Cosmology: Past, present and future*

Jayant V. Narlikar

This is a broad-brush review of the development of cosmology during the twentieth century. The 'past' deals with the first nine decades of the century while the 'present' deals with the last decade. Although technological achievements have helped the astronomer in better viewing the universe, a 'final' understanding still eludes the search for the correct cosmological model. The article ends with a list of unsolved questions which the 'future' may eventually answer.

LET me at the outset register a mild protest that this seminar is being convened a year too soon. Many commercial enterprises, probably misguided by the much-publicized Y2K syndrome, have declared that the third millennium begins on 1 January 2000 AD, whereas by all logic of counting, it should begin a year later, on 1 January 2001 AD. I had hoped that a community of physicists like ours which sets a premium on exactness, should have got our calendar right. So, modulo whatever exciting that might happen during the year 2000 AD, the last one of the 20th century, (but which has to be left out as I have no crystal ball) here is my brief account of the highlights in the subject of cosmology during this century. Towards the end of this talk I will mention a few outstanding issues that hopefully will be resolved in the future years.

To be more specific, I will make the following break up of the past, present and future: Past: 1901–1990; Present: 1991–2000; Future: 2001– . . . .

I will deal with these time zones in that order.

Expanse of the universe

This century saw a remarkable turn around in the views of scientists about how vast our universe is. Two major views held sacrosanct by the majority of the astronomical community over the nineteenth century, fell by the wayside as the horizons of observational astronomy expanded and theorists became bolder in pushing their extrapolations of laboratory physics to larger systems. Here is a timetable of important highlights.

Observational developments

1900–1915: The first belief to go was that the solar system is at the centre of the Milky Way as originally claimed by William Herschel (see Figure 1). Thanks to more accurate measurements of distances of stars and globular clusters, Harlow Shapley was able to show that the Galactic Centre is considerably farther from the Sun. The currently estimated distance is around 30,000 light years.

1900–1920: A change of viewpoint from the Milky Way-based universe to Kant's island universe hypothesis took place. This was the second of the two long-held beliefs to go. Immannuel Kant (1724–1804) had argued that our Milky Way was just one of the innumerable galaxies populating the universe, all distributed like islands in a vast ocean. This notion was violently resisted by most astronomers, who believed that everything that they observed was part of our Milky Way Galaxy. An example of how the community still resisted the Kantian viewpoint at the turn of the century can be seen from the following quote from a popular book of astronomy of the time:

'. . . The question whether nebulae are external galaxies hardly any longer needs discussion. It has been answered by the progress of research. No competent thinker, with the whole of the available evidence before him, can now, it is safe to say, maintain any single nebula to be a star system of co-ordinate rank with the Milky Way . . .' (Agnes Clerke, *The System of the Stars*, 1905, p. 349).

These nebulae were diffuse, cloud-like in appearance and were widely believed to be systems in our own Galaxy. There was considerable debate between Shapley and Curtis, with Shapley this time on the conservative side.

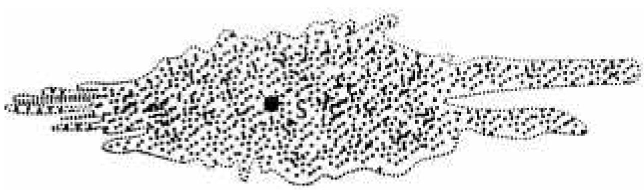

Figure 1. Herschel's map of the Milky Way with the sun (S) shown at the centre.

*Based on a talk delivered at a Seminar on 'Physics in the 20th Century and Trends for the New Millennium' Indian Physics Association.

J. V. Narlikar is at the Inter-University Centre for Astronomy and Astrophysics, Post Bag 4, Ganeshkhind, Pune 411 007, India.

His view is summarized in the following quote:

'. . . Observation and discussion of the radial velocities, internal motions, and distribution of spiral nebulae, of real and apparent brightness of novae, of the maximum luminosity of galactic and cluster stars, and finally of the dimensions of our galactic system, all seem definitely to oppose the "island universe" hypothesis of the spiral nebulae . . .'. (H. Shapley, *Publications of the Astronomical Society of the Pacific*, 1919, **31**, 261).

One reason for the conservative viewpoint to be maintained was the considerable observational work by the senior astronomer Van Maanen who reported significant transverse angular motion of these nebulae. This meant that if the nebulae were really distant, then their physical velocities would be too enormous to be real. And so Van Maanen's measurements of transverse motions implied that nebulae could not be extragalactic. However, eventually astronomers came to discount these measurements, as they could not be verified by any subsequent observations.

1920–1930: Starting with the early spectroscopic measurements of Slipher, and others and the detection of spectral shifts, mostly towards the red end of the spectrum, culminating in the work of Humason and Hubble, the extragalactic nature of nebulae became accepted. The spectral shifts, interpreted as Doppler shifts led to the picture that most of these nebulae are receding from us. Hubble gradually established that these nebulae are galaxies of stars, just like the Milky Way, thus confirming the Kantian hypothesis.

1929: Hubble's Law relating the radial velocity (V) to distance (D) of a typical galaxy was put forward for the first time. Written today as $V = HD$, the constant H is called 'Hubble's constant' (see Figure 2).

1930: The concept of the *Expanding Universe* was established and this was to form the basis for future development of cosmology.

Theoretical developments

1917: Einstein proposed in 1915 his general theory of relativity and in 1917 he applied the theory to construct a mathematical model describing a static, finite but also unbounded universe. He then required a non-zero cosmological constant (λ). He had hoped this to emerge as a unique model of the universe. However, within a few months, De Sitter showed that an empty but expanding universe was also a solution of Einstein's modified field equations. Whereas Einstein's solution had matter without motion, De Sitter's universe had motion without matter!

1922–1924: Alexander Friedmann produced expanding non-empty world models, but these were ignored as mathematical curiosities by Einstein and others. In these models, the space was taken to be of constant curvature, positive, zero or negative. In modern terminology we denote these by a curvature parameter k which takes values $+1, 0, -1$.

