. (Sept 2002) Titanium armor plate, *Advanced materials and processes*, p 35.

Franzen, R.R; Orphal, D.L; Anderson, C.E. Jr. (1997) The influence of experimental design on depth of penetration test results and derived ballistic efficiencies, *Int. J. Impact Engng,* Vol.19, pp 727-737.

Geiger, G. (2004) Progress in ceramic armour, *American Ceramic Society*.

Goldsmith, W. (1999) Non ideal projectile impact on targets, *Int. J. Impact Engng,* Vol. 22, pp 95-395.

Gupta, N.K; Madhu, V. (1997) An experimental study of normal and oblique impact of a projectile on single and layered plates, *International Journal of Impact Engineering*, Vol. 19, No. 5-6, pp 395-414.

Haung, W. (1995) Metallurgical studies of deformation and failure patterns from target impacted by the long rod penetraters, *Shock wave and high strain rate phenomena*, (New York), Dekker publications, pp 265.

Held, M. (2001) Disturbance of shaped charge jets by bulging armour, *Propellants Explosives Pyrotechnics*, Vol. 26, pp 191-195.

Held, M. (2005) Shaped charge optimization against bulging targets, *Propellants Explosives Pyrotechnics*, Vol. 30, pp 363-368.

Hewish; Mark; Leland, N. (1996) Shoot First, Ask Questions Later: Smart Tanks Learn To Fend For Themselves, *Jane's International Defense Review*, pp 34.

Jain, L.C; Ichalkaranje N.S; Tonfoni, G. (Nov 2002) Advances in intelligent systems for defence, *World Scientific Pub*.

James, J. F. (Sep 2003) Aluminium alloy in military vehicles, *Advanced materials and processes*, pp 43.

James, W McCauley. (2002) Ceramic Armour Materials by Design, *Ceramic Transactions*, v134.

Jane's Armour and Artillery Upgrades 1997-1998.

Jeffrey, J. S. (2005) Advances in ceramic armour, *Ceramic Engineering and Science Proceedings*, Vol. 26, No. 7.

Joseph, R. Pickens. (May 2004) Low cost titanium for ships and tanks, *Advanced materials and processes,* p 37.

Kashin, V. (Jun 1996) Arena: Active Protection System For Tanks, *Military Parade*, p 2.

Kennedy, C; Murr, L.E. (2002) Comparison of tungsten heavy alloy rod penetration into ductile and hard metal targets : microstructural analysis and computer simulations. *Material Science and Engg*., Vol. A325, p 13.

Lichtenberger, A; Scharf, M. (Nov 1999) Metallographic investigation of Titanium alloy impacted by long rod penetraters, *18th International Symposium on Ballistics*, (Antonio, TX).

Littlefield, D.L; Anderson, C.E. Jr.; Partom, Y; Bless, S.J. (1997) The penetration of steel targets finite in radial extent, *Int. J. Impact Engng*. Vol.19, pp 49-62.

Madhu, V; Dikshit, S.N; Balakrishna Bhat, T. (2003) Armour for futuristic infantry combat vehicles, *National Seminar on Emerging trends in ICV technology,* VRDE, (Ahmednagar, India).

Mayseless, M; Goldsmith, W; Virostek, S.P; Finnegan, S.A. (1987) Impact on ceramic targets, *J. Applied Mech*, Vol. 54, pp 373-378.

Mescall, J.F. (1986) On the relative roles of Strain hardening and thermal softening in adiabatic shear bands, *Metallurgical application of shock wave and high strain rate phenomena*. (New York), Dekker publications, p 689.

Monal, S; Ajit, K. Banthia. (2005) Low temperature synthetic route for boron carbide, *J European Ceramic Society*, Vol. 25, pp 287-291.

Munjal, Chheda; Michael, J. Normandia; James, Shih. (Jan 2006) Improving ceramic armor performance, *Ceramic Industry*, pp 124-126.

O'Donnell, R.G. (1991) An investigation of the fragmentation behaviour of impacted ceramics, *J. Materials Science Letters*, Vol.10, pp 685-688.

Ogorkiewicz, R.M. (Oct 1997) Transforming the Tank, *Jane's International Defense Review*, p 40.

Peter, T.B. Shaffer. (1992) Engineering properties of carbides, *Engineered Materials Handbook : Ceramics and Glasses ASM*, Vol. 4, pp 804-805.

Shockey, D.E; Marchand, A.H; Skaggs, S.R; Cort, G.E; Burkett, M.W; Parker, R. (1990) Failure phenomenology of confined ceramic targets and impacting rods, *Int. J. Impact Engng*, Vol. 9, No. 3, pp 63-275.

Sivakumar, K; Balakrishna Bhat, T. (1998) High velocity impact response of fibre glass reinforced plastic laminates, *17th International Symposium on Ballistics*, Midrand, (South Africa), p 175.

Soon-Kil, Chung. (1990) Fracture characterization of armour ceramics, *Amer Ceram. Soc. Bul*, Vol. 69, No. 3, pp 358-366.

Stelly, M; Dormeval, R. (1986) Adiabatic shearing, *Metallurgical application of shock wave and high strain rate phenomena,* (New York), Dekker publications, p 607.

Thevenot, F, (1990) The Physics and chemistry of carbides, nitrides and borides, Freer R (ed), *ASI series E*, (The Netherlands), Vol. 185, pp 87-96.

Tracy, C; Slavin, M; Viechnicki, D. (1988) Ceramic failure during ballistic impact, *Advances in Ceramics: Fractography of glasses and ceramics*, Vol. 22, pp 295-306.

Viechnicki, D.J; Slavin, M.J; Kliman, M.I. (1991) Development and Current Status of Armor Ceramics, *Am. Ceram. Soc. Bull.*, Vol. 70, pp 1035-1039.

Woodward, R.L; Gooch, W.A; O'Donnell, R.G; Baxter, B; Jand Pattie, S.D. (1994) A study of fragmentation in the ballistic impact of ceramics, *Int. J. Impact Engng*, Vol. 15, pp 605-618.

Yaziv, D; Partom Y. (1993) The ballistic efficiency of thick alumina targets against long rod penetrators, *14th Int. Sym on Ballistics*, Quebec, (Canada), Vol.2, pp 331-340.

Zukas, J.A; Gaskill, P. (1996) Ricochet of deforming projectiles from deforming plates, *Int. J. Impact Engng*., Vol. 18, pp 601-610.

Zukas, J.A. (1990) *High velocity impact dynamics*, John Wiley & Sons, (New York).

Zukas, J.A; Nicholas, Swift; Curran, G. (1981) *Impact Dynamics*, John Wiley & Sons, (New York).

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

SURFACE ENGINEERING OF CYLINDER COMPONENTS

N. GOWRISHANKAR*

ABSTRACT

The rapid, world-wide spread of automobiles has necessitated the development of improved technologies for environmental protection, resource utilization and customer satisfaction. As the designs of automotive components have changed, emphasis has shifted to tribological approaches. New tribo-materials are needed for automotive components that must run at higher temperatures, higher pressures and higher velocity conditions in highly corrosive environments. This paper outlines current issues related to tribo-materials and coatings practiced/proposed for cylinder components in passenger cars and commercial vehicles. It provides a glimpse of some R&D work done in our laboratory on piston ring coatings.

1. INTRODUCTION

In the automotive industry, the need for lower manufacturing costs, the use of less strategic materials, as well as easier, quicker and more flexible routes for manufacturing are of high importance. Engine design and tribology engineers are constantly challenged to innovate advanced products to meet more demanding emissions and fuel economy targets.

Further, the automotive companies are looking for increasing the efficiency of the engine and reducing the weight of the components. In the engine and engine components, power losses due to friction add up to 15% of the total energy losses in a vehicle. Cylinder friction is a major contributor to overall frictional force occurring in the engine. Future engines are aimed towards more fuel efficiency which necessitates reduction in the power cylinder friction. From the overall friction occurring in the IC engine, around 30% is contributed by the cylinder components. Figure 1 depicts the amount of friction caused by each component in the internal combustion (IC) engine.

Figure 2 illustrates the various components present inside the cylinder. For each IC engine, the cylinder comprises a cylinder wall or sleeve and a piston fitted with rings on its grooves to seal the combustion gas. On the underside of the piston, a tube-shaped pin is placed which is connected to the crankshaft using the connecting rod. The whole system moves back and forth against the

*IP Rings Ltd., D11/12, Industrial Estate, Maraimalai Nagar – 603 209, Tamil Nadu

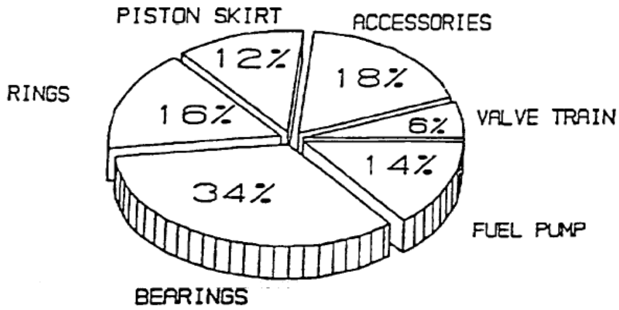


FIG. 1 FRICTION SHARE OF EACH COMPONENT IN THE ENGINE

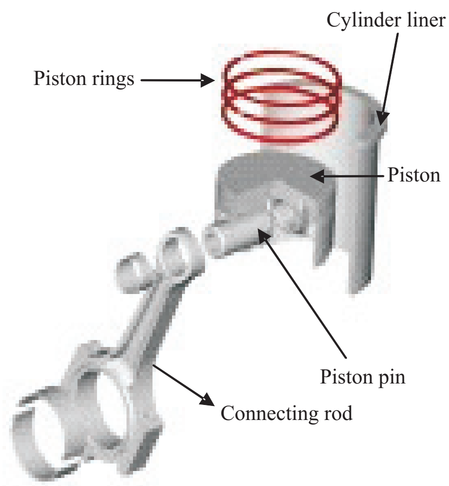


FIG. 2 COMPONENTS PRESENT INSIDE THE CYLINDER WALL

cylinder wall at velocities depending on the engine power. The movement creates lots of friction on each component that requires considerable surface engineering attention to avoid scuffing/seizure of the engine.

The cylinder components are also subjected to aggressive environments such as corrosive media, wear and thermal cycling. It is very difficult to choose a single material that possesses both good mechanical and tribological properties at low cost. Moreover, each component demands specific functional requirements depending on its role in operation. For instance, piston ring primarily requires wear resistant coating, whereas piston skirt coating should impart lubrication property. There are several surface engineering processes being practiced for the cylinder components viz. piston, piston rings, cylinder wall, piston pin and the connecting rod to the crankshaft.

The present paper aims to elaborate the various types of coatings that can perform smoothly in engine operation. The futuristic coatings for high-end engines are also outlined.

2. CYLINDER BORES

2.1 Special Features

Smooth cylinder bores with special features can reduce oil film thickness and thus minimize oil consumption and particulate emissions. [1].

Conventional honing machines are used for normal honing, plateau honing and fine honing (without pocket features).

Special honing methods are used for improved finishes as listed below:

a) Spiral honing

This is a refined variant of plateau honing. Deep grooves are produced during intermediate honing at an angle of about 150°. A honing machine with higher axial speed and high torque at low rotational speed is used. The plateaus are cut during finishing with special diamond honing tool at conventional honing angle of about 50°. The deep honing grooves are reduced to required size during this process. This results in low peak roughness value, reflecting the cylinder surface in a run-in engine, thus avoiding the running-in effects. With spiral honing, oil consumption is initially 25-50% that with plateau honing and with increased running up to say 13000 Kms, decreases to 20%.

b) Laser pocket structuring

Different patterns in all or part of the bore surface is possible by this process. Fine honing is carried out after laser pocket structuring to remove ‘overshoot’ caused by the process. With laser pocket structuring, oil consumption is initially about 30-50% that with plateau honing and with increased running up to say 14000 kms, decreases to 25-40%.

The smooth honing variants show reduction in both partial-load and rated-power oil consumption and reduce the soluble organic fraction (SOF) of the particulates by more than 50% compared to conventional honing.

2.2 Thermal Spray Processes

There is an increasing use of aluminum alloy engine blocks to reduce vehicle weight. Cast iron liners are used for those cylinders to improve the wear resistance. Using advanced plasma spray technology, a thin (150ìm) coating can be deposited directly onto aluminum alloy cylinder bores eliminating the need for cast iron or composite liners. These coatings have many advantages over conventional bore treatments. The residual porosity in the coating helps to reduce significantly the coefficient of friction by creating a micro-cavity lubrication system. Compared to cast iron liners, these micro-cavities extend the hydrodynamic lubrication area, leading to greater performance and reduced oil consumption. The wear resistance of these coatings is much better than cast iron by several factors and the heat flow between the combustion chamber and engine block is much more predictable. The weight of the block, as well as the pitch distance between bores, is further reduced.

2.2.1 *Process selection*

Different thermal spray processes are being used to provide coatings for cylinder bores [2-5]. The atmospheric plasma spraying process (APS) shows several advantages in comparison with high velocity oxy-fuel (HVOF) spraying process or electrical wire arc spraying. The high degree of liberty regarding the choice of materials, the reliability of the melting process and the low heat transfer to engine blocks are some of the key advantages of the APS process.

The low cost for high volume production makes the process very attractive [6, 7]. The coating is deposited using a rotating plasma torch while the engine block remains stationary. Figure 3 shows the photograph of plasma spray coating on the cylinder bore. Prior to coating deposition, the surface is activated using grit blasting with alumina-based material. This plasma process for the coating of cylinder bores was first introduced into production in 2000 in Europe while the competitive processes HVOF and electrical arc still are at the prototype stage. Table 1 illustrates the salient features of each thermal spray coating process.

2.2.2 *Coating characteristics*

Different types of materials can be used for the coating of cylinder bores in engine blocks [8]. For metallic alloys or composites, the following characteristics are measured for coated cylinder bores of 75 to 100 mm diameter.

FIG. 3 PHOTOGRAPH OF PLASMA SPRAY PROCESS ON AL-SI CYLINDER BORE. THE PLASMA TORCH ROTATES TO COAT THE INTERNAL DIAMETER OF THE BORE WHILE THE CYLINDER BLOCK REMAINS STATIONARY

Table 1. *Capabilities of each thermal spray process for cylinder bore*

Criterion	*Plasma Transferred (wire arc process)*	*HVOF(wire or powder)*	*Rotating Plasma (powder)*
Versatility in the choice of material	Metallic alloys Restricted choice of materials	Metals, alloys, Carbides, composites. Limitation for refractory materials	Metals, alloys,Carbides, ceramics,composites. High versatility
Heat transfer into the engine block	Low	Medium	Low
Reliability of the melting process	Medium	Very high	High
Control over formation of the melted particle	Medium	High (powder) Medium (wire)	High
Coating properties for cylinder bores	Medium	High	High
Process cost	Low	Very high	Low

- Residual porosity 2% typical
- Oxygen content 2% in carbon steel
- Microhardness $HV_{0.3}$ 350 - 550

Hardness strongly depends on the spray material. The lower limit is determined by coating performance; the upper limit is determined by the machinability costs of the coating. The minimum value of 30 MPa bond strength is obtained without problems. Figure 4 shows typical microstructures of a low alloyed carbon steel coating and composite sprayed using the plasma process in a cylinder bore of 80 mm diameter. During the coating deposition, the formation of FeO (wustite) and Fe_3O_4 (magnetite) can be controlled. These types of oxides also act as solid lubricants [8]. However the formation of Fe_2O_3 (hematite) should be avoided because of the abrasive effect of this type of oxide.

2.3 Nikasil Coating on Cylinder Liners

Electroplated Ni coatings with embedded hard, wear resistant particles offer a valuable method of extending life of Al cylinder walls in 2-and 4-cycle piston engines. Despite encouraging trials of such coatings in 1964 in the U.S. they had not adopted this process. Whereas, in Europe, about 1 million sq.dm of Al bearing surfaces have been coated with SiC particles dispersed in electroplated Ni. Watts and sulfamate baths have been used to describe the “Nikasil” (Nickel+ Silicon carbide) process which offers equal or better efficiency with fewer waste treatment problems. Size and quantity of particles influence wear results. Thickness of coatings ranges from 80 to 90 microns. Numerous investigations on test stands plus extensive practical experience have shown the “Nikasil” coated cylinders in 2-cycle, 4-cycle and diesel engines offer technical advantages over hard Cr plating. The friction co-efficients of Nikasil (Ni-P-SiC) plating contribute to superior

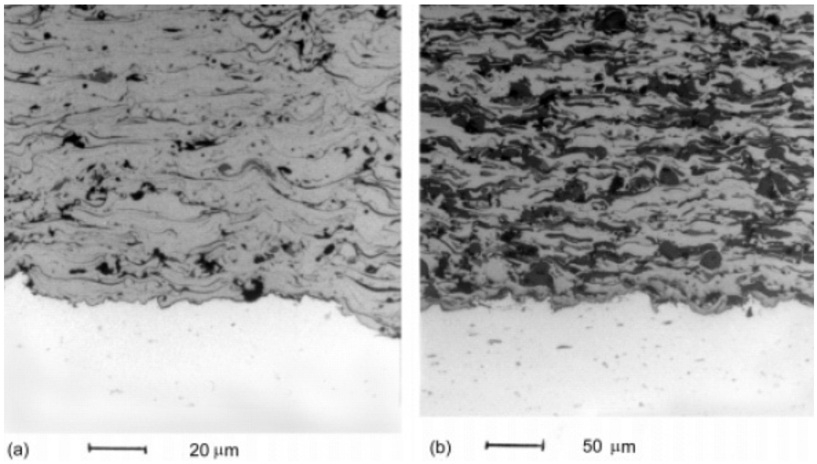


FIG. 4 TYPICAL INTERNAL COATINGS SPRAYED ON AL-SI CAST ALLOY. (a) LOW ALLOYED CARBON STEEL WITH BUILT-IN SOLID LUBRICANT ($HV_{0.3}$ = 400); (b) COMPOSITE METAL CERAMIC ($HV_{0.3}$ = 530)

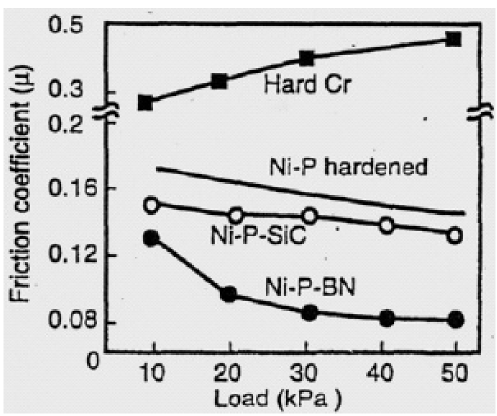


FIG. 5 FRICTION COEFFICIENTS OF VARIOUS NI COMPOSITE PLATINGS IN COMPARISON WITH CR PLATING

performance than hard chrome plating. Figure 5 illustrates the friction coefficient of nickel composite plating in comparison with hard chrome plating [9].

3. PISTON

In reciprocating machines, such as modern high speed internal combustion engines, it is desirable to reduce rubbing or sliding friction between the walls of the cylinder and the piston to the greatest possible extent. Such a reduction in friction not only results in decreasing the amount of wear between the parts but also increases the efficiency of the engine. Piston skirt/cylinder bore scuffing is one of the major contributors to the failure of automotive engines. Scuffing is usually accompanied by dramatic friction increase and rapid temperature rise that accelerates lubricant degradation. The contact and rubbing of the scuffed surfaces may cause vibration and noise, or even seizure of the interface. Figure 6 shows the wear scar on the liner sample due to the rubbing of the piston skirt [10].

Several coatings on the piston skirts are being attempted to overcome the friction and wear problems. As an initial attempt, poly tetra-fluroethylene (PTFE) was coated on the piston skirts which showed positive effect [11]. However, in recent times, the engine power has increased drastically resulting in increase in the cylinder temperature. Operating temperatures of cylinders

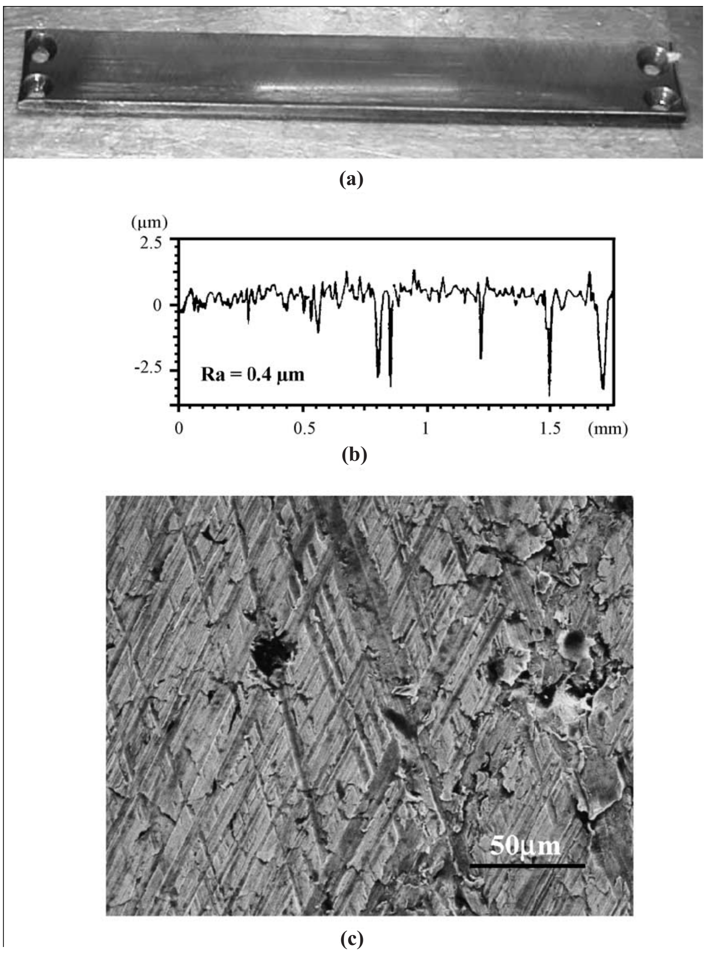


FIG. 6 A LINER SAMPLE AND ITS SURFACE PROFILE DUE TO FRICTION WITH PISTON SKIRT: (a) A LINER SAMPLE; (b) SURFACE PROFILE IN THE SLIDING DIRECTION; (c) AN SEM MICROGRAPH OF WORN LINER SURFACE AREA

above 260°C will lead to PTFE degradation and hence these coatings cannot be used on such systems.

Solid lubricant coating such as molybdenum disulfide is widely used as coating material for piston skirts. The coating is applied either by aerosol spray or by screen printing. Apart from MoS_2 coatings, there are several other solid lubricants including polyimide of 2,2-bis[4-(4-aminophenoxy)phenyl]hexafluoropropane ("4-BDAF") and an aromatic tetra carboxylic acid being attempted [12].

The other potential area of surface engineering in the piston is thermal barrier coating on the piston crowns. Future internal combustion engines will operate at higher temperatures and pressures than present day engines. For example, commercial diesel engines may operate at the cylinder temperatures of 760°C to about 860°C and brake mean effective pressure averaging about 1030 kPa. This will restrict the usage of Al-Si base piston which has low melting point. Figure 7 shows one such case in which the piston crown has melted due to excess heat produced inside the cylinder.

To operate in such conditions, critical engine parts must be insulated. Insulation lowers the temperature of the parts and reduces the amount of heat rejected to the environment. To be cost effective, the insulation should have a service life greater than about 20,000 hours. Thermal barrier coatings are proposed for such applications. There have been numerous research papers in recent years describing the theoretical benefits obtained from the use of ceramic components in reciprocating engines. Recently, successes have been reported and ceramic components are now in service in production engines, mainly for reduced in-cylinder heat rejection, in turn avoiding failures or a drop in engine performance [13].

4. PISTON RING

Among all the cylinder components, piston rings require maximum attention on surface engineering. In the cylinder assembly, piston rings are in contact with the cylinder wall and are

FIG. 7 FAILURE OF PISTON DUE TO EXCESS HEAT PRODUCED IN CYLINDER BORE. THE CROWN STARTED MELTING CAUSING ENGINE SEIZURE

moved back and forth. During the operation, the rings are prone to high wear along with friction. Hence, the rings require high wear-resistant coatings that possess low friction coefficients. Presently, there are a number of surface modification processes being practiced depending on the engine condition. The choice of the surface engineering strongly depends on the engine type and cost factor. For example, gas nitrided steel rings are very common for passenger cars operated with gasoline engines. At high temperatures, the gas nitriding rings had poor scuff resistance which resulted in development of scoring marks on the cast iron cylinder liner. In such cases, the thermal spray coatings containing molybdenum are most preferred. Mo is known for its excellent scuff resistance that can withstand the harsh conditions created in the diesel engine.

Following are the most preferred surface modifications for piston ring applications.

4.1 Electro Deposition

Hard chromium plating on the cast iron rings are the first stage surface engineering in the piston ring applications. Increased pollution norms restricted the usage of chromium plating baths in recent times. However, the chrome composite plating is getting much attention now. The composite plating is tailored by generating the cracks in the chrome plating and impregnating the hard oxide/nitride particle into the cracks. These composite chrome platings show improved wear resistance compared to the normal chrome plating. Figure 8 shows the optical micrograph of chrome composite coating in comparison with normal chrome plating. The cracks were generated by etching the coating at various stages. Hard particles such as alumina powder are added in the plating bath to incorporate them into the cracks.

4.2 Gas Nitriding(GN)/Liquid Nitriding

Gas nitrided piston rings are most commonly used in passenger car applications. For the gas nitriding process, special steels such as martensitic stainless steels are used to get the required

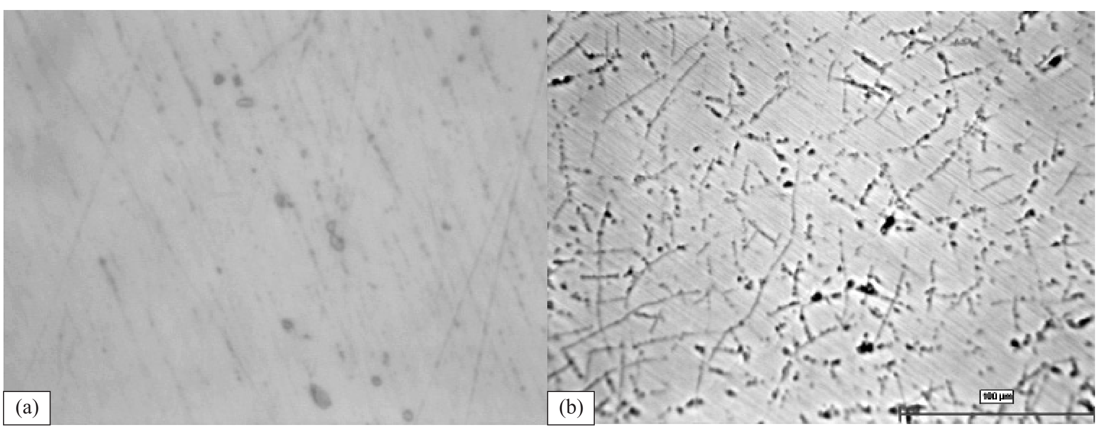


FIG. 8 OPTICAL MICROGRAPH OF (a) NORMAL HARD CHROME PLATING (b) COMPOSITE CHROME PLATING

surface hardness. The steel should contain any one of the nitralloying elements namely, Cr, Mo, V, and/or Al. During nitriding, the steel develops a brittle layer on the surface which is commonly known as white layer. The finished component should be free of this white layer.

4.3 Thermal Spray Process

Thermal spray processes such as flame, plasma and HVOF are employed for the piston ring coatings. Thermal spray coatings are proven to be vital surface engineering processes for automotive and aerospace applications [3,4]. The advantages of this technique include offering a wide range of materials with desired coating thickness, eco-friendliness compared to conventional electroplating and low operating time and cost compared to physical vapor deposition. Plasma spray process is widely used in the industrial sector due to low capital and operating cost with moderate coating characteristics such as bond strength and porosity. However, spray parameters required for each given coating depends on both the powder used for the coating and the base substrate [5].

The bond strength is very crucial for the plasma spray coatings. Poor bond strength leads to delamination of the coating in service. Plasma spray coatings employed for piston rings should ensure adequate bond strength even at the harsh operating environments. In our laboratory, extensive studies were carried out on preparation of the surface, powder composition and post finishing operation. We devised a test, namely twist fatigue life test to assess the bond strength of the coating. From our results, we concluded that to improve the bond strength, each and every stage of the coating and post finishing process plays a significant role. Figure 9 reveals the effect of different powder compositions on bond strength. Addition of nickel in the coating drastically improves the twist fatigue life of the rings. However, higher amount of nickel decreases the wear resistance of the coating. From our trials we identified the composition that results in excellent bond strength with improved wear resistance as well.

4.4 Physical Vapour Deposition

The future piston ring coatings solely depend on physical vapour deposition process. The process offers highest wear resistant coating with low friction coefficient. Several metal nitride coatings are being attempted and among them, chromium nitride is most suitable for the piston ring application as it has good corrosion resistance and low internal stresses. The coating thickness would ideally be 15 microns for passenger cars and 30-60 microns for commercial vehicles. The life expectancy of the PVD coating would be over three times the existing coatings. We have conducted the wear tests for various surface treatments using the ring-on-disc wear tester. Figure 10 represents the wear loss of ring coating as well as the cast iron disc which is used as the cylinder liner. Among the various surface treatments, PVD coatings showed the least wear on both ring as well as on disc.

5. PISTON PIN

Piston pins must have a high fatigue resistance as well as wear and seizure resistance during contact with the piston and the connecting rod. Presently, steels with carburizing or nitriding are being used for most of the commercial vehicles as well as for the passenger cars. These case hardened steels

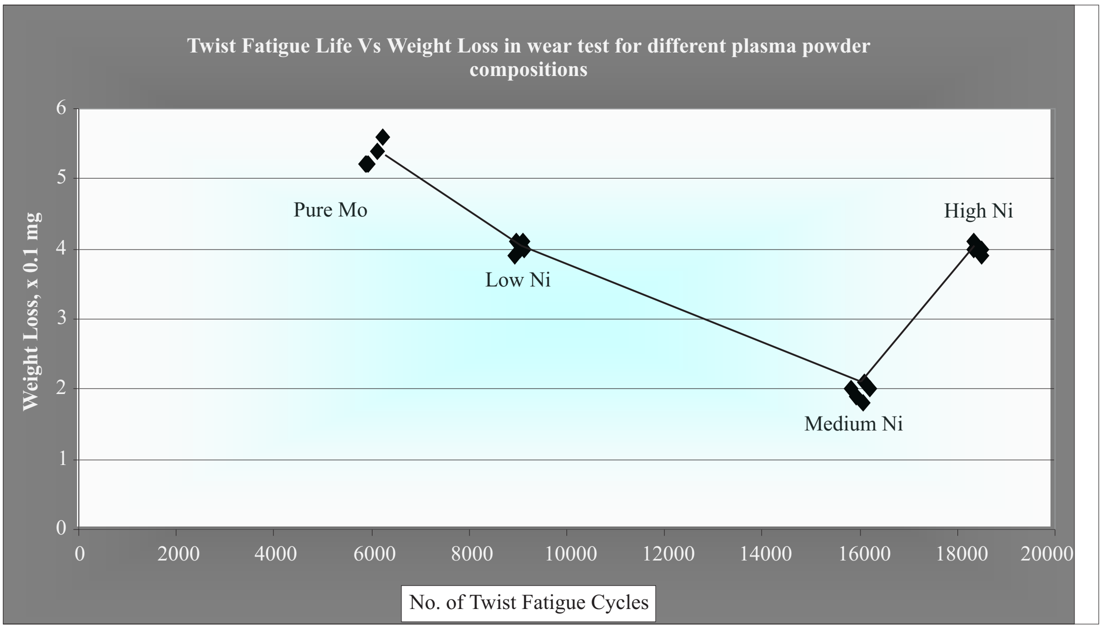


FIG. 9 TWIST FATIGUE LIFE VS WEAR LOSS IN WEAR TEST FOR DIFFERENT PLASMA POWDER COMPOSITIONS

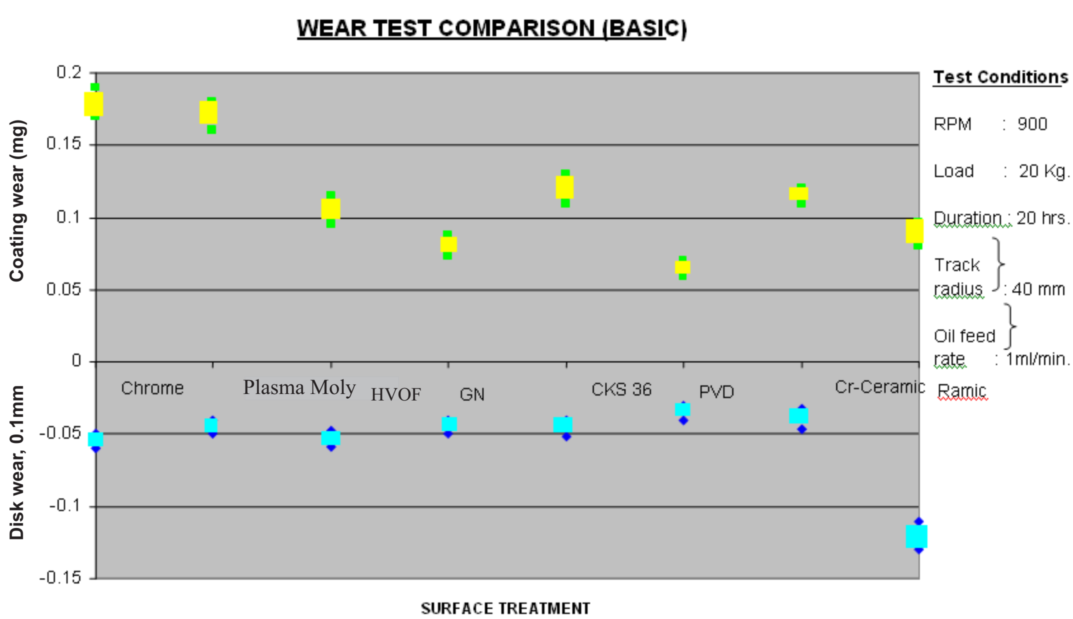


FIG. 10 WEAR TEST DATA FOR VARIOUS COATINGS USING RING-ON-DISC WEAR TESTING MACHINE

reveal optimum properties for achieving these requirements. Operating temperatures above 200°C seem to be critical for case-carburized steel, as high temperatures lead to a reduction of surface hardness and core strength. In such conditions, nitrided steel can be used because increased temperatures will not reduce the core strength or surface hardness.

Nevertheless, bushings in the bores are often necessary in order to attain acceptable tribological conditions. For both weight and cost reduction, engine manufacturers pursue the aim of superseding bushings. The resulting stress conditions and contact materials often make use of a coated piston pin necessary. Furthermore, a positive influence on the friction performance is expected. In the last decade, physical vapour deposition of Cr-N has been used in racing applications. However, today this technology is shifting to the ordinary vehicles as well, especially coatings of Diamond Like Carbon (DLC), which have proven their merit. These are being coated over the case carburized/nitrided piston pins.

A PVD-coated piston pin increases wear resistance, scuffing resistance, seizure resistance and dry-running properties while reducing friction. In recent years, apart from conventional metal nitrided (CrN, TiN) coatings, Mo-based high hardness coatings are being tried which have better wear resistance. Figure 11 shows the wear scars of the various coatings introduced in reciprocating wear tests. MoN coating showed better wear resistance than TiN coating. Also, addition of Cu and Ag drastically improved the wear resistance by showing the least wear scars. The wear volumes of balls decreased 2-3 fold with the addition of Cu and Ag. The width of the wear tracks on Mo-N-5% Ag coating was the narrowest of all the coatings indicating minimum wear on balls [14].

These results showed the important role of tribofilms on wear behavior. TiN with a hardness of 30 GPa could not resist the severe conditions. However MoN which is known to form lubricious molybdenum oxides performed satisfactorily. By the addition of copper to the coating, wear performance became better indicating better lubricating properties of molydenum-copper oxides (molybdates). In the case of 5% at silver addition, further improvement is achieved in the wear performance. Although this coating had a lower hardness of 33 GPa compared to all Mo-N-Ag-Cu series of coatings, it gave the best performance showing the more dominant role of tribo-films on wear. These results indicated that these coatings can be used for improving the tribological behaviour under fretting and galling conditions.

The unique chemical compositions of the coatings enable them to synergistically interact with the environment and lubricant additive packages to achieve low friction coefficients under boundary lubrication regimes – a feature that translates into direct energy savings and improved durability/reliability.

6. CONCLUSION

Cylinder components such as cylinder bore, piston, piston ring and pin require various surface engineering depending on their design and cost.

The need for surface engineering will increase with increase in the cylinder operating temperature and power, and requirements of reduced noise and fuel efficiency.

Though several coatings are practiced in the present time, the future trends for the coatings are likely to be as follows:

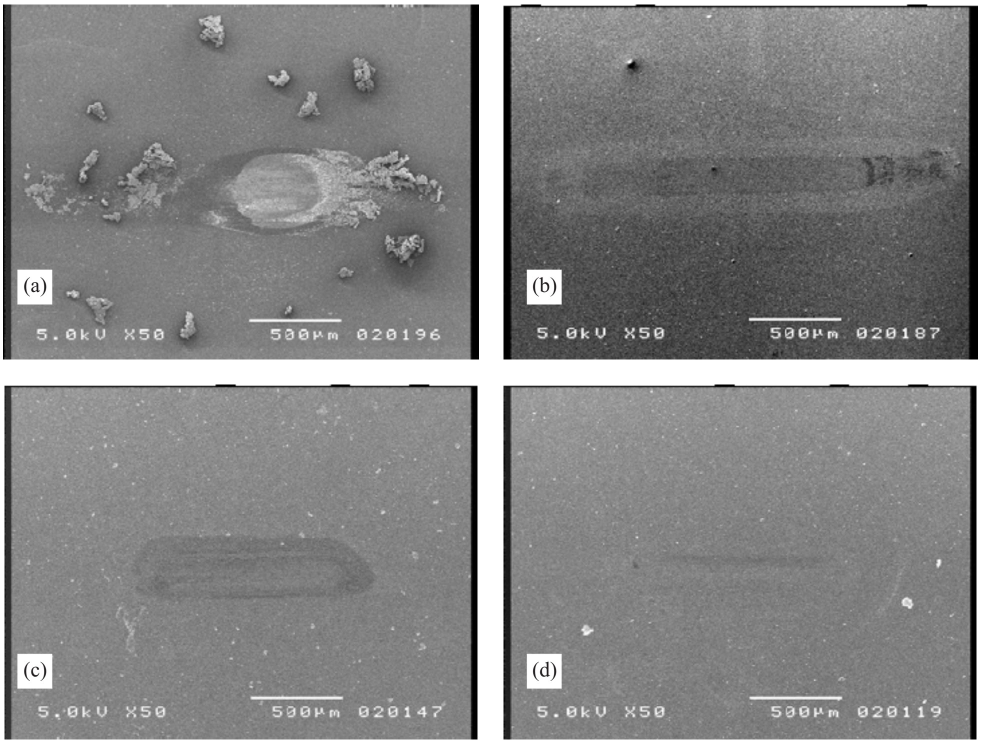


FIG. 11 SEM IMAGES OF WEAR SCARS AFTER RECIPROCATING TESTS A) TIN COATING, B) MON COATING C) MON-2 %CU COATING D) MON-5%AG COATING

- Cylinder bore → Thermal spray
- Piston → Thermal barrier coating on the crown and solid lubricant on the skirt
- Piston ring → Gas nitriding + Physical vapour deposition of ceramic/DLC
- Piston Pin → Nitriding + Physical vapour deposition of DLC

REFERENCES

[1] F. Zwein and A. Robota, Influence of cylinder bore topography on the oil consumption and emissions of a DI TCI diesel engine, Technical Symposium- Piston Technology Perspectives 1st February 2000.

[2] Barbezat G., Keller S., Wegner K.H.: Rotaplasma application in the automotive industry. ITSC 95, Kobe, 1995 proceeding, pp. 9-13.

[3] Mc Cune: Thermal spraying of cylinder bore surfaces for aluminum engine blocks. Welding Journal, August 1, 1995, pp. 41-47.

[4] Rao V.D.N., Kabat D.M., Rose R., Leong D.Y.: Performance of plasma spray coated bore 4.6 L-V 8 aluminum block engines in dynamometer and fleet vehicle durability tests. SAE paper no. 970008.

[5] Byrnes, Kramer M.: Method and apparatus for the application of thermal spray coating onto aluminum engine cylinder bores. Proceedings NTSC, Boston, 1994, p. 39.

[6] Barbezat G., Wuest G.: Advantages for automotive industry of plasma spray coating of AlSi cast alloy cylinder bores. Surface Engineering, London, 1998, 14, 2, pp. 113-116.

[7] Rao V.D.N., Cikaneth H.A., Wuest G.: Material systems for cylinder bore applications – plasma spray technology. SAE paper no. 970023.

[8] Barbezat G.: The state of the art of the internal plasma spraying on cylinder bore in AlSi cast alloys. Int. Journal of Automotive Technology, vol.2, no.2, June 2001, pp 47-52

[9] Donahue R. and Fabiyi P.A.: SAE Transactions: Journal of Materials & Manufacturing, 2000, 00M-89.

[10] Ye Z., Zhang C., Wang Y., Cheng H.S., Tung S., Wang Q. J., He X., Wear 257 (2004) 8-31.

[11] Schenkel J. - US Patent 4,987,865, 1991.

[12] Presswood J.K - US Patent 4,831,977, 1989.

[13] Chan S.H., Khor K.A. - Journal of Materials Engineering and Performance, 9(1), 2000, 103-109.

[14] Ürgen M., Production and Characterization of MeN-X Type Nanocomposite Coatings Produced by Cathodic Arc Based Hybrid Techniques, Presented at ASM, International Surface Engineering Congress and Exposition, August 2-4, 2004, Orlando Fl.

NUCLEAR POWER AND ENERGY TECHNOLOGIES

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

APPLICATION OF CONVECTIVE HEAT TRANSFER IN NUCLEAR REACTORS

S.K. GUPTA

TYPES OF NUCLEAR REACTORS AND THEIR BASIC COMPONENTS

The basic types of reactors are as follows:

a) Based on energy of neutrons causing fission: Thermal Reactor (Moderator is there), Fast Reactor (No moderator)
b) Based on core structure: Homogeneous, Heterogeneous
c) Based on the moderator used: Light water, Heavy water, Graphite, Organic
d) Based on the heat transport (HT) fluid used: Pressurized Light water, Pressurized Heavy Water, Boiling Light Water, Gaseous (CO_2 or He), Liquid Metals, Organic Liquid

Basic components of reactor are core, reflector, coolant, moderator and control rod etc. In India a wide variety of nuclear reactors are in operation and in different stages of construction. The main stay of Indian nuclear power programme today is 'Pressurized Heavy Water Reactors (PHWRs)'. These reactors are of CANDU (CANadian Deuteruim Uranium) type. There are 17 operating PHWRs and several others in different stages of construction. These reactors are either of 220 MWe or 540 MWe capacities. A typical Indian PHWR is heavy water moderated, pressurized heavy water cooled, natural uranium fuelled and horizontal pressure tube type reactor.

BASIC CORE DESIGN OF PHWR

The pressurized heavy water coolant passes through a tube called pressure tube (PT) inside which fuel bundles are located. In a typical 220 MWe Indian PHWR there are 12 fuel bundles inside a pressure tube. The pressure tube is surrounded by a concentric tube called calandria tube (CT). The annulus between PT and CT is filled with CO_2 to reduce the heat loss. Heavy water, which acts as both moderator and reflector, is contained in a horizontal vessel called calandria vessel .There are 306 channels (PT) in a 220MWe reactor. Figure 1 shows the schematic of core of a PHWR.

Each fuel bundle in 220 MWe PHWR contains 19 fuel pins. The length of each bundle is approximately half meter. The pins are arranged in rings having 1, 6 and 12 pins. To avoid contact in between the pins helical spacer are wrapped on the inner pins. To avoid contact of outer fuel pins with PT, bearing pads are provided on outer pins. The pins are held together by endplates on both the ends.

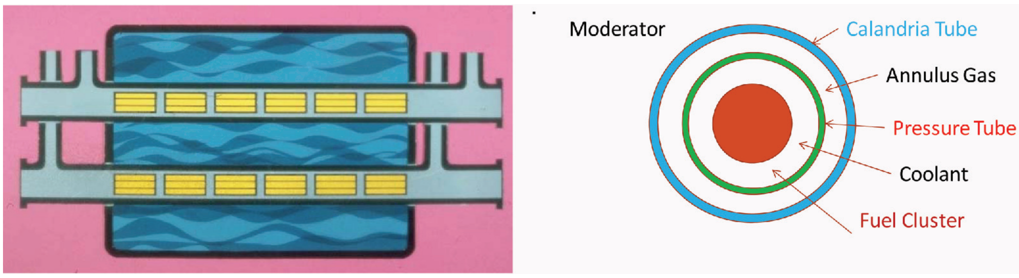


FIG. 1 SCHEMATIC OF CORE OF A PHWR

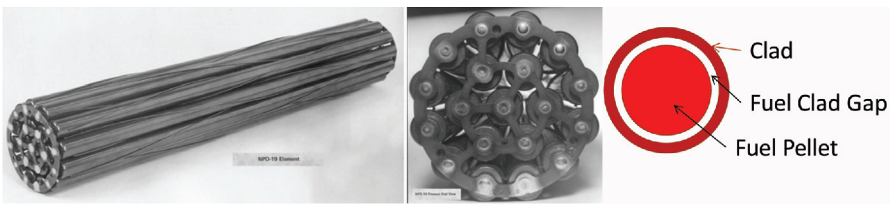


FIG. 2 FUEL BUNDLE DESIGN OF PHWR

Each fuel pin contains about 24 UO_2 pellets. The pellets are encapsulated by a tube called clad. The gap between pellets and clad is filled with helium. The gap accommodates the expansion of pellets. A typical fuel bundle is shown in Figure 2. The different materials of core components in PHWR are given in Table 1.

Table 1. *Core materials of PHWR*

Component	*Material*
Fuel Pellet	Natural UO_2
Fuel-Clad Gap	He
Clad, End plate	Zircaloy-4
Coolant	D_2O
Pressure Tube	Zr-2.5% Nb
Annulus Gas	CO_2
Calandria Tube	Zircaloy-2/Zircaloy-4
Moderator	D_2O

CONVERSION OF NUCLEAR ENERGY INTO ELECTRICAL ENERGY

The heat generated in fuel is ultimately converted to electrical energy in generator. The following schematic diagram (Figure 3) shows the process of conversion of nuclear energy into electrical energy in a CANDU PHWR.

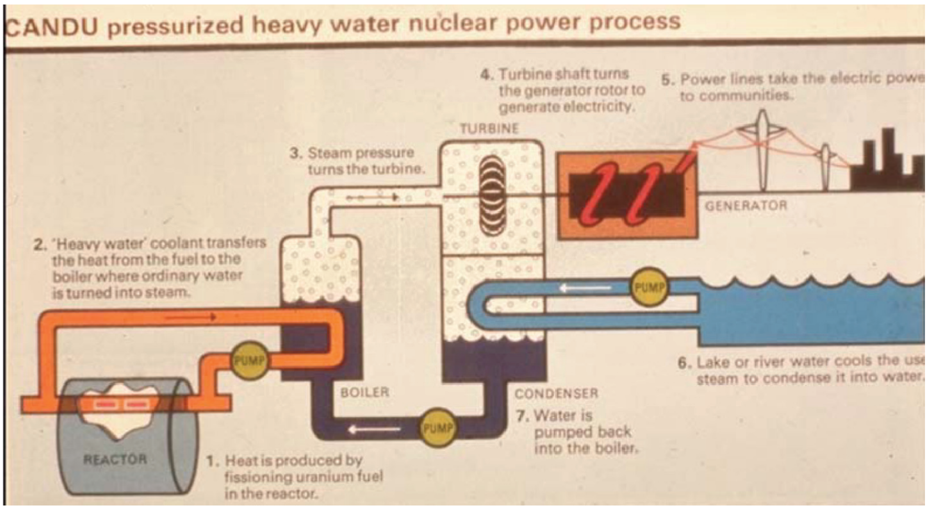


FIG. 3 CONVERSION OF NUCLEAR ENERGY TO ELECTRICAL ENERGY

In a PHWR heat is produced by fissioning of natural uranium fuel. Heavy water coolant transfers the heat from the fuel to the steam generator where ordinary water is converted into steam. The steam produced runs the turbine. The turbine shaft turns the generator rotor to generate electricity. The low pressure steam coming from turbine gets condensed in condenser and is pumped back into steam generator. The coolant in PHWR is pressurized and boiling of coolant is avoided in channels. But in boiling water reactors (BWR) where coolant is light water, the coolant gets converted into steam inside the core. In PHWR, the secondary water becomes steam in steam generator before entering turbine. To study the heat transfer process at different locations it is essential to have knowledge of heat transfer zones that occur during pool boiling.

BOILING CURVE

The boiling curve described below pertains to water at 1 atmosphere although similar trends characterize the behavior of other fluids. Surface heat flux q_S'' depends on the convection coefficient h, as well as on the excess temperature ΔT_e. Different boiling regimes may be delineated according to the value of ΔT_e.

Free Convection Boiling

Free convection boiling is said to exist if $\Delta T_e \le \Delta T_{e,A}$, where $\Delta T_{e,A} \approx 5°C$. The surface temperature must be somewhat above the saturation temperature in order to sustain bubble formation. As the excess temperature is increased, bubble inception will eventually occur, but below point A (referred to as the onset of nucleate boiling, ONB), fluid motion is determined principally by free convection effects.

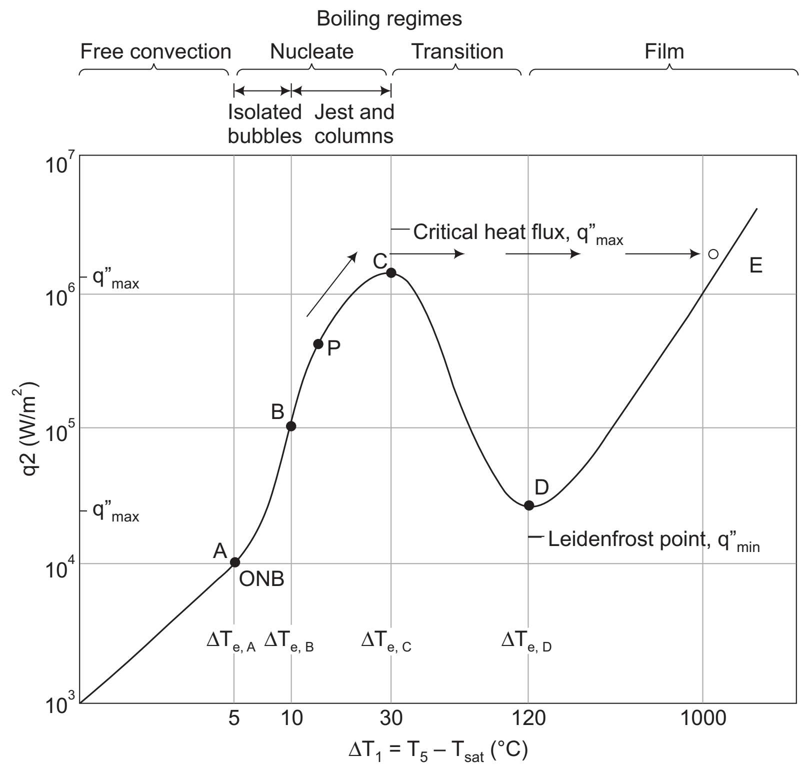


FIG. 4 TYPICAL BOILING CURVE FOR WATER AT 1 ATM: SURFACE HEAT FLUX q_S'' AS A FUNCTION OF EXCESS TEMPERATURE, $\Delta T_e = T_S - T_{sat}$ [5]

NUCLEATE BOILING

Nucleate boiling exists in the range $\Delta T_{e,A} \leq \Delta T_e \leq \Delta T_{e,C}$, where $\Delta T_{e,C} \approx 30°C$. In this range, two different flow regimes may be distinguished. In region A-B, isolated bubbles form at nucleation sites and separate from the surface. This separation induces considerable fluid mixing near the surface, substantially increasing h and q_S''. In this regime most of the heat exchange is through direct transfer from the surface to liquid in motion at the surface, and not through the vapor bubbles rising from the surface. As ΔT_e is increased beyond $\Delta T_{e,B}$ more nucleation sites become active and increased bubble formation causes bubble interference and coalescence. In the region B-C, the vapor escapes as jets or columns, which subsequently merge into slugs of the vapor. Interference between the densely populated bubbles inhibits the motion of liquid near the surface. Point P corresponds to an inflection in the boiling curve at which the heat transfer coefficient is maximum. At this point h begins to decrease with increasing ΔT_e, although q_S'' which is the product of h and ΔT_e, continues to increase. The maximum heat flux $q''_{s,C} = q''_{max}$ is usually termed the *critical heat flux* and in water at atmospheric pressure it exceeds 1 MW/m^2. At the point of this maximum, considerable vapor is being formed, making it difficult for liquid to continuously wet the surface. Because high heat transfer rates and convection coefficients are associated with small values of the excess temperature, it is desirable to operate many engineering devices in the nucleate boiling regime.

Transition Boiling

The region corresponding to $\Delta T_{e,C} \leq \Delta T_e \leq \Delta T_{e,D}$ where $\Delta T_{e,D} \approx 120°C$, is termed transition boiling, unstable film boiling, or partial film boiling. Bubble formation is now so rapid that a vapor film or blanket begins to form on the surface. At any point on the surface, conditions may oscillate between film and nucleate boiling, but the fraction of the total surface covered by the film increases with increasing ΔT_e. Because the thermal conductivity of the vapor is much less than that of the liquid, h (and q_S'') must decrease with increasing ΔT_e.

Film Boiling

Film boiling exists for $\Delta T_e \geq \Delta T_{e,D}$. At point D of the boiling curve, referred to as the *Leidenfrost* point, the heat flux is a minimum $q''_{s,D} = q''_{\min}$, and the surface is completely covered by a vapor blanket. Heat transfer from the surface to the liquid occurs by conduction and radiation through the vapor. As the surface temperature is increased, radiation through the vapor film becomes more significant and the heat flux increases with increasing ΔT_e.

In the foregoing discussion of the boiling curve, it is assumed that control may be maintained over T_s. However nuclear reactor involves controlling of q_S''. Let us consider starting at some point P in gradually increasing q_s''. The value of ΔT_e and hence the value of T_s will also increase, following the boiling curve to point C. However, any increase in q_s'' beyond this point will induce a sharp departure from the boiling curve in which surface conditions change abruptly from $\Delta T_{e,C}$ *to* $\Delta T_{e,E} \equiv T_{s,E} - T_{sat}$. Because $T_{s,E}$ may exceed the melting point of the solid, destruction or failure of the system may occur. For this reason point C often termed the burnout point or the boiling crisis, and accurate knowledge of the critical heat flux (CHF) is important. In nuclear reactor, instead of pool boiling, there exists forced convection boiling and hence it is important to study the flow regimes that occur during forced convection boiling.

FLOW REGIMES IN FORCED CONVECTION BOILING

In pool boiling fluid flow is due primarily to the buoyancy-driven motion of bubbles originating from the heated surface. In contrast, for forced convection boiling, flow is due to a directed (bulk) motion of the fluid, as well as to buoyancy effects. Conditions depend strongly on geometry, which may involve external flow over heated plates and cylinders or internal (duct) flow. Internal, forced convection boiling is commonly referred to as two-phase flow and is characterized by rapid changes from liquid to vapor in the flow direction.

Internal forced convection boiling is associated with bubble formation at the inner surface of a heated tube through which a liquid is flowing. Bubble growth and separation are strongly influenced by the flow velocity, and hydrodynamic effects differ significantly from those corresponding to pool boiling. The process is accompanied by the existence of a variety of two-phase flow patterns. Let us consider flow development in a vertical tube that is subjected to a

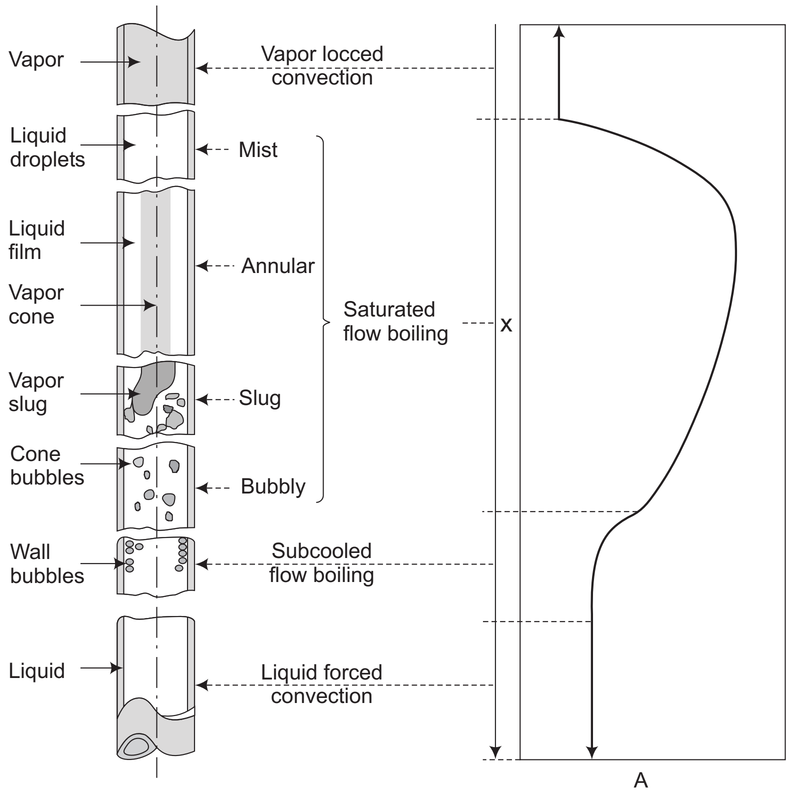


FIG. 5 FLOW REGIMES FOR FORCED CONVECTION BOILING IN A TUBE [5]

constant surface heat flux, as shown in Figure 5. Heat transfer to the sub-cooled liquid that enters the tube is initially by *single-phase forced convection*. Farther down the tube, the wall temperature exceeds the saturation temperature of the liquid, and vaporization is initiated in the *sub-cooled flow boiling region*. This region is characterized by significant radial temperature gradients, with bubbles forming adjacent to the heated wall and sub-cooled liquid flowing near the center of the tube. The thickness of the bubble region increases farther downstream, and eventually, the core of the liquid reach the saturation temperature of the fluid. Bubbles can then exist at any radial location, and the time-averaged mass fraction of vapor in the fluid (i.e., quality), X, exceeds zero at any radial location. This marks the beginning of the *saturated flow boiling region*. Within the saturated flow boiling region, the mean vapor mass fraction defined as

$$\overline{X} \equiv \frac{\int_{A_C} \rho u(r, x) X dA_C}{m}$$

increases and, due to the large density difference between the vapor and liquid phases, the mean velocity of the fluid, u_m, increases significantly. The first stage of the saturated flow boiling region corresponds to the *bubbly flow regime*. As $\overline{X}$ increases further, individual bubbles coalesce to form slugs of vapor. This *slug-flow regime* is followed by an *annular-flow regime* in which the liquid forms a film on the tube wall. This film moves along the inner surface of the tube, while vapor moves at a larger velocity through the core of the tube. Dry spots eventually appear on the inner surface of the tube and grow in size within a *transition regime*. Eventually, the entire tube surface is completely dry, and all remaining liquid is in the form of droplets that travel at high velocity within the core of the tube in the *mist regime*. After the droplets are completely vaporized, the fluid consists of superheated vapor in a *second single phase forced convection region*. The increase in the vapor fraction along the tube length, along with the significant difference in the densities of the liquid and vapor phases, increases the mean velocity of the fluid by several orders of magnitude between the first and the second single phase forced convection regions. The local heat transfer coefficient varies significantly as $\overline{X}$ and u_m decrease and increase, respectively, along the length of the tube, x. In general, the heat transfer coefficient can increase by approximately an order of magnitude through the sub-cooled flow boiling region. Heat transfer coefficients are further increased in the early stages of the saturated flow boiling region.

It is also important to study the effect of mass flux on the boiling curve. In forced convection regime, for the same heat flux the wall temperature is higher for low mass flux as shown in the Figure 2.6. (G = mass flux). In fully developed nucleate boiling, the wall temperature primarily is determined by heat flux and pressure and only slightly affected by liquid velocity.

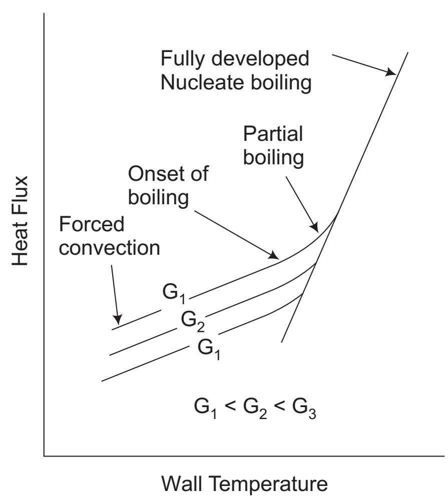


FIG. 6 INDEPENDENCE OF FULLY DEVELOPED NUCLEATE BOILING ON LIQUID VELOCITY [1]

Convective heat transfer and fluid flow are characterized by a number of dimensionless parameters. A set of mostly used parameters are provided below.

DIMENSIONLESS GROUPS

Reynolds Number (Re)

$$\text{Re} = \frac{\rho V L_c}{\mu} = \frac{V L_c}{v} = \frac{\text{inertial forces}}{\text{viscous forces}}$$

μ is the viscosity (kg/(s·m)) and v is the kinematic viscosity (m^2/s)

Prandtl Number (Pr)

$$\text{Pr} = \frac{v}{\alpha} = \frac{\text{momentum diffusivity}}{\text{mass diffusivity}}$$

α is the thermal diffusivity (m^2/s)

Schmidt Number (Sc)

$$Sc = \frac{v}{D_{AB}} = \frac{\text{momentum diffusivity}}{\text{mass diffusivity}}$$

D_{AB} is the mass diffusivity (m^2/s)

Nusselt Number (Nu)

$$\text{Nu} = \frac{hL}{k_f} = \left.\frac{\delta T^*}{\delta y^*}\right|_{y^*=0} = \text{dimensionless heat transfer coefficient}$$

Sherwood Number (Sh)

$$\text{Sh}_L = \frac{h_m L}{D_{AB}} = \left.\frac{\delta \rho_A{}^*}{\delta y^*}\right|_{y^*=0} = \text{dimensionless mass transfer coefficient}$$

Lewis Number (Le)

$$\text{Le} = \frac{a}{D_{AB}} = \frac{\text{thermal diffusivity}}{\text{mass diffusivity}} = \frac{Sc}{Pr}$$

The evaluation of pressure drops in two phase systems is complex. To simplify this following methodology is followed. The frictional component of two phase pressure gradients is often expressed in terms of a two-phase multiplier.

Two phase pressure gradient = single phase pressure gradient × two-phase multiplier

$$\left(-\frac{dp}{dz}\right)_F = \left(-\frac{dp}{dz}\right)_{lo} \phi^2{}_{lo}$$

Where $(-(dp/dz))_{lo}$ is a single phase frictional pressure gradient. The subscript *lo* means

- It is a liquid single-phase flow
- It is calculated at a liquid mass flux of G = total mass flux of liquid and gas in two-phase flow

ϕ^2_{lo} is the two-phase multiplier. There are different methods for evaluating ϕ^2_{lo}.

The calculation of heat transfer coefficients and other thermal-hydraulic parameters are based on correlations. A number of correlations are available. Care should be taken while choosing correlations.

CHOICE OF CORRELATIONS

The correlations for heat and mass transfer should be selected based on the following considerations.

- Range of application
- Developing/developed flow
- Geometry consideration
- Data points in the original development
- Trend outside the range
- Semi empirical based on phenomena

While using any correlation it should be noted that properties are calculated at the temperature mentioned by the developer.

LOCA (Loss of Coolant Accident) is one of the important accident conditions of PHWRs. During LOCA, steam is formed in the system. In water and steam system, the density of steam is very less than that of water. Hence steam is called void. The relation of void fraction and quality is important and is described below.

VOID FRACTION AND QUALITY

If the cross-sectional area of the channel is A and the cross-sectional areas occupied by the gas and liquid phases are A_g and A_f respectively then the void fraction is given by

$$\alpha = \frac{A_g}{A_f}$$

If the total mass flow rate is W which is the sum of liquid mass flow rate W_f and gas mass flow rate W_g, the mass quality, x, is defined as

$$x = \frac{W_g}{W_g + W_f}$$

Void formation is to be avoided in PHWRs because of its effect on reactivity. The definition of reactivity and the effect of void on reactivity are brought out in subsequent paragraphs.

REACTIVITY

When neutron multiplication factor $k = 1$ and the effects of source neutrons are negligible, the neutron flux and the power level will remain constant. It is important to note that a reactor may be critical at any power level. Under normal operational conditions, a reactor is operating close to criticality (that is, k is nearly equal to 1). It is more convenient under these circumstances to talk in terms of the amount by which k differs from 1 than it is to keep quoting the value of k itself. When we do a precise analysis of the equations that describe the behavior of the reactor power when k is changed, it turns out that the solutions involve a quantity, called reactor reactivity, which is defined by the relation.

$$\text{Reactivity} = \frac{k-1}{k}$$

If, as is normally the case, we are dealing with values of k which are close to 1, the above expression is nearly equal to $k - 1$ and we can use the following approximation:

Reactivity = $k - 1 = \Delta k$

Because the reactivity changes involved in normal reactor control are always quite small, they are measured in a smaller unit, the *milli-k* (abbreviated *mk*).

RELEVANCE OF VOIDS TO PHWR BASED NPPS

Voids will be formed if the moderator or the heat transport system fluid boils. Since void formation is more likely to occur in the coolant than in the moderator, the effects of loss of coolant are discussed. The possible causes of coolant boiling are low pressure (pipe rupture, pressurization system failure), low flow (blockage, pipe rupture, pump failure) or excess power (flux distortion, regulating system failure). Under these circumstances, the coolant will gradually be displaced by steam and eventually the channel (s) will become totally depleted of liquid coolant. The fuel channel causes a decrease in the moderation of neutrons in the immediate neighborhood of the fuel elements. When the channels are voided, there is no moderation so that-higher energy neutrons interact with fuel elements. The overall effect is that voiding the coolant results in a positive reactivity, which is greatest with fresh fuel. The total reactivity change for full core voiding is typically in the range of 7 to 13 mk. depending on the degree of fuel burn up.

FLOW INSTABILITY

Flow instability refers to flow oscillations of constant or variable amplitude. Flow oscillations can be aggravated when there is thermo-hydrodynamic coupling between heat transfer, void, flow pattern and flow rate. Flow oscillations are undesirable for several reasons:

- Sustained flow oscillations can cause undesirable forced mechanical vibration of components.
- Flow oscillations can cause system control problems of water-cooled reactors where coolant also acts as a moderator.
- Flow oscillations affect local heat transfer characteristics and boiling crisis.

The last point has far reaching effects on reactor safety. Reactor acts as a constant heat flux source. If CHF is exceeded once, the operating point on flow boiling may shift from nucleate boiling zone (Point '*P*') to film boiling zone (point '*E*') leading to a high fuel surface temperature which is unacceptable from safety considerations. The operating point does not shift back to '*P*' even when oscillations disappear. The ratio of operating heat flux to the CHF is called Critical Heat Flux Ratio (CHFR). MCHFR, the minimum value of CHFR is a regulatory requirement.

CRITICAL FLOW

A sudden change in fluid velocity causes a large pressure change that is seen as a pressure wave, across which there is discontinuity in pressure and velocity. The pressure wave, which can be either a compression wave or rarefaction wave, propagates at sonic velocity relative to flow. When a one-dimensional acoustic wave travels in a fluid of equal and opposite velocity, the wave becomes stationary with respect to earth. There exist critical flow at this point (fluid at sonic velocity) and downstream pressure signals can no longer be transmitted to the upstream fluid. Thus, critical flow rate is limited by the upstream and critical point conditions, but not by downstream conditions.

STRATIFIED FLOW

Stratified flow occurs at very low liquid and vapor velocities. The two phases flow separately with a relatively smooth interface.

Stratified flow condition exists in NPP at the following locations:

- Vapor pull through in a header where pipe branches exist
- Hydro accumulators including advanced versions
- Pressure tube small breaks
- Stagnation channel breaks in a channel:

Final – \\server\e\book work
3-06-10

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

APPLICATION OF RADIATIVE HEAT TRANSFER IN NUCLEAR REACTOR

S.K. GUPTA

THERMAL RADIATION

Radiative heat transfer is concerned with the exchange of thermal radiation energy between two or more bodies. Thermal radiation is defined as electromagnetic radiation in the wavelength range of 0.1 to 100 microns (which encompasses the visible light regime), and arises as a result of a temperature difference between two bodies. No medium is required between the two bodies for heat transfer to take place (as is needed by conduction and convection). The heat transferred into or out of an object by thermal radiation is a function of several components. These include its surface reflectivity, emissivity, surface area, temperature, and geometric orientation with respect to other thermally participating objects. In turn, an object's surface reflectivity and emissivity is a function of its surface conditions (roughness, finish, etc.) and composition.

EMISSIVE POWER

The total amount of energy flux emitted by a surface at a given temperature is the emissive power. It depends on the temperature of the surface, and the surface characteristics. At a defined temperature there is a maximum limit to the emissive power of a surface. The maximum emissive power at a given temperature is the black body emissive power.

ABSORPTANCE, REFLECTANCE AND TRANSMITTANCE

Suppose that a radiant heat flux, q, falls upon a translucent plate that is not black. A fraction, α, of the total incident energy, called the *absorptance or aborptivity,* is absorbed in the body; a fraction, ρ, called the *reflectance or reflectivity,* is reflected from it; and a fraction, τ, called the *transmittance or transmitivity*, passes through. Thus

$$\alpha + \rho + \tau = 1. \tag{1}$$

All radiant energy incident on a black body is absorbed, so that $\alpha_b = 1$ and $\rho_b = \tau_b = 0$.

MONOCHROMATIC EMISSIVE POWER

At a given temperature, a body emits a unique distribution of energy in wavelength. At each temperature, a black body yields the highest value of e_λ that a body can attain. The very accurate measurements of the black-body energy spectrum by Lummer and Pringsheim (1899) are shown in Figure 1. The locus of maxima of the curves is also plotted. It obeys a relation called *Wien's law:*

$$(\lambda T)_{e\lambda = \max} = 2898\ \mu m.K \tag{2}$$

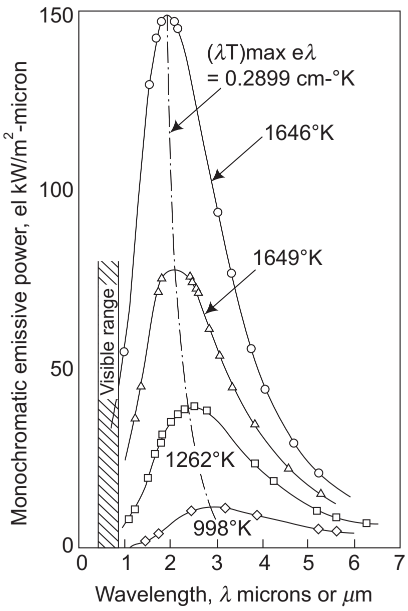


FIG. 1 MONOCHROMATIC EMISSIVE POWER OF A BLACK BODY AT SEVERAL TEMPERATURES [6]

In 1901, Max Planck predicted monochromatic emissive power of a black body as a function of λ and *T*. He found that

$$e_{\lambda b} = \frac{2\pi h c_0^2}{l^5[\exp(hc_0/k_B T\lambda) - 1]} \tag{3}$$

where c_o is the speed of light, 2.99792458×10^8 m/s; h is Planck's constant, $6.62606876 \times 10^{-34}$ J·s; and k_B is Boltzmann's constant, $1.3806503 \times 10^{-23}$ J/K.

EMITTANCE OR EMISSIVITY

A real body at temperature *T* does not emit with the black body emissive power e_b but rather with some fraction, ε, of e_b. The same is true of the monochromatic emissive power, $e_\lambda(\lambda, T)$, which is always lower for a real body than the black body value given by Planck's law. The symbol $e_\lambda(\lambda, T)$ designates the distribution function of radiative flux in λ, or the *monochromatic emissive powe*.

Thus monochromatic emittance, $\varepsilon\lambda$ is defined as

$$\varepsilon_\lambda = \frac{e_\lambda(\lambda, T)}{e_{\lambda_b}(\lambda, T)} \tag{4}$$

or the total emittance, ε defined as

$$\varepsilon = \frac{e(T)}{e_b(T)} = \frac{\int_0^\infty e_\lambda(\lambda, T)d\lambda}{\int_0^\infty e_{\lambda_b}(\lambda, T)d\lambda} = \frac{\int_0^\infty \varepsilon_\lambda e_{\lambda_b}(\lambda, T)d\lambda}{\int_0^\infty e_{\lambda_b}(\lambda, T)d\lambda} \tag{5}$$

For real bodies, both ε and $\varepsilon\lambda$ are greater than zero and less than one; for black bodies, $\varepsilon = \varepsilon\lambda = 1$. The emittance is determined entirely by the properties of the surface of the particular body and its temperature. It is independent of the environment of the body.

THE STEFAN-BOLTZMANN LAW

The monochromatic emissive power $e_\lambda(\lambda, T)$ and total emissive power $e(T)$ are related as follows:

$$e_\lambda(\lambda, T) = \frac{de(\lambda, T)}{d\lambda} \tag{6}$$

$$\Rightarrow \qquad e(\lambda, T) = e_\lambda(\lambda, T)d\lambda$$

$$\Rightarrow \qquad e(T) = E(\infty, T) = e_\lambda(\lambda, T)d\lambda$$

Putting the value of monochromatic emissive power from Planck's law, the total emissive power of a black body becomes

$$e_b(T) = \sigma T^4 \tag{7}$$

This is called Stefan-Boltzmann law.

σ = Stefan-Boltzmann constant = 5.670400 × 10+8 W/m^2·K^4 and T is the absolute temperature

SHAPE FACTOR

Radiative heat transfer rate between two black bodies can be calculated by the equation stated below.

$$Q_{1\text{–}2} = \sigma A_1 F_{1\text{–}2}(T_1^4 - T_2^4) \tag{8}$$

Where: A_1 = surface area of back body 1

$F_{1\text{–}2}$ = shape factor, which depends on the spatial arrangement of the two objects (dimensionless)

Similarly heat transfer from black body 2 to 1 is as follows

$$Q_{2\text{–}1} = \sigma A_2 F_{2\text{–}1}(T_2^4 - T_1^4) \tag{9}$$

Since $Q_{1\text{–}2} = -Q_{2\text{–}1}$, it follows that

$$A_1F_{1-2} = A_2F_{2-1} \tag{10}$$

This is called law of reciprocity of shape factor.

