FIFTY FIRST ANNUAL MEETING OF THE INDIAN ACADEMY OF SCIENCES, HELD AT MADURAI KAMARAJ UNIVERSITY CAMPUS, MADURAI. SUMMARIES OF LECTURES DELIVERED

A Computer Study of the Indus Script by I. Mahadevan, *International Association of Tamil Research, Madras. (Residence: Vyjayanthi, 112, Chamiers Road, Nandanam, Madras 600 035, India).*

This paper is a brief report of the ongoing programme of a computer study of the Indus Script carried out by the present author and Mythili Ranga Rao at the Tata Institute of Fundamental Research, Bombay. A Database for the Indus Script has been compiled on the CYBER 170/730 Computer System. Preliminary results of the statistical analysis indicate that none of the claims of decipherment of the Indus Script made so far is wholly successful.

1. *Database for the Indus Script*

A Database for the Indus Script has been compiled by the authors on the CYBER 170/730 Computer System at the TIFR, Bombay. The database consists presently of 3573 lines of text found written on 2906 artefacts from 19 Harappan and 5 West Asian sites. The artefacts (designated as 'inscribed objects') have been classified broadly into eight types based on the materials of manufacture and modes of writing. A summary of the distribution of the inscribed objects according to sites and types is given in table 1.

The database has been divided into two parts namely, background data and the Texts. The background data provide information on (i) site, (ii) locus of occurrence within the site, (iii) Stratigraphic level, (iv) object type, (v) associated field symbols (pictorial motifs), (vi) direction of writing of each line of text and (vii) the number of signs in each line. A reference number, which is uniquely defined, is assigned to each line of text. The Texts reproduce the Indus Signs occurring in the inscriptions in both numerical and literal (pictorial) forms to facilitate statistical study and reporting respectively. A brief description of the elements of the database is given in table 2. The database also includes a library of signs in the Indus Script for graphic reproduction on a CALCOMP Drum Plotter (examples in figure 1).

Table 1 *Distribution of inscribed objects according to types and sites*

Types of inscribed Objects	No. of Occurrences at Sites							
	MD	HP	CD	LL	KB	OS	WA	CORP. Total
Seals	1232	350	58	89	56	13	16	1814
Sealings	119	288	3	75	21	4	1	511
Miniature Tablets	–	272	–	–	–	–	–	272
Pottery Graffiti	13	64	4	1	20	17	–	119
Copper Tablets	135	–	–	–	–	–	–	135
Bronze implements	5	3	1	–	2	–	–	11
Ivory/bone rods	28	1	–	–	–	–	–	29
Misc. Obj.	8	7	–	–	–	–	–	15
Total	1540	985	66	165	99	34	17	2906

[Notes: MD: Mohenjodaro, HP: Harappa, CD: Chanhudaro, LL: Lothal, KB: Kalibangan, OS: Other sites, WA: West Asian Sites, CORP.: Corpus of Texts. Misc. Obj: Miscellaneous Inscribed objects. See statements 1 and 2 for analysis.]

Table 2 *Data description of the fields in the data base*

Attributes	Description
Reference Number	Each line of text has a unique Ref. No. in 6 digits comprising the Site Number (col. 1), the Object Number (cols. 2–4), the Side Number *i.e.* the number of inscribed faces of the object (col. 5) and the Line Number, *i.e.* the number of lines of text on each inscribed side of the object (col. 6).
Locus	Area, Section or sub-section of the site as determined by the excavator.
Level	The Level in ft. at which the object was found above (+) or below (−) the datum (in Mackay's excavations at Mohenjodaro and Chanhudaro), or below (−) the surface in Marshall's excavation of Mohenjodaro and Vats' excavation of Harappa. (The data on levels on other sites are not available.) The levels are rounded off to the nearest foot.
Type	The typology of the inscribed objects. (See Table 1 for list of types)
Field Symbol	The pictorial motif in the field on each side of the inscribed object. (See I. Mahadevan 1977, pp. 793–813 for the list of field symbols and illustrations)
Direction of Writing	Mostly from right, occationally from the left and rarely from top to bottom. (The direction of writing was determined by the criteria discussed in I. Mahadevan 1977, pp. 10–14.)
No. of positions in a Line of text	This number records the total number of signs and text breaks (or illegible portions) in a line of text for computational processes.
No. of Signs in a Line of Text	This number indicates the total of extant and legible signs in a line of text.
Line of Text	Each line of text is coded as a series of 3-digit numbers each uniquely defining a sign. (For the Sign List of the Indus Script, see I. Mahadevan 1977, pp. 32–35). Doubtful signs are marked by asterisks. Breaks and illegible portions are also indicated by a special symbol.

Note: The Corpus of Texts published by I. Mahadevan (1977) is based on this Input Data; but the format in the book has been slightly re-arranged. Data on Locus and Level, and the number of 'positions' and signs have been omitted, the Field Symbol codes abridged.