1927: Abbe' Lemaitre from Belgium produced similar theoretical solutions, being unaware of Friedmann's work.

1932: Realizing that in the context of Hubble's discovery, a static model was no longer relevant, Einstein abandoned the cosmological constant and in a joint paper with De Sitter, favoured the flat space expanding model from Friedmann's solutions. Therefore this model is often called the *Einstein–De Sitter model*. This model is the simplest of all Friedmann models and has $\lambda = 0$, as well as the curvature parameter $k = 0$.

1933–1936: Eddington and Lemaitre, however, continued working with models having non-zero cosmological constant as they felt that a larger parameter space will be helpful to account for all the observed features including the formation of galaxies.

A range of Friedmann models with or without λ is shown in Figure 3.

Can observations help choose the right Friedmann model?

The next three decades were used by cosmologists to extend their observations to *test* cosmological models, with the hope that observations would single out a specific model as *the* model of the universe. The time table of some

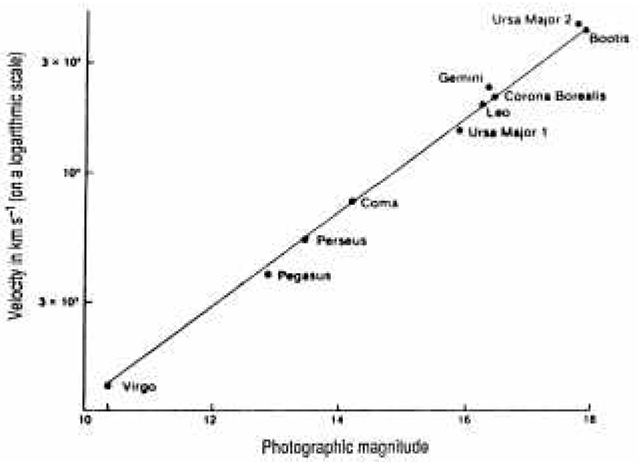

Figure 2. Hubble's plot for the fifth brightest member in clusters of galaxies. The photographic magnitude can be related to distance on the logarithmic scale, while the recessional velocity is obtained by multiplying the observed redshift by the velocity of light c.

specific developments is given next.

1935–1940: Hubble hoped to determine the correct model by counting galaxies up to increasing level of faintness, as it gave a radius–volume relation, that could be compared with the model-dependent theoretical relation. This project was doomed to failure as the number of galaxies to be counted up to distances where curvature differences are noticeable, was too large.

1940–1945: The Palomar Telescope of 200 inch aperture was built for the above key project. However, by the time the telescope was completed, it became clear that the project was unworkable.

1945–1955: The emerging science of radio astronomy went through an early period when radio astronomers thought that all radio sources are stars in the Galaxy. Tommy Gold had argued that a large population of the radio sources may be extragalactic, a conclusion that was violently resisted by the Cambridge radio astronomer Martin Ryle. A few years after the Gold–Ryle controversy, it was realized that a majority of sources was indeed extragalactic and this led to optimism that one could solve the cosmological problem by counting radio sources instead of galaxies. (Radio sources are not as numerous as galaxies.)

1955–1965: Radio source counts were used by Ryle as a disproof of the Steady State Cosmology (SSC). The SSC was proposed in 1948 by Hermann Bondi and Tommy Gold and by Fred Hoyle as a reaction to the apparent shortcomings of the Friedmann Cosmology, namely:

(i) *A singular origin*: The model had a beginning in a primordial event often called the *Big Bang*, a name due to Hoyle himself. We shall refer to the various Friedmann models as part of the Standard Big Bang Cosmology (SBBC). The big bang itself is a physically undefinable and mathematically singular event. That is, all theoretical machinery breaks down at this instant, labelled by the time coordinate at $t = 0$. In a physical theory such an event is therefore indicative of some serious shortcoming of basic formalism.

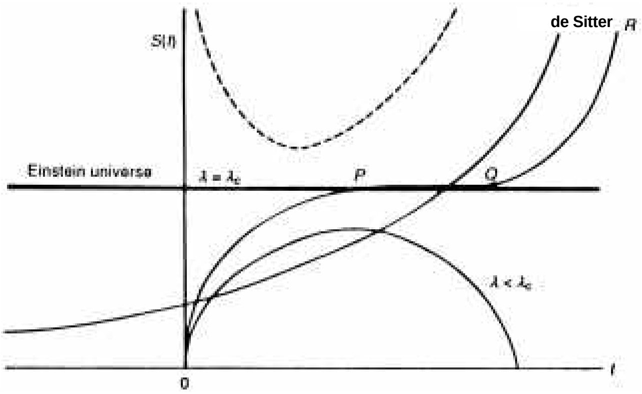

Figure 3. Expanding world models for different values of l and a positive curvature parameter.

(ii) *The age problem*: Counting the time since the above epoch of 'beginning', the age of the universe at present can be determined for any model in terms of the measured value of Hubble's constant. The answer came out smaller than the ages of many of the oldest stars.

(iii) *The origin of matter not explained*: The epoch of big bang represents creation of the universe. At this epoch the law of conservation of matter and energy breaks down and so the most fundamental of the cosmological issues, viz. the origin of all the matter we see today, is not addressed.

The Hoyle–Ryle controversy of 1961 on source counts and their interpretation marked a major confrontation between the attackers and defenders of the SSC. Eventually, Hoyle's approach turned out to be closer to reality, although not so realized at the time. The SSC, however, made several useful contributions to cosmology:

(i) *Ideas on matter creation and baryon non-conservation*: The theory sought to explain matter creation in the form of baryons, through the agency of a scalar field. At the time a scalar field was not popular with the field theorists, nor could they stomach the idea of the baryon number not being conserved. On both these counts today's theoretical physicists have changed and come round closer to what the SSC had said in the 1950s and 1960s.

(ii) *Massive collapsed objects in galactic nuclei*: One frequently hears of the discovery of collapsed massive objects (glamorized as *black holes*) in the nuclei of galaxies. The idea was in fact first proposed by Fred Hoyle and the author in 1966, at which point the notion was considered bizarre.