ELECTRICAL ANALOGY FOR GRAY BODY HEAT EXCHANGE

To derive the electrical analogy of heat exchange between two gray bodies the following two terms are defined:

H (W/m^2) ≡ *irradiance = flux of energy that irradiates the surface*

B (W/m^2) ≡ *radiosity = total flux of radiative energy away from the surface*

The radiosity can be expressed as the sum of the irradiated energy that is reflected by the surface and the radiation emitted by it. Thus,

$$B = \rho H + \varepsilon e_b \tag{11}$$

The net heat flux leaving any particular surface is the difference between B and H for that surface. Hence,

$$q_{net} = B - H = B - \frac{B - \varepsilon e_b}{\rho} \tag{12}$$

$$q_{net} = \frac{e}{\rho} e_b - \frac{1-\rho}{\rho} B \tag{13}$$

If the surface is opaque $(\tau = 0)$, $1 - \rho = \alpha$, and if it is gray, $\alpha = \varepsilon$. Then,

$$q_{net}A = Q_{net} = \frac{e_b - B}{\rho / eA} = \frac{e_b - B}{(1-\varepsilon)/\varepsilon A} \tag{14}$$

This equation is a form of Ohm's law, where $(e_b - B)$ can be viewed as a driving potential for transferring heat away from a surface through an effective surface resistance, $(1+ \varepsilon)/\varepsilon A$.

Let us consider heat transfer from one infinite gray plate to another parallel to it. Radiant energy flows past an imaginary surface, parallel to the first infinite plate and quite close to it, as shown as a dotted line in Figure 2.

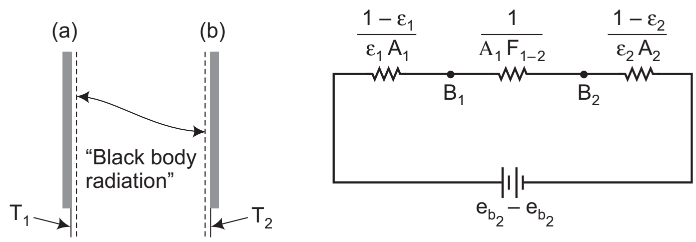


FIG. 2 THE ELECTRICAL CIRCUIT ANALOGY FOR RADIATION BETWEEN TWO GRAY INFINITE PLATES. [6]

If the gray plate is diffuse, its radiation has the same geometrical distribution as that from a black body, and it will travel to other objects in the same way that black body radiation would. Therefore, the radiation leaving the imaginary surface (the radiosity) can be treated as though it were black body radiation travelling to an imaginary surface above the other plate.

Hence,

$$Q_{net1-2} = A_1F_{1-2}(B_1 - B_2) = \frac{B_1 - B_2}{\left(\dfrac{1}{A_1F_{1-2}}\right)} \tag{15}$$

This is also a form of Ohm's law: the radiosity difference ($B_1 - B_2$), can be said to drive heat through the geometrical resistance, $1/A_1F_{1-2}$, that describes the field of view between the two surfaces.

When two gray surfaces exchange radiation only with each other, the net radiation flows through a surface resistance for each surface and a geometric resistance for the configuration. The electrical circuit shown in Figure 3.2 expresses the analogy and provided means for calculating Q_{net1-2} from Ohm's law. Recalling that $e_b = \sigma T^4$, we obtain

$$Q_{net1-2} = \frac{eb_1 - eb_2}{\sum \text{resistances}} = \frac{\sigma(T_1^4 - T_2^4)}{\left(\dfrac{1-\varepsilon}{\varepsilon A}\right)_1 + \dfrac{1}{A_1F_{1-2}} + \left(\dfrac{1-\varepsilon}{\varepsilon A}\right)_2} \tag{16}$$

RADIATIVE HEAT TRANSFER IN NUCLEAR POWER PLANTS

Radiative heat transfer is significant when convective cooling is significantly reduced and fuel rods are dry. Such situations occur during severe accidents only. It is required that even when the probability of such situations is extremely low the core configuration be estimated so that appropriate emergency management actions be taken to mitigate/limit the consequences. In PHWR when the coolant channels are full of voids and the temperature is high, radiative heat transfer becomes prominent. For calculating the temperature of fuels, pin to pin interaction is considered. There are 19 pins in 220 MWe PHWR. The evaluation of shape factors for one pin to the other is complex process. Different techniques are available for calculating the shape factors.

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

APPLICATION OF CONDUCTION HEAT TRANSFER TO NUCLEAR REACTOR

S.K. GUPTA

THERMAL CONDUCTION

Thermal conduction in solids is governed by Fourier's law of heat conduction. It states that the heat flow per unit normal area in any direction is proportional to the temperature gradient in that direction. The constant of proportionality is known as the transport property of the medium and is usually represented by 'k'. The unit for 'k' is w/m K. The heat flow per unit normal area is known as heat flux in that direction. It is a vector quantity and the unit is w/m^2. The gradient is represented by K/m. Conduction is primarily controlled by the random molecular and atomic motions of different types. The Fourier's law is mathematically expressed as:

$$q = -kA\frac{dT}{dx} \tag{1}$$

Where 'q' is heat flow in any direction, A is area normal to the direction of heat flow; dT/dx is temperature gradient in the direction of heat flow. The minus sign in the equation indicates that temperature gradient and heat flow are in opposite direction.

$$q'' = q/A, \text{ where } q'' \text{ is heat flux}$$

The heat transfer processes are of two types. In the first type, the temperature difference across two surfaces is maintained constant and heat flux is controlled by it. In the other type, the heat flux is maintained constant and the temperature difference gets controlled by it. In practical systems, it is a combination of the two. The heat transfer through the walls of the tube of a heat exchanger is dominantly, temperature controlled. On the other hand, the temperature difference in the fuel of a nuclear reactor is dominantly heat flux controlled.

Thermal conductivity though a property of the material depends on the temperature of the material. For ceramics it generally decreases with temperature. However, for fuel materials of a nuclear reactor such as uranium dioxide it again starts increasing with temperature at higher temperature. A schematic of variation of thermal conductivity of UO_2 with temperature is given in Figure 1. For UO_2, a commonly used fuel for thermal reactors, thermal conductivity is also a function of fuel burn up. Fuel burn up is a measure of how much energy has been obtained from the fuel due to fission processes. The dependence on burn up is relatively weak and is at times neglected for low burn up fuel. If 'k' is assumed to be constant (as is the case in most of the application where temperature variation is not significant) then equation (1) can be integrated and rewritten as:

$$\frac{T_2 - T_1}{q} = \frac{l}{kA} = R_T \tag{2}$$

Where T_2 and T_1 are the temperature of two surfaces of a rectangular prism. 'l' is the normal distance between two surfaces. The term 'l/kA 'is known as thermal resistance between the two surfaces of a rectangular prism with length of prism being 'l'. This equation is similar to the equation (1.3) developed by Ohm's law for linear electrical networks. Consequently thermal conduction networks with constant 'k' follow all the laws of linear electrical network.

$$\frac{V}{l} = R \tag{3}$$

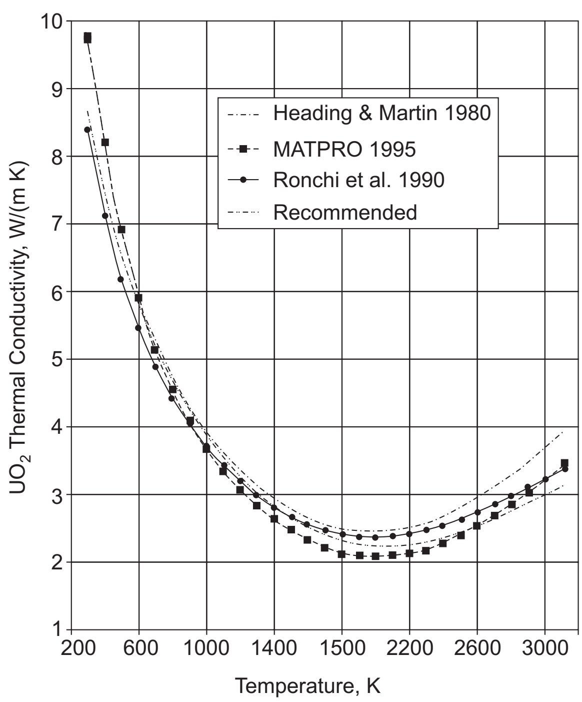


FIG. 1 THERMAL CONDUCTIVITY OF UO_2 [7]

On the basis of similarity it can be seen that voltage '*V*' corresponds to heat flow in the normal direction to the surfaces. A similar analogy can be drawn with hydraulic networks where pressure difference corresponds to temperature difference, flow corresponds to heat flow and frictional resistance corresponds to thermal resistance. The major difference between the two is that hydraulic resistances are highly non-linear.

The Fourier equation can be written in cylindrical coordinates as follows:

$$q = -k.2\pi rl.\frac{dT}{dr} \tag{4}$$

Integrating the equation and putting it in thermal resistance form leads to the following equation:

$$\frac{T_2 - T_1}{q} = \frac{\ln(r_2/r_1)}{2\pi/k} = R_1 \tag{5}$$

Where '*l*' is the length of the cylinder, r_2 and *r1* are the outer and inner radius of the cylinder.

Similarly for a hollow sphere thermal resistance is given by,

$$R_T = \frac{1}{4\pi k}\left[\frac{1}{r_1} - \frac{1}{r_2}\right] \tag{6}$$

If the thermal conductivity of a material is constant then, the variation of temperature in the rectangular, cylindrical and spherical wall can be easily derived. A typical temperature variation is given in Figure 2.

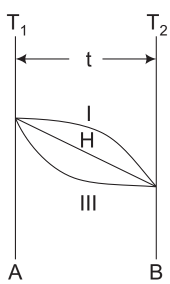


FIG. 2(a)

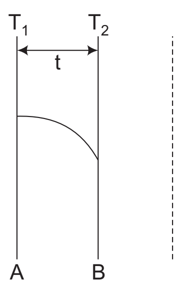


FIG. 2(b)

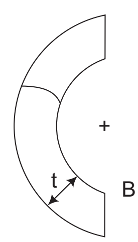


FIG. 2(c)

For the same gradient at wall '*A*', the gradient is smallest at *B* in Fig. 2(a) while it is highest at 2(c). It should be noted that if in Fig. 2(a), the observed temperature variation in the wall is either type I or II or III, it indirectly indicates the type of variation in thermal conductivity with temperature. In case I, it decreases with temperature, in case II it is independent of temperature and in case III it increases with temperature.

It should be noted that since thermal **linear** networks and electrical **linear** networks are similar, the Kirchhoff's laws can be applied to both the networks. The equivalent resistance for a set of resistances in series or in parallel or in any combination of these can be calculated in a similar manner as that for electrical networks.

Figure 3(a) shows thermal slabs in series. Their equivalent resistance R_{Teq} is

$$R_{Teq} = \frac{l_1}{k_1 A} + \frac{l_2}{k_2 A} + \frac{l_3}{k_3 A} \tag{7}$$

Figure 3(b) shows thermal slabs in parallel and their equivalent resistance R_{Teq}:

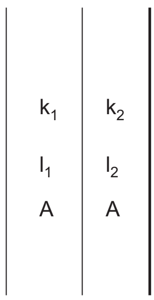


FIG. 3(a)

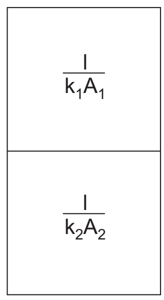


FIG. 3(b)

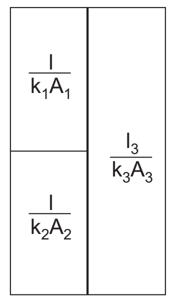


FIG. 3(c)

$$R_{Teq} = \frac{\frac{l_1}{k_1 A_1} \cdot \frac{l_2}{k_2 A_2}}{\frac{l_1}{k_1 A_1} + \frac{l_2}{k_2 A_2}} \quad \text{Where } l_1 = l_2 = l \tag{8}$$

Similarly Figure 3(c) shows thermal slabs in parallel-series combination and their equivalent resistance is

$$R_{Teq} = \frac{\frac{l_1}{k_1 A_1} \cdot \frac{l_2}{k_2 A_2}}{\frac{l_1}{k_1 A_1} + \frac{l_2}{k_2 A_2}} + \frac{l_3}{k_3 (A_1 + A_2)} \quad \text{Where } l_1 = l_2 = l,\ A_1 + A_2 = A_3 \tag{9}$$

In steam generators the tubes are designed to withstand a large pressure difference across it. Consequently the tubes are relatively thick. The thermal resistance of these tubes is very large as compared to the convective heat flow resistance on the two sides of the tubes. The total thermal resistance for the heat flow from primary fluid to secondary fluid consists of thermal conduction resistance of the tubes in series with convective thermal resistance on the two sides **(inner and outer)** of the tube. Any attempt to reduce the total resistance should aim at reducing conduction resistance. Consequently no fins are provided on steam generator tubes as these fins reduce onl convective resistance on the two sides of the tube.

In nuclear components, the heat is generated because of either absorption of radiation as in case of shield or because of the presence of radioactive fission products. This source of heat is generally volumetric in nature. However, there are also cases where surface heat generation sources are created as is the case in metal water reaction. In such case Fourier's equation may not be sufficient. An elementary differential equation with volumetric sources of heat generation may have to be derived. This steady state differential equation with volumetric sources of heat generation for rectangular, cylindrical and spherical geometry is given below (for constant thermal conductivity).

$$k\left(\frac{\delta^2 T}{\delta x^2}+\frac{\delta^2 T}{\delta y^2}+\frac{\delta^2 T}{\delta z^2}\right)+q'''=0$$

$$k\left(\frac{\delta^2 T}{\delta r^2}+\frac{1}{r}\frac{\delta T}{\delta r}+\frac{1}{r^2}\frac{\delta^2 T}{\delta \theta^2}+\frac{\delta^2 T}{\delta z^2}\right)+q'''=0 \tag{10}$$

For cases where 'k' is not constant the equation in rectangular system becomes

$$\frac{\delta}{\delta x}\left(k\frac{\delta T}{\delta x}\right)+\frac{\delta}{\delta y}\left(k\frac{\delta T}{\delta y}\right)+\frac{\delta}{\delta z}\left(k\frac{\delta T}{\delta z}\right)+q'''=0 \tag{11}$$

For cases where 'k' is a function of temperature alone the equation reduces to

$$\frac{\delta k}{\delta t}\left(\left(\frac{\delta T}{\delta x}\right)^2+\left(\frac{\delta T}{\delta y}\right)^2+\left(\frac{\delta T}{\delta z}\right)^2\right)+k\left(\frac{\delta^2 T}{\delta x^2}+\frac{\delta^2 T}{\delta y^2}+\frac{\delta^2 T}{\delta z^2}\right)+q'''=0 \tag{12}$$

The steady state differential equation for heat conduction can be easily derived by applying Fourier's law of heat conduction to different faces of a rectangular prism with elemental dimensions and subsequently make these dimensions tend to zero.

The volumetric source of heat generation in a nuclear reactor depends of the neutron flux causing fissions. Normally it is assumed to be proportional to the distribution of neutron flux causing most of the fissions. Strictly speaking this may not be true. In any reactor the fissions are caused by all fast, thermal and epithermal neutrons and the distribution of each type may be significantly different from the other. In a thermal reactor, the thermal neutrons causing most of the fissions undergo a depression towards the centre of a fuel pin as shown in Figure 4(a). On the other hand for a fast reactor, the fast neutron population is highest at the centre of the fuel pin as shown in Figure 4(b). This is because the fuel acts as a strong sink for thermal neutrons and a strong source of fast neutrons. The distributions themselves are a strong function of the location of fuel rod inside the reactor, the fuel burn up and reactor power. Different types of distributions generally possible in a reactor are shown in Figure 4. These distributions are generally difficult to estimate. Consequently it is assumed the volumetric heat generation rate q''' is uniform at any cross-section of fuel as shown in Figure 4(c). The assumption is generally conservative for thermal reactors and is not so optimistic for fast reactors. Later in this chapter it will be shown how one can estimate if this assumption is conservative.

For estimation of temperature distribution the differential equation for conduction needs to be integrated. Boundary conditions need to be provided. The boundary conditions could be of three types.

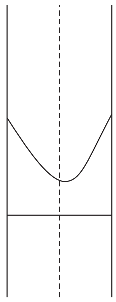


FIG. 4(a) THERMAL REACTORS

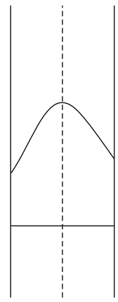


FIG. 4(b) FAST REACTORS

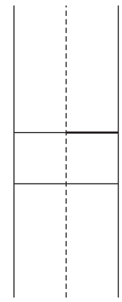


FIG. 4(c) UNIFORM HEAT GENERATION

(i) Dirichlet Boundary Condition

The temperature at the boundaries is maintained constant. This condition is normally applicable if a phase change is taking place which occurs at constant temperature. This is represented by

$$T_{\text{surface}} = Constant$$

(ii) Neumann Boundary Condition

The heat flux at the boundary is maintained constant. This is normally the case for a steady constant heat flux processes such as taking place in a nuclear fuel pin surface.

This is represented by

$$q''_S = -k\frac{dT}{dx}\bigg|_{\text{surface}} = Constant.$$

q_S'' equal to zero represents adiabatic or insulated surfaces. This condition can also be applied at the centre of fuel pin under symmetric conditions. At the interface of two materials of different properties, the Neumann Boundary Condition can be applied in the following form

Heat flux from left side of the interface = Heat flux from the right side of the interface

The above statement is valid as there is no heat accumulation/sink/source at the surface.

(iii) Cauchy's Boundary Conditions

This is convective surface boundary condition where heat flux at the surface of a solid cooled by a liquid is proportional to the temperature difference between surface temperature and liquid temperature.

TEMPERATURE DISTRIBUTION IN A FUEL ROD (STEADY STATE)

Almost all current power reactors use cylindrical fuel rods. There is a need to calculate the temperature distribution in a fuel rod (pin) at maximum operating power to estimate highest temperature and stored heat in the fuel. Significance of stored heat from reactor safety consideration is explained later in this chapter. Highest temperature is used to estimate the safety margins from fuel melting considerations. During reactor operation the effective melting point does not change significantly due to the presence of fission products. However, during accidents where bulk of fuel sees higher temperatures, a change in melting point is foreseen due to diffusion of oxygen and zircaloy. This is being discussed later.

Following are some of the assumptions made in the derivation of expression for temperature distribution.

ASSUMPTIONS

1) 1-D differential equation in radial direction for conduction is a suitable model. This is because the temperature distribution is generally symmetric about the axis of the fuel rod due to symmetric boundary conditions and uniform source of volumetric heat generation (see assumption 2). The axial heat conduction is a very small as temperature gradients encountered in radial direction are order of magnitude higher than in axial direction. However, this assumption may not be valid during accident conditions where fuel temperatures are high and rewetting of fuel is progressing in axial direction.
2) The volumetric source of heat generation is assumed to be uniform. As is shown later in this chapter, this is conservative assumption for thermal rectors and may not be strictly conservative for fast reactors.
3) Thermal conductivity of fuel is assumed to be function of temperature alone and isnot a function of burn up of fuel. The assumption is valid for the commonly used fuel such as uranium dioxide. Figure 5 shows the variation with temperature at different burn up for UO2. It is observed that '*k*' is a weak function of burn up.

1-D Differential Equation for a Fuel Rod

$$\frac{1}{r}\frac{\delta}{\delta r}\left(kr\frac{\delta T}{\delta r}\right) + q''' = 0 \tag{13}$$

Integrating it

$$Kr\frac{dT}{dr} + q'''\frac{r^2}{2} + C_1 = 0$$

or

$$k\frac{dT}{dr} + q'''\frac{r}{2} + \frac{C_1}{r} = 0 \tag{14}$$

or

$$\int k\delta T + q''' + C_1 \ln r + C_2 = 0 \tag{15}$$

If the distribution is symmetric about central line then at $r = 0$, $\frac{\delta T}{\delta r} = 0$

Substituting this is in equation (14) $C1 = 0$.

Similarly substituting Dirichlet boundary condition in equation (15), we get

$$C_2 = -q''' \frac{a^2}{4}$$

On integrating between surface temperature T_S and temperature T_r we get

$$k(T_r - T_S) = \frac{q'''}{4}(a^2 - r^2)$$

Maximum temperature drop in a fuel pin is from surface to centre

$$T_0 - T_S = \frac{q'''}{4k}(a^2) \tag{16}$$

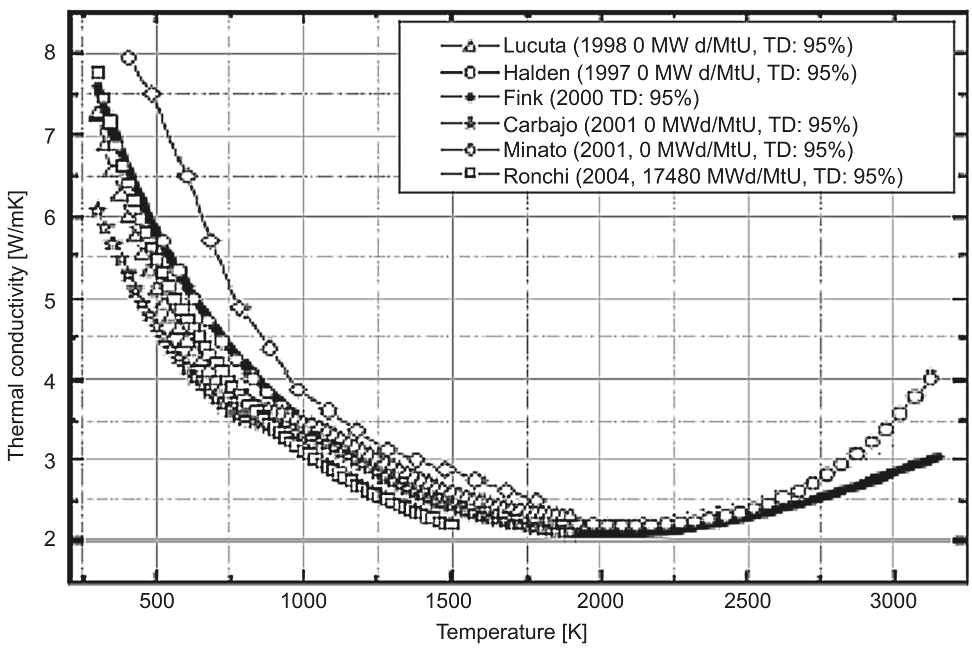


FIG. 5(a) VARIATION OF THERMAL CONDUCTIVITY WITH TEMPERATURE AT ZERO BURNUP [8]

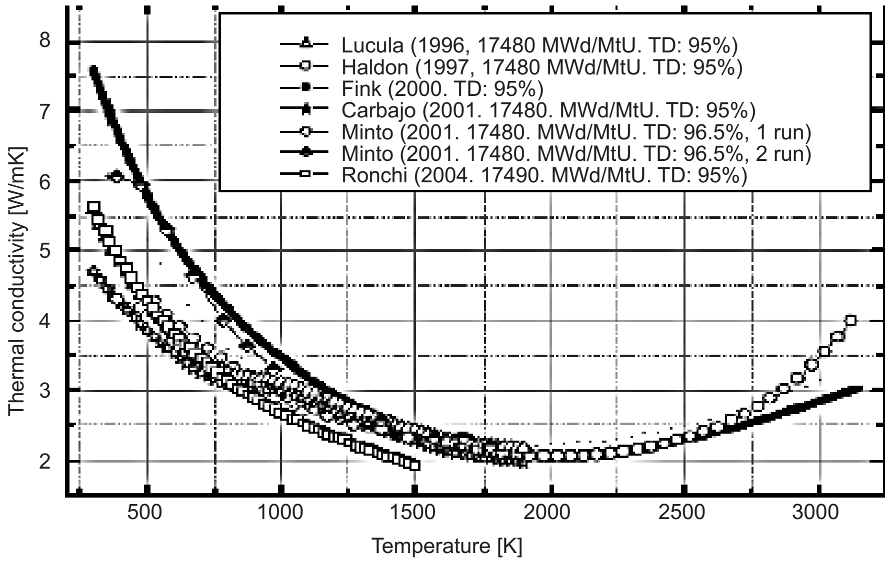


FIG. 5(b) VARIATION OF THERMAL CONDUCTIVITY WITH TEMP. AT 17480 MWD/MTU BURNUP [8]

It is prudent at this stage to define certain terms in context of fuel rods.

(1) Linear heat generation rate (LHGR): It is the amount of heat generated per unit length of the fuel rod. It is represented by 'q_L'. At any point along the length of the fuel

$$q_L = \pi a^2 q'''(x)$$

Where $q'''(x)$ is assumed uniform volumetric source of heat generation along length at point 'x'. Expressing equation (1.16) in terms of LHGR

$$T_0 - T_S = \frac{q_L}{4\pi k} \tag{17}$$

It is observed from equation (17) that the temperature drop (radial) depends only on the linear heat generation rate of the fuel and does not depend on fuel radius.

Eq. (15) can be rewritten in the following form after substituting the values for C1 and C2

$$\int kdT = \frac{1}{4} q'''(a^2 - r^2)$$

or

$$\int_{T_s}^{T_0} kdT = \frac{q_L}{4\pi} = \int_0^{T_0} kdT - kdT \tag{18}$$

Thus integral of kdT is a measure of linear heat generation rate. Earlier this concept was used for measuring temperature drop in the fuel pin. $\int kdT$ Vs. temperature graph for a fuel can be plotted by integrating 'k' with respect to temperature. It is nothing but the area under the graph 'k' vs. 'T'. From this graph fuel control line temperature can be obtained.

In the Fig. 6, the surface temperature T_S is taken on temperature axis and corresponding $\int kdT$ is found out. If LHGR is known then $\int kdT$ at T_0 is found from equation (18). Once $\int kdT$ at center is known, the corresponding center line temperature can be found from the graph. Similar procedure can be adopted for estimating radial temperature distribution in the fuel rod. From Figure 14. read $\int kdT$ at surface. Calculate the heat produced by the fuel up to distance 'r' from the centre ($\pi r^2 q'''$). Subtract it from LHGR ($\pi a^2 q'''$). Divide this by 'π' and add the resultant to $\int kdT$ at surface. This gives $\int kdT$ at 'r'. Find temperature T_r corresponding to this $\int kdT$ from the figure. This is the temperature at a distance 'r' from the centre. It should be noted that the procedure for obtaining temperature distribution in a hollow pin is also similar.

In some of the reactors the fuel pins are hollow cylinder. The basic differential equation can be integrated twice to get the temperature distribution. However, the boundary conditions for the hollow rod for the symmetric case are

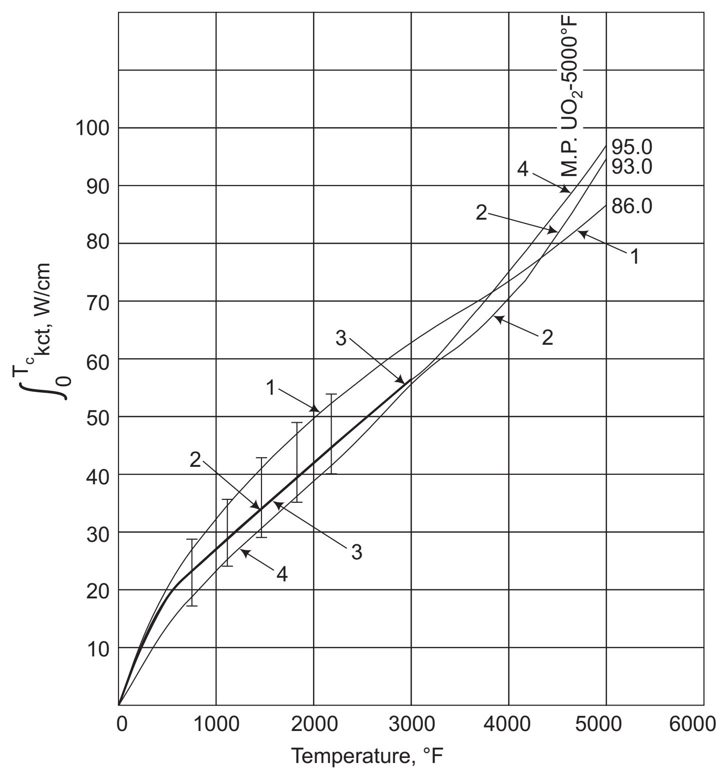


FIG. 6 VARIATION OF INTEGRAL KDT WITH TEMPERATURE [1]

$$\left.\frac{dT}{dr}\right|_{r=b} = 0 \text{ and } T_{r=a} = T_S$$

Where 'b' is the inner radius and inner surface is assumed insulated.

A mathematical derivation would show that for a given linear heat generation rate and a given surface temperature, the highest temperature is observed in the solid rod for uniform heat generation. This could also be proved by Fourier's law of heat conduction. Consider two rods 'A' and 'B' both having an outer radius 'a'. 'A' is solid rod and 'B' is a hollow rod with inner radius 'b'. The heat conduction is always radially outwards. Consider an imaginary cylindrical surface of radius 'c' where 'c' is greater than 'b' and less than 'a'. Heat flowing across 'c' in the two cases will be $\frac{q_L}{\pi a^2}\pi c^2$ and $\frac{q_L}{\pi(a^2-b^2)}\pi(c^2-b^2)$ respectively. Since $\frac{c^2}{a^2}$ and $\frac{c^2-b^2}{a^2-b^2}$ are less than 1, $\frac{c^2}{a^2}$ will always be greater than $\frac{c^2-b^2}{a^2-b^2}$.

If the heat flow through the same surface 'c' is larger in solid cylinder; the temperature gradients will also be larger according to Fourier's law. Consequently for a given surface temperature, the temperature at all similarly placed points will be higher for the solid cylinder as compared to hollow fuel rod.

Uniform volumetric source of heat generation is considered for thermal reactors as not only that it is convenient but also it is conservative. This can be easily shown from Fourier's law. Consider two identical fuel rods with same LHGR. The thermal flux is uniform in one (Case I) and has a depression in the centre (Case II). It is to be shown that case I will always give a higher centre line temperature. For this consider an imaginary cylindrical surface of radius 'b' which is smaller than radius 'a'. Let us also consider that 'c' is the radial distance from centre where q''' for the two cases is same. (It is to be noted that for the same LHGR in the two cases, q'''_{caseI} is greater than q'''_{caseI} for all points at radial distances smaller 'c' and is smaller for radial distances greater than 'c'). For all 'b' smaller than 'c' the heat crossing the imaginary surface will be larger in case I as q'''_{caseI} is greater than q'''_{caseII} everywhere. Total heat reaching 'a' is same in the two cases as LHGR is same in both. Hence even for cases where 'b' is greater than 'c' and so at points between c and b, q'''_{caseI} is smaller than q'''_{caseII}, the total heat crossing 'b' is smaller for case II. From this it can be concluded that heat crossing the surface of radius 'b' is always smaller for case II irrespective of whether 'b' is greater that 'c' or less than 'c'. Consequently from Fourier's law, for the same area of heat transfer in the two cases, the temperature gradients will be more in case I. From this it can be seen that temperature drop within the fuel will always be larger for case I. The exemptions for this would occur when fuel rod does not have symmetrical thermal flux about the axis due to fuel rod location. Figure 7(a) shows such a case. Even for Fast reactors (Figure 7(b)) the case I will not be conservative.

A fuel rod consists of bare fuel rod surrounded by fuel clad with gap resistance between the fuel and clad. This gap between the fuel and clad , meant for containing fission gases acts as a thermal resistance and fuel clad acts as a container of fuel pellets. All the logics given above hold true for the temperature drop between fuel central line and clad surface temperature. A typical temperature distribution in a fuel pin is given in Figure 8. Fuel clad also acts as a thermal resistance.

Although it is assumed after that no heat is generated in the clad, but strictly speaking about five percent of heat is generated in the clad due to absorption of radiations. From the logic given earlier it can be shown that, the temperature drop between fuel central line and fuel clad surface is the maximum for thermal reactors assuming a uniform volumetric source of heat generation in the fuel.

Temperature Distribution in Rectangular Geometry

In power reactors, the fuel is in cylindrical shape. Only for some special applications such as nuclear submarines, the fuel is plate type. The temperature distribution in plate type fuel can be estimated in a similar as that of a cylindrical fuel along with similar assumptions. The 1-D differential equation is

$$\frac{\delta}{\delta x}\left(k\frac{\delta T}{\delta x}\right) + q''' = 0$$

or,
$$k\frac{\delta T}{\delta x} + q'''x + C_1 = 0$$

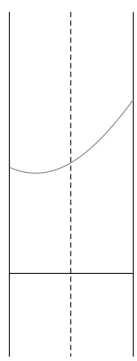


FIG. 7(a) NON-SYMMETRICAL CASE

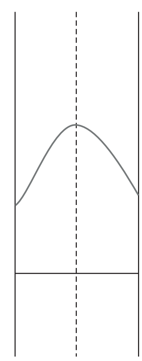


FIG. 7(b) FAST REACTOR CASE

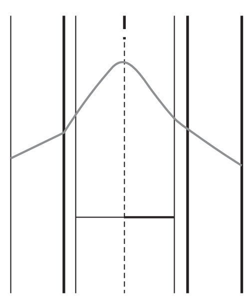


FIG. 8 TYPICAL TEMPERATURE DISTRIBUTION IN A FUEL PIN

or, $$k(T_0 - T) + q'''\frac{x^2}{2} + C_1 x + C_2 = 0$$

T_0 = center temperature. If the boundary conditions are symmetric and plate is of thickness $2a$ then,

$$x = 0,\ T = T_0,\ \frac{\delta T}{\delta x} = 0 \Rightarrow C_1 = 0$$

$$x = a,\ T = T_S \Rightarrow C_2 = -k(T_0 - T_S) - q'''\frac{a^2}{2}$$

$$\therefore \qquad T_S - T = \frac{q'''(a^2 - x^2)}{2k}$$

In reactor shields q''' may not be constant. The heat generated in the shield is due to absorption of radiation which may vary exponentially then,

$$q'''(x) = q_0'''e^{-\mu x}$$

Where 'μ' is absorption co-efficient.

It is important to define boundary conditions carefully. In some cases, though heat generation is less the temperature may be high. Take the following case.

Temperature Distribution in Shields

Shields are provided around a nuclear reactor to protect man and equipments from the ill effects of radiations. These radiations, when absorbed by the shields produce heat. Each shield material has to be cooled within a safety temperature limit. The heat produced decreases exponentially with distance. Let us take a rectangular radiation shield in which 1-D heat conduction is applicable. The shield is insulated from one end and the radiations are coming from the other end as shown in Figure 9.

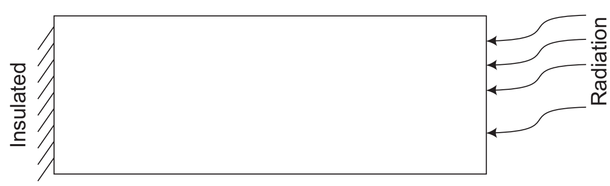


FIG. 9 A SHIELD HAVING LEFT END INSULATED AND RIGHT END EXPOSED TO RADIATION

For estimating temperature distribution in this, one dimensional differential equation for heat conduction is to be solved. The Dirichlet boundary condition is to be applied on left face. Right face can have any of the three boundary conditions. It should be noted that temperature in this case is highest where the heat production is least. The shields may be multi material layers. The heat production may be discontinuous at the interface. However the temperature distribution will be continuous. Same heat flux at the two sides is the interface boundary condition. Figure 10 gives the temperature distribution in two material shields.

Partial Differential Equation for Heat Conduction

If the temperature is changing with time then the partial differential equation as given below is to be used.

$$\frac{1}{r}\frac{\delta}{\delta r}\left(kr\frac{\delta T}{\delta r}\right)+\frac{1}{r^2}\frac{\delta}{\delta \varphi}\left(k\frac{\delta T}{\delta \varphi}\right)+\frac{\delta}{\delta z}\left(k\frac{\delta T}{\delta z}\right)+q'''=\rho c_p\frac{\delta T}{\delta t} \tag{19}$$

For k = constant, the equation becomes

$$\frac{\delta^2 T}{\delta r^2}+\frac{1}{r}\frac{\delta T}{\delta r}+\frac{1}{r^2}\frac{\delta^2 T}{\delta \varphi^2}+\frac{\delta^2 T}{\delta z^2}+q'''=\frac{\rho c_p}{k}\frac{\delta T}{\delta t}$$

or,
$$\frac{\delta^2 T}{\delta r^2}+\frac{1}{r}\frac{\delta T}{\delta r}+\frac{1}{r^2}\frac{\delta^2 T}{\delta \varphi^2}+\frac{\delta^2 T}{\delta z^2}+q'''=\frac{1}{\alpha}\frac{\delta T}{\delta t} \tag{20}$$

where 'α' is known as thermal diffusivity and has the dimension of m^2/s.

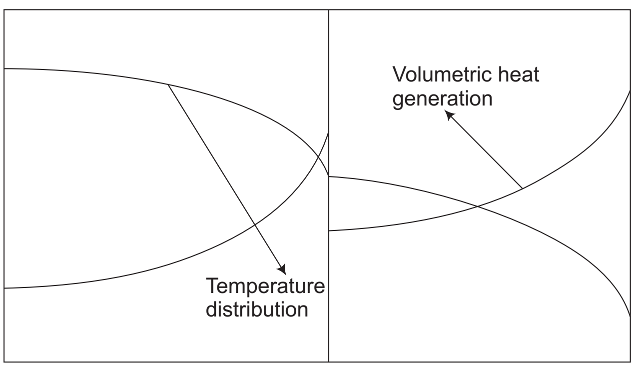


FIG. 10 TYPICAL TEMPERATURE DISTRIBUTION IN A TWO-MATERIAL SHIELD

This is a parabolic type partial differential equation. It can be easily integrated using central differencing schemes of numerical integration.

Effect of Reactivity Excursion on Temperature Distribution

In a reactivity excursion accident the power of the reactor increases almost exponentially and very rapidly as compared to the time constant for readjustment of temperature distribution in the fuel rod. Consequently the local temperature rises according to the local heat generation which is absorbed locally. For the case of uniform volumetric source of heat generation the distribution becomes more and more steep in the initial one or two seconds. Figure 11 shows the temperature distribution in a reactivity excursion.

The threat to the fuel integrity during this accident is integrity of the fuel. The heat does not have sufficient time to get conducted away to coolant. This adiabatic deposition of energy leads to sudden expansion of fission gas bubbles which are at the grain boundaries and do not have sufficient time to diffuse out. Consequently the fuel gets fragmented at much lower temperatures.

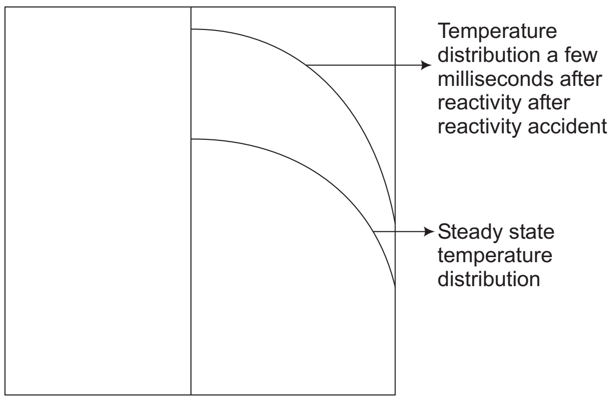


FIG. 11 TYPICAL TEMPERATURE DISTRIBUTION DURING REACTIVITY EXCURSION

Fuel temperature Distribution during a Loss of Coolant Accident

A loss of coolant accident (LOCA) is the one where the cooling of the fuel rapidly deteriorates due to the loss of coolant from the system. Under such accidents the reactor is rapidly shut down. Although main heat generation reduces due to reactor shut down, decay heat continues to be produced for a very long time. Engineered safeguard system are installed to pump coolant to remove the decay heat so that fuel temperatures remain within safe limits. Subsequently a long term recirculation system establishes a sustained cooling for the fuel. This entire transient from temperature distribution within the fuel considerations can be divided into several phases.

1. Redistribution of stored heat

This is the phase when the reactor heat due to fission has significantly reduced due to reactor trip (shut down). Heat transfer to coolant is reducing by order of magnitudes due to coolant loss. The heat within the fuel starts redistributing itself. The centre line temperature comes down towards the average temperature of the fuel while the surface temperature rises due to heat flow from one part to another. Consequently the temperature distribution within the fuel becomes relatively flat. Redistribution is over in about two to three fuel time constants (explained later). Then the flat temperature starts rising slowly due to decay heat. For understanding this phase let us consider that heat production is zero during this phase and heat removal from the fuel has also reduced to zero. Quickly, the full fuel will come to one average temperature. All temperatures higher than average temperatures will come down and those lower than average temperatures will raise. It is the similar phenomena which is taking place when the heat production and heat removal both have been significantly reduced.

2. Core heat up

In this phase the temperature of the fuel starts rising at almost uniform rate due to the generation of decay heat. When the surface temperature (the temperature distribution inside the fuel is almost flat) reaches a value which is threshold of metal water reaction, zircaloy of clad starts reacting with the steam present. This is an exothermic reaction. Significant amount of heat is produced as a surface source of heat generation. This reaction is Arrhenius type and the rate of reaction increases with the temperature. This causes temperature of fuel to rise rapidly. The reaction for zircaloy starts at about 900C and is literally uncontrollable at about 1400C. Radiative heat transfer dominates. In fuel which has heat sinks for radiation heat, significant circumferential and radial temperature gradients can be seen.

3. Reflooding

In this phase the engineered safeguard system start coolant injection for cooling the fuel. Fuel temperatures come down and distribution tends to remain flat. However, during the process of rewetting the axial gradients become large.

REFERENCES

1. Thermal Analysis of Pressurized Water Reactors, L.S.Tong and J. Weisman, 2nd Edition, American Nuclear Society, 1979
2. Two-Phase Flow and Heat Transfer, P.B.Whalley, Oxford Science Publications
3. Convective Boiling and Condensation, Collier J.G and Thome J R, 3rd edition, Clarendon Press, Oxford, 1994,
4. Heat Transfer, A basic Approach, M .N.Ozisik, International Edition, McGraw-Hill Book Company, 1985
5. Fundamentals of Heat and Mass Transfer, Incropera, Dewitt, Bergman, Lavine, 6th edition
6. A heat transfer textbook, John H. Lienhard IV / John H. Lienhard V, 3rd edition, Phlogiston Press Cambridge, Massachusetts, U.S.A., 2006
7. INTERNATIONAL NUCLEAR SAFETY CENTER (INSC) Material Properties Database, Argonne National Laboratory
8. INTERNATIONAL ATOMIC ENERGY AGENCY, , Thermophysical properties database of materials for light water reactors and heavy water reactors, IAEA-TECDOC-1496, June, 2006
9. DOE Fundamentals Handbook, Thermodynamics, Heat Transfer, and Fluid Flow, U.S. Department Of Energy, Washington D.C., DOE-HDBK-1012/1-92, June, 1992

INTERDISCIPLINARY ENGINEERING AND TECHNOLOGY

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

REQUIREMENT OF HYDRO GAS SUSPENSION SYSTEM FOR TRACKED ARMOURED FIGHTING VEHICLES

DR. D. HANUMANNA[1] U. SOLOMON[2] AND SEKAR. N.S.[3]

ABSTRACT

The modern day Tracked Fighting Vehicles are equipped with Hydro Gas Suspensions, in lieu of conventional mechanical Suspensions like Torsion Bar and coil spring bogie suspensions. The uniqueness of Hydro Gas Suspension is that it offers a non-linear spring rate, which is highly required for the cross-country running of a tracked vehicle. These Tracked Vehicles are capable enough to negotiate different cross country terrains like sandy, rocky, river bed, etc. at an enhanced speed to meet the shoot-and-scoot war doctrine.

Hydro Gas Suspension was considered for MBT as a frontier technology and ensures high level performance of on board systems, consistent with the level of their development well in 2000's. This type of Suspension are under consideration of various futuristic MBTs of the world as an ultimate solution which will lead to semi-active, and active control and are under various stages of development.

The Hydro Gas Suspension system permits increased vehicle speed thereby reducing vertical acceleration of the vehicle, which may become intolerable to the structure, the Cargo and to the occupants. The suspension system supports the vehicle body at a desired height above the ground. It permits the vehicle to negotiate obstacles like deep ditch, snow, soft sand, to ford in shallow waters, boulders, logs, deep rubs and holes etc. The suspension system allows a uniform weight distribution in accordance with a predetermined plan. The provision of crew and passenger comfort is another function of suspension. Suspension provides traction with ground and transmits driving and breaking torque. And above all, the suspension system provides a stable gun platform, which is a prerequisite for a fighting vehicle.

Keywords: Hydrogas, Suspension, Tracked Vehicle, Cross Country Vehicle, Spring, Damper, Vibration, Off-road, Non-linear Suspension, Torsion Bar

---

[1]Outstanding Scientist (Rtd) [2]Scientist - F [3]Scientist – C
Combat Vehicles Research & Development Establishment, (DRDO, Ministry of Defence), Avadi, Chennai 600 054, India

1. INTRODUCTION

The basic requirements of a Battle Tank are Mobility, Protection and Fire Power. Increasing the battle tank mobility is as important as its firepower and protection. High mobility in general and high off road mobility in particular gives definite advantages to a battle tank over an opposing tank. Requirement in terms of off-road performance for a Main Battle Tank (MBT) is continuously growing due to reduced concentration of tanks and wide coverage area war theatre. The increasing demand for the cross-country speed made the development of innovative Hydro Gas Suspension system.

It is obvious that any vehicle needs a suitable Suspension system to flexibly support the mass of the vehicle and to offer good vibration isolation to the goods & passengers that are being transported. Mobility is one of the important characteristics of a tracked vehicle and the mobility is determined mainly by the power of the Engine & drive train and the Running Gear sub-system, on which the vehicle is mounted. The Suspension is one of the important constituents of the Running Gear, which is shown in Fig.1 for a typical tracked vehicle. The Suspension essentially possesses a vibration isolation means and a damper mechanism to reduce the transmissibility of the terrain inputs to the crew and to reduce the successive amplitudes caused due to springing, respectively. The selection of a right kind of Suspension is a very critical issue as it requires a deep understanding about the system functioning, terrain deployment, forces acting upon, etc.

2. CHARACTERISTICS OF TRACKED FIGHTING VEHICLES

The Tracked Fighting Vehicle is designed primarily based on the war doctrine that is being adapted in the respective country. The mobility requirement is based on the war field ranging from strategic mobility to tactical mobility and tactical mobility to battle field mobility. While engaging the enemy, the battle field mobility dominates compared to other. The battle field engagement with enemy is performed in a cross-country environment with high agility, accurate firing capability and maximum protection. The firing causes heavy forces due to gun recoiling and the protection requires a heavy impact absorbing. The mobility of the vehicle encounters considerable forces due to terrain bounces and ditches, braking, steering and gradient climbing.

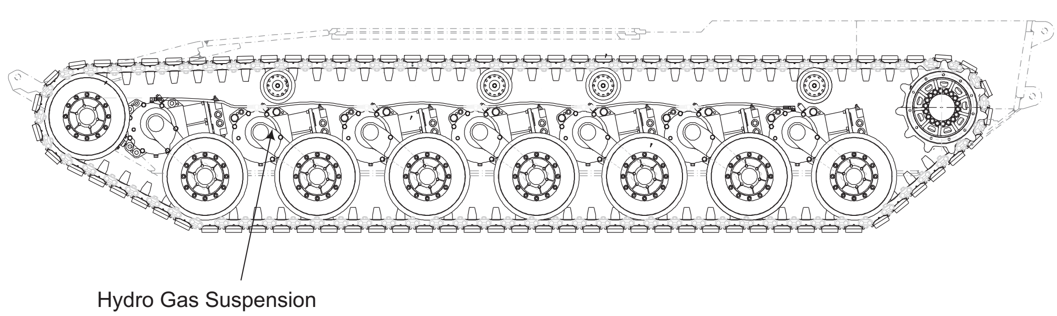


FIG. 1 RUNNING GEAR FOR A TYPICAL TRACKED VEHICLE

3. VEHICLE DYNAMICS

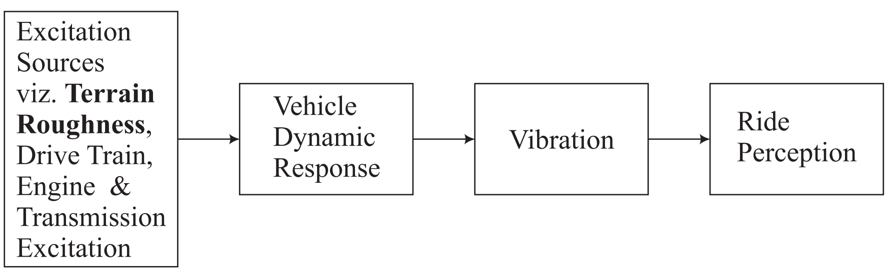
3.1 Necessity of Suspension in Tracked Vehicle

The ultimate aim of providing a flexible support through the Suspension is to negotiate the obstacles and ditches that are encountered during a cross-country run to provide a better ride comfort to the crew. During the operation, the Suspension components have to encounter high resisting loads in dynamic conditions and are subjected to higher level of vibration. To augment an increased mobility, the enhancement on mobility does not merely imply increasing power to weight ratio, but it demands a superior Suspension system, which limits the vibration, that crew and equipment can tolerate for prolonged period under those conditions. Ideally a battle tank must be equipped with a Suspension system such that it allows the driver to maintain the speed unchanged over uneven ground, without sacrificing combat efficiency.

The majority of cross-country vehicles are equipped with power trains capable of developing maximum speeds, yet these vehicles are able to operate off-the-road on typical hard rough terrain at much lesser speeds. This apparent contradiction results because the sustainable speed of off-road vehicle is not governed by power alone, but also by the geometry and dynamic properties of the vehicle. As a vehicle travels over rough terrain it induces vertical Acceleration, pitching and bouncing in the vehicles.

3.2 Ride Perception Model

The Tracked Fighting Vehicles travelling at high speed will experience a broad spectrum of vibration as depicted in the following Model:



Being a cross-country vehicle, the chassis is encountering various excitation sources viz. terrain undulations, Running Gear drive system and Engine & Transmission power sources. This in-turn leads the vehicle for a dynamic response resulting induced vibration. The crew and the passenger on the vehicle are subjected to the effect of the vibration, which is known as '*Ride Perception*'. The excitation sources are the large non-uniformities on the terrain. Terrain Roughness is described by the elevation profile along the terrain the wheel tracks over which the vehicle passes. Terrain profile fits the general category of broadband random signals and represented as Power Spectral Density (PSD) function.

3.3 Need of Human Comfort on Vehicles

Every human organ is having a maximum tolerable vibration level. The sensitivity varies from organ to organ and person to person. Human tolerance to the vibration is measured in terms of acceleration that is coming on to the vehicle body. The human body is sensitive to both the vertical and fore/aft acceleration. The studies show minimum tolerance i.e. max sensitivity of human body to vertical vibration is in frequency range of 4 - 8Hz (Figure 2), within which all the natural frequencies of a number of principal parts of the body including the abdomen.

For frequencies above and below this range, the tolerance increases in proportion to the frequency. The actual shape of the boundaries will often show small deflections in 10 - 20 Hz range due to other organ resonances especially resonance near 10 Hz. The human response to the fore/aft vibrations is some what different from that of the vertical. The region of maximum sensitivity occurs in 1 - 2 Hz range. Another fundamental method of combining vertical and fore/aft vibration adopted by U.S Army AMM – 75 is the absorbed power for each direction and simply adding it. The Suspension, thus, plays vital role in eliminating/ reducing the vibration level so that the human organs are least affected while riding on the Tracked Vehicle.

3.4 The Suspension Model

The Suspension essentially consists of two elements viz. Damper, an energy dissipation medium and Spring, an energy absorbing medium.

Equation of Motion:

$$m\ddot{x} + C\dot{x} + kx = F_o \operatorname{Cos}(\acute{\omega} t) \tag{1}$$

The mass ‘m’ is the vehicle sprung mass, acceleration & velocity based on displacement ‘x’ and the input forces F_o in time domain. The ‘C’ and ‘k’ are the damping co-efficient and spring stiffness

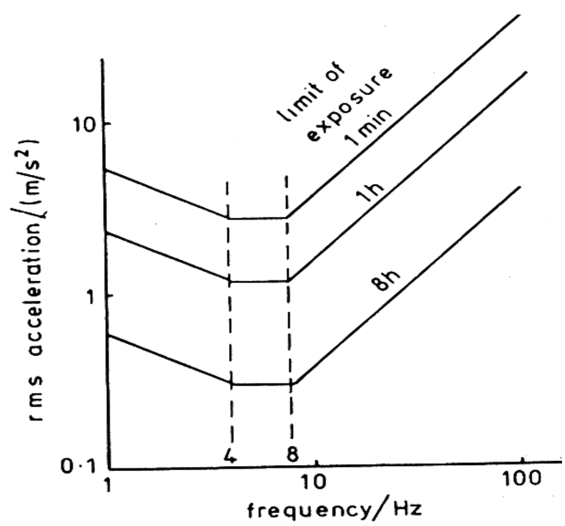


FIG. 2 HUMAN TOLERANCE LEVEL (ISO:2631)

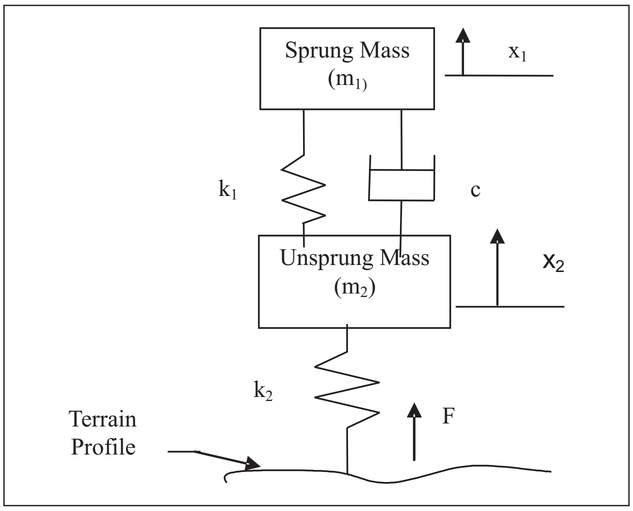


FIG. 3 A TYPICAL SUSPENSION MODEL

of the operating mediums (Figure 3). The velocity and displacement of the sprung mass governs the output i.e. acceleration, which needs to be reduced at an optimum level so that the crew fatigue is reduced. Also the suspension can have rubberised rolling element with some stiffness to support the Unsprung mass. From the above equation, it is obvious that 'x' is the terrain input and the requirement is to have reduced acceleration. So, the design of suspension dictates us to manipulate with 'C' and 'k', which are under design control.

4. REQUIREMENT OF SPRING STIFFNESS AND DAMPING RATIO

To cater to the requirement of Human Comfort and to reduce the Crew Fatigue the vibration level coming to the vehicle body has to be minimized. Also, for protecting the on-board equipments from shock & vibration and to get high level of agility on cross-country terrains.

From the Figure 4 and Eqn. (1) it is obvious that the damper domination on ride perception is high in the zone where the natural frequency of the systems nearly equal to the induced frequency (i.e. $ø/ø_n = 0.7 – 1.4$). The spring dominates on the ride perception till $ø/ø_n = 0.7$. Likewise the mass of the vehicle gains better control if the $ø/ø_n$ exceeds 1.4. The 'ø' is directly depends on the vehicle speed & the terrain conditions and hence a proper selection of optimised 'k' and 'C' are quite important to reduce the transmissibility of vibration under all vehicle speeds.

4.1 Application of Spring in Suspension

An ideal suspension characteristic is one which transmits no acceleration to the hull, but provides a constant force, just sufficient to carry its share of weight, independent of wheel movement. This is termed as zero rate suspension system or ideal suspension system or perfect suspension. Such a suspension poses a number of problems in achieving it practically. This is because the purposeful

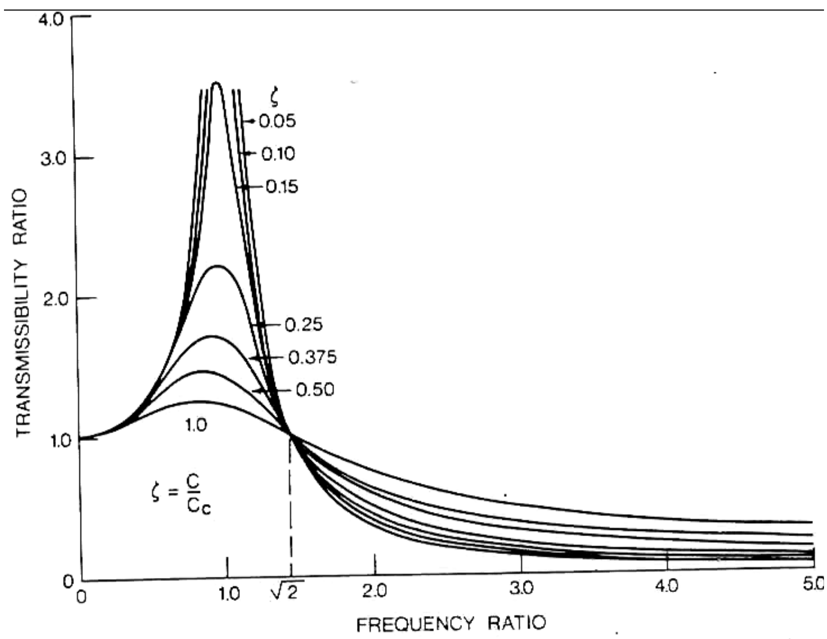


FIG. 4 VIBRATION TRANSMISSIBILITY

negotiations like gradient climbing, impact absorption, etc. needs a harder spring also. Hence it is necessity to have a combination of Soft and Harder Springs for Cross-country manoeuvre.

The Suspension spring characteristics depicted in Figure 5 reveals that the non-linear springing with softer spring rate is required up to 2/3 of the travel and harder spring rate up to 1/3 of the travel. The Suspension thus should be designed to cater to the dynamic loading of 4 – 5 times the static loading to counter the terrain inputs during dynamic conditions.

4.2 Application of Damper in Suspension

The variation of damping force with velocity and displacement may be different for compression and rebound strokes. The system is designed to obtain the higher damping during bump than rebound. Because of the flexibility in damper setting, one can set the bound stroke damping and the rebound stroke damping independently. The damper setting and flow rate characteristic is depicted in damper flow rate curve (Figure 6). Series of tests were conducted for different parameters of orifice size, damping obtained and heat built up. With the results obtained an optimum size of orifice has been chosen which suits most of the operating conditions.

The critical damping ratio of $\zeta = 1.0$ is not preferable as it could offer more add-on resistance to the vehicle and a very low damping $\zeta < 0.3$ may lead to resonance. Hence for a typical cross-country terrain ζ ranging from 0.30 – 0.35 is ideally preferred as the optimal HRV coincides with this. The damper is designed to have medium damping with $\xi = 0.3$ by keeping required orifice diameter and it is an arrangement to dissipate the energy. It can be either hydraulic or friction.

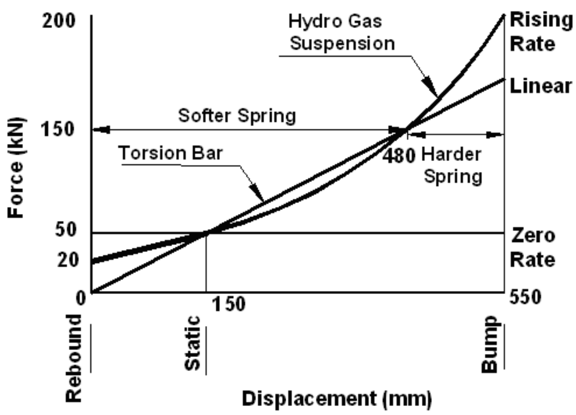


FIG. 5 SPRING CHARACTERISTICS (LINEAR & NON-LINEAR)

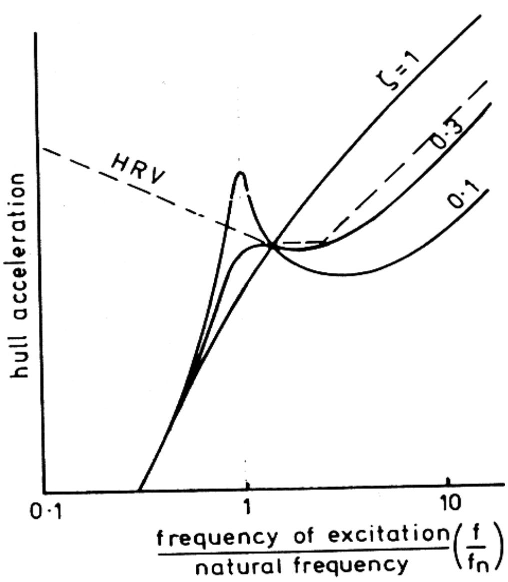


FIG. 6 HRV CURVE AND DAMPING RATIO DETERMINATION

Hydraulic dampers are provided with pressure relief valve so that when the vehicle negotiates a larger obstacle at high speed, the Pressure Relief Valve (PRV) opens the passage without dampening the flow i.e. truncating to reduce the excessive energy dissipation.

4.3 Conflicting Requirements of a Suspension System

From the above discussions, it is clear that a combination of soft and hard spring rates are essential to meet the requirements. The following model depicts the conflicting requirement of the Suspension:

Conflicting Requirement of Suspension

For good vibration isolation for the sprung mass over a wide range of frequencies	**Soft spring** is required
To provide good road holding capability at a frequency close to natural frequency of un sprung mass	**Stiff spring** is required
To reduce the amplitude of the vibration of the sprung mass close to resonance	**High Damping** ratio is required
In the high frequency range to provide good vibration isolation for the sprung mass	**Low Damping** ratio is required

The non-linearity property of the gas meets the requirement of the first two aspects and the bi-directional damping with PRV certainly meets the remaining aspects of the conflict requirements to the extent of acceptable in passive type Suspension.

5. SELECTION OF SUSPENSION

The selection of Suspension needs a thorough knowledge on (i) Role-High Speed, Silence & Minimum Logistic backing, (ii) Terrain deployment viz. Desert, Semi-desert, Ploughed soils and terrain with boulders, (iii) Amphibious requirement and (iv) Vulnerability on Battle Field.

Thus the selection of the Suspension is aimed at:

i. To reduce acceleration levels in vehicle within bearable limits of the crew for the duration of the mission.
ii. To limit acceleration and to protect sensitive on-board equipments from shock and vibration.
iii. To modify or isolate terrain inputs to the vehicle such as to provide stable platform for firing on the move.

The Selection of Suspension thus leads to improved mobility while supporting the vehicle body and hull. The Suspension thus provides lateral and longitudinal stability with selective distribution of weight and maintains the track in contact with ground. The ultimate purpose of deploying the Suspension is to achieve a higher level of passenger/ crew comfort, vibration free transportation for the on-board sensitive equipments and enable a stable platform for firing on the move.

Traditionally different types of Suspensions are being used for the tracked vehicles ranging from (i) Bogie Suspensions, (ii) Coil spring Suspension, (iii) Torsion Bar Suspension and (iv) Hydro Gas Suspension. The Bogie & Coil Suspensions suffer from a limited wheel travel, while the Torsion Bar is undesirable due to high linear stiffness due to material property. The Hydro Gas Suspension is the order of the day due to its inherent non-linear characteristics. The Suspensions are otherwise classified in terms of their response as Passive, Active Suspension and Semi-Active Suspensions. All our discussions are related only to Passive response Suspensions.

5.1 Comparison of Different Suspensions

Traditionally various types of Suspension (Figure 7) viz. Coil Spring, Bogie Wheel Suspension and Torsion Bar Suspension are being used for cross-country vehicles. As the need of enhanced agility arose, the Suspension needs to augment this, necessitating a design to opt for non-linear suspension like Hydro Gas Suspension. The formers are basically of material based systems while the later one is based on gaseous principle. i.e. $PV^n = C$ and $PV = mRT$.

(a) Bogie Suspension (b) Torsion Bar Suspension

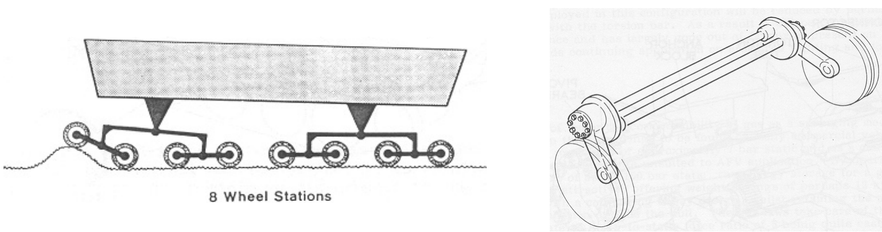


FIG. 7 CONVENTIONAL TYPES OF SUSPENSION

5.1.1 Bogie and coil spring suspension

This type of suspension system was used extensively in early tracked vehicles. This system consists of Links, Arms and Springs are interconnected in a manner to permit two or more road wheels to function together in tandem. The bogie system divides the load equally between a pair of tandem wheels. When one of the wheels is displaced or is subjected to a vertical track force, an equal force is reflected on the wheel of the bogie unit. This arrangement permits vehicle to negotiate substantial terrain irregularities. The bogie suspension system is only for low speed operation at speeds about 10 Kmph. Hortsmann Suspension is another form of Suspension using the coil spring. Coil Spring may think of as being a Torsion Bar wound into helix. Although the length problem is not there but the space in the centre makes it bulky and as a result coil spring have normally been found and externally mounted Suspension units. With these springs, limitation on the cross country is the weight of tanks is going up.

5.1.2 Torsion bar

A Torsion Bar is an elongated member which in normal usage is fixed at one end and loaded by applying a couple acting in a plane normal to the axis of the bar to the opposite end. The torsion bar springs used in existing military suspension systems are straight bars of circular cross section splined at each end. The Torsion Bar suffers from a serious drawback of increasing the tank height from the ground, offers constant spring stiffness and needs an externally mounted damper.

5.2 The Hydro Gas Suspension

In Hydro Gas Suspension system instead of conventional mechanical Torsion Bar, gas is used to act as a spring medium. Hydraulic oil is filled-in to provide hydraulic damping to dissipate the energy thus diminishing the successive amplitudes. A typical Hydro Gas Suspension system consists of a stationary Casing, a rotating Crank and sliding Pistons inside the Cylinders are depicted in the Figure 8. The Suspension is mounted with an Axle Arm and a Hub is mounted over the Arm through a Stub Axle. On this Hub, Road Wheels are fastened to transmit the terrain inputs to the Suspension. The Casing, Crank, Piston Rod and the Cylinder form a four bar single slider crank mechanism to convert the rotary movement of the Axle Arm to result in Piston displacement to compress the Gas medium.

When the tracked vehicle is fielded into operation, the track drags up the vehicle utilising the Engine power. Due to this, the Suspension Road Wheels rotate and while the vehicle negotiates an obstacle, the Road Wheel along with the Axle Arm swings up. The kinematic arrangement results in the displacement of the filled-in hydraulic oil. Ultimately this fluid compresses the Accumulator Piston against the entrapped gas. This raises the pressure of the gas offering sufficient load to take the vibrations induced. A Damper is an important sub-system of a Suspension interposed in between Actuator and Accumulator to dampen out the vibration by dissipating the energy so as to eliminate the chances of resonance in the system.

In this Suspension, gas is used as a spring medium and hydraulic oil is utilised for force transmitting and dampening out the oscillations. The advantage of employing the hydro gas Suspension in a tracked vehicle is not only to isolate the primary vibrations induced into the Suspension system but also it can offer better ride comfort through non-linear springing action of

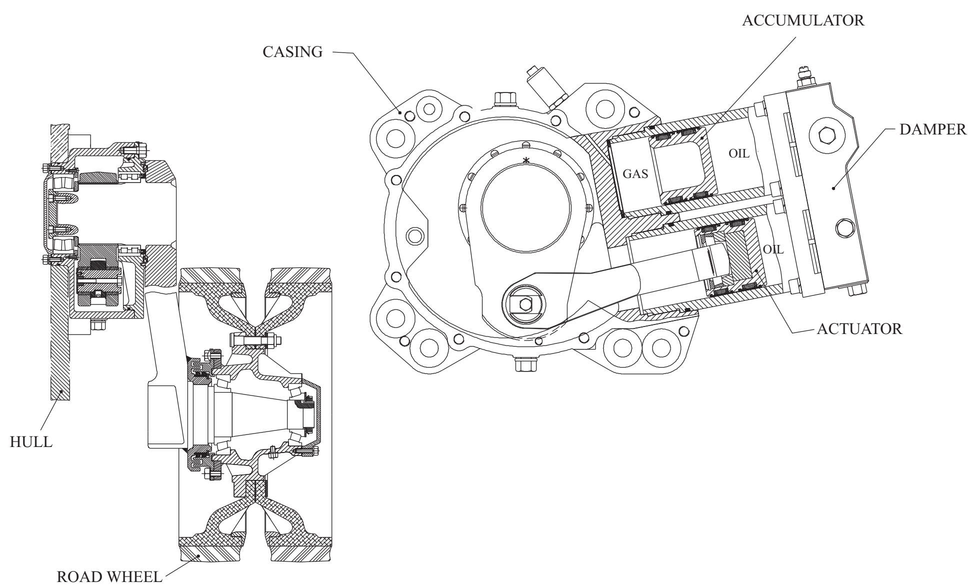


FIG. 8 TYPICAL HYDRO GAS SUSPENSION

the gas. To take the maximum advantage of the Suspension, it is expected that the system should have less spring rate at the same time it should be able to vary according to the load. This is because at lower amplitudes the behaviour of the spring should be very soft where as after two third of the wheel travel the spring should offer more resistance with high spring rate. This is because the Suspension should have a minimal transmissibility ratio but at the same time it should able to negotiate larger obstacles & gradients and to counter the gun re-coiling forces with high spring rate.

5.2.1 Types of constructions for hydro gas suspension

Depending upon the space availability, actuation need, forces acting upon the system, Ground Clearance requirement, etc, the Hydro Gas Suspensions are configured in various forms. The working is similar in all other than the physical movement of the actuation. Following are some types being used in the industry:

i) Opposed Cylinder (Actuator & Accumulator in oppose)
ii) Parallel Cylinder (as shown in Fig. 8)
iii) In-Arm (Integral Arm & Casing used at restricted space)
iv) Telescopic Strut (Mostly used as add-on hybrid Suspension)

5.3 Uniqueness of Hydro Gas Suspension

In Suspension system, the principle compressibility of gas is utilised as a spring medium. Moreover Suspensions have non-linear load deflection characteristics, which result in lower natural frequencies under the most common operating conditions. They also offer the possibility of damping at all wheel stations. Therefore there is a strong reason for incorporating Suspensions for optimum performance. The systems pressure that is spring rate can be independently adjusted and the in-built damper enabling a better damping distribution.

6. DESIGN CONSIDERATIONS OF HYDRO GAS SUSPENSION

The Suspension has to perform its intended functions viz. (i) to convert the Road Wheel & Axle Arm rotation into linear displacement of Piston, (ii) force transfer to the compensating Energy Absorbing Medium, (iii) to have minimum side thrust in turn to reduce Seal side loads, (iv) to effect larger fluid compression for softer non-linear spring rate, (v) optimised fluid flow to effect required damping, (vi) absorbs shocks and vibrations due to lateral movement of the linkages and (vii) to provide intended vertical wheel.

6.1 Forces Acting on Hydro Gas Suspension

The Suspension encounters various loads viz. (i) dynamic load acting vertically due to terrain inputs and acceleration/ braking, (ii) dynamic load acting axially due to turning manoeuvre and slope of the obstacle contour, (iii) static load due to vehicle weight, (iv) speed of the vehicle at which it operates as an input in terms of time and (v) terrain conditions due to obstacles and ditches resulting in high input to the kinematic chain.

The dynamic loads acting on the Suspension is complex owing to the input forces occurring from vehicle acceleration, braking, turning manoeuvres, gun recoiling, ballistic impact, etc. The mass and speed of the vehicle offer higher inertial forces and the kinetic energy varies directly with the mass and with square of the speed. Due to this heavy forces are set up in Suspension system, which are to be encountered with appropriate design of the Kinematic linkages. Also the Centre of Gravity due to variation in mass distribution of the vehicle plays a vital role in offering the loads to the Suspensions.

6.2 Wheel Geometry for Hydro Gas Suspension

In static the tank should have required Ground Clearance (G.C.). The swing Arm Length, its mounting, road wheel diameter, etc. are controlling parameters determining the G.C.

Three positions of Suspension:

(i) Rebound - To enable contact with the ground while crossing the ditch

(ii) Static - Nominal Position

(iii) Bump - Maximum wheel travel position

The suspension system should be designed to offer soft springing for lower terrain inputs and hard springing for higher terrain inputs. The ratio of Static to Bump is decided based on dynamic loading requirement. Acceleration level of 4.5 to 5.0 g is usually taken for dynamic loading conditions. The off-loading of the Suspension from the static to rebound is about $1/4^{th}$ of the travel available for on-loading from static to bump, which is $3/4^{th}$ of the total travel. These travels are ultimately deciding the spring rate requirement and capability of the vehicle to negotiate the obstacles.

7. CONCLUSION

In this paper, an attempt has been made to highlight the requirement for a Hydro Gas Suspension to meet the cross-country manoeuvre as well as providing stable platform for the gunnery equipments. As a transformation from the traditional approach in the application of passive Suspension system for a off-road vehicle is changing to meet the newest challenges, the battle field is posed to face. The need for a Hydro Gas Suspension system for the tracked vehicles arose due to its inherent advantages over the conventional Suspension system.

The Hydro Gas Suspension is a state-of-the-art Suspension system offering smoother ride quality with reduced crew fatigue. The Hydro Gas Suspension not only improves the cross-country mobility but also enhances the weapon system accuracy and reliability significantly in case of Tracked Fighting Vehicle application as it provides a more stable gun platform compared to conventional Suspension systems. The uniqueness of the Hydro Gas Suspension with the non-linear springing and bi-directional damper with by-passing Pressure Relief Valve finds immense application in the modern day battle tanks. These types of Suspensions are under consideration for various futuristic tracked vehicles of the world for various applications ranging from earth moving to combating due to their inherent advantages.