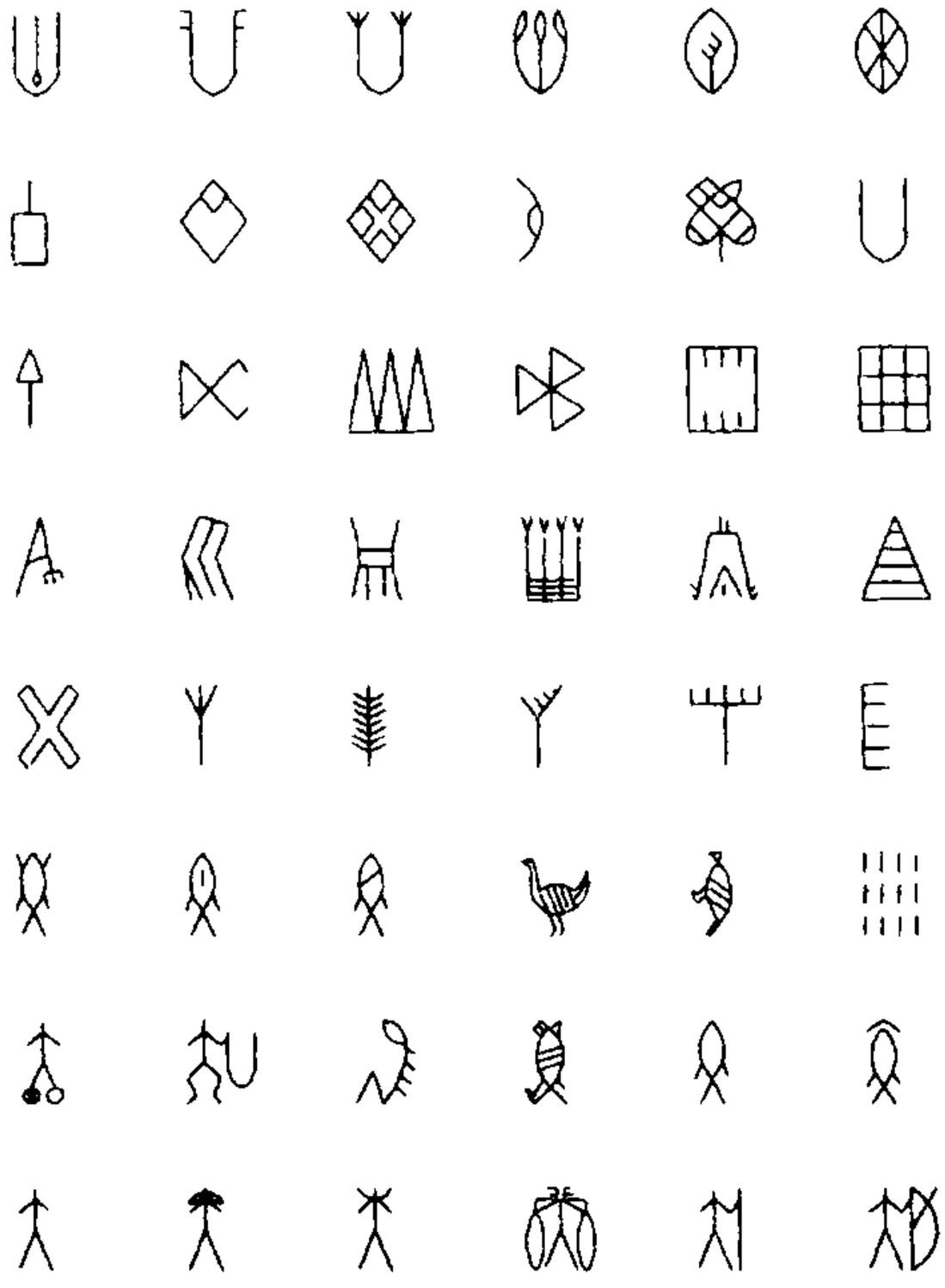

Figure 1. Specimens from the Library of Indus Signs produced on Calcomp Plotter

2. Application of the database: Preliminary results

A Corpus of Texts, Sign-Concordance and Statistical Tables (in an abridged form) compiled from the database have already been published[1]. (For another independent effort to reproduce the Indus Texts and compile a Sign-Concordance with the aid of computer, a reference may be made to the publications of a Finnish Group[2]). The algorithm to prepare the concordance, and methods to determine the direction of writing and for segmentation of texts into probable 'words' and 'phrases' have also been described by the authors in earlier papers[3,4].

Currently work is being carried out in the following areas:— (a) Study of the inscriptions in their archaeological context of occurrence[5], (b) Formal analysis of the texts to recognise grammatical features of the language[6] and (c) Evaluation of proposed claims of decipherment by trying to match the known frequency-distribution characteristics of the Indus signs with the features of the script and the language proposed by would-be decipherers[7]. None of the claims tested so far has come out successfully. But it should be emphasised that the tests are largely negative in character, serving to eliminate the unlikely but unable to pick out the uniquely correct solution!

1. Mahadevan, I., *The Indus Script: Texts, Concordance and Tables*, Archaeological Survey of India, New Delhi, 1977.
2. Koskenniemi, K. and Parpola, A., *Corpus of Texts in the Indus Script; A Concordance to the texts in the Indus Script*, Research Reports 1 and 3, University of Helsinki, 1979, 1982.
3. Rao, Mythili, R. and Mahadevan, I., *Computerised Concordance to the Texts in the Indus Script*, Paper read at the III Int. Conf. on Computing in Humanities, University of Waterloo, Canada, 1977.
4. Mahadevan, I., Recent Advances in the study of the Indus Script, *Puratattva*, 9 (1977–8), New Delhi, 1980, p. 34.
5. Rao, Mythili, R. and Mahadevan, I., *Archaeological Context of the Indus Texts at Mohenjodaro*, paper read at the Int. Conf. on Databases in the Humanities and Social Sciences, Grinnell College, Iowa, June, 1985.
6. Mahadevan, I., *Towards a Grammar of the Indus Texts*, paper read at the Seminar on Indus Script, Tamil University, Thanjavur, Sept. 1983.
7. Mahadevan, I., *Claims of decipherment of the Indus Script: Some objective methods to test their validity*, paper read at the South Asian Regional Conf. (SARC) Workshop on Epigraphy, Mysore, March 1985.

Computer Engineering

Fifth Generation Computers by V. Rajaraman, *Computer Centre, Indian Institute of Science, Bangalore 560012, India.*

INTRODUCTION

The term *Computer Generation* has been widely used most often in relation to the electronic devices used to construct computers. We will, however, take a broader perspective and consider, besides electronic devices used, the storage technology used, the mean time between failures (MTBF), the evolution of software and applications through the various generations. In the first generation the devices used were vacuum tubes. The memory was made up of acoustic delay lines and later magnetic drums. Tubes were not very reliable and consequently the mean time between failures of the computer was about an hour. The machines were programmed using machine codes or assembly languages which were one to one transformation of machine code. Early applications were mostly for scientific and engineering calculations. A big revolution took place in the second generation of computers due to the advent of transistors. Reliability of computers dramatically increased. This coupled with the invention of magnetic core memories made the use of high level machine independent languages such as FORTRAN feasible. This led to a rapid growth in computer usage in both science and business. The next step was more evolutionary. Transistors were replaced by integrated circuits with a consequent reduction in cost and ten fold increase in reliability. High level languages improved and were standardised by international standards organization. From a user's point of view a major advance was time sharing of computers with consequent interactive use of computers. The interactive use led to tremendous increase in the productivity of users as they could build models in close association with the computer. Currently we are in the fourth generation which is characterised by the microcomputer revolution. Large number of microprocessors are used in personal computers, in sophisticated instruments and even in household appliances. Other major characteristics of this generation are the advent of graphics and of computer networks. Table 1 is a comparative summary of the characteristics of the four generations of computers.

VON NEUMANN ARCHITECTURE

Computers have become faster, smaller and cheaper but one fundamental feature has remained invariant throughout the four generations. This is the basic logical structure of the computer which was proposed by Von Neumann in 1946. This logical structure also known as *architecture* is characterised by the program being stored in a central main memory and interpreted sequentially by the central processing unit. Another central concept in this architecture is the identification of a variable by a storage location which gets updated by program overwriting the contents. As a consequence of this logical structure of the machine, mostly sequential algorithms have been designed. Many algorithms are characterised by extensive use of loops, which are inherently sequential and involve dynamically changing the state of the memory. The correctness of algorithms is difficult to establish as these algorithms do not satisfy many universally understood mathematical laws. Computer languages are designed to efficiently represent algorithms and consequently are sequential languages.