(iii) *Superclustering of galaxies*: The hot universe model of Gold and Hoyle in 1958 had shown that structure formation in the SSC would take place through thermal pressure gradients, resulting in typical units of size 50–100 Mpc, characteristic of superclustering of galaxies. Hoyle and the author had used inhomogeneity on this scale to explain Ryle's source counts. In the early 1960s, inhomogeneities on this scale were not considered likely; today they are an accepted part of reality.

Can all nuclei of elements be made in a primordial process just after the big bang?

Parallel to the development of the SSC, a new direction was being provided to the SBBC by George Gamow who attempted to show that nuclei of *all* chemical elements were formed in the first few minutes after the big bang. The landmarks in this branch of physical cosmology were as follows.

1946: George Gamow initiated work on this problem with his student Ralph Alpher and later joined by another, Robert Hermann.

1948: Affirmative answer by Alpher, Bethe and Gamow to the question as to whether atomic nuclei could be synthesized in the early hot era. This work became known as the α–β–γ (alpha–beta–gamma) theory, after its authors! A modern version of this calculation yields abundances shown in Figure 4. Only light nuclei can be made this way. For all heavier nuclei the appropriate location is inside stars.

1948: Prediction of relic black body radiation background in microwaves was made by Alpher and Herman. This radiation of the early hot era was expected to cool down as the universe expanded: Alpher and Hermann guessed the present temperature of the background as ~ 5 K.

As physicists five decades ago did not take cosmology seriously (nor did the astronomers!) this important prediction was largely ignored both by theorists and observers.

What is the significance of the cosmic microwave background radiation?

The microwave background today is considered the strongest evidence for the SBBC. Here, however is a time table of how information about this important component of the universe was put together piece by piece.

1941: McKeller found that CN–molecular transitions imply a radiation background of 2.3 K. This result was, however, largely ignored, partly because of the wartime preoccupations and partly because it was published in an obscure journal.

1948: Prediction by Alpher and Herman came up as reported earlier.

1955: Bondi, Gold and Hoyle estimated the energy density of stellar radiation in the SSC, assuming all helium found in the universe to be of stellar origin. They found that if thermalized, that energy density would be like a black body radiation of temperature ~ 3 K. However, they did not press this point further, partly because they did not see an obvious process of thermalization.

1965: Penzias and Wilson serendipitously found the CMBR of temperature ~ 3.5 K. Their observation was of course at a single wavelength of around 7 cm. But the uniformity of the background was taken to identify it with the relic radiation of the SBBC.

1965–1990: Various surveys culminating in COBE in 1990 subsequently confirmed a black body spectrum of the CMBR with a temperature of ~ 2.7 K. The COBE spectrum is shown in Figure 5.

1977: Dipole anisotropy in the radiation background was discovered and interpreted as arising from the earth's motion against the isotropic background.

1992: First detection (by COBE) of small-scale inhomogeneities in the CMBR generated considerable excitement and euphoria in the big bang community as these were perceived as the indicators left on the background by the proc-

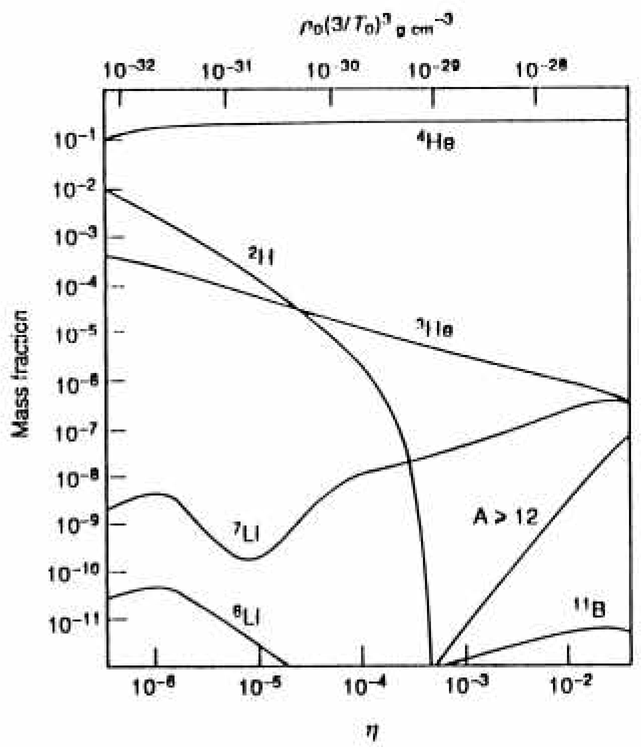

Figure 4. Primordial abundances of light nuclei plotted as a function of baryon density $\rho = \eta\,(T/10^9)^3$, with T measured on absolute scale.

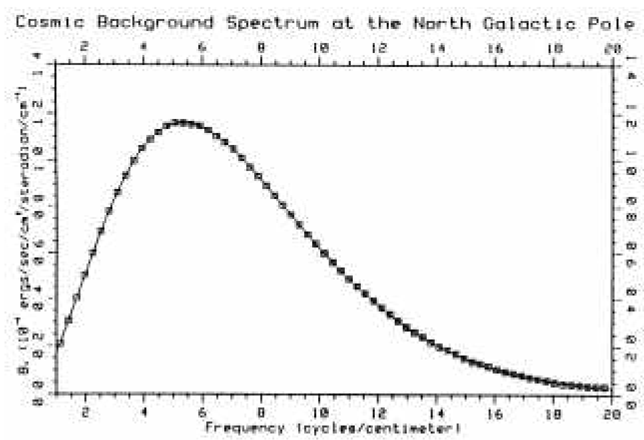

Figure 5. COBE measurements of CMBR spectrum. The continuous curve describes a 2.73 K black body curve passing through error rectangles.

ess of structure formation.

Can surveys of the universe to high redshifts determine the correct world model?

1950–1980: Two groups, Sandage *et al*. and de Vaucouleurs *et al*., continued to improve the determination of Hubble's constant, but also continued to differ by a factor 2, with Sandage and co-workers advocating a lower value.