REFERENCE

[1] Head Quarters, USA Army Materiel Command, "Engineering Design Hand Book-Automotive Series – Automotive Suspensions", 1967

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

HAZARDOUS WASTE MANAGEMENT: CLEANER PRODUCTION/WASTE MINIMIZATION

A.K.A. RATHI*

1. INTRODUCTION

After independence, the industrial development was considered to be a major tool for achieving economic growth in India. The Government of India, through industrial licensing policies, influenced types of industrial projects and their locations. The capacities for licensing manufacture were fragmented with an objective of balanced growth of different regions of the country [1]. Small-scale industries, irrespective of the type of products, were encouraged because of high employment potential per unit of capital invested. As a result there are a large number of small industrial units spread out in different parts of the country. Such industries find it difficult in mobilizing resources for proper pollution control. With increasing public awareness, public interest litigation and judicial interventions, the industry has been under tremendous pressure to take effective pollution control measures especially since mid-1990s. The new approach should aim at environmental management, and not just pollution control, using a broad mix of incentives and pressures to achieve sustainable improvements. Cleaner production techniques offer improvements in the manufacturing processes and their management in reducing the quantum of pollution generated while achieving increased production efficiencies and reduced operating costs [2].

2. HAZARDOUS WASTE

It may be appreciated that natural resources are used in the generation of energy, manufacture of goods and consumption of goods. Resources are required for products to be manufactured (in the form of feedstock, additives, *etc*), production processes involved (in the form of utilities), packaging, transportation (in the form of fuel, lubes, *etc*) and energy generation (in the form of fuel). From each of these activities, waste gets generated, which can cause air, water and/ or land pollution, Figure 1 and Figure 2 respectively. From a typical material balance carried out for a manufacturing operation, it would be observed that *waste* gets generated, Figure 3. Since it gets

* Former Chief Technical Adviser, Government of Gujarat, Industries Department, Gandhinagar 382017,
INAE-AICTE Distinguished visiting Professor, National Institute of Technology Karnataka, Surathkal 575025
Visiting Professor, Nirma University, Ahmedabad 382481, *drakarathi@gmail.com*

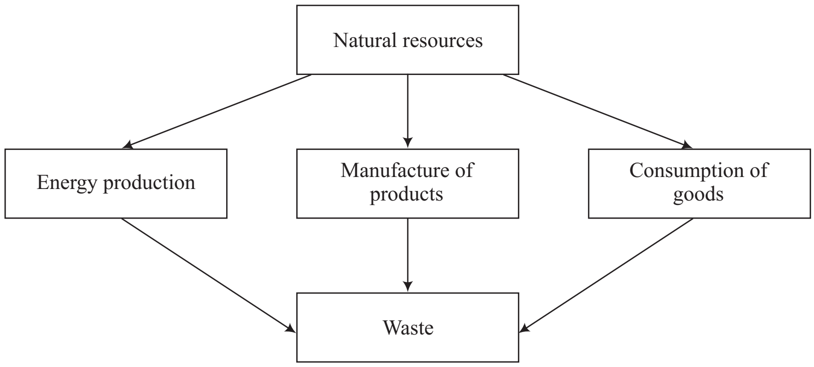


FIG. 1 WASTE GENERATION

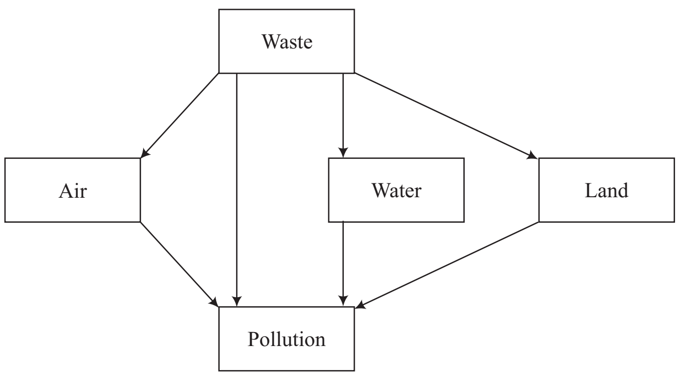


FIG. 2 WASTE IN DIFFERENT MEDIA

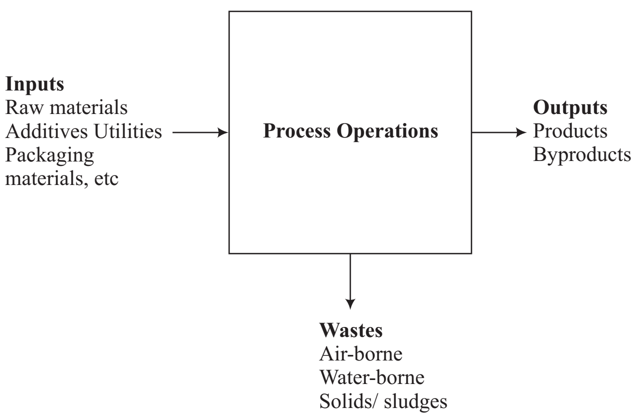


FIG. 3 INDUSTRIAL WASTE AS BY-PRODUCT

produced from manufacturing process, it should strictly be considered as a *byproduct arising out of inefficiencies* in the manufacturing processes. Thus *waste* should be treated as an undesirable byproduct. It could as well be considered as a resource, which is at a wrong place and at wrong time [3]. Waste generated at one location may be a useful resource at some other location. Waste generated at a given time may become a resource at some other time when different technologies get developed and new products/ applications get developed. Industrial waste could be in the form of contaminated gases, liquid waste, solid waste, sludges, and waste generated from wastewater treatment facilities.

Hazardous waste can be defined in a number of ways including

- hazardous characteristics *e.g.* toxicity, flammability, reactivity, *etc*.
- certain toxic compounds *e.g.* arsenic, PCBs, *etc*.
- types of materials *e.g.* organic solvents, explosives, *etc*.
- processes from which they originate *e.g.* clinical waste, residue from pesticide manufacture, *etc*.
- specific waste streams e.g. wastewater treatment plant sludge, incinerator ash, *etc*.

As per the working definition given by UNEP in 1985, hazardous wastes mean wastes (solids, sludges, liquids, and contaminated gases) other than radioactive (and infectious) wastes which, by reason of their chemical activity or toxic, explosive, corrosive, or other characteristics, cause danger or likely will cause danger to health or the environment, whether alone or when coming into contact with other waste. In USA, any of the following four characteristics make a waste hazardous:

- Corrosivity (waste that is highly acidic or alkaline)
- Ignitability (waste easily ignited and thus posing a fire hazard during routine operations)
- Reactivity (waste capable of potentially harmful, sudden reactions such as explosions)
- Toxicity (waste capable of releasing specified substances to water in significant concentrations)

Resource Conservation and Recovery Act, 1976 of USA defines hazardous waste as a solid waste or combination of solid wastes, which because of its quantity, concentration, or physical, chemical, or infectious characteristics may cause or significantly contribute to an increase in mortality or an increase in serious irreversible, or incapacitating reversible, illness; or pose a substantial present or potential hazard to human health or the environment when improperly treated, stored, transported, or disposed of, or otherwise managed.

Govt of India notified Hazardous Waste (Management and Handling) Rules, 1989 under the Environment (Protection) Act, 1986. Major amendments to these Rules were notified in 2000 and 2003. Hazardous waste defined in the Rules is any waste which by reason of any of its physical, chemical, reactive, toxic, flammable, explosive or corrosive characteristics, causes danger or is likely to cause danger to health or environment, whether alone or when in contact with other wastes or substances.

The regulations allow several exemptions and exclusions when determining whether a waste is hazardous. Chemical process industry contributes to about 75% of the total hazardous waste generated. The hazardous waste generated could form 1-5% of the output of the chemical industry.

2.1 Why Hazardous Waste Management?

Hazardous waste management is necessary because improper and careless management of hazardous waste has often created problems; to cite few instances [4]:

i. Minimata Bay methyl mercury disaster in Japan in 1952 in which several fishermen were killed and thousands suffered permanent brain damage and paralysis due to poisoning by methyl mercury.
ii. Love Canal chemical disaster in USA in 1976 which caused deformation in newly born children, contamination of surface as well as ground water in a large area with chlorophenols, dioxins and other hazardous chemicals.
iii. Basel chemical disaster of river Rhine in Switzerland in 1986 in which 1300 tons of hazardous chemicals got released in river destroying all life form which did not regenerate even after two decades.

Hazardous waste residue of production processes may cause significant damage to environment and/or human health unless handled, stored, transported, treated and disposed off scientifically using environmentally sound technologies and practices. The nearly uninhibited movement, activity, and reactivity of hazardous waste in the atmosphere are well established, and movement from one medium to another is evident. Hazardous wastes may enter the body through ingestion, inhalation, and dermal absorption or puncture wounds. Impacts on health of human beings and other living beings could be felt immediately, ecosystem are felt later, and global environment are felt still later. Some effects cannot be seen but could be felt only. Precautionary principle aims at preventing short-term as well as long-term harm to the environment and to human health

The approach to Environmental Management has undergone a major shift. Till 1970s or even early 80s, focus was on reacting to a problem, compliance, end-of-pipe control, and isolation of environment, health and safety functions in an organization. However, new mantras that came into being from 1990s onwards include anticipation and prevention, life cycle approach, muti-function integration, mainly because aspirational horizon shifted from local to *global* and immediate and short term to *the needs of next generation*. The drivers of actions on environmental/ waste management have mainly been international treaties, legal requirements in different countries, public interest litigations, and voluntary initiatives taken by industry sectors.

Sustainable development is defined in Brundtland Report as “development that meets the needs of the present without compromising the ability of the future generations to meet their own needs”. The Supreme Court of India has laid in its judgments *the precautionary principle*, and *polluter pays principle* as essential features of sustainable development. Sustainable consumption is defined as the provision of services and related products which respond to basic needs and bring a better quality of life while minimizing the use of natural resources and toxic materials as well as the emissions of waste and pollutants over the life-cycle of the service or products, with a view not to jeopardize the needs of future generations. With all these developments, hazardous waste management has become an integral part of manufacturing management.

3. WASTE REDUCTION

Waste reduction is considered as the first pro-active measure in hazardous waste management. Different terminologies are being used for achieving the objective of waste reduction.

3.1 Terminologies

A number of related terms being used for conveying the concept of waste reduction include

- low or no-waste technologies
- waste minimization (India)
- waste and emissions prevention (Netherlands)
- source reduction (USA)
- eco- efficiency
- environmentally sound technologies

An ideal and thus the most preferred waste management hierarchy should be the one shown in Figure 4. In the hierarchical choices for increasing resource utilization, obviously the first choice should be reduction at source. This could be achieved by input materials change- purification, substitution, and dilution; process technology changes- process change, equipment, piping or layout change, operational setting changes, process control and instrumentation; procedural changes- operating procedures, loss prevention, segregation of waste streams, material handling, inventory control, production scheduling, preventive maintenance, product changes- product substitution, product reformulation, product conservation, etc. When further source reduction is not

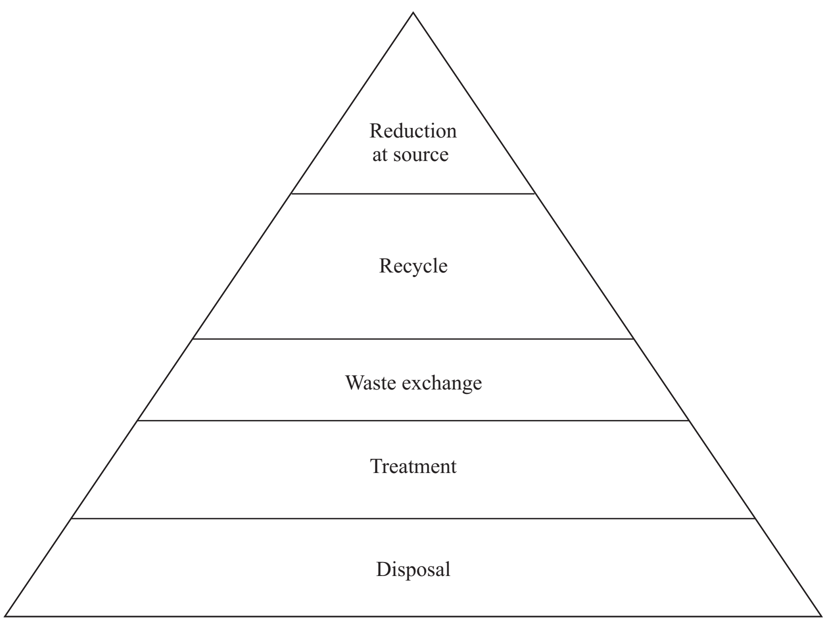


FIG. 4 WASTE MANAGEMENT HIERARCHY

possible, recycling for direct reuse on-site, preferably in the same process with or without some makeup is the next-best option. Additional recovery on-site for recycling/ reuse in some other processes could be considered thereafter. After exhausting recycling/ reuse on-site options, recovery off-site is considered as the next best option. This includes reclamation of desired material for some application or material and energy recovery for reuse off-site. If material and energy recovery is not possible, waste exchange for using the waste as a resource in some other off-site application is the next best option. Thereafter only, treatment is considered as the next option, followed by disposal opted as a last resort. In practice, however, it is observed that most of our industries are exercising only last two options *viz*. treatment of waste generated by them, and land disposal of solid waste. It is not being appreciated that waste treatment and disposal options involve cost, which is avoidable to a large extent. A typical common integrated hazardous waste management facility often involves large capital and has potential environmental risks because the facilities include

- Storage and handling of segregated wastes received from different sources.
- Recovery/recycling facilities for recovering some material as a saleable product like solvents, oils, acids, metals, energy value in the form of waste derived fuels for use in boilers, cement kiln, *etc*.
- Treatment facilities for thermal destruction, aqueous treatment, stabilization, biological treatment, *etc*.
- Disposal facilities like secured landfill with leachate handling, and long term monitoring facilities.

Considering capital intensiveness and perceived environmental liabilities of such facilities, the cost of treatment and disposal of waste becomes substantial. Therefore the option of waste reduction at source should be explored first by the generators of waste.

3.2 Pollution prevention

Pollution prevention, as defined under the USA Pollution Prevention Act of 1990, means source reduction and other practices that reduce or eliminate the creation of pollutants through increased efficiency in the use of raw materials, energy, water, or other resources, or protection of natural resources by conservation. Recycling, energy recovery, treatment and disposal are not included within the definition of pollution prevention [5]. Pollution prevention has a broader coverage; it includes managing chemicals to reduce risk, identifying and estimating all releases, waste minimization, and continuous environmental improvement approach. A properly implemented pollution prevention strategy would give industry the opportunity to have best of both the worlds i.e. increased profitability, and reduced environmental liabilities.

3.3 Waste minimization

US EPA defined waste minimization in 1986 as the reduction, to the extent feasible, of hazardous waste that is generated or subsequently treated, stored or disposed of. It includes any source reduction or recycling activity undertaken by generator that results in either the reduction in total volume or quantity of hazardous waste, or the reduction of toxicity of hazardous waste, or both, so long as the reduction is consistent with the goal of minimizing present and future threats to human health and environment. The adoption of waste minimization can reap several benefits, *e.g.* waste generated gets reduced thereby pollution potential is reduced, and environmental performance gets improved; environmental risks and liabilities get reduced; cost of waste treatment and/ or disposal gets reduced, working conditions in industry get improved; the consumption of feedstock gets reduced and to that extent input costs get reduced. These result into improved productivity, positive impact on competitiveness, and thereby on profitability of the enterprise. Thus waste minimization results into a situation that is a *win-win situation* for all the stakeholders.

Waste management audit methodology [6] consists of four steps, shown in Figure 5. The output from the first step of "Planning and organization" shall be in the form of a preliminary assessment of the organization and getting commitment from its management to proceed. After the second step of "Assessment phase", the output shall be in the form of an assessment report of the selected options. The third step of "Feasibility analysis phase" shall result in the output in the form of final report containing recommended options. The result of the last step of "Implementation phase" shall be successfully implemented waste minimization project. The steps described in this methodology form a close loop, and after completing a waste minimization project, one should continue this as an ongoing exercise by selecting new assessment targets and re-evaluating the options considered earlier, and repeating all the steps.

FIG. 5 WASTE MANAGEMENT AUDIT METHODOLOGY

Waste minimization can be primarily undertaken in the following stages suggested by Hirshorn and Oldenburg:

1. *Commonsense waste reduction*- immediate opportunities without having to change production processes e.g. segregation of hazardous and non- hazardous wastes, reduction in water usage in cleaning equipment, using final wash as first was in the next cycle, *etc*.
2. *Information-driven waste reduction*- relatively easy opportunities resulting from simple changes in processes based on information on successful implementation in similar industries e.g. replacing a solvent, recovery from a waste stream, installing improved process controls and instrumentation, *etc*.
3. *Audit-dependent waste reduction*- opportunities requiring formal audit methodology, described above.
4. *Research and development-based waste reduction*- opportunities arising out of technological developments and innovations in manufacturing process/equipments.

There are several constraints in the implementation of waste minimization. The prime constraint is that of mindset. There are general perceptions among industry that adoption of waste minimization is expensive, it requires technology, my technology (which is supposed to be secret) would get leaked to others, my process and other operations are the best, as long as I make profits where is the need of improvements, management tools are expensive to implement, and that I know everything. For taking up waste minimization, an enterprise should be having open mind that there is always some scope of improvements, it should be open to ideas and participation with team-work spirit, with a clear understanding that pollution control alone is not environmental management, waste minimization helps in reduction of cost of pollution control, it helps in increasing productivity, and that it is a voluntary effort.

3.4 Industrial ecology

In a traditional model, an industrial activity consists of individual manufacturing processes that take feedstock, generate products for sale and waste for disposal. In an integrated model, industrial system is considered in which consumption of energy and materials is optimized, waste generation is minimized, and wastes of one process serve as feedstock for another process. The central concept of industrial ecology is the evolution of the industrial system from an open (linear) system, where resources are consumed and harmful wastes are discharged into the environment to a more closed (cyclic) system in such a way that no waste would leave the industrial system or have negative impact on natural system. Such symbiotic industrial arrangements would have positive effects *viz.* reduction in the usage of the virgin materials as resource inputs, reduction in pollution, increasing systemic energy efficiency leading to reduce energy use, reduction in volume of waste products requiring disposal, with the added benefit of preventing disposal-related pollution, and increasing the amount and types of process outputs that have market value.

3.5 Green chemistry

Paul Anastas/ John Warner defined green chemistry in 1991 as the utilization of a set of principles that reduces or eliminates the use or generation of hazardous substances in the design, manufacture and application of chemical products. Green chemistry is based on the following twelve principles:

1. Prevention
2. Atom Economy
3. Less Hazardous Chemical Syntheses
4. Designing
5. Safer Solvents & Auxiliaries
6. Design for Energy Efficiency
7. Use of Renewable Feedstocks
8. Reduce Derivatives
9. Catalysis
10. Design for Degradation
11. Real-time analysis for Pollution Prevention
12. Inherently Safer Chemistry for Accident Prevention

3.6 Total quality management

The use of basic principles of total quality management helps industry in pollution prevention by way of reducing the requirement of inputs including raw materials, energy and other utilities, reducing rejects, improving quality of products, and reducing generation of waste and thereby reducing cost of treatment and disposal. This also results in better compliance of environmental regulations, improvement of productivity, and lowering of cost of production.

3.7 Cleaner production

The term cleaner production came into general use through the efforts of United Nations Environmental Program (UNEP) Cleaner Production program, established in 1989. Cleaner production is continuous application of an integrated preventive environmental strategy to processes and products to reduce risks to humans and environment. It includes the development and use of hard technologies and management tools.

For production processes, cleaner production includes conserving raw materials and energy, eliminating toxic raw materials/intermediates, and reducing the quantity and toxicity of all emissions/ discharges/ wastes before they leave a process. For products, the strategy focuses on reducing impacts along the entire life cycle from raw material extraction to the ultimate disposal of the product after usage. The objectives of cleaner production are to strive for efficiency and quality in products and manufacturing processes by avoiding generating waste in the first place, and minimizing the use of raw materials and energy. The scope of cleaner production includes product modification/changes, inputs substitution, process/technology changes/ innovations, equipment modifications, better process control, on-site recycling and reuse, production of useful by-products, operation and maintenance (practices) improvement, good house-keeping, and training of personnel. The special features of cleaner production, which may be highlighted, include preventive approach, multi-media consideration- air, water, and land, reduced risk to workers, cost savings, improved efficiency, lower environmental handling costs, lower regulatory compliance cost, and reduced long term liability. It often involves low- or no-technology approaches especially for reaping low hanging fruits.

The cleaner production audit methodology essentially aims at understanding basic issues like where wastes are generated, why wastes are generated, and how wastes can be eliminated or minimized, as shown in Figure 6. The cleaner production audit methodology [7] consists of six steps, shown in Figure 7.

The first step in the audit is to make a beginning. It consists of designating a cleaner production team representing different functional areas, which is capable of identifying opportunities and developing solutions. Involving an external expert in the team may be useful. The team should enlist all the processes including production, utilities, material handling, and storage, and occasional processes like cleaning and maintenance, identify all the inputs and outputs *viz*. materials, water, energy, and waste in different forms; identify and prioritize wasteful process steps related to economics (monetary losses associated with waste streams), environment (quantity and composition of waste streams), and technical (opportunities for improvements). With some output from the first step in the form of selection of audit focus, the second step consists of analysis of processes. For this, the team should prepare process flow chart for audit focus by identifying all the unit operations, linking these operations with material flows, and material and energy balance. The data sources to be used may include on-site measurements, production records, and procurement and sales records. It has to be ensured that the data used is representative, accurate and reliable. All the waste streams should be assigned cost, including internal costs (to reflect waste collection and handling, treatment, and loss of raw materials, products and utilities) and external costs (charges for discharge, consent, authorization, etc). Thereafter processes should be analyzed to understand waste resulting out of product specifications, quality of inputs, production technology, process efficiency, equipment design, layout, operation and maintenance, etc.

With the output from the second step in the form of list of waste sources and causes, the third step involves generation of cleaner production opportunities. For this purpose, the team holds brainstorming sessions; ideas are solicited from different stakeholders, technology surveys and benchmarking is carried out. Some of the preventive measures are considered by way of product modification, input change, technology change, equipment modification, improved process control, on-site recycling and/ or recovery and reuse, production of byproducts, good house keeping, etc. as opportunities. The workable opportunities may fall in categories like feasible, non-feasible, and those options requiring further feasibility analysis. The preliminary evaluation should be carried out by the respective experts on the basis of expected waste reduction, technical feasibility, financial viability and ease of implementation.

Using the output from the third step in the form of list of cleaner production opportunities, the fourth step involves selecting cleaner production solutions by detailed assessment of technical feasibility and financial viability. The environmental improvements by way of reduction in the generation of pollutants, toxicity, energy consumption, other inputs, and the pollution load are

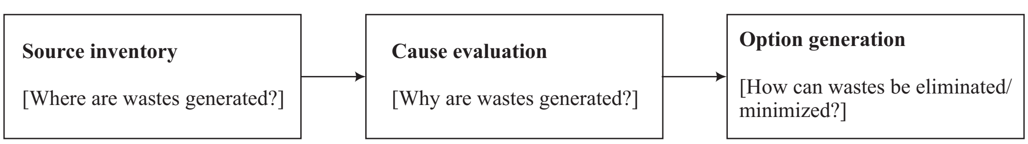


FIG. 6 CLEANER PRODUCTION AUDIT - BASIC INGREDIENTS

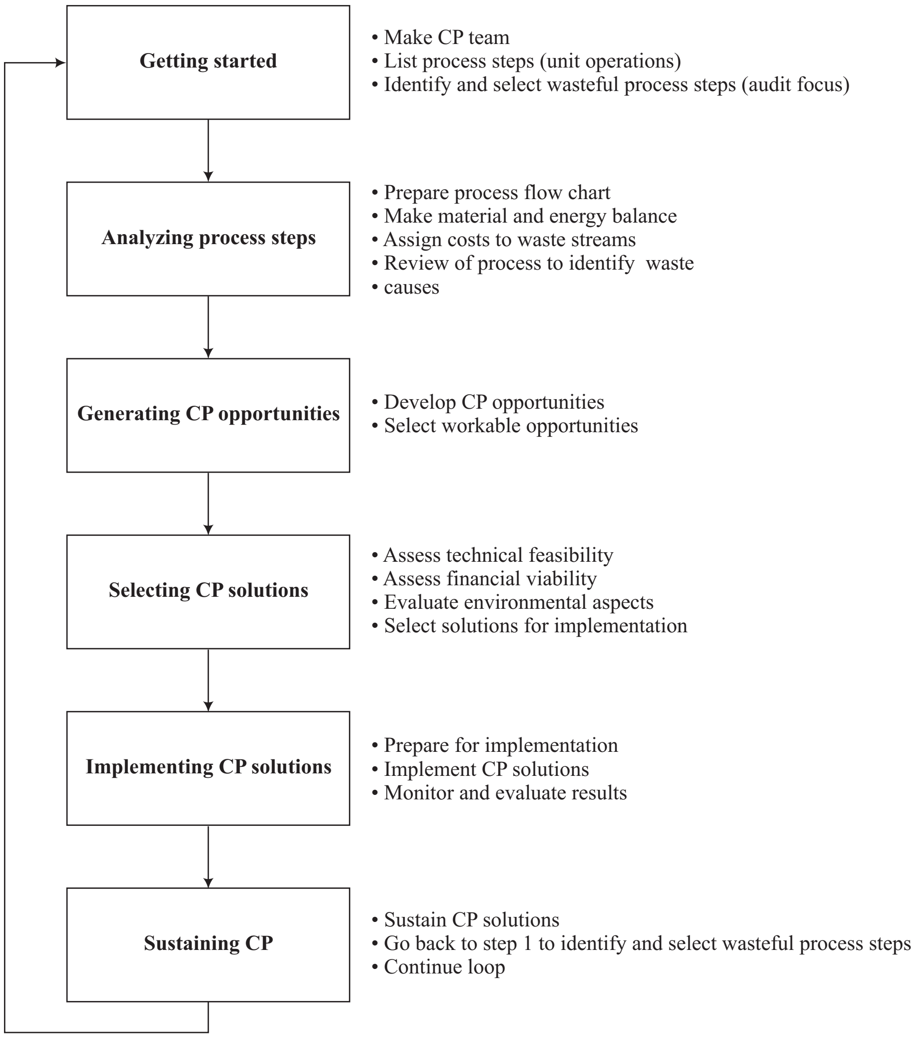


FIG. 7 CLEANER PRODUCTION AUDIT METHODOLOGY

evaluated. Based on the evaluation of technical, financial and environmental factors, cleaner production solutions are selected for implementation. At this stage, anticipated results and benefits are properly documented to present them as bankable projects and for subsequent monitoring. The fifth step is implementation of cleaner production solutions. It involves activities *viz.* detailed specifications and design, procurement, construction, and commissioning. Monitoring is carried out with the objective of evaluation of benefits vis-à-vis those anticipated before implementation. With the successfully implemented cleaner production solutions, the sixth step should aim at sustaining cleaner production initiative. This would require putting in place a mechanism that will facilitate forming long-term strategies and policies for cleaner production. Such a mechanism should ensure that the concepts of cleaner production become integral to every activity carried out in/ by the organization, and cleaner production becomes an ongoing process.

Some general prerequisites for cleaner production audit include common sense, understanding of law of conservation of mass *viz*. material balance and water balance, law of conservation of energy *viz*. energy balance, laws of thermodynamics, and knowledge of technology and operations.

Cleaner production merits consideration because it is the most cost-effective way to operate processes and to develop and produce products, to avoid/ minimize costs of wastes and emissions, and negative environmental and health impacts. The implementation experience has revealed that as a rough guide, 20-30% reductions in pollution can often be achieved with no capital investment required, and further 20% or more reduction can be obtained with investments which have a payback period of few months, and in many cases the adoption of cleaner production can reduce or even eliminate the need of effluent treatment facilities.

The implementation of cleaner production requires a different approach to plant design, manufacturing operations, and maintenance, and emphasis on the development of innovative technologies to cut emissions/ discharges rather than applying expensive control technologies. One should start right from the strategies of manufacturing operations i.e. product design, process design, plant/ equipment lay-out, data and control, personnel, organization, research and development, and interaction with vendors. It should be realized that waste streams (whether solid, liquid, or gaseous) often contain unreacted raw materials, impurities present in the reactants/ other inputs, undesirable by-products, spent auxiliary materials like catalysts, solvents, and lubricants, off-specification products, maintenance wastes and materials, material generated during start-up and shut-down, material caused by upsets in process or by spills, material generated during product and/ or waste handling, sampling, storage or treatment, and fugitive sources. These aspects should be taken into consideration right from the process design stage.

Wastes arising out of process operations could broadly be classified as variable wastes and fixed wastes. The quantum of variable wastes generated is linked to production i.e. these are intrinsic wastes, and the quantum of fixed wastes generated normally follows some standard for a particular operation i.e. these are extrinsic wastes. Intrinsic wastes are thus built into the original product and process design, and extrinsic wastes are more of functional in nature. Intrinsic wastes could be eliminated or reduced by several means e.g. R&D appraisals, major equipment modifications, major process modifications, improving manufacturing processes, better process control, better separation technology, and by changing the flow-sheet, whereas extrinsic wastes could be reduced by better administrative controls, improved maintenance procedures, additional maintenance, minor modifications in equipment/ materials of construction, training of plant personnel, better supervision, and good house keeping.

Organizational aspects are extremely important for successful implementation of cleaner production. Cleaner production should be the commitment at all the levels, starting right with the top management. It is a matter of focusing attention on where wastes come from, looking for opportunities to reduce them, and tightening up operations. Therefore, it can be cost effective. It is not the absence of technology that keeps holding up the things but it is a level of commitment, engineers' mindset, and matter of management of organization that is responsible for maintaining *status quo*. Plant personnel can reduce extrinsic waste but the reduction of intrinsic waste requires coordinated team work of operators, engineers, R&D personnel and top management, and also the support from other staff functions like purchase, marketing, financial and personnel in addition to close interaction with the suppliers of raw materials, equipment and other items.

Whereas cleaner production can be readily adopted by large enterprises, there are some barriers to its introduction in small and medium enterprises; these could be categorized as:

- Organizational - non-involvement, and high turnover of technical and production personnel
- Systemic - poor record-keeping, weak management, lack of training
- Technical - lack of monitoring or analytical facilities, limited trained personnel, lack of access to technical information
- Economic - lack of finance, weaknesses in traditional accounting systems adopted
- Attitudinal - poor housekeeping, resistance to change, job insecurity
- Governmental - counteractive resource pricing policies, emphasis on end-of-pipe approach, lack of incentives for proactive measures
- Other - lack of institutional support, lack of public pressure, lack of enforcement of environmental regulations

Some practical tips would be helpful to the professionals wanting to implement cleaner production. These are: following good industrial practices by checking the quality of all input materials prior to their use, emptying all the containers completely, employing bulk handling facilities, following good working practices, avoiding spillages through worker training, improved handling equipment and accident prevention program, and carrying out preventive maintenance. Elaborate process assessment should be carried out to include some of the probing questions like: What operations contribute most to hydraulic load? Can washing, rinsing be reduced? Is it possible to recycle wash water? What is the level of fugitive emissions? What components pose greatest problem in effluent? Can these be replaced? Can scrubber liquors or collected dust in filters be recycled/reused? Can excess reactants, precipitating agents, etc be reduced? Can cycle times be reduced (to save heating energy)? Can by-products be recovered for reuse or sale? How much off-specification product is produced? Can this be reduced by better process control? Is proper temperature control achievable? For optimizing washing of filter cake are test parameters available? One may explore whether some of the prescriptions like using pressure wash equipment for clean-outs, conserving heat (recover dryer heat, improve oven insulation, etc.), minimizing sample size, and returning residual samples to process, running longer campaigns to reduce clean-outs, better packaging design for purchased as well as sale of products to minimize dusting and damage during storage, handling and transportation, bulk supplies in reusable containers (*e.g.* ISO containers), greening supply chain, etc. could be implemented.

The framework of Environmental Management Systems (EMS) as per ISO 14001 involves environmental policy, planning, implementation and operation, checking and corrective action, management review, and continued improvement. The continued improvements could be achieved by following the principles of cleaner production. Thus cleaner production is an integral part of EMS.

3.8 Institutional framework for cleaner production in India

Several agencies have taken initiatives in propagating the concepts of cleaner production in India. In addition to the initiatives taken by the Ministry of Environment & Forest Government of India, National Productivity Council has prepared an action plan for the country on cleaner production under a project funded by Asian Development Bank. National Cleaner Production Centre (NCPC), set up jointly by National Productivity Council and UNIDO, has been advocating the concept of cleaner production. State Cleaner Production Centres have been set up with technical assistance from NCPC in the states of Karnataka, Gujarat, Punjab and West Bengal. Gujarat has taken initiative in promoting Regional Cleaner Production Centres in some industrial estates [8]. The cleaner production is expected to be more readily acceptable to industry if international as well as multilateral funding agencies work with the industrial development agencies, project/ process professionals, and the financial institutions [9]. The cleaner production movement in the country has no doubt picked up but slowly. The visionary and proactive policies, missionary zeal and dedicated efforts at different levels are required for wider and quicker acceptance of cleaner production for deriving desired benefits.

Central Pollution Control Board has brought out Charter on corporate responsibility for environmental protection (CREP), addressed to corporate bodies as well as regulatory agencies, for coming to a voluntary agreement. CREP is aimed at enhancing environmental performance of some industrial sub-sectors like cement, pesticides, petrochemicals, pharmaceuticals, pulp and paper, sugar, tannery, chlor-alkali, distillery, dyes and dye intermediates, fertilizers, and oil refineries. CREP is aimed at getting commitment for partnership and participatory action of the concerned stakeholders in preparing road map for progressive improvement in environmental management systems not necessarily limited to compliance of end-of-the-pipe effluent and emission standards. Under this agreement targets are set ahead of effluent and emission standards. CREP is also expected to help achieving waste minimization from industrial sector.

3.9 Cleaner technologies

"Cleaner technology" has become a *buzzword* for policy makers as well as environmentalists in developing economies. Scientists and technologists have always been researching *interalia* on newer technologies for existing products with some specific objectives such as improving yields, operating the plant under less severe conditions of temperatures and pressures, generating less waste, reducing cost of energy and utilities, and improving safety features. Process industry has witnessed several developments as well as improvements in several production processes and products without using the *cleaner technology* terminology. Technological gaps between developed and developing economies continue to be wide in addition to strong perceptions that the cost of development of technology is higher than the gains likely to accrue. There are no benchmarks on the consumption of resources and generation of wastes for producing many products in our country, possibly because of quite broad based manufacturing base i.e. a product may be getting manufactured by a small-scale industry (which in fact is a tiny unit in the global context), a medium, and a relatively large scale enterprise, often using primitive batch operations on one end and sophisticated process control and instrumentation in continuous operations on the other. A number of products are often manufactured in a given facility, and data is not properly recorded for each product especially with respect to the usage of energy, utilities, and waste generation. Cleaner technology includes both the manufacturing technology, and environmental technology. For quantum jumps in achieving cleaner production, cleaner technologies are essential. The induction of cleaner technologies in developing countries [10] should start with the identification of dirty technologies/ products, followed by a systematic and conscious action plan of phasing out such technologies and products as illustrated in Figure 8. The next stage of induction of cleaner technology that merits consideration is development of suitable cleaner environmental technologies for the utilization of large quantity of wastes generated from different industrial activities [11], as depicted in Figure 9.

Under the Basel Convention on the control of transboundary movement of hazardous wastes and their safe disposal, 1989, international cooperation is to be extended to developing countries in transferring technology and management systems, developing and implementing new environmentally sound low-waste technologies, and improving existing ones to eliminate the generation of waste, as far as practicable, and studying the economic, social and environmental effects of adopting the new technologies, and developing and promoting environmentally sound management of hazardous and other wastes.

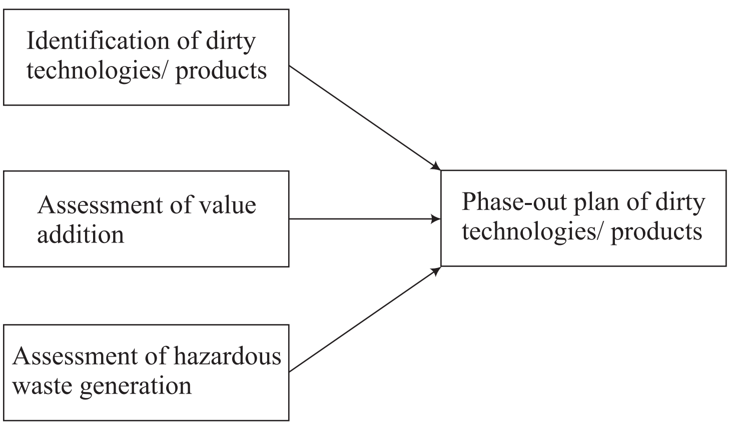


FIG. 8 TECHNOLOGIES/ PRODUCTS PHASE OUT

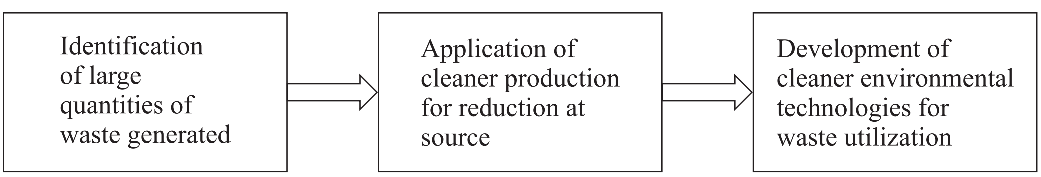


FIG. 9 CLEANER ENVIRONMENTAL TECHNOLOGIES DEVELOPMENT

4. MANAGEMENT OF BULK INDUSTRIAL WASTES

Bulk of industrial waste generated by thermal power plants and chemical industry consists of fly ash, gypsum sludge, iron oxide, spent activated carbon, and sludge from industrial wastewater treatment (both primary as well as secondary operations). Secure land filling of hazardous waste cannot be the only option on a long-term basis [12] since land is scarce, precious, and has several competing alternate uses especially in highly populated regions. Further, long-term implications of maintenance of such hazardous waste sites need to be considered seriously. Land filling option for all types of solid hazardous waste is thus not sustainable.

The problem of disposal of fly ash from thermal power plants is being discussed for a very long time. Several efforts are being made for finding applications of fly ash in making some useful products. Considering alumina, iron and silica content in the fly ash, one of the potential and promising uses of fly ash could be as a feedstock (raw-mix) along with limestone and clays for manufacturing cement [10]. If this usage could be suitably established by doing some developmental work, fairly large quantity of fly ash could get utilized in cement industry in addition to the well-accepted practice of blending of some quantity of fly ash during grinding of clinkers for manufacturing pozzolana cement. A large quantity of gypsum waste containing some traces of organic chemicals gets generated from chemical industries. Cement industries have started using such gypsum if it is colorless. However, large quantities continue to be disposed of in the secured landfills. As a long-term measure, instead of continuing the present mode of disposal, efforts should be directed in gainfully utilizing the basic constituents of gypsum viz. calcium and sulfur. It is learnt that the then East Germany had a plant consisting of rotary kiln, similar to the one in cement plants, to produce from gypsum calcium oxide aggregates that could be used for road construction, and sulfur dioxide that could go in the manufacture of sulfuric acid. In addition to rotary kiln technology, vertical shaft kiln technology could also be explored, based on the experience of small cement plants in India, for decomposing gypsum sludge into oxides of calcium and sulfur. A large quantity of chemically contaminated iron oxide gets generated in chemical industries. This has potential of utilization in the existing steel plant furnaces, and in bricks manufacturing in addition to red oxide primer.

The sludge from secondary treatment of industrial wastewater is essentially dead biomass having low calorific value compared to fossil fuels. In addition to its direct use as low-grade fuel, or as fuel supplemented with other conventional or agro-waste fuels in boilers or high temperature furnaces, it could be digested for biogas generation. This, however, would require development of suitable technology to overcome inhibitory effects of heavy metals, if any. Rock phosphate is used as a feedstock for the manufacture of inorganic chemicals like phosphoric acid. No organic chemicals are involved during the manufacturing process. Waste generated from such a process need not be considered as hazardous waste for secure land filling. There is a need of following a different approach for the management of such wastes. This necessitates developmental work for proper treatment, finding suitable uses, and for disposal of residual waste. Some proprietary powders are available in the market that could be used alongwith portland cement to stabilize and harden the hazardous waste. The hardened mass can be used as a building material for pavements, construction of walls, roads, *etc*. It is claimed that these treatment processes can avoid incineration and thereby all the associated problems, while recycling the materials. Such treatment processes could be evaluated to ensure their environmental friendly applicability in the management of hazardous waste.

Large quantities of dilute spent acids as well as spent solvents get generated from several industrial processes. The dilute spent acid streams are often neutralized with lime solution, generating sludge that gets disposed in secure landfills. The dilute spent solvents are often mixed with wastewater streams, increasing organic loading in the wastewater treatment plants. These acids as well as solvents could be reused if properly concentrated in common facilities. This will also help in the conservation of natural resources like sulfur and hydrocarbons. Used packaging materials provide scope for reclaiming basic polymers, and/ or energy, if properly collected and segregated. Waste plastics can be converted into green/ clean fuels, having low sulfur content. Flakes/ pellets of plastics can be prepared for boilers, cement kilns, etc. The development of suitable techniques for converting hazardous solid waste into non-leachable substance using locally available materials, and into vitreous form will help in reducing pressure on secure landfill facilities.

5. EXAMPLES OF CLEANER PRODUCTION IN PRACTICE

Some examples of cleaner production being practiced by industries by taking some simple and well-known measures are highlighted below:

Equipment Modification

- Replacing shell and tube condensers with plate heat exchangers in solvent systems to reduce solvent loss.
- Installation of high efficiency burners in boilers to save fuel.
- Installation of high efficiency motors to reduce energy consumption.

Input Substitution

- Isoproturon, a herbicide, was being manufactured by multinational companies using methyl isocyanate (MIC). A process was developed by an Indian company that replaced MIC with urea, which is much less toxic.
- Methanol was conventionally used as a solvent in the manufacture of monocrotophos, an insecticide. This was replaced with water.
- The usage of dry hydrochloric acid instead of using concentrated sulfuric acid for removing moisture from wet hydrochloric acid avoided generation of spent sulfuric acid.
- Replacing activated carbon by ion exchange resins avoided hazardous waste generation in the form of spent carbon.
- Chloroflurocarbons were replaced with carbon dioxide in compressors and refrigeration systems to minimize greenhouse gas effect.
- Replacing naphtha with natural gas in a captive power plant resulted in the reduction of carbon dioxide emissions substantially, enabling the company in getting carbon credits under Kyoto Protocol.

Better Quality Inputs

- The usage of good quality salt reduced brine sludge generation in caustic chlorine plants.
- The usage of lime with more than 90% purity in place of low cost low purity lime in effluent treatment plant reduced sludge generation substantially.

Recycling of Resources

- Two stage scrubbing of hydrochloric acid vapors with water followed by caustic solution scrubbing helped in recycling hydrochloric acid, reducing caustic consumption, and reducing waste in the form of sodium chloride.
- In washing operations, second wash recycled as first wash minimized water consumption.
- By developing a suitable technology for grinding ceramic fired waste, it was possible to recycle the finely ground powder, resulting in conservation of ceramic raw materials and minimization of waste to be disposed.

Reuse/Recovery of By-products

- Hydrogen sulphide generated in phorate manufacture is absorbed in caustic solution to produce saleable sodium hydrogen sulphide.
- Sulfur dioxide generated in sulfonation reactions is scrubbed with caustic solution to produce sodium bisulphite.
- Sulfur dioxide generated in the manufacture of pivolic acid is recovered and bottled for reuse in the thionyl chloride production.
- Methyl chloride generated in agrochemical manufacturing is recovered, which was earlier getting vented off.
- Ammonium chloride generated in trimethyl phosphite manufacture is converted into ammonium/ calcium chloride, and dicalcium phosphate is also recovered.
- In the manufacture of acephate, ammonium sulfate and ammonium acetate also get produced. By reacting this waste liquid stream with sulfuric acid, it was possible to recover acetic acid, and thereby reduce wastewater quantity.
- Soda ash manufacture produces wastewater containing sodium chloride and calcium chloride. By fractional crystallization in multiple-effect evaporators, it was possible to recover water and sodium chloride for recycling, and saleable calcium chloride, and thereby reducing pollution load.
- Use of membrane technology helped in concentration and recovery of by products from wastewater and reuse of water.

Waste Exchange

- Waste generated in the manufacture of phosphorus from rock phosphate is used as feedstock in some other industries *viz.* silica slag in cement industry and ferro slag in foundries.
- Acidic wastewater generated in the manufacture of specialty chemicals is often neutralized

with lime, generating a large quantity of gypsum waste containing some traces of organic chemicals. Cement industry uses such colorless gypsum in blending with cement.

6. SUMMARY

It is now realized that fundamental solutions to the industrial pollution problems lie in improvements in operating practices, work culture, product design, and manufacturing technology. The experience of various industries world-over, which have implemented cleaner production measures, reveals that with proper house keeping and some minor modifications in the operating practices and techniques, it is possible to achieve 20-30% improvement in productivity. For further improvements, technology interventions are required. The rate of return on investments in waste minimization projects is often found to be very attractive. After adoption of cleaner technologies in manufacturing, the next stage that merits consideration is development of cleaner environmental technologies for the utilization of large quantity of wastes generated from various industrial activities. For improving productivity and profitability, several management tools are employed in business and industrial operations *viz*. financial management - financial audit and cost audit, environmental audit, energy audit, bench marking, etc. Environmental audit reports should be used by industries as a tool for improvements, and by Pollution Control Boards for benchmarking, monitoring and information dissemination. There are immense opportunities for improvements in unit operations, unit processes, material handling, operating practices, house keeping and working conditions. These would go a long way in waste reduction and thereby reduction in cost and also environmental risks. Indian industry has become quite mature. Industry should plan to go beyond compliance of statutory requirements. There is enormous potential of waste reduction in almost every sub sector of industry. Waste reduction would invariably be cost-effective in addition to leading to several intangible benefits. Pollution Control Boards should act as facilitators for waste reduction while enforcing environmental regulations under relevant Acts and Notifications. The authorities should appropriately price all the resources including power and water to reflect their true price. Govt should phase out financial incentives for end-of-pipe treatment facilities, and extend such incentives for waste minimization and resource conservation efforts of industries-individually or on a collective basis. Industries associations and/ or the agencies promoted by the industries associations could be more suitable for promotion of cleaner production. They should take up such an activity especially for small and medium enterprises. Academic institutions can play a vital role in propagating the concept of cleaner production in industry if they get interested in taking up cleaner production-related activities as a part of industrial projects for their students. This will also help students getting exposure to real issues of industry. Adoption of cleaner production especially in developing countries facing resource crunch is in their national interest and everyone should support it by taking few steps. This would lead towards sustainable development.

One may propagate the concept and practice of cleaner production/ waste minimization, taking cue from a Chinese proverb

If you are thinking one year ahead, plant rice,

If you are thinking ten years ahead, plant trees and

If you are thinking hundred years ahead, educate the people.

7. REFERENCES

(1) A.K.A. Rathi, "Issues facing Indian industry in implementing cleaner production", *Proc. of National Conference on Pollution Prevention & Control in India*, Nagpur, India, 2-3 March 2002, pp 113-116.

(2) *Pollution Prevention and Abatement Handbook- Towards Cleaner Production*, Annual meetings edition, The World Bank Group, 1997.

(3) A.K.A. Rathi, "Environmental management: some issues and policies". *Indian J. Environmental Health*, Vol 43, No.3, July 2001, pp 119-127.

(4) Rajiv K. Sinha & Sunil Heart, *Industrial and Hazardous Wastes – Health impacts and management plans*, 1st edition, Pointer Publishers: Jaipur, 2004.

(5) David H.F. Liu & Bela G. Liptak, *Environmental Engineers' Handbook*, 2nd edition, Lewis Publishers: New York, 1997.

(6) Michael D. LaGrega, Phillip L. Buckingham & Jeffrey C. Evans, *Hazardous Waste Management*, 2nd edition, McGraw Hill: Singapore, 2001.

(7) *Training Package (Advanced Level) for Intensive Training Program on Cleaner Production*, National Cleaner Production Center: New Delhi, 2000.

(8) A.K.A. Rathi, "Cleaner production for industrial pollution abatement - Gujarat experience". *Chemical Engineering World*, Vol 36, No.11, November 2001, pp 58-68.

(9) A.K.A. Rathi "Promotion of cleaner production for industrial pollution abatement in Gujarat (India)". *J. Cleaner Production*, Vol 11, No.5, May 2003, pp 583-590.

(10) A.K.A. Rathi & K.V. Bhanujan, "Cleaner technology development - some practical steps for developing countries". *J. Environmental Science and Engineering*, Vol 46, No. 4, October 2004, pp 257-260.

(11) A.K.A. Rathi, "Waste minimization: Gujarat experience", *Proc. of USEFI Fulbright Indo American Environmental Leadership Program on Hazardous Waste: Strategies for Minimization,* Ahmedabad, India 9-11, December 2004, pp 32-33.

(12) A.K.A. Rathi, "Out of landfill- hazardous waste management". *The Economic Times*, 11 January 1998.

Final – \\server\e\book work
3-06-10

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

ENVIRONMENTAL IMPACT ASSESSMENT: AN OVERVIEW

A.K.A. RATHI*

1. HISTORICAL DEVELOPMENTS

Environmental concerns arose from historical global developments including

- Industrial revolution
- Agricultural revolution
- Population increase

Industrial revolution in the late eighteenth century led to usage of machinery for mass production requiring movement of raw materials and other inputs from the places of origin to distant places where manufacturing facilities were installed. This gave a thrust to transportation and infrastructure sectors. It resulted in the migration of skilled as well as unskilled labor force. These developments had significant impact on the social and cultural values of societies. The influence of industrialization had a profound effect on the socioeconomic and cultural conditions in UK, and the changes subsequently spread throughout Europe and North America and eventually the world, a process that continues as industrialization increases. Trade expansion was enabled by the introduction of water transport, improved roads and railways. Urbanization was a fall out. Agricultural revolution was born from the necessity of feeding increasing population. It gave thrust to water management, soil management, agrochemicals, transportation and agro-processing sectors. The onset of these revolutions marked a major turning point in human society; almost every aspect of daily life was eventually affected in some way. Such a kind of profound changes in the lifestyle, working as well as living conditions, increased requirement of amenities, *etc*. gave rise to societal concerns on different aspects including human health, availability and sustenance of resources and environment in general.

The approach to environmental management in the developed economies till 1970s was essentially reacting to a problem when it arose and end-of-pipe control for compliance. Over a

* Former Chief Technical Adviser, Government of Gujarat, Industries Department, Gandhinagar 382017,
INAE-AICTE Distinguished visiting Professor, National Institute of Technology Karnataka, Surathkal 575025
Visiting Professor, Nirma University, Ahmedabad 382481, *drakarathi@gmail.com*

period of time, based on the experience gained and lessons learnt, the approach has undergone drastic changes. At present environmental management focuses on anticipation and prevention of the problems likely to arise, keeping in view the fact that aspirational horizon has shifted from local to *global,* and immediate and short term to *the needs of next generation.*

2. CONFLICTS AND CHALLENGES

The entire gamut of development and environment has been giving rise to a variety of conflicts from time to time, which in turn has been posing challenges. In earlier days, there was debate on whether the society wanted economic development or environmental protection. Based on the level of technological developments especially in developing countries, even till late twentieth century, it was accepted that environmental degradation was an unavoidable offshoot of economic development and one has to live with that evil. Over a period of time with globalization, free access to information, global concerns leading to international treaties, *etc*. it is realized that economic development and environmental protection can go hand in hand. This has posed major challenges to the planners, policy makers and technologists. With massive developments taking place almost on every front, human aspirations at times are in conflict with natural processes. Another conflict which poses a challenge is balancing the interests of individuals vis-à-vis that of society at large *i.e.* social equity. Often it is perceived that individuals tend to gain by implementing developmental projects at the cost of society. While enormous amount of efforts and funds are spent for environmental management, little is done in providing basic amenities to improve quality of life of people around. One school of thought is of the opinion that instead of spending so much for environmental management of a facility, priority may be given to providing basic services especially to the lower strata of the society to improve their quality of life, which may turn out to be a less costly proposition while having visible positive impacts. A conflict that is perpetual and needs to be resolved amicably is on different priorities of developed and developing/ under- developed countries. While the developed countries have already ensured basic amenities to its citizens who have high per capita income levels, very high level of consumption of resources and concomitant waste generation, the priority of developing and under- developed countries is to make provisions for basic needs of food, water, shelter, sanitation, transportation, employment, communication, *etc.* to their masses so that living standards of its citizens improve. Thus economic developments by way of massive infrastructural development, industrialization, urbanization, extensive agriculture, development of service sector and so on are undertaken. These measures, in turn, result into anthropogenic pressure on environment. In a fast developing country like India population growth, faster pace of urbanization, industrialization and infrastructural development are observed to have direct impact on air quality, water bodies, land, social and economic aspects of society, and health of living beings. Thus while the developed economies aim at further improvement of quality of life, the developing economies aim at improving living standards by way of rapid economic development. So when it comes to international treaties on long term global issues, the perceptions and priorities often differ very widely and the developed world is asked for providing special considerations, concessions and exemptions.

3. ENVIRONMENT

Merriam-Webster's Collegiate Dictionary defines "environment" as the complex of physical, chemical, and biotic factors (as climate, soil, and living things) that act upon an organism or an ecological community and ultimately determine its form and survival. Broader concept of environment, defined by Canter considers

- physical-chemical
- biological
- cultural and
- socio-economic

components of the total environment. A comprehensive definition of environment given by European Community Directive includes

- Human beings, fauna and flora
- Soil, water, air, climate and landscape
- Material assets and the cultural heritage
- Interaction between any of the above factors

The perception of the word *environment* differs among haves and have-nots. Considering various developmental projects and rapid urbanization in the developing and under-developed economies, 'environment' is often considered synonymous to 'pollution'. It is, however, now being realized that 'pollution' is only a subset of 'environment'.

4. SUSTAINABLE DEVELOPMENT

In the present context, sustainable development is *buzz word*. Sustainable development, defined in Brundtland Report, and widely accepted is *development that meets the needs of the present without compromising the ability of the future generations to meet their own needs*.

For moving in the direction of sustainable development, the following key sustainability principles have to be followed in word as well as spirit while planning developmental projects whether infrastructural, manufacturing, energy, urban or even social projects:

- Anticipate and prevent
- Exercise precaution
- Remain within source and sink constraints
- Maintain natural capital at or near current levels
- Avoid conversion of land to more intensive uses
- Make the polluter pay

Anticipation and prevention of potential adverse impacts on environment at the project conception stage itself is most desirable and cost effective. Exercising precaution rather than making efforts to cure will be much more beneficial. From Figure 1 it may be observed that for sustenance of living beings, resources are obtained from environment and waste is returned back to environment *i.e.* source and sink are not different. Therefore it is essential that one should remain

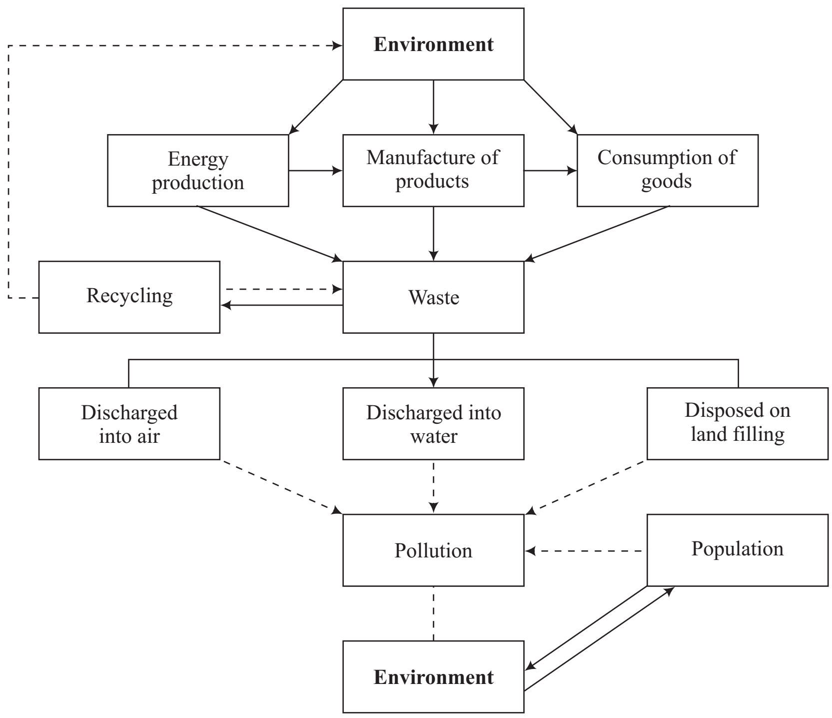


FIG. 1 ENVIRONMENT: SOURCE AND SINK

within the constraint of source and sink. Centuries back when human race was in Stone Age, environment had enormous and unlimited capacity to provide resources and also assimilate waste. Anyone could take anything from environment, whether forests or rivers and throw back waste anywhere without any adverse impacts. With increasing population and competing uses of environmental resources, resources have become finite, and the capacity of environment to act as sink for waste has also become finite. The astronauts in space crafts understand the importance of remaining within source and sink, considering the limited space available to them. This leads to the concept of 'carrying capacity' of environment. Resource management is an essential ingredient of sustainable development, which requires that natural capital does not diminish significantly. Land is a finite resource with several competing uses. Intensive use of land is likely to result into infertile land, which may turn out to be an irreversible process and thereby loss of a precious resource. Intensive use of land for aquaculture on the eastern coast of India in late 1990s has left its scars, taking a toll of several farmers. While industrialization is considered as a means of providing employment and generating revenue for government by way of taxes and duties, generation of pollution was considered unavoidable outcome of manufacturing. Pollution abatement by way of treatment (end-of-pipe control) was referred as a social responsibility by the corporate about a decade back till the judiciary enunciated 'polluter pays principle'. As a consequence, the generator of pollution is expected to manage it suitably, by incurring necessary expenditure, in accordance with the environmental regulations which are getting more and more stricter from time to time.

5. ENVIRONMENTAL MANAGEMENT PHILOSOPHY IN INDIA

The environmental dimension into the development planning in the country was introduced as far back as in the Fourth Five-Year Plan: “It is an obligation for each generation to maintain the productive capacity of land, air, water, and wildlife in a manner, which leaves is successors some choice in the creation of a healthy environment.” The Seventh Plan document further reiterated this while introducing the concept of resource management for sustainable development: “The degree to which a nation can prosper depends on its productivity, which is the efficiency with which it is able to utilize the resources of the environment to satisfy human needs and expectations…while providing for current needs, the resources base be managed so as to enable sustainable development.” The umbrella act for environmental management got enacted in the form of Environment Protection Act in 1986 by Government of India. For environmental impact assessment, a notification was issued in 1994, which was superseded by Environmental Clearance notification in 2006. Supreme Court of India, in its judgment in the case of *Vellore Citizens Welfare Forum v/s Union of India and others* observed: the traditional concept that development and ecology are opposed to each other is no longer acceptable. Sustainable development, defined by Brundtland Report… is the answer. The Court was of the view that the *precautionary principle* and the *polluter pays principle* are essential features of sustainable development. Most of the Indian environmental legislations incorporate these basic principles.

6. ENVIRONMENTAL IMPACT ASSESSMENT

Environmental impact assessment (EIA) responds to the precautionary principle for environmental management aimed at the ultimate goal of sustainable development. Considering that the focus of EIA is on environment and health and well-being of living beings, it is expected to ensure that environment is considered an essential part of development-related decisions and that impact mitigation is addressed right at the early stage of a project. An objectively carried out EIA is expected to help improved environmental decision-making, especially with respect to its

- openness
- comprehensiveness
- transparency
- robustness

It may be recalled that EIA was conceived as a *tool* to aid decisions in late 1960s in response to *ecocentric* (an extreme view, possibly over-cautious one, in which environmental degradation was suspected from each developmental activity proposed) concerns, challenging *technocentric* (another extreme view, in which it was believed that technology could offer a solution to every issue) view. EIA is considered as a *process* having the ultimate objective of providing decision-makers with an indication of the likely consequences of their actions. In actual practice EIA shall get affected by the attitudes and capabilities of those who

- take decisions
- seek to influence decisions

EIA is defined as a *technique* or *a systematic process* by which information about the environmental effects of a project is collected both by the developer and from other sources, and taken into account by the planning authority in forming their judgments on whether development may be permitted. The International Association for Impact Assessment (IAIA) defines an EIA as the process of identifying, predicting, evaluating and mitigating the biophysical, social, and other relevant effects of development proposals prior to major decisions being taken and commitments made. The objective of EIA may be summed up as

- Prediction and assessment of impacts from human activities
- Understand utility of EIA as a tool in decision making
- Balancing of environmental, economic and social concerns
- Potential of EIA to enhance sustainable development

Human induced *changes* are sometimes equated with *effects,* and *consequences* are referred to as *impacts*. Changes in *spatial* and *temporal* characteristics because of the proposed project compared with the situation without the proposed activity *i.e.* post-project and pre-project (business as usual) scenarios are to be considered. Impacts resulting from proposed project can be *beneficial* or *detrimental* to different people and/ or elements of the environment in the short- or long- term; naturally reversible or irreversible, reparable with managerial interventions or irreparable, primary (direct) or secondary (indirect), *etc*. While carrying out impact assessment, it may be appreciated that impacts on the health of human beings would be very evident from the ill effects of environmental degradation, impacts on ecosystem are recognized later and those on global environment cannot be seen but felt much later. Thus it is important to broaden one's perspective while considering environmental issues. Global concerns are being addressed on the issues including *global warming, greenhouse effect, rising sea level, depleting forests, ozone layer depletion, etc.* with an objective of minimizing further degradation of environment and of containing adverse impacts.

6.1 EIA process

Main stages involved in EIA process, primarily a decision- support system which should optimize both intrasystem decisions as well as the primary decision to implement the project or not, are:

- Activity definition/ Preliminary assessment
- EIA report preparation/ Detailed assessment
- Decision making
- Implementation and follow-up

Activity definition includes

- Consideration of need of the proposed project and alternatives - types of manufacturing processes, technologies and sites for locating the project.
- Screening- whether EIA needs to be carried out, depending on legal requirements or prima-facie significant potential impacts of the project.
- Scoping- contents of EIA to include various issues likely to be involved and different types of impacts

The detailed report preparation includes

- Detailed scoping to plan specific or site-specific characteristics
- Base-line survey/ assessment to reflect pre-project status of environment
- Impact prediction superimposed on base-line conditions - spatial as well as temporal
- Estimating/ evaluating significance of impacts
- Drafting report with environmental impact statement (EIS), a summary of significant impacts, and environmental management plan (EMP), encompassing impact mitigation measures at different stages of project implementation and post-project monitoring.

The decision making includes

- Review of EIA report by experts and decision makers
- Formal consultation with experts and/ or public
- Approval from different authorities *e.g.* Forest, Maritime, Civil aviation, Defense, *etc*., wherever required
- Approval mechanism under the concerned legal requirements *i.e.* Environmental Clearance Notification of Government of India (Ministry of Environment and Forests or State-level approval)
- Communication of decision to the project proponent whether the project could be implemented or not.

Implementation and follow up includes

- Making necessary changes in the project to incorporate conditions imposed by the decision maker
- Construction and commissioning the project while complying with all the statutory requirements and the conditions imposed
- Simultaneous implementation of the relevant provisions of EMP, and compliance monitoring- during construction as well as operational phases
- Periodic monitoring and auditing of EMP
- Revision of EMP, if required

All the above stages are closely linked with each other as depicted in figure 2. There is a close interrelationship between each of the above stages, requiring ongoing communication and participation at every step.

Major components of EIA are

- Air quality assessment
- Water (quantity and quality) assessment
- Aesthetic/ visual, landscape and land use assessment
- Ecological assessment
- Social, economic, cultural assessment
- Risk assessment
- Cumulative effect assessment
- Mitigation measures and monitoring

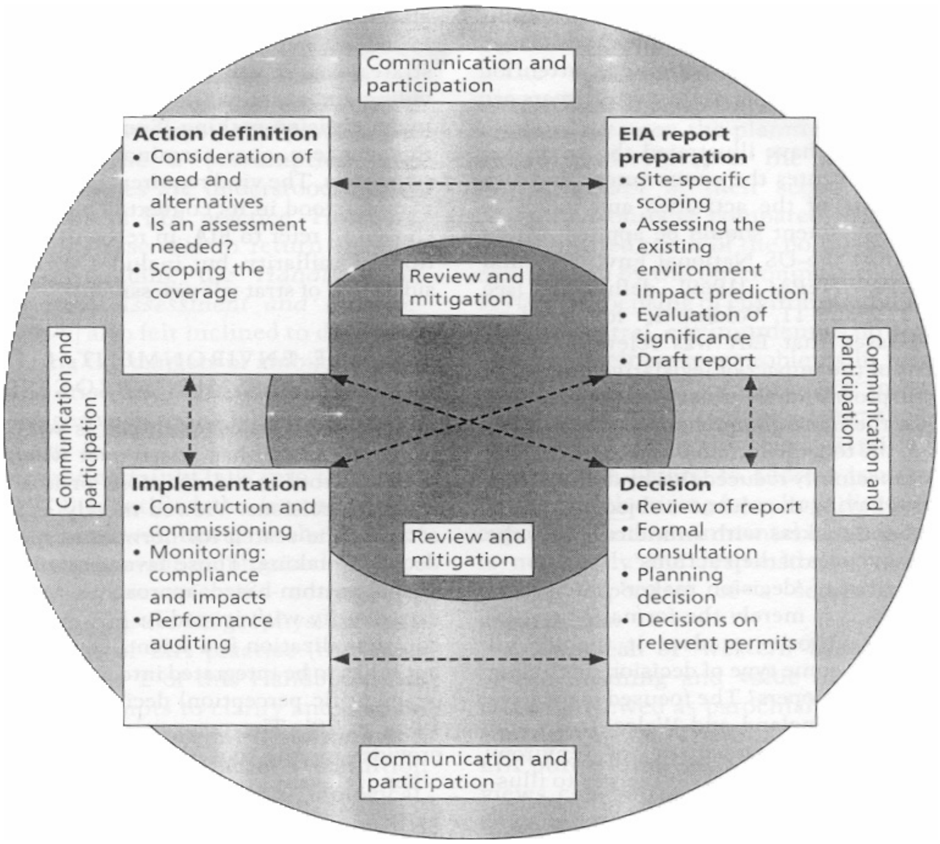


FIG. 2 EIA DECISION-SUPPORT SYSTEM [1]

6.2 EIA experience

It is reported that EIA as a formal process is used in more than 100 countries. Canadian Environmental Assessment Agency with International Association for Impact Assessment reported in 1996 that practitioners and managers must position environmental assessment as a sustainability mechanism for the 21st century. It also observed that EIA is still seen as a

- mere hurdle to overcome on the way to development
- means of appeasing environmental concerns

Indian perceptions have not been different. In India we have experience of more than 14 years of implementing EIA related regulation *i.e.* Environmental Impact Assessment notification, 1994 and several amendments thereof, superseded by Environmental Clearance Notification, September 2006. The experience shows that EIA

- is still considered as an obstacle to development
- is very time consuming

- consists of too bulky documents, more like research thesis at times
- often focuses more on mitigation measures rather than on the changes in design and location of the proposed project
- includes predictions and evaluation of multiple impacts, sometimes without focusing on the most important and significant ones
- justifies the project approval on the basis of benefits likely to accrue using subjective analysis
- is prepared just to meet the legal requirements after taking all decisions about the project

EIA is thus hardly being considered by project proponents as a long-term necessity for sustainability.

7. PRACTICAL DECISION MAKING

After assessment of each of the components of EIA, the next step is establishing environmental significance. It is not necessary that every impact is significant. To ensure that the focus is not lost, only significant impacts should be considered for mitigation and for decision making. *Significant* requires considerations of both *context* and *intensity*. Context means that significance must be analyzed relative to society as a whole, the affected region, the affected interests, the locality, and whether the effects are short- or long- term. Intensity refers to the severity of impact. The factors influencing judgments of significance include:

- Character of the receiving environment
- Nature and magnitude of the impacts
- Resilience of the impacted environment
- Confidence in the predicted impacts (data and models employed)
- Potential for mitigations
- Level of public concern

It may, however, be appreciated that

- Criteria and standards of significance are not often documented whereas regulatory standards of air emissions, water and waste discharge are specified
- Several uncertainties are involved in impact assessment, and validity/change of assumptions may cause change in environmental significance
- Qualitative criteria related to social, economic and cultural impacts require deliberations to come to some broad agreement

Practical decision making involves an ongoing complex process of *trade offs* among the objectives of the project with respect to

- Economic efficiency
- Social equity
- Environmental capacity

This may involve lot of subjectivity since the criteria to be adopted would depend on the location and status of its social as well as economic development. Any developmental project is expected to contribute something acceptable to each of the three dimensions *viz.* economy, society and the environment, as illustrated in figure 3. Contribution to any one of the aspects at the cost of some other(s) shall not be acceptable in a democratic society. Thus practical decision making involves trade offs among different objectives for integrated development that is beneficial to the society at large in the long term.

8. STRATEGIC ENVIRONMENTAL ASSESSMENT

It is being recognized that environmental impact assessment of individual developmental projects alone is not sufficient since cumulative impacts get neglected and alternatives with respect to sites, technologies, resources, *etc.* are often not considered seriously in this exercise. The project level EIA is, therefore, not adequate for achieving sustainable development, for which the concept of Strategic Environmental Assessment (SEA) needs to be applied. SEA can be viewed as the formalized, systematic and comprehensive process of evaluating the environmental impacts of a policy, plan or program and its alternatives, and the preparation of written report on the findings of that evaluation and using the findings through a transparent process of decision making. SEA is basically application of the principles of EIA to

- existing and proposed policies (both environmental and non-environmental)
- plans (sectoral and spatial), and
- programs (for implementation)

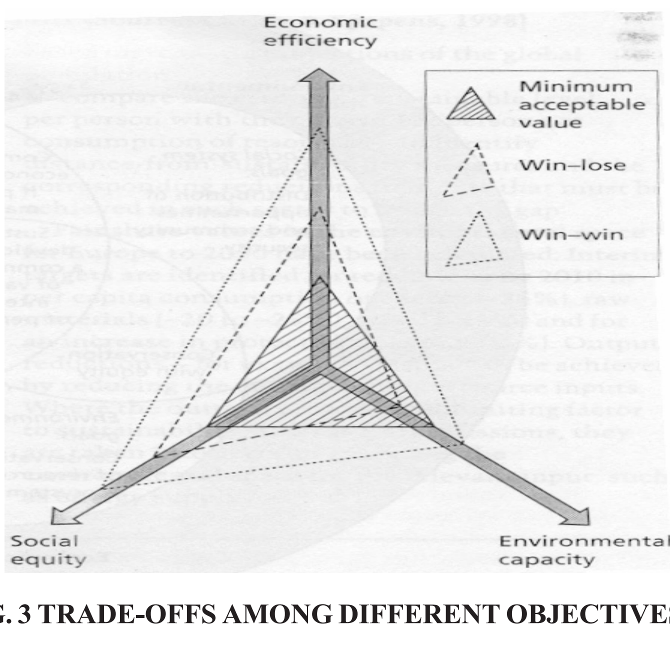


FIG. 3 TRADE-OFFS AMONG DIFFERENT OBJECTIVES [1]

This concept of SEA is being linked in the developed countries to evolve Local Agenda 21, national sustainability development plans and sustainability indicators and other approaches being adopted to implement the post- Rio Agenda. SEA can make a significant contribution towards a comprehensive and coherent system of resource management. By addressing the policies adopted for environmental protection and economic development and assessing them in a clearly structured fashion, potential conflicts can be identified, avoided or mitigated. SEA, also understood as regional environmental assessment, provides consistency across the planning system allowing individual projects to be effectively assessed in the context of overall development strategy.

9. DEVELOPMENTAL SCENARIO IN INDIA

Some of the following captions are often seen in recent business dailies/magazines:

> "In the last few years, India has arrived on the world stage and now figures in the big league. There is virtually no Fortune 500 company on this planet that doesn't think about India. If you're not thinking about India, you're doing so at your own peril", Arun Sarin, Vodafone.
>
> "India has displaced the US to become the second-most favored destination for FDI, after China".
>
> "Indian economy is second fastest growing economy in the world".
>
> "India is emerging as one of the world's largest economies".

It may be observed that Indian GDP growth rate in the financial year 2006-07 was 9.6% and it is anticipated to be around 9% in the years to come. The growth in manufacturing and services sectors is 10-11% per annum. More than 340 SEZs are approved with proposed investment in excess of Rs. 3000 billion and 4 million additional jobs. Several special corridors as well as investment regions are proposed. In the current scenario, lack of proper infrastructure is identified as a major bottleneck in the growth of the Indian economy. Infrastructure is the pillar of economic as well as social development. Sectors like power, rail, roads and ports have to improve substantially if they want to meet the ever rising demand from India Inc. Some of the following per capita statistics of China and India reveal substantial gap in some of the economic indicators:

	India	China
Per capita consumption		
Electricity, *KWH*	618	1684
Steel, *Kg*	34	244
Petroleum, *Kg*	108	287
Petroleum refining capacity, *Kg*	131	248
Cargo handled ports, *Kg*	572	4265
Passengers handled airport, *per 1000 persons*	71	151

The potential for development in India is evident from the above if Indian economy is bench-marked against that of China.

In the 11th five-year plan (2007 – 12), the expenditure proposed for infrastructural development proposed is as follows:

	Rs., Cr.
Electricity	616,526
Roads	311,816
Telecom	267,001
Railways	258,001
Irrigation	223,131
Water supply	199,127
Ports	73,941
Airports	34,748
Storage	22,378
Gas	20,500

[Eleventh Five-year Plan document, Planning Commission]

There is one opinion that the infrastructure growth in India has reached an inflexion point and historical growth may no longer act as guidance for the future. However, looking at the revolution that has taken place in the telecom sector in India, revolution in infrastructure cannot be ruled out. Infrastructure growth will have a cascading effect across a number of sectors, ranging from cement, metals and construction to capital equipment, project finance, manufacturing, tourism, *etc.*

10. CONCLUSION

With all round developments being witnessed and those planned, India is poised for a total face lift on various fronts including infrastructural, manufacturing as well as services. It is realized that massive developments are required for upliftment of the masses. For these developments to be sustainable and beneficial to all sections of the society, it is necessary that environmental management related aspects are seriously taken into consideration at policy and planning stages in addition to every stage of a project; the bottom line being environmental regulations (including *EIA as precautionary principle*) are religiously followed in spirit as well letter.

REFERENCES

1. Judith Petts, *Handbook of Environmental Impact Assessment* Blackwell Science Ltd, Oxford, 1999.
2. A.K.A. Rathi, Environmental management: some issues and policies. *Indian Journal of Environmental Health*, Vol. 43, No.3, July 2001, pp 119-127.
3. David H.F. Liu & Bela G. Liptak, *Environmental Engineers' Handbook* Lewis Publishers, New York, 1997.
4. L.W. Canter, *Environmental Impact Assessment* McGraw-Hill Inc, Boston, 1995.