Table 1 *Computer generations–comparison*

Generation	Years	Switching device	Storage device	Switching time	MTBF/	Software	Applications
First	1949–55	Vacuum Tubes	Acoustic delay lines and later magnetic drums. 1 Kbyte memory	0.1 to 1 milli-second	30 mts. to 1 hour	Machine and Assembly languages. Simple monitors	Mostly scientific. Later simple business systems.
Second	1956–65	Transistors	Magnetic Core main memory, tapes and disk as peripheral memory. 100 Kbyte main memory	1 to 10 microseconds	About 10 hr.	High level languages. FORTRAN, COBOL, Algol. Batch operating systems	Extensive Business Applications. Engineering Design optimization, scientific research
Third	1966–75	Integrated Circuits (IC)	High speed magnetic cores. Large Disks (100 MB). 1 Mbyte main memory	0.1 to 1 microseconds	About 100 hr	FORTRAN IV, COBOL 68, PL/I.Time-shared operating system.	Data Base Management systems. On-line systems
Fourth	1975–present	Large Scale integrated circuits. Micro processors (LSI)	Semi-conductor memory. Winchester disk. 10 Mbyte main memory. 1000 Mbyte disks	10 to 100 nanoseconds	About 1000 hr	FORTRAN 77, Pascal, ADA, COBOL-74, concurrent PASCAL.	Personal computers. Distributed systems. Integrated CAD/CAM. Real time control. Graphics oriented systems.

•MTBF—Mean time between failures of the processor

SOFTWARE CRISIS

A major problem in the use of computers has been the escalating cost of software due to the need for skilled programmers whose productivity is constrained by the available languages which require attention to clerical details. The major criticism of the current state-of-the-art of computers is that even though machines have become cheaper, very powerful and highly reliable, they have not become easier to use in the same proportion. In fact current architectures inhibit thinking in "parallel", *i.e.* perceiving what could be done simultaneously and what need to be done sequentially. They also inhibit program proof procedures and consequently lead to unreliable software. Further the current methods of program development have too much emphasis on how to procedurally solve a problem and mixes up the two concerns: What problem is to be solved and how to solve it. It has now been realised that if the productivity of programmers is to improve dramatically there is a need to shift the emphasis of programming to precise formulation of the problem to be solved – namely problem specifications. A built-in translator must then convert the specification to detailed procedure. It would also be useful if program acceptance criteria are stated along with program specification so that the testing of the translated "object program" is easily accomplished. Thus new languages are needed to specify algorithms without the inhibition of a preconceived logical structure of a computer. This is one of the goals of fifth generation computers – namely to develop specification oriented languages. If a logically good notation for such a language is developed then one may be able to use the same language to specify application – translate it for implementation and prove its correctness.

MULTI-PROCESSOR ARCHITECTURE

The second major problem with the current computer architectures is the limit to their speed imposed by physical constraints such as the velocity of light. There are two ways of increasing the speed of computers. Both of them involve the use of multiple

processors to do a single job. One method is to break up a job into a number of independent sub-tasks and allocate each sub-task to one processor in a chain of processors. The sub-tasks are done sequentially by processors in the chain. A stream of jobs are sent to the processors in the chain. As all processors work simultaneously on different sub-tasks the total time taken to do the jobs will be reduced. This is known as *pipe lining*. This method will work only if we do a large number of identical jobs. Another method would be to do several jobs simultaneously by allocating one job each to many processors. All the processors can work simultaneously and independently thereby reducing the time to carry out the job. Both these ideas are being explored extensively for building fifth generation computers.

Another significant development in technology is the feasibility of designing custom Very Large Scale integrated circuit chips for special applications. This has made it economical to build special purpose computers extremely inexpensively. Such machines have been built for solving partial differential equations, simulate planetary motion etc.

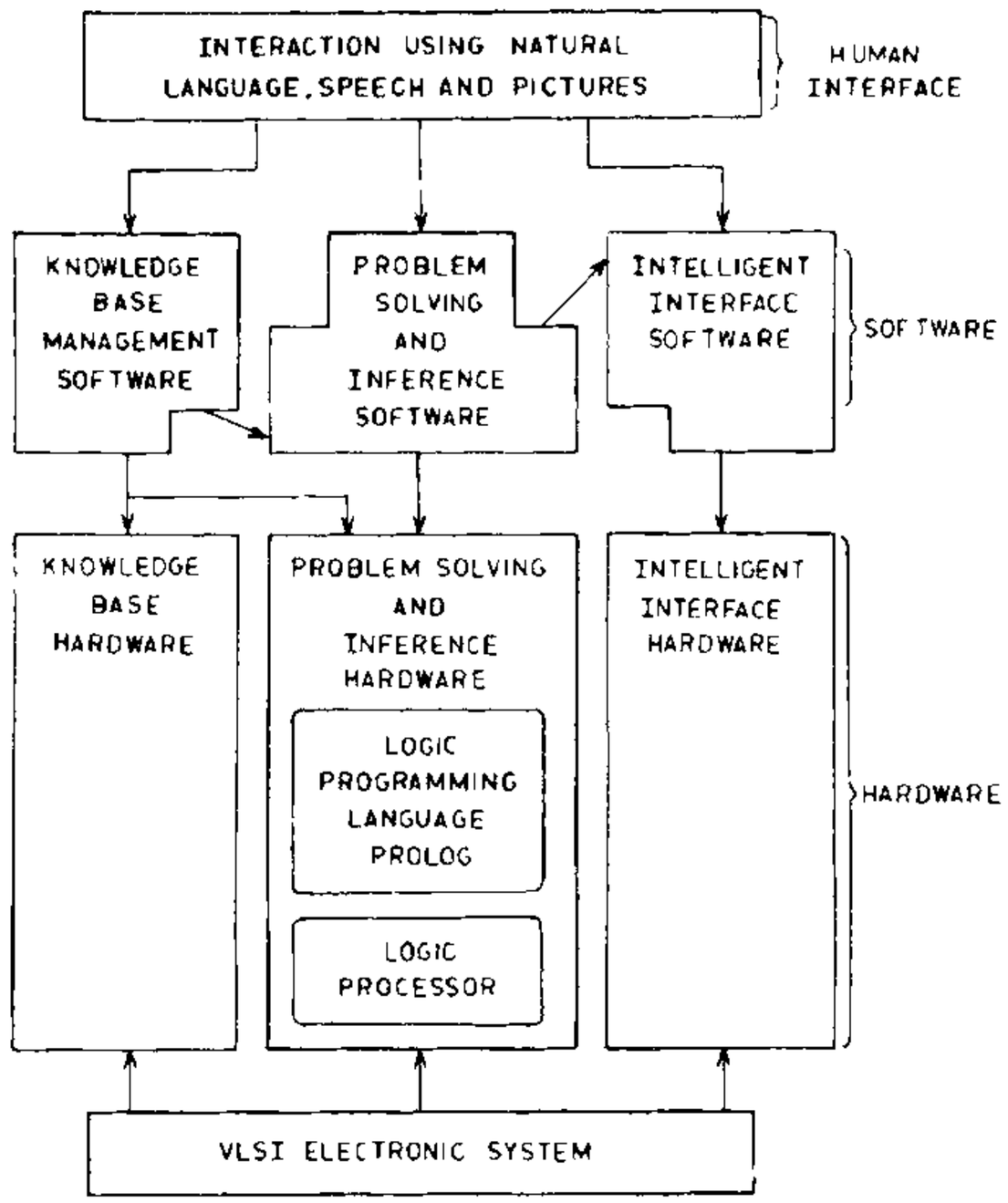

Figure 1. Proposed Japanese knowledge processing system.