1960–1980: Sandage carried on measurements of the Hubble relation to high redshifts with the hope of measuring the deceleration parameter, i.e. how fast the universe was slowing down; but systematic errors and evolutionary corrections proved insurmountable. Thus the goal of determining the correct model still eluded observers.

1960–1990: The angular diameter–redshift relation also was beset with many uncertainties and evolutionary effects and could not settle the cosmological problem.

Thus the improvement of observational techniques only served to remind the observer that several pitfalls lie between observations and interpretation. Even the counts of galaxies obtained by computerized reading of plates for a large number of galactic images made it clear that a clear-cut conclusion of the kind expected by Hubble fifty years earlier is still not possible.

Can high energy physics usefully interact with the SBBC to resolve mutual problems?

1968: The electroweak unification raised hopes of a *grand unified theory* (GUT), but particle physicists needed a high energy laboratory where such a theory could be tested.

1977: The only such laboratory was provided by the SBBC if one could confidently extrapolate close to the big bang. Thus particle physicists teamed up with cosmologists.

1980–1981: Out of such wedlock was born the idea of *inflation* first suggested by Kazanas, Guth and Sato independently; and it has played a key role in the agenda of the SBBC.

1970–1990: Astronomical observations indicated existence of a large quantity of dark matter which the SBBC required to be largely non-baryonic and hence cosmologists began to get inputs from various ideas in particle physics, ideas like GUT, super-symmetry, strings, etc. for candidates for such matter.

This now brings me to the present decade.

Thrust areas in cosmology in the last decade of the 20th century

The present work in cosmology is mainly in the following areas:

Structure formation: Given the primordial seeds in the pre-inflationary era, attempts are made to see how they would grow and lead to the presently observed galaxy → cluster → supercluster format of large-scale structure, together with their peculiar motions as well as their imprints on the CMBR. This is a multi-parameter exercise which folds in such items as the nature of dark matter, biasing, N-body simulations, etc.

Redshift surveys: These will tell us how matter is distributed around us out to greater distances so as to know about structural hierarchy.

Universe at large redshifts: Observations of discrete source populations at redshifts going up to $z \sim 5$ tell us about how the universe has evolved in the last few Gigayears and thus put constraints on cosmological theories.

Baryogenesis: A fundamental issue has been to understand how baryons, etc. formed in the early universe. A particularly interesting issue still to be understood is the apparent predominance of matter over anti-matter, and the overall dominance of radiation as exemplified by the large photon to baryon number ratio.

Alternative cosmologies: As the present observational constraints are already proving severe for the SBBC, it is worth exploring alternative cosmologies.

In 1993, Hoyle, Burbidge and Narlikar proposed an alternative cosmology called the *Quasi-Steady State Cosmology* (QSSC) which has the following positive features:

1. It explains creation of matter in non-singular minibangs in a universe without a beginning, and without violating any conservation laws. The minibangs essentially 'drive' the universe which has a long-term exponential expansion superposed with short-term oscillations. The oscillations are generated by the on/off switching of mini-creation events. A typical oscillation lasts for ~ 50 Gyr while ~ 20 oscillations take place in one *e*-folding time of the long-term expansion. (see Figure 6)
2. It accounts for the origin of the CMBR along with its observed temperature as thermalized relic starlight. Stars are born and burn out during one oscillation. Thus there is relic starlight from all previous cycles.
3. It explains dark matter as relic stars of earlier generations.
4. It accounts for light nuclei as created in minibangs and in stars, with abundances consistent with observations.
5. It is consistent with large redshift observations of dis-

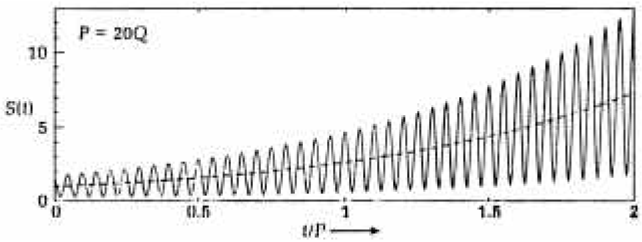

Figure 6. Scale factor of the Quasi-Steady State Cosmology with short-term oscillations (of time scale *Q*) coupled with long-term expansion (of time scale *P*).

crete source populations.

6. It seems to have a viable theory of structure formation through minibangs.

Naturally this cosmology needs to be further investigated for conformity with all the available data about the universe.

Issues for the future

I now end with a few issues that will need attention in the coming years. Future work will tell us many new facts about the universe, and hopefully answer some outstanding questions of the present, for example:

1. What, if anything do minute-scale inhomogeneities of the CMBR tell us about how large-scale structure formed?
2. Did an inflationary phase occur in the very early universe?
3. Is a cosmological constant necessary? If so, how did it originate?
4. How did the universe develop an asymmetry between matter and antimatter?
5. Is the Hubble interpretation of redshift universally applicable to all extragalactic redshifts? The cases of *anomalous redshifts, redshift periodicities*, etc. reported by Arp, Tifft and others are growing in number. These are difficult to fit within the framework of Hubble's law.
6. Will the SBBC survive with minimal modifications, or will we need radically different alternatives like the QSSC for our understanding of the universe?

Perhaps, for those cosmologists who think that they have everything settled and worked out about the universe, I should end with a cautionary remark of J. B. S. Haldane:

'The universe is not only queerer than we suppose, it is queerer than we can suppose.'

Received 30 December 1999; revised accepted 7 January 2000

Importance of small and moderate size optical telescopes

Ram Sagar

Small and moderate size optical telescopes have advantages over large and giant ones in the areas of efficiency, availability, flexibility and serendipitous and speculative observations. Recent developments in astronomical detectors and instrumentation along with the growth in computers and softwares have increased their capabilities many-fold. They are therefore responsible not only for a number of recent discoveries in astronomy, e.g. detection of microlensing phenomenon, but also for providing valuable optical observations of celestial objects and phenomena discovered at other wavebands, such as radio, infrared, X-ray and γ-ray. All these factors make well-instrumented, small and moderate size optical telescopes highly relevant in contemporary astronomy despite competition from 6 to10 m class ground-based optical telescopes and from the 2.3-m Hubble Space Telescope. Such telescopes in India have an added advantage of geographical location.