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

GLOBALIZATION OF ENGINEERS ETHICS AND CODE OF CONDUCT

C.G. KRISHNADAS NAIR*

1. INTRODUCTION

Engineers play a significant role in the application of science and technology for growth of human civilization enhancing quality of life wealth and welfare. However unwise and unethical application of science and technology can lead to harmful effects causing injury to humans, animals, and environment and affecting sustainability. It is imperative that engineering education in addition to imparting technical knowledge and skills, must educate engineers on their ethical responsibilities and to act responsibly. Engineers ethical code of conduct should make engineers responsible to avoid harm to society, animals and environment and also ensure sustainable development. Many professional societies have formulated engineers ethical code of conduct. However these do not form part of engineering curricula. Engineering education mostly concern with imparting scientific and technical knowledge and skills and do not train engineers to resolve moral dilemmas, pressures from vested interests, and conflicts of interests and act responsibly. With the globalization of business, more and more engineers work in countries other than their home country and face new working environment with different cultural and religious values customs traditions and practices influencing the local moral standards. Professional societies across the world should endeavour to develop an internationally accepted code of ethics for engineers and introduce the same in the engineers' curricula. A common strategy is to be evolved to certify competence of engineers and engineering organizations for undertaking professional activity and to lay down common code of ethics for engineers in the context of globalization of engineering profession.

2. GENERAL MORALITY AND ENGINEERS' PROFESSIONAL ETHICS

Human civilizations evolved in different parts of the world at different times, and these set their standards of morality. Communities depending on their circumstances modified and adapted their

---

*Former Chairman, Hindustan Aeronautics Limited (HAL), President, Society of Indian Aerospace Technologies & Industries (SIATI), Aeronautical Society Building, Suranjandas Road, Off Old Madras Road, Bangalore 560 075

own particular set of rules of conduct. Religions had a profound influence on setting standards of general morality. Village and tribal chiefs, kings, emperors and various socio- political organizations influenced general morality from time to time. In the modern era communism, capitalism and liberalism have also made their impacts. General morality is not standard and universal. General morality appears to be diverse from community to community and place to place, although there is a great convergence with respect to the fundamentals. All religions advise human beings to be good natured and pure in thought and action and be 'GOD like'. The fundamental moral codes of conduct are really secular and are of virtuous commonsense values. Good personal ethics are developed on the basis of these secular codes.

Professional ethics are moral code of conduct as applicable to persons belonging to a particular profession, for example engineering. Professionals are empowered with knowledge & skills and training in providing value addition to society in the form of products & services.

This empowerment can also be misused adversely affecting the well being of others, society and environment. Hence professionals need to be regulated by ethical standards. These form the codes of professional ethics and are promulgated by each of the professional societies and some times regulated by the State Laws.

Most of the engineer's professional associations across the world who have published engineer's code of ethics emphasize engineers' paramount responsibility as protecting the human beings and human society from harmful effects of technology and to do so while fulfilling their professional obligation for adding value to society and enable the advancement of civilization. Some do not put similar emphasize on protection of animals and environment. There is a need to evolve a globally accepted code of ethics with respect to engineering responsibility for protecting the society and environment including animals.

Engineers must be responsible environmentalists. The engineer's work (application of technology) is essential for progress and growth of our civilization, bringing comfort and prosperity. But it can also lead to environmental problems. For example design and execution of a hydro-electric project, may destroy forests. An irrigation project, may destroy forests, land and adversely affect the eco-system. Design and execution of a fertilizer plant, very much needed for improving agricultural yield and eradication of poverty, may cause pollution to the neighboring environment. Mining of coal, minerals and metals, very much needed for industrial growth, will also disturb the local eco-system and destroy forests. Automobile exhausts continuously poison the air. Yet can we live without automobiles, without electricity, without coal, minerals, metals, chemical & fertilizers, dams and irrigation? While technology is needed for progress, and in the process it causes environmental issues, technology can also be used to reduce/eliminate the environmental problems. For example the poisonous gaseous emissions from the automobile engine exhaust can be controlled by design improvements of the engine and change of fuel. The pollutants from industries can be neutralized by chemical treatments & disposed. The lost forests can be regenerated. Engineers, while implementing technology for progress, must endeavour to design and execute projects, processes, & products in such a way to eliminate or reduce the threats to environmental integrity. Engineers as private citizens have the responsibility like other citizens to protect the environment from degradation. In addition as agents of application of technology they have professional responsibility for protecting environment from degradation & destruction.

Engineers are expected to uphold steadfastedly the safety of environment and the principle of sustainable development. While performing their professional duties engineers are also required to inform their superiors, employers, and clients of possible harmful consequences to safety, health, and environment as and when their professional judgment and actions are overruled by vested interests.

Sustainable development is a development in which the technology and resources are used to meet the present needs and aspirations without endangering the opportunities and capabilities to meet the needs and aspirations of future generations. Conversation of natural resources and protection of environment are central to sustainable development. Some people believe that nature is passive and is to be exploited for the benefit of human beings. Indiscriminate exploitation of natural resources is against sustainable development and is unethical.

The belief that nature is passive is fundamentally wrong. Man itself is part of nature. Earth with all its living organisms and the inanimate world form the earth's ecosystem which is like a living organism. In the Indian civilization, earth is worshiped as Bhoomidevi and the Hindus praised and prayed to Mother Earth for nourishing and sustaining us. The Greeks too used to worship earth as the goddess "Gaia" The 'Gaia theory', a recent scientific hypothesis, articulates this ancient wisdom and consider earth as a living organism, with all its eco-systems man, animals, birds, trees and seas and the innumerable bacteria and virus etc as part of this living earth, just like millions of living and dying cells in the human body. As the immune system in the human body, the earth also fights the dangers to its ecosystem, inflicted by human beings. Man is a part of the web of life, this eco-system. Any disturbance to the web is a disturbance to man. If humans exploit or pollute nature indiscriminately beyond the nature's capability to regenerate, then its immune system will act and destroy the very cause of irritation that is man itself. Sustainable development is thus fundamental requirement for sustaining human civilization on earth.

3. ENGINEER'S PROFESSIONAL CAREER & GROWING ETHICAL RESPONSIBILITIES

Most engineers after their academic course and training start their job in the first level of managerial hierarchy, working as supervisors, process planners, junior engineers, inspection engineers, designers, maintenance, and service engineers. As they advance in career, they take up positions of greater responsibility and authority, as middle level managers, senior managers, General Managers, Directors on the Board, and CEOs. Some engineers pursue a formal management education and graduate as MBA, and start their career as managers and advance in career in the managerial hierarchy. Some pursue research and development, or teaching and advance as scientists/academicians and become professors, engineer- scientists, directors of research labs etc. Some enter State/Central Government Administrative Services. Some many become entrepreneurs and start their own engineering firms, consulting firms and industries and become employers of other engineers, As the power and authority grow, the negative impact of wrong decisions/actions and devastation to the society and environment will grow in greater intensity and the size of the affected people and environment will also be enormously larger.

4. ENGINEERS AS MANAGERS

The term manager is used in broad sense to include supervisors and managers of employees, works and projects etc and also management at the higher level such as General Management Board level Executives, CEO's etc. Managers are often concerned with speedy task completion, reducing expenditure and maximizing the profit. There will be temptations for cutting corners with respect to production processes, testing, & quality control, pollution control etc. and compromise on safety at work place. They may direct engineer employees to comply with these demands in the name of loyalty to the organization. This is unfair to the employees and is unethical. Engineer managers are responsible for ethical conduct of their subordinates and should act as responsible leaders. They should never persuade their subordinates to do what is ethically wrong. They should not hide mistakes/lapses on ethical conduct either committed by them or by their subordinates. However the emphasis should not be on fixing the people who made the mistake but on correcting the systems to avoid hazards to society and environment. Ask why it went wrong and how it went wrong and correct the same instead of looking for people to punish. Manager- Engineers are responsible for the safety of the people working with them and they should periodically review the work environment and recommend to higher management for continuously improving safety in the work place and also safety for the environment.

Engineer managers at the corporate level as Directors and CEO's must ensure good Corporate Governance fulfilling their responsibility to the society in addition to other stake holders such as the shareholders, customers, and employees. Their strategy for growth of the industries/organizations should take into consideration respect to society environment and sustainable development. They must ensure compliance with the statutory provisions with regard to safety, pollution control and anti corruption measures. They must evolve and publish a code of conduct for all employees and disseminate the same through publications and awareness programmes. They must build a corporate culture which will value 'straight talk' and reward integrity and encourage 'dissent with discipline'. Dissent with discipline is the articulation of opposite views without fear of ridicule or reappraisal so that managers and corporate management can get honest opinions from their engineers and experts on the ethical aspects of projects, plans, or actions. In this context they must also institute an appropriate policy for protecting 'Whistle Blowers'.

5. ENGINEERS AS ADMINISTRATORS

Administrators have an ethical responsibility to formulate and implement appropriate policies for ethical conduct of engineers. They must endevour to ensure good ethical, individual, and corporate conduct and protect society and environment from hazards. They should develop and implement State laws against corruption/bribery, extortion and ensure pollution control. Administrator-engineers who take up administration as a career may also encourage professional societies of engineers to develop a code of conduct and adopt the same at the State level. They should consider their position of power and authority as a position of responsibility and service.

6. ENGINEERS AS ENTREPRENEURS/EMPLOYERS

An engineer as an employer and entrepreneur will be responsible for several employees. In this role as the entrepreneur/employer, the engineer will have considerable power, over people. This power should be used to encourage ethical conduct of employees. They should not exploit the loyalty of employees engineers to persuade them to act in favor of gains for the employer with adverse effect on the society and environment. For example, over exploitation of ground water to enhance production of soft drinks is against sustainable development. Inadequate measures for ensuring safety and pollution control to save expenditure and enhance profit will be at the cost of causing hazard to people and environment. Engineer- Entrepreneur is in the business to make money and to make profit, but it should not be at the cost of society, environment, and sustainable development.

7. RISKS, SAFETY AND LIABILITY

Engineers have a paramount responsibility to protect life and environment from hazards. Risks of hazard and safety in the work place are well recognized and there are statutory provisions in many countries to be complied with to minimize risk and maximize safety. Engineers, managers and owners of business are liable for legal action for violations. Engineers must prepare for dealing effectively and responsibility with issues of risks, safety and liability.

Concern for risk and safety has a prominent place in the engineer's professional code of ethics. Engineers must design products and structures which are safe for public to use. Engineers must make the work place including machines, tools, plant and equipment and processes safe to avoid injury and harm to health and environment. Maintenance of buildings, machines, plants & equipments vehicles etc must be done in conformity with laid down procedures and engineering standards. Engineers must accept responsibility to make engineering analysis, decisions, and actions consistent with safety, health, and welfare of workers and the public and disclose factors which may a adversely affect public and environment. Engineers are obliged to inform their superiors, employers, clients and appropriate governmental/public authorities if and when their professional judgment and advice on safety are over ruled exposing the employees, or public, or the environment to risk. Engineers must be ethically responsible for risks. They must keep themselves aware of the risks related to products, processes technology etc and the approaches to the decision on acceptable limits of risks and ensure fair play and adhere to professional ethics focused on respect to people and environment. As new knowledge and experience is acquired: acceptable limits of a particular risk may undergo change. Engineers must act responsibility keeping themselves updated on this aspect, and work towards reducing risk through technological innovations

8. CLIENT PROFESSIONAL AGREEMENTS

Engineers on their own or employee engineers on behalf of their employer provide design, consultancy, audit and such other professional services. In such cases the sensitive/confidential information provided by the client and information generated by the consultant during the specific work are protected through confidentiality agreements. The engineer's code of ethics requires strict compliance to such agreement. For example such information in the case of one client should not be revealed to another client. The engineer must refuse to break such confidentiality even under threat, or under monetary or other types of inducements. But an engineer's paramount obligation is to the safety of the public and may break confidentiality and reveal to public such information if it will impact the safety and wellbeing of the public.

9. NON DISCLOSURE AGREEMENTS (NDA)/KNOWN-HOW TRANSFER ETC

When engineering firms negotiate to collaborate, or purchase know- how of a product information is exchanged on the basis of NDAs. Both parties, and engineers involved are bound to maintain confidentiality of such information exchanged. Know how purchased from the licensor should be used only for the license, and it should not be transferred to others unless rights for transfer to third parties are specifically mentioned in the know- how transfer agreements. Engineers involved in the know- how utilization should adhere to the terms and conditions of such agreements, and act as per the intellectual property rights.

10. INTELLECTUAL PROPERTY RIGHTS

Intellectual properties are generated through research design and development. It may be a product or technology and is protected in many ways such as trade secrets and patents. Most companies make their engineers/scientists to sign agreements by which all such trade secrets and patents, are the properties of the employer, even though these are developed by the employee engineers/ scientists. Even in the case of organization which share the IP rights with the employee inventors/ designers the right to sell/transfer etc vests with the organization, the employee sharing only the sale value and royalties. As per such agreements and also as per code of ethics, engineers are to act with integrity and loyalty to the employer organization in IPR matters. Also employee engineers will also have access to such vital data and information if they are involved in the application of such inventions/designs/technologies. Engineers should not reveal such data to others, except with the permission of the IPR holding organization.

11. INTERNATIONAL CONTEXT

Different countries and communities may have different values and practices with respect to common morality. This may lead to conflicts in adhering to engineer’s professional code of conduct. Some countries pay less to female employees compared to male employees, some discriminate women from men for employment and for holding superior positions. In some countries business relations are built upon and nurtured through personal relations involving social visits, get- togethers, and exchange of gifts. Corruption is prevalent in some countries. Under such circumstances, engineers are sometime confused or misled on their responsible actions in accordance with their professional code of ethics. Some recommend questionable compromises on the basis of economic conditions, religious sentiments and cultural traditions and prevailing levels of corruption which need to be satisfied to get things done. While there is a point for consideration in striking a balance between the extremes, all such compromises cannot be justified, and may be unethical. There can be exceptions to the rule, but those exceptions must be morally justifiable. More and more engineers are working in countries other than their home countries. They are engaged in design, manufacturing, construction, marketing and in other services and management. There is an increasing need for evolving a common code of ethics for engineers which is accepted worldwide as standards of quality such as ISO 9000.

12. ECONOMIC CONDITIONS

Sometime lower economic development of a country is used to justify applications of lower standards for safety, health, and environment compared to economically advanced countries. This prescription is used by some unscrupulous entrepreneurs/managers and engineers from the advanced countries, to reduce their expenditure and enhance profits and it is clearly unethical A responsible engineer from an industrially and economically developed country while engaged in a project in a less developed country must aim for the same high standards for safety and environment as in his own country. But if it is an informed consent by the Government and people of the developing country, restrained temporarily due to economic constraints, some flexibility in standards could be accepted. For example automobiles produced/used in many Indian cities now comply to emission standards which are lower than specified in many other countries. But there is a plan to come to world standards of Euro 3 and Euro 4 eventually. While automobile manufacturers from other countries, who are setting up manufacturing facilities in India, may use this lower standard it is nobler for these companies to straight away adopt the higher standards just as they comply with, in their home countries. Similar relaxations may be made with respect to an irrigation project or a hydroelectric project, a fertilizer plant etc on a cost-benefit/utilitarian approach. But it must be a conscious decision of the Government and people of the concerned country, with a plan for revising the standard at a future date.

13. CULTURAL VALUES, TRADITIONS, AND PRACTICES

Engineers working in foreign country with different cultural values, traditions, and practices may have difficulties in deciding on ethical issues and responsible professional conduct as they would in their own country. Giving and accepting gifts and building business relations based on personal relations is a tradition and practice in some countries. Even, modern management education emphasizes the need for networking with business associates for mutually rewarding strategic relations. However the engineers ‘code of ethics in many countries generally prohibit giving and accepting gifts with respect to one’s vendors and customers and consider it a mild form of bribe. Similarly consuming alcohol is strictly prohibited at all times in some countries. The engineer’s code of ethics does not permit working under intoxication, and consuming alcohol at work place. But in some countries consuming alcohol during long business lunch sessions or while entertaining or being entertained by business associate is a practice. In such cases, the engineer must try and adhere to the engineer’s code of ethics, and respect the values and practice of the country in which he works, as long as it does not adversely affect others. However this should not be taken to the extreme and justify giving/accepting bribes and justifying that it is a ‘fee’ and is practiced in the country. It is unethical and is a crime against the people of that country.

14. ENGINEER’S RIGHTS

Engineer’s professional responsibility to adhere to professional ethics must be supported by legal and moral rights. Responsibility without authority and legal rights and recognition by all concerned will be difficult to implement. All concerned include the government, society, employers, colleagues and engineers professional associations. Engineer's must be provided not only legal protection, but also financial and social protection against vengeful action by vested interests, Engineers' rights with respect to professional responsibility include:

- Right of refusal to involve directly or indirectly in activity in- violation of professional ethics
- Right of professional judgment and advise/inform all concerned.
- Right of speaking, writing and acting in public interest, in accordance with engineers professional ethics.
- Right to protect client and employer confidentiality obligations, without sacrificing public interest.
- Right to professional recognition and to engage in activities of engineers professional associations
- Right to protect environment and public from harmful effects of technology in general and more specifically from own, employers and client's work.
- Right for legal, financial, and professional protection from threats, coercion, attacks, retribution, loss of job and such other activities by clients, employers and their agents.
- Right for claiming support in respect of the above from public, the state and engineers professional associations.

15. ROLES OF PROFESSIONAL ASSOCIATIONS

Professional societies advance and promote their individual professions, protect the interests of their members and their image in the society and ensure that the members are competent to undertake their professional activities, and perform responsibility following a strict code of ethics. Since there are various disciplines of engineering and there are associations/societies for each such discipline, they should evolve a common code of ethics. Professional societies have a responsibility in developing and promoting the practice of professional ethics. Ethical conduct by engineers and employers can be encouraged through awareness and motivational (commitment) workshops, rewards and punishments. Punishment to enforce ethical conduct has limitations as far as Professional associations are concerned, as the major punishments for ethical violations which can be imposed by the professional associations are only suspension or expulsion from the association's membership. However, membership in a professional association is not mandatory for engineers to be employed. Hence such punitive actions will not have an impact on the employability and professional activities of the suspended/expelled member. Professional Associations of Engineers may have to consider making it a statutory requirement (by law) for engineers to be registered with the respective association as for example a "charted engineer" before they can practice as a professional. This is being done in other professions, such as for lawyers and medical doctors.

Professional associations may institute awards for engineers and employers for exemplary ethical conduct. Professional societies should provide moral, physical and financial support to engineers who are unfairly treated by their unscrupulous managers/employers for adhering to high ethical standards. Professional societies can play a major role in educating the public on the risks and benefits of new technologies and on safe practices, and sustainable development, etc.

Professional societies must also interact with similar societies of other countries and endeavour to develop common code of ethics applicable internationally. More and more engineers work in countries other than their home country and face new working environment with different cultural, religious values, customs, traditions and practices, influencing the local moral standards. Professional societies across the world should endeavor to develop an internationally accepted code of ethics for engineers.

16. CONCLUSIONS

Engineer's professional associations across the world should develop an internationally accepted professional code of ethics for engineers.

It should take into account the increasing concern for the environment and sustainable development.

It should facilitate engineers to be certified as per internationally laid down and recognized standards, enabling their global acceptance for professional practice/employment.

It should cover not only engineers' responsibility towards society and environment but also professional responsibility towards the employer, the share holders, the employees, and customers and compliance to client professional confidentiality agreements, non disclosure agreements, intellectual property rights, license and other contractual agreements in the international context.

Clear guidance should be included with respective engineers' compliance with the code of ethics and at the same time respecting the local cultural and traditional values and customs and the aspirations of the economically and industrially developing countries.

Code of conduct should provide for the certified engineers, legal and financial support against unfair treatment by unscrupulous vested interests for adhering to the prescribed ethical and moral standards.

Final – \\server\e\book work
3-06-10

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

DISCOVER AND DEVELOP CREATIVITY

C.G. KRISHNADAS NAIR*

1. INTRODUCTION

Dictionary meaning of the word "Create" is to produce, to bring into existence something of novelty, special and striking out of one's imagination. Webster's Thesaurus describes the word creative as Imaginative, Ingenious, Original, Resourceful, Clever, Inventive and Innovative etc.

Creativity is fascinating as it makes our work more interesting, our life more enjoyable and meaningful. The excitement of a Scientist or Engineer at the discovery of a novel idea, process or product and that of an artist when he or she creates a new sculpture or painting is difficult to describe, but it gives a feeling of living fully. Man has a creative instinct and it is this, which made him look for novelties, and contributed richly in the development of human civilization right from storage when he created tools out of stone to the modern age of Information Technology, Intelligent Materials, Satellites and Space Travel. In today's complex Society the Scientists, Engineers and Managers have synergetic role and so is the Artists, Craftsmen and Business Managers, in the exploitation of creativity for the benefit of Society and Environment.

2. CREATIVE PROCESS

Creativity results by the combined action of three elements the Domain, the Creative Person and the Field (Figure 1). The domain is the domain of knowledge or culture like Art, Science, and Engineering etc. The domain itself has grown very vast and specialized for eg: if we take Science we have the domains of Chemistry, Physics, Mathematics, Life Sciences etc., within Physics again we have specialization such as Astro Physics, Theoretical Physics, Nuclear Physics and so on.

The field consists of Experts, Society, Commercial Organizations etc, who are to recognize and validate the innovation.

All the three are necessary for a creative idea or product to get implemented. Creative accomplishment is never the result of a sudden flash of insight in the mind of a creative person. It is the result of perseverance and effort on his part and also on the part of the field.

---

*Former Chairman, Hindustan Aeronautics Limited (HAL), President – Society of Indian Aerospace Technologies & Industries (SIATI), Aeronautical Society of India Building, Suranjandas Road, Off Old Madras Road, Bangalore – 560075.

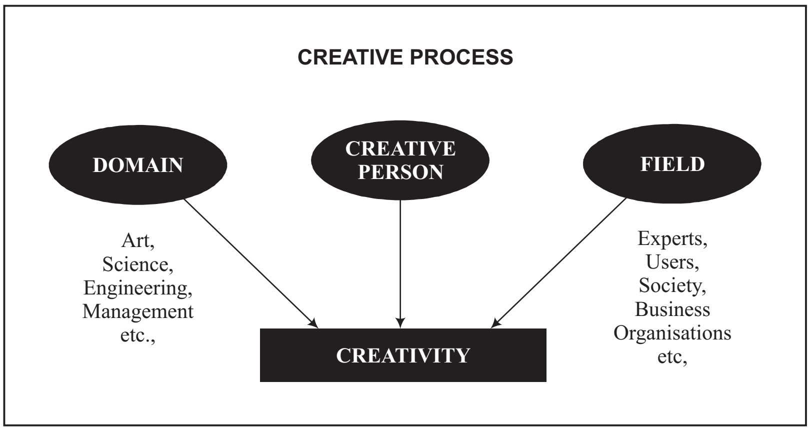


FIG. 1 CREATIVE PROCESS

Thomas Edison's invention of Electricity or Einstein's Discovery of the Theory of Relativity were conceived out of the vast domain knowledge, and intellectual and social network that stimulated their thinking and result of positive interaction by the field, the social mechanisms that recognized and spread their innovation. Such is the story with respect to steam engines, rail road and to give a more contemporary eg., the mobile telephones. The spark is necessary but without air and fuel there will be no fire and flame. Creative idea is like the spark; efforts on the part of the creative person and person's team and the field are the air and tinder/fuel. Creativity results where a person with a novel idea pattern, process, product add value to domain such as Engineering, Science, Music, Business etc. Value addition to the domain need to be accepted by the field such as Experts, Followers, Dealers etc. When VAN GOUGH and PICASSO made novelties in the conservative domain of painting art, this was not immediately accepted. Only when a number of their followers practiced this novel art and eventually, art experts and dealers (which form the field) accepted such paintings as something of a novel contribution to the domain of art, these became modern art paintings. Without the filed acceptance these would have been dubbed as strange meaningless paintings.

The domain knowledge is very important for the creative person to make his contribution. All the examples I sighted earlier, be it Picasso, Thomas Edison or Einstein, these persons had tremendous knowledge in the domain to be able to add value to domain. Knowledge in associated domains can assist in creativity. An example is sir M.Vishveshwaraiah's creation of the Brindavan gardens when he built the Civil Engineering marvel namely Krishna Raja Sagar Dam.

3. DEVELOPING CREATIVITY

Creativity is adding value to a Domain, and so domain knowledge is to be developed, if one has to develop creativity and put it into action. Domain knowledge is one of the three elements; others being the creative person and the 'field' as we discussed earlier. One should develop knowledge in the particular domain of interest, be it physics, mathematics, chemistry, life science, engineering or management etc. Unless we know the domain we cannot add value to the domain.

Develop Creativity

- Develop spirit of innovation
- Acquire domain knowledge
- Develop field (social contact), with passionate commitment for service to society and environment

One does not have to be a genius or brilliant or even talented to be creative. Of course being any or all of these will greatly enhance creativity. But the fact is that an ordinary person can also be creative. The creative instinct is inbuilt. Along with the creative instinct there is also conservative instinct built into our genes for self-protection. The child is born with a lot of curiosity. But as the child grows the curiosity and creative instinct becomes restricted and conditioned by environment and society. As we grow up we get more inhibited. No longer we are able to wonder as a child. No longer we have the same curiosity as a child. So we should undo the limits or restriction put on our imagination and inquisitiveness and liberates the creative Psychic energy.

To be Creative

- One need not be **GENIUS**
- One need not be **BRILLIANT**
- One need not be **TALENTED**
- An average person can be creative
- Being Creative is everyone's business as it makes one's work and life more enjoyable and meaningful

Creative individuals are remarkably adaptive. They have uncanny ability to adapt to circumstances and do with what ever they have to reach their goals, mobilizing additional resources as they go on and on. Preservance and hard work, ability to put in long hours as they enjoy what they do and committed to their goals. They are indeed passionately committed to their work but objective. Because of passionate commitment to work they never loose interest even when the task becomes difficult. Pain is overcome by pleasure. Being objective gives focus and realism.

Traits of a Creative Person

- Remarkable ability to adapt
- Passionate commitment for service to society
- Curiosity to observe and discover new ideas
- Dreamers with vision and mission
- Willingness to work hard for long hours
- Imaginative to the point of fantasy
- Firm conviction in realization
- Going beyond what is considered real and creating new reality

- Simultaneously humble and proud
- Love what they do
- Adventurous in thought and action
- Passionate in their work, but objective
- Diverse interest
- Humanitarian, Compassionate
- Positive response to negative inputs

Creative individuals have a high level of curiosity and observe novelty and get novel ideas. They are excellent in observation of nature's wonders. They are imaginative to the point of fantasy but with firm conviction in reality; they go beyond what is considered real and create new reality.

Creative individuals are adventurists in thought and action and are dreamers with a vision and mission. They are passionately committed to serve Society. They are doing what they do for the joy of accomplishment not for fame not for money. Creative people are simultaneously humble and proud. While they rejoice the product of their creativity and accomplishment they do not take credit and become proud. They are aware of the long line of previous contributors in the domain and aware that "They stand on the shoulders of Giants" and see their own contribution in perspective. Creative persons have often Interest and knowledge in diverse domains such as Scientists having interest in Music, Art, Social Service and Engineers etc. Creative persons have less of negative feelings and more of positive feelings. Jamshed Tata is quoted, have said "Never react to an insult, give a reply through positive and creative action." It is said that once he was not allowed in a posh western hotel. He did not take it as insult. He created the Taj at Bombay, the forerunner of all Magnificent Hotels in India. Mahatma Gandhi contributed to the world one of the greatest creative philosophy "Ahimsa" again as a positive response to insult and intimidation he, and many suffered.

4. STEPS FOR ENHANCING CREATIVITY

Every person has inherent energy and potential to be creative. But there are obstacles to lead a creative life. We are drawn by forces of mundane demands of life and career and distracted by these demands we often have less and less time to tap the inner psychic energy to be creative. In order to liberate the creative energy and put into action, we must let go of our obsession with mundane demands and explore the world around with passion for service and creative living.

Cultivation of curiosity and interest is vital to develop the inherent creative energy. Just like the child, which explores with interest everything insight, we must look at things around. We must develop a habit of observing nature with a sense of wonder. We need to develop curiosity and ability to feel wonder and delight at novelties. As we acquire more and more knowledge in particular domain, say science for example, we must remember that there is much more to learn and wonder. A physics scholar who have learnt and acquired considerable knowledge in physics may choose to wonder at sub-atomic structure and activities or galaxies and inter stellar space. An aeronautical engineer may observe the birds flying with constantly changing shape of wings and wonder and perhaps design and develop mission adaptive wings. There is so much to learn from the innovations of creators by just looking and observing the nature and creatures of creation.

Develop the habit of looking at things even if you have seen them many times before. Every time you look you may find some thing new. We must get rid of 'know all' feeling and be humble to learn more and observe novelties. Look for new patterns, ideas, and experience what you have not seen and thought before. There will always be some thing that you have not observed before and be surprised.

When some spark of idea, pattern design flash through your mind conscious or unconscious (as in a dream) pursue it with determination. Write it down; reflect on it, many times. If you do not do this you will soon forget it as we get too busy in our routines.

Before you go to sleep each night you put your mind in website on a search mode. You think of an unsolved problem in art science, engineer or management. Think of a puzzle in design or a modification for making a design etc. Play it out briefly in your mind. When you are at sleep your unconscious mind will start a search, very much like a search on the website and it may even link up most likely with other minds and the divine forces of creativity and you may get a flash of an answer. You should catch this spark and pursue, develop and act.

Develop a complex personality. By this I mean develop multi interest but be integrated. There is difference between a complex system and a complicated system. Complex system is a differentiated system but totally integrated. The system that is differentiated but not integrated is complicated and not complex. Living organisms, as per Darwin's theory of evolution become more and more complex as the evolution progresses. Likewise the creative mind in addition to being specialized in depth and in breath in any particular domain must also develop interest in others and cultivate relationship with other minds with knowledge in other domains and also with the society. Thus a scientist or an engineer may develop interest in music, art and develop friends from other fields and be socially active. It is the net working of these minds and net working with other people society, that will enable the 'field' to give support to creativity.

Steps for Enhancing Creativity

- Liberate the creative psychic energy
- Cultivate curiosity and interest
- Observe nature, wonder and learn
- Observe novelty in the things you see and be humble and surprised
- Follow the spark of an idea, pattern, design, flashing through your mind
- Look forward to each day and the joy of accomplishment practice creativity in daily life
- Set your minds energy website on a search machine every night
- Discover what you like and develop domain knowledge
- Develop interest and knowledge in other domains
- Develop social contacts and network
- Develop parallel thinking and analysis
- Chitha Shudhi exercise to develop positive feelings

Develop the capability of parallel thinking and analysis. Consider several solutions to a problem from many angles. This will liberate the mind from thinking in one track way. Micro innovations at shop floor and suggestion schemes etc are simple applications of this procedure. Look forward to each day, at the dawn with curiosity and interest and the joy of new encounters and the joy of doing the tasks of the day and new opportunities.

5. CONCLUSIONS

Engineers must discover and develop the inherent creativity in them. Creativity can be enhanced by various means such as observing nature, cultivating curiosity, parallel and out of box thinking, acquiring domain knowledge in multiple fields, interest in fine arts and passionate commitment for service to society etc. by being creative engineers will become more innovative and useful to society, work will become more enjoyable, satisfactory and successful.

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

VALUE ENGINEERING

C.G. KRISHNADAS NAIR*

1. DEFINATION

Value Engineering is the application of Value Analysis in the manufacturing of a product or in providing an engineering service. Value Analysis is the systematic analysis of all factors involved in providing the services or in the manufacture of a product with the aim of overall cost reduction without affecting the quality or performance of the product or service. It is not just cost cutting, as it is imperative that whatever changes are made in the design or material or processes of manufacture and inspection for cost reduction, those changes should not alter the quality performance and reliability of product/services.

2. PROCEDURE

A step-by-step approach to the procedure involved in Value Engineering is give below:

- Understanding the product/service, its specifications, quality requirements, utility and reliability in service.
- Collection of data pertaining to design, materials, manufacturing/ inspection technologies, processes and equipments, packaging, delivery, etc.
- Decision of maintaining, improving or revising the quality requirements and specifications as per customer's feed back.
- Step-by-step analysis of the existing design, materials, manufacturing and inspection methods etc., with a view to conserve materials, energy, labour etc., to improve productivity and reduce costs with an analysis on the impact of such changes on the quality of product/service.
- Implementation of changes, which do not affect the quality and performance.
- Testing results through field trials and customer feed back.
- Implementation in production/service.

*Former Chairman, Hindustan Aeronautics Limited (HAL), President – Society of Indian Aerospace Technologies & Industries (SIATI), Aeronautical Society of India Building, Suranjandas Road, Off Old Madras Road, Bangalore – 560075.

Basic requirements in Value Engineering are applications of common sense, logical and lateral thinking, willingness to experiment and innovate, domain knowledge, imagination and creativity. Persons involved in Value Engineering must have a questioning mind as they must seek and find answers to the questions such as: Can the designs be changed? Can a cheaper alternate material be used? Can a different method of fabrication be used, reducing the time for manufacture and reducing material, wastage and energy consumption etc., and all these without any compromise on the performance, reliability and maintainability?

3. CONSTRAINTS

Few product designers and production engineers readily accept changes in their design/processes, once they have been proved to function adequately. In their view thus they have been optimized. Sometimes there will also be resistance from quality assurance authorities as any change in process or material or design will require very detailed quality audit, as some of the changes especially in manufacture/inspection technology and equipments may require additional investment and there could be questions on the economic viability. Thus changes are always looked upon with suspicion from each department when looked in isolation.

4. THE TEAM

Value Engineering, become more practical and achieve best results when it is taken up as a team effort. The team must necessarily question the existing practices with a view to invite and improve to reach the organizational goal of greater excellence in providing quality product/service at lower and lower cost. Composition of a typical team in an engineering environment will be a multi-disciplinary group drawn from various departments such as Design Engineering, Production Engineering, Manufacturing shop, Marketing, Commerce and Finance, Laboratory and Quality Control.

5. CASE STUDIES

Case – 1: Fin Forging – Saving in Material

Foundry & Forge Division, Hindustan Aeronautics Limited, Bangalore Complex had undertaken the manufacture of an Aluminum-Copper alloy close tolerance Fin forgings required for anti-tank missiles (See Figure 1) Approximate annual requirement was one lakh pieces.

The forgings were being produced from extruded bar stock of section 41mm × 13mm and 230mm long. The forging stock weighed 327gms whereas the forging weighed only 80gms, resulting in heavy loss of material as forging flash. A study of the geometry of the fin forging showed that the section drastically reduced from the fixed end to the other end. Hence, the same bar stock split into two could easily be forged to the shape required. Further, it was possible to reduce the thickness from 13mm to 10mm. Additional material saving was also proposed by suggesting to utilize a profiled extrusion. The various improvements and effects are shown in Table 1.

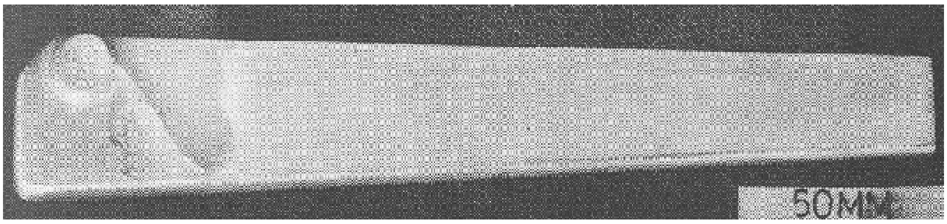
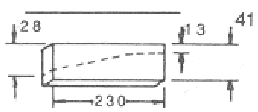
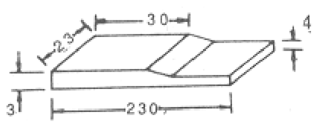


FIG. 1 FIN FORGING

Table 1. *Materials Economy in Fin Forging*

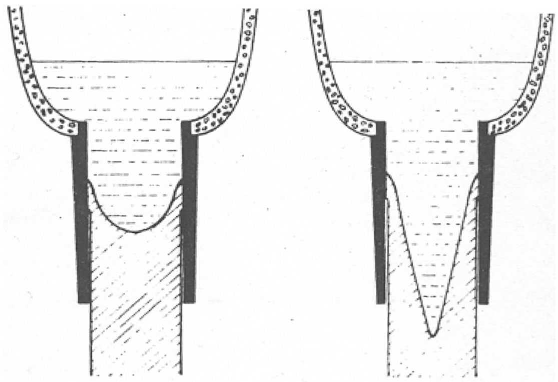
Method	*Forging stock sizing in mm*	*Stock weight in Gms.*	*Material saving per piece in Gms.*	*Cumulative saving per annum (One lakh Pcs)*	
				Material in Tons.	*Value Rs. in lakhs*
Original	$41 \times 12 \times 230$	328	–	–	–
Phase-1 improvement		164	164	16.4	6.2
Phase-2 improvement	$\frac{41 \times 13 \times 230}{2}$	134	30	19.4	7.37
Phase-3 improvement		100	34	22.8	8.66

Even though a few gms. of metal only were saved at each phase of improvement in process, the impact on mass production was enormous. Value of cumulative savings has been worked out taking Rs.48/kg for the alloy, and a scrap recovery value of Rs.10/kg.

Case – 2: Semi-continuous Aluminum Alloy Billet – Saving in Material and Labour

Foundry & Forge Division had developed Semi-continuous cast billets in high strength Al-Cu alloy for supply to a sub-contractor for manufacture of profiled extrusions for aircraft structural applications. The approximate requirement of these billets had been 30 tonnes/annum. These billets were required to be of 150mm dia to suit the particular extrusion press. The division was however casting 190mm dia billets and machining the same to 150mm dia, resulting in considerable wastage of labour and material. A study revealed that this method was adopted due to the poor surface quality of the billet, as shown in Figure 2. The circular ripple defect on the surface extended quite deep and at times resulted in puncturing the billet. Under constant conditions of water-cooling of

FIG. 2 AS CAST BILLET SHOWING POOR SURFACE QUALITY, 190 MM DIA



(A) (B)

FIG. 3 GEOMETRY OF SOLIDIFICATION FRONT IN CONTINUOUS BILLET CASTING (A) SLOW WITHDRAWAL RATE (B) FAST WITHDRAWAL RATE

the mould and constant level of liquid metal, the billet withdrawal rate affected the geometry of solidification front as shown in Fig. 3. Careful observations during billet casting revealed that surface defects were related to fluctuations of level of molten metal in the mould and fluctuations of billet withdrawal rate. A 165mm dia mould was made and casting parameters were standardized to obtain billets of satisfactory surface finish (Fig. 4) so that a light machining was adequate prior to extrusion.

This cut down the cost of machining in addition to saving valuable material, and resulted in a total saving of about Rs.1.7 lakhs/annum as given in Table 2.

Case – 3: Centrifugally Cast Cylinder Liner – Saving in Material and Labour

Foundry & Forge Division had developed an alloy cast iron Cylinder Liner casting for the L-60 engine of Vijaynta Tank. It was a high quality casting with 100% radiographic inspection. At the time of this work, forty liners per month were being produced, and there was a plan to increase the

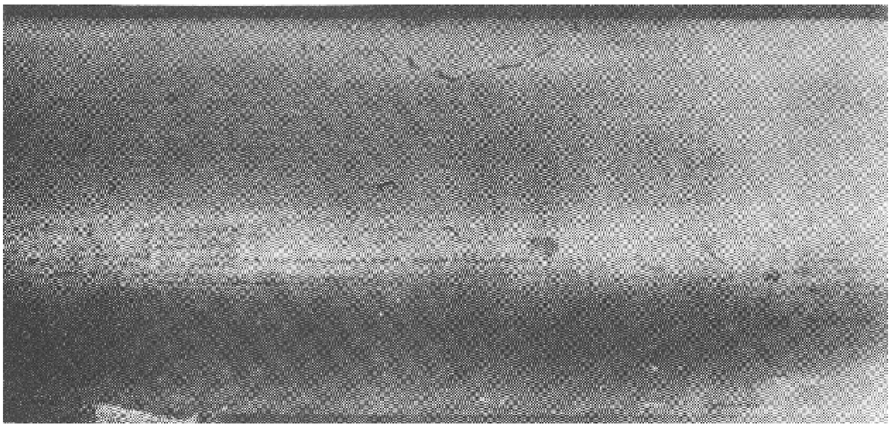


FIG. 4 AS CAST BILLET WITH GOOD SURFACE QUALITY AFTER PROCESS IMPROVEMENT, 165MM DIA

Table 2. *The Case of Semi-continuous Cast Billet*

Saving in Material:	
Saving in Material/billet	25%
Saving in Material/annum	= 25% of 30 tonnes = 7.5 tonnes
At the rate of Rs.30/kg for billet and Rs.10/kg	
Scrap recovery, saving (Rs.)	= (30-10) x 7500 = Rs. 1,50,000
Saving in Machining:	
Cost of Machining	
190mm dia billet to 150mm dia	= Rs. 20/billet (300mm)
Cost of Machining	
165mm dia billet to 150mm dia	= Rs.7/billet
Saving per 30 tonnes (1800 billets)	= Rs.13 x 1800 = Rs.23,400
Total Saving per annum	**= Rs. 1.73,400 say Rs. 1.73 lakhs**

production to 150 per month. The alloy cast iron was melted in induction furnace and centrifugally cast into hollow cylinders. These were then heat-treated, rough machined, X-rayed to check soundness and subsequently finish machined. Fig. 5 shows, as cast, rough machined and fully machined liners.

The as cast part weighed 75kgs, and rough machined part and finish machined part weighed 26kgs and 15kgs respectively. As there were many costly alloying elements, the alloy cast iron costed about Rs.12/kg, and it appeared therefore desirable to reduced the weight of the as cast part and this would also reduce the extent of rough machining. The cost of rough machining was Rs.530/- per liner.

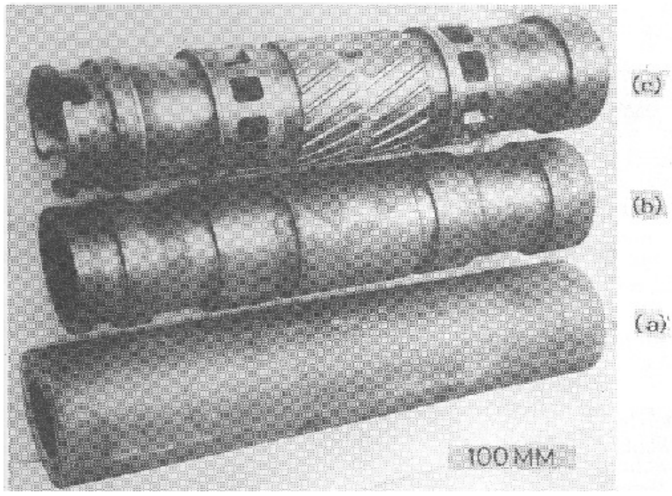


FIG. 5 CYLINDER LINER CASTING (a) AS CAST (b) ROUGH MACHINED, AND (c) FULLY MACHINED

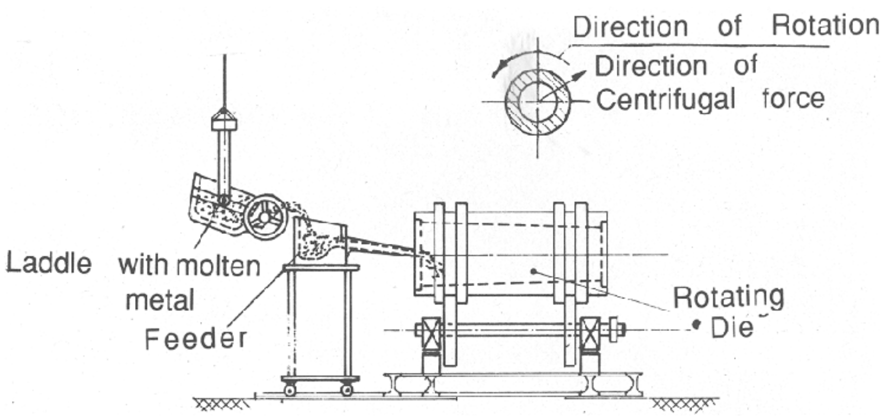


FIG. 6 SCHEMATIC VIEW OFF THE CENTRIFUGAL CASTING

A schematic of centrifugal casting is shown in Figure 6. Molten alloy is poured into a rotating mould, and the metal is distributed to the circumference of the cylindrical mould radically by the centrifugal force. The centrifugal force however drives the denser alloy to a greater distance (radically) than the less dense non-metallic inclusions and impurities. Thus the impurities get concentrated towards the inside surface of the cast cylinder.

This was why a liberal wall thickness and machining was provided for the cast cylinder. However, with better control over melt quality it would be possible to increase ID and thus reduce the casting weight. This was implemented and a saving of 10kg per casting was realized. This, at the rate of Rs.12/kg resulted in saving of Rs.120 per casting, with a bonus saving of Rs.40 for machining. Later the production rate was increased from 40 to 150 per month and saving on account of this improvement was Rs.2.88 lakhs/annum, as illustrated in Table 3.

Case – 4: Valve Block Forging – A Method Change

A valve block in HE-15 (Aluminum-Copper) alloy being manufactured and supplied by Foundry and Forge Division for a defence project was made traditionally from an under sized extrusion with the process involving a number of upsetting and drawing in a two stage forging followed by milling to shape. Since the forging was a simple rectangular section (108mm x 95mm) it was proposed that it may be converted into a near-net shape extrusion, which may be cut to the correct length, thus saving considerable amount of labour. However, the supplier of extrusion declined to guarantee the properties for such a large section of extrusion. His press was capable of extruding only 230mm dia cast billet and for an extrusion of 108mm x 95mm, this allowed a reduction ratio of only 4, which in his opinion was probably too low to breakdown the cast structure and improve properties in this alloy. By a series of carefully planned experiments, using a lab-scale extrusion press, the microstructure and properties of as extruded alloy at different extrusion ratios were studied. From this it was concluded that satisfactory properties can be achieved at the reduction ratio for the particular extrusion in question. The proposal was successfully implemented after obtaining approval from the customer based on performance evaluation. Table-4 illustrates the various operations when the part was made as a forging and the reduction in labour/machine minutes achieved by converting it into an extrusion. This work was done when an order for 600 Nos. was outstanding (to be supplied at the rate of 200 per year). A total saving of Rs.1.57 lakhs was realized from this work.

Subsequent to this, a number of such other valve blocks which were originally forgings were converted to extrusions with similar substantial cost saving.

Table 3. *Cylinder Liner Casting – Cost Saving by Change of Casting Geometry*

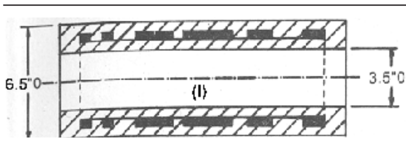
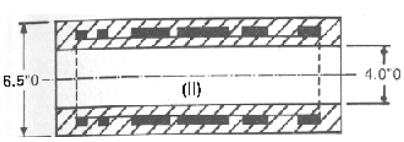
Casting geometry	*Weight (Kg)*	*Saving*	
		Per Casting	*Per annum (1,800 Pcs)*
	75	–	–
	65	Material (10Kg) Rs. 120 Machin ing Rs. 40 Total Rs. 160	Rs. 2.88 lakhs

Table 4. *Comparison of Labour Required for Valve Block Forging/Extrusion*

Operations	*Method-I* Forging to shape (108 × 95 × 495 mm) from under sized extrusion		*Method-II* Extrusion (108 × 95 mm) and cut to length
	Manu Minutes	*Machine Minutes*	*Man Minutes*
Bar stock cutting	9.7	–	4.1
Debut and Stamp	3.6	3.6	2.1
Degrease	18.0	–	18.0
Forge I Operation	69.0	14.0) 3000	–
Forge II Operation	50.0	10.0) Tonne Hyd.	–
Degrease & etch	18.0	Press	–
Test bar forging	3.0		2.3
Mill to size	55.0		–
Vibro identification	1.0		1.0
Degrease & etch after heat treatment	18.0		18.0
Cut Micro specimen	2.0		2.0
Turn tensile specimen	3.0		3.0
Ardrox (NDT)	15.0		15.0
	265.0		65.5

Case – 5: Latch & Safety Plug Body Forging – A Design Change

Foundry & Forge Division, Hindustan Aeronautics Limited, Bangalore Complex was to manufacture the subject forging in a high strength Aluminum alloy, for a defence equipment under design and development. On examination of a part drawing of the component (shown in Fig. 7), it was found to be designed as a die forging, whereas its shape was also conducive to its manufacture as an extrusion.

It was considered that the functional requirements of the component could also be met by an extrusion, and this was confirmed during discussions with the designer. Since the manufacture of the item as an extrusion would result in considerably less expenditure in terms of tool and labour costs, the design was changed to an extrusion, as suggested by us. A comparison of the cost of production by the two methods of manufacture is given in Table 6. Timely feed back from the manufacturer to the designer considerably reduced the cost of the component, resulting in a cheaper product to the satisfaction of the designer and user, without any sacrifice to quality.

Case – 6: Supercharger Bucket blades – Saving in Material and Labour

A major task taken up recently by the Foundry & Forge Division has been the manufacture of precision forged turbine bucket, blades (Figure 7) in nickel base x 750 super alloy (70% Ni, 14% Cu-T-Al Fe-Co-Cb-Ta alloy) for super charger of Diesel Locomotive engines. Firm projected

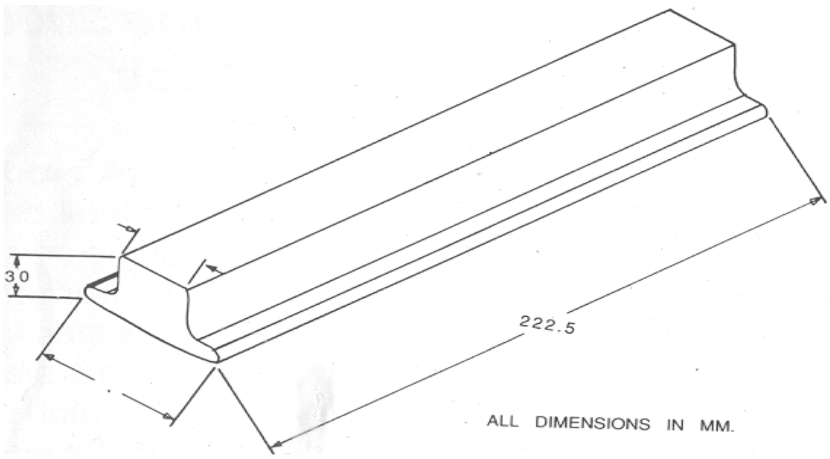


FIG. 7 LATCH AND SAFETY PLUG BODY FORGING (EXTRUSION?)

Table 5. *Comparison of Costs – Latch & Safety Plug Body Forging*

	Forging	*Extrusion*
Tooling Cost	Rs. 40,000/-	Rs. 2,500/-
Production Cost:		
Material cost per piece	Rs. 55/-	Rs.30/-
Labour cost per piece	Rs. 190/-	Rs.44/-
Net Saving per piece		Rs.171/-
Saving for total requirements (5,600 pcs)	–	Rs.9.576 lakhs

requirements were 1000 during development, 6000 in the first year of production and thereafter 12,000 per annum for the next 10 to 12 years. Although the metallurgical process employed was indigenous, the tools/dies and thus the technology was imported. During development/trial production considerable die wear was experienced, resulting in poor die life and excessive cost over run.

The forging was being produced from a rolled stock of 29mm diameter and 45mm long. The forging stock weighed 220 gms, whereas the weight of the blade forging was only 120 gms, thus giving an yield of 55%. Nickel based alloys are very tough to forge and excessive material necessitated higher deformation loads, causing over loading of dies resulting in the premature die failure by wear/breakage.

The sequence of forging operation is illustrated in Figure 8. By systematic analysis of metal flow during extrusion and forging, it was found possible to re-design the sequence and use a smaller stock size (25mm dia and 40mm length). The next set of dies was made to suit the new method and the process is now fully implemented, as properties of the blades were found satisfactory.

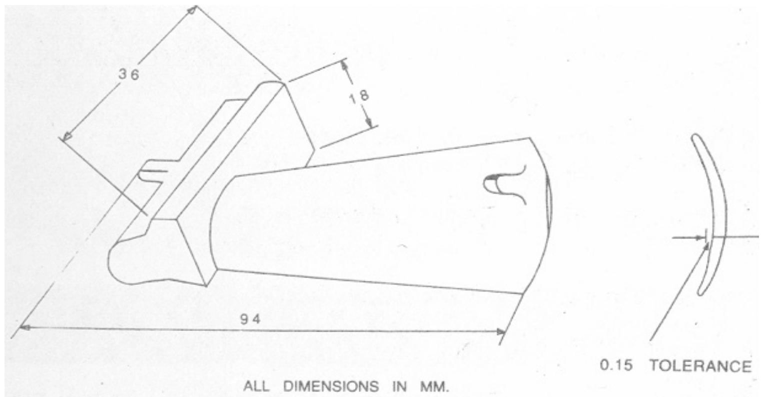


FIG. 8 BUCKET BLADE FORGING

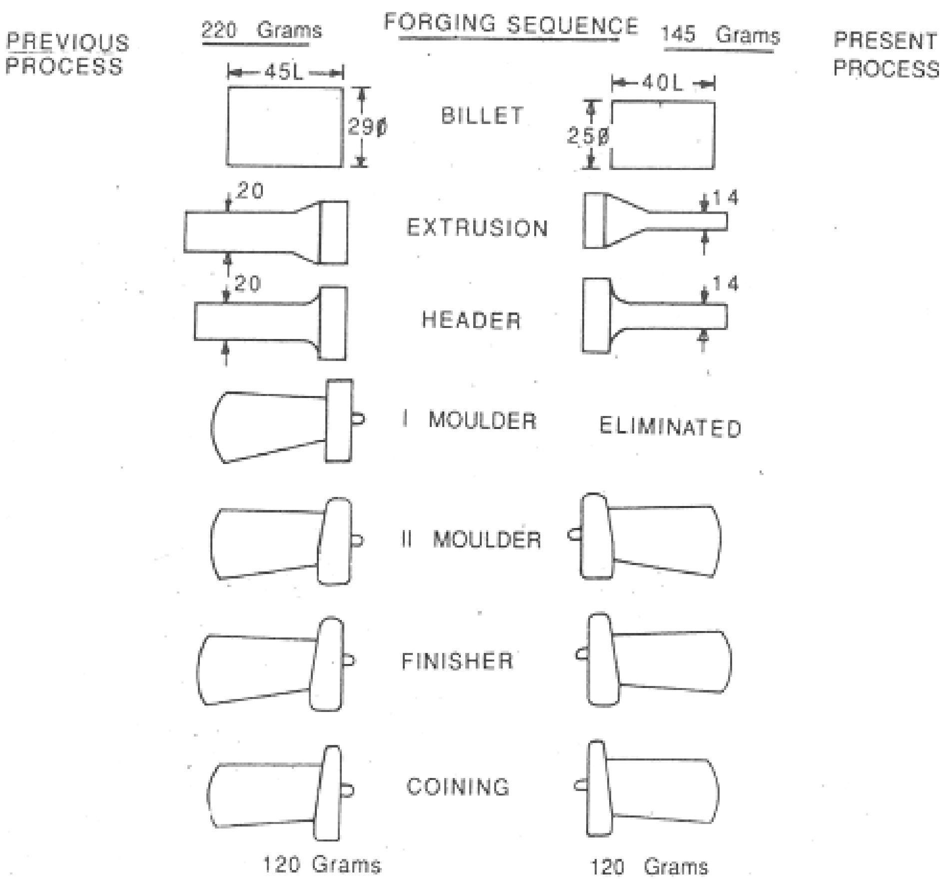


FIG. 9 SEQUENCE OF FORGING OPERATIONS

Table 6. *Die Life (In Terms of Average No. of Pieces Forged)*

Process	*Extrusion*	*Header*	*Moulder*	*Finishing*	*Coining*
Previous	100	1500	500	275	500
Improved	200	2000	800	275	800

Table 7. *Cumulative Saving*

		Rs. P
1.	Saving due to less Input materials per piece (220g – 145g):	
	75g @ Rs.535 per kg.	40. 13
2.	Saving in manufacturing process 65 min @ Rs.70 per manhour.	75. 83
3.	Saving due to increased die life per piece	196. 00
	Say Rs.312	311. 96
	Cumulative saving at the rate of 12,000 pieces/annum	

Considerable saving in material and labour hours and improved die life and consequent saving were achieved as shown in Table 6 and Table 7.

Case – 7: Machining of Inlet Flange and Body for a Liquid Propellant Engine - - Improvement in Method of Manufacture/product design

Three components viz., Inlet Flange 'O', Inlet Flange 'F' and Body required for a liquid propellant engine powering a missile (as per the earlier design and method) were planned to be produced by machining out of bar stock material. Keeping in view the anticipated programme of 810 engines, it was felt that the present method involving extensive machining merited consideration for change. Even certain features of the components were complicated in nature and required high skill and time to machine as per the drawing. It was therefore proposed to produce these components out of forgings which will considerably reduce the machining time and facilitate easy manufacture. The feasibility for forging and machining was since established and product/method redesigned to machine from shaped forging. The old and new method are illustrated in Figure 10 for one component (body) and a comparison of costs and estimated saving is given in Table 8.

Case – 8: Air Seal Front – Reduction in Material Consumption and reduced Wastage due to Change in Tool Design

The subject component for a Jet engine is made by machining from a ring forging in chromium alloy steel to specification MSRR 6503. The ring forging and the machined component are shown in Figure 11. Five components were being machined out of one forging. The balance end piece which was just short in thickness required for one more component used to be discarded as waste.

It was observed that considerable material was wasted during machining due to the excessive thickness of the 'part off' tool. With change in design of the 'Part –off' tool, it was possible to make components out of the same forging as illustrated in Figure 12. The forgings are presently being

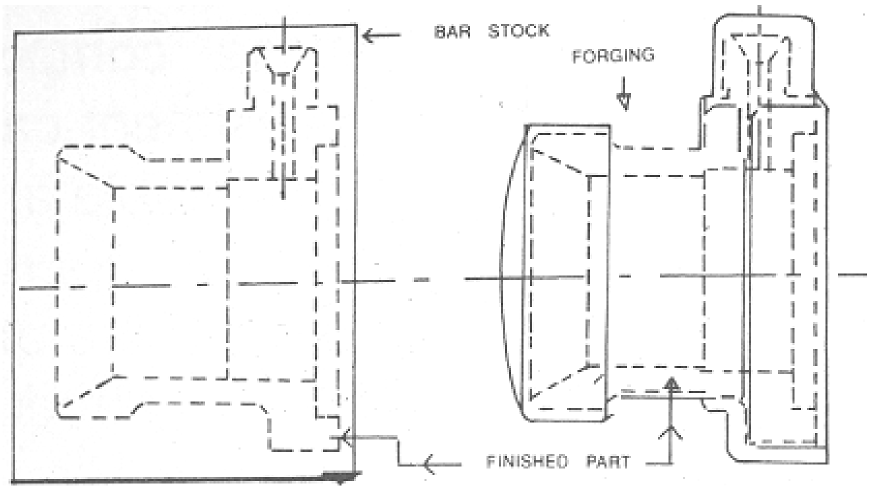


FIG. 10 MACHINING OF INLET FLANGE 'O' FROM BAR STOCK AS COMPARED TO FROM A SHAPED FORGING, SAVING IN MATERIAL AND MACHINING TIME.

Table 8. *Cost Comparison and Saving*

Component	*Old Method*	*New Method*	*Saving*
1. Inlet Flange 'F' (Qty. per engine-1)			
Material Cost (AIS!304/04Cr 18 Nil0) IS : 6603	Bar Stock dia 90mm × 75mm long- Rs.800/-	Forging Rs.1,695/-	— Rs. 895/-
Machining Cost @ Rs.130/hr.	20 hrs. Rs.2,600	7 hrs. Rs.910/-	+ Rs.1,690/-
Saving per part			**Rs. 795/-**
2. Inlet Flange 'O' (Qty. per engine – 1)	Saving in Cost: same as above		Rs.795/-
3. Body 'F' (Qty. per engine – 1) Material cost (Al Alloy IS:733 65032WP condition)	Bar stock dia. 120mm × 150 mm long-Rs.120/-	Forging Rs.1,226/-	—Rs.1,106/-
Labour @ Rs.130/hr	21 hrs. Rs.2,730/-	8 hrs. Rs.1,040	+Rs.1,690/-
Saving per part in cost			Rs. 584/-
Saving per engine set = (Rs. 795 = 795 = 584) =			Rs. 2,174/-

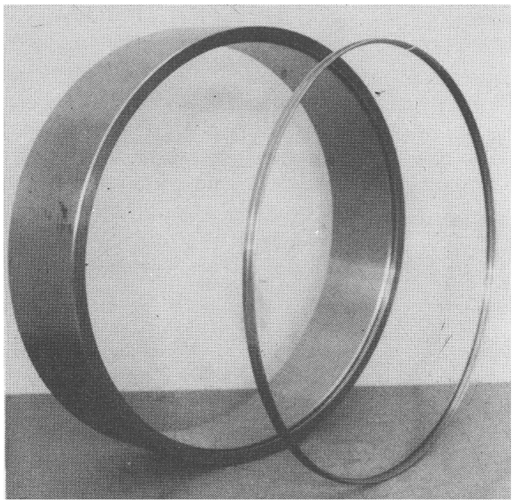


FIG. 11 RING FORGING AND MACHINED COMPONENT (AIR SEAL FRONT)

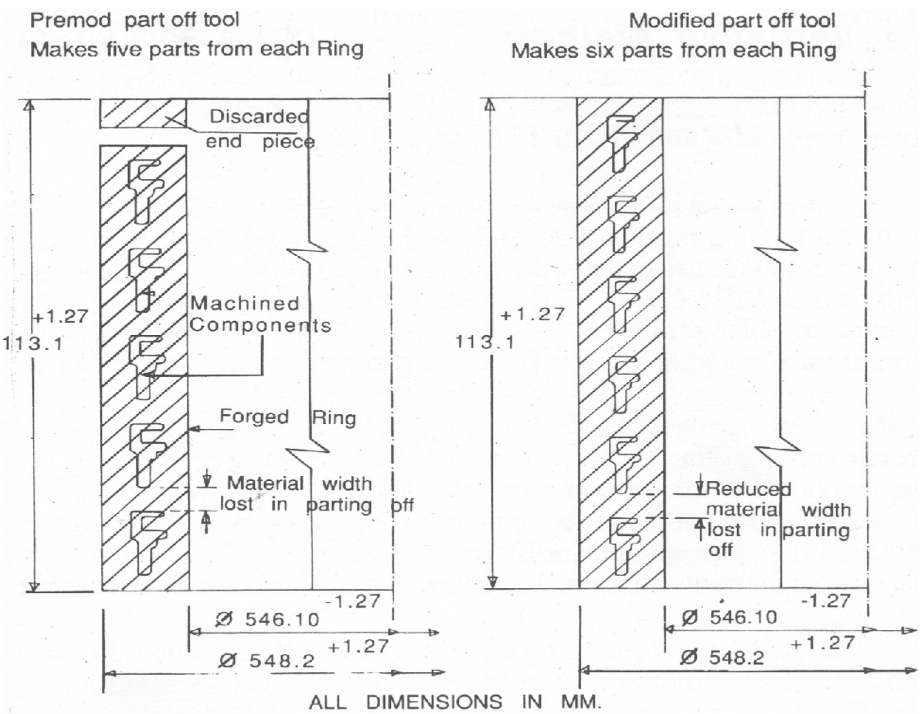


FIG. 12 SKETCH SHOWING POSITION AND NUMBER OF 'AIR SEAL FRONT' COMPONENT MACHINED FROM FORGED RING

imported and cost of each forging is 584 pound sterlings (Rs.13,265/-) at 1987 price level. The changed method reduced consumption of forging by one number per annum to meet the production requirement of 24 components per annum. The saving thus works out to approximately to Rs.13,265/- per annum in Foreign Exchange.

Final – \\server\e\book work
3-06-10

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

DEFENCE ELECTRONICS SYSTEMS

S. MADIVAANAN*

It is decided to give an overview on "Defence Electronic Systems" in the beginning and in subsequent lectures each sub topic will be dealt in grater detail .The systems covered under this subject are as given below

VOL. I

1. INTRODUCTION
2. SENSORS
3. COMMUNICATION SYSTEMS

VOL. II

1 C^4I^2 SYSTEMS
2 WEAPON PLATFORMS
3 FUTURE WEAPONS

DEFENCE ELECTRONICS SYSTEMS

1. INTRODUCTION

It was in 30 April 1897, Sir Joseph John Thompson discovered the existence of electron. This discovery no doubt stands as the most singular one in the history of modern science. Today the electronics engineering field has grown to such a level, even Sir JJ Thompson would have not dreamt.

It encompasses the fields of Electronic devises, Computers, Microprocessors, Communication Systems, Electro optical and Acoustic Systems, Guidance and Control Systems, Satellite & Remote sensing Systems. These Electronic systems are required to Sense, transmit, compile, analyse and process information and then launch, guide, control and trigger the various MUNITION SYSTEMS. The most recent wars have clearly demonstrated the key to a successful campaign in the warfare are high tech system like RADAR, LASER, SONARS, and Precision guided weapons and like sophisticated sensor and weapon carrying platforms all of which fundamentally depend on

*Scientist 'F', Head (ATL&N Division), CVRDE, AVADI, Chennai-54

Electronics Systems. In the gulf war in 1991 the most lethal 7 most expensive weapon was the Electronic Information System developed by multinational troops led by United states of America. It is a fact that most of these technologies would not have materialized if Sir JJ Thompson had not discovered 'ELECTRON' a Century ago.

Major Electronic system used in defence generally fall in to three major categories

- Sensor Systems
- Communication Systems and
- Weapon System

2. SENSORS

A Sensor is a device that receives signals or stimulus and responds with an electrical signal while a transducer is a converter of one type of energy into another. In practice however the terms are used interchangeably. Sensor systems include active and passive ones that sense the presence of an object and determine some of the characteristics even in adverse weather conditions where normal human sensors are inefficient or can't be used directly for operational reasons. Active transducers are those that require a current or voltage excitation. Passive transducers are those that requires an excitation. Active sensors illuminate a target with their own energy in infrared, microwave or optical frequency bands, whereas the passive systems operate on the radiation emanating from the target itself. Both types of systems are essential in today's warfare.

Table 1

Active	*Passive*
1 Radar	1 EW sensor
2 Laser designator	2 Laser warning systems
3 IR imager	3 Thermal imager
4 Sonar	4 IR seeker
	5 Thermo couple,

2.1 Radar And Its Defence Applications

RADAR stands for Radio Detection and Ranging. It is one of the most common sensor System employed in modern warfare and it operates in microwave to milli meter wave with frequency range 0.5 GHz to 100 GHz. It is a critical element of many Surveillance Systems that prevails in to days Military and Civilian Environments . Its role in these environments includes target detection and identification, navigation and mapping, target tracking and weapon guidance. Modern RADAR is capable of extracting accurate parametric information about its target including range, bearing velocity and its size.

A Radar is essentially a ranging or distance measuring device .It consists fundamentally of a transmitter, a receiver an antenna and an electronics system to process and record the data. The transmitter generates successive short bursts (or a pulse of microwave (A) a regular intervals which are focused by the antenna in to beam (B).The Radar beam illuminates the surface obliquely at a right angles to the motion platform. The antenna receives a portion of the transmitted energy

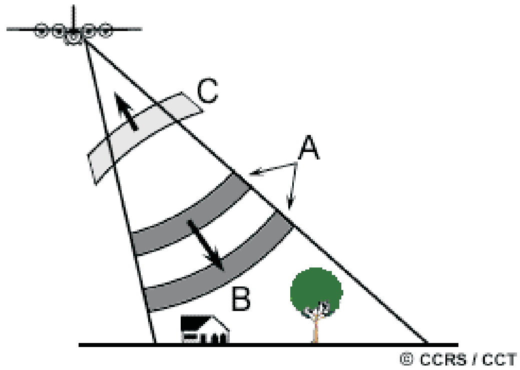


FIG 1. PRINCIPLE OF OPERATION OF RADAR

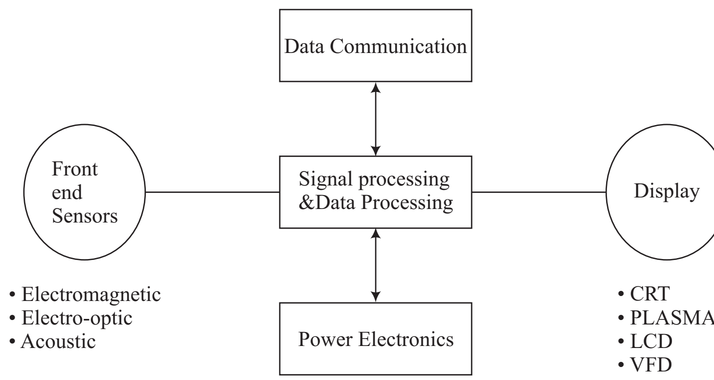


FIG. 2 TECHNOLOGIES OF RADAR

reflected (or backscattered) from various objects within the illuminated beam(C).By measuring the time delay between the transmission of pulse and the reception of the backscattered "echo" from different targets, their distance from the radar and thus their location can be determined. As the sensor platform moves forward, recording and processing of the backscattered signals builds up a two-dimensional image of the surface.

2.2 Sonar

SONAR, abbreviation of Sound Navigation And Ranging, is used by ships and Submarines generally to detect, localize(range, depth and bearing) and classify under water objects by acoustics means. These Systems operate on the principle of acoustic wave propagation in water. Unlike a microwave RADAR a SONAR can have both active and passive modes of operation. An active SONAR, similar to RADAR transmit a modulated pulse and looks for the echo from the target. An Analysis of this echo provides important information regarding a target like its location, Doppler and other Signature. A typical SONAR can do interception ranging, Passive Surveillance, underwater telephone, Fire control and launching weapons.

2.3 Electro Optics Sensors

As in the basics of optics, Electro optics is concerned with the generation, control of propagation and detection of electromagnetic waves. The main difference is that Electro-optics specializes in components devices and systems which operate by modification of the optical properties of a material by an Electric field. Typically they are in the Infra red region of electro magnetic spectrum. The basic issues are:

How is Electro optics used in the generation of light?

How does it control the propagation of radiation?

How is Electro Optics used for the detection of light

Electro optics is concerned with the generation of electro magnetic waves by means of Electrical fields. Certainly, radio and television waves and micro waves are created electronically as Laser. But in the case of Laser wavelength Electro optics is not concerned with then creation radiation as it is concerned with the propagation and detection.

2.4 Controlling Propagation

There are electronic devices that can control the propagation of non-visible light, but they are not the mainstay of Electro-optics usually Exotic material are used to direct and focus the various non-visible wavelengths.

Detection is the heart of Electro optics. Photo detectors respond to Electro magnetic radiations and electrically record its characteristics.

Electronic devices such as focal plane arrays and charge coupled devices (CCD) are often used to detect and directly change the light into an electrical signal.

2.5 Laser

Laser is an acronym for Light Amplification by Stimulated Emission of Radiation. The word Radiation in the acronym does not refer to ionizing radiation as produced by X-ray but the Thermal Radiation. Thermal radiation is non ionizing and does not require monitoring devices. Basically, laser light is focused beam of intense, single coloured light that does not disperse and is uniformly absorbed with minimal loss of energy.

Types of lasers

Lasers are characterised by the type of material used to produce Laser light also called the lasing material. For example CO2 laser produces laser light by exciting atoms of carbon di oxide gas. Er. YAG laser produce energy by exiting Erbium(Er) atoms in a crystal of Yitrium Aluminium and Garnet abbreviated as YAG. Lasing materials can be crystals, of a semiconductor(diode) materials each determining a unique wavelength or color of light emitted from the laser. The wavelength of each lasing material is unique and is described in terms of nanometers(nm) or one billionth of a meter from the visible(400 nm – 700 nm) to the invisible infrared(700 – 10,000 nm).

2.6 Laser Target Designator

LTD operates at 24V dc providing 5MV Laser pulse at 1064 nm with a nominal repetition rate of 20 Hz and can designate main battle tanks from a distance of more than 5Km.

2.7 Laser Range Finderrange Finder (LRF)

The principle used in most of LRF is pulse echo or time of flight measurement. Most of LRFs are designed for ranging non co-operative targets. Pulsed LRFs enable target from several tens of kilometers to several tens of kilometers. These work on the echo principle very much like RADARS in which the time taken to receive the reflective pulse from the target is measured and converted to the range since the velocity of light is constant and known. LRFs are highly compact owing to the small emitting required to produce a capability of 10 to 20 Km with an accuracy of + or – 5 m is typical for these military LRFs which may operate at repetition rates ranging from 12 pulses/min to more than 15 pulses/sec depending on the application.

The design of any LRF system involves several critical technologies relating to transmitter design, low level signal detection, signal processing and design of various optical and Electronic Systems.

The Laser transmitter commonly employs xenon lamp pumped solid state LASER operating in the near Infrared such as the neodymium.

3. COMMUNICATION SYSTEMS

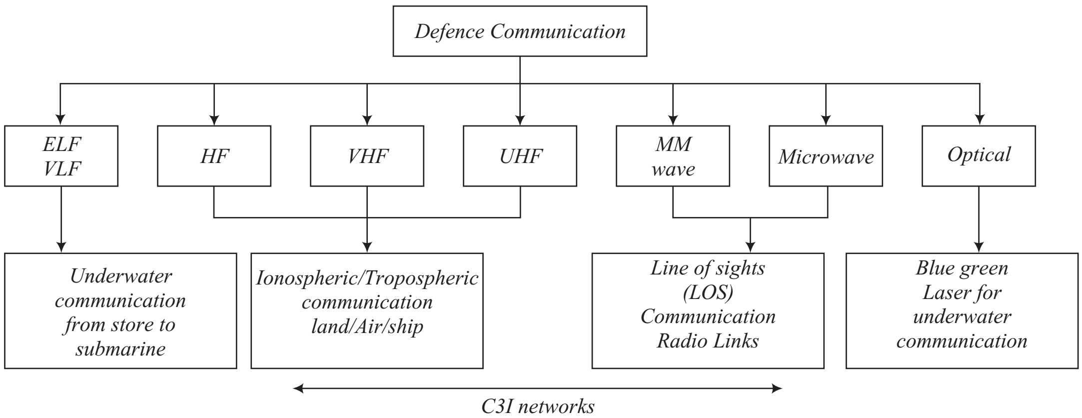
Communication plays a vital role in Defence. Analog to Digital technology, radio wave to microwave carriers and Optical Communication technology have totally revolutionized today's battlefield scenario. The influence of Communication technology in the Gulf war, closely observed and analyzed by every commander and researcher clearly indicates this influence. Future wars will definitely decided by technological supremacy and Communication technology is one of the primary ones.