KNOWLEDGE BASED SYSTEMS

The phrase '*fifth generation computers*' was coined by the Japanese to set goals for a nationally coordinated research and development programme with the goal to become world-leaders in computer technology by 1990. An organization called ICOT (Institute for new generation Computing) was set up with cooperation from industries, research laboratories, universities and the government. The Japanese foresee that in the next century *Knowledge Processing Industry* will be the dominant industry replacing heavy industry of today. Thus their main goal is to develop a new computer which will not merely be a data processor but will also be a knowledge information processor.

The main goal of the project is to develop tools and techniques for knowledge information processing. As a first step in developing such systems, *knowledge bases* are required. In current computers we basically develop data bases and process these data bases. The difference between a data base and a knowledge base is brought out by the following example. Suppose a large number of experiments are carried out and the data collected. The data can be organised systematically as a data base for easy retrieval of the desired data. If the data base is combined with a set of *inference rules* which will allow us to explain the experimentally obtained data base as a consequence of physical law(s), then these *inference rules* along with the data base is much more powerful and constitute a knowledge base. The inference rules can be derived only by knowledgeable experts in the area and will be a valuable asset. An example of such an inference rule would be

> "If car does not start and battery is OK and ignition system OK and carburettor OK then check petrol line"

A collection of such knowledge bases in diverse areas will increase the competitive edge of a nation enormously.

To achieve the main goal of the project, it is necessary to develop super speed computers which would be able to process large knowledge bases. The main output obtained by processing knowledge bases would be *logical inferences*. It is necessary to carry out dozens of instructions for each logical inference. Thus very high speed computers would be required to process such knowledge bases. As was pointed out earlier in this article, distributed asynchronous computers would be needed to achieve these high speeds.

Such architectures known as *non-Von Neumann* architectures will achieve speeds in the range of billion inferences per second.

INPUT-OUTPUT

Currently the types of input and output devices which can be connected to a computer are restricted. Inputs are mainly via keyboards or magnetically recorded media such as floppy disks. Outputs are either printed pages or displays of both alphanumeric information and graphics on video terminals. A more natural way of human communication would be by speech, handwritten documents and pictures. One of the goals of the fifth generation computers is to develop new input/output devices which would recognize speech, pictures and handwritten inputs and give answers in spoken as well as pictorial form, besides printed form. To meet this goal would require building complex speech and picture recognition systems. Such systems require complex inference mechanisms as most human speech is not well structured and there is considerable difference in speech patterns of individuals.

Lastly to meet the goal of high speed computing, extensive developments in electronic device technology and computer aided design of electronic systems are needed. Japanese goal is also to invest massively in this area to obtain the required electronic systems.

FIFTH GENERATION

A block diagram of the proposed fifth generation computing system is shown in figure 1. This figure succinctly summarises the proposed features of the machine to be developed. Currently forty of the best Japanese engineers drawn from many industries and research laboratories are working in Tokyo in the Institute for new generation computing and their goal is to build this machine by 1990. As a response to this Japanese challenge, groups have been formed in USA, UK and Europe to develop their own versions of such a machine. It is hoped that by the turn of this century such machines will become a commercial reality.

CONCLUSION

To conclude, table 2 shows a summary of what may be the fifth generation. This Table does not fully bring out the true spirit of the fifth generation, which is to let scientists and engineers concentrate on formulating clearly what questions they want to ask and why they want to ask them rather than spend a lot of time on working out in detail how the computations are to be carried out. The more intelligent computers of the future will assist in working out the details of computation by picking the right algorithms, the right programs and knowledge bases using expertise built into them.

Table 2 *Fifth Generation Computer characteristics*

Year	Switching Device	Storage Device	Switching time	MTBF
1990 (?)	V^{n}LSI Ga As (?)	Associative Memories, Optical Disks. Tens of Giga bytes of distributed storage	~ 0.1 nsec. speed measured in Giga Lips	10000 hr.

Software	*Applications*
Functional Programming (Specification Language) PROLOG Personalised expert systems	Primarily knowledge processing oriented applications

1. Rajaraman, V., *Fundamentals of Computers*, Prentice-Hall of India (Pvt.) Ltd., New Delhi, 1985.
2. Feigenbaum, E. A. and McCorduck, P., *The Fifth Generation*, Pan Books Ltd., London, 1984.
3. Fredrick Hayes-Roth, Waterman, D. A. and Lenat, D. B. (eds.), *Building Expert Systems*, Addison-Wesley, Reading MA, U.S.A., 1983

Voice Input and Response by Computer by Prof. P. V. S. Rao, *Tata Institute of Fundamental Research, Bombay 400 005, India.*

Man-Machine Interaction

Use of computers requires a man machine interface. This interface needs to be capable of accepting inputs by way of instructions and data from the human user in some mode or form that the user is capable of

understanding, and conveying it to the machine in a form that it can interpret. For outputs, the interface should be capable of accepting information in a form that can be provided by the computer and presenting it to the user in a form that he can assimilate.

Input and Output in the Speech Mode

Even in pre-Fifth-Generation-Computers, there are important applications justifying the need for at least a limited capability for interaction in the speech mode. Among these are: point of sales, postal sorting and assembly line type of situations and on-line computers associated with sophisticated process systems such as large power plants, factories, high performance aircraft, etc where the hands and eyes of the operator are engaged in control and monitoring functions but his voice channel (speech and hearing) is free.

Implementation

There would be no major difficulties in using such a system if speech recognition were possible with 100% reliability. The state of the art of speech recognition today however is significantly short of this level. There is, therefore, need for a scheme for effective interaction even in such a situation.

One solution is to provide visual feedback by way of textual display of the speech as recognised by the computer. Keyboard input can be used as back up to correct errors in speech recognition. This, however, would be impractical if the user is not very literate. The solution suggested here is as follows: The operator speaks into a microphone connected to the computer. His speech is recognised. A transcript is flashed on the video screen. However, in addition to this, the computer vocalises the same utterance, using its own internal speech synthesis system. The user therefore hears his message being repeated by the computer and identifies mistakes if any. He may correct these by repeating the relevant portion (the wrong word or phrase) more slowly, more clearly, etc. if the mistake is repeated, the word may be spelt out in some form. Keyboard input may be needed only in very rare cases. The various sub systems required in such a system are described below.

Speech Recognition

The central part of such a system would be the speech recognition software. A comprehensive and powerful speech recognition system has been evolved for continuous speech. This scheme utilises information about interphoneme contextual effects contained in forment transitions and employs trial synthesis and feedback comparison as a means for recognition.

Speech Synthesis

Synthesis by rule has been attempted by several workers. Rules are usually formulated for each possible adjacent pair of phonemes and used to synthesise understandable speech. The difficulty with this approach is that for N phonemes, the system requires N × N rules.