AN optical telescope is classified after the size of its objective which is either reflector-(mirror) or refractor-(lens) type. In this article, optical telescopes are arbitrarily classified according to their sizes into four groups namely 'small' for telescopes of sizes up to 1 m; 'moderate' for sizes between 1 and 3 m; 'large' for sizes between 3 and 6 m and 'giant' for sizes larger than 6 m. Throughout the world there are many large but a few giant size optical telescopes but more than a dozen of moderate and a large number of small size optical telescopes. In our country, there are four one-metre class telescopes (two 1.2-m located at Japal–Rangapur near Hyderabad and Gurushikhar near Mount Abu; two 1-m located at Kavalur and Nainital) and only one 2-metre class (2.34-m Vainu Bappu Telescope (VBT) at Kavalur) optical telescopes. Even the three upcoming ones, namely the 2-m telescope at the Inter University Centre of Astronomy and Astrophysics at Giravali near Pune[1], the 2-m telescope at the Indian Astronomi-

*For correspondence. (e-mail:)Ram Sagar is at the Uttar Pradesh State Observatory, Manora Peak, Nainital 263 129, India.

cal Observatory at Hanle in the high-altitude cold desert of south-eastern Ladakh[2] and about 3-m telescope at the UP State Observatory (UPSO), Nainital and Tata Institute of Fundamental Research, Mumbai[3] belong to moderate size only. They are expected to become operational in the next few years. Locations of these new and other existing 1- and 2-m class Indian optical telescopes are shown in Figure 1. They are almost evenly distributed from north to south in the country.

Interest in the design and construction of very large optical telescopes has never been greater than at the present time and there is no shortage of research projects for such instruments. However, financial constraints limit the possibilities to what can be achieved within tight budgets. Fortunately, technology has come to the aid of the astronomer and produced ways and means of doing something worthwhile. New technologies have made telescopes lighter and cheaper. As a result, there are many small and moderate size optical telescopes all over the globe and some are still being built. The obvious question is whether such telescopes produce worthwhile research in the presence of large and giant ground-based optical telescopes and 2.3 m Hubble Space Telescope (HST). The aim of this article is to discuss the role of small and moderate size optical telescopes in contemporary astronomical research with special emphasis on Indian telescopes.

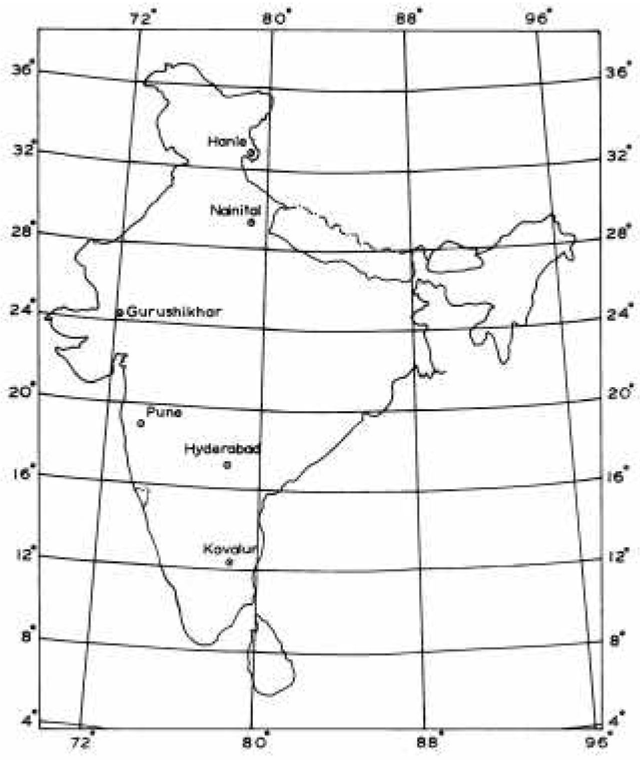

Figure 1. Location of existing as well as proposed one metre and larger size optical telescopes in India.

Purpose of an optical telescope

An optical telescope is used to image celestial objects at its focus. It is well known that, for sky background limited observations, signal-to-noise ratio at a frequency,

$$\nu \text{ is } \propto \sqrt{\frac{A_{\text{eff}} \times t}{\theta_D \times B(\nu)}},$$

where A_{eff} is the light-gathering power of the telescope which includes the losses due to optics and the quantum efficiency of the detector used at the focus of the telescope; $B(\nu)$ is the sky background intensity; t is the integration time and θ_D the solid angle formed by the diffraction limited image of a telescope of diameter D at wavelength λ, is an Airy's disc of size ~ λ/D. For many purposes the power of a telescope is therefore $\propto (A_{\text{eff}}/\theta_D)$. A 2.5-m telescope with 0.5″ image is thus equivalent in performance to a 5-m telescope with 1.0″ image, if other conditions are similar. This is true only for telescopes located in space like HST. For ground-based telescopes, however, the image degrades due to turbulence in the earth's atmosphere which is measured in terms of angle and is called *seeing* by astronomers. Most of the recorded seeing measurements indicate that during a large fraction of ground-based observing time, seeing is more than 1″ at visual wavelengths. The seeing spreads the image thereby diluting the concentrating action of ground-based optical telescopes of sizes larger than about 15 cm. The seeing limited stellar image thus formed is much bigger than θ_D. More turbulence in the earth's atmosphere at a place, therefore, increases the deterioration due to seeing and hence degrades the image formed by an optical telescope. This results in much longer hours of observations for recording any faint image or spectrum by a ground-based optical telescope in comparison with its counterpart in space. For example, HST records images of celestial objects in a relatively shorter time than the ground-based telescopes of similar size. It is therefore clear that small and moderate size optical telescopes can provide neither angular resolution better than HST nor collect more photons per image element than the large and giant size optical telescopes. Even then they not only contribute significantly to the optical observations of present-day research but are also cost effective, as assessed recently by Gopal-Krishna and Barve[4] and earlier by others[5,6]. It has been found that in the optical band almost half of the first rate results were obtained using moderate size optical telescopes. The main reasons for this are given below.