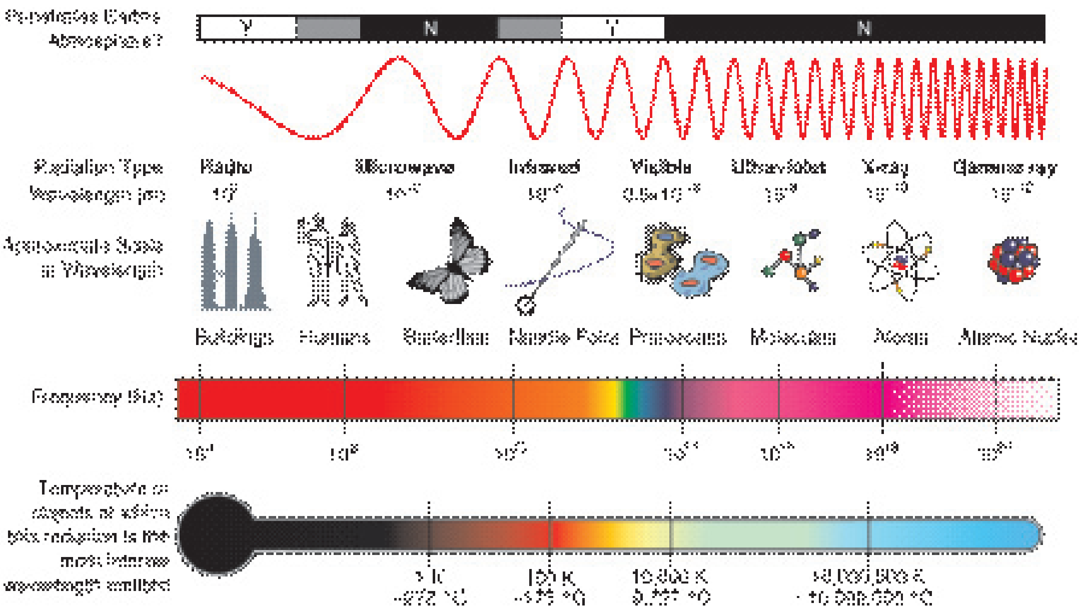
Defense Communication Systems cover almost the entire Electro Magnetic Spectrum ranging from ELF through microwave to Optical Frequency bands. Underwater Communication is rather a challenging task because conventional radio communication techniques are not at all feasible as high frequency electromagnetic fields are highly attenuated by the sea water. Special Communication Systems are designed at very low EM frequency (10 Khz to 30 Khz) for carrying data at the rate of about 100 baud in a broadcast mode. Normal underwater communication systems are centered around acoustic wave probagation in sea water and use varieties of Sonars.

3.1 Electromagnetic Spectrum

Although some radiations are marked as *N* for no in the diagram, some waves do in fact penetrate the atmosphere, although extremely minimally compared to the other radiations.

The **electromagnetic (EM) spectrum** is the range of all possible electromagnetic radiation.[1] The "electromagnetic spectrum" (usually just *spectrum*) of an object is the characteristic distribution of electromagnetic radiation from that particular object.





The electromagnetic spectrum extends from below the frequencies used for modern radio (at the long-wavelength end) through gamma radiation (at the short-wavelength end), covering wavelengths from thousands of kilometres down to a fraction the size of an atom. It's thought that the short wavelength limit is the vicinity of the Planck length, and the long wavelength limit is the size of the universe itself (see physical cosmology), although in principle the spectrum is infinite and continuous.

3.2 AM Radio Band

The Amplitude Modulated (AM) radio carrier frequencies are in the frequency range 535-1605 kHz. ***The frequencies 30-535 kHz are used for maritime communication and navigation and for aircraft navigation.*** Carrier frequencies of 540 to 1600 kHz are assigned at 10 kHz intervals.

CLASS	FREQUENCY	WAVELENGTH	ENERGY
γ	300 EHz	1 pm	1.24 MeV
HX	30 EHz	10 pm	124 keV
	3 EHz	100 pm	12.4 keV
SX	300 PHz	1 nm	1.24 keV
EUV	30 PHz	10 nm	124 eV
NUV	3 PHz	100 nm	12.4 eV
NIR	300 THz	1 μm	1.24 eV
MIR	30 THz	10 μm	124 meV
FIR	3 THz	100 μm	12.4 meV
EHF	300 GHz	1 mm	1.24 meV
SHF	30 GHz	1 cm	124 μeV
UHF	3 GHz	1 dm	12.4 μeV
VHF	300 MHz	1 m	1.24 μeV
HF	30 MHz	10 m	124 neV
MF	3 MHz	100 m	12.4 neV
LF	300 kHz	1 km	1.24 neV
VLF	30 kHz	10 km	124 peV
VF/ULF	3 kHz	100 km	12.4 peV
SLF	300 Hz	1 Mm	1.24 peV
ELF	30 Hz	10 Mm	124 feV
	3 Hz	100 Mm	12.4 feV

Legend

γ= Gamma rays
HX= Hard X-Rays
SX= Soft X-Rays
EUV= Extreme ultraviolet
NUV= Near ultraviolet
Visible light
NIR= Near Infrared
MIR= Mid infrared
FIR= Far infrared
Radio waves
EHF= Extremely high freq.
SHF= Super high freq.
UHF= Ultra high freq.
VHF= Very high freq.
Freq=Frequency
HF= High freq.
MF= Medium freq.
LF= Low freq.
VLF= Very low freq.
VF/ULF= Voice freq.
SLF= Super low freq.
ELF= Extremely low freq.

Frequencies: 500-1500 kHz
Wavelengths: 600 - 200 m
Quantum energies: 2 - 6 $\times$ 10^{-9} eV

3.3 Short Wave

The frequencies from the top end of the AM band to the bottom of the VHF television band are generally called the "short wave" range, a historical term. They are part of the general range referred to as "radio frequencies" or RF. The range from 1605 kHz to 54 MHz has multiple communication uses.

1,605 kHz - 30 MHz	Amateur radio, government radio, international shortwave broadcast, fixed and mobile communications.
30-50 MHz	Government and non-government, fixed and mobile. Includes police, fire, forestry, highway, and railroad services.
50-54 MHz	Amateur

The RF frequency range around 40-50 MHz is important as the proton resonance frequency range used in nuclear magnetic resonance (NMR) and magnetic resonance imaging (MRI).

Frequencies: 1.605 - 54 MHz

Wavelengths: 187 - 5.55 m

Quantum energies: $.66 \times 10^{-8}$ - $.22 \times 10^{-6}$ eV

3.4 TV And FM Radio Band

The carrier frequencies for VHF television Channels 2-4 cover the frequency range 54 to 72 MHz. There is a band from 72-76 MHz which is reserved for government and non-government services, including a standard aeronautical beacon at 75 MHz. VHF TV channels 5 and 6 are between 76 and 88 MHz. The FM radio band is from 88 to 108 MHz between VHF television Channels 6 and 7.Above the FM is a range 108-122 MHz for aeronautical navigation including localizers, radio ranging and airport control. From 122 to 174 MHz is another general service band for both government and non-government signals. It includes fixed and mobile units and amateur broadcast. Channels 7 through 13 span the frequency range 174-216 MHz. 216-470 MHz includes a number of fixed and mobile communication modes, including some aeronautical navigation and citizens radio. 470-890 MHz includes UHF television channels 14 to 83. Frequencies 890-3000 MHz include a variety of aeronautical and amateur uses, studio-transmitter relays, etc. There are radar bands 1,300-1,600 MHz.

The FM stations are assigned center frequencies at 200 kHz separation starting at 88.1 MHz, for a maximum of 100 stations. These FM stations have a 75 kHz maximum deviation from the center frequency, which leaves 25 kHz upper and lower “gaurd bands” to minimize interaction with the adjacent frequency band. Television channels have 5 MHz separation.

The frequency range for mobile cellular telephones is listed as 824.040 - 848.970 MHz.

Frequencies: 54-1600 MHz

Wavelengths: 5.55 m - 0.187 m

Quantum energies: 0.22×10^{-6} - 0.66×10^{-5} eV

L-Band for Satellite Communication

The range 390-1550 MHz in the ultrahigh radio frequency range is designated as the L-Band and is used for a variety of satellite communication purposes.

For example, the Global Positioning System uses two carrier frequencies in this band for broadcasting navigation data.

3.5 Microwaves, Radar

While there are some radar bands from 1,300 to 1,600 MHz, most microwave applications fall in the range 3,000 to 30,000 MHz (3-30 GHz). Current microwave ovens operate at a nominal frequency of 2450 MHz, a band assigned by the FCC. There are also some amateur and radio navigation uses of the 3-30 GHz range. In interactions with matter, microwave radiation primarily acts to set produce molecular rotation and torsion, which manifests itself by heat. Molecular structure information can be obtained from the analysis of molecular rotational spectra, the most precise way to determine bond lengths and angles of molecules. Microwave radiation is also used in electron spin resonance spectroscopy.

For microwave ovens and some radar applications, the microwaves are produced by magnetrons.

Of great astrophysical significance is the 3K background radiation in the universe, which is in the microwave region. It has recently been mapped with great precision by the WMAP probe.

Frequencies: 1.6-30 GHz

Wavelengths: 187 - 10 mm

Quantum energies: 0.66×10^{-5} - 0.12×10^{-3} eV

3.6 Millimeter Waves, Telemetry

The range 30-300 GHz is used for a variety of experimental, government and amateur purposes in communication.

Frequencies: 30-300 GHz

Wavelengths: 10 - 1 mm

Quantum energies: 0.12×10^{-3} - 0.12×10^{-2} eV

3.7 Software Radio

HF/VHF/UHF Band of frequencies are extensively used by the defense forces for ground to ground (static and mobile), ground to air, and air to air communications. Networking of defense units for operational data exchange and command also rely on Digital Communication Systems. These Systems are developed with advanced ECCM features. Key based synchronous switching of channel frequency (frequency hopping) for frequency agility and Spread Spectrum Techniques for avoiding detection and jamming are very common of present day Communication systems. Special secrecy systems further aid secure Communication information Channels for Voice and data transmission .These ruggedised Systems by necessity employ State-of-the- art Electronic hardware and Software Technologies.

Need for multiple features of such Radios and the need to adapt to varieties of requirements by the Defense have enthused developmental activities of a Software controlled, Digital and field configurable integrated Radio called Compact Net Radio(CNR). The CNR covers HF/VHF/UHF bands in a single unit and suitable for FM,SSB and Digital Data Communication. This is a fully programmable futuristic Radio generally termed as a ***"Software Radio"*** as no hardware tuning is involved in the System. The design is mainly based on DSP chips which emulates the functioning of IF filters, RF detectors and Local Oscillators.

4. CONCLUSION

Defence electronic systems are sophisticated and complex.But the developments taking place in sensors and Communication systems are going to be more sensitive and versatile. Weapon systems and C^4I^2 will be covered in the next paper under VOL. II.

ACKNOWLEDGEMENTS

The author is thankful to the Director, CVRDE and Associate Director, CVRDE for giving permission to carry out the work. The author is grateful to Mrs. JeyaPradha Ganesh Sc 'D', Mr Praveen M. JRF, Ms. Rekha J. JRF and all the officers and staff of ATL&N division, CVRDE for their support and valuable suggestions. The author wishes to thank Prof. Sneh Anand, IIT and Brig SC. Marwaha, VSM(Retd.) Executive Secretary, AICTE-INAE for giving this wonderful opportunity.

5. APPENDIX A

Range of the spectrum

The spectrum covers EM wave energy having wavelengths from thousands of meters down to fractions of the size of an atom. Frequencies of 30 Hz and below can be produced by and are important in the study of certain stellar nebulae and frequencies as high as 2.9×10^{27} Hz have been detected from astrophysical sources.

Electromagnetic energy at a particular wavelength λ (in vacuum) has an associated frequency f and photon energy E. Thus, the electromagnetic spectrum may be expressed equally well in terms of any of these three quantities. They are related by the equations:

$$c = \text{frequency} \times \text{wavelength or} \quad \lambda = \frac{c}{f} \quad \text{and} \quad E = hf \quad \text{or} \quad E = \frac{hc}{\lambda}$$

where c = 299,792,458 m/s (speed of light) and h is Planck's constant, ($h \approx 6.626069 \cdot 10^{-34}$ J·s ≈ μeV/Ghz)

So, high-frequency electromagnetic waves have a short wavelength and high energy; low-frequency waves have a long wavelength and low energy.

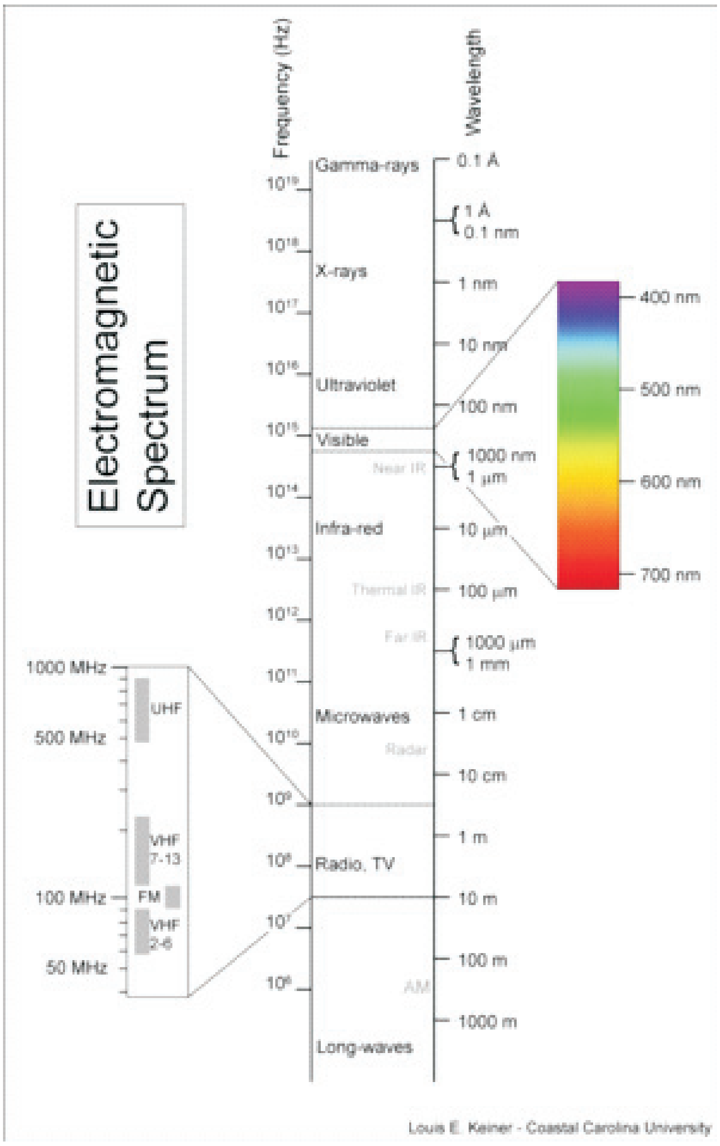
Whenever light waves (and other electromagnetic waves) exist in a medium (matter), their wavelength is decreased. Wavelengths of electromagnetic radiation, no matter what medium they are traveling through, are usually quoted in terms of the *vacuum* wavelength, although this is not always explicitly stated.

Generally, EM radiation is classified by coiled wavelength into radio wave, microwave, infrared, the visible region we perceive as light, ultraviolet, X-rays and gamma rays.

The behavior of EM radiation depends on its wavelength. Higher frequencies have shorter wavelengths, and lower frequencies have longer wavelengths. When EM radiation interacts with single atoms and molecules, its behavior also depends on the amount of energy per quantum it carries. Electromagnetic radiation can be divided into octaves — as sound waves are.

Spectroscopy can detect a much wider region of the EM spectrum than the visible range of 400 nm to 700 nm. A common laboratory spectroscope can detect wavelengths from 2 nm to 2500 nm. Detailed information about the physical properties of objects, gases, or even stars can be obtained from this type of device. It is widely used in astrophysics. For example, many hydrogen atoms emit radio waves which have a wavelength of 21.12 cm.

5.1 Types of Radiation



The electromagnetic spectrum

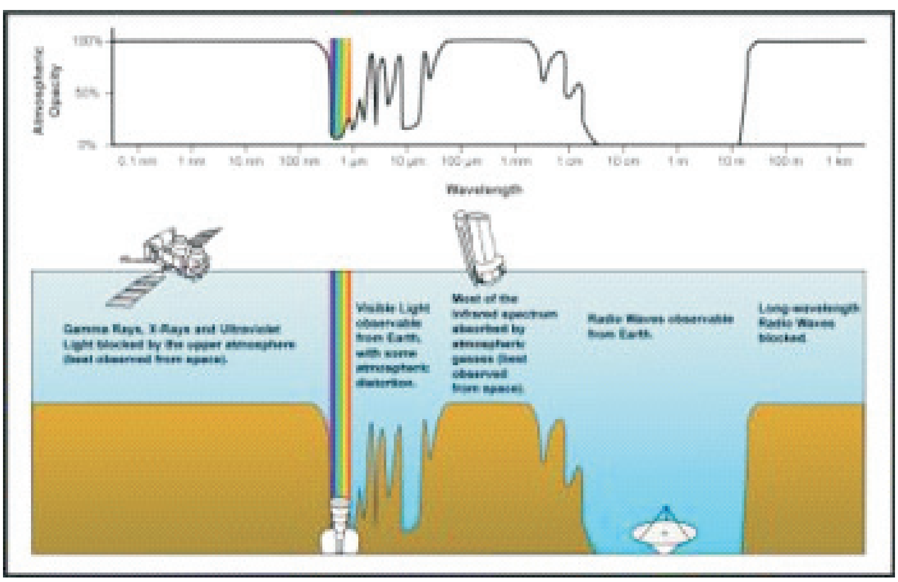
While the classification scheme is generally accurate, in reality there is often some overlap between neighboring types of electromagnetic energy. For example, SLF radio waves at 60 Hz may be received and studied by astronomers, or may be ducted along wires as electric power. Also, some low-energy gamma rays actually have a longer wavelength than some high-energy X-rays. This is possible because "gamma ray" is the name given to the photons generated from nuclear decay or other nuclear and subnuclear processes, whereas X-rays on the other hand are generated by electronic transitions involving highly energetic inner electrons. Therefore the distinction between gamma ray and X-ray is related to the radiation source rather than the radiation wavelength. Generally, nuclear transitions are much more energetic than electronic transitions, so usually, gamma-rays are more energetic than X-rays. However, there are a few low-energy nuclear transitions (e.g. the 14.4 keV nuclear transition of Fe-57) that produce gamma rays that are less energetic than some of the higher energy X-rays.

5.2 Radio Frequency

Radio waves generally are utilized by antennas of appropriate size (according to the principle of resonance), with wavelengths ranging from hundreds of meters to about one millimeter. They are used for transmission of data, via modulation. Television, mobile phones, *MRI, wireless networking* and amateur radio all use radio waves.

Radio waves can be made to carry information by varying a combination of the amplitude, frequency and phase of the wave within a frequency band and the use of the radio spectrum is regulated by many governments through frequency allocation. When EM radiation impinges upon a conductor, it couples to the conductor, travels along it, and induces an electric current on the surface of that conductor by exciting the electrons of the conducting material. This effect (the skin effect) is used in antennas. **EM** radiation may also cause certain molecules to absorb energy and thus to heat up; this is exploited in microwave ovens.

5.3 Microwaves



Plot of Earth's atmospheric transmittance (or opacity) to various wavelengths of electromagnetic radiation.

The super high frequency (SHF) and extremely high frequency (EHF) of Microwaves come next up the frequency scale. Microwaves are waves which are typically short enough to employ tubular metal waveguides of reasonable diameter. Microwave energy is produced with klystron and magnetron tubes, and with solid state diodes such as Gunn and IMPATT devices. Microwaves are absorbed by molecules that have a dipole moment in liquids. In a microwave oven, this effect is used to heat food. Low-intensity microwave radiation is used in Wi-Fi.

When active, the average microwave oven is powerful enough to cause interference at close range with poorly shielded electromagnetic fields such as those found in mobile medical devices and cheap consumer electronics.

5.4 Terahertz Radiation

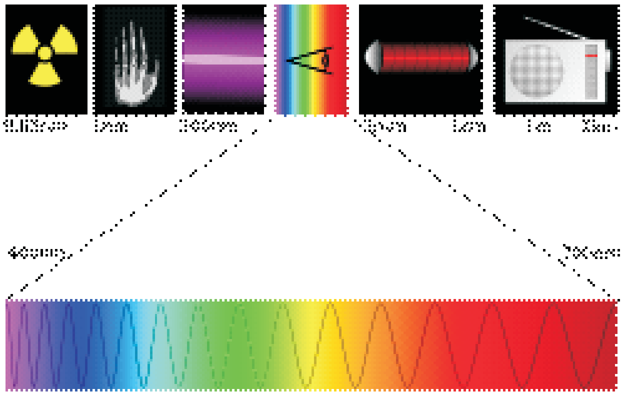
Terahertz radiation is a region of the spectrum between far infrared and microwaves. Until recently, the range was rarely studied and few sources existed for microwave energy at the high end of the band (sub-millimetre waves or so-called terahertz waves), but applications such as imaging and communications are now appearing. Scientists are also looking to apply Terahertz technology in the armed forces, where high frequency waves might be directed at enemy troops to incapacitate their electronic equipment.

5.5 Infrared Radiation

The infrared part of the electromagnetic spectrum covers the range from roughly 300 GHz (1 mm) to 400 THz (750 nm). It can be divided into three parts:

- **Far-infrared**, from 300 GHz (1 mm) to 30 THz (10 ìm). The lower part of this range may also be called microwaves. This radiation is typically absorbed by so-called rotational modes in gas-phase molecules, by molecular motions in liquids, and by phonons in solids. The water in the Earth's atmosphere absorbs so strongly in this range that it renders the atmosphere effectively opaque. However, there are certain wavelength ranges ("windows") within the opaque range which allow partial transmission, and can be used for astronomy. The wavelength range from approximately 200 ìm up to a few mm is often referred to as "sub-millimetre" in astronomy, reserving far infrared for wavelengths below 200 ìm.
- **Mid-infrared**, from 30 to 120 THz (10 to 2.5 ìm). Hot objects (black-body radiators) can radiate strongly in this range. It is absorbed by molecular vibrations, where the different atoms in a molecule vibrate around their equilibrium positions. This range is sometimes called the *fingerprint region* since the mid-infrared absorption spectrum of a compound is very specific for that compound.

Near-infrared**, from 120 to 400 THz (2,500 to 750 nm). Physical processes that are relevant for this Visible radiation (light)**



Visible Electromagnetic spectrum illustration.

The light spectrums of different grow lamps

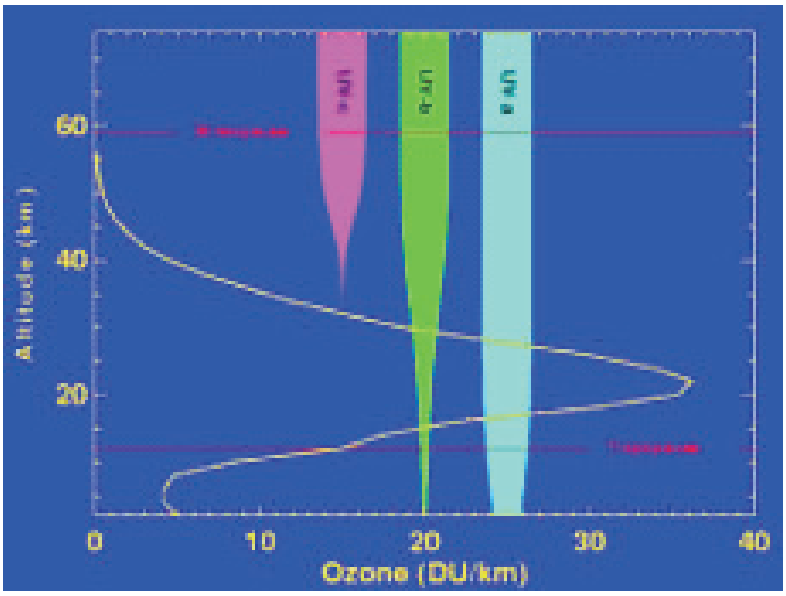
Above infrared in frequency comes visible light. This is the range in which the sun and stars similar to it emit most of their radiation. It is probably not a coincidence that the human eye is sensitive to the wavelengths that the sun emits most strongly. Visible light (and near-infrared light) is typically absorbed and emitted by electrons in molecules and atoms that move from one energy level to another. The light we see with our eyes is really a very small portion of the electromagnetic spectrum. A rainbow shows the optical (visible) part of the electromagnetic spectrum; infrared (if you could see it) would be located just beyond the red side of the rainbow with ultraviolet appearing just beyond the violet end.

EM radiation with a wavelength between approximately 400 nm and 700 nm is detected by the human eye and perceived as visible light. Other wavelengths, especially nearby infrared (longer than 700 nm) and ultraviolet (shorter than 400 nm) are also sometimes referred to as light, especially when the visibility to humans is not relevant.

If radiation having a frequency in the visible region of the EM spectrum reflects off of an object, say, a bowl of fruit, and then strikes our eyes, this results in our visual perception of the scene. Our brain's visual system processes the multitude of reflected frequencies into different shades and hues, and through this not-entirely-understood psychophysical phenomenon, most people perceive a bowl of fruit.

At most wavelengths, however, the information carried by electromagnetic radiation is not directly detected by human senses. Natural sources produce EM radiation across the spectrum, and our technology can also manipulate a broad range of wavelengths. Optical fiber transmits light which, although not suitable for direct viewing, can carry data that can be translated into sound or an image. The coding used in such data is similar to that used with radio waves.

5.6 Ultraviolet Light



The amount of penetration of UV relative to altitude in Earth's ozone

Next in frequency comes ultraviolet (UV). This is radiation whose wavelength is shorter than the violet end of the visible spectrum.

Being very energetic, UV can break chemical bonds, making molecules unusually reactive or ionizing them, in general changing their mutual behavior. Sunburn, for example, is caused by the disruptive effects of UV radiation on skin cells, which can even cause skin cancer, if the radiation damages the complex DNA molecules in the cells (UV radiation is a proven mutagen). The Sun emits a large amount of UV radiation, which could quickly turn Earth into a barren desert; however, most of it is absorbed by the atmosphere's ozone layer before reaching the surface.

5.7 X-rays

After UV come X-rays. Hard X-rays have shorter wavelengths than soft X-rays. X-rays are used for seeing through some things and not others, as well as for high-energy physics and astronomy. Neutron stars and accretion disks around black holes emit X-rays, which enable us to study them.

X-rays will pass through most substances, and this makes them useful in medicine and industry. X-rays are given off by stars, and strongly by some types of nebulae. An X-ray machine works by firing a beam of electrons at a "target". If the electrons were fired with enough energy, X-rays will be produced.

5.8 Gamma Rays

After hard X-rays come gamma rays. These are the most energetic photons, having no defined lower limit to their wavelength. It's uncertain what the physical lower limit of their wavelengths would be. They are useful to astronomers in the study of high-energy objects or regions and find a use with physicists thanks to their penetrative ability and their production from radioisotopes. The wavelength of gamma rays can be measured with high accuracy by means of Compton scattering.

Note that there are no precisely defined boundaries between the bands of the electromagnetic spectrum. Radiation of some types have a mixture of the properties of those in two regions of the spectrum. For example, red light resembles infra-red radiation in that it can resonate some chemical bonds.

REFERENCES

1. {{cite web|url=http://imagine.gsfc.nasa.gov/docs/dict_ei.html#em_spectrum|title= Imagine the Universe! Dictionary
2. What is Light? – UC Davis lecture slides
3. The Electromagnetic Spectrum, The Physics Hypertextbook
4. Definition of frequency bands on vlf.it
5. J. J. Condon and S. M. Ransom. Essential Radio Astronomy: Pulsar Properties. National Radio Astronomy Observatory. Retrieved on 2008-01-05.
6. A. A. Abdo; B. Allen; D. Berley; E. Blaufuss; S. Casanova; C. Chen; D. G. Coyne; R. S. Delay; B. L. Dingus; R. W. Ellsworth; L. Fleysher; R. Fleysher; I. Gebauer; M. M. Gonzalez; J. A. Goodman; E. Hays; C. M. Hoffman; B. E. Kolterman; L. A. Kelley; C. P. Lansdell; J. T. Linnemann; J. E. Mc Enery; A. I. Mincer; I. V. Moskalenko; P. Nemethy; D. Noyes; J. M. Ryan&#x A;;&#x A; F. W. Samuelson&#x A;;&#x A; P. M. Saz Parkinson; M. Schneider; A. Shoup&#x A;;&#x A; G. Sinnis&#x A;;&#x A; A. J. Smith; A. W. Strong; G. W. Sullivan; V. Vasileiou; G. P. Walker; D. A. Williams; X. W. Xu; G. B. Yodh (2007 March 20). "Discovery of TeV Gamma Ray Emission from the Cygnus Region of the Galaxy". *The Astrophysical Journal Letters* **658**: L33. doi:10.1086/513696.
7. Isaac Asimov, *Isaac Asimov's Book of Facts*. Hastingshouse/Daytrips Publ., 1992. Page 389.
8. Hyperphysics (see Gamma-Rays)

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

GUN CONTROL SYSTEM FOR FUTURISTIC MAIN BATTLE TANK (2020)

S. MADIVAANAN*

ABSTRACT

State of art, All Electric Gun Control System (GCS) is proposed as an essential feature for the much envisaged MBT 2020. This has the requisite long time battle standing features including high RAM-D(Reliability, Availability, Maintainability and Durability). All Electric GCS is a complete electrical system consisting of all electrical and electronic devices and components viz Electrical Motors, Solid State Power drives, Digital Controllers, Sensors etc. Thus it will not have any hydraulic oil or components and will enable the crew to overcome the psychological apprehension of oil leakages and fire hazards. This shall improve "Survivability under stringent battle conditions".

State of Art Electric gun control system includes static DC-AC inverter, Capacitor accumulator, Light weight brushless electric motors, stabilization electronic and Electromagnetic manual drive mechanism. This system provides an edge over the conventional DC electric drives with commutator and brush assemblies since the motors are controlled electronically. BLDC motors can potentially be deployed in any area currently fulfilled by Brushed DC motors. Cost and control complexity prevents BLDC motors so far from replacing Brushed motors in most common areas of use. Nevertheless BLDC motors have come to dominate many applications viz Computer Hard drives, CD/DVD players and PC cooling fans, electric vehicles etc. It offers higher efficiency and reliability, reduced noise, longer life time (no brush erosion), elimination of ionizing sparks from the commutator and overall Electromagnetic interference (EMI). Thus system maintenance is minimized.

A true silent watch capability without switching engine ON provides most useful mode for the crew. Its feature like fast acceleration and ease of laying gun on target makes the choice mandatory for incorporation in MBT beyond 2020.

1. INTRODUCTION

1.1 Main Battle Tank (MBT)

- MBT is the state of the art Tank adopting most advanced technologies for superior fire power, high mobility, excellent protection and crew comfort.

---

*Scientist 'F', Head (ATL&N Division), CVRDE, AVADI, Chennai-54

FIG. 1 MAIN BATTLE TANK

Generally It has two major subsystems

1) Turret and weapon system
    - Turret structure
    - Ordinance (gun barrel, breaching, fume extractor and thermal jacket)
    - Gun control and fire control assemblies
    - Communication system
    - Special systems viz. NBC, global position system, smoke grade system
2) Basic hull structure
    - Power pack
    - Running gear

Main battle tanks are evaluated by three key parameters. They are:

a) Mobility
b) Fire Power
c) Protection

This fire power is taken care by reliable Gun Control System.

The basic characteristics of an ideal GCS is to ensure the first round hit in the target in the least possible time. The fire must be effective under all such conditions when both the task and target are stationary, when both task and target are moving and when one task is stationary and the other is moving. Moreover, coming across multiple targets simultaneously may be a routine affair. This will require a definition of priorities of engagements and fast reliable/acquisition time. In order to achieve these requirements, stabilized and fully automated gun traverse/elevation, ballistic compute etc are absolute necessities.

2. GUN CONTROL SYSTEM (GCS)

Gun Control System is a built in feature of every battle tank ever since the tanks were employed in the wars. Initially simple gear and ram along with a rotating handle was sufficient for the crew to move the gun in up and down directions. Later on, with the advent of rotating turret and gun mounted on it, moving mechanisms for Gun and Turret were introduced.

As the gun barrels and Turrets started increasing in caliber and size/weight in order to add more fire power to the tank, simple hydraulic actuators like cylinder and hydraulic motors were used initially. Later on, Electrical Machines followed to replace cumbersome hydraulics by Amplidyne and DC motors for obvious reasons like leakages, risk of fire etc.

With the advent of sophisticated Gunners Main Sight and requirement of responding to moving targets from moving tanks of 60 tons class MBT's, hydraulic systems again found the place to meet higher load inertia of 20 ton Turret and 120mm barrel and to stabilize the Gun/Turret with respect to vehicles pitch, roll and yaw disturbances across country. Even though these are in use for decades, there are inherent problems such as fumes due to high pressures of the order of 160 bar leakages, safety, noise due to pump and slow emergency manual drives etc. So it has become imperative to look for alternatives, like All Electric Gun and Turret Drive Systems (EGTD).

Components of Gun Control System

- Hydraulic Power Pack
- Gear box
- Elevation Cylinder
- Gunners Control Handle
- Commander Control Handle
- Gyro
- Servo Package
- Distribution Block
- Shut off Valve
- Transfer Valve
- Digital Combat Electronic Unit
- Gunner's Control Box

Now with the availability of Technology for various capacities of Inverters, compact brushless (AC Synchronous) motors, digital controls, sensors etc., it is possible to adapt EGTD.

Present environment in the battlefields calls for maintenance force, fast acting weapon to ensure very high first round hit probabilities. Safety also becomes paramount importance due to chemical/ nuclear warfare. This requires closed hatch operations for the crew for long hours. All EGTD ensures better survivability and instill confidence in the crew in the absence of hazards of fire fumes

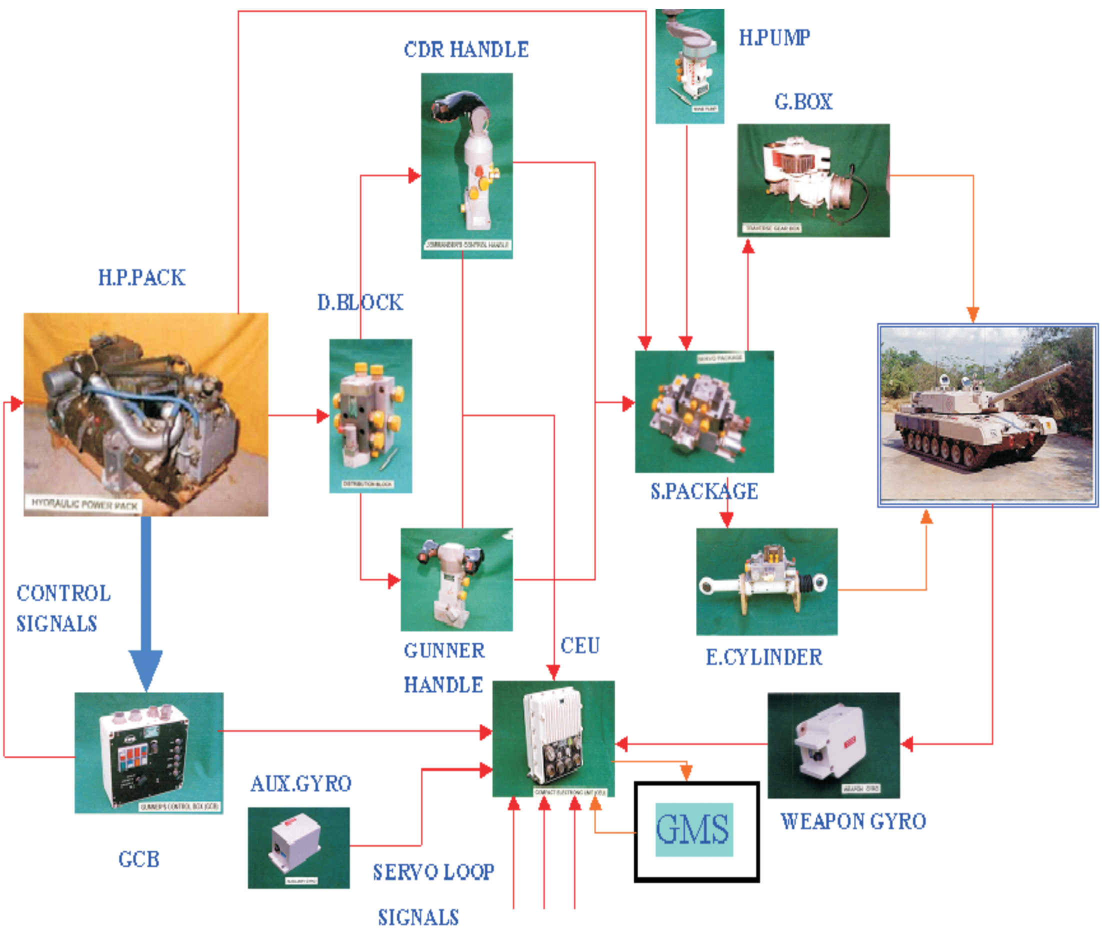


FIG. 2 ELECTRO HYDRAULIC GUN CONTROL SYSTEMS–BLOCK DIAGRAM

or oil leakages. Further due to its simplicity and ease of interface with modern fire control systems the operation becomes simple and easy.

2.1 Feature of All Electric Gun Control System

Basic System configuration is shown in Figure 3 . The system consists of DC-AC Booster inverter for boosting tank battery voltage to around 270V AC. This 3 phase power is supplied to the brushless synchronous motors through power control modules. These modules have IGBT devices, controlled by Electronic Switching as per control signals from either gunner handle or commander control handle.

Between the Inverter and power modules, a capacitor bank is connected to meet the high peak current demands during slewing accelerations. Capacitor banks provide additional advantage of energy recovery and getting recharged during the breaking phases from the gun and Turret kinetic

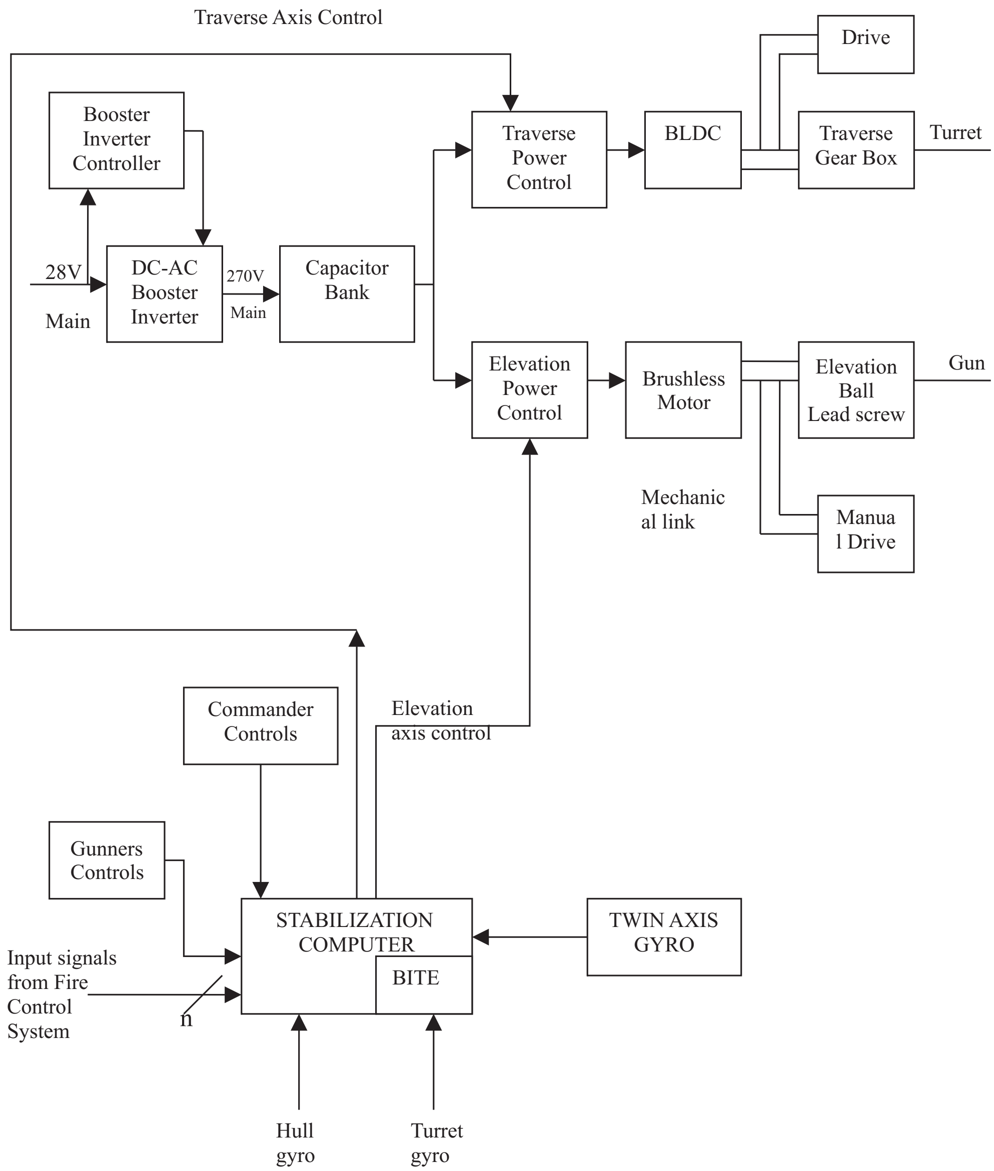


FIG. 3 GUN CONTROL SYSTEM FOR FUTURISTIC MBT 2020

energies. This saves lot of energy from getting otherwise dissipated in thermal form. Enhanced dynamic performance is achieved through its high acceleration capability, combined with servo configuration.

Stabilization computer provides correction signals to isolate the gun and turret platform from the Vehicular disturbances and stabilize the gun to keep pointing in the desired position. The computer system performs this by using the signals from the twin axis gyro package mounted underneath the gun bridge, two single axis gyro fitted in Hull and Turret and the two control handles.

Combined performance of stabilization computer along with the dynamic drive results in fire laying the gun as well as tracking the moving target while tank itself is also moving. The capability of fire on the move requires through knowledge of vehicle motions depending upon its characteristics including suspension system and the structural data of Gun and Turret which are very complex nature.

System has capability to get interfaced with the fire control system so as to implement the super corrections in traverse and gun angles, in relation with ballistic characteristics of ammunition and meteorological parameters.

The key elements of All Electric Drive are:

- High performance brushless dc motors
- Backlash compensated compact two stage reduction gear boxes.
- Advanced control and power electronics

2.2 Advantages of All Electric Gun and Turret Drive System

As it is well established, All Electric Gun and Turret Drive System have many advantages over hydraulic systems. Salient points are given below.

1. No cumbersome Pump Motor combination makes operation totally noise free.
2. Silent watch operation, without Engine ON.
3. Stabilization requires less than 30% of Electric power than that of Hydraulic system.
4. No parts subject to wear as in Hydraulic Systems and no filters for cleaning.
5. Continuous upgradation and retrofitment in other tanks is easy.
6. Space and weight efficient design (400 kg in place of 800 kg hydraulic system)
7. Easy maintainability and reliability
8. Increased MTBF and survivability
9. Capacitive Energy Accumulator permits
    - Delivery of high peak power to the Turret.
    - Recovery of Gun/Turret kinetic energy due to regenerative braking.
10. Enhanced Performance

	Traverse	Elevation
Maximum Slewing speed	>0.8 rad/sec	>0.8 rad/sec
Maximum acceleration	>0.6 rad/sec^2	>2.2 rad/ sec^2
Laying time	< 6 sec for 180^o	<1.5 sec for 30 days
Stabilization accuracy	<0.2 mil	<0.2 mil

11. Longer life span
12. Simple operation and imparting training to crew is easy

3. CONCLUSION

With the fast changing battlefield scenario, involving nuclear, biological, chemical weapony tanks remains the most needed and main advancing force of the country. Their effectiveness is important to get an edge over enemy. To achieve this state of art weapon control system is mandatory for the tanks beyond 2010 and AEGTD, with inherent features only can fulfill these demands.

4. ACKNOWLEDGEMENTS

The author is thankful to the Director, CVRDE and Associate Director, CVRDE for giving permission to carry out the work. The author is grateful to Mrs.JeyaPradha Ganesh Sc ‘D’, Mr. Praveen M. JRF, Ms. Rekha J. JRF and all the officers and staff of ATL&N division, CVRDE for their support and valuable suggestions. The author wishes to thank Prof. Sneh Anand, IIT and Brig SC Marwaha, VSM (Retd) Executive Secretary, AICTE-INAE for giving this wonderful opportunity.

5. REFERENCES

[1]. http://publications.drdo.gov.in
[2]. www.guncontrol.ca
[3]. www.machinedesign.com
[4]. www.security-int.com
[5]. www.israeli-weapons.com
[6]. www.bel-india.com
[7]. www.berr.gov.uk

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

TQM FOR ENHANCED AUTOMOTIVE PERFORMANCE

N. GOWRISHANKAR*

ABSTRACT

The paper explains the difference between conventional quality control and TQM. The models of TQM and Tools of TQM are discussed. With specific reference to piston ring industry, the importance of strategic decisions, process sequence in manufacture, effect of minor dimensions on performance, and criticality of process knowledge and component performance knowledge are discussed with examples.

WHAT IS TQM?

The word "Total" in Total Quality Management also means that everyone in the organization must be involved in the continuous improvement effort, the word "Quality" shows a concern for customer satisfaction, and the word "Management" refers to the people and processes needed to achieve the quality.

The word Customer in TQM refers not only to external customers but also to internal customers. In an organization with many interacting processes that ultimately produce the product, there will be a chain of internal suppliers and customers. Each one should function as a supplier or the customer depending upon his role in the value chain. The involvement of each and everyone concerned with such processes and their treating their next process as their customer alone will bring real customer satisfaction to the ultimate customer.

One important aspect of TQM is that it centers round employee and his involvement.

TQM will be successful only when there is Total Employee Involvement (TEI).

TQM involves systematic implementation of certain steps and sustaining these on a continuous basis. This requires a systematic review of effectiveness of the actions through the popularly known PDCA cycle or the Deming Cycle.

*IP Rings Ltd, D11/12 – Industrial Estate, Maraimalai Nagar – 603 209, Tamil Nadu

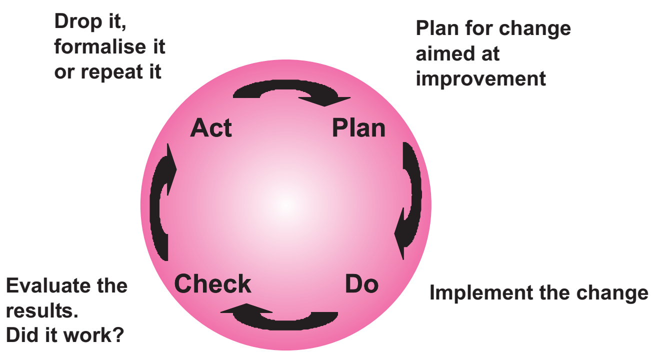


FIG. 1 PDCA CYCLE

Conventional Quality Control Vs. TQM

Conventional Quality control is focused on satisfying customer requirements and hence is Static.

Total Quality Management is focused on satisfying customer's total requirements and also quality from the manufacturer's point of view, therefore it is Dynamic.

The customer requirements in the first case are as given by the customer, whereas in the latter case it is the *manufacturer's understanding* of the customer's requirements and the commitment to improve on it without customer provocation.

Conventional Quality Control focuses on building Quality through Inspection, whereas TQM addresses the inherent capability and robustness of the manufacturing process to produce consistent, reliable quality at optimal cost.

The following models show how TQM addresses and helps organizations to achieve total customer satisfaction.

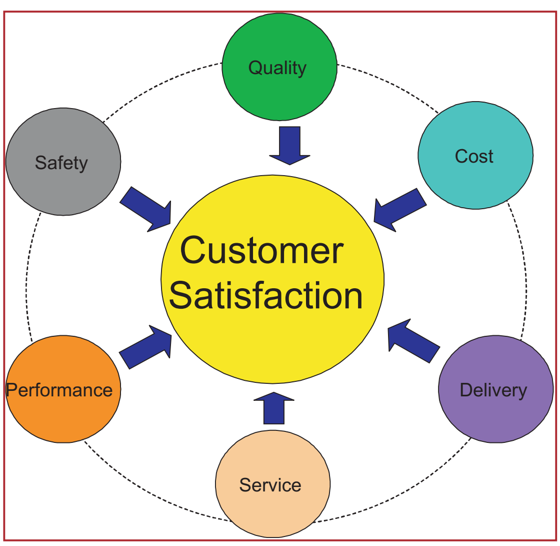


FIG. 2 CUSTOMER SATISFACTION

QUALITY

It is not mere conformance to specifications. Customer satisfaction is achieved by ensuring and building quality up to end of its life cycle. The Quality Function has moved from Inspection to Quality Control to Quality Assurance. Today the producer is responsible for quality of his product. Inspection as a separate activity by an independent team is a non value added activity. The process ensures quality and the quality person has to do only an audit.

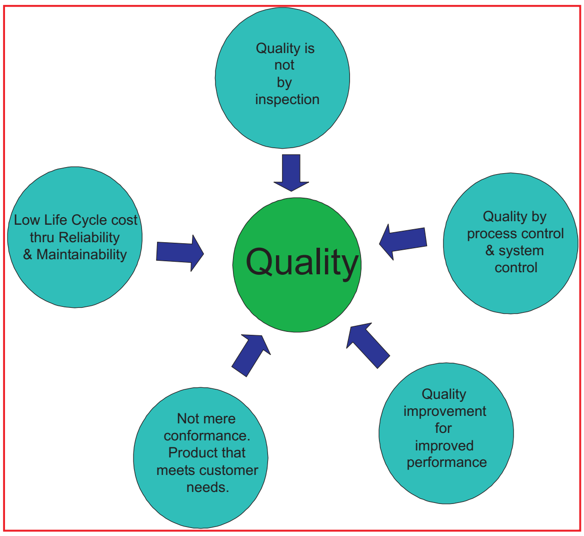


FIG. 3 QUALITY FUNCTION

COST

The customer expectation that the product should be of highest quality with lowest cost can be achieved only if there is a conscious and continued effort to eliminate losses and non value added activities in the manufacturing chain. TQM emphasizes this concept.

DELIVERY

With customers focusing on Just-In-Time Deliveries, Customer satisfaction can be achieved only through delivery of products in right quantity at right time.

QUALITY IN DOCUMENTATION DURING DELIVERY- EXAMPLE

We were procuring plasma Mo powder (for plasma spraying for piston ring coating) from source A in the USA. Mistakes were observed in their documentation resulting in untold misery for us with our Customs clearance. We appraised source A of our difficulties.

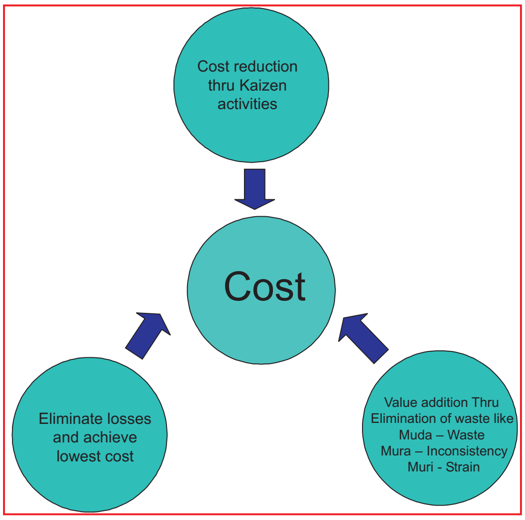


FIG. 4 COST

FIG. 5 DELIVERY

But their mistakes in documentation (typing errors) continued, causing more problems for us with our Customs.

Hence we developed alternate Source B from the USA for the same plasma Mo powder.

Source A has thus lost our huge orders because of their typing errors in documents.

PERFORMANCE

The life expectancy of the product should not only be met but bettered and for this, the supplier should constantly work and devise suitable Online and Offline testing of his products such that there will be continuous enhancement of product quality. The supplier has to use such testing methods and evaluate his products with that of his competitors so that he can improve his losses and always deliver the best to his customers. In the manufacturing process, variation reduction is important since product performance depends upon this parameter. While the customer specifies "a tolerance", the supplier has to ensure that he constantly studies his process and improves it for giving product to a closer tolerance. Statistical Process Control is a pro active tool which can be successfully used for 1) Variation reduction and 2) Minimizing rejections.

SERVICE

Gone are the days when the supplier had only to ensure that his product passed the inward inspection at the customer end. Today most of the customers consider inward inspection as a non value added activity and demand that products are directly supplied to the line. Also the customer expects that the supplier understands the voice of the customer, the technical requirement of the product from assembly to the field and his changing needs. The Quality Function Deployment addresses these requirements and the supplier has to comply with this.

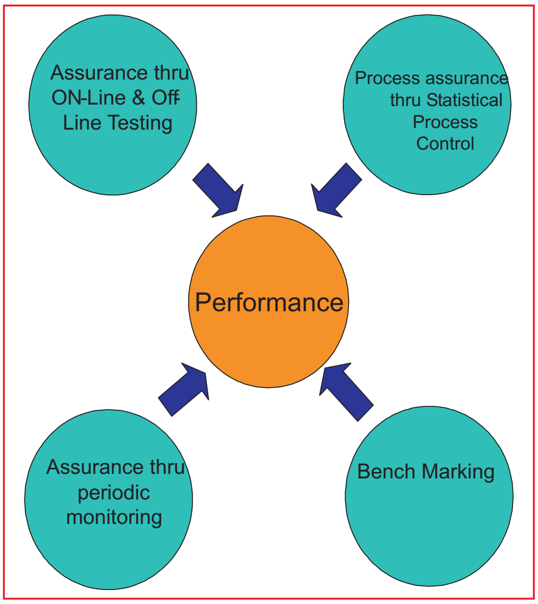


FIG. 6 PERFORMANCE

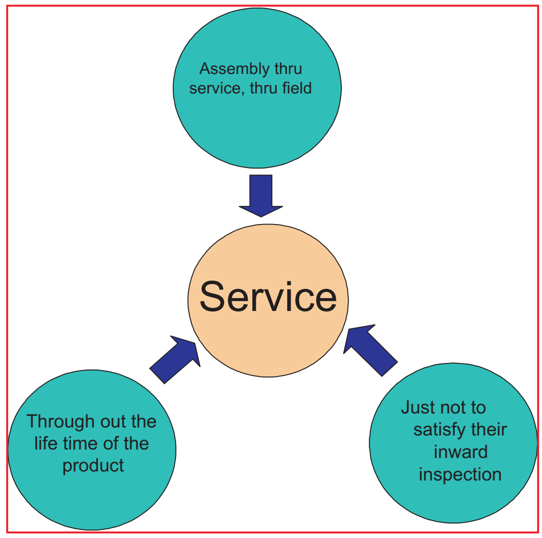


FIG. 7 SERVICE

SAFETY

When we talk about safety it is not the product safety alone. The safety starts from Design to process of manufacture till it is assembled and used. The supplier has to ensure that the process, the material he employs do not affect the environment and the personnel in the manufacturing and assembly lines.

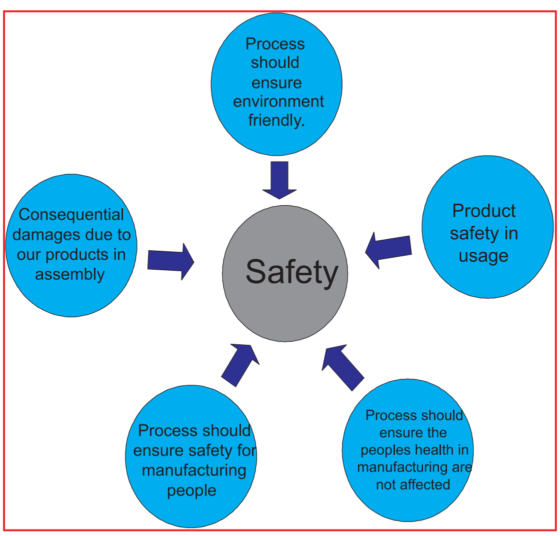


FIG. 8 SAFETY

TQM UMBRELLA

TQM is not a stand alone process.

It encompasses all initiatives for a good work place management and good business management.

If the TQM umbrella has to be effective and protect an industry, all initiatives are important.

5S is the foundation and back bone for all improvement processes and it is the Binding stick of the umbrella.

Various other initiatives listed are like the spokes which are vital to hold the protective umbrella cover in place.

For the umbrella (TQM) to function effectively, both the stick and spokes have to be strong.

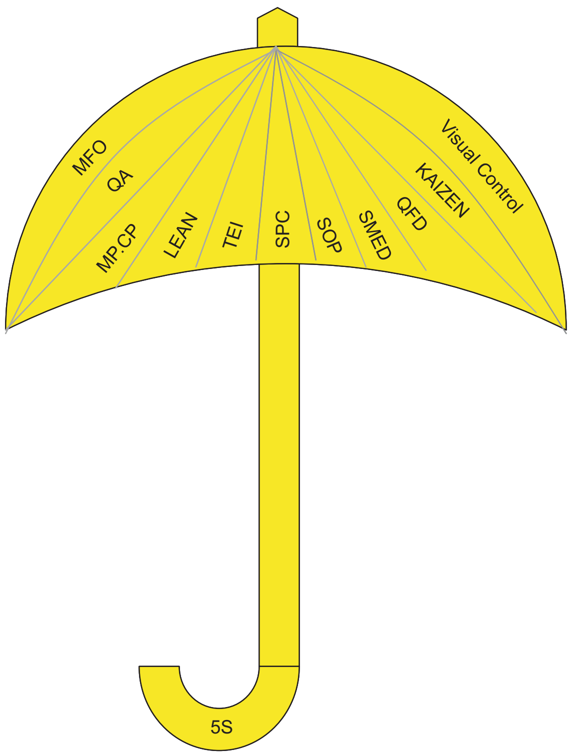


FIG. 9 TQM UMBRELLA

TQM IMPLEMENTATION

Starts with creating the organization's Vision and Mission and ends with converting these missions into reality, thus achieving the company's Vision.

The Vision & Mission of the Top Management is converted into goals which help in framing the company objectives.

Subsequently these are synthesized to department objectives and actions.

Thus it is a team effort with total synergy.

IMPORTANT TQM TOOLS

- 5S
- Quality Mgt System(QMS)
- Inventory Management
- Lean Manufacturing
- Daily Works Management(DWM)
- Management For Objectives(MFO)
- Managing and Checking Points(MP CP)

1. TQM Tools- 5 S

Table 1

TQM Tool	*Through*	*Result*	*Impact on the organization*
5 S	Total Employee involvement	• Daily management discipline. • Waste elimination. • Operator ownership.	• Improvement culture. • Neat workplace.

2. TQM Tools-QMS

Table 2

TQM Tool	*Through*	*Result*	*Impact on the organization*
Quality Management System	• Standards • Process capability • Standard operating procedures • Process mapping • FMEA • QFD	• Reducing Customer complaints and warranty claims • Reduction in process rejection.	• Quality Excellence

3. TQM Tools - Inventory Management

Table 3

TQM Tool	*Through*	*Result*	*Impact on the organization*
Inventory management	• Just In Time • Leveled production • FIFO	• Improved cash flow • Good quality • Work place area saving	• High Inventory Turn Ratio • Working capital management

4. TQM Tools - Lean Manufacturing

Table 4

TQM Tool	*Through*	*Result*	*Impact on the organization*
Lean manufacturing	• Lean management • Cellular manufacturing • SMED • Elimination of NVA	• Stability in process • Reduced lead time • Elimination of all types of wastes	• Lean thinking

5. TQM Tools - Daily Works Management

Table 5

TQM Tool	*Through*	*Result*	*Impact on the organization*
Daily works management	• Total Employee Involvement • Maintenance of Exactness	• Standardized production quantity • Predictable results • Reduced minor stops • Autonomous production	• Retainment and Improvement

EXACTNESS AND DAILY WORKS MANAGEMENT

Management has three key functions – (RIB) Retainment, Improvement and Breakthroughs. These are on a maturity scale, i.e. it is not possible to improve without retainment, and no breakthrough is possible until improvement is a culture.

In Total Quality Management, the phase of retainment is built on two key factors: exactness and daily works management. Becoming 'purposeful' or 'exact' in operations means having full control over the quantity, quality and delivery performance areas. Exactness of operations comes through exactness in: Man, Method, Material, Machine and Environment.

Most important factor to MAINTAIN THE EXACTNESS IS WHEN WE ARE HANDLING EXCEPTIONS. Who will decide the actions if there is a crisis such as sudden changes in the product or quantity (unplanned), quality to be produced, changes in inputs (different supplier) etc.

must be identified. Experience of dealing with crisis should lead to more flexible and responsive systems.

One easy method of ensuring exactness is usage of a checklist for monitoring the 4M condition.

The checklist can be tailor made to suit the industries' requirement. An example of a typical checklist is shown in the figure below:

IP Rings Ltd, MM Nagar		TENTATIVE Autonomous Maintenance Checklist																									STATION :				1		
SL	Circle Leader:Sudhir Rao	Machine :**Double Slide Press**																Cost centre :									Month :				**Jun-06**		
No	ITEM	FREQ.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
	Cleaning																																
1	Die block area	Daily	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
2	Cutting position area	Daily	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
3	Cutting burr collection tray	*	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
4	Overall body	Daily	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
5	Oil collection guide	Daily	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
6	Back side of vent punch	Daily	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O

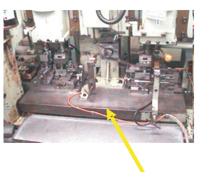
IP Rings Ltd, MM Nagar		**TENTATIVE AU TONOMOUS MAINTENANCE STANDARD**						STATION : 1	
Circle Leader : Sudir B Rao		**Machine : Double Slide Press**			Cost Centre :			DATE MADE : 01.02.03	
SI No.	**ILLUSTRATION**	**ACTION**	**ITEM**	**REQUIRED CONDITION**	**METHODS, TOOLS & MATERIALS**	**RESP**	**FREQ**	**TIME REQD**	**REVISION No & Date**
1		Cleaning	Die block area	Free from dust and oil	Machine **"OFF"** use Dry cloth, brush	Operator	Daily	1 min	***1 /1.2.03*** ***2 /5.6.03*** *inspection pt no 8 added* ***3 /1.10.03*** *inspection*

FIG. 10 TYPICAL CHECKLIST

The second factor in the 'Retainment phase' namely Daily Works management is expected to make flow of material smooth in the operations. Daily Works management is built on a strong foundation of 'exactness' where a sense of 'purpose' is visible at the operation level.

The objective of Daily Works Management is also to develop a degree of comfort for people to deal with data. The operating people collect the performance data like production quantity, scrap quantity, machine downtime, etc., and react to bridge the gap between the target and actual.

What we expect to see in a DWM practicing company:

- All losses like Breakdown, Setup, Tool change are captured.
- Permanent countermeasures for abnormalities.
- Productivity and Quality improvement.

WINDING CELL

CHROME PLATING

2 PIECE OIL RING CELL

FIG. 11 OWNERSHIP TAKEN BY OPERATORS

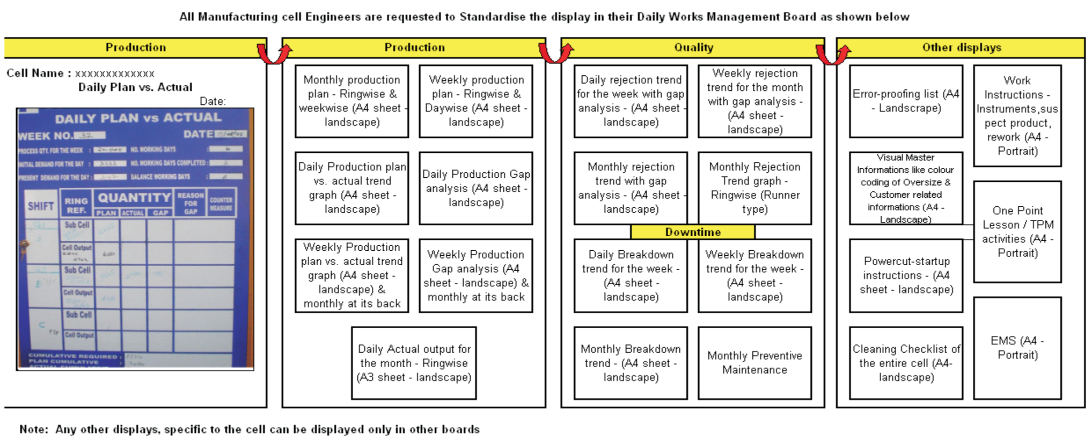
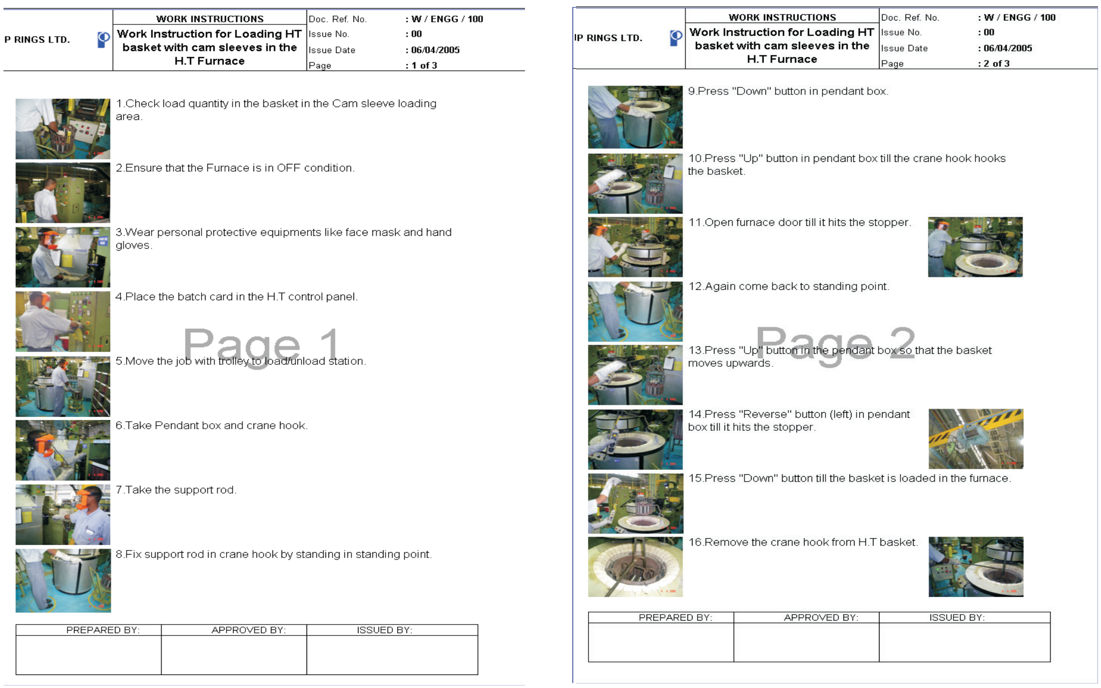


FIG. 12 DAILY WORKS MANAGEMENT BOARD

- Close monitoring of the performance and timely remedial actions help achieving the plan and there by delivery commitments to customers.

STANDARD OPERATING PROCEDURES (SOPS)

Standard operating procedures are essential to ensure that the required product quality level, consistency, effectiveness and efficiency are realized in a process. SOP s explain the step by step process required to be performed to carry out a job in the documented form.


P RINGS LTD.
WORK INSTRUCTIONS
Work Instruction for Loading HT basket with cam sleeves in the H.T Furnace
Doc. Ref. No. : W / ENGG / 100
Issue No. : 00
Issue Date : 06/04/2005
Page : 1 of 3

1.Check load quantity in the basket in the Cam sleeve loading area.
2.Ensure that the Furnace is in OFF condition.
3.Wear personal protective equipments like face mask and hand gloves.
4.Place the batch card in the H.T control panel.
Page 1
5.Move the job with trolley to load/unload station.
6.Take Pendant box and crane hook.
7.Take the support rod.
8.Fix support rod in crane hook by standing in standing point.

PREPARED BY:	APPROVED BY:	ISSUED BY:

IP RINGS LTD.
WORK INSTRUCTIONS
Work Instruction for Loading HT basket with cam sleeves in the H.T Furnace
Doc. Ref. No. : W / ENGG / 100
Issue No. : 00
Issue Date : 06/04/2005
Page : 2 of 3

9.Press "Down" button in pendant box.
10.Press "Up" button in pendant box till the crane hook hooks the basket.
11.Open furnace door till it hits the stopper.
12.Again come back to standing point.
Page 2
13.Press "Up" button in the pendant box so that the basket moves upwards.
14.Press "Reverse" button (left) in pendant box till it hits the stopper.
15.Press "Down" button till the basket is loaded in the furnace.
16.Remove the crane hook from H.T basket.

PREPARED BY:	APPROVED BY:	ISSUED BY:

FIG. 13 WORK INSTRUCTIONS

Since verbal instructions change over time and are forgotten easily, documented SOPs provide a baseline to assure controlled work process, irrespective of people or shift or any other variable.

Good SOPs can be used in every industry to help improve almost every process.

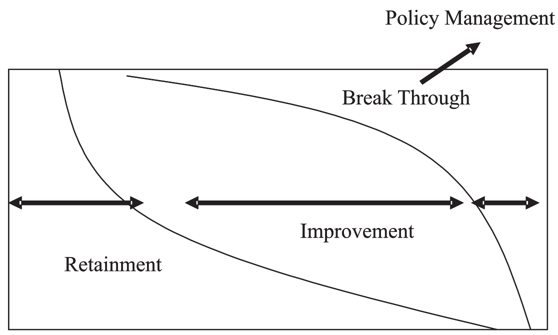
Characteristics of a good SOP

- Easy to read (taking into consideration the language skills of the workers)
- Visual
- Only includes the provided tools and materials.
- Tested and approved by workers and management.
- Meets safety and quality standards.

6. TQM Tools -MFO

Table 6

TQM Tool	*Through*	*Result*	*Impact on the organization*
MFO (Mgt for Objectives)	• Business planning	• Enhances the creativity and innovativeness of managers.	• Break thro' results
	• PDCA		• Business Excellence



- Top Mgmt
- Div. Head
- Dept. Managers.
- Supervisors
- Operators
- Share of Responsibility
- Retainment : To maintain Quality level already achieved
- Improvement : Improve Quality level more than that of achieved
- Breakthrough : Innovation of Improvement

FIG. 14 RETAINMENT, IMPROVEMENT AND BREAKTHROUGH

MANAGEMENT FOR OBJECTIVES (MFO)

MFO can be initiated in a company to set and achieve breakthrough objectives once there is an infrastructure in place for management of daily routine work. While DWM raises the capability of managers to be able to retain and improve upon their performance areas by setting standards and improving standards, MFO raises further capabilities to build up creative and innovative talents of managers for breakthroughs.

Step 1: Identify a company objective that we want to deploy. This objective must be a stretch target – which you cannot achieve through normal means that we adopt every day

Step 2: Develop company guidelines – there must be involvement of senior people representing interests of business and understanding the function of each department. These guidelines must be such that each department can interpret them in a creative manner and help to bring a focus in the creativity. E.g. 'Use of Information technology' as against 'use e-mails for communication'.

Step 3: Identify departmental objectives. These are not a list of ideas. These are measurable and targeted – time bound objectives. These must also be in terms of the same measures set for the company objective for which they are derived.

Step 4: Identify departmental action plans. These action plans must arise from the responsibilities and initiative of the department people themselves. Actions must be derived from: The analysis of the gaps from last years plan vs. achievements

Step 5: Put in a system for monitoring all action plans

Step 6: Analysis of performance against company objective, departmental objectives and action plans.

Step 7: Select the next cycle objective(s). From bottom-up there needs to be an analysis of last year's performance against each objective in each department and plans for improvements proposals coming up to the top from each department/division/section of the company.

X-MATRIX

The X – matrix is only a reference for showing all the elements up to Step 4. The X matrix indicates the correlation of the department objective to the company objective(s).

Typical X-Matrix is shown below:

7. TQM Tools -MP CP

Table 7

TQM Tool	*Through*	*Result*	*Impact on the organization*
MP-CP (Managing point and checking point)	• Action plans aligned to the business plan targets	• Actuating PDCA in real life.	• Aligning people thro PDCA thinking
	• Cascading the objectives thro different levels of the hierarchy and realizing thro various actions		• Realizing business goals
	• PDCA		• Business Excellence

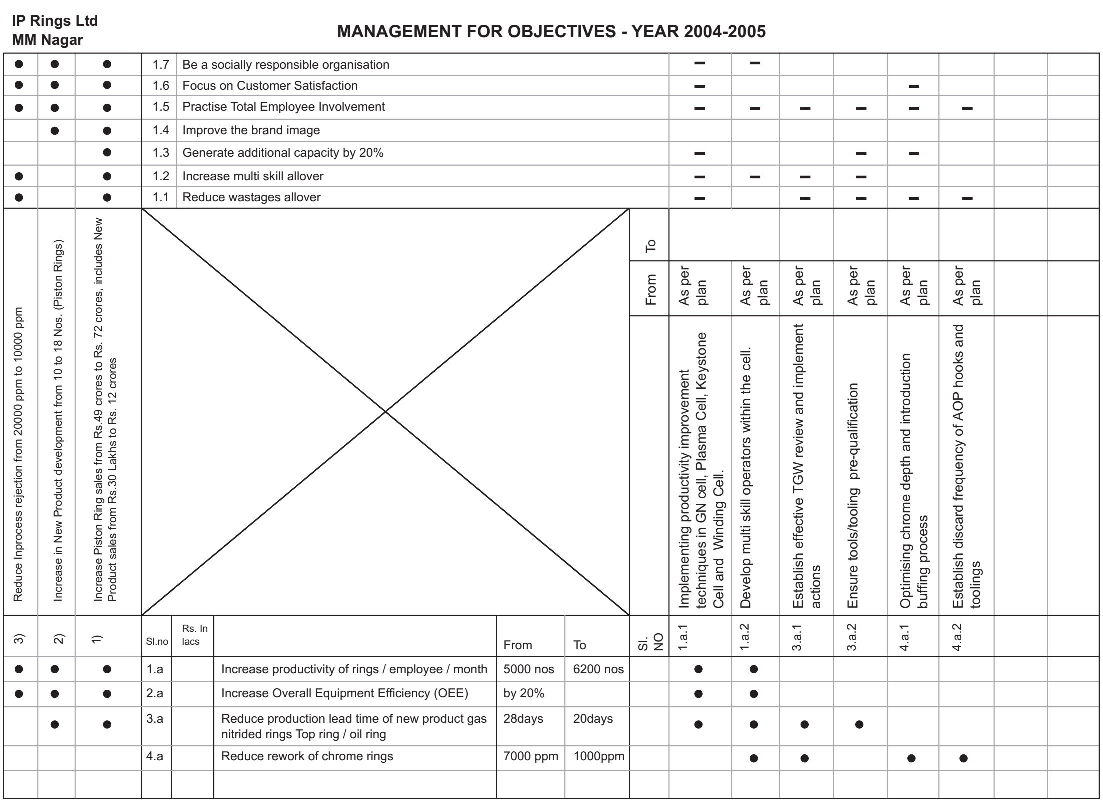


FIG. 15 X-MATRIX

MANAGING POINTS (MP) AND CHECKING POINTS (CP)

The practice of the PDCA cycle is a logical way of thinking which should reflect in everything that an organisation does to achieve its' business results. The question is how to actualize the PDCA in real life. TQM has a system supported with appropriate tools and techniques to make it happen. The purpose of TQM is to align people with the PDCA way of thinking for achievement of business excellence.