Intonation of the synthesised speech

A special feature of the current application is that the computer is not often required to speak on its own. It is either repeating a statement made by the user or giving him a response which is essentially one of a limited number of pre-determined sentence types.

The problem of intonation of messages independently put out by the computer would remain. In the absence of a full-fledged intonation generation scheme, the choice would be between two options (1) to conform to a fairly restricted set of syntactic structures for which the intonation pattern can be pre-specified and (2) accept a monotonic output. It would be appropriate to treat (1) as the preferred alternative and to resort to (2) only in cases not covered by it.

Conclusion

Practically all the subunits required for the implementation of this system have already been realised and implemented. On the basis of the performance of these subunits, the system is expected to be able to operate in an untreated office room with some reasonable restrictions on noise level.

Applications of such a system would be several. As mentioned already, such a system could be used as a data entry device in rural areas. Insertion of information regarding revenue and land data, population, crop, rainfall, occurrence of diseases in man, animals and crops, farm outputs, and so on, can be easily input using such a terminal. A further application of such a system would be its use even as an environment for the evaluation of various speech recognition and synthesis approaches.

Mathematics

Binary Recursive Sequences by T. N. Shorey *Tata Institute of Fundamental Research, Bombay 400 005, India.*

In the following equations we have

$$1 + 3 + 3^2 + 3^3 + 3^4 = 121 = 11^2$$
$$1 + 7 + 7^2 + 7^3 = 400 = 20^2$$
$$1 + 18 + 18^2 = 343 = 7^3$$

The numbers 121, 400 and 343 have all the digits equal to one with respect to base 3, 7 and 18 respectively. Furthermore, they are all perfect powers. This leads us to the question of determining perfect powers in integers with all the digits equal to one with respect to a given base. Let x > 1 be an integer. Tijdeman and the present author (1976) showed that there are only finitely many perfect powers with all the digits equal to one with respect to fixed base x; say $x = 10$. For a sufficiently large x and an integer $q > 1$, the author has, recently, shown that the number of q-th perfect powers $(= 2^q, 3^q, \ldots)$ with all the digits equal to one with respect to the base x is less than q.

Goormaghtigh observed that

$$31 = \frac{2^5 - 1}{2 - 1} = \frac{5^3 - 1}{5 - 1},$$

$$8191 = \frac{2^{13} - 1}{2 - 1} = \frac{90^3 - 1}{90 - 1}$$

Thus the numbers 31 and 8191 have all the digits equal to one with respect to two distinct bases. It has been conjectured that these are the only ones with this property. For distinct integers $x > 1$ and $y > 1$, the author has, recently, shown that there are at most 17 integers with all the digits equal to one with respect to the base x as well as base y.

In the above results, we are concerned with expressions $(x^m - 1)/(x - 1)$, $m = 0, 1, 2, \ldots$ which is binary recursive sequence given by

$$u_0 = 0, u_1 = 1, u_m = (x + 1)u_{m-1} - xu_{m-2}, \quad m = 2, 3, \ldots$$

Under necessary restrictions, it has been proved by Petho (1983) and Stewart and the author (1983), independently, that there are only finitely many perfect powers in a binary recursive sequence.

$$u_0 = 0, u_1 = 1, u_m = (x + 1)u_{m-1} - xu_{m-2}, \quad m = 2, 3, \ldots$$

Curves in space by N. Mohan Kumar, *Tata Institute of Fundamental Research, Bombay 400 005, India.*

It is a very basic problem in Algebraic Geometry to determine the number of equations necessary to define algebraic varieties in n-spaces (over, say, the complex numbers.) Naively speaking, one should expect that a d-dimensional variety in n-space can be defined by n-d equations. But this is false even when $d = 2$. But the case of $d = 1$ (curves) has attracted a lot of attention; it was conjectured by Kronecker that curves are defined by n-1 equations. Any variety can be defined by n equations in n-space, by a theorem of D. Eisenbud and G. Evans. The first break-through was a result of D. Ferrand and L. Szpiro, when they proved that Kronecker's conjecture is indeed true when $n = 3$ and the curve is smooth. (*i.e.* it is a one-dimensional complex manifold.) The present author proved it for arbitrary n when the curve is smooth. R. C. Cowsik and M. V. Nori proved this result without the smoothness hypothesis, but over a field of positive characteristic. Inspite of vigorous activity in this direction, till to-day, the general problem over complex numbers remains open.

Astrophysics

Black Holes are not Forever by B. R. Iyer, *Raman Research Institute, Bangalore 560 080, India.*

A description is given of black hole solutions, in general relativity and some of their characteristic features like event horizons, infinite redshift surface and static limits. In the context of rotating or charged black holes, dynamical processes like Penrose process, Superradiance and Zeldovich-Starobinsky-Unruh emission are explained. An account is given of the no hair result, Hawkings area theorem and the conceptual problems that arise when one considers thermodynamics of systems containing black holes. This is used to elucidate the concept of black hole entropy and bring out the analogy between laws of black hole mechanics and thermodynamics. Hawking's discovery of black hole evaporation that brought about a complete consistency in the thermodynamic description of black holes is explained. Conceptual issues arising from the discovery of Hawking radiation relating to breakdown of predictability, observer dependent quantum field theory and the general notion of gravitational entropy are indicated in conclusion.

Mitochondrial Genome: Structure and Distinctive Participation in Eukaryotic Gene Functions by D. S. Pradhan and K. Pasupathy, *Biochemistry Division, Bhabha Atomic Research Centre, Bombay 400085, India.*

One unique feature of eukaryotic gene expression is the extranuclear dispersion of a part of genetic information concerned with elaboration of mitochondria and chloroplasts. The knowledge on structural attributes of DNAs from these organelles is increasing rapidly and is proving useful in our understanding of eukaryotic cell regulation. At the same time, it also promises to give insights into some facets of 'Genome in flux'. Mitochondrial DNA-usually a closed circular duplex molecule with superhelical twists–exhibits a wide variation in nucleotide composition and size in different organisms. The mitochondrial genome of wild-type *Saccharomyces cerevisiae* cells (approximately 78,000 base-pairs in size) has GC content of 18%, the lowest reported so far for a functional genome. It is also characterized by the presence of AT spacers constituting about 50% of the genome and GC clusters accounting for about 10% of the genome. Human mitochondrial DNA, whose complete sequence of 16,569 base-pairs has been elucidated, has no such non-translatable elements and displays an amazing economy. While some genes in *S. cerevisiae* mitochondrial genome contain long introns, those in mammalian mitochondrial genome have no introns; in fact, some genes overlap. The high degree of fluidity of yeast mitochondrial genome structure is manifested in its extensive re-arrangement in petite mutations. The rho mitochondria contain circles of DNA that are much smaller than the usual genome with complexity ranging from 0.2 to 30% of the original genome. The sequence retained in rho$^-$ petite is often amplified to generate large number of copies. Recent studies have shed light on the mechanism of mitochondrial DNA polymerisation in mouse L cells. Replication is unidirectional starting from two *ori* sequences, one for each strand. The primer sequence of newly synthesized H strand corresponds to the transcription of an oligopyrimidine stretch, which is conserved in the D-loop regions of mammalian mitochondrial DNAs. Yeast mitochondrial DNA replication is an RNA-primed bidirectional process and has a basic similarity with mammalian mitochondrial DNA replication. Besides 2 active *ori* sequences, there are atleast 5 other *ori* sequences, on *S. cerevisiae* mitochondrial DNA which are perhaps non-functional in the wild-type cells but become functional in petite mutants.