Utility of small and moderate size optical telescopes

In present-day astronomical research, utility of small and moderate size optical telescopes has increased many-fold mainly due to the need of multi-wavelength astronomical observations and because of the advent of modern detectors like charge coupled devices (CCDs) and the advancement in technology of optical telescopes, instruments and computer facilities.

Modern astronomy: Marriage of multiple wavelengths

While the optical band remained our sole window to the universe throughout most part of history, the past 5–6 decades have witnessed the genesis and dramatic progress of other branches of astronomy, encompassing the radio, submillimetre, infrared, X-ray and γ-ray bands. The short wavelength astronomy owes its rapid growth, in particular, to the advent of the satellite era in the recent decades, and space astronomy is likely to play an increasingly prominent role at practically all the wavebands in the coming decades. As celestial objects radiate across the entire electromagnetic spectrum, the various wavebands provide complementary information. This has been the impetus behind the present era of multi-wavelength astronomy. The advent of new observational astronomy at wavelengths other than optical has enabled the discovery of a number of new celestial objects and phenomena, e.g. the possibility of finding a black hole, the discovery of binary X-ray sources, γ-ray bursts and quasars, etc. However, to establish their identity and meaning in astrophysical terms and also to arrive at more definitive and clearer conclusions, optical observations are indispensable, as the optical band contains extra-ordinarily rich concentration of physical diagnostics which have developed from the accumulation of over a century of observations, associated laboratory experiments and theoretical work, though it covers quite a narrow wavelength range. A wealth of knowledge of atomic and molecular transitions of common species that are found in typical astrophysical environments occurs at the optical/infrared wavelengths. Because of its unique ability to measure red shifts, wide angle coverage with high angular resolution, and high spectral throughput, optical astronomy continues to play a key role in unravelling the structure and physical properties of the astronomical objects on all scales. The crucial determination of distance, age and chemical abundance of nearly all astronomical objects remains firmly rooted in the techniques of optical astronomy. For example, distances and other important parameters of γ-ray bursts could be estimated only recently when optical observations of their afterglows could be taken[7,8], though the phenomenon was discovered in the late sixties. Therefore, optical telescopes are needed not only for the observations motivated by optical properties of the objects, but also for the objects and phenomena which are discovered/observed at other wavebands. The demand on the optical observations is often acute enough to keep even small and moderate size (but well-instrumented) optical telescopes fully occupied. For example, observing time on the 2.34-m VBT located in Kavalur (see Figure 1) is generally over-subscribed by a factor of 4 or more.

Limitation of ground-based optical observations

Another contributing factor to the *time squeeze* on optical telescopes is the simple fact that ground-based optical observations of celestial objects are mostly confined to night time and even there, a good deal of time is lost due to clouds or poor seeing conditions. In contrast, radio observations, for instance, are far less susceptible to such natural factors. A reflection of this contrast is the fact that so many small and moderate size optical telescopes that were set up a few decades ago (but have been upgraded with modern instruments and detectors) continue to be in demand even today. In short, the productive life-span of optical telescopes turns out to be remarkably long.

Cost effectiveness

The cost effectiveness, E, of a telescope or collection of telescopes can be considered from several points of view. A detailed comparison of a large telescope versus an array of smaller ones working on the same kind of observations by Disney[5] indicates that large and giant telescopes are disproportionately expensive to build and require equipment which are more complicated to use to yield the best return. Also, the flexibility of operation of the array will yield much more first-rate astronomy than a single large telescope of the same cost. Warner[6] has defined E as the annual return on capital investment. Construction cost of a telescope generally follows the power law of its size. The exponent is 3 for traditional designs but may be 2 for new technology, as it has made telescopes cheaper. Considering the number of significant publications produced by a telescope in a year, it is concluded that to build four half-size telescopes is at least as profitable as one full-size telescope. These factors perhaps prompted astronomers of European Southern Observatory to build the world's largest 16-m optical telescope in the form of an array of four 8-m telescopes.

Efficiency

Large telescopes are often designed to be as versatile as possible and hence multipurpose instruments are used. This makes them not always efficient at their tasks. In contrast, a small or moderate size dedicated telescope can be optimized for a single purpose using, for example, special optics, super-reflective coatings or distinctive mounting. Furthermore, the permanent installation of an instrument on a dedicated telescope ensures stability of performance, improved monitoring of calibrations and reduced loss of time from instrument malfunctions associated with equipment changes. A site with several small or moderate size telescopes, each dedicated and optimized, is an ideal arrangement; there can even be a large telescope if desired.

Need of observations over extended period

It is common at major observatories to find that the large and giant telescopes are inevitably over-subscribed and as such time on them is at a premium. As small and moderate size optical telescopes are larger in number and also less in demand, the observing time available on them constitutes a major research resource. It is here that they outclass the large and giant telescopes. There are several astronomical studies which require sustained observations of the same source over a long period or systematic surveys of large areas of the sky for specific type of sources; for example, studies of long-period variables, short-period variables with complicated frequency spectra, for photometry of standard stars or for general photometric and spectroscopic survey. These observations are as much needed for progress in our understanding of the universe as the studies of the more exotic objects reported by the Keck 10-m telescope or the 2.3-m HST. The best example of this is the discovery of microlensing events towards the Magellanic Clouds and the Galactic bulge[9–12] using 1- to 2-m class optical telescopes. The very low microlensing probability requires several millions stars to be monitored daily to observe a significant luminosity increase and hence observations are required for a short time each night but over a long period. Also, serendipitous and speculative observations are excluded from large and giant telescopes but they can be carried out on small and moderate telescopes. The discovery of the Crab pulsar is a fine example of the reward of speculative observation.