Managing and Checking points should be chosen once the organisation has a proper organisation structure with some hierarchy and functions represented.

Managing points and checking points systematically ensures the connectivity of various levels in the hierarchy and their reporting functions. Thus the Managing point of the subordinate will be the checking point for his immediate superior. In this way all the actions initiated towards achieving the overall objective are converted into easily measurable targets and monitored for achievements.

QFD

QFD (Quality Function Deployment) is an important part of TQM activity.

The voice of the customer and the stated and unstated needs of the customer are of paramount importance.

Here one has to remember the story of the reportedly highly nutritional dog biscuits which were a failure in the market because the dogs did not like them.

If the functionality of the product is not understood, we can end up in two possibilities, both of which are bad for the organization.

1) A substandard product – which will not sell
2) A costly or wasteful product – which may sell but will not bring profits to the manufacturer.

COMPANY NAME:	M/s IP Rings Ltd							MANAGING POINTS & CHECKING POINTS				
DEPT/SECTION			NAME R. SRIRAM		DESIGNATION MANAGER - PRODUCTION				21.5.2004 ISSUE-04			
	ACCOUNTABILITY		MANAGING POINTS	TARGET (UM)	VCs REF:	FREQ.		CHECKING POINT	TARGET (UM)	VCs REF:	RESP.FOR C.P	FREQ
A1	DELIVERY	M1	Production Plan adherence Keystone Cell Gn Cell Oil Cell Nifflex Cell Vc & 2nd ring cell	90 % & No. of rings	Bar graph/ Trend Graph	Weekly	C1	Production plan Adherence Keystone Cell Gn Cell Oil Cell Nifflex Cell Vc & 2nd ring Cell Winding cell Plasma cell Cr. Plating	Plan Vs Actual	Table	Shift In-Charge	Daily
		M2	OEE of bottleneck machines	Above 70%	Trend graph	Weekly	C2	Breakdowns due to poor JH	Zero	Trend graph	Dy. Mgr.-Prod.	Weekly
							c3	Machinewise Downtime due to				
								Startup	Std. Vs Actual			
								Changeover	Std. Vs Actual	Table	Dy. Mgr. Prod.	Weekly
								Operator absent in bottleneck m/cs	Std. Vs Actual			
							C4	No. of abnormality tags pending	Pending tags<30	Trend graph	Cell Enginner	Weekly

FIG. 16 EXAMPLE OF MP – CP

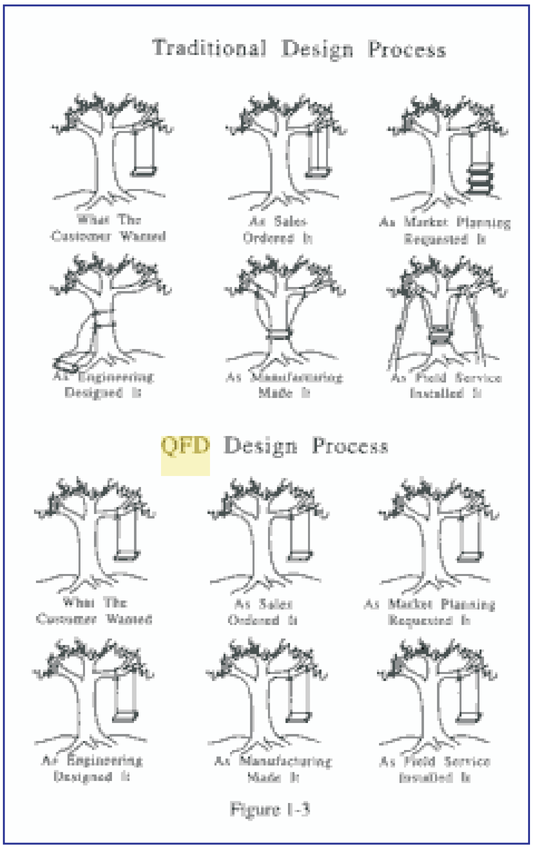


FIG. 17 QFD VS. TRADITIONAL DESIGN PROCESS IN A SWING

IMPORTANCE OF PLANNING ACTIVITY

Often it is said that well begun is half done. In our view, well begun is almost fully done.

Here also we have the frequently quoted example of the Japanese way of creating new products which gives most importance to planning – in fact, 80% of the time could be spent on planning and only balance 20% of the time in actual manufacture of the prototype with first time right and minimal testing.

On the other hand, the earlier American way was to plan only for 20% of the time, then produce a sub-standard prototype which has to be honed into a good finished product after several prolonged rounds of testing.

INNOVATION

The success of TQM depends considerably on innovation.

For innovation, one must not be a conformist and must **break all the rules**.

One must question everything-often things are done in an organization as that is how it was done earlier.

There could be many wasteful activities which could be cut down and there may be many different better ways of performing the same activity.

It must be clearly understood that a total conformist cannot innovate.

After all it was the Buddha's apparent irreverence to the then customary rituals which led him on to Nirvana.

CASE STUDIES IN A PISTON RING INDUSTRY

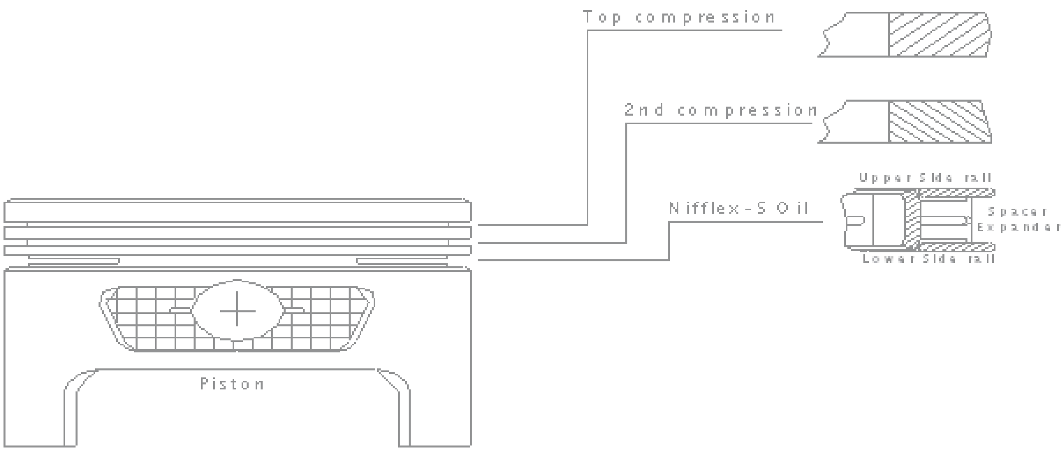


FIG. 18 PISTON RINGS ASSEMBLY

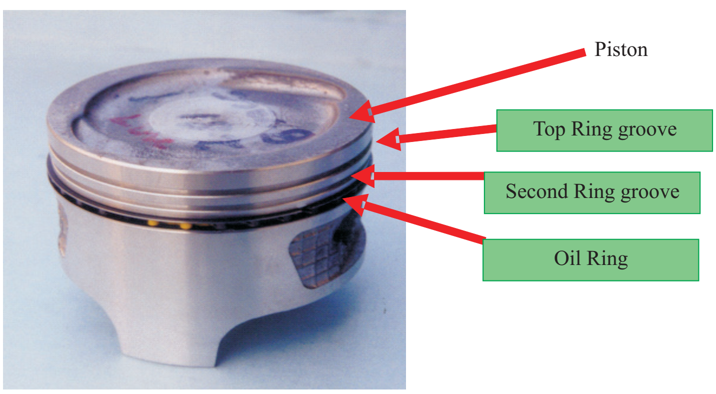


FIG. 19 PISTON RINGS LOCATION

1. Strategic Decisions

e.g. machinery procurement

Strategic decisions on procurement of machinery for specific processes have to be done with great care. For example, a thermally sprayed ring outer periphery needs to be ground.

For this the obvious method is to traverse grind the outer periphery of the rings mounted on the mandrel. Hence initially a traverse grinding machine was procured, with caution.

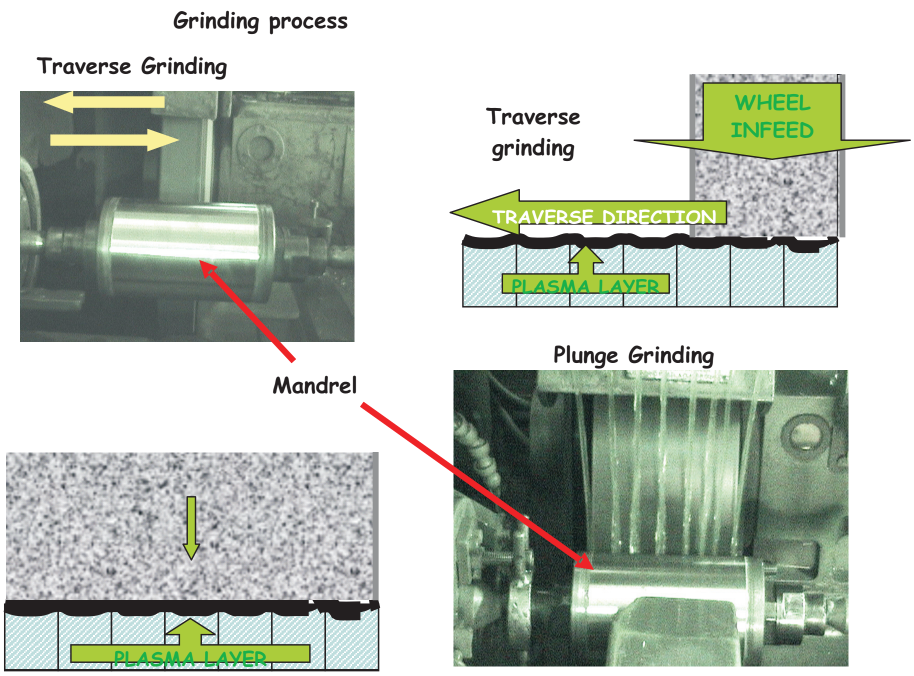


FIG. 20 GRINDING PROCESS

However after several trials, it was found that this traverse grinding was in effect disturbing the bond between the substrate and the coating because of the very nature of the operation tending to disturb the bond by generation of interface tensile stresses.

Hence it was decided to try plunge grinding instead as this will involve only compressive stresses at the critical interface between the substrate and the coating.

Accordingly a plunge grinding machine was procured at a cost of Rs. 60 lakhs and it was found to be superior to traverse grinding.

Now the question is – what if we had blindly procured several traverse grinding machines?

Then, even if we had come to the understanding that plunge grinding is better than traverse grinding, we could not have implemented plunge grinding as that would have meant increased cost, and we could not have given the better product to the customer.

2. Process Sequence

The order of operations or processes can also be an important factor in obtaining a good reliable product.

For example, in an internal stepped cut (ISC) piston ring, it is essential that the ISC operation is done finally after OD lapping and not vice versa.

The ISC operation actually provides the twist required in the ring during engine running (to control oil consumption and blowby).

The OD lapping operation ensures a perfect round ring with full contact all around the periphery. Hence if lapping is done after ISC, it will nullify the effect of the ISC operation.

Outwardly the piston ring will look the same, but there will be a considerable difference in its performance.

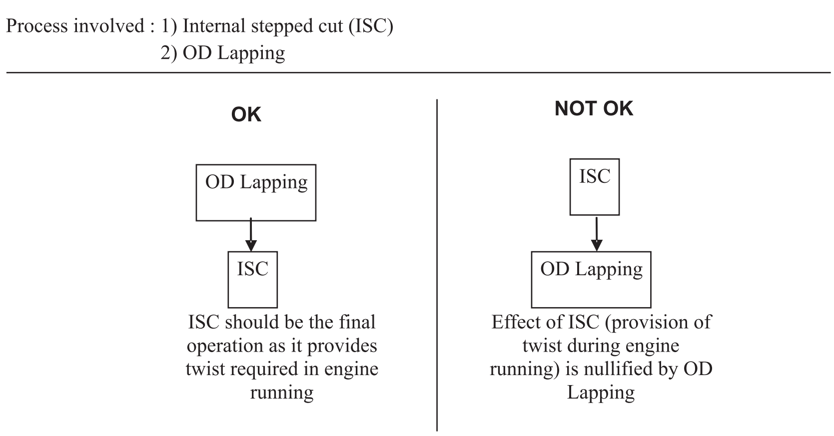


FIG. 21 PROCESS SEQUENCE

3. Effect of Minor Dimensions

In our specific product of piston ring, every taper or chamfer of a few minutes or microns respectively could be very important or equally very wasteful or even counter-productive .

A matter of 0.05mm additional chamfer in a critical location can make the piston ring completely unacceptable with reference to its key parameters of performance, namely oil consumption and blowby.

Table 8

Parameter	*Customer spec*	*IPR Performance*
O/C – gms/h	< 20	< 10
Blowby - Lpm	< 14	18 -high

Competitor ring test result (Tested at our end for 100 Hrs):

Avg O/C : 6 Gm/h

Avg Blowby : 9 – 10 Lpm

Engine dismantled and competitor rings were critically examined for dimensional aspects. This revealed following differences in top ring alone.

Table 9

Parameter	*Competitor*	*IPR*
Bottom OD Corner Radius - mm	0.10 – 0.15	0.10 - 0.20
Gap end chamfer - mm	0.03 – 0.08	0.15 – 0.25

It was decided to change one parameter at a time and carry out trials on the same cylinder block and piston

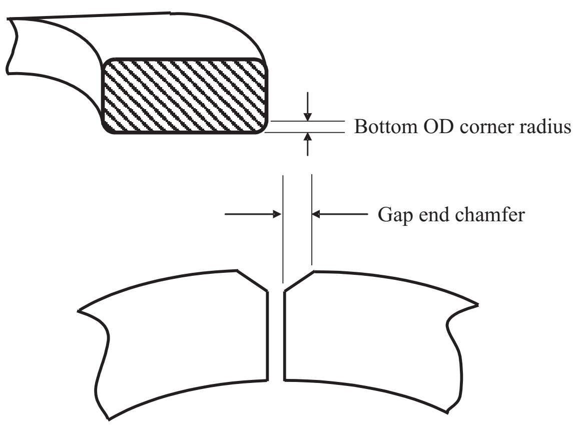


FIG. 22 RING BOTTOM OD CORNER & GAP END CHAMFER

Test results at IPRL R&D

Table 10

Trial	*Bottom Radius-mm*	*Gap End Chamfer – mm*	*O/C – Gms/h*	*BBY - Lpm*	*Remarks*
1	0.20	0.15~0.20	8.6	16	Blowby level is out of spec.
2	0.15	0.15~0.20	8.9	17	Blowby level is out of spec.
3	0.10	0.15~0.20	9	16	Blowby level is out of spec.
4	0.15	0.10 max.	8	11	Comparable to competitor performance.

4. Importance of Closed Gap

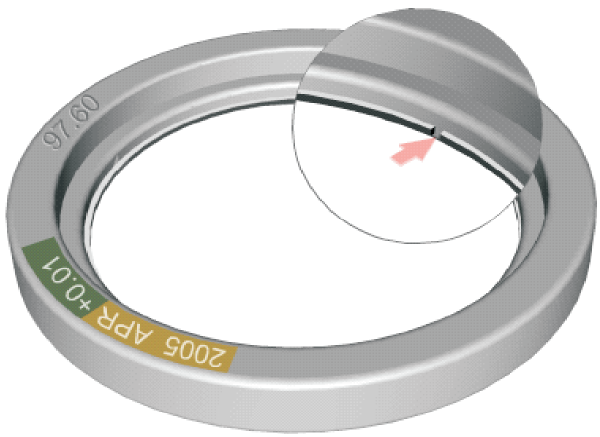


FIG. 23 CLOSED GAP IN PISTON RING

Controlled specification tolerances: example top ring closed gap

Table 11

Actual Range of closed gap, mm	Tolerance of closed gap, mm	Remarks
0.20 ~ 0.40	0.20	Common specification, results in blowby of 45 lpm
0.20 ~ 0.30	0.10	Closer toleranced specification demanded by customer, for a reduced blowby of 37 lpm

For closer toleranced specification

Process capability is to be improved by modifying the process (e.g. changeover from dry gap sizing to wet gap sizing) or by modifying the toolings (e.g. Shape and material and rigidity of grinding wheel).

5. Process Knowledge and Component Performance Knowledge

Example: top ring closed gap and second ring closed gap interactions.

It is advantageous to have lower closed gap to control blowby.

Hence R&D efforts were made to improve wear resistance of top ring coating so that wear rate is slower. These are described below.

During the course of development or improvement in the plasma coating, the wear loss has decreased by 94% from a level of 1.10mg to 0.18mg. The twist fatigue life (which is a measure of the bonding of coating and substrate) has increased by a factor of 96 times from a level of 250

Table 12

	Top Ring closed gap, mm	*Second Ring closed gap, mm*	*Remarks*
During initial running in engine	0.20/0.40	0.40/0.60	Top ring wears faster due to closeness to combustion zone. Pressure around top ring is 10 times as much as pressure around second ring.
After 1 lakh kms running in engine	0.50/0.70	0.48/0.68	Closed gap of both rings are somewhat close to each other, but blowby will be higher.

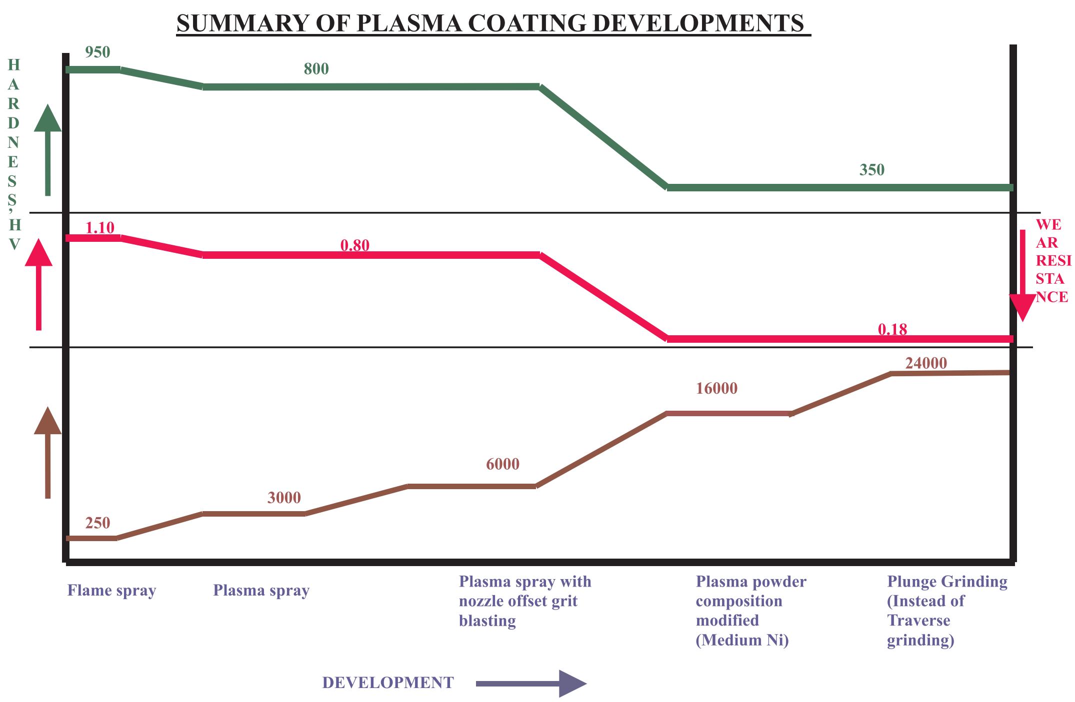


FIG. 24 SUMMARY OF PLASMA COATING DEVELOPMENTS

cycles to 24000 cycles. During the same period, the hardness has dropped from a high of 950 HV to a low of 350 HV. Thus even though the hardness has dropped, the wear resistance has considerably increased (instead of decreasing, possibly because it is a sprayed material and not a compact material).

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

"WORLD CLASS MANUFACTURING" [WCM] THROUGH STATE OF ART TECHNOLOGIES, TOOLS, SYSTEMS & BEST PRACTICES

S.R. SEETHARAM*

INTRODUCTION

This article highlights the importance of using the state of art technologies along with some of the tools, techniques, systems that are practiced by successful manufacturing companies in order to position themselves as world class manufacturers.

WHAT NECESSITATES WCM?

- Globalization relates to the ability to develop and produce products for both domestic and international markets and offers a unique challenge to manufacturing firms.
- Business firms compete with one another in different ways, namely price, quality, product differentiation, dependability, flexibility, time to market, customer service, productivity, managerial competence etc. For this, successful companies have adopted highly efficient operations, which enabled them to produce high quality products, apart from being competitive. Competitiveness, therefore, is in meeting all the requirements of the international markets, while maintaining the profitability.
- The world class companies were able to produce innovative products of high quality and develop the capability to bring new products to the market at faster speed and established newer benchmarks. Hence the central focus should be in the formulation of long term strategies to achieve the desired core / distinctive competencies.
- Porter classifies these competitive business strategies as:

*Formerly Executive Director-BEML, Distinguished Visiting Professor- AICTE-INAE Scheme
Professor- Department of Management studies-JSS Academy of Technical Education-Bangalore

1) Overall Cost leadership strategy
2) Differentiation strategy and
3) Focus or Market segmentation strategy

- Five basic competitive priorities that supports the above said generic strategies are:
  1) Cost [Low cost-Competency]
  2) Quality [High Performance design, Consistent Quality]
  3) Delivery [Rapid Delivery]
  4) Flexibility [In variety & volume]
  5) Service [Superior customer service]
- To establish manufacturing competitiveness, capability and competencies, WCM's should master many programs in their operational strategies. To name a few:
  1) Improving product design and manufacturability through CAD & CAM,
  2) Adopting Group Technology concept,
  3) Encouraging Quality Circle to solve problems & employee participation
  4) Evolving Zero Defect Programs,
  5) Continuous Improvements as way of life in production and operations
  6) Just-in-time [JIT] for inventory and cost controls
  7) TQM programs, ISO-9000-Quality management systems for building Quality
  8) Business Process Re-engineering, Concurrent Engineering etc
  9) Bringing automated operations, wherever cost economics workout better
  10) Reduction in direct labor content
  11) Using state of art technologies like Flexible Manufacturing System [FMS]
  12) Practice "Total productive Maintenance" to improve process capability
  13) Training and enlarging worker's participation by empowerment etc.
  14) Responding quickly to changes in demand by adopting suitable production-technology, Improvements in process, Quality and in communication.
  15) Differentiating their strengths in knowledge, process superiority & innovation
- Some of the above said Tools, Techniques, Philosophies, Systems and the best Practices, that are basically used for WCM, among many, are briefly discussed here.

I. LEAN PRODUCTION SYSTEM

- This term was used first by Womach, Jones and Roos in their book "The Machine that Changed the World" after their study in automobile industries. This is a new kind of production system adopted by leading automobile industries, where in, the process is designed to minimize their inherent wastes of time, materials and money.
- Lean production programs are those that use minimal resources to produce high volume, high quality and variety products. It implies less space, less inventory and higher productivity with few workers to produce the desired output. Hence the lean is about getting

the right things, to the right place, at the right time, in the right quantity while minimizing waste and being flexible and open to change.

- Lean manufacturing is a management philosophy that focuses on the reduction of wastes to improve overall customer value in the following areas:

  1) Transportation 2) Inventory 3) Motion 4) Waiting time 5) Over-production
  6) Processing Itself 7) Defective Products

- Toyota defined seven categories of waste. In fact there are now nine 'deadly wastes':
  1) Overproduction (making more than needed, or earlier than needed)
  2) Transportation (moving products farther than is minimally required)
  3) Waiting (products waiting at next production step/people waiting for work)
  4) Inventory (Excess Inventory are Deadliest type of waste)
  5) Motion (people moving more than minimally required)
  6) Processing (processes that are not linked to upstream or downstream)
  7) Defects (the effort involved in inspecting for and fixing defects)
  8) Safety (unsafe work areas creates lost work hours and expenses)
  9) Information (too much of data or too many than required)
- Therefore, the lean manufacturing demands:
  1) Small lot production just adequate to meet the customer demands
  2) Less setup time in between batches
  3) Perfectly maintaining machine performance
  4) Use of pull production system for speedy delivery
  5) Use of group technology for high volume production,
  6) Focusing on improvements which reduces time, money, materials etc
  7) Using the state of art technology, as feasible & appropriate,
  8) Using or improving the process / product layout production
  9) Built-in flexibility, variety and high quality of goods into the market
  10) Production systems, which caters to the varying demands etc.

  - **KEY PRINCIPLES INCLUDE:**
    - Pull processing – products are pulled from the consumer end (demand)
    - Perfect first-time quality – quest for zero defects
    - Waste minimization – elimination of activities that do not add value
    - Continuous improvement – reducing costs, improve quality, productivity etc
    - Flexibility – producing different mixes or greater diversity of products quickly,
    - Building and maintaining a long term relationship with suppliers through collaborative risk sharing, cost sharing and information sharing
- Lean implementation program is inculcated by many manufacturing units after the same is approved by the senior management and set of objectives of the company is firmed up after a thorough brainstorming process. Lean Implementation team normally consisting of 5 to 7

members, select a pilot project and run for a short period and then measure, evaluate and review the results. Once it is successful, others are trained in the system, etc and make it habitual to run many such projects. While selecting the project, the one's that gives the biggest return is prioritized.

- Now a days Companies are looking beyond the shop-floor to find opportunities for improving overall company's cost and performance to compete & remain world class.

II. JIT [JUST-IN-TIME] PHILOSOPHY

- JIT concept of production was first introduced in Japan under the name Kanban, where in materials and components are supplied to the workstation just at the time that is required for use. JIT is also referred to as **'LEAN PRODUCTION'**.
- JIT & TQM are management philosophies through which Co' s focuses on:
  1) Continuous improvement of processes by identifying and eliminating wastes
  2) Understanding the needs and wants of customers and supply the products
  3) Increased involvement of people in identifying & eliminating obstructions
  4) Problem analysis and solving & Implementing improvements in operations
  5) Empower to take and carry out decisions in their area of operations
  6) Reducing the inventory
  7) Identify the hindrance factors in the flow of speedy production.
- JIT is a logistic approach for keeping inventory to bare minimum as they arrive when just needed. JIT uses many types of systems/processes namely 1) Kanban System or Pull scheduling 2) Kaizen system 3) Reduction process in setups 4) Lean Production system 5) Poka-Yoke [Fool Proofing] 6) In-built Quality, Standardization & Simplification 7) Partnership with vendors 8) Economical processing.
- **JIT IS DEMANDED WHEN:**
  1) Every time the production is changed from one item to another, there will be expected delays, disturbances and costs, as these may waste resources.
  2) A stable environment for use of specialized equipment.
  3) JIT works best with high volume, mass production operations.
  4) Production system must give a smooth flow of products through the process.
  5) Deliveries of materials are made just at the time when they are needed.
  6) Lead-time as far as possible must be shorter.
- **ELEMENTS IN JIT OPERATIONS:** [Q=Quality related elements]
  1) Standard product with few variations (Q)
  2) Automated high-volume operations & continuous production(Q)
  3) A balance process that fully utilizes resources
  4) Reliable production equipment- (Q)
  5) Minimum stocks & Short lead times for materials & Reliable suppliers

6) Low set up and delivery cost
7) Efficient materials handling- (Q)
8) Flexible workforce

- **BENEFITS OF JIT**
  1) Lower stocks of raw materials and work in progress
  2) Shorter lead time needed to make a product
  3) Higher equipment utilization and higher productivity
  4) Simplified planning and scheduling
  5) Improved quality of materials and products
  6) Less scrap and wastage
  7) Better morale in the workforce
  8) Better relations with suppliers
- **SPECIFIC PROBLEMS / DISADVANTAGES IDENTIFIED BY JIT:**
  1) Initial higher investment and cost of implementation
  2) Reliance on perfect quality materials from suppliers & their adoption to JIT
  3) Need for stable production when demands from customers may vary
  4) Reduction in flexibility to change products
  5) Difficulty of reducing set-up times
  6) Need to change layout of facilities

III. TOTAL QUALITY MANAGEMENT [TQM]-PHILOSOPHY

- For lean manufacturing and JIT practices, the TQM or Company wide Quality Control [CWQC] should accomplish as a philosophy. As JIT has the objectives of producing products for fast delivery and overall customer satisfaction, the objective of TQM is to produce products of perfect quality
- TQM is a management approach, centered on quality, based on employee's total participation and aims at long-term success through customer satisfaction. It is a philosophy that helps change the attitude and culture within the organization to achieve the continuous improvement of products
- TQM requires that the company maintain high <u>quality standard</u> in all aspects of its business, i.e. things are done right the first time and that defects and waste are eliminated from operations. Improved Quality through TQM will result in increased profits thro' sales gains improved response, higher prices, improved reputation, cost reduction thro' increased productivity, lower reworks, lower warranty costs etc.
- TQM has been widely used in manufacturing, education, Government and services
- **TQM FOCUSES ON:**
  (1) Customer satisfaction [a) fitness for use b) reliability, life aspect of quality, c) value of money d) after sales service e) packaging, f) speed of service g) maintainability h) customer confidence in the Co

(2) **Leadership** [1) Identify and encourage employee potential b) accept the responsibility c) be a role model d) remove all the road blocks

(3) **Quality policy** a) build enthusiasm and loyalty of people b) serve as a standard for evaluating the performance c) support missions d) encourage innovation e) promote teamwork & leadership f) serves the objectives of customer satisfaction

(4) **Organization structure** [structure helps to harness the potential at each level]

(5) **Employee involvement** [helps change the work culture]

(6) **Quality cost** [measures of all costs associated with achievement of quality]

(7) **Supplier selection & development** [Zero defect concepts at supplier's end- buying from qualified vendors, vendor rating, analysis on supplies and educating]

(8) **Recognition & rewards** a) to show the company's appreciation for better performance b) to improve morale c) to reinforce the behavioral patterns d) to create satisfied workforce and e) to create highly motivated.

IV. ISO 9000- A QUALITY MANAGEMENT SYSTEM

- **NEED OF THE HOUR**: Customer needs are becoming more & more stringent. For global market, the uniformity is through product standardization. As the Product certification involves post production inspection for conformance, a well administered QMS certification can get a company the required international recognition & acceptance as it ensures that non conforming products do not come out.
- ISO (the international organization for standardization) is a worldwide federation of national standards bodies (ISO member bodies) and its mission is to provide international standardization to facilitate worldwide exchange of goods and services.
- ISO 9000 is a series of generic standards for building, operating & document a QMS-quality management system . This series of standards is being applied across manufacturers of various domain , banks, legal firms, health, service providers, educational institutions, chemical processing plants etc
- ISO 9000 suggests that supplier quality will improve with the consistent use of a defined quality system. Its objectives "what is defined/said is what is practiced"
- A QMS is what ISO 9000 is defined: " the organizational structure , responsibilities, procedures, processes and resources for implementing quality management necessary to achieve the quality objectives stated in the quality policy ".
- TQM is about continuous improvement and hence QMS that is compliant with ISO-9000 is viewed as enabling TQM
- **OBJECTIVES OF ISO 9000 QUALITY SYSTEM**
  - An organization should achieve and sustain the quality of the products produced so as to continuously meet purchaser's stated or implied needs.
  - An organization should provide confidence to its own management that the intended quality is being achieved and sustained. An organization should also provide confidence to the purchaser that the intended quality is being achieved in the delivered product

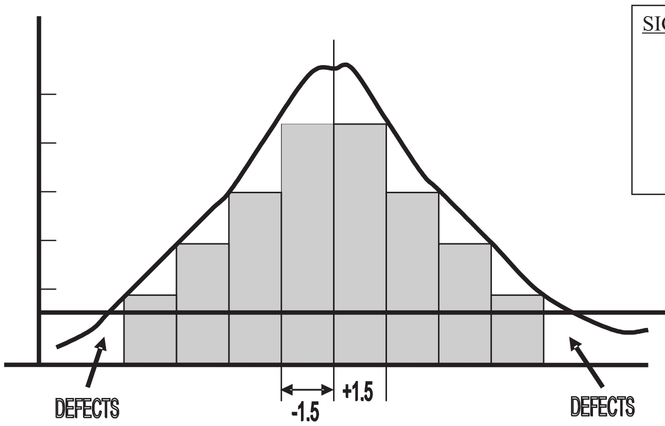
- **QUALITY MANAGEMENT IS BUILT AROUND EIGHT PRINCIPLES**:
  1). Customer Focus 2) Leadership 3) Involvement of People 4) Process Approach 5) System Approach to Management 6) Continual Improvement 7) Factual Approach to Decision Making 8) Mutually Beneficial Supplier Relationships
- **REQUIREMENTS TO ACHIEVE THE SAID OBJECTIVES:** The Implementation of a formal quality system with proper documentation is very demanding on time & efforts. To make the system cost effective there is a need to change attitude of people to accept the new working practices and make the standards user friendly.

V. SIX SIGMA SYSTEM

- **SIGMA** (the lower-case Greek letter ó) is used to represent standard deviation (a measure of variation) of a population.
- The term "six sigma process" comes from the notion that if one has six standard deviations between the mean of a process and the nearest specification limit and there will be practically no items that exceed the specifications. This is the basis of the process capability studies.
- **Six Sigma** is a vision of quality, which equates with only 3.4 defects per million opportunities [DPMO] for each product transaction & strives for perfection.
- For world class manufacturing, companies have to ensure a quality level of five to six Sigma as detailed below. Also indicated the cost of quality for each sigma.

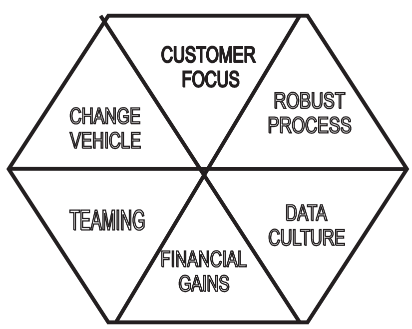
Sigma	*Defect-rate [ppm]*	*Cost of quality*		*Competitive level*
6	3.4	< 10%		World class
5	233	10-15%		
4	6210	15-20%		Industry average
3	66807	20-30%		
2	308537	30-40%		Non competitive
1	690000	>40%		

- **SIX SIGMA DIAGRAM**:
- In a Capability Study, sigma refers to the number of standard deviations between the process mean and the nearest specification limit-upper or Lower
- It is the best measurement of process variability, because smaller the deviation-value, lesser the variability in the process. For + or – Six Sigma, 99.9999998% of the product or service will be between specification and the non-conformance will be only 0.002 parts per million.
- Within the range of -1S and +1S, i.e. above and below the target, 99.99996600% of the product or service will be between specifications and non-conformance rate will be 3.4 PPM. This off-center situation gives process capability index Cpk=1.5 with 1 being the de-facto standard



SIGMA	DEFECTS [PPM]
1	1697692
2	2308537
3	366810
4	46210
5	5233
6	63.4

- Six sigma is a statistical term and derived from parameters related to the process capability, which is = [Upper specification limit-Lower specification limit] / [Six Sigma]
- Process, that is normally distributed as depicted in the sketch above where process is centered, will have 3.4 parts per million beyond a point that is 4.5 standard deviations above or below the mean (one-sided Capability Study). This implies that 3.4 DPMO corresponds to 4.5 Sigma's, not six as the name would imply.
- When the specification limits are 6S from the target, and sigma (the standard deviation) value is low enough to ensure that the probability of the process producing outside the spec limit is only 3.4 parts per million for perfect six sigma level
- **TRANFORMATIONS DUE TO SIX SIGMA:** Important areas of transformation in a company who is practicing Six Sigma are detailed in the sketch below. These transformations will lead the organizations to grow in their stature and in the process.



- The key methodology adopted is DMAIC (Define, Measure, Analyze, Improve and Control) is used to improve an existing business process -a process for continued improvement. It is systematic, scientific and fact based. This closed-loop process eliminates unproductive steps, often focuses on new measurements, and applies technology for improvement.

- **DFSS** (Design for Six Sigma) - A systematic methodology utilizing tools, training and measurements to enable us to design products and processes that meets the customer expectations and produced at Six Sigma Quality levels.
- **MOTOROLA EXPERIENCE:**
  - The term Six Sigma is a system of practices developed by Motorola to systematically improve processes by eliminating defect. Defects are defined as units that are not members of the intended population.
  - The process was pioneered by Bill Smith at Motorola in 1986 and was defined as a metric for measuring defects and improving quality, and a methodology to reduce defect levels below 3.4 defects per (one) DPMO.
  - Six Sigma is a registered service mark and trademark of Motorola
  - Motorola has reported over US$17 billion in savings from Six Sigma in 2006, by practicing it as a "Six Sigma-movement". This was started with an aim of achieving near zero-defect level of manufacturing competence. They defined the zero-defect as 99.999% defect free manufacturing capability, which was translated to DPMO at six different sigma-levels. At a perfect six-sigma level, this meant DPMO of 3.4, as low figure compared to what was prevalent in manufacturing. Second phase movement was called 'Second Generation' quality movement, Objectives of which is to move closer to zero-defect and also to improve the level of customer satisfaction level
  - **OTHER SUCCESSFUL CO'S WITH SIX SIGMA PRACTICES**:

    a) General Electric, 3M, Bank of America, Caterpillars, Honeywell International, Boeing Company- to name a few of MNC's and Mumbai's Dabbawala's association [only organization in the world to achieve Six Sigma and yet have no written documentation]. In fact they are above the six sigma standard, because their DPMO is zero. In fact astandard better than six-sigma is required for them, as the recent survey showed only one mistake in every 6,000,000 deliveries]

VI. BUSINESS PROCESS REENGINEERING [BPR]

- BPR is a management approach of improving efficiency and effectiveness of the processes that exist within the organizations. It is also known as Business Process Redesign, Business Transformation, or Business Process Change Management.
- Reengineering was adopted at an accelerating pace by 1990's and as many as 65% of the Fortune 500 Co' s have initiated reengineering efforts.
- "BPR is the fundamental rethinking and radical redesign of business processes to achieve dramatic improvements in critical contemporary measures of performance, such as cost, quality, service, & speed."- Hammer & Champy.
- "BPR encompasses the envisioning of new work strategies, the actual process design activity, and the implementation of the change in all its complex techno-logical, human, and organizational dimensions"- Thomas. H. Davenport.

- "BPR seeks radical changes apart from continuous improvement. It escalates the efforts of JIT and TQM to make process orientation a strategic tool and a core competence of the organization. BPR concentrates on core business processes, and uses the specific techniques within the JIT and TQM "toolboxes" as enablers, while broadening the process vision."-Johansson
- Four major areas that are subjected to change in BPR are the organization, technology, strategy, and people. Human resources deals with aspects such as education, training, motivation and reward systems based on process ownership, customer focus, value-adding, cross-functionality etc
- In BPR, IT plays a major role as enabler of new forms of organizing and collaborating, apart from supporting the existing business functions.
- Methodology: 1) Envision new processes 2) Initiating change 3) Process diagnosis 4) Process redesign 5) Reconstruction. 6) Process monitoring 7) Structural orgn.
- **SOME SUCCESS STORIES :**
    - **General Motor Corporation:** By implementing BPR, GM could overcome the financial drawbacks due to stiff competition. By replacing numerous brands of desktop systems, network operating systems and application development tools etc to have a common communication system, GM have saved 10% to 25% on support costs, 3% to 5% on hardware, 40% to 60% on software licensing fees, and increased operational efficiency.
    - **American Airlines:** It helped them to get back on track from the bad debt that was affecting the business and its practices.
    - **DELL Incorporated,** the world's fastest growing major PC manufacturer, was successful in keeping the smallest inventory possible by having a direct link with the manufacturer. This reduced the cost for inventory tracking and massive warehouse maintenance.
    - **Ford** reengineered their business and manufacturing process for manufacturing quality cars, thus saving millions on recalls & warranty repairs.
    - **Procter and Gamble** which carries 300 brand products has met the challenge through rapid pace of innovations in digital technology etc. Due to Re-engineering, the pace of innovation has doubled in the last three years.
    - **DISADVANTAGES:** As BPR focus strictly on efficiency and technology, it carries the label that it was used for major workforce reductions. Lack of management support may result in poor acceptance. Exaggerated expectations of the potential benefits of BPR & consequently its failure may demoralize employees.

VII. BENCHMARKING [BM]-A PROCESS APPROACH

- BM is a systematic method by which organizations measures their own performance with the best practices within their sector.

- BM is defined as "measuring performance against the best-in-class companies, comparing the performance levels, use those information and decide on moving from where we are to where we want to be".
- As a result of the above measurement, company's are able to set new goals, adopt new and best practices and in the process achieve improvement in certain aspect of performance namely best quality, cost, best product, and satisfy the customers.
- BM or "Process benchmarking" is a process used in strategic management, in which organizations evaluate various aspects of their processes in relation to best practice, usually within their own sector.
- **ADVANTAGES OF BM:**
    - BM is a powerful management tool because it overcomes Blindness [thinking-that what they are doing is best, no other method can do]
    - "BM opens to new methods, ideas, tools to improve their effectiveness.
    - It helps crack through resistance to change by demonstrating other methods of solving problems than the one currently employed, and demonstrate.
- **PROCEDURE FOR BM:**
    - **Identify problem areas:** Because BM can be applied to any business process or function, a range of research techniques used include: informal conversations with customers, employees, or suppliers; marketing research, quantitative research, surveys, questionnaires, process mapping, quality control variance reports, or financial ratio analysis.
    - **Identify organizations that are leaders in these areas -** Look for the very best in any industry and in any country. Consult customers, suppliers, financial analysts, trade associations to determine company to be studied.
    - **Survey companies for measures and practices -** Companies target specific business processes using detailed surveys of measures and practices used to identify business process alternatives at leading companies.
    - **Visit the "best practice" companies to identify leading edge practices -** Companies typically agree to mutually exchange information beneficial to all parties in a benchmarking group and share the results within the group.
    - **Implement new and improved business practices -** Develop implementation plans which include identification of specific opportunities, funding the project and selling the ideas for gaining demonstrated value from the process.
- **XEROX & MOTOROLA EXPERIENCE:**
    - **XEROX** initiated the BM in the 1980's because of the fact that their market share of the products shrunk to 35%. They were pioneers in establishing sound BM processes & eventual winners of the Prestigious "Malcolm Baldrige National Quality award". Results: they could reduce the Quality problems by 65%, manufacturing cost by almost 50%, Development-time for new products by 60%, Labor costs-direct- by 50%, Corporate staff number by 35% Vendor list to 300 from 5000 and Concurrent

engineering was practiced and Commonality of parts increased to 20% through standardization.

- **MOTOROLA** had set an internal goal of achieving improvement in quality attributes by tenfold in five years. By BM and following the best practices with leading companies like Walmart, Domino, and other companies from Japan, they could achieve the same BM in three years.

VIII. TOTAL PRODUCTIVITY MAINTENANCE [TPM]

- TPM is the concept originated & developed by JIPM (Japan Institute of plant maintenance), Tokyo, since late sixties. TPM combines the American practice of TPM with Japanese total quality control (TQC) and total employee involvement(TEI)
- TPM paves way for an excellent planning, organizing, monitoring and controlling practices through a unique 8-pillars method. TPM is one of the most valuable strategies for those who want to be competitive & meet the World Class Competition
- **TPM BACKGROUND:**
  - TPM had its genesis in the Japanese car industry in the 1970's. It first evolved in Nippon Denso a major supplier of the Toyota car company
  - The new quality approach of "prevention at source" was translated to the maintenance environment through the concept of TPM.
  - TPM is a method designed to eliminate the losses caused by break-down of machines and equipments by identifying and attacking all causes of its breakdowns and system down time caused by such breakdowns.
  - Six types of losses are: 1) down time due to equipment failure, 2) down time for setups and adjustments, 3) speed losses due idling or minor stoppages 4) speed losses due to discrepancies between designed and actual speed of equipment 5) defect losses caused due to process defects that cause scrap and quality related problems 6) defects due to reduced yields in the time between machine start up and stable production
  - TPM is a means to achieve high level of productivity, efficiency, effectiveness with zero loss concept, through total participation of all employees with self- managing abilities in practice, to achieve total customer satisfaction

- **OBJECTIVES OF TPM:**
  - Restoring equipment to a like-new condition
  - To have operators involved in total maintenance activities
  - Improve maintenance efficiency and effectiveness
  - On job training of the labor to improve their job skills
  - Have a sound equipment maintenance management and prevent all sorts of breakdowns under reliability strategy of the company
  - Effective use of preventive and predictive maintenance technology

- Achieve maintenance prevention concept through planned maintenance, to have trouble free machines and equipment, producing defect free products for total customer satisfaction (in quality, cost, delivery and services)
- Achieve zero loss concepts (zero breakdown, defect, losses, waste and accidents) in all the resources, over the life cycle of a production system
- Achieve TPM with active participation & involvement in all levels
- "Value Added" activity that the equipment is contributing to your products.

- Achieving manufacturing excellence through TPM, means the elimination of micro level operational issues and achieve defects free operations, no / least breakdowns, accidents-free operations, increase plant efficiency, reduction in manufacturing costs, reduction in equipment life cycle cost, boosting morale of employees etc

IX. USE OF AUTOMATIONS IN PRODUCTION SYSTEMS

a) COMPUTER AIDED DESIGN (CAD)
b) COMPUTER AIDED MANUFACTURE (CAM)
c) FLEXIBLE MANUFACTURING SYSTEM [FMS]

a) & b) CAD & CAM

- Since late 1950's, the applications of computers and computing techniques in Engineering disciplines have increased dramatically, because computers have become larger in memory capacity and faster in processing speed.
- As a result, Computer Aided Design (CAD) and Computer Aided Manufacture (CAM) are adopted to create major improvements in Productivity and Quality.
- CAD/CAM can be defined as the use of computers to aid the design and manufacturing process and the technology represents an efficient, accurate and consistent method to design and manufacture high quality products.
- The main purpose of CAD is to produce a definition of the part or system to be manufactured in the form of a geometric database or a drawing from this database, which establishes the physical configuration of the part or system. The main purpose of CAM is to translate this definition to tangible hardware based on that database.
- In Cam, computers play many roles in manufacturing functions such as numerical control, process planning, robotic and factory management.
- The characteristics of man and computer and its application in the design of a CAD system is important, especially in the following fields:
    1) Design Construction Logics (method of constructing the design)
    2) Information Handling (storing and communication of design information)
    3) Modification (handling of errors and design changes)
    4) Analysis (examination of the design and factors influencing it)

- Hence, the computer has the main functions:
  1) Serve as an extension to the memory of the designer
  2) Enhance the analytical and logical power of the designer
  3) Relieve the designer from routine repetitious tasks
- Similarly the designer has the following activities:
  1) Control of the design process in information distribution
  2) Application of creativity, ingenuity and experience
  3) Organization of design information.
- CAD allows a user to interact with the computer through a graphic terminal to define a design configuration, analyze the structure and its mechanical behavior, perform kinetic study and model testing and automatically produce engineering drawings.
- Production people will make use of the geometric description provided by the CAD as a starting point in CAM to create NC program for machine tools, determine process plans for fabricating the assembly, instructs robots to handle tools and work pieces and schedule plant operation with a factory management system.
- **BENEFITS OF CAD AND CAM**
  - **CAD** has drastically reduced the number of steps involved in the design process of products, it is easier and less tedious to designer, it increased the work output, it saved enormous time between initial conception of an idea to final implementation, it provided the designer with a versatility tool to graphically manipulate and obtain deep insight into complex problems in the design so that he can take quick decision and arrive at an optimum solution, which leads to improvement in quality of the product and hence become more competitive.
  - **CAM** has greatly increased productivity in the shop floor, as It generates NC programs with fewer errors as they can be checked in the graphic display, Scheduling of components and tools through workstations is made easier, thus reducing the delivery time of products.
  - **FOR MANAGEMENT:** The combination of CAD and CAM can improve capacity planning, production scheduling, better inventory control and working capital, simplifies product forecasting, estimation and price fixing becomes simpler. The other main benefit is from the information stored in the computer instead of the paper work & transfer of data becomes quicker, more reliable and less redundant.
- **APPLICATIONS OF CAD**

  Interactive computer graphics has been widely applied in many areas of science and technology. Its application includes: 1) Study of molecular structures in chemistry. 2) Medical research 3) Animation 4) Aircraft flight simulation 5) Structural design in aircraft 6) Ship building 7) Automobile industries 8) Integrated circuits and printed circuit board design in the electronics industry 9) Town planning and architectural design 10) Pipe routing 11) Layout in chemical plant design 12) Mesh data preparation for infinite element analysis and draughting 13) Design of the integrated electronic circuits in electronic industries.

- **USE OF CAD AND CAM AS AN INTEGRATED SYSTEM:**

  For most manufacturers, the promise of increasing productivity in drafting and NC programming is just enough to go ahead with investment on CAD AND CAM systems. Errors in designing and manufacturing can be reduced in integrated CAD/CAM systems by passing data electronically between disciplines instead of recreating it from scratch, hand carrying it from person to person and machine to machine or copying from existing drawings.
- **BENEFITS ACCRUED BY USING INTEGRATED CAD & CAM SYSTEMS:**
  - Shorter Product Cycle
  - Earlier recovery of errors
  - Integration of design and analysis
  - Increased design productivity and shorter lead time
  - Better coordination of tools and fixture design
  - More efficient material handling
  - Reduced manufacturing set up times and lead times
  - Reduced manufacturing costs
  - Competitive pricing of products

c) FLEXIBLE MANUFACTURING SYSTEMS (FMS)

  - FMS is a system that combines micro-electronics and mechanical engineering to being economics of scale to batch work. A central on line computer controls the machine tools and other stations and the transfer of component and tooling.
  - An FMC/FMS is a manufacturing cell/system consisting of one or more CNC machines connected by an automatic material handling system, all operated under control of a central computer with other auxiliary subsystems like component load/unload station, automatic tool handling system, tool pre-setter, component measuring equipment, wash system etc.
- Flexibility is applied to the machine because of CNC control and flow of product from one to another is through flexible transport system.
- Although production volume and variety will greatly define the areas of application, other factors that affect and modify the system design are:
  1) Work piece accuracy and configuration
  2) Assembly in a common unit (to produce one-to-one relationships)
  3) Product life cycle
  4) Phasing upward mobility (phased implementation and adaptability to change)
  5) The classic conflict between productivity and flexibility can be resolved by the application of a mid volume, mid variety manufacturing system concept.

- **FLEXIBILITY HAS THREE LEVELS:**
  1) Complete flexibility: universal machinery, tooling, fixtures and processes)
  2) Process specialization: tooling, fixtures and processing to increase productivity.
  3) Machine specialization: is restrictive, as the machinery is tailored to its application like tooling and fixtures of the previous level.
- **CHARACTERISTIC FEATURES**
  1) Solves production problems of mid volume and mid variety parts.
  2) Designed to process several types in a defined mix.
  3) Equipped with flexible machine tools that are capable of processing a different parts with negligible tool change over time.
- **BENEFITS OF FMS**
  1) Flexible to change part variety
  2) Higher machine utilization and hence higher Productivity
  3) Balanced output
  4) Less rejection and High Product Quality
  5) Reduced work-in-process and inventory
  6) Better control over production
  7) Just in time manufacturing
  8) Minimally manned operation & Unmanned operation possible
  9) Easier to expand
  10) Reduced plant size
  11) Reduced set up time
  12) Quicker model change
  13) Quicker response to market changes
  14) Shorter delivery time
  15) Consistent accuracy
  16) A route to CIM

X. OTHER OPERATIONAL PROCESSES IN MANUFACTURING

There are many production improvement processes that are adopted by leading industries. The following three widely used as value added systems are discussed briefly

a) Kanban system
b) Kaizen processes in production
c) Supply Chain Management

a) & b) KANBAN AND KAIZEN SYSTEMS

Kanban system

- Japanese took the lead in practicing product and process innovations for improvements, like **"Kanban System"** of production control, **Kaizen** –the continuous improvement, Taguchi method of Quality control and market driven manufacturing.
- Basically production is classified into:
    - Produce on demand (Pull system)
    - Produce to forecasts of demand or make to stock (Push system)
- Kanban production system is a Japanese information system that "harmoniously" controls the production quantities in each process.
- Kanban- means cards or visual record in Japanese. A Kanban is a card of two types, a withdrawal Kanban and a production ordering Kanban. The withdrawing Kanban shows the quantity of items that the subsequent process should withdraw from the preceding and the production ordering Kanban shows the quantity that the preceding process should produce. All operations are controlled by Kanban, which 'pulls' materials through the process with MRP and others, & 'pushed' through the system.

Kaizen philosophy

- Kaizen is a Japanese word for the philosophy that defines management's role in continuously encouraging and implementing small improvement involving everyone.
- It is the process of continuous improvements in small increments that make the process more effective, efficient, under control and adaptable.
- Improvements are accomplished at little or no expense, without sophisticated techniques or expensive equipment. It focuses on simplification by breaking down complex processes into their sub-processes and the improving them.
- The kaizen improvement focuses on the use of:
    1) Value added and non-value added work activities.
    2) Seven classes of waste-over production, delay, transportation, processing, inventory, wasted motion, and defective parts.
    3) Principles of motion study and the use of cellular production technology.
    4) Principles of material handling and use of one-piece flow.
    5) Documentation of standard operating procedures.
    6) The five S's for workplace-Five Japanese words that mean proper arrangement (Seiko), orderliness (Seiton), Personal cleanliness (Seiketsu), cleanup (Seiso) and discipline (Shitsuke).
    7) Visual management by means of visual displays that everyone in the plant can use for better communications
    8) JIT principles to produce only the units in the right quantities, at the right time, and with the right resources.

9) Poke-Yoke to prevent or detect errors.
10) Team dynamics like problem solving, communication skills, conflict resolution.

- Kaizen relies heavily on a culture that encourages suggestions by operators who continually try to incrementally improve their job or process. This change results in a small improvement in quality but substantial improvement in operator satisfaction.

  [Ex: Toyota production system has pioneered the Kaizen philosophy and achieved remarkable improvements in their operations

c) SUPPLY CHAIN MANAGEMENT [SCM]

- SCM is a process adopted and practiced by successful companies for development and management of its total supply system, which includes all the activities of purchasing function and process of procurement.
- SCM consists of four phases of the generation of requirements, sourcing, pricing and post award processes. All these are interdependent as the activities of conducting the purchase and procurement functions, involvement of cross functional teams in selection and qualifying suppliers and developing close linkages with the selected suppliers/vendors in the value chain apart from control of quality and costs.
- World class SCM contributes in increasing sales by enhancing the quantity of the products, endures on-time performance, reduce time to market enabling technology absorption and supports sales and marketing in maximizing the firm's net revenue through price monitoring in the entire process.
- SCM will enable the integration process between production, marketing and finance departments and helps in developing the business by understanding the business environment apart from formalizing the market-driven plans for all the materials bought-out. Therefore, the major steps in a sound SCM are the development of a good cross functional teams, creating a good supply chain, developing the required supply network, harness a work culture of partnerships and alliances with suppliers and take decisions on cost and quality effective purchasing and procurements.

CONCLUSION

- Today's quest of all manufacturing industries to produce products by practicing better, faster, more agile, cost effective production, perhaps, is a result of the changes that happened in the auto industry. The US industries dominated the global market and became giants in world economy with mass production techniques. To meet the challenges thrown by US companies, Japanese firms started adopting and practicing in "Lean manufacturing systems" based on the JIT and Lean production. Toyota Motors gave a stiff competition to US automobiles giants like General Motors, Ford, and Chrysler and started monopolizing the global market.
- It was observed that successful companies were able to adopt highly efficient operations which resulted in producing high quality products at reasonable costs and could meet the competition thrown across the globe.

- These world class manufacturers started establishing new bench-marks in almost all the areas of production, quality, costs, customer satisfaction etc and established the required competitive weapons in the process.
- Some of the established best practices, systems, technologies as discussed above, has made the business units to compete with each other in price, quality, product differentiation, flexibility, time to market customer satisfaction, higher employee morale, productivity, managerial expertise etc.

REFERENCES

1) Elwood S. Buffa, "Modern Production/Operations Mnagament" 7th Edition, Wiley
2) Lee J. Krajawski, Lary P. Ritzman "Operations Management" Addison Wesley
3) Dr. K. Srdhara Bhat, "Total Quality Management" 1st Edition, Himalaya Publishing

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

OVERVIEW OF INDUSTRIAL PROCESS AUTOMATION

K.L.S. SHARMA*

ABSTRACT

The material presented in this article is meant to impart some basic concepts in the present day industrial process automation. Hence, there may be some deviations/differences in reality. This article introduces, in general, the current industrial process automation, physical processes and signals, instrumentation and signal interfacing, data acquisition and control unit, human machine interface, automation system structures, and supervisory control and data acquisition.

Key words: Automation, Process, Process control, DACU, Input/Output, Open and closed loop control, PLC, Controller, RTU, HMI, DCS, NCS, and SCADA.

1. INTRODUCTION

Over the years, the industrial process automation technology has moved from **pneumatic** to **analog** to **digital**, and in digital, from **proprietary** to **open**. Today, the industrial process automation technology is highly influenced by the advancement in **information, communication,** and **related** technologies. The future industrial process automation technology is moving towards a **distributed network** of **embedded** systems.

Before we discuss the subject, it is necessary to understand the meaning of the industrial process which we want to automate. The industrial processes are built to produce the products and the services of interest for use or consumption. An industrial process, unlike a natural process, is a man-made one, not a self regulating one, and need be continuously monitored and guided to achieve the desired results to ensure the **quality, consistency, and cost-effectiveness** of the service and the products. While doing this, the process counters the presence of external disturbances and the internal disturbances (effects arising out of its own dynamics) which can affect the desired or the expected results. Also, while performing, the process should not enter into unpleasant conditions such as oscillations leading to instability or sluggishness leading to slow response. Figure 1 illustrates a process without automation.

*Professor, International Institute of Information Technology, 26/C, Electronics City, Bangalore 560 100 (kls.sharma@iiitb.ac.in)

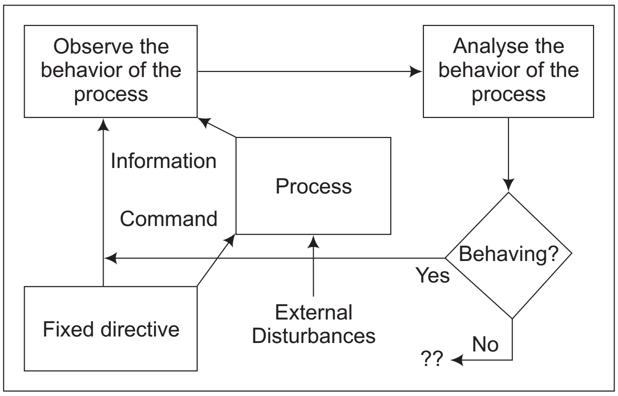


FIG. 1 PROCESS WITHOUT AUTOMATION

The above means that the process may not always follow the intended course to deliver the intended results. Further, the presence of the undesirable effects cannot be totally eliminated, but can be minimised to a large extent through automation.

In order to ensure that the process follows the directives to produce the desired results, it becomes necessary to continuously vary or adjust the inputs (directives) for not only to counter the effects external disturbances, but also to take care of the internal effects of process dynamics. Figure 2 illustrates the processes with automation.

2. NEED FOR PROCESS AUTOMATION

Today, apart from producing quality, consistent, and cost effective service and products, the industrial process automation has become necessary to manage:

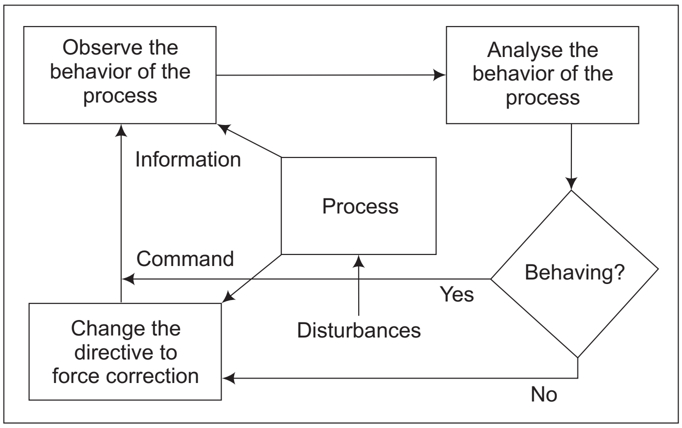


FIG. 2 PROCESS WITH AUTOMATION

a. Hazardous processes (nuclear, high voltage, toxic chemicals, etc.) where human operation is quite dangerous and is not desirable.
b. Repetitive operations (traffic management system, etc.), where continuous repetitive manual operations can lead to failure due human fatigue.
c. Sequential operations with interlocks (start up and shut down of complex plants, etc.), where many sequential operations, at every step, need to check for many conditions before continuing the operation, which is time consuming and prone to mistakes if done manually.
d. Complex operations (involving decisions based on heavy computing in very short time such as aircraft tracking, etc.) is not humanly possible.

Process automation can deliver the following benefits:

a. Fast response as there is no human angle involved.
b. Quicker decisions and faster restoration of plants after a break-down (restarting, reconfiguration, etc.).
c. Reduction of losses as the plant is brought on-line quickly.
d. Optimisation of resources as either total or substantial reduction in the dependence on high performance skills.
e. Higher safety of the personnel and the equipment.
f. Higher security, availability, reliability, and efficiency
g. Lesser overall operational costs.

3. INDUSTRIAL PROCESSES

In general, the industrial processes can be broadly divided into two categories, namely, the localised and the distributed, based on their nature. The localised processes are generally concentrated in a relatively small physical area while the distributed processes are spread over a large physical area (can be even geographical area). **The distributed process can also be defined as a network of localised processes distributed over a wide physical area**. This broad division is due to the philosophy of automation of these processes differing significantly, at least for now. However, it is difficult to strictly draw a line between these two as, due to various technical and commercial considerations, an automation philosophy generally meant for distributed process may be a better choice for a localised process and vice-versa.

Also, another way of categorising the processes is based on their areas of applications such as utility industries and process industries. Some typical examples in utility and process industry segments (both localised and distributed) are given in the Table 1.

4. AUTOMATION STEPS

After getting an introduction to the process, let us look into the details of an automation system. As seen earlier, the steps involved in a typical automation system are:

a. Observing the behaviour of the process by reading certain parameters from the process. This step is **Process Data Acquisition** and the parameters are **Process Inputs**.

Table 1. *Examples of processes*

Localised Processes	Utility industries	Power	Power plants, Transmission and Distribution substations
		Gas	Pumping, Repeating, and Distribution stations
		Water	Treatment and Distribution plants , Pumping, and Repeater stations
	Process industries	Oil and Gas	Oil wells, Refineries, Terminals, Pumping, Repeater stations
		Metal	Furnace, Mills
Distributed Processes	Utility industries	Power	Power transmission and distribution networks
		Gas	Gas transmission and distribution pipeline networks
		Water	Water transmission and distribution pipeline networks, Canal networks
	Process industries	Oil/Gas	Oil/gas transmission pipeline networks
		Metal	Slurry transmission pipeline networks

b. Analysing the behaviour of the process by studying the acquired inputs and deciding on new directives. This step is **Process Data Analysis** and **Control Generation**.
c. Controlling the behaviour of the process by sending the commands to the process to change its behaviour. This step is **Process Control** and the directives are **Process Outputs**.

The above inputs and outputs here are in relation to automation system and not the process.

5. PROCESS SIGNALS

Physically, these **Process Inputs and Process Outputs**, or **Process Signals** exist in one of the following three forms:

a. **Continuous (Analog):** The signal amplitude varies continuously with time
b. **Discrete (Digital):** The signal state changes discretely with time
c. **Fluctuating (Pulse):** A variation of discrete (digital) signal changing its state more frequently with time. The change of state may be periodic or non-periodic.

The above process signals can be in any physical form. Table 2 gives examples of some typical process signals in both utility and process industries (localised and distributed). To be in line with current technology, we, henceforth, use the words **analog, digital,** and **pulse** to represent the process signals.

The basic automation systems consists of **Instrumentation** and **Control** systems as explained below:

6. INSTRUMENTATION SYSTEM

Table 2. *Examples of input and output signals*

Process data acquisition (input signals)	Continuous (Analog)	Current, Voltage, Power (Active, Reactive, and Apparent), Power factor, Frequency, Speed, Pressure, Temperature, Flow, Level, Density
	Discrete (Digital)	Status of switchgears and Alarms, Position of Transformer taps and Limit Switches, Status of Valves, Pumps, Motors
	Fluctuating (Pulse)	Consumption of Energy, Water, Gas, Steam
Process control (Output signals)	Continuous (Analog)	Current/Voltage for Set Point Control
	Discrete (Digital)	Close/trip switchgears, Switch on/off devices, Start/stop devices
	Fluctuating (Pulse)	Increase/decrease speed of stepper motors

The present day automation systems are electronics based while the physical process signals (both inputs and outputs) are generally in other forms such as, electrical, mechanical, pneumatic, hydraulic, etc. These are not compatible with the electronic systems. In view of this:

a. Physical input signals from the process need be converted into appropriate electronic signals, **without any loss of information**, acceptable to the control system.

b. In the reverse direction, the electronic output signals (directives or commands) generated by the control system need be converted into appropriate physical signals, **without any loss of information**, acceptable to the process devices.

c. Also, sometimes, there is a need to isolate and protect the automation system from the process (and vice-versa) for the safety of the automation equipment, operating personnel, and the process plants in case of hazards.

The above interface is called **Instrumentation**.

Figure 3a illustrates the mechanism involved in conventional or hardwired instrumentation for analog signals. Here, during data acquisition, the physical input signal is first converted by a sensor into an electrical signal (transduction process) equivalent to the **value** of the physical signal. The converted electrical signal, then, passes through a signal conditioner (to prepare the same acceptable to the next stage) and a transducer to get finally converted into **compatible electronic signal** representing the **value** of the original physical signal.

In the reverse direction, the automation system generated electronic output signal passes through an actuator, signal conditioner, and the final control element to generate a **compatible physical signal** representing the **value** of the original electronic signal.

Figure 3b illustrates the mechanism involved in conventional or hardwired instrumentation for discrete or fluctuating (digital or pulse) signals. Here, during data acquisition, the physical input signal is first converted by a sensor into an electrical signal representing the **state** of the physical signal. The converted electrical signal, then, passes through a level converter to get finally converted into **compatible electronic signal** representing the **state** of the original physical signal.

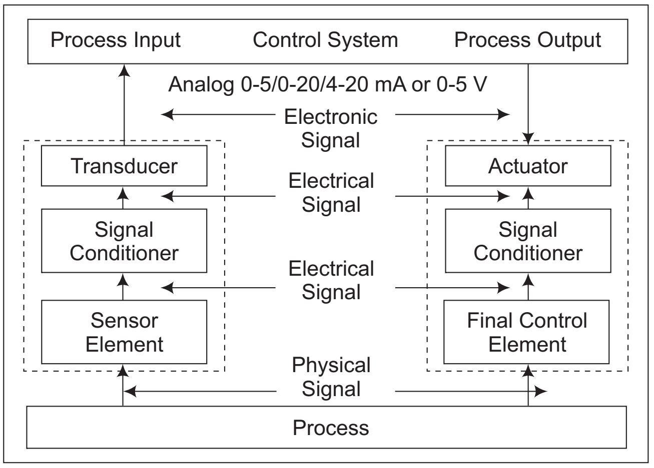


FIG. 3a INSTRUMENTATION FOR ANALOG SIGNALS

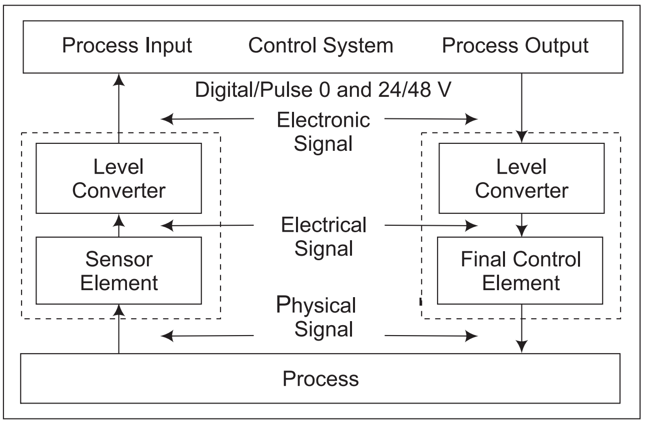


FIG. 3b INSTRUMENTATION FOR DIGITAL/PULSE SIGNALS

In the reverse direction, the automation system generated electronic output signal passes through a level converter, and final control element to generate a **compatible physical signal** representing the **state** of the original electronic signal.

Figure 3c summarises the signal flow between the process and the control system in both the directions, without any loss of information even though the same goes through many forms.

While there are specific industry standards for various types of physical signals, the industry standards for input and output electronic signals are:

a. Analog: 0 to 1/5/10V DC or 0 to 5mA/10mA/20mA DC, or 4 to 20mA DC
b. Digital/Pulse: 0 and 24 V DC or 0 and 48V DC

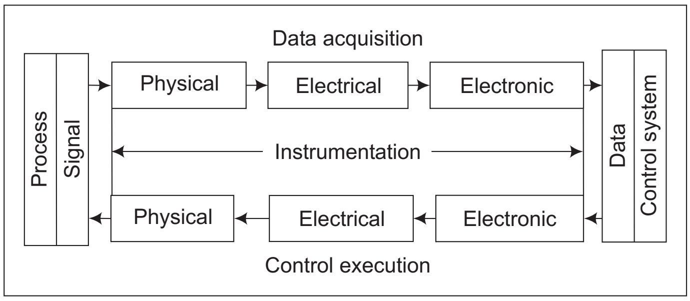


FIG. 3c INFORMATION FLOW BETWEEN THE PROCESS AND CONTROL SYSTEM

The above standards provide the **compatibility, inter-changeability and inter-operability** among the products by control system and instrumentation vendors.

7. CONTROL SYSTEM

In order to facilitate easy handling of input and output signals, the control system further divides its data acquisition and control hardware into the following functional modules:

a. Digital input module and Digital output module
b. Analog input module and Analog output module
c. Pulse input module and Pulse output module

Each module, normally, is designed to support one of 1, 2, 4, 8, 16, or 32 signals. With this modular structure, the control systems can be configured or built to meet the specific process input and output requirement through addition of required modules. Further, the present day electronics facilitates the digital input and output modules to handle pulse inputs and outputs also unless the pulse rates are too high calling for special purpose high speed pulse input and pulse output modules.

As the present day control systems are predominantly computer/microprocessor based, they cannot handle the analog signals unless converted into a machine readable digital values. In view of this, analog input module converts the electronic analog signal from the process into its digital representation (using **Analog** to **Digital Converter-ADC**) without any loss of information which can be read and manipulated by the control system. Similarly, in the reverse direction, analog output module converts the digital representation (using **Digital** to **Analog Converter–DAC**) of the analog value generated by the control system into its representative electronic signal for passing on to the process without any loss of information.

However, there is no need of such conversions in case of digital/pulse input signals as they are already in the digital form and requires to be just transformed to the voltage levels acceptable to the control system. Similarly, the control system generated digital/pulse output signals need no conversion except transformed to the voltage levels acceptable to the instrumentation.

The control system with data acquisition by input modules and control execution by output modules acts as a front-end to the process (via instrumentation) and is called **Data Acquisition** and

Control Unit (DACU).

DACU, also called as **Process Computer,** is primarily a microprocessor or a computer based equipment for performing the data acquisition and control functions as per the strategy programmed and stored in its memory. The various modular components of DACU are:

a. **Rack** – A mechanical structure to physically house all the electronic functional modules
b. **Bus** – Receives all the functional modules (via the connectors) electronically to provide the physical connectivity/path for communication (bi-directional) and supplies power to the functional modules
c. **Power supply**: Receives the power from outside source (mains, charger, UPS, etc.) and provides the power to all the modules in the rack via the bus at the required voltage levels
d. **CPU/Memory**: This is the heart of DACU and performs the **Data Acquisition** and **Control** functions programmed and stored in the memory.
e. **Input/Output (I/O)**: Functional electronic modules to receive input signals from and drive output signals to the process (via instrumentation).

Figure 4 illustrates the block schematic of general purpose DACU. In the above arrangement, all the functional modules communicate with CPU over the bus. The communication is parallel. All I/O modules can communicate only with CPU and not among themselves. This DACU is **Standalone** as the same cannot communicate with external devices/systems.

If we remove the output modules from DACU, the same becomes only **Data Acquisition Unit** (**DAU**) meant for only acquiring the process data and with no control functions.

7.1 Communicable DACU

Sometimes, the standalone DACU needs to communicate with other external compatible devices/systems for sharing the information. To meet this additional requirement, the standalone DACU can be extended to support local as well as remote communication facility through the addition of communication modules (**LAN Interface** module for local communication and **Serial Interface**

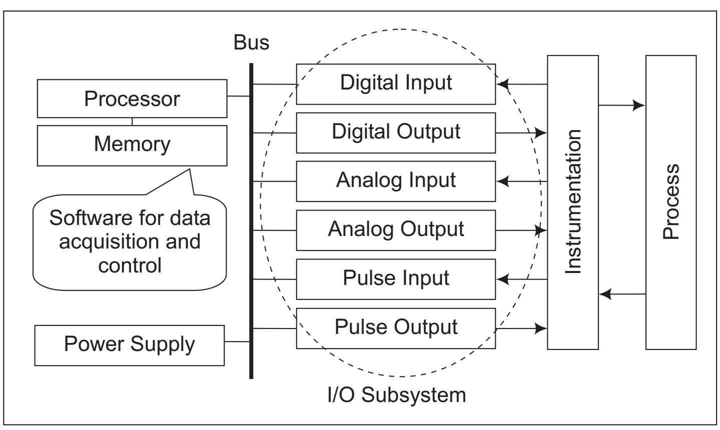


FIG. 4 STANDALONE DACU

module for remote communication). Figure 5 illustrates the block schematic of **Communicable** DACU and Figure 6 illustrates its physical structure.

7.2 Centralised I/O

In the figures shown above, all the I/O modules are accommodated in the same rack, are close to CPU/Memory via the bus, and are called **Centralised I/O (CIO)**. This arrangement has a serious disadvantage when the process signals are far away from DACU and are spread over a large area as all the signals need to be brought to DACU through long cabling. Apart from signal quality degradation, this is also expensive, and difficult to install and maintain.