The alternative life-styles of *S. cerevisiae* cells during growth under anaerobic and aerobic conditions have greatly facilitated research on mitochondrial DNA functions. Some studies on these aspects carried out in the authors' laboratories will be presented. Only a few copies of mitochondrial DNA are present in the anaerobic *S. cerevisiae* cells; replication of this DNA, as well as transcription, is essential for the *de novo* biogenesis of functional mitochondria during aerobiosis. We have found that during the anaerobic-aerobic transition, *de novo* biogenesis of mitochondrial DNA polymerase takes place. It would be of interest to know whether other enzymes involved in mitochondrial DNA polymerization especially RNA primase are also induced in similar manner. The fragility of mitochondrial DNA to very low doses of DNA-acting agents including ultraviolet radiation and chemical mutagens can be easily demonstrated. Mature mitochondria of aerobically grown *S. cerevisiae* cells do not seem to possess excision repair machinery. The excision repair machinery in nuclei cannot also repair ultraviolet radiation-damaged DNA in mature mitochondria, presumably on account of the double-membrane barrier. On the other hand, ultraviolet radiation-damaged mitochondrial DNAs in anaerobic cells is repaired by nuclear DNA repair machinery. It would be of interest to know whether similar situation exists in respect of recombinational repair of damaged mitochondrial DNA. The advances in research on organelle genomes should clarify our concepts on nucleocytoplasmic inter-relationships.

Random Walks by V. Balakrishnan, *Department of Physics, Indian Institute of Technology, Madras 600036, India.*

Random walks find application in numerous and diverse areas, ranging from physics to demography, ecology and economics. In addition to its continuing role in the development of the mathematical theory of probability, random walk theory is relevant—to name a few instances—to diffusion in fluids and lattices, the configurations and dynamics of polymers, stochastic formulations of quantum mechanics and field quantization, energy transport in amorphous solids, various problems in percolation, population dynamics in inhomogeneous environments, the motion of bacteria (*e.g.*, chemotaxis), market fluctuations in economics,

dam theory. Monte Carlo techniques in computer simulation, and so on: in nearly every area, indeed, in which probabilistic and statistical methods are required.

Random walks may be considered in discrete or continuous time, with discrete or continuous spatial variables, depending on the application. Beyond the primary problem of deducing the probability distribution (and its moments) characterizing the random walk, there is a host of problems of physical and mathematical interest: distributions of the first passage time and time of escape from a given region; probability of return to a given position (recurrence); the number of distinct sites visited; the number of returns to a given site; the distribution of the maximum displacement and the span of a walk; random walks in the presence of traps, and the associated survival and extinction probabilities; directed and biased random walks; clustered walks and other spatially correlated walks; persistent walks and other temporally correlated walks, such as (non-Markovian) 'continuous time random walks'—walks whose time evolution is governed by a renewal process, relevant in the physical problem of stochastic transport in disordered media; self-avoiding walks of various kinds; walks under various kinds of static and moving boundary conditions; walks with internal degrees of freedom (multi-state random walks); walks involving several interacting 'walkers'; walks on regular and random fractals, as well as walks on random structures ('random' random walks), leading to various kinds of 'anomalous' diffusion.

The modern era in the development of the subject is just about eighty years old. A great deal has been done, but new, interesting and physically relevant questions emerge continually at a rate which indicates no diminution in the freshness of the random walk problem.

Materials Science

Synthesis of Phase Diagrams by P. Ramachandra Rao, *School of Materials Science & Technology, Institute of Technology, Banaras Hindu University, Varanasi 221005, India.*

The Metallurgical community is now making an international and concerted effort towards synthesising, analysing, compiling and dissemenating all available information on phase diagrams. The present paper cites several examples highlighting the need for accurate phase diagrams of binary, ternary and multi component alloys in hydrogen storage devices, in understanding morphology of rapidly solidified alloys, in development of epitaxial crystals for optoelectronics etc. The principles underlying the synthesis of phase diagrams from thermodynamic data are presented. Computational methods are discussed with particular reference to the available transformation and solution thermodynamics, analytical expressions for excess thermodynamic properties and computer storage and retrieval of data. Current views on lattice stability and *ab initio* calculation of phase diagrams are incorporated. A few examples are given of the successful synthesis of binary phase diagrams such as Al-Zn, Mg-Sn and In-Sb.

Chemistry

Theoretical Study of an Organometallic Reaction by E. D. Jemmis, *School of Chemistry, University of Hyderabad, hyderabad 500 134, India.*

Even though theoretical chemistry has come of age, the chemical problems for which exact numerical solutions are available are restricted in size. Reactions of binuclear transition metal complexes do not yet yield to rigorous *ab initio* Quantum Chemical Studies. Methods based on symmetry, overlap, and perturbation theory often give useful insights in such cases. The reaction of $L_2W(\mu\text{-}CR)_2WL_2$, (1) (L $= (CH_3)_3SiCH_2$, R $= (CH_3)_3SiC$) with alkynes is an example. We study the electronic structure of the models $(CH_3)_2W(\mu\text{-}CH)_2W(CH_3)_2$ (2), the product of the reaction of 2 with acetylene, $(CH_3)_2W(\mu\text{-}CH)(\mu\text{-}C_3H_3)W(CH_3)_2$, (3), and the intermediate $(CH_3)_2W(\mu\text{-}CH)_2W(CH_3)_2(\eta^2\text{-}C_2H_2)$ (4) using Fragment Molecular Orbital Theory. The following observations can be explained by us. (a) Only one equivalent of C_2H_2 reacts with 1. (b) Only acetylene and allene, but not ethylene, react with 1. (c) Ta and Nb analogs of 1 do not react with acetylene. (d) The W_2C_2 ring in 1 is an unusual case of a four membered ring with four electrons and yet aromatic. As most of the arguments developed here depend on symmetry, overlap, electron count and electronegativity differences, our conclusions should depend minimally on the method of calculation.

Applications of superabsorbent polymers in the separation processes by M. G. Kulkarni, *National Chemical Laboratory, Pune 411 008, India.*

Superabsorbents belong to a family of synthetic and modified natural polymers, which in contact with water absorb 200–600 times their weight of water to form gels. They are distinguished from other water soluble polymers by their swelling capacity. Equilibrium swelling capacity of these polymers is governed by polymer-polymer interaction, network elasticity and hydrogen ion concentration. The gels undergo deswelling in the presence of the nonsolvents, and electrolytes at appropriate temperature and pH. The phenomenon is analogous to the critical phenomenon.