Advantages of geographical location of India

The longitude of India locates it in the middle of about 180 degree wide longitude band having modern astronomical facilities between Canary Islands (~ 20°W) and Eastern Australia (~ 157°E). Because of this and existence of good astronomical sites, small and moderate size optical telescopes located in India can make a unique contribution to astronomical research, particularly involving time critical phenomena. The observations which are not possible in Canary Islands or Australia (during daylight hours), can be obtained from India. Two examples when this was vital were when the rings of uranus were discovered[13,14] and the optical observations of GRB afterglows were made[8]. Furthermore, time series observations of astronomical objects sometimes require a 24-h coverage to understand complex phenomena, e.g. pulsation of white dwarfs[15]. Such coverage is possible, as a few 1-m and one 2-m class optical telescopes exist in India. This situation will improve significantly once all the up-coming Indian moderate size optical telescopes become operational.

Impact of modern detectors and instruments

The enormous improvement in astronomical detectors, the availability of highly sensitive spectrographs, the active and adaptive optics, all make a modern optical telescope far more efficient today than its counterpart a couple of decades ago.

Role of modern astronomical detectors, computers and softwares

The kinds of detectors available to astronomers have always limited the accuracy and efficiency of measuring the light from stars that can be gathered by a telescope. They have constrained, if not dictated, the direction of the advance of knowledge in the area of astronomy. Though Galileo invented the telescope in the early 1600s as a tool to collect more light, the detector at its focus was only the eye until late in the 19th century. Then came the era of photographic plates/films which are not more sensitive than the eye, but they have the great advantage that they can build up a picture of a faint object by accumulating light for a long time. After World War II, photomultiplier tubes became widely available for measurements of the brightness of astronomical objects. They have several advantages over photographic emulsions, for example, unlimited exposure times, larger sensitivity and linear output. The major disadvantage of the photomultiplier tube is that it can observe only a small part of the sky; as a result, stars must be measured one at a time, and extended objects such as galaxies must be sampled point by point, a task that is very costly in terms of telescope observing time. More recently the technology of television and electronic image amplification has been adapted to astronomy, with an aim of combining the accuracy and unlimited exposure time of the photomultiplier tube with the extended field of view of the photographic plate. The CCDs are one of such devices and in fact were developed for astronomical use first. Now they are also used in many other areas. In recent years, photographic emulsions are therefore rarely used for recording astronomical information and in fact they have become out-dated. A comparison of the most important intrinsic properties of these two detectors is given in Table 1.

Table 1. Comparison of photographic emulsion with CCD.

Property	Photography	CCD
Response	Nonlinear	Linear
Size (single piece)	> 30 cm × 30 cm	Max 3.1 cm × 3.1 cm 24 cm × 24 cm with mosaicing
Dynamic range	Small (~ 1 : 100)	Large (~ 1 : 10,000)
Usability	Only once	Re-usable
Quantum efficiency	~ 4%	60–90% at peak sensitivity
Cost	Low	High
Resolution	Smallest grain size (few µm)	Generally ≥ 10 µm

In spite of several drawbacks of CCD in comparison to photographic plate/film, namely their small areas, the high rate of detection of cosmic-ray induced events, and producing enormous amount of data which need more sophisticated computer hardware and software for their aquisition and analysis, respectively the former are preferred over the latter because they offer a combination of qualities like excellent linearity, high quantum efficiency, large dynamic range, low system noise and dark current, and good overall system stability. Hence CCDs in combination with 1-m class optical telescopes are capable of capturing as faint objects as photographic plates can record on a 3–4-m size telescope in a relatively shorter time.

In order to extract even the last bit of information as well as to produce the best possible astronomical results from the CCD imaging, powerful computers and reduction procedures are as important as modern detectors. The best examples of this are doing stellar photometry in crowded regions like globular clusters and estimating completeness of data as a function of brightness in the studies of luminosity functions. Such studies could not be done earlier. The key ingredient in the new software is replacement of fixed-aperture techniques with a point spread function semi-empirically defined from uncrowded images in each frame or image. In this way, the signal-to-noise ratio is maximized and deconvolution of overlapping images becomes feasible. The extensive and time-consuming numerical analyses used in these data reduction procedures could be carried out due to availability of powerful and cheap computers in recent times.

Impact of new technology telescopes

Towards the end of 1980 there was a major step forward in the technological capabilities of telescopes. First, availability of sophisticated computer control changed the telescope mounting from equatorial to the much more compact alt-azimuth. Second, thin mirror technology took a quantum leap forward with active control of the shape of the primary mirror and positional control of the secondary, together enabling better imaging performance[16]. Third, the developments in dome and building design and environmental control have reduced the earth's atmospheric turbulence around the telescope and hence improved the seeing and thus the quality of the images. Fourth, major breakthroughs are taking place in adaptive optics, which effectively control flexible mirrors at frequencies of 10–20 Hz in order to remove the effects of the blurring of images caused by the turbulence of the earth's atmosphere[17]. A combination of these improvements means that a new technology ground-based optical telescope has the capability of approaching the ultimate in angular resolution, namely the diffraction limit of the telescope itself. This kind of performance is normally associated with telescopes in space, but is now achievable with ground-based optical telescopes at a fraction of the cost of a space mission. In the optical region the gap between the capabilities of space and ground-based telescopes is therefore gradually being narrowed. The introduction of active optics in the European Southern Observatory's 3.6-m new technology telescope which was completed in 1988 has improved ground-based image quality significantly. Wilson[16] and Ortolani *et al.*[18] have clearly demonstrated what is achievable with ground-based new technology telescopes on a night of excellent seeing. In fact, it has revolutionized the ground-based imaging.

During the next few years it is expected that the use of adaptive optics will enable us to obtain diffraction-limited images with ground-based optical telescopes. This will make a vast difference in the faintness of the object that can be recorded with them. Increased angular resolution benefits many areas of astronomy.