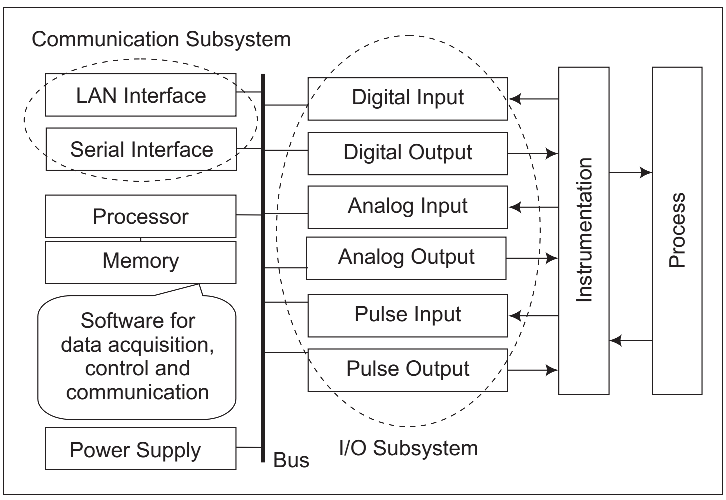


FIG. 5 COMMUNICABLE DACU

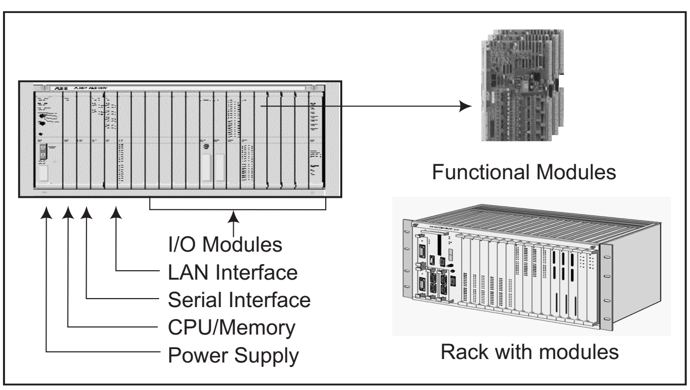


FIG. 6 TYPICAL PHYSICAL STRUCTURE OF DACU

7.3 Remote I/O

In order to overcome the above problem, **Remote I/O (RIO**) is developed. RIO is nothing but a specialised and dedicated communicable DACU whose main job is to receive the data from the process through its own local input modules and communicate the same to main DACU over the serial line and vice-versa for the output signals. The RIOs can be placed close to the process. With this arrangement, the signal cabling is only between the local input-output modules of RIO and the process. The link between RIO and the main DACU is only a serial communication cable with either 2 or 4 wires (data cable). Apart from cable reduction, **data transfer** takes place between DACU and RIO instead of **signal transfer**. Figure 7 illustrates block schematic of RIO while Figure 8 illustrates physical structure of a communicable DACU with both CIO and RIO.

A couple of disadvantages of remote I/O are:

a. The data acquisition and control operations are relatively slow because of serial communication.
b. Being intelligent and communicable, they are expensive.

While the first one is not that serious as communication in present day electronics is much faster than the process response requirements. However, the second one is important and needs a cost-

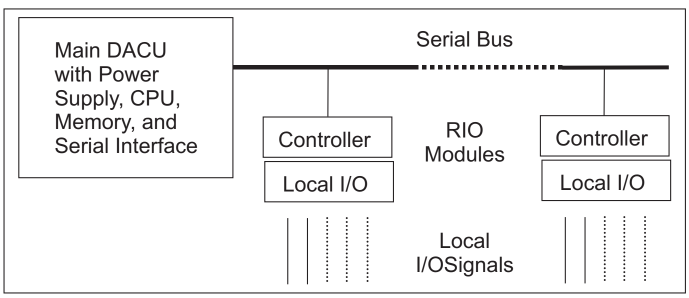


FIG. 7 REMOTE I/O

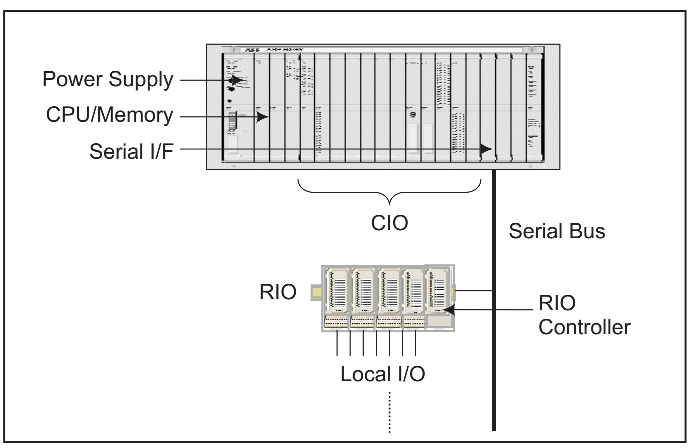


FIG. 8 COMMUNICABLE DACU WITH CIO AND RIO

benefit analysis to decide on CIO, RIO, or a combination by comparing their costs along with the associated cabling cost (cables, installation, maintenance, etc.).

8. PROCESS CONTROL STRATEGIES

Before we proceed further, let us look into some aspects of commonly used process control philosophies as given below:

- Open loop control
- Closed loop (Feedback) control
    - Discrete control
    - Sequential control with interlocks
- Continuous control
    - Two step (On-off) and multi step control
    - Analog loop control

8.1 Open Loop Control

The advantages of this control are: this is simple, economical, and gives a pre-defined response. The disadvantages are: this provides no corrections for external and/or internal disturbances. Hence, this is generally used for less demanding applications.

For example, if we switch a bulb on, the bulb glows. However, there is no correction if the intensity is less than the intended if there is a fall in the input voltage.

Figure 9 illustrates the typical open loop control system block schematic.

8.2 Closed Loop Control

To overcome the disadvantage of open loop control, the correction to the response is provided to continuously take care of the disturbances as well as the effects of process dynamics. The advantages are: the response follows the reference input (reducing the error/difference between the reference input and the actual output). The disadvantages are: this is complex and expensive. Hence, this is generally used for more demanding applications.

For example, let us take a car which has to move at a set speed irrespective of the disturbances on the road such as wind resistance, uphill, down hill, etc. The speed of the car is continuously

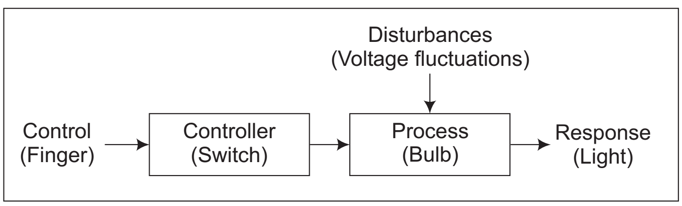


FIG. 9 OPEN-LOOP CONTROL

measured, compared with the set speed, and appropriate correction is introduced based on the difference (error) between the set speed and the actual speed either to increase the speed or to decrease the speed as required. Figure 10 illustrates the typical closed loop control system block schematic.

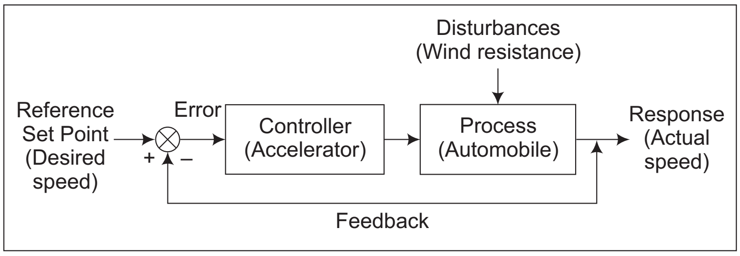


FIG. 10 CLOSED LOOP CONTROL

8.2.1 Discrete control

8.2.1.1 Sequential control with interlocks

This deals only with the discrete inputs and discrete outputs. In this control, the control sequences are executed sequentially (one after the other) only after **ensuring** that certain conditions (**interlocks**) are met with at each step before executing the step. The advantage of this control is of doing the sequential steps with interlocks quite fast and reliably without any manual intervention.

For example, let us take the case of operation of a lift in a building. The sequences of commands and the interlocks followed are:

a. The starting conditions are: the lift is not moving, lift is resting in a floor, and the lift cabin door is open.
b. A user, going up or down, enters the lift cabin and presses the request button (**main command**).
c. A sequential command is generated to close the lift cabin door **provided** “there is a request, door is already open, and lift cabin is not moving and is resting in a floor (interlocks)” conditions are satisfied.
d. Next sequential command is generated to start moving the lift cabin **provided** “there is a request, cabin door is closed, and cabin is not moving and is already in a floor (interlocks)” conditions are satisfied.
e. Next sequential command is generated to stop moving the lift cabin **provided** the “floor is reached (interlock)” condition is satisfied.
f. Next sequential command is generated to open the cabin door **provided** the “door is already closed, floor is reached, and cabin is not moving (interlocks)” conditions are satisfied.

In the above illustration, the lift control logic is simplified to impart the concept of sequential control with interlocks. In reality, the lift has many more parameters and the logic is quite complex.

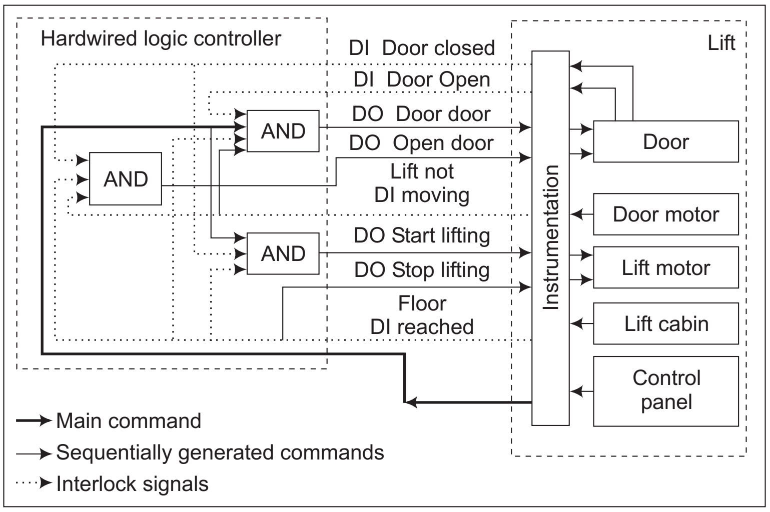


FIG. 11 LIFT CONTROL – SIGNAL FLOW

Figure 11 illustrates the signal flow of this process control. The instrumentation devices here are the push button switch for the main command; proximity/limit switches for recognising door open, door closed, floor reached, lift moving, lift not moving, etc. while control relays for issuing commands to start and stop moving the cabin, open/close door, etc.

8.2.2 Continuous control

8.2.2.1 On-off (two step) control

This employs a combination of analog and digital signals. In this case of feedback control, the process output is continuously compared with the set point and the control is switched **on** or **off** depending on whether the output is higher or lower than the set point. However, this leads to instability (controller output hunting) as the controller operates for even minor deviations. One way to improve the performance is to force the output to stay within acceptable band of higher and lower limits (**dead band**) and switch on or switch off of the control input only when the output crosses the dead band. The advantages of this control are: this is simple, cost effective, and a move towards continuous control. The disadvantage is: this control is executed in frequent on-off steps leading to faster wear and tear of the final control gear. Further, the smaller the dead band, even though provide a finer control, lead to hunting (oscillation) or instability of the control output reducing the life of control gears (final control elements).

For example, in an air conditioner, the compressor motor is stopped when the temperature of the chamber falls below a set low limit (say 24°C) and started when the temperature goes above the set high limit (say 26°C). No control action takes place when the temperature is in between 24°C and 26°C (dead band). Figure 12 illustrates the signal flow of this control without the dead band. Here the instrumentation devices are temperature transducer for measuring the actual temperature, potentiometer for setting the reference temperature and control relays for start/stop of compressor.

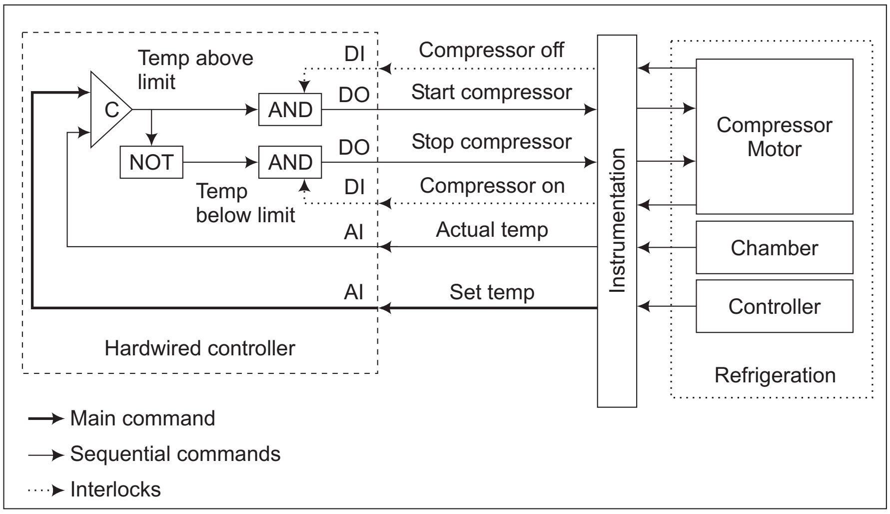


FIG. 12 AIR CONDITIONER CONTROL (WITHOUT DEAD BAND) – SIGNAL FLOW

One can extend the above logic to multi-step control to take the same even more closer to the continuous control. One of the common examples of the multi-step control is the house hold voltage stabiliser to maintain the output voltage (within a pre-defined band) irrespective of the input and/or load variations. The voltage output of the transformer is increased or decreased in multiple steps using its taps. Higher the number of taps, finer or smoother is the control.

Theoretically, the two step control with a **very small dead band** can work as continuous control. However, the same is not practically feasible.

8.2.2.2 Analog loop control

As the two step or multi step control, in practice, cannot give a perfect continuous control, the obvious next step is to adjust the control input **continuously** based on the actual process output to force the process output to follow the reference input continuously. This is known as **Analog loop control**. This employs only analog input and analog output modules. The advantage of this control is: this is finer, smoother, and the output continuously follows the set point or the reference. The disadvantages are: this leads to steady state error, long settling time, oscillations, etc. in the response due to the effects of process dynamics.

Figure 13 illustrates signal flow in the speed control of a motor. A Tachometer is used to measure the actual speed while a variable frequency drive (VFD) is used to control the speed of the motor as instrumentation devices. The frequency of the output voltage from VFD is continually adjusted over the nominal frequency (proportional to the deviation) by comparing the actual speed of the motor with the set speed.

Figure 14 illustrates the process response, in analog loop control, to a step control input. For every step change in the input, the process, due to its inertia takes some time to settle down. Apart

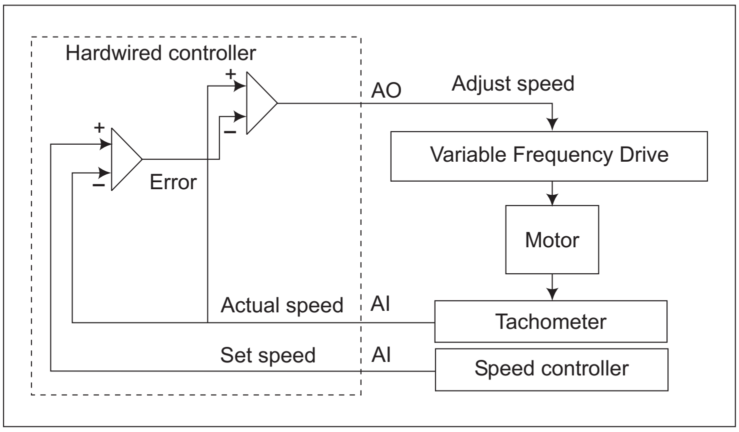


FIG. 13 MOTOR CONTROL – SIGNAL FLOW

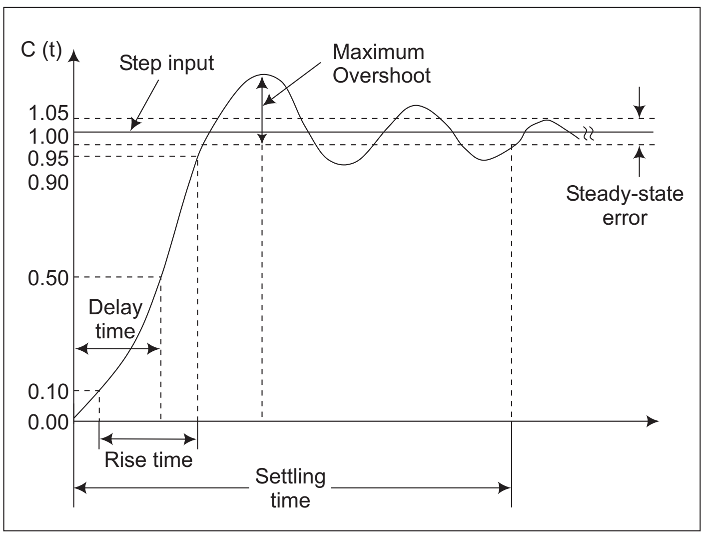


FIG. 14 PROCESS RESPONSE TO A CONTROL INPUT

from leaving a steady state error in the response, lower process inertia leads to increased oscillations (instability) and longer settling time while higher process inertia leads to more time taken to reach the steady state (sluggishness). All these conditions are not desirable and need some corrections to ensure zero steady state error, minimum oscillations, and minimum settling time so that process is ready to accept the next step change in the input. The above can be accomplished by employing **integral** and **derivative** control, in addition to **proportional** control.

In other words, the correction signal is computed as the summation of **Proportional, Integral,** and **Derivative (PID)** of the error signal with appropriate weightages. Figure 15 illustrates the

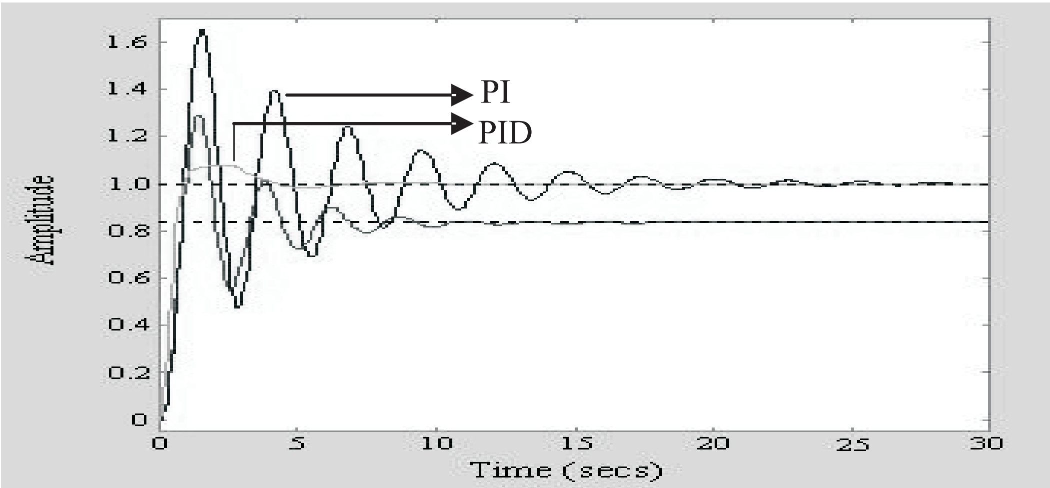


FIG. 15 PROCESS RESPONSE TO PID CONTROL

process response to PID signal. As seen in the figure, Proportional (P) control introduces steady state error. Proportional plus integral (PI) control eliminates the steady state error, but introduces the oscillations and increases the settling time.

In summary, the proportional plus integral plus derivative (PID) control eliminates the steady state error, reduces the oscillations and settling time. Through a proper mix of P, I, and D signals, it is possible to achieve the desired response. The latter is known as control system tuning.

There are many other control strategies such as feed-forward control, ratio control, cascade control, etc. which are not discussed in this article.

9. EVOLUTION OF CONTROL SYSTEMS

Originally, in discrete control, the control logic was based on **Relay logic** for sequential control with interlocks as simple relays (normally open or **NO** and normally closed or **NC**) were wired to produce the required logic.

The relay logic was subsequently replaced by **Solid state logic** which made the control hardware compact, reliable, less power consuming, and maintenance free. Similarly, **Solid state circuits** were used for implementing continuous control. The block schematic of the arrangement is given in Figure 16.

The next step obviously is to replace the hardwired controller by a DACU with required number of I/O and a suitable control program. This provided the **flexibility** to reuse the controller for variety of applications irrespective of the type of control just by reconfiguring the I/O subsystem and changing its control software. Additionally, DACU, being micro-processor based, is also capable of **self supervising** its own health.

Figure 17 illustrates the block schematic of the equivalent of the hardwired controller by replacing the same by a general purpose DACU (stand-alone or communicable) equipped with

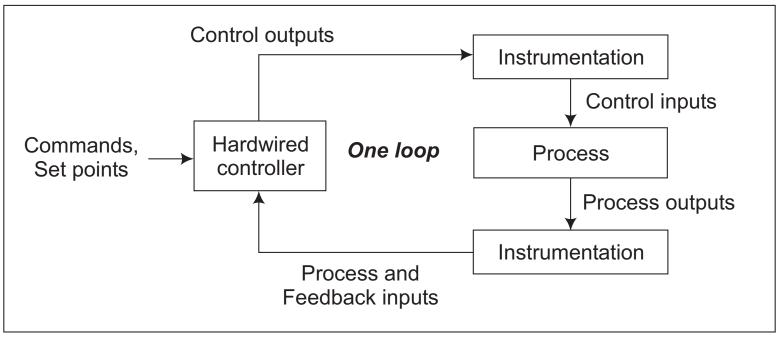


FIG. 16 HARDWIRED CONTROLLER

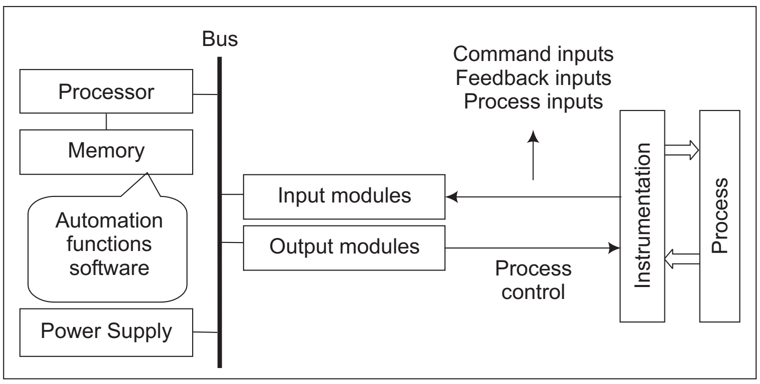


FIG. 17 DACU BASED CONTROLLER

required types and quantity of I/O modules and with the software for **relevant automation functions.** The latter can be configured to support (through suitable control software) all control philosophies with multiple inputs/outputs and multiple loops with **interdependence** among them, if required. Creating interdependence among various loops is not possible in hardwired controllers.

In practice, there are many compact micro-processor based controllers specifically designed to replace the existing hardwired controllers on **one-to-one basis** (maintaining the physical size and terminal compatibility) to support protection of related investments.

9.1 Special Purpose DACUs

Further evolutions of DACU are based on the concepts of data acquisition and control associated with localised and distributed processes as explained in the following sections.

9.1.1 Programmable logic controller

Programmable Logic Controller (PLC) is an application of general purpose DACU with only digital inputs and digital outputs to replace the traditional relay logic (or subsequent solid state logic) to execute the sequential control with interlocks. PLC may be communicable or non-communicable and is for used for automation of the localised discrete processes.

9.1.2 Programmable controller

Programmable Controller (Controller) is an extension of PLC to support analog inputs and analog outputs to perform continuous control also in addition to sequential control with interlocks with digital input and analog I/O. In other words, the controller merges the functions of PLC and analog loop control.

A controller may be communicable or non-communicable. Once again, controllers are for the automation of the localised processes.

9.1.3 Remote terminal unit

The distributed processes called for acquisition of both analog and digital signals from a distance, and process control from a distance (tele-metering, tele-signalling, and tele-control). To facilitate this, a communicable DACU with all types of I/O modules called **Remote Terminal Unit** or **RTU**, is employed.

Figure 18 gives the pictures of some commercially available PLC, Analog loop controller, Controller, and RTU.

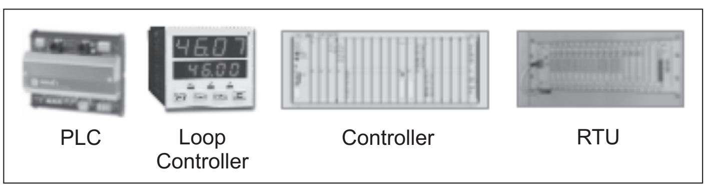


FIG. 18 COMMERCIAL DACUS

9.1.4 Difference between Controller and RTU

Physically and hardware-wise, both these look alike as they have all the components of a general purpose communicable DACU. The differences are mainly in their applications and the software.

The controller, while performing logic and continuous control (automation functions) adopts the following sequence:

- Data acquisition
- Execution of local automation functions
- Control execution
- Data transmission to/receipt from remote places (optional)

RTU, on the other hand, is functionally similar to remote I/O and adopts the following sequence:

- Data acquisition from the process, and transmission to the remote place
- Data (commands) receipt from the remote place and transfer the same to the process
- Execution of automation functions (limited and optional)

To sum up, both the controller and RTU support all the functionalities of DACU but only with different priorities.

When RTU was evolved, the same was only meant for process data acquisition and transmission to the control center and in the reverse direction, receipt of commands from the control center and transmission of the same to the process. RTUs were generally dumb and did not have enough computing power to execute the automation functions. However, with the availability of powerful and cost-effective micro-processors, things changed and the present day RTUs provide automation functions also (not as powerful as of controllers).

In practice, it is possible to use the controllers as RTUs and vice-versa provided the choice meets the application, functional, performance, and cost.

9.1.5 Programming of controllers and RTUs

Controllers, in additions to configuring, need extensive application programming for the implementation of application related automation functions. However, there is no need of any such programming if RTU is used for its basic functions (only configuring is required). Originally, PLCs used **Ladder Logic** programming which is similar to the relay logic wiring. Special compilers are available to support ladder logic programs. However, the ladder logic supports only the sequential control with interlocks and does not support continuous or analog loop control. The next step is to have **Functional Block** approach which supports both sequential control with interlocks and continuous control.

Subsequently, many computer-like languages are also developed to support programming of automation functions. In order to make the application programming independent of the hardware platforms and to support software portability, IEC brought all the languages under **IEC 61131-3** standard. The supported languages are:

a. Ladder logic diagram (LLD)
b. Functional Block Diagram (FBD)
c. Instruction List (IL)
d. Structured Text (ST)
e. Sequential Function Chart (SFC)

The structure of the above is shown in Figure 19. Each of the above has its own advantages. However, most commonly used programming languages by automation engineers are LLD and FBD.

10. HUMAN MACHINE INTERFACE

The simple applications of DACU did not call for any monitoring of the process and/or control of the process manually. Once DACU is programmed and installed, the same worked like a typical hardwired controller doing the assigned job not requiring any manual intervention. The need of an

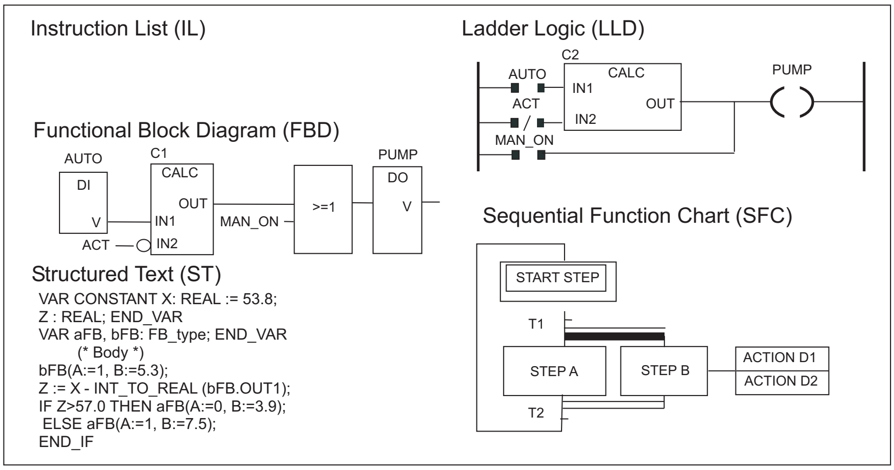


FIG. 19 STRUCTURE OF PROGRAMMING LANGUAGES

interface between DACU and the operator, or **Human Machine Interface (HMI)** has now become necessary for manual monitoring of the behavior of the process and effecting the manual control whenever needed. There are two ways to accomplish this function as given below.

10.1 Hardware based HMI

In early days and in the absence of sophisticated computers, a conventional **Mimic Diagram Board** or **MDB** based HMI was employed which essentially consists of:

a. Passive depiction of the full process diagram on its surface.
b. Dynamic display of the object states, values, etc. of process parameters using lamps, meters, indicators, etc. at appropriate places on process diagram on MDB.
c. Active control elements, such as push buttons, control switches, etc. for effecting the process control.

The above active elements are wired to the appropriate I/O modules of DACU. In other words, DACU, not only doing automation functions, but also interfaces the active elements of MDB for display of process parameters and for facilitating manual control. The selection of process objects is un-ambiguous here as all the dynamic elements corresponding to various process objects are placed at the appropriate location on the process diagram in MDB. Operator can watch the current status and the values of the process objects from a distance. However, he needs go near MDB to execute the manual control as the control elements are placed on MDB. This is also called **Operator panel**. MDB is interfaced with DACU as shown in Figure 20.

The main disadvantages of MDB are:

a. Extra I/O modules with wiring between DACU and MDB (hardware interfacing).
b. Expanding and/or modifying MDB is too difficult and expensive even though the construction of MDB with mosaic tiles reduced this problem to some extent.

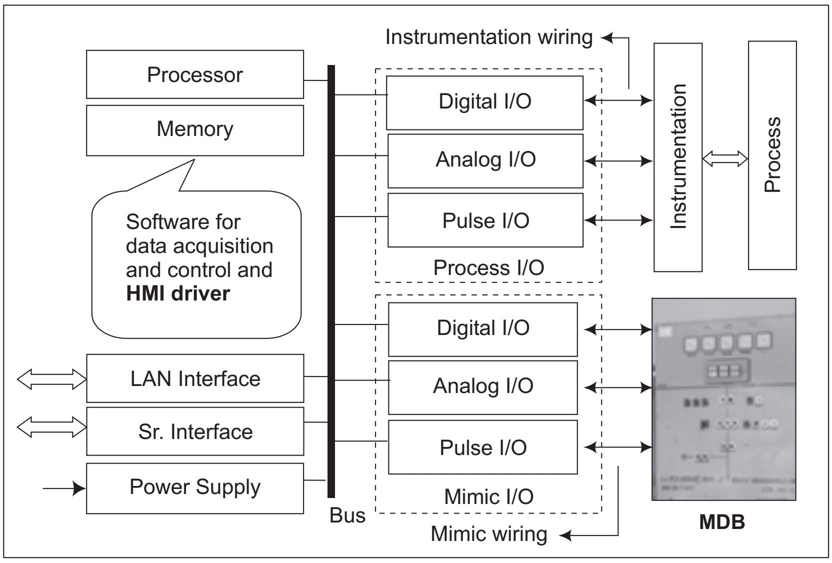


FIG. 20 MDB INTERFACING WITH DACU

c. Frequent movement of the operator between MDB and the control desk for control operations.

10.2 Software based HMI

To overcome the above disadvantages of the traditional MDB, software driven HMI is developed by employing the computer graphic terminals. This is also known as **Operator station**. In this, the screen consists of:

a. Passive display of the process diagram on the screen
b. Dynamic display of process objects through symbols at appropriate locations on process diagram on the screen
c. Mouse for selection of process objects for interaction
d. Mouse/keyboard for executing the process control

Even though the above retains most of traditional MDB features, the biggest disadvantage of this approach is its inability to present the panoramic view of the process as the screen size is limited. However, this can be made good to some extent by having multi-monitor based displays for different parts of process diagram and/or other frequently required information or going in for large screen display.

This is also called **Work Station**. Operator station is interfaced with DACU as shown in Figure 21.

The advantages are:

a. Needs much less space and power.
b. DACU drives the computer graphic terminal over LAN (software interfacing).

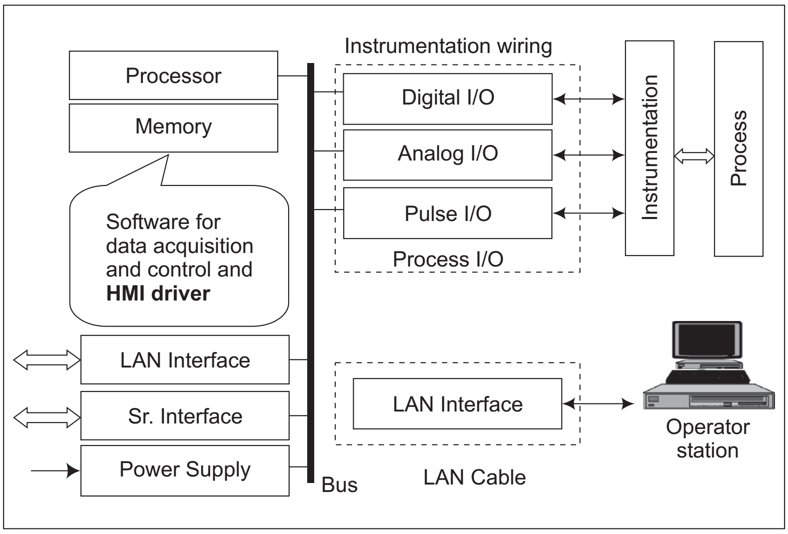


FIG. 21 OPERATOR STATION INTERFACING WITH DACU

c. Expanding and/or modifying HMI is easy and less expensive as only the software needs to be changed.
d. Operator can stay put near the control desk for all monitoring and control operations.

Figure 22 shows some typical displays (process diagrams, alarms/events, trends, etc.). MDB can display only the process diagram with dynamic objects while operator stations can present many other views such as alarm list, event list, trends, etc.

Figure 23 illustrates a HMI with both MDB and operator stations. In practice, in complex process plants, MDB is still employed to have a panoramic view of the process with a display of only a few selected dynamic process objects. However, the basic HMI operations (both monitoring as well as control) are done through the operator stations only.

11 AUTOMATION SYSTEM STRUCTURE

Earlier, we discussed two types of process, namely, lolalised and distributed. The automation system structure for a localised process is either **Centralised Control System (CCS)** or **Distributed Control System (DCS)** while the same for a distributed process is called **Network Control System (NCS)**. Various functional layers involved in the structures of CCS/DCS and NCS are shown in Figure 24.

11.1 CCS/DCS

In CCS, only one controller is employed for the monitoring and control of the entire process while in DCS, more than one controller is employed to provide improved availability of the overall control system through dedicated controllers to different sub-processes or sub-functions, by

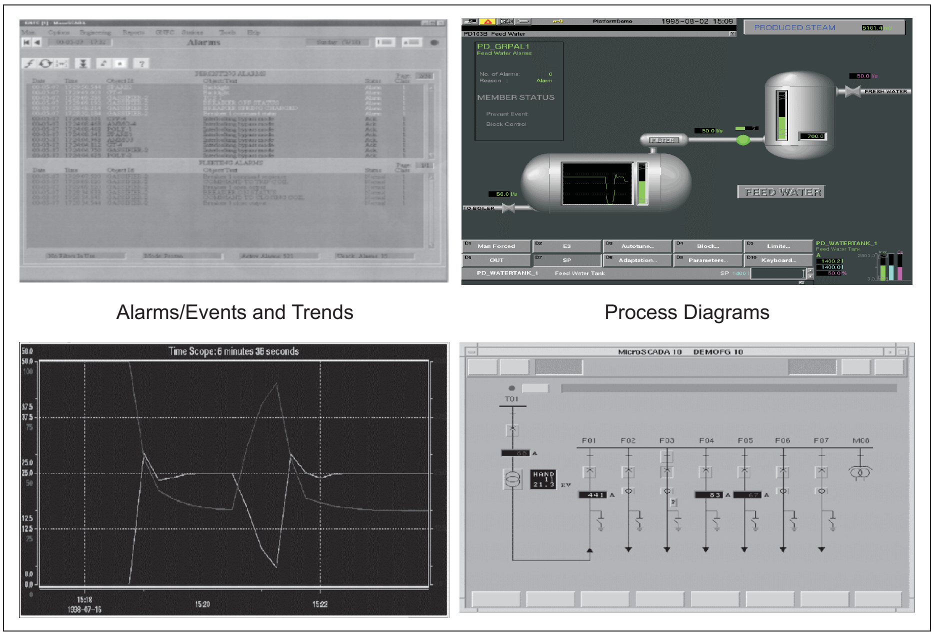


FIG. 22 OPERATOR STATION DISPLAYS

FIG. 23 HMI WITH BOTH MDB AND OPERATOR STATIONS

creating a distributed architecture of the controllers tightly coupled over a local area network (LAN). Also, there may be more than one HMI to distribute the load to many operators. The structure of CCS and DCS are illustrated in Figures 25a and 25b.

For example, in a power plant, independent controllers for various sub-processes such as boiler, turbine, generator, auxiliaries, etc. can be provided and connected over a LAN to share the information among themselves and with the operator stations.

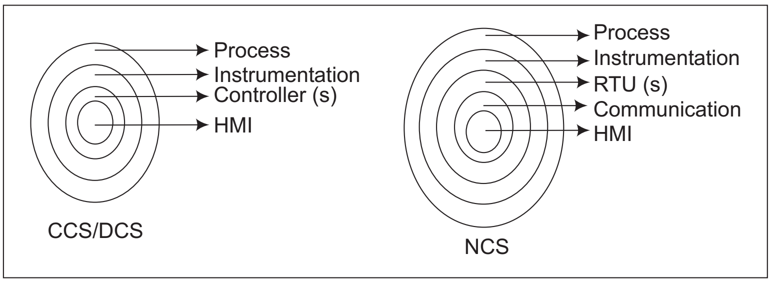


FIG. 24 FUNCTIONAL LAYERS OF CCS/DCS AND NCS

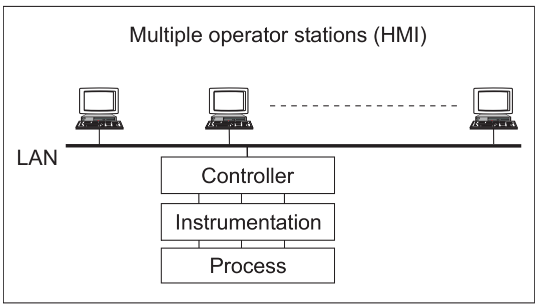


FIG. 25a CCS WITH MULTIPLE OPERATOR STATIONS

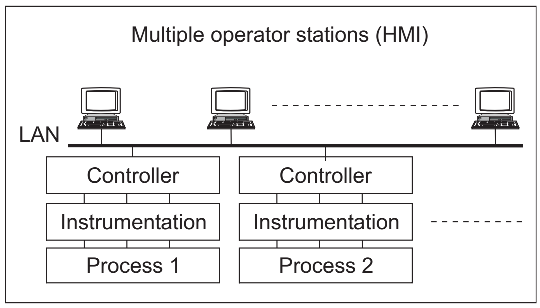


FIG. 25b DCS WITH MULTIPLE OPERATOR STATIONS

11.2 NCS

In a geographically distributed process, RTUs are distributed over a wide geographical area and are loosely connected over a wide area network (WAN) to HMI at the control center. The structure of NCS with multiple operator stations is illustrated in Figure 26a.

In the above basic scheme, performance of HMIs is seriously affected as HMIs, in addition to their own functions, have to perform a lot of additional routine communication functions to interact with RTUs. This is overcome by employing a **Front-end Processor (FEP)** to take away the additional routine communication load from HMIs as shown in Figure 26b.

FEP is a special and independent subsystem to deal with the communication with RTUs. In NCS, the network refers to the process networks and not the computer network at the control center.

An example of NCS is, in an electrical transmission network, all the geographically distributed electrical substations are interconnected through a transmission network. RTUs installed at all these substations are connected to FEP at the control center over WAN.

11.3 Supervisory Control and Data Acquisition

Earlier, the hardwired controllers worked independently and could not provide any interaction among themselves as they were non-communicable. However, there always existed a requirement

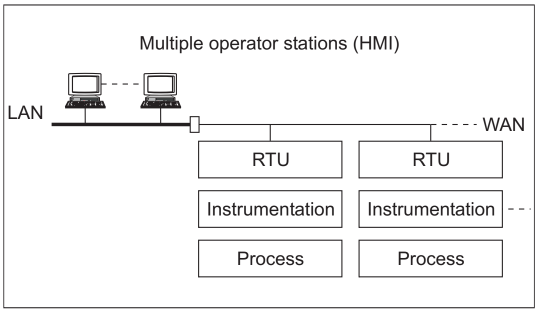


FIG. 26a NCS WITH MULTIPLE OPERATOR STATIONS

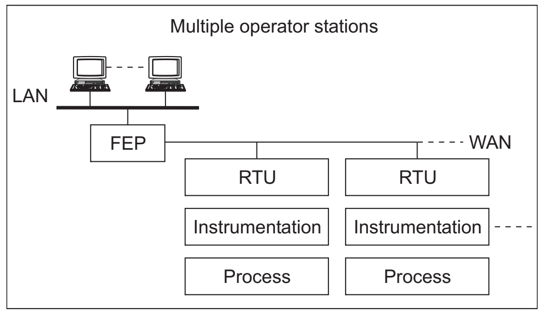


FIG. 26b NCS WITH FEP

to coordinate the functions of these distributed and independent hardwired controllers (especially the analog loop controllers) in the complex processes. Also, the computers were not first considered for process automation as they did not have the acceptable reliability. In view of this, as an intermediate step, the computers were introduced only to monitor, coordinate, and optimise the functions of the hardwired controllers. As hardwired controllers functioned independent of the supervisory computer, the failure of the computer did not affect the process control. This is called **Supervisory Control and Data Acquisition (SCADA)**.

An example of SCADA is to use a supervisory computer to monitor, co-ordinate, and optimize the set points of a group of analog controllers as shown in Figure 27.

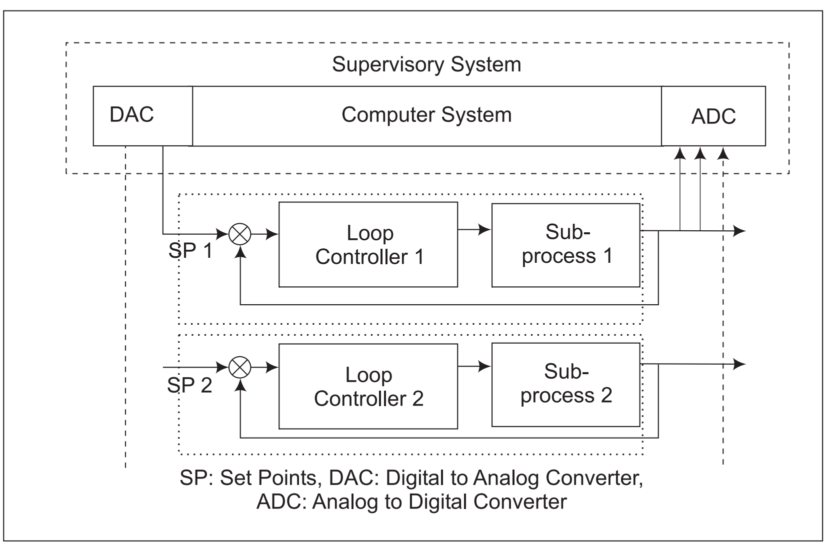


FIG. 27 TRADITIONAL SCADA SYSTEM

The philosophy used in NCS is also quite similar to SCADA as all the RTUs, functionally similar to the hardwired controllers (tele-metering, tele-signaling, and tele-control equipment), even though intelligent, are coordinated by a central supervisory computer system. In view of this, even today, NCS, sometime, is referred to as SCADA.

To sum up, in today's automation context, SCADA has become more a generic term and an integral part of the automation system supporting the following basic and supervisory functionalities:

a. **Data acquisition** – bring the process data (both from near/local and far/remote) be it analog or digital.
b. **Data supervision** – monitoring the received data for their abnormalities, if any, in the control room itself.
c. **Process control** – controlling the process parameters effecting the changes in the process through commands issued from the control room and receiving the result of the control action (back indication).
d. **Process survey** – surveying the entire process from the control room for all the happenings in the process.
e. **Process studies and optimisation** – utilising the data received/generated for performing process studies to arrive at optimum control strategies.
f. **Human machine interaction** – extensive interaction with all the process parameters to change their limits, values, blocking, de-blocking, etc.
g. **Data logging and history generation** – alarm, event, and periodic logging to create a history bank for future reference.
h. **Supervisory functions** – implementing hierarchical control for better co-ordination of distributed controllers and FEPs.

The above are supported in today's automation systems (in both DCS and NCS).

Also, in today's computer based automation systems, we can have a large amount of data acquired, stored, and accessed when required as shown in Figure 28.

In a stable process, the data required for normal operations is very small. However, a little more data is required to handle an emergency situation. The advantage is that we have the entire data at our finger tips to meet any contingency.

12 REDUNDANCY (STANDBY) CONCEPTS

Failures of any vital part in an automation system cannot be ruled out however rare they are. In the above DCS and NCS configurations, the failure of any vital subsystem takes full or part of the process out of monitoring and control.

Further process monitoring and control is not possible till the failure is attended to. Figure 29 and the following sections address various methods available to overcome this problem.

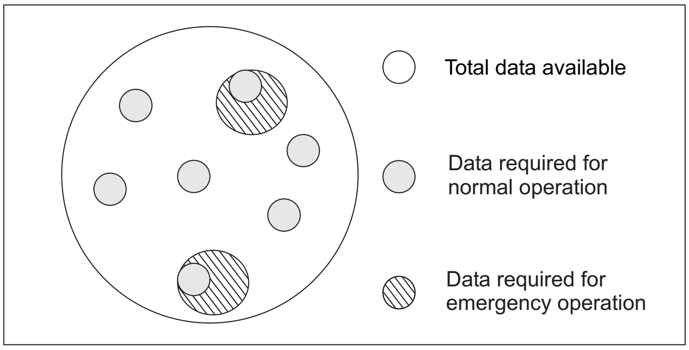


FIG. 28 DATA UTILISATION IN AUTOMATION SYSTEM

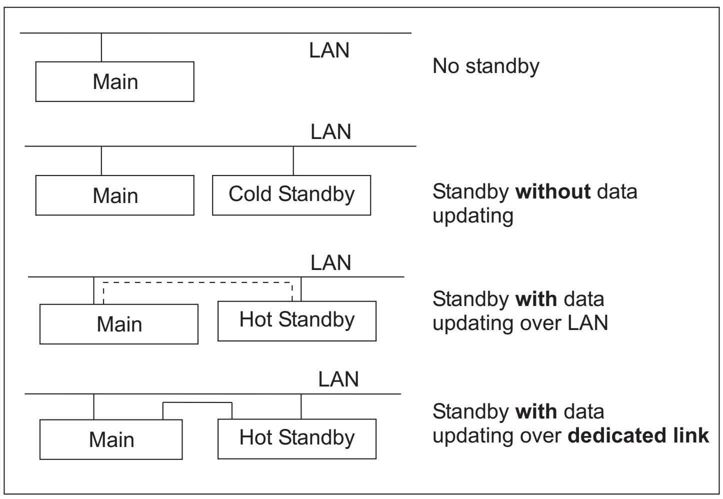


FIG. 29 REDUNDANCY (STANDBY) CONFIGURATIONS

12.1 Controllers and FEP

12.1.1 No standby

In some non-critical processes, there is only one system (Controller or FEP) and its non-availability for some time is acceptable. This means that the process monitoring and control can not be done till the problem is attended to. In this case, MTR (Mean Time to Repair) is quite high as only after repairing and re-starting the failed unit can resume the operation. The advantages are: this is simple and cost-effective. The disadvantage is its non-availability during failed period.

12.1.2 Cold standby

In this scheme, a standby is provided physically and, on failure of the main, the standby is brought on-line manually (start up). The advantages are: this is still simpler and cost-effective. The disadvantages are: There is a time gap between the failure of the main and starting up of the standby. Also, the standby starts afresh and does not carry the history available in the failed system.

12.1.3 Hot standby

This scheme overcomes the limitations in the cold standby by keeping the standby always **on-line** and ready to take over from the main. The hot standby, continuously supervises, through queries, the health of the main at regular pre-defined intervals. If the main responds positively to this query, the latest data in the main is transferred (data shadowing) to the standby over the LAN. If the main does not respond, the standby declares that the main has failed and immediately takes over the functions of the main, changes its own state to main, and starts functioning. The amount of data lost is limited to the data during the changeover interval. DACU accords highest priority for this **change over** function over its other functions. The lower the change-over interval, lesser is the data lost. However, this is with the increased computational load on DACU leading to the possibility of other programs not meeting their response time.

The above has one disadvantage. The LAN, being used for data shadowing, might slow down its performance because of this additional periodic data transfer load. To overcome this, an additional dedicated link is provided for this periodic and fast data transfer sparing the LAN. However, all these come with a price and technical complexities.

12.2 Communication

12.2.1 LAN

If LAN has no redundancy, the failure of a LAN component does not allow inter-subsystem communication partly or fully. Following are some methods to provide redundancy to overcome this problem.

First step in this case is to go in for physical duplication of LAN as shown in Figure 30. In Configuration 1, both main and hot standby are not connected to both LANs. Hence, simultaneous failure of **Main and LAN 2** or **Standby and LAN 1** does lead to a total failure irrespective of duplication of controller/FEP and LAN. In Configuration 2, this problem is overcome by connecting both main and standby to both LANs. Of course, all these come with a high price and increased complexities.

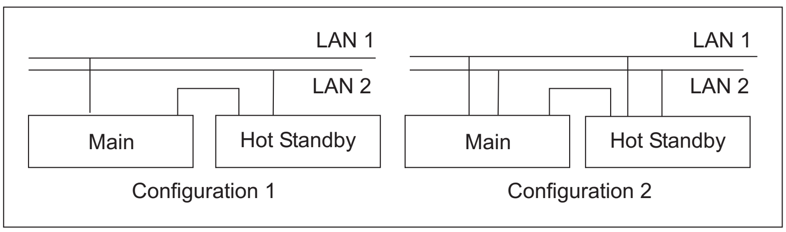


FIG. 30 LAN REDUNDANCY

12.2.2 WAN

If WAN has no redundancy, the failure of a WAN component at the control center paralyses all the RTUs. Figure 31 shows a method to provide redundancy to overcome this problem.

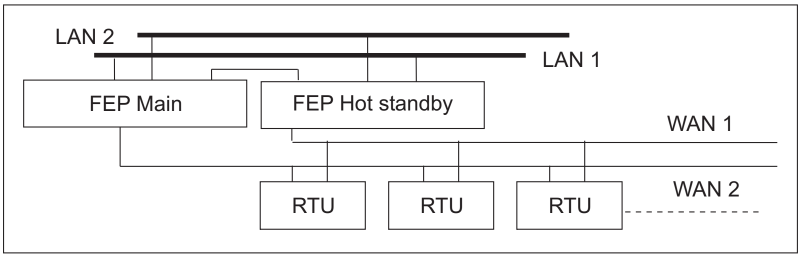


FIG. 31 WAN REDUNDANCY

In this case, apart from providing the redundancy for FEP and LAN, redundancy is provided for WAN also along with its interfaces at RTUs. This ensures that all the RTUs are always connected to the control center even when a WAN fails.

12.3 Some General System Configurations

There are many configurations of automation systems (simple to complicated) meeting various application requirements as given below.

12.3.1 Independent operator stations

In the configuration shown in Figure 32, failure of any one operator station can not lead to a total failure of monitoring and control as the functions of the failed operator station can be performed from the other. The only disadvantage is that the work allocated to two operator stations (for monitoring and control of specific areas or functions of a process) need to be done by a single operator station till the failed operator station is rectified and brought on-line.

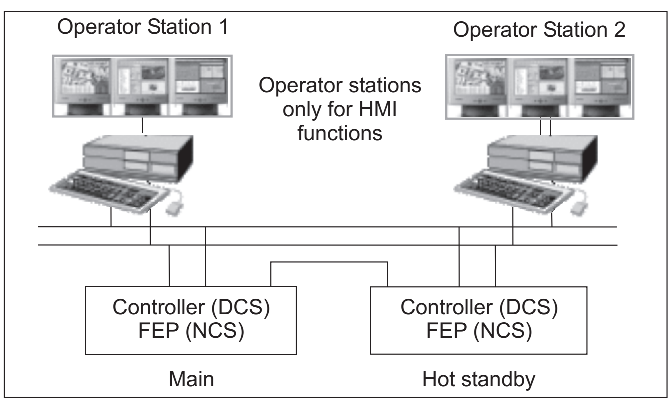


FIG. 32 INDEPENDENT OPERATOR STATIONS

12.3.2 Operator stations with supervisory functions

In some situations, where many controllers are present in a DCS or many FEP's in a NCS, there may be a necessity to have some supervisory functions to coordinate these controllers/FEPs. These supervisory functions can be accommodated in one of the operator stations. However, this has a drawback if the particular operator station fails thereby losing the supervisory functions. To overcome this problem, hot standby concept can be employed having two identical operator stations each equipped with the supervisory functions as shown in Figure 33.

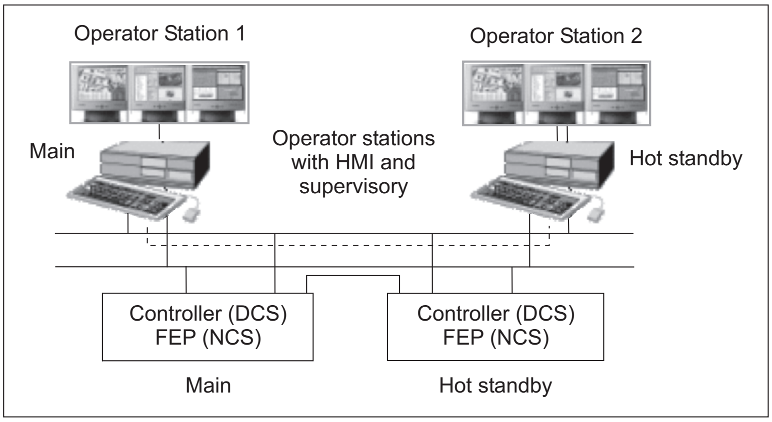


FIG. 33 OPERATOR STATIONS WITH SUPERVISORY FUNCTIONS

12.3.3 Independent supervisory stations

If we load the operator stations with the supervisory functions, the same may load the operator stations and bring down their performance (response time). This is not desirable as operators need a quick response from HMI to attend to the emergencies. To overcome this, the supervisory functions can be shifted to an independent station. However, these functions, always being critical, need hot standby with data shadowing feature as shown in Figure 33.

12.3.4 Independent supervisory stations

If we load the operator stations with the supervisory functions, the same may load the operator stations and bring down their performance (response time). This is not desirable as operators need a quick response from HMI to attend to the emergencies. To overcome this, the supervisory functions can be shifted to an independent station. However, these functions, always being critical, need hot standby with data shadowing feature as shown in Figure 34.

12.3.5 Independent stations for applications

In some cases, DCS or NCS may have to support special application software which is computational intensive and accommodating them in any of the existing stations (operator stations or supervisory stations) is not feasible due to the degradation in the latter's performance. Hence,

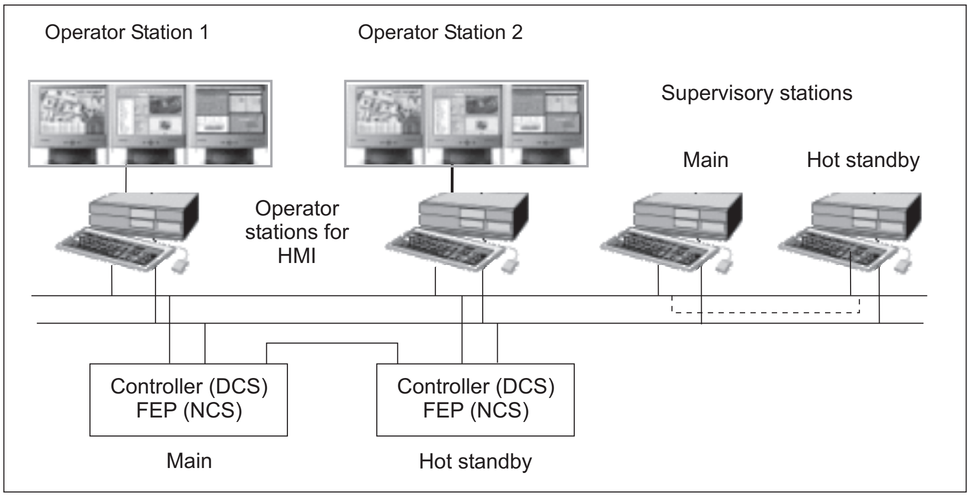


FIG. 34 INDEPENDENT SUPERVISORY STATIONS

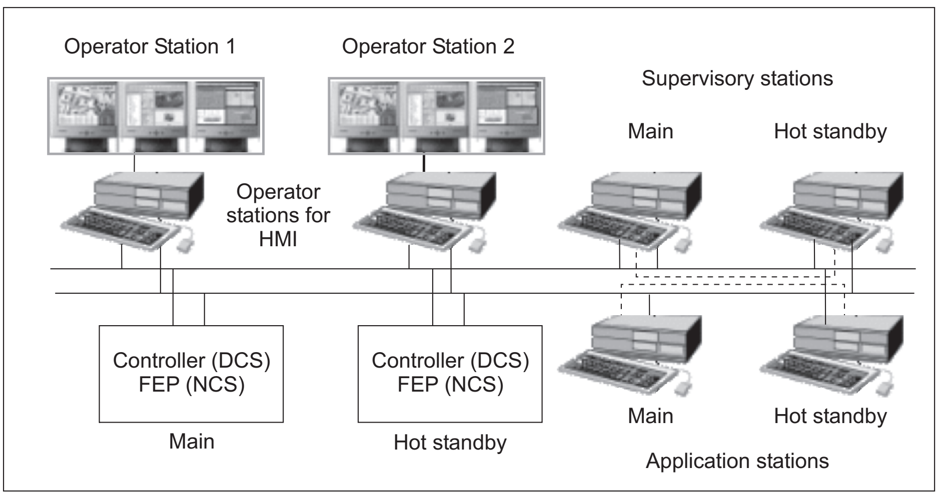


FIG. 35 INDEPENDENT APPLICATION STATIONS

independent stations are provided (with or without standby) to support these applications as shown in Figure 35.

All the above configurations are supported in today's automation systems (both in DCS and in NCS).

Finally, Figure 36 shows a picture of a complete modern control center.

FIG. 36 MODERN CONTROL CENTER

13 CONCLUDING REMARKS

In the above article, based on author's 26 years of learning, practicing, and training experience in ABB, India, an attempt is made to give an overall picture of the industrial process automation which is in line with the current technology and practices. Hence, one can expect some ABB-centric features in the article.

With the above general introduction to the subject, it possible to study further the working platforms of the major automation companies such as ABB, Honeywell, Siemens, Emerson, Invensys, etc. All these platforms generally follow more or less a similar philosophy while maintaining their own specific technical differences.

REFERENCES

Various, books, papers, articles, websites, manufacturer's published information, etc. on Industrial Process Automation

ACKNOWLEDGEMENTS

Author likes to gratefully thank:

a. ABB, Bangalore - where the author learnt, practiced, and taught industrial process automation. Some of the concepts and the pictures are based on ABB products, systems, and solutions
b. International Institute of Information Technology Bangalore (IIITB) - where the author is presently teaching deployment of information technology in industrial process automation and related subjects
c. National Institute of Technology Karnataka (NITK), Surathkal - where author taught basics of industrial process automation under AICTE/INAE Distinguished Visiting Professor program

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

CONSTRUCTION OF NPP CONTAINMENT STRUCTURES WITH HPC

PRABIR C. BASU*

1.0 INTRODUCTION

The reactor building of pressurized heavy water reactor (PHWR) based Indian nuclear power plant (NPP) has double containment structural systems, Figure 1. The inner containment structure (ICS) is made of pre-stressed concrete and consists of cylindrical wall (IC wall) and capped with segmented hemispherical dome. The outer containment structure (OCS) is a reinforced concrete structure having similar configuration. Both the containment structures are built on a common reinforced concrete raft. The inner containment dome (IC dome) of Kaiga Atomic Power Project, Unit-2 (Kaiga-2) is the first concrete structure, which was constructed in India with engineered high performance concrete (HPC) using silica fume. This was followed by the construction of three more IC domes; two domes of Rajasthan Atomic Power Projects, Units-3&4 (RAPP-3&4) and one dome of Kaiga-1. Complete inner containment structures of Tarapur Atomic Power Project, Units-3&4 (TAPP-3&4) including raft was constructed with silica fume based HPC.

The mechanism, which leads to development of HPC has been discuss elsewhere [1] and in the accompanying article of the author published in this volume [2]. The mechanism has basically three components,

- Reaction mechanism,
- Physical process, and
- Curing

Reaction mechanism is the basic component in developing the HPC, and conceptually it continues for a long time. This is principally based upon chemical reaction between the cementitious material, i.e. cement and mineral admixtures and related physical phenomena [3,4]. Hydration of cement is the basic feature of reaction mechanism. Mineral admixture such as silica fume in the present case improves the transition zone by grain refinement [5] as well as pore refinement through pozzolanic reaction, and acting as filler material. Superplasticiser creates conducive condition for complete hydration of cementitious materials principally by deflocculating

*Civil & Structural Engineering Division, AERB, Mumbai.

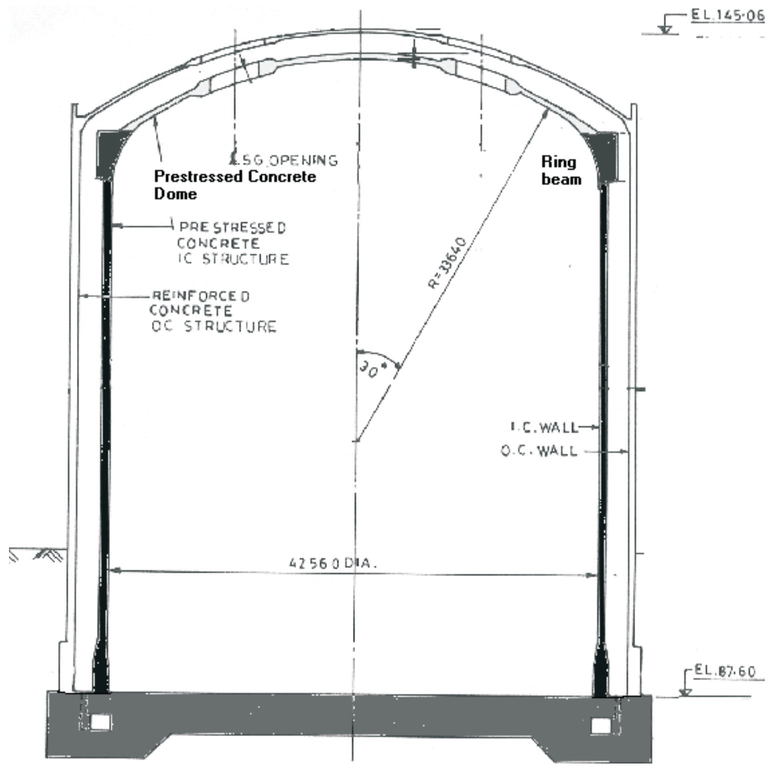


FIG. 1 CONTAINMENT STRUCTURES OF INDIAN PHWR BASED NPP

the lumps (which form after mixing cementitious material with water) and making cementitious material water mixture well disperse system that brings almost all cementitious particles fully in contact with water [3,4].

Physical process results in creating satisfactory condition for the reaction mechanism to take place appropriately and imparting desired attributes the concrete mix both at fresh and hardened state. Mixing of ingredients for production of concrete, transportation, placement, compaction, etc, are part of the physical process.

Curing is important for successful completion of reaction mechanism among the ingredients to desirable state. Curing protects the hardening concrete till it reaches a certain degree of hydration so that long-term properties can develop in storage or during service condition. Curing of concrete aims at the maintenance of a satisfactory moisture content and temperature in concrete during its early stages.

Physical process and curing are primarily supportive components. The reaction mechanism of HPC is more sensitive to various items of construction activities in comparison to conventional normal strength concrete (NSC). Therefore, a successful construction method of concrete structure using HPC should aim at adopting appropriately desired physical process and curing condition so that reaction mechanism takes place without hindrance in “in-situ condition” during and after construction of structures.

The above discussion implies significant modification in approach and in the construction methods for concrete structures, especially those of nuclear power plant structures, when HPC is used. Construction activities should be systematically carried out to achieve the in-situ properties of HPC, which are specified from its intended use [6]. Figure 2 summarize different construction activities related to the three principals/components of mechanism related to developing HPC.

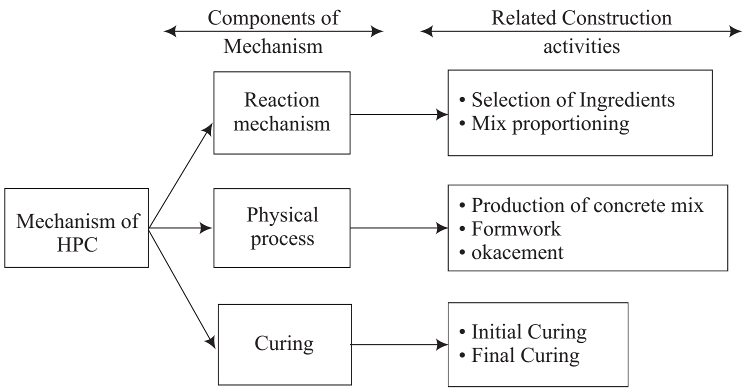


FIG. 2 MECHANISM OF HPC

Proportioning of HPC mixes for Kaiga-1&2, RAPP-3&4 and TAPP-3&4 are presented in the accompanying article [2] and other publications [7,8] and not discussed here. Present paper explains different aspects considered and approach adopted in the construction of the IC structure of Indian NPP using silica fume based HPC. All these activities were carried out following a well laid down construction methodology. The construction methodology, at first stage, was prepared based on previous experience, information available from published literature and design requirements. A number of field trials in the form of mock-up studies were carried in order to examine the constructibility of the IC structure using HPC. The construction methodology was then finalized incorporating the information and experience gained from these mock up studies.

2. FIELD TESTS AND TRIALS FOR PHYSICAL PROCESS

The high performance concrete mix, designed for the IC domes of Kaiga-1&2 and RAPP-3&4 and TAPP-3&4, were tested in detail to ensure that the mix satisfies the specified requirements. Laboratory tests and field tests were carried out to assess primarily the workability, strength and permeability under all probable conditions. Field tests were carried out to ensure that the in-situ strength of the mix would meet the structural design requirements. These tests were undertaken also to ascertain the pumpability of mix and to determine the properties of the pumped concrete, especially the strength. Table 1 contains the pumpability trial test results on strength of Kaiga-HPC. Concrete samples were taken before and after pumping for wet sieve analysis and it was observed that mix proportion of ingredients remained almost identical after pumping. It was also observed that HPC mix was more cohesive, had low slump retention capability, and became sticky as compared to NSC.

Table 1. *Pumpability trials test results on strength of Kaiga HPC*

Curing regime	*N*	*28-day compressive strength (N/mm^2)*			*Split tensile strength(N/mm^2)*		
		μ	σ	f_{ck}	μ	σ	F_{st}
Laboratory curing	34	77.0	3.59	71.10	4.325	0.234	3.940
Site curing	34	74.10	3.47	68.40	4.342	0.246	3.937

Note:

N: No. of samples; m: Average strength; s: Standard deviation of strength; f_{ck} and f_{st}: characteristics value of compressive and split tensile strength respectively.

A number of field trials were carried out in the form of full-scale mock-ups to examine the constructibility aspects. Another outcome was to fine tune the concrete mix if found necessary from the observations of field trials. In addition to examination of constructability aspects, these field trials were also helpful in studying various other parameters such as reinforcement detailing, construction joint pattern and the efficacy of using top formwork. Further, the field trials also helped in identifying the congested areas and problem spots, thereby enabling the designers to carry out suitable modifications in the detailing. The construction methodology for IC structures were finalised based on the information gained from these field trials. The various parameters studied during the field trials as regards to HPC were

i) Mixing method
ii) Production of cool concrete
iii) Constructibility
iv) Construction joint preparation
v) Placement, spread and compaction
vi) Curing

Most of the filed trials were carried out at Kaiga project site. Experiences gained there were used in the construction at RAPP-3&4 and TAPP-3&4. The field trials were carried out following well laid down methodology and strict quality control procedure. Detailed discussion on these field trials is presented by Jagdish et al., 1999 [9].

2.1 Mixing Method

Field trials were carried out for optimising mixing sequence and duration of mixing. Finally two stage mixing method was found to be efficient. Silica fume was mixed dry along with aggregates, in the first stage for 5 seconds, followed by final mixing of 45 seconds after adding all the ingredients. Total cycle time for each batch of 0.5 M^3 worked out to 80 seconds, which is nearly one, and half times of that NSC.

2.2 Production of Cool Concrete and Transportation

Cement content in HPC is much higher than NSC. To overcome the problem of heat of hydration, concrete temperature was restricted to 23°C and 19°C at placement point for Kaiga-1&2 / RAPP-3&4 and TAPP-3&4 respectively. Temperature of concrete at production point was worked out to be 13-15°C considering the ambient temperature of project site (maximum of 40°C – 42°C); possible temperature rise in the time gap (maximum value of 90 minutes) between concrete production, transportation and pumping; and additional margin of 4°C in order to account for unforeseen situation. During trials, it was observed that a maximum of 90% of the quantity of water could be replaced with ice flakes, which was not sufficient to achieve the requisite temperature of concrete at production point. After studying the efficacy of various methods for cooling, a combination of addition of ice flakes (as partial replacement of water) and cooling of aggregates by flowing chilled air through aggregates were adopted in all sites. It was possible to maintain the required limits of temperature of concrete by adopting the above method of cooling the concrete. To reduce temperature rise in concrete during transportation, transit mixers were covered with wet burlap and sponge.

2.3 Examination of Constructibility by Mock-up Studies

IC structures were constructed by pumping the concrete. Slump retention of any concrete mix is an important characteristic for pumping. HPC has less slump retention characteristic than that of NSC. Also, HPC becomes stickier with comparatively less time and it is difficult to pump and place sticky mix. Two field trials were conducted to study the slump retention, pumpability, flowabiliy and other construction aspects by means of full-scale mock-ups of a portion of the ring beam (field trial FT-1) and a portion of the dome near the steam generator (SG) opening (field trial FT-2). The concrete mix was designed for a slump of 175±25 mm at batching plant with exposed time zero.

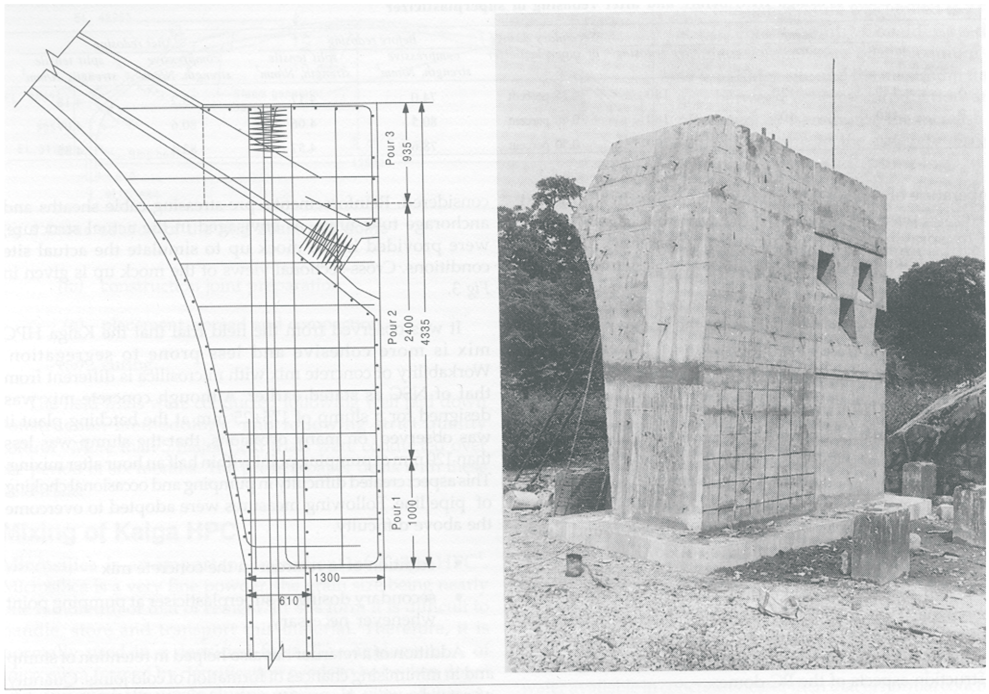
Field trial FT-1

This field trial was carried out by simulating a full-scale mock up of a portion of a ring beam having maximum congestion with reinforcement, pre-stressing cables, and anchorages. The full height, 4.335m of the ring beam along with 1.5m stretch of the springing of dome, and a length of 3.0 m was considered. Reinforcements, pre-stressing cable sheaths and anchorage tube units, as envisaged in the actual structure, were provided in the mock up to simulate the actual site conditions, Figure 3.

It was observed from the field trial that the Kaiga HPC mix was more cohesive and less prone to segregation. Although concrete mix was designed for a slump of 175±25 mm at the batching plant, it was observed on many occasions that the slump was less than 120 mm at pumping point within half an hour after mixing. This led to difficulty in pumping and occasional choking of pipeline. Following measures were adopted to overcome the above difficulty.

- Addition of a retarder in the concrete mix
- Secondary dosing of superplasticiser at pumping point whenever necessary.

Whenever the slump of concrete was found to be below 140 mm at the pumping point, second dosage of admixture was mixed to improve the workability of mix. To study the effect of secondary dosage on strength of concrete, a few additional trials were carried out. Three Kaiga HPC mixes were produced in the batching plant and transported to site in transit mixers. These transit mixers were retained for 30 (trial-A), 45 (trial-B), and 60 (trial-C) minutes to allow drop in workability and then based on the slump of the concrete secondary dosage of superplasticiser was added. Test results of workability and strength before and after re-dosing are given in Table 2. Secondary



(a) **Cross section of field trial-1 (ring beam)** *(b)* **View of ring beam after concreting**

FIG. 3 FULL-SCALE MOCK-UP OF A PORTION OF RING BEAM HAVING MAXIMUM CONGESTION OF REINFORCEMENT (FT-1)

dosage of superplasticiser was observed effective in increasing the slump of concrete mix without having any detrimental effect on strength of concrete.