The reversible swelling and deswelling of superabsorbents forms the basis of the technique of concentration of macromolecules from aqueous solutions. Apart from water, other low molecular weight solutes are also absorbed to the exclusion of macromolecules. The concentrations can be carried out in a batchwise, semibatchwise or in continuous operation. Proteins such as albumin, myoglobin, hemoglobin, and enzymes such as glucose isomerase, alkaline protease, xylanase were concentrated using the technique.

The fluxes are comparable to those realized during conventional ultrafiltration separations. In contrast to some of the newly emerging techniques which offer higher fluxes but lower recoveries, more than 90% recoveries are possible using the superabsorbents.

The technique is particularly useful for the concentration of the biological macromolecules which undergo deactivation during concentration by conventional techniques such as precipitation and ultrafiltration.

Zeolite Catalysts: Performance and Promise by R. A. Rajadhyaksha, *Department of Chemical Technology, University of Bombay, Bombay 400 019, India.*

Zeolite catalysis is now regarded as one of the frontier areas of heterogeneous catalysis. This is due to their exceptional catalytic properties and some of the novel concepts introduced by zeolite catalysts. One of the very important and unique features of zeolite catalysts is 'shape selective catalysis'. The crystalline structure of zeolites include channels of dimensions comparable to the dimensions of molecules encountered in common reactions. This has an interesting effect on selectivity of zeolite catalysts. Three types of selectivities have been observed; reactant selectivity, product selectivity and transition state selectivity. Reactant selectivity is observed when the reactant stream consists of a mixture of species of different molecular sizes. It is observed that only those molecules undergo reaction which can enter the channels of the zeolite catalysts. This selectivity is exploited commercially in selective cracking of linear paraffins in a mixture containing linear, branched and cyclic hydrocarbons. Product selectivity is observed when a reactant is capable of forming a variety of products of different molecular size. Zeolite catalysts permit the formation of only those products which can be contained in the channels and which can diffuse out of the channel structure at a reasonable rate. Very interesting example is the conversion of methanol to a hydrocarbon mixture in the gasoline range containing a large proportion of monocyclic aromatic hydrocarbons. The reaction can potentially continue further to naphthelenic and higher hydrocarbons but is restricted to the monocyclic stage due to the selectivity mentioned above. Another very exciting example of this type of selectivity is alkylation of toluene with methanol to give more than 95% yield of para xylene. The third type of selectivity arises due to the restrictions imposed by the channel structure on the interaction of the reactant molecules in the necessary orientation. This feature has also been exploited commercially in isomerisation of xylenes where zeolites of smaller channel size have been observed to supress the undesirable disproportionation reaction almost completely.

One of the most exciting aspects of the shape selectivity discussed above is that it is predictable since the sizes of molecules and the structure of zeolites are known *a priori*. Although the role of shape selectivity in a variety of reactions has been studied, the concept still remains to be applied to many other reactions. Considerable promise also exists in the synthesis of new zeolites which can exhibit new types of shape selective properties. Yet another direction in which efforts appear worthwhile is to explore the catalytic properties of known zeolites for novel catalytic reactions.

Role of Oncogenes and Growth Factors in Cell Proliferation by G. Shanmugam, *Cancer Biology Division, School of Biological Sciences, Madurai Kamaraj University, Madurai 625021, India.*

Oncogenes were discovered a decade ago as part of the replicative intermediate-genome of transforming retroviruses. The reverse-transcribed genome of RNA tumour viruses gets integrated with cellular DNA. Using restriction enzymes one can dissect the viral oncogene and use it for transforming normal cells into tumour cells. Recent studies show that the oncogenes indeed exist apparently in all normal cells *viz.* human, simian, feline, murine, avian, drosophila, dictostelium and yeast. The fact that the oncogenes are there in such evolutionarily divergent species, (without being eleminated) shows that they may have some useful function in the cells' life. The present consensus is that the cellular oncogenes were acquired by transducing viruses millions of years ago from some common progenitors of vertebrates. The role of cellular oncogenes in cell proliferation has become evident from the findings of sequence similarities between oncogene products and growth factors. The product of a simian sarcoma virus-related cellular oncogene (c-sis) is related to platelet derived growth factor (PDGF). Both the c-sis product and PDGF induce DNA synthesis and tyrosine phosphorylation. Our results[1, 2] with mouse embryo fibroblasts indicate that both serum and PDGF induce transiently a set of proteins during the transition of cells from quiescent to proliferation stage. Two of these proteins were studied in detail. A 29,000 dalton serum (PDGF) induced protein was found to be tightly associated with chromatin and satisfy some of the criteria of a competence protein for DNA synthesis. These criteria are: (a) association with chromatin, (b) synthesis before DNA is made and (c) short half-life. Actin is the other mitogen induced protein we studied. Although actin is believed to be a cytoskeletal protein, recent studies imply its role in gene activation. It is involved in the transcription of lamp-brush chromosomes. Furthermore, a part of the viral oncogene of Gardner-Rasheed sarcoma virus has been shown to be derived from actin gene. The gene for epidermal growth factor receptor is related to the oncogene of erythroblastosis virus. The expression of certain cellular oncogenes like c-fos and c-myc are induced in immediate response of the addition of mitogens to resting cells. The cell cycle specific response of c-myc expression is lost in transformed cells. The normal PDGF gene when cloned appropriately, is capable of transforming normal cells. The relationships in structure and function of some of the cellular oncogenes and the genes for growth factors (receptors) support the view that the normal function of the oncogene is growth control and probably the abnormal production/function of the oncogenes leads to loss of cellular regulation resulting cancer.

1. Subramaniam, M. and Shanmugam, G., *Cell. Biol. Int. Rep.* 1985, 9, 51.
2. Subramaniam, M. and Shanmugam, G., *J. Cell. Physiol.* (in press).

Biology

Food Nitrogen as a Predictor of Digestive Efficiency by T. J. Pandian, *School of Biological Sciences, Madurai Kamaraj University, Madurai 625021, India.*

Digestive efficiency is an index of the proportion of the consumed food that is transferred from the gut lumen into the body of an animal. Absorption is measured as the difference between ingestion and egestion. Gravimetric estimation of digestive efficiency is a time-consuming process and still subjected to technical errors. The difficulties involved in the estimations of ingestion and egestion have led several authors to resort to indirect procedures to estimate digestive efficiency. Available indirect marker methods require no quantitative estimate on ingestion or egestion but their applicability is limited (*e.g.* Chromic oxide method), or their reliability is questionable (*e.g.* Conover's ash ratio method). From published information, it has been observed that nitrogen content of food holds a positive correlation to digestive efficiency of fishes, amphibians, reptiles, aquatic insects and polychaetes; the relation is ($P < 0.001$) significantly correlated (> 0.09) in these animal groups. Hence digestive efficiency of these animal groups is predictable from the food N with less than 10% error using the following formulae, in which Ae or Ase refers to digestive efficiency:

Animal group	Formula
Fishes	$\log Ae = 1.3700 + 0.5807 \log N$
Amphibians	$Ae = 49.615 + 4.596\ N$
Reptiles	$Ase = 14.424 + 8.210\ N$
Aquatic insects	$Ase = 9.29 + 8.82\ N$
Polychaetes	$Ae = 11.74\ N + 6.51\ N^2 + 0.46\ N^3$

Unlike in these aquatic animal groups, water content of the food significantly influences the digestive efficiency of terrestrial lepidopterans; hence the inclusion of food water as a co-predictor of food N improves the precision of prediction ($r = 0.868$; SE = 9.7). The multiple regression thus obtained predicts the digestive efficiency of lepidopterans: $11.610 + 8.246\,N + 0{\cdot}509\,W$ more precisely.