Multi-object spectroscopy

Most of the astronomical spectrographs were used to take spectra of a single object till about two decades ago. As the field of view of a telescope is generally more than 30′, the observing efficiency was thus very low. The image of a star at the focal plane of the telescope (i.e. at the slit of the spectrograph) is extremely small (only ~ 100 µm) compared to the total length of the slit which is generally more than a couple of centimetres. With the advent of new technologies, it has become possible to use full slit length of the spectrograph and take spectra of more than one object at a time. Such spectrographs are therefore called multi-object spectrographs. Presently, spectra of a maximum of 400 objects located within ~ 2° can be taken simultaneously with the facilities available at the 3.9-m Anglo-Australian Telescope located at Siding Spring, Australia. In this, developments in fibre optics system, computers and electronics have played important roles. With the use of this type of spectrographs, the observing efficiency of a telescope has increased many-fold.

In order to cover a wide spectral range, at many observatories astronomers are using double beam spectrographs.

This also increases the observing efficiency of a telescope.

To achieve high spectral resolution, astronomers generally install heavy and large size spectrographs at the Coudé focus of a telescope. This requires many reflections and thus the loss of light collected by the telescope. To overcome the difficulties with such arrangements, fibres have recently been deployed in astronomical spectrographs.

In the light of above, one can say that astronomical spectrographs have taken full advantage of the latest development in technologies. This has increased the overall spectroscopic observing efficiency of optical telescopes.

Observational programme

It is impossible even to list here all the observational programmes which can be carried out with the modern small and moderate size optical telescopes. They have been discussed extensively in the literature as proceedings of symposia and workshops[19]. The programmes range from solar to cosmological studies. The topics covered are studies on planetary systems, star formation and stellar evolution, astroseismology, galaxy formation and its evolution, structure and kinematics of the galaxy and gravitational lensing, etc. Studies related to large-scale structure and dynamics of the universe and its evolution and cosmology can also be carried out.

Conclusions

Improvement in the observing efficiency of an optical telescope due to use of present-day technology in electronics, computers and light detectors has made the small and moderate size telescopes very useful instruments in the contemporary context. There has been significant improvement in the angular resolution capabilities and signal-to-noise ratio of the image/spectrum obtained using modern techniques. It has also pushed the limiting magnitude of observations to much fainter objects than what was possible a decade before, with a particular size of optical telescope. Consequently, those measurements that were undreamt of 50 years ago and were only a vague hope about 2 decades ago or could not be carried out due to lack of observing time on large and giant telescopes can now be carried out using small and moderate size optical telescopes. Hence, there are big opportunities for the small and moderate size optical telescopes in the contemporary astronomical research. In fact, research programmes on these telescopes not only complement but also go hand in hand with those on large and giant telescopes. Funding to build more moderate size optical telescopes in India is therefore timely and fully justified.

1. Das, H. K., Menon, S. M., Paranjpye, A. and Tandon, S. N., *Bull. Astron. Soc. India*, 1999, **27**, 609–618.
2. Anupama, G. C., *Curr. Sci.*, 1999, **77**, 1127–1130.
3. Mohan, V., *Bull. Astron. Soc. India*, 1998, **26**, 367–370.
4. Gopal-Krishna and Barve, S., *Bull. Astron. Soc. India*, 1998, **26**, 417–424.
5. Disney, M. J., *Mon. Not. R. Astron. Soc.*, 1972, **160**, 213–232.
6. Warner, B., *IAU Symp.*, 1986, **118**, 3–16.
7. Castro-tirado, A. J. *et al.*, *Science*, 1999, **283**, 2069–2072.
8. Sagar, R., *Curr. Sci.*, 1999, **76**, 865; Sagar, R., *Curr. Sci.*, 2000, **78**, 539; Sagar, R., Pandey, A. K., Mohan, V., Yadav, R. K. S., Nilakshi, Bhattacharya, D. and Castro-tirado, A. J., *Bull. Astron. Soc. India*, 1999, **27**, 3–7.
9. Alcock, Ch. *et al.*, *Nature*, 1993, **365**, 621–623.
10. Aubourg, E. *et al.*, *Nature*, 1993, **365**, 623–625.
11. Udalski, A. *et al.*, *Acta Astron.*, 1993, **43**, 289–294.
12. Ansari, R. *et al.*, *Astron. Astrophys.*, 1997, **324**, 843–856.
13. Pandey, A. K., Mahra, H. S. and Mohan, V., *Bull. Astron. Soc. India*, 1984, **12**, 258–262.
14. Mahra, H. S., Pandey, A. K., Joshi, U. C. and Mohan, V., *Bull. Astron. Soc. India*, 1983, **11**, 152–158.
15. Marar, T. M. K., Seetha, S. and Ashoka, B. N., *Bull. Astron. Soc. India*, 1998, **26**, 559–565.
16. Wilson, R., *Messanger*, 1989, **56**, 1–5.
17. Chinnappan, V., Saxena, A. K. and Sreenivasan, A., *Bull. Astron. Soc. India*, 1998, **26**, 371–375.
18. Ortolani, S., Barbuy, B. and Bica, E., *Messanger*, 1995, **82**, 20–22.
19. Proceedings published in *Bull. Astron. Soc. India*, 1998, **26**, 351–614; *IAU Symp.*, 1986, **118**, 3–457.

Received 18 December 1999; revised accepted 11 February 2000

Catalysts for plastics – New science for new materials

Sumit Bhaduri and Virendra Kumar Gupta*

The purpose of this article is two-fold. It starts with the premise that, in all probability, the science and technology of the twenty-first century will be far more inter-linked than they have ever been. The article aims to show how critically important oriented basic science has been for the development of the enormously successful manufacturing technologies for polyolefins. These are materials more commonly known as plastics. It also aims to show how the reverse is equally true, i.e. how the demands of technology in turn have propelled and continue to inspire creative academic research. This is one of those clear examples where the paths of science and technology, theory and experiment, discovery and innovation, knowledge and commerce continually criss-cross. An article along these lines, at a time when terms like 'academy industry linkages', the 'knowledge society', etc. are being constantly used, may be of some special value to the Indian scientific community.

CHEMISTRY is called a 'molecular science'. In practice, among other things, this means