Addition of a retarder has also helped in retention of slump and in minimising chances of formation of cold joints. Quantity of retarder mixed was determined by trials. Substantial loss of slump was also observed in Rajasthan and was taken care by increasing the superplasticiser dosage (than that determined by laboratory trials) inthe mix proportion. Such phenomenon was not observed in TAPP-3&4 as the mix design was developed with superplasticiser dosage more than that of Kaiga HPC.

Field trial FT-2

A 5.0 x 5.0 m plan area of a portion of the IC-dome engulfing a quadrant of the steam generator (SG) opening including the transition zone from general thickness to increased thickness near the opening was considered to study the constructibility aspect at critical locations of the dome using *Kaiga-HPC*. Reinforcements, pre-stressing cable sheaths, anchorage tube units and embedments, as designed in the actual structure, were provided in the mock-up to simulate the actual site conditions, fig.4 delineates the portion of the dome containing SG opening considered for field trial 2. Slope of the dome in some locations (for example, near S.G. opening area, Figure 4) and

Table 2. *Test results of Kaiga HPC before and after re-dosing of superplasticiser*

Time[1] (min.)	*Re-dose[1](%)*	*Slump (mm)*			*Strength (N/mm²)*			
					Case-II		*Case-III*	
		Case-I	*Case-II*	*Case-III*	f_c^4	f_{st}^4	f_c^4	f_{st}^4
30	0.25	170	120	180	74.0	4.17	75.2	4.18
45	0.40	180	100	180	80.5	4.06	80.6	4.07
60	0.50	160	75	190	78.6	4.57	83.4	4.55

Note:

1. Time after production of concrete
2. By weight of cement.
3. Case-I = intial at batching plant; Case-II = before re-dosing at pumping point; Case-III = after re-dosing at pumping point.
4. f_c = Compressive strength, f_{st} = Split tensile strength.

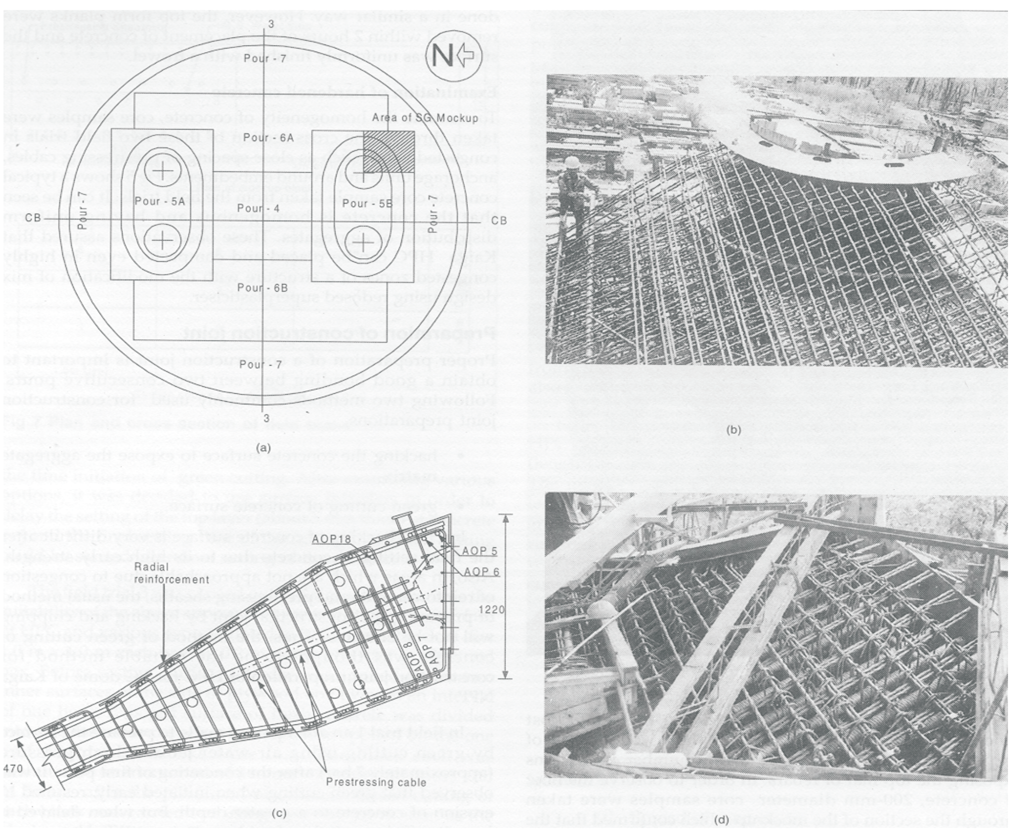


FIG. 4 MOCKUP FOR CONCRETING OF DOME NEAR SG OPENING (FT-2)

springing levels are more than 15°C. Top formwork was used in these locations. Arrangement for fixing the top shutter is also shown in the figure. To support the shutters, trusses for the formwork were arranged in orthogonal direction.

Post concreting observations indicated that there were pockets of unconcreted areas on the top surface at number of locations exposing the top layer of reinforcements. 200-mm diameter core samples were taken through the section of the mock-up, which confirmed that the uneven flow of concrete was a phenomenon of only the top layer and the distribution of aggregates and compaction of concrete in the remaining section was very good.

Field trial FT-2R

In view of the undesirable quality of the surface finish obtained in the field trial 2, a repeat mock-up of the same portion (field trial FT-2R) was conducted with a modification of construction methodology. In a significant change from the earlier mock-up (field trial FT-2), the pattern of the top truss was revised from orthogonal direction to radial direction (in plan), thus allowing the placement of planks along the natural contours of concrete. In addition, minor modification in reinforcement detailing was incorporated. Concreting was done in a similar way. The top from planks was removed within 2 hours of the placement and the concrete surface was found to be of acceptable quality.

2.4 Construction Joint Preparation

Field Trial FT-3

The field trial 3 was carried out at Kaiga site to examine the method of green cutting of the surface of HPC and various aspects related to it. Concrete block of size 1.0 m x 1.0 m, each having 0.2 m depth, was cast in plywood box. Before casting, surface retarder was applied on the two inner surfaces of the side of plywood box twice at an interval of one hour. Exposed surface of fresh concrete was divided into two halves and surface retarder was applied on the one half. Eight such blocks were cast. One block each was de-shuttered 8, 12, 16 and 24 hours after the placement of concrete. Green cutting of each block was carried out using water-air jet as soon as the block was de-shuttered. The trial was carried out using surface retarder procured from two manufacturers, mentioned as “A” and “B” in Table 3. The method provided good green cut surface, Figure 5.

The surfaces applied with surface retarder of brand ‘B’ yielded good results, which essentially indicate the necessity of using appropriate type of surface retarder. The optimum time for initiation of green cutting was found to be 8 to 12 hours, which were after elapsing of final setting time of concrete.

2.5 Placement, Spread and Compaction of HPC

As stated earlier, HPC is cohesive and does not tend to segregate, but loss of slump with time is very fast and it becomes stiff and highly adhesive. Above factors affect placement, spread and compaction of concrete. To study these aspects, field trial 4, was conducted. Some of the observations made from this field trial were also used for developing curing specifications.

Table 3. *Observations made from field trial FT-3*

Time (after placement of HPC)	*Depth, quality of green cutting and exposureof aggregates*			
	Surface Retarder-A		*Surface Retarder-B*	
	Withretarder	*Withoutretarder*	*Withretarder*	*Withoutretarder*
8 hours	5-7mm, uniform	1-2 mm, non uniform	5-7 mm, uniform	1-2 mm, non uniform
12 hours	2-3 mm, non uniform	1 mm, non uniform	4-5 mm, uniform	1 mm, non uniform
16 hours	1-2 mm, non uniform	No exposure of aggregates	4-5 mm, uniform	Non exposure of aggregates
24 hours	No exposure of aggregates	No exposure of aggregates	1-2 mm, uniform	No exposure of aggregates

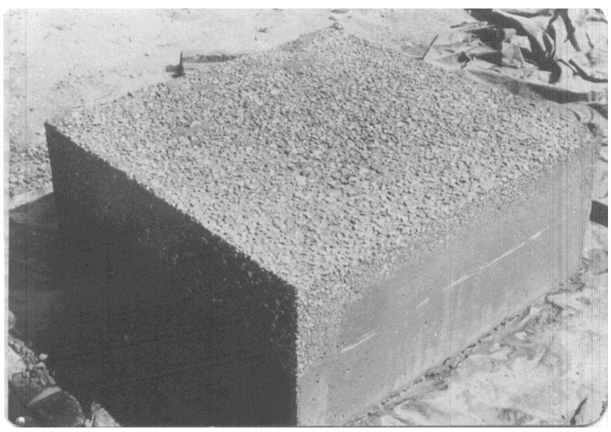


FIG. 5 GREEN CUTTING USING SURFACE RETARDER AND WATER JET (FT-3)

Field Trial FT-4

Schematic presentation of this field trial is given in Figure 6. HPC mix was cast in mould, of size 6.0 m (length) x 1.2 m (height) x 1.3 m (width), made by plywood. The mould was divided into two parts each having 3.0 m length. The first part contains reinforcements and cable sheaths arranged in the similar manner as that of the first pour of the ring beam, see Figure 3. The second part was kept as unreinforced. HPC was placed in the mould in layers of 300 mm. The data of the fresh concrete were as follows:

- slump of concrete 185 to 200 mm
- temperature of concrete 13 to 15°C
- ambient temperature 33°C
- duration of placement of concrete 1 hour

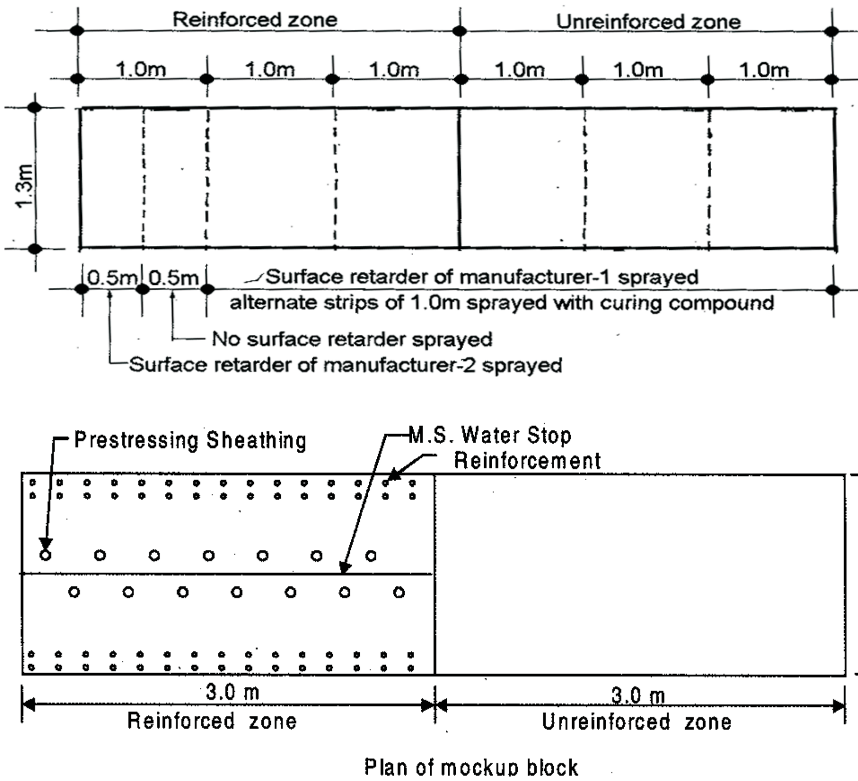


FIG. 6 MOCKUP FOR PLACEMENT SPREAD, AND COMPACTION (FT-4)

The observations made during the field trials were as follows:

i)	Spread of concrete in reinforced zone	1.5 m (before vibration), and 2.0 m (after vibration).
ii)	Spread of concrete in un-reinforced zone	2.0 m (before vibration), and 2.5 m (after vibration).

This field trial indicated that the spread of Kaiga-HPC was less in congested zone compared to that in less congested zone, which is obvious; it could however be placed without any difficulty in both the parts. The spread of concrete was mainly due to free fall from pipeline (effect of high slump and good slump retention capability of the mix) but appropriate vibration was necessary for proper compaction. It was also observed that in order to maintain same apparent workability, HPC requires approximately 50 mm higher slump than that of NSC. Even with excessive vibration, HPC mix did neither segregate nor cause any formation of laitance at the top of the concrete layer.

Surface retarder, procured from two manufacturers, was applied 15 to 30 minutes after completion of concreting of field trial-4. Water-based curing compound was sprayed half an hour after the application of the surface retarder in alternate strip of 1.0 m width starting from the second strip. Green cutting of the surface was done with water-air jet after 12 hours of concreting. Nature of surface appeared after green cutting is similar for both the portions where curing compound was applied and where it was not applied. Quality of surface appeared after green cutting was also found here too depended on the quality of surface retarder. Curing compound appeared to have very little impact on the green cutting. This observation confirms that for effective green cutting, curing compound is not required. Curing compound was found to stick with reinforcement and it was very difficult to remove afterwards indicating that curing compound is not suitable for the surface with dowel bars/reinforcement.

2.6 Curing

Present author described efficacy, various considerations, and approach to be adopted for curing of HPC [10]. HPC is cured on two stages – initial and final. Objective of HPC curing at initial stage is to prevent loss of moisture from fresh concrete without using water. In final stage, wet curing is resorted to. Experience of wet curing of concrete is substantial and a number of methodologies are available. Unfortunately, available information on initial curing is rather scanty.

Two important aspects of initial curing are method of initial curing and its duration. Method to be adopted for initial curing depends on the type of surface to be cured. For the purpose of specifying initial curing procedure systematically, Exposed surfaces of fresh concrete is categorized in to two types, Type –1 surface (to be exposed during service period) and Type-2 surface (surface of construction joints) [10]. Type-1 is further divided into Type 1A for surface with finishing and Type-1B without finishing.

It was observed from field trial FT-4 that water based curing compound is not efficient for initial curing. However, application of surface retarder could effectively prevent the formation of plastic shrinkage cracks on Type-2 surface till the time of final setting.

Field Trial FT-5

Purpose of this field trial is to study the phenomena of appearance of plastic shrinkage cracks on the surface of HPC exposed to various curing regimes without using curing compound. HPC was cast in a box of 3 m x 1 m covered area, which was divided into four blocks of 0.75 m width, using wooden shutter. Depth of concrete block was 1.0 m, while that of the box and the shutter was 2.0 m. This arrangement was made in order to simulate condition similar to ring beam in which depth of first two pours is less than the height of shutter. The schematic arrangements of this field trial are shown in Figure 7 and observations made from this field are given in Table 4.

Moisture loss is less when concrete is covered, which is obvious. Potential of appearing plastic shrinkage cracks was found to be higher in the finished surface compared to the one without finishing. Initial curing could be best effected when done by spreading PVC sheet in contact over the fresh concrete surface over laid by wet burlap. Initial curing was followed by final wet curing after final set of concrete.

3. CONCRETING OF CONTAINMENT STRUCTURES

Several new approaches have been adopted in the construction of containment structures using HPC. The construction work was carried out following well laid down construction methodology that was finalized incorporating the observations made from the mock up studies of field trials.

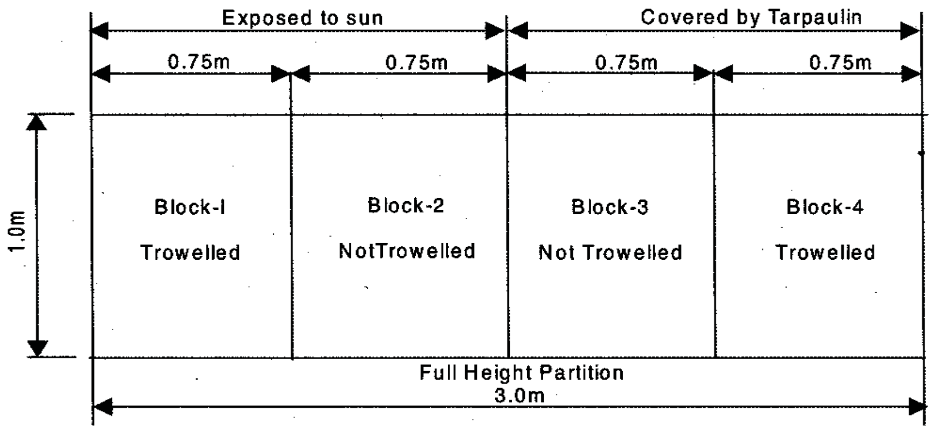


FIG. 7 MOCKUP FOR APPEARANCE OF PLASTIC SHRINKAGE CRACKS (FT-5)

Table 4. *Observation made from field trial 5*

Time# (hrs)	*Observations*			
	Case –1	*Case-2*	*Case-3*	*Case-4*
4.0	Water sheen disappeared	Water sheen disappeared	-	-
6.0	-	-	-	Water sheen disappeared
7.5	-	-	Water sheen disappeared	-
9.5	2 cracks appeared, l = 2.5-3 cm, w < 0.1mm			One crack appeared, l =2.5-3 cm, w < 0.1mm
12.75	Effective green cutting	Effective green cutting	Effective green cutting	Effectivegreen cutting

Note:
#Time after placement
Case 1: Block 1, surface exposed to sun and trowelled
Case 2: Block 2, surface exposed to sun and not trowelled.
Case 3: Block 3, surface covered by tarpaulin and not trawelled
Case 4: Block 4, surface covered by tarpaulin and trowelled.
Time of concreting: 12.00 hrs
In Case-4 surface was green cut with ease and cracks disappeared after green cut.

3.1 Production of Cool Concrete

Fresh concrete was produced following the two stage mixing method as described earlier using pan type mixer. The batching plant of TAPP-3&4 is computer based and fully automated. The batching plants of Kaiga-1&2 and RAPP-3&4 were also automated. Ice flakes in the mix replaced 90% of water. Aggregates were cooled by painting the aggregate bins with light reflecting paints and covering the bins and aggregate stacks against sunlight. Chilled air was passed through the aggregate storage bins by installing air handling unit. This has not been resorted to in TAPP-3&4 as ambient temperature is lesser than that of Kaiga-1&2 and RAPP-3&4 sites.

3.2 Transportation of Fresh Concrete

Transit mixers were used to transport concrete from batching plant to pumping point. The transit mixers drums were insulated with foam and coir and wrapped with jute. Cold water was sprayed on the transit mixer drum at the batching plant and at the pumping point. Concrete pipelines were also covered with jute bags and kept wet by spraying chilled water during concreting to reduce the rise in temperature during pumping of concrete. Average temperature was observed generally 13°C at production point and 19°C or less at placement point during actual construction. At TAPP-3&4, transit mixer drums were insulated using poly urathene foam (PUF).

3.3 Placing and Compaction

Concrete was placed using pump. Schematic diagram of concrete pipeline for pumping is shown in Figure 8, and actual arrangement in Figure 9. The HPC was placed in layers of 300 mm adopting built up scheme. Placement scheme was developed based on the observations of field trial FT-4. Adequate arrangement was made for proper vibration of the concrete. Vibrators were qualified and identified (with reference to the locations of pouring) for this purpose.

Based on the results of field trial FT-1 whenever slump of concrete was found to be less than 140 mm, secondary dosage of superplasticizer was added in Kaiga HPC. A re-dosing tower was installed at Kaiga near the pumping point to dispense the superplasticiser into the transit mixer and the drum of the transit mixer was rotated at a high speed for three minutes to ensure proper mixing of re-dosed superplasticiser with the concrete. At RAPP-3&4 and TAPP-3&4; dosage of admixture was increased at the production point to ensure sufficient workability upto placement time and avoiding the redosing of admixture.

3.4 Shuttering

Engineering of appropriate formwork shuttering for NPP containment structure is of immense importance. Comparatively higher rigidity of formwork is essential to avoid out of shape concrete surface after hardening, as HPC is less amenable to post concreting repair. Moreover field trial FT-2 indicated that right supporting arrangement is necessary for proper concreting the curved structural concrete element with top shutter. In addition to achieve good quality of curing, leak proof shuttering is also needed.

Concrete structures of circular/ring shape are more susceptible to shrinkage deformation than that having other structural shapes. In addition, HPC mix generally has low water cement ratio, and it being richer mix with high cement content liberate more heat per unit volume during hydration process. These make HPC structure more vulnerable to shrinkage, especially at early stage. Following measures was found effective in minimizing the effect of shrinkage on cylindrical wall at early stage,

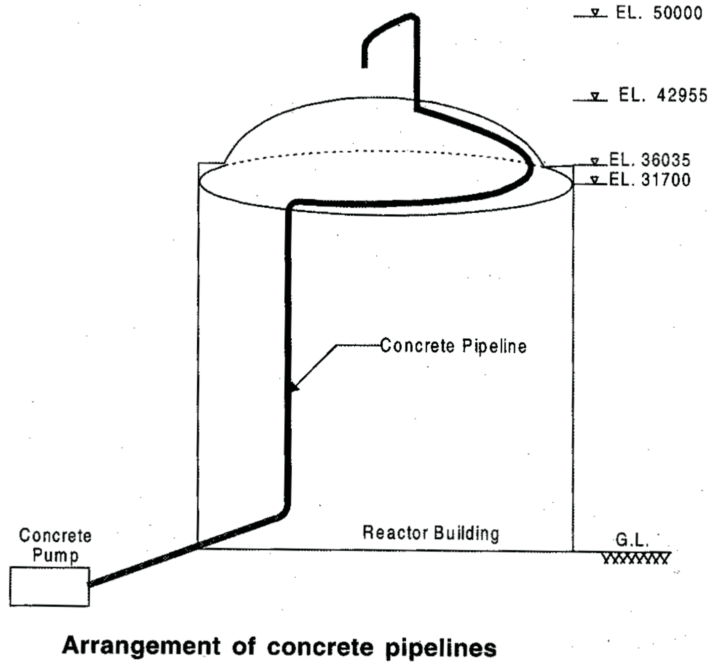


FIG. 8 SCHEMEATIC DIAGRAM OF CONCRETE PUMPING

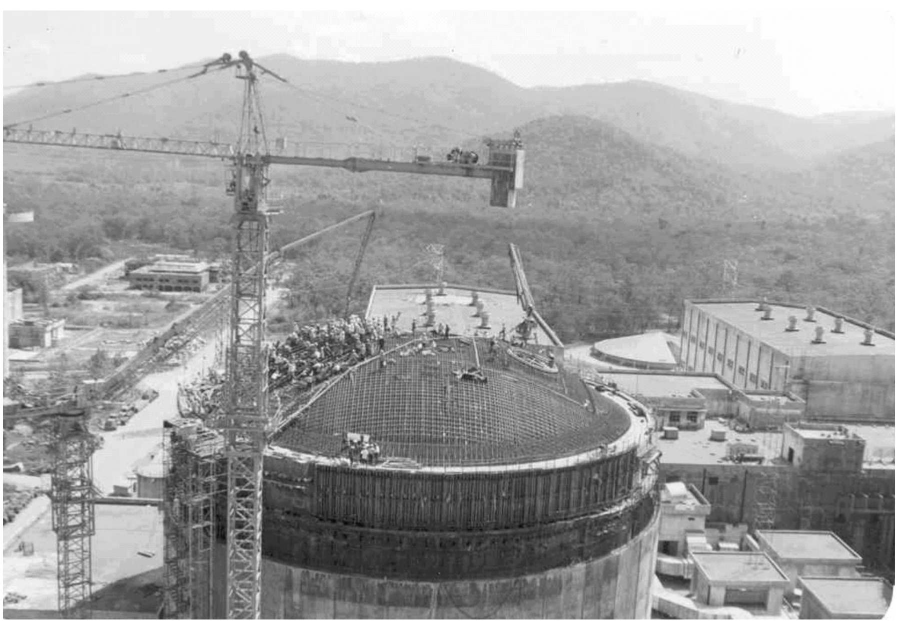


FIG. 9 PUMPING OF CONCRETE (KAIGA–2 IC DOME)

i) Application of water resistant paint on inner surface of shuttering so that shuttering material does not absorb water from fresh concrete.

ii) Stripping of shuttering after 48-72 hours of casting (in case of peak temperature due to heat of hydration and ambient temperature is expected to be more than 20°C). If such difference is very high, external surface of shuttering better be insulated suitably.

iii) Use of wire mesh in cover region if the clear cover is more than 20 mm.

3.5 Curing

Considering the dry and hot ambient conditions, a number of measures were taken to protect fresh HPC such as insulating the shuttering, covering the fresh concrete, placement of fresh concrete during cool hours of the day and night.

For initial curing of Type-1 surface of Kaiga-2 IC dome, curing compound was first applied on the exposed surface and allowed to set for some time. Opaque color plastic sheet was spread over the fresh HPC contacting the exposed concrete surfaces having curing. Wet burlap was spread over the plastic sheet for reducing the temperature of concrete surface. Finishing work of Type-1A surface was completed prior to initial setting of the HPC. It was observed that curing compound did not have significant effect on the initial curing at Kaiga-2. Based on this experience, curing compound was not used in RAPP-3&4, Kaiga-1 and TAPP-3&4; only opec color PVC sheet overlaid by wet burlap was used.

Based on the observations of field trial FT-5, it was concluded that providing shed using opaque color PVC sheet after spraying of surface retarder and subsequent green cutting after the final setting of concrete was found to be a practical approach for the initial curing as well as the treatment of Type-2 surface. Surface retarder was mixed with die in order to ensure uniform spraying by visual inspection. In hot and dry ambient condition, fog spray may be resorted to maintain required humidity around the surface of the fresh HPC. Fog spray worked satisfactorily to negotiate the harsh ambient condition (maximum temperature of about 42°C and humidity about 62%) at RAPP-3&4.

Above procedure of curing for both Type-1 and Type-2 surfaces of IC Dome at Kaiga-1&2 and RAPP-3&4 was found quite satisfactory. Initial curing was continued about 8 to 10 hrs. In TAPP-3&4 similar procedure of curing was adopted.

Conventional methods like pounding water, using wet burlap and sprinkling water were adopted for curing of silica fume based HPC at Kaiga-1&2, RAPP-3&4 and TAPP-3&4. Duration of about 10 days for wet curing is sufficient for HPC. Wet curing of 10 days was adopted in all these projects for HPC

4. QUALITY ASSURANCE

Each stage of work, starting from the characterization of ingredients to curing, was carried out following a stringent quality assurance program.

Moving average chart for compressive strength for the samples taken during the construction of the IC dome and cured under laboratory and field conditions is given in Figure 10 [8]. Variation of

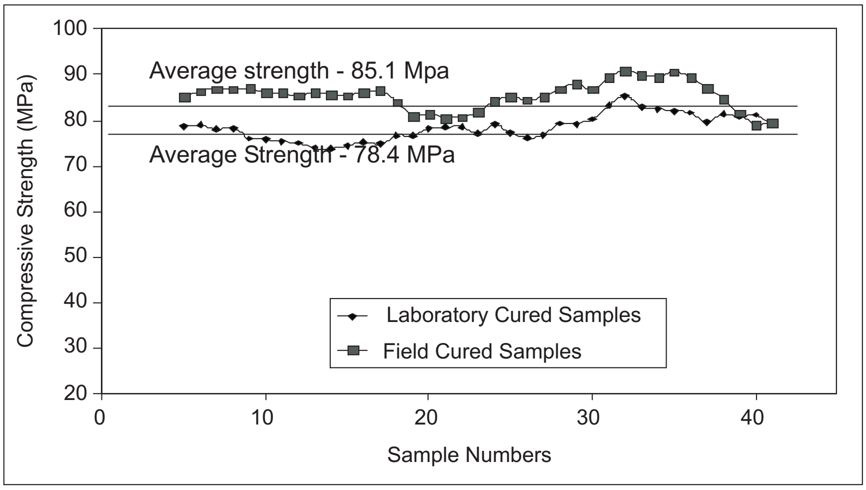


FIG. 10 MOVING AVERAGE FOR KAIGA 2 IC DOME FOR COMPRESSIVE STRENGTH FOR SAMPLES CURED UNDER LABORATORY CONDITIONS AND FIELD CONDITIONS

compressive strength with sample number is reasonably flat without any peak for samples cured in both conditions. Similar observations were made from laboratory tests. This confirmed the consistency in quality of the system starting from characterization of ingredients, production of fresh concrete to placement, compaction and curing. Interesting observation here is that the strength of site-cured concrete is always higher than that of standard at laboratory cure concrete. This may be due to higher site curing temperature than the standard curing temperature in laboratory.

5. SUMMARY

1) A number of new and innovative measures have been adopted in the construction of containment structure of Indian NPP (IC Dome in case of Kaiga-1&2 and RAPP-3&4 and complete containment structure of TAPP-3&4) using HPC. Construction methodology of containment structures of Indian NPP is developed keeping in mind three basic components of mechanism of HPC, i.e. reaction mechanism, physical process, and curing.
2) Laboratory tests and field tests were carried out to assess primarily the workability, strength and permeability under all probable conditions. Mock-up studies for field trials were carried out to examine the constructibility resulting in developing the construction methodology. Based on the results of field tests and outcome of mock-up studies, a few measures were adopted in the construction methodology.
3) For controlling heat of hydration, cooled concrete was produced at temperature of 13°C to 15°C. Protection of aggregates from hot ambient condition ...... Pre-cooling of aggregates in addition to replacement of water with ice flakes was found suitable in controlling the

temperature of fresh concrete. Also insulation of transit mixers and concrete pipelines helped in controlling the rise in temperature during transit and placing by pumping.

4) HPC requires higher slump and slump retention characteristics than that of normal strength to maintain the same apparent workability for pumping and spread of concrete. This was achieved either by re-dosing of retarder at the pumping point or by using higher doses of chemical admixture during mix proportioning. Due to the fact that HPC is not prone to segregation, it can be placed and compacted even in very congested areas. Vibrators were qualified and identified (with reference to the location of pouring) to ensure proper compaction of concrete.
5) HPC is required to be cured in two stages; initial curing and final (wet) curing. Initial curing of HPC for Type-1 surface could be best effected when done by spreading PVC sheet in contact over the fresh concrete surface overlaid by wet burlap. Final (wet) curing followed initial curing after the final set of concrete.
6) Application of surface retarder followed by high-pressure air water jet green cutting was found to be an effective method for initial curing of surface for extended reinforcement (Type-2 surface) construction joint preparation.
7) Application of water resistant paint on the inner surface of the shuttering, stripping of shuttering after 48-72 hours of casting and use of wire mesh in clear cover area (in case clear cover is more than 20 mm) are effective measures to minimize the shrinkage effect at early age of cylindrical wall of the containment structure.

ACKNOWLEDGEMENTS

This paper is text version of the presentation titled, " Construction of Containment Structures of Indian NPP using HPC" made in ACI Centennial Convention held in Washington D.C. on 14th –18th March 2004, which was co-authored with Shri Amit Mittal, formerly ACE(QA-Civil), TAPP-&4, NPCIL. The HPC was developed at the project sites of Kaiga-1&2, RAPP-3&4 and TAPP-3&4 with the participation of experts from Atomic Energy Regulatory Board and Nuclear Power Corporation of India Ltd. All trial mixes and laboratory tests were carried out at the concrete testing laboratories of the project sites. The execution work of the field trials and mock-up studies were carried out by M/s. ECC of L&T (at Kaiga-1&2 and TAPP-3&4) and M/s. HCC (at RAPP-3&4). Author thankfully acknowledge the contribution of the experts and supporting staffs of these organizations. The author also acknowledges the permission given by AERB to publish the paper.

REFERENCES

1) Basu Prabir C. (2001), "High Performance Concrete: Mechanism and Application", ICI Journal, Vol.2, No.2, April – June 2001
2) Basu Prabir C. (...........................), "High Performance Concrete", ...
3) Neville A.M., "Properties of Concrete", 4th edition, ELBS with Longman, Singapore.
4) Mehta P.K., Monterio Paulo J.M. (1997), "Concrete; Microstructure, Properties and materials", Indian edition, Indian Concrete Institutes, Chennai, India.

5) Monterio P.J.M., Mehta P.K. (1986), “Improvement of the Aggregate Cement Paste Transition Zone by Grain Refinement of Hydration Product”, Proceedings of the VIIth Int. Congress on the Chemistry of Cement, Rio-de-Janeiro, Vol. 2, pp. 433-437.
6) Basu Prabir C. (1999a), “Performance Requirements of HPC for Indian NPP structures”, The Indian Concrete Journal, Vol, 73, pp. 539-546.
7) Mittal Amit, Basu Prabir C. (1999), “Development of HPC for PC Dome of NPP, Kaiga”. The Indian Concrete Journal, Vol. 73, No. 4, pp 548-556.
8) Mittal Amit and Kamath Radhakrishna R (1999), “Properties of HPC for PC dome of NPP, Kaiga”, The Indian Concrete Journal, Vol. 73, No.9, pp. 561-568.
9) Jagdish H.T., Ram Mohan V, Suresh M.S., Ramesh Rao and Kamath Radhakrishna R (1999), “Field trials of HPC for PC dome of NPP, Kaiga”, The Indian Concrete Journal, Vol. 73, No.9, pp. 571-579.
10) Basu Prabir C., (1999b) “Curing of High Performance Concrete, Indian Experience”, March 1999, Vol. 73, No.3, pp. 342-348.
11) Bapat SG., Agarwal Neeraj (1999), “Experience in the Development and Production of HPC for RAPP-3&4”, The Indian Concrete Journal, Vo.73, No.9, pp. 582-586.

Current Trends in Engineering Practice Vol. II
Editor: Sneh Anand

FLY ASH CONCRETE

PRABIR C. BASU*

1. INTRODUCTION

Fly ash is produced by burning of coal or lignite in boiler of a thermal power plant [1,2,3]. It is a fine-grained mineral admixture with pozzolanic characteristics. ASTM C 618-03 [3] defines the fly ash, in similar way of IS 3812 [1,2], as finely divided residue that results from the combustion of ground and powdered coal and that is transported by flue gas. SEM images reveal that fly ash consists of either solid and/or hollow spheres, fig.1. Fly ash is most probably best-suited mineral admixture to be used as supplementary cementitious material in concrete [4]. Fly ash concrete is the concrete mix in which fly ash is incorporated as supplementary cementitious material [5]. This class of concrete is also termed as ‘fly ash based concrete’ or ‘fly ash incorporated concrete’.

Use of fly ash in concrete has many beneficial effects that would contribute towards the sustainable development of India,

(a) Producing concrete of better rheology, higher strength and enhanced durability,
(b) Saving of energy in the production of OPC,
(c) Preservation of limestone and coal reserves,
(d) Minimizing greenhouse gas emissions associated with manufacturing of OPC, and
(e) Environment friendly and economical disposal of millions of tones of fly ash.

Considering the above benefits, fly ash may be viewed as a resource material.

When fly ash is used as a supplementary cementitious material, it is generally incorporated to replace cement. In low volume fly ash concrete (LVFAC), the cement replacement level (CRL) is in the range of 10% to 30 % by mass of total cementations material (c_m. The Advance Concrete Technology Group at CANMET developed high volume fly ash concrete (HVFAC) [5]. This class of concrete has low water content (water-cementitious material ratio, w/b <0.4), possible due to use of weight is replaced with siliceous fly ash (SFA) of IS 3812, or Class F fly ash of ASTM-618.

There are plenty of applications of LVFAC in Indian construction industries, where as use of HVAFC is not much. Efforts are being made now for constructing structures with HVFAC mixes using Indian ingredients, cement and fly ash. Present paper discusses different aspects of fly ash

*Civil & Structural Engineering Division, AERB, Mumbai

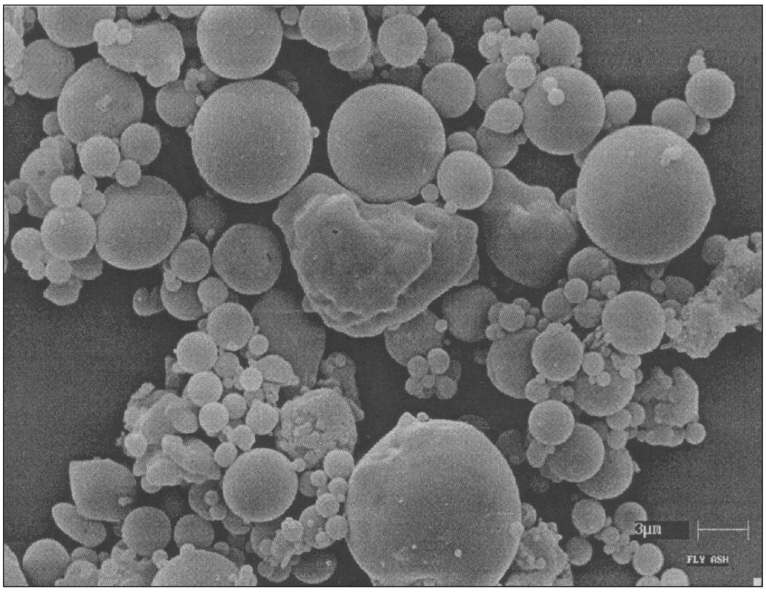


FIG. 1 MICROGRAPH OF FLY ASH PARTICLES

concrete covering approach for characterization of fly ash, incorporation of fly ash in concrete mix; mix proportioning of fly ash concrete and construction aspects. Paper also includes examples of application of fly ash concrete, both LVFAC and HVFAC, in Indian construction industries, especially in nuclear power plants.

2. CHARACTERIZATION OF FLY ASH

Results presented in reference-6 indicate that Indian fly ash possesses all requisite characteristics needed for producing HVFAC of acceptable quality. The independent work of Fournier *et al.* and Sivasundaram confirms this [7,8]. Basu and Saraswati examined the suitability of existing IS codes for engineering of HVFAC [9]. Bureau of Indian Standards (BIS) has published the specifications of pulverized fuel ash, IS 3812: 2003 in two parts [1, 2].

- Part-1: for use as pozzolana in cement, cement mortar and concrete
- Part-2: for use as admixture in cement mortar and concrete.

This code can be adopted for characterization of fly ash depending on its use as pozzolana or mineral admixture. Fly ash is categorized into two types, namely, siliceous fly ash (SFA) and calcareous fly ash (CFA).

Siliceous Fly Ash (SFA): Fly ash with reactive calcium oxide less than 10% by mass falls under this category. Such fly ashes are normally produced from burning anthracite or bituminous coal and have pozzolanic properties.

Calcareous Fly Ash (CFA): This category of fly ash contains reactive calcium oxide not less than 10% by mass; the content could be as high as 25%. Such fly ash is normally produced from lignite or sub-bituminous coal and has both pozzolanic and hydraulic properties. Excess calcium oxide of CFA will combine with silica and alumina portion of the ash; as a result, there will be less of these compounds to react with the lime liberated by the primary hydration of ordinary Portland cement.

ASTM C 618-03 categorises fly ash into two classes; class F and class C, which are equivalent to SFA and CFA respectively of IS3812.

IS 3812 allows the use of fly ash in concrete both as pozzolana (Part-I) and as mineral admixture (Part-II). Terminologies, pozzolana and mineral admixture, are sometimes ambiguously used as synonymous to each other. It is important to clearly understand the difference between these two types of usage, as there exist certain variations in specification given in the two parts of IS 3812: 2003.

As per ASTM C125-02, admixture is a material other than water, aggregates, hydraulic cementitious material, and fibre reinforcement that is used as an ingredient of a cementitious mix to modify its freshly mixed, setting or hardened properties and which is added to the batch before or during its mixing. Chemical admixtures like superplasticiser are water-soluble. The term mineral admixture has been used to refer to different types of water insoluble finely divided materials. These materials are not similar and it is not correct to group them under a single term. Mineral admixtures may be reactive or inactive in terms of cement hydration. Pozzolans or cementitious materials are reactive mineral admixtures, while finely divided aggregates are inactive. Pozzolan is siliceous or siliceous and aluminous material, which itself possesses little or no cementitious value but will, in finely divided form and in the presence of moisture, chemically react with calcium hydroxide at ordinary temperatures to form compounds possessing cementious properties

Both the parts of IS 3812 specify detailed requirements on physical and chemical characteristics of fly ash. Amongst these, two requirements, the one on loss of ignition and the other one on the fineness are important. Requirement on LOI has significant bearing on durability; higher LOI could impede durability. Therefore, strict adherence to the requirement of LOI ($\leq$ 5%) is very important. Fineness has influence on pozzolanic characteristics of fly ash thus improving strength and durability of the concrete mix.

IS 3812 specifies 34 percent retention of particles coarser than 45 ì in case of fly ash used as pozzolanic material (Part-I) and 50 percent retention for its use as mineral admixture (Part-II). Additionally, the standard also specifies the average fineness requirement as 320 sq.m/kg for the use as pozzolana (Part – I) and 200 sq.m/kg for mineral admixture. Contribution of particles coarser than 45 ì to secondary hydration by means of pozzolanic reaction is not significant. Their incorporation in concrete mix may be viewed as replacement of sand. Therefore, requirement with respect to particle size of 45 ì or less is more important for pozzolanic action than the average fineness of fly ash.

3. INCORPORATION OF FLY ASH IN CONCRETE MIX

The resource material fly ash is introduced into concrete in one of the following two ways.

- Through blended cement, Portland pozolan cement (PPC), manufactured by intergrinding fly ash with OPC clinkers, and

- As supplementary cementitious material at the concrete mixing batching plant.

The issues related to use of PPC vs. fly ash as partial replacement in the batching plant was debated in length during the HVFAC conclave organized by confederation of Indian Industries (CII) and the following consensus were arrived at [10],

(i) In *unorganized sector* where hand mixing is employed for production of concrete and adequate resources are not available for proper quality control, the use of PPC is recommended. To encourage more use of fly ash in this sector, PPC with higher percentage of fly ash (near to 35%) should be made available.

(ii) In the *organized sector* where adequate quality control regime is implemented, use of HVFAC is recommended. Even though, PPC could also be used in this sector.

The use of blended cement is easier, since it is free from the complications of batching additional materials at the construction site and may ensure more uniform control. Upper limit of 35% on fly ash content in PPC has been incorporated recently in BIS documents [11]. Further increase in this limit could be recommended after gaining experience on the behaviour of PPC with 35% fly ash in Indian scenario. After extensive deliberations it was agreed that the existing provisions related to the quantity of fly ash to be incorporated in PPC, as per Cl.5, IS 1489 Part I: 1999, Amendment No.3, July 2003, i.e. minimum 15% and maximum 35% should be retained for the present time [12].

The addition of fly ash at the concrete mixing stage is flexible and allows exploitation of higher quantities of fly ash as an ingredient of concrete. However, for a concrete mix in which fly ash is incorporated at batching plant, it is not essential to put any limitation on the quantity of fly ash in the mixture proportions. Specific concrete mixture proportions could be developed with quantity of FA as desired by the designer depending on the requirements of the design and construction schedules.

There is a school of thought which subscribes to the view that IS codes do not allow more than 35 percent of fly ash in concrete. No IS code limits the quantity of fly ash to be used either as pozzolana or mineral admixture for replacement of OPC in proportioning the mix for structural concrete. However, IS 1489 (Part I):1991 specifies 35 percent as the upper limit of fly ash for blended cement, (Portland pozzolanic cement). The confusion might have originated from the footnote of Table 5 of IS 456. The first footnote of the table states,

> "Cement content prescribed in this table is irrespective of the grades of cement and it is inclusive of additions mentioned in 5.2. The additions such as fly ash or ground granulated blast furnace slag may be taken into account in the concrete composition with respect to the cement content and water-cement ratio if the suitability is established and as long as the maximum amounts taken into account do not exceed the limit of pozzolana and slag specified in IS 1489 (Part-I) and IS 455, respectively".

Intent of the footnote is only related to determination of the minimum cement content, maximum water-cement ratio and minimum grade of concrete for different exposures with normal weight aggregates of 20 mm nominal maximum size. As per this footnote, contribution of fly ash should not be taken more than 35 percent of the total cement content (which is OPC + fly ash) in calculating the minimum cement content when CRL by fly ash is more than 35 percent. For example, in case of severe exposure condition, minimum grade of concrete and corresponding minimum cement content specified in Table 5, IS 456 is M30 and 320 kg/m^3, respectively. If HVFAC mix is to be proportioned with CRL of 50 percent by fly ash, then cementitious material (OPC + fly ash) content of 377.00 kg/m^3 is required to satisfy the minimum requirement of 320-kg/m^3 cement (OPC: 0.5X377.00+fly ash: 0.35X377.0).

As inter-grinding of FA and clinker is often adopted in manufacturing PPC, and general requirement of fineness for PPC is to be satisfied as per IS 1489:1999; further limitations on fineness, lime reactivity, pozzolonic activities of FA to be blended in PPC is unnecessary as long as LOI satisfy the codal requirement and bottom or pond ash is not used. Outcome of research in this area indicates that grinding of coarse fly ash; particle size greater than 45m, to finer size enhances the pozolanic characteristics of fly ash [13,14,15]. This opens up the opportunity of using effectively fly ash collected in the lower fields of ESP of a thermal power plant in PPC leaving the finer material of higher field for using in the batch mode.

4. MIX PROPORTIONS

A number of standard mix proportioning methods/guidelines are available for conventional normal strength concrete (NSC). IS10262 is used in mix proportioning of NSC in India [16]. Mix proportioning of HVFAC is a more critical process than that of NSC in view of the high fines content and the low w/c_m ratio. Jiang and Malhotra suggested a mix proportioning method, based on the combination of empirical results and the absolute volume method [17]. It is felt that the procedure adopted for mix proportioning of HVFAC is similar to that of NSC, except that in the case of the former, adjustments have to be made for the chemical admixture, and mineral admixture content.

For mix design of fly ash concrete, incorporating low or high volume of fly ash as cement replacement, the basic principles of low water content and low w/c_m is to be maintained. For HVFAC, w/c_m is limited to 0.4. In order to achieve high performance with excellent durability characteristics, it is essential that high slump and also good rheology be achieved by using superplasticizer of appropriate dosage.

The water demand for a given value of workability and good rheology is controlled by particle size distribution, particle packing effect and smoothness of surface texture. Fly ash particles are spherical in shape that has positive effect on reducing the water requirement vis-à-vis rheology of concrete mixtures. Depending on the quality of fly ash and the amount of cement replaced, up to 20% reduction in water requirements can be achieved [17].

Siliceous fly ashes retard the setting of cement. The small size and the essentially spherical form of low calcium fly ash particles influence the rheological properties of cement pastes, causing a reduction in the water requirement or an increase in workability compared with that of an equivalent part without fly ash. As Davis et al [18] noted fly ash differs from other pozzolans, which usually increase the water requirement of concrete mixtures. According to Owens [19], the major factor influencing the effects of ash on the workability of concrete is the proportion of coarse material (> 45 m) in the ash. Concrete using fly ash generally shows reduced segregation and bleeding and is more satisfactory than plain concrete when placed by pumping [20]. In line with the improved rheological properties and as a result of fine particulate content, some fly ashes give a markedly improved finish. Consideration of strength and durability in proportioning fly ash concrete mix has been discussed in reference 21.

Basu and Saraswati developed fly ash mix proportion with CRL 20% to 70% [6]. More than 100 trials were carried out to examine suitability of Indian ingredients, especially the cement and fly ash for developing HVFAC mixes for Indian conditions. These mixes were proportioned with total cementitious material content of 400kg/m^3, 450 Kg/m^3, 500 kg/m^3, 525 kg/m^3 and 550 kg/m^3, and w/c$_m$ ranging from 0.3 to 0.375. In addition, a few control concrete mixes (CCM), without fly ash were proportioned and examined for better understanding the properties of HVFAC. Slump was taken as the parameter to examine the workability of fresh concrete though it does not cover all aspects of this attribute. Both cube compressive strength (f_c) and split tensile strength (f_{st}) were studied. The durability aspect was assessed by rapid chloride permeability test (RCPT).

Tables 1 and 2 contain some of the mix proportions and their properties respectively. The mixes are proportioned using the information available in existing literature on HVFAC mixes with varying cement content, varying content of pozzolanic admixture, and varying water-binder ratio were considered. The quantities of aggregates were determined following the available mix proportioning method. The fines to total aggregate ratio were taken as 36.4%. Superplasticizer dosage was adjusted during actual trials. All fly ash concrete mixes were cohesive, non sticky and no difficulties was encountered in handling. Initial slump increases, in general, with fly ash content for a given total quantity of cementitious material, which is expected [22]. The strength parameters, f_c and f_{st} at 28 days decrease with increase in fly ash content for a given total quantity of cementitious material. This observation has support of other published results [22, 23, 24]. RCPT value increases with fly ash content at 28 days indicating increase in permeability at initial age, which is also expected [22].

The mixes are FA02, FA07, FA12, FA20 and FA21, which have total cementitious material content of 500 kg/m^3 and CRL of 20%, 33.3%, 50%, 60% and 70% respectively. For all these mixes, w/c$_m$ is 0.3 and superplasticizer dosage is within 0.7% - 1.0%. The first two mixes are LVFAC and remaining three is HVFAC. Reference CCM is C01.

Initial slump of all fly ash concrete mixes is higher than that of CCM excepting FA02, fig.2.Fly ash concrete mixes with 33.3%, 60% and 70% CRL always exhibited better slump retention than CCM. Slump retention of FA02 was poorer than the CCM during entire period of 60 minutes. HVFAC with 50% CRL by fly ash possessed lower slump after 30 minute. The CCM, C01, contained both water and superplasticizer of higher quantity than the mixes with fly ash. This could be the reason of CCM exhibiting better slump and slum retention compared to that of FA02 and FA12 respectively.

Cube compressive strength (f_c) of both LVFAC and HVFAC is lower than that of CCM at early age – 7 and 28 days, but strength gain with age is predominant in case of both types of fly ash concrete mixes, fig.3. CCM exhibited insignificant strength gain after 28 days. This observation is common experience for any control concrete mixes having superplasticizer and OPC of higher fineness and C_3S content. Similar observation is made from fig.4 in case of split tensile strength (f_{st}). Lower strength for fly ash concrete compared to CCM is expected. The quantity CSH gel generated from primary hydration of OPC is more than that generated from secondary hydration due to pozzolanic action of fly ash. Higher the fly ash content in concrete mix, lower would be the CSH gel and hydrated cement paste resulting lower strength.

Table 1. *Details of fly ash concrete mixes*

Mix.	*Cement*	*Fly ash*	*Coarse*	*Fine*	*W/b*	*CA (%)*
#C01	500.0		1199	686	.300	.300
#C02	500.0		1177	673	.325	.325
#C03	500.0		500	660	.350	.350
#C04	450.0		450	694	.350	.350
#C06	400.0		400	708	.400	.400
!FA22	200.0	$350.0 (%63.6)	1071	612	.300	0.70
!FA09	300.0	225.0 (42.8)	1124	643	.300	0.90
*FA02	400.0	100.0 (20.0)	1178	674	.300	0.70
*FA04	400.0	100.0 (20.0)	1156	661	.325	0.80
*FA05	400.0	100.0 (20.0)	1134	641	.350	0.50
*FA07	333.3	166.7 (33.3)	1164	666	.300	0.70
!FA12	250.0	250.0 (50.0)	1147	656	.300	1.00
!FA20	200.0	300.0 (60.0)	1136	650	.300	0.70
!FA21	150.0	350.0 (70.0)	1126	6434	.300	0.80
*FA10	300.0	150.0 (33.3)	1203	688	.325	1.25
!FA14	200.0	250.0 (55.5)	1202	687	.300	1.25
!FA15	200.0	250.0 (55.5)	1182	676	.325	0.50
!FA18	200.0	250.0 (55.5)	1142	653	.375	0.50
!FA19	200.0	250.0 (55.5)	1162	664	.350	0.60
!FA16	200.0	200.0 (50.0)	1232	704	.350	1.25
!FA17	200.0	200.0 (50.0)	1214	694	.375	0.90

Note:
#Normal or control concrete mix (CCM): Mix without fly ash.
!LVFAC mix: CRL @ 30%.
*HVFAC mix: CRL $\geq$ 40%.
$Total quantity of fly ash in Kg.
%Cement replacement level (CRL) in %.

The gain in both types of strength, f_c and f_{st}, with ages increases with the increase in CRL by fly ash. Ratio of f_c at 28 days to that at 120 days is 1.1 for 20% replacement, and about 1.78 for 70%. Another important observation is that the rate of strength gain for both LVFAC and HVFAC decreases after the age of 56 days. Both f_c and f_{st} of CCM and LVFAC at 120 days are almost same. This clearly indicates for a given binder content strength of CCM and LVFAC (20% and 33.3% CRL) tends to equal at a latter age. f_c of HVFA mixes with CRL of 60% and 70% were almost same after 28 days of age. Trend of similar strength gain was exhibited by these two mixes, e.g. HVFAC mix having 50% CRL at different ages lies between the cluster containing LVFAC and HVFAC with CRL of 60% and 70%. The observations made on the strength gain of fly ash concrete are supported by existing publications [5, 23, 25, 26, 27].

Figure 5 contains the results of RCPT at different ages. Seven days RCPT value of mixes with fly ash is, in general, higher than that of control concrete; significantly higher in case of 60% and 70%

Table 2. *Properties of fly ash concrete mixes*

Mix.	*#Initial Slump(mm)*	**Strength (N/mm²)*		*RCPT Value*(Culomb)*
		f_c	f_{st}	
C01	150	82.7	4.78	1663
C02	195	77.8	4.33	1751
C03	220	72.2	4.04	1965
C04	165	65.2	4.16	2082
C06	145	54.1	3.87	2213
FA22	190	40.0	2.38	3816
FA09	180	67.0	4.22	1493
FA02	100	78.7	4.42	1048
FA04	190	67.0	3.94	1726
FA05	190	62.5	3.88	1578
FA07	175	76.4	4.31	1122
FA12	190	58.8	3.85	2672
FA20	180	41.1	2.92	2468
FA21	180	28.2	2.24	4069
FA10	150	65.0	3.54	1020
FA14	180	57.6	3.77	1757
FA15	165	56.7	3.68	2014
FA18	190	37.0	3.44	2476
FA19	170	48.1	3.96	1827
FA16	170	48.1	3.48	1827
FA17	175	36.2	3.44	2576

Note:
#Value at production point of batching plant
• 28 day's value

CRL. In case of 28 days, LVFAC mixes has lower RCPT value than that of CCM, but it is higher for HVFAC mixes. RCPT value for both LVFAC and HVFAC is lower than that of CCM from the age 56 days onwards. RCPT values of HVFAC having different CRL by fly ash tend to converge to a similar value with ages. Similar observation is made for LVFAC mixes also. The above observation indicates that the permeability of HVFAC decreases significantly with age and is lower than that of CCM after 56 days. The phenomenon of decreasing permeability of HVFAC is supported by the work of other research in this field.

The results in effect, would produce information on fly ash concrete having CRL 0% to 70%. The strength gain of fly ash concrete is substantial upto 56 days. For optimal utilization of this material, the acceptance criteria of fly ash concrete is better to be based on 56 days strength rather than 28 days strength. Canadian code adopted 90 days strength for acceptance of this class of concrete [28].

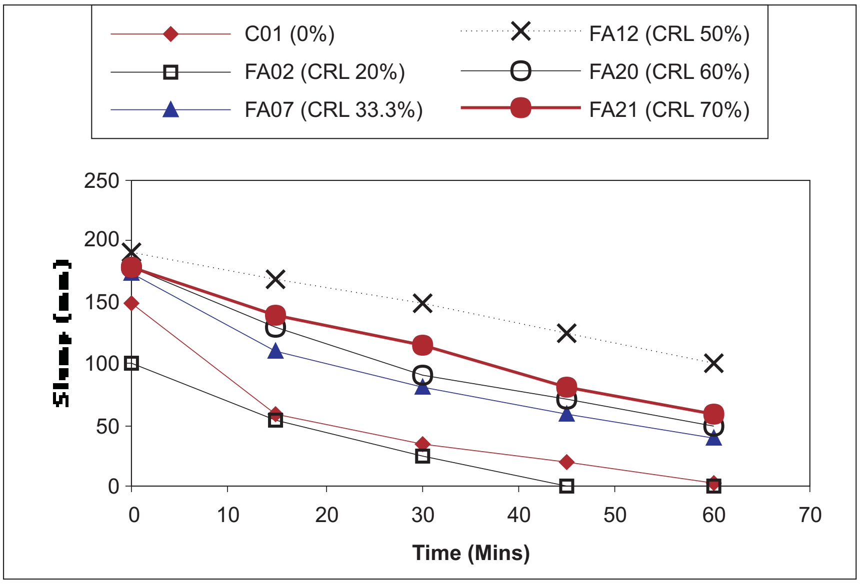


FIG. 2 VARIATION OF SLUMP WITH AGE FOR DIFFERENT CEMENT REPLACEMENT LEVEL BY FLY ASH

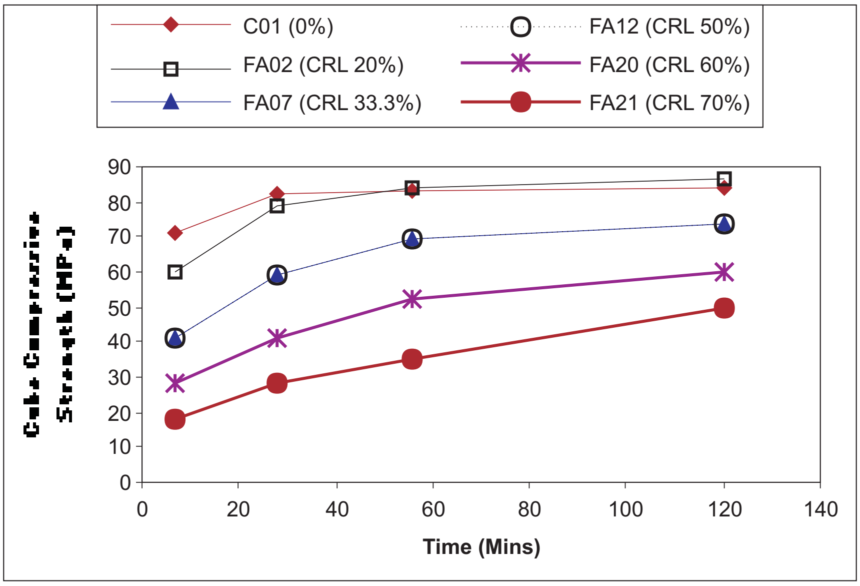


FIG. 3 VARIATION OF CUBE COMPRESSIVE STRENGTH WITH AGE FOR DIFFERENT CEMENT REPLACEMENT LEVEL BY FLY ASH

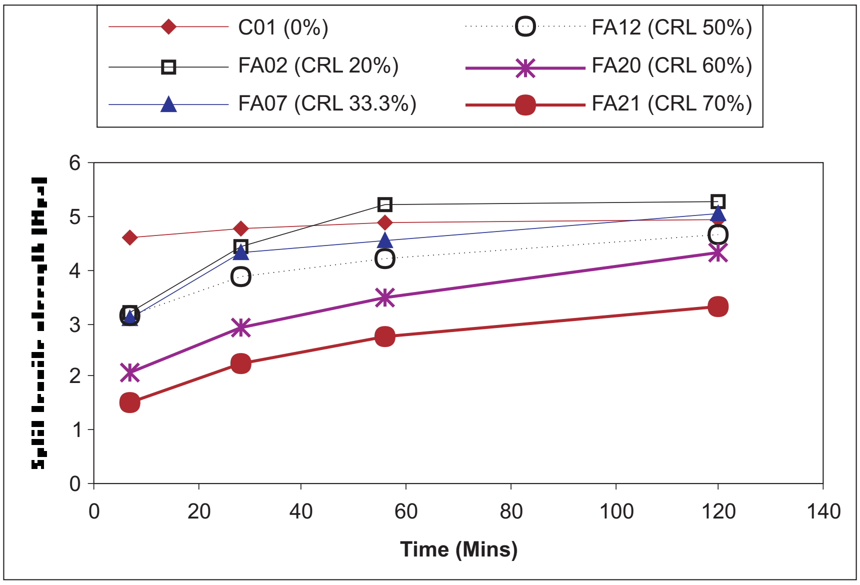


FIG. 4 VARIATION OF SPLIT TENSILE STRENGTH WITH AGE FOR DIFFERENT CEMENT REPLACEMENT LEVEL BY FLY ASH

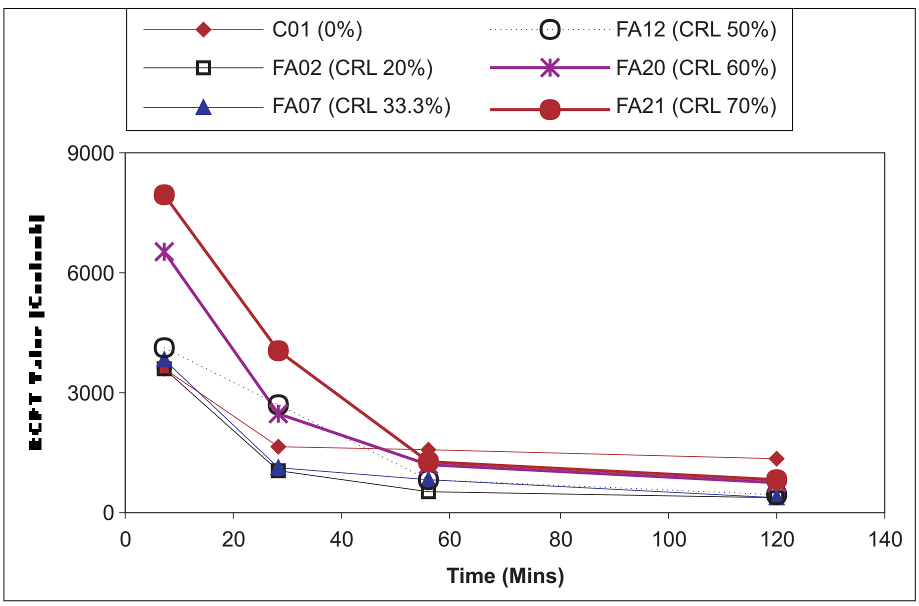


FIG. 5 VARIATION OF RCPT VALUE WITH AGE FOR DIFFERENT CEMENT REPLACEMENT LEVEL BY FLY ASH

5. CONSTRUCTION ASPECTS

The activities related to construction can broadly be grouped into the following:

- Batching and mixing methods
- Transportation and placement
- Curing
- Stripping of shuttering

5.1 Batching Mixing

It is generally accepted that weigh batching is essential for the production of fly ash concrete to maintain control on the use of cementitious material content and w/c_m in line with the specified mix proportion. In large applications, batching plants with a separate fly ash silo and weighing scale will ensure proper quality in rapid production of fly ash-based concrete, especially for HVFAC.

The mixer, which is successful for production of normal concrete, or CCM, is equally suitable for fly ash concrete. Hand mixing shall not be adopted when fly ash is incorporated at site during production of concrete. Rotary drum mixers that are capable of producing homogeneous OPC concrete can also give satisfactory result for fly ash-based concrete provided ingredients are batched by weight.

HVFAC normally has low w/c_m; therefore, use of a superplasticizer becomes necessary to achieve low water content. Mixing sequence adopted should also take into consideration for the proper dispensation and the mixing of a superplasticizer.

Requirements of fly ash storage are not as stringent as that for cement as SFA type of fly ash (ASTM Class F) is not reactive in presence of moisture. BIS specification is adequate for this purpose. In case of formation of lumps in fly ash, these will be broken during the mixing without any detrimental effect on the properties of concrete. The storage requirements for CFA type of fly ash (ASTM Class C) should be the same as for cement.

It is established that method of mixing and the time of agitation play a crucial role in governing the properties of the mix especially in case of HPC. Mixing method of fly ash concrete greatly depends on the batching plant especially the mixer. Result presented in reference-4 would provide useful guidelines in this context. A multi stage mixing sequence with varying time of agitation at each step is suitable for fly ash concrete. The mixes of Table-1 produced with laboratory version pan type mixer having 30-rpm blade speed adopting following multistage sequence,

i) Coarse aggregate, fine aggregate, cement, fly ash and 33% of total water were added to the mixer and allowed to mix for 60 seconds.
ii) Chemical admixture and 67% of total water were then added to the mix and the mixing was continued further for 210 seconds.

Total mixing time was 270 seconds.

5.2 Transportation

The specification of transportation and placement for HVFAC is similar to that of normal OPC concrete or CCM with lower w/c_m, and no special precaution is required.

5.3 Curing

Fly ash concrete is cured in two stages – initial curing and final curing. Initial curing should be started immediately after placing of concrete by covering surfaces with impervious sheet or by applying of curing compounds as in the case for concrete pavements by slip-forming methods. It is dry curing and helps in protecting the fresh concrete from the potential of plastic shrinkage crack by preventing moisture loss during early age. After final setting of concrete, water curing should be commenced. In case fly ash content is less than 35% and environment is humid, water curing must be continued for at least 10 days. Water curing shall be continued up to 14 days for dry environment or when fly ash content is more than 35%. This is inline with the requirement of IS456 [11].

5.4 Stripping of shuttering

Stripping of shutter is generally related to strength development. It was agreed that stripping of shutter for fly ash concrete shall be as per the requirements specified in IS456: 2000. Sometimes shuttering provides a good means of curing in sealed condition; for example soffit of slab. In cases where de-shuttering is adopted before the stipulated period of final curing, alternate curing measures shall be provided and be continued up to the stipulated time period.

6. APPLICATION

At SERC, Chennai, a demonstration building was constructed in the early 1970's utilizing fly ash in every part of construction. Concrete grades of upto M45 were developed using fly ash and cement available during that time. This building, even after three decades of exposure to coastal environmental conditions (Chennai), is in excellent condition compared to other building made of RCC in the CSIR campus. This building is testimony of how good and useful fly ash concrete could be in Indian condition, especially in coastal environment.

In India, PPC is in use nowadays. Experiences of using LVFAC by producing in batch mode are also considerable in India. In contrast, little information could be found on application of HVFAC in India.

Information on limited application of HVFAC in road pavement is available in literature [29, 30]. Recently HVFAC is being used in the construction of tremie seal and pylon pile cap of Bandra-Worli sea link, in Mumbai [31]. However, India is yet to start using HVFAC in reasonable quantity to reap its benefit.

Fly ash concrete, both LVFAC and HVFAC, had been used in the construction of Indian nuclear power plant structures [32]. Traditionally placed concrete (TPC) and self compacting concrete (SCC), incorporating fly ash with CRL ranging from 20% to 50%, were used in the construction of important structures like control building, induced draft cooling tower, pump house, turbine buildings, etc. of Kaiga Project, Units – 3&4 (Kaiga-3&4); Rajastha Atomic Power Project, Units – 5&6 (RAPP-5&6), Tarapur Atomic Power Projects, Units – 3&4 (TAPP-3&4). TPC is that class of concrete, which is compacted after pouring by vibration or other means of external energy. SCC gets compacted by its self-weight.

FIG. 6 PUMP HOUSE COLUMN (KAIGA-3&4)

FIG. 7 HIGHLY CONGESTED COLUMNS OF TURBINE BUILDING CONCRETED WITH HVFAC (RAPP-5&6)

FIG. 8 TG DECK OF TURBINE BUILDING OF UNIT 6 CONCRETED WITH HVFAC (RAPP-5&6)

FIG. 9 IDCT CONCRETED USING SCC (RAPP-5&6)

Figures 6, 7, 8 & 9 depict view of some of the structures of Indian recent NPPs constructed with fly ash concrete. About 1600 cum, 37600 cum and 200 cum of TPC type fly ash concrete were used in Kaiga-3&4, RAPP-5&6 and TAPP-3&4 respectively. The quantity of SCC type of fly ash concrete used in these projects was 4200 cum, 600 cum and 230 cum respectively. In addition, approximately 10000 cum of TPC type fly ash concrete were used as bedding concrete in RAPP-5&6.

7. SUMMARY

1. Use of fly is one of the efficient measures for sustainable development of country like India. Fly ash should be considered as resource material considering its beneficial effects of using it in the material concrete for construction industries.
2. Existing IS codes are suitable for characterization of concrete ingredients for HVFAC. Major observations on is 3812 (Part-I and II) : 2003 specifications for characterization of fly ash are,
   (a) The standard specifies suitable requirements for characterization of fly ash. Requirement mentioned about average fineness is not necessary and may be deleted.
   (b) Most important requirement for characterization of fly ash is to restrain the LOI to 5 percent.
3. Fly ash could be incorporated in concrete either through Portland pozzolana cement (PPC), or by using it as supplementary cementitious material replacing OPC in batching plant. It is recommended to use PPC in unorganized sector.
4. IS codes impose limitation of 35 percent on the maximum usage of fly ash in PPC but there exists no limitation on the quantity of fly ash in concrete mix, if it is incorporated separately in site batching.
5. High volume fly ash concrete (HVFAC) can be developed using commercially available Indian cement and fly ash with cement replacement level at least upto 70 percent.
6. The gain in strength with ages is higher in case of fly ash concrete. Such gain is more predominant in case of HVFAC than that of LVFAC. The rate of strength gain with age decreases for both LVFAC and HVFAC after 56 days.
7. From the point of view of economy, HVFAC should be characterized for 56-day strength. However, the mix should have required strength for construction purpose after the minimum period required for curing.
8. Fly ash concrete is cured in two stages – initial and final. Initial curing is done by covering the concrete surface with membrane or plastic sheet and continued, at least, till final setting time. Final curing is the conventional moist curing and shall be continued till 10 days for LVFAC and 14 days for HVFAC.
9. Shuttering can be stripped following the guidelines of IS456. However, alternate arrangement of curing till the specified period shall be arranged for the concrete surface that is covered by shuttering to achieve sealed curing. Example of such surface is soffit of slab.

REFERENCES

1. ………….., "Pulverised Fuel Ash – specification, Part-1: For Use as Pozzolana in Cement, Cement Mortar and Concrete", IS 3812 (Part 1): 2003, Bureau of Indian Standard, New Delhi, India.
2. ………….. "Pulverised Fuel Ash – specification, Part-2: For Use as Admixture in Cement, Mortar and Concrete", IS 3812 (Part 2): 2003, Bureau of Indian Standard, New Delhi, India.

3. …………….. "Standard specification for Coal fly-ash and raw or calcined natural pozzolan for use in concrete", ASTM C 618-03.
4. Basu, P. C., "High Performance Concrete", ...
5. Malhotra, V. M, and Ramezanianpour, "A. A.,Fly Ash In Concrete, 2nd edition,CANMET, 1994.
6. Basu, P. C., Saraswati, S., "High volume fly ash concrete with Indian ingredients", *The Indian Concrete Journal*, March 2006, Vol. 80, No. 3, pp. 37-48.
7. Fournier, b., Lu., D., Charland, J.P. and Li, J., "Evaluation of Indian fly ashes for use in HVFA Concretes - Part I": Characterisation, *The Indian oncrete Journal*, November 2004, Vol. 78, No. 11, pp. 22-30.
8. SIVASUNDARAM, V., Evaluation of Indian fly ashes for the HVFA concrete - Part II: Behaviour in Concrete, *The Indian Concrete Journal*, November 2004, Vol. 78, No. 11, pp. 41-50.
9. Basu, P., C., Saraswati, S., "
10. CII Doc.
11. …………., "Indian Standard: Plain and Reinforced Concrete – Code of Practice", IS 456, 4th revision, September 2000.
12. ………, 'Specification for Portland Pozzolana Cement : Part 1 Fly ash based', IS 1489 (Part I): 1991, Bureau of Indian Standard, New Delhi, India.
13. Riding, K. A., Garci Juenger, M. C., " Potential Use of Beneficiated Fly ash In High Performance Concrete", *High-Performance Cement-Based Concrete Composites, Material Science of Concrete,* Special Volume, 2005,The American Ceramic Society, Ohio, USA,
14. Bouzoubaa, B., Zhang, M. H., Bilodeau, A., Malhotra, V. M., " The Effect of Grinding on the Physical Properties of Fly Ashes and a Portland Cement Clinker", *Cement and Concrete Research*, Vol. 27, No. 12, pp1861-1874, 1997.
15. Bouzoubaa, B., Zhang, M. H., Malhotra, V. M., " Mechanical Properties and Durability of Concrete made with High Volume fly ash blended cemen iusing coarse fly Ash", *Cement and Concrete Research 31 (2001) 1393-1403.*
16. ……., 'Recommended Guidelines for Concrete Mix Design', IS 10262:1982, Bureau of Indian Standard, New Delhi, India.
17. Jiang, L.H., and Malhotra, V. M., "Reduction in Water Demand of Non Air-Entrained Concrete Incorporating Large Volume of Fly Ash"; Cement and Concrete Research. Vol. 30, 2000, pp. 1785-1789.
18. Davis, R.E., Carlson, R.W., Kelly, J.W., and Davis, H.E. 1937. Properties of cements and concrtes containing fly ash. Journal of the American Concrete Institute, 33:577-612.
19. Owens, P. L., Fly ash and its usage in concrete, The Journal of the Concrete Society, 1979
20. Central Electricity Generating Board, 1967, PFA data book, London, UK.
21. Tikalsky, P. J., Carrasquillo, P. M., and Carrasquillo, R. L., Strength and durability considerations affecting mix proportioning of concrete containing fly ash, ACI Materials Journal, pp 505-511,1998.

22. Malhotra, V. M, and Mehta, P. K., High Performance, High-Volume Fly-Ash Concrete, 1st edition, Supplementary Cementing Materials for sustainable Development Inc, Ottawa, Canada, 2002.
23. Mehta, P.K. and Monteiro, P.J.M., Concrete –Micorstructure, Properties and Materials, First International Edition, Indian Concrete Institute, 1997.
24. Nagataki, S., Tomosawa F., Kanazu. T. and Yamamoto T., 'Classification of Fly-ash according to Japan Industiral Standard, JISA 6201-1999', Seventh CANMET/ACI International conference on Fly-ash, Silica Fume, Slag and Natural Pozzolanas in Concrete, ACI SP 199, Vol.-1, pp 29-42, Chennai 2001.
25. Sivasundaran, V., Malhotra, V.M., "High-performance high volume fly ash concrete", The Indian Concrete Journal, Vol. 78, No. 11, pp13-21.
26. AITCIN, P. C., High Performance Concrete, 1st edition, 1998, E & FN Spon.
27. Sivasundaram, V., Carette, G. G., and Malhotra, V. M., Long Term Strength Development of High Volume Fly Ash Concrete., Cement and Concrete Composites, 12:263-270,1990.
28. Canadian code..
29. Srinivasan P., Tiwari A.K., Banchhor A., "Suitability of HVFAC Concrete Pavements", The Indian Concrete Journal, Vol. 78, Number-11, PP 58-61.
30. Desai J. "Construction of HVFA Concrete Pavements in India: Four case studies", The Indian Concrete Journal, Vol. 78, Number-11, pp67-71.
31. Ashraf, M., Sharma Amit, Mokal Manish, "HVFA Concrete for Termie seal and Pylon Pile Cap of Bandra-Worli Sea Link Project", The Indian Concrete Journal, Vol. 78, Number-11, pp31-36.
32. Basu, P.C., Bapat, S.G., Mittal, A., Gupta, S.K., Kulkarni, S.B, "Use of fly ash concrete in Indian nuclear power plants", The Indian Concrete Journal, Vol. 81, No. 11, Novmber 2007, pp 7-24.