The different biokinetic ranges of food N for these animal groups were recognized for the first time. Increasing food nitrogen content by 1% results in the increase of the efficiency by a factor of $7 \pm 2\%$ in these animal groups. When a food with equal nitrogen content is given the amphibians display significantly higher efficiency than fishes and reptiles.

Molecular Biology of Cancer by M. R. Das, *Centre for Cellular & Molecular Biology, Hyderabad 500 007, India.*

With the development of tissue culture methods, it was realized that viral model systems provided distinct advantages in making quantitative observations on carcinogenesis compared with the use of chemical or physical agents. By the 70's it was demonstrated that DNA tumour viruses induce cancer in their hosts after the integration of viral DNA with the chromosomal DNA of the hosts followed by viral specific gene expression. The discovery of reverse transcriptase led to further understanding of the molecular basis of malignancy induced by transformation competent retroviruses. Analysis of the replication cycle of retroviruses along with the genetic experiments revealed the existence of viral oncogenes (v-*oncs*) in transformation competent retroviruses. The v-*oncs* found in retroviruses were shown to be transfected from cellular homologues, and the existence of cellular oncogenes in vertebrate DNAs has been clearly established.

In the light of current information available on the regulation of expression of viral and cellular oncogenes, results from our laboratory in the following areas are significant in terms of basic understanding of the molecular biology of cancer.

(i) The discovery and studies on the structure function relationship of reverse transcriptase.
(ii) Studies on nucleic acid homology of RNA-tumour viral genomes and cellular DNA, and the study of oncogene expression in a chemically induced rat ascitic hepatoma (the Zajdela Ascitic Hepatoma, ZAH) and the significance of these observations.
(iii) The isolation and characterization of a tumour specific transplantation antigen from ZAH and the significance of the study of the TSTA in inducing tumour immunity.
(iv) The discovery, physical separation by Percoll gradient and characterization by flow cytofluorimetry of two tumour cell subpopulations from ZAH, one which kills host animals and the other which causes tumours that regress in three months. This system provides a handle to separate the 'killer' property from the proliferation characteristics of the tumour cells and makes it a convenient model to study oncogene expression, tumour immunology and tumour heterogeneity.
(v) The isolation and characterization of a ribonuclease from human milk and the rationale for the potential use of antibodies raised against the enzyme for diagnostic/prognostic use for human mammary neoplasia.

DNA Sequence Analysis–Biological and Structural Implications by A. S. Kolaskar, *Centre for Cellular and Molecular Biology, Hyderabad 500 007, India.*

The recent development of rapid methods of sequencing DNA has enabled many biological laboratories to determine the sequence of large number of molecules of DNA. Analysis of these sequences has been carried out to get an insight into some problems in molecular biology which attempts to relate structure with function of nucleic acids. We have analysed DNA sequences from *E. coli* phages MS2, M13, ϕX174 fd, G4, λ, fl, and T7 (only for these phages complete genome sequences are known). Our analysis has shown that 13 triplets AAA, CTG, CAG, TAT, TAC, TGG, CCC, CGG, CAT, AGG, TCG, GGG and AAT, occur in the coding frame with a frequency that is very different from that in non-coding frames. We have used this information to determine whether a particular polynucleotide will code for a protein or not. Further, analysis is carried out around ATG/GTG codon (start codons) and a weight matrix is prepared which is used to decide whether a particular ATG/GTG will be an initiator codon or an internal codon. Thus the approach 'gene search by content' and 'gene search by signal' was combined and a new

algorithm was formulated to locate protein coding sequences in prokaryotic DNA. This algorithm was applied to prokaryotic DNA sequences available in European Molecular Biology Laboratory (EMBL) nucleic acid sequence data base. Our results have shown that the method developed has an overall accuracy of 90%. Further, we can extend the approach to suggest whether particular DNA sequence would code with high expressivity or low expressivity. We have also analysed distribution of near neighbour codon pairs and shown that constituent codons in a pair interact strongly in large number of cases. These interactions are mainly due to the need that two tRNA molecules should be present on the ribosome to give rise to constraints in the choice of synonymous codons.

Zoology

Changes in conformation and expression in the genome of the rat during its life span by M. S. Kanungo, *Department of Zoology, Banaras Hindu University, Varanasi 221005, India.*

Specific timings of birth, maturity and cessation of reproductive ability in mammals, more or less fixed life spans of individuals of a species, sequential changes in the isoenzymes, and the ability of genes to alter their expression under the influence of hormones, show that various components of the genome governing these functions are not static. An analysis of the chromatin housing the genes, by digestion with endonuclease DNase I, that cuts the DNA at 10 base pair intervals, shows that it undergoes increasing condensation with age. Digestion of chromatin by DNase I and EcoRI, followed by nick-translation to quantitate incorporation of dTMP into DNA, for the measurement of DNA synthesis shows that the incorporation is 50% less in the old. Digestion by the isoschizomers MspI/HpaII which quantitate methylation in CCGG sequences, followed by nick-translation shows that the incorporation after Hpa II digestion is 50% less in the old. This indicates increasing methylation with progressive age. Since increasing methylation is correlated with compaction, the inference drawn from these studies is that chromatin becomes more compact with increasing age. This may account for the lower level of several enzymes that may lead to old age.

If this is so, one should see decreasing transcription/expression of genes, and correspondingly less messenger RNAs. Indeed, normal transcription of mRNA which is carried out by RNA polymerase II is lower in the old. Even when exogenous eukaryotic RNA polymerase II is added to old nuclei, the transcription is not raised to the level of that of the young. This shows that either the genes are inaccessible to RNA polymerase II or that they are no longer in active conformation that is required for transcription. Use of probes for specific genes gives similar results.

Thus the genes undergo continuous and sequential changes in activity during the life span of an organism.

ANNOUNCEMENT

KALINGA PRIZE

Sir Peter Medawar, British Biologist and a Nobel Prize winner for medicine has been awarded the Kalinga Prize for popularisation of science. The Kalinga Prize was instituted by UNESCO on the initiative of Kalinga Foundation in India in 1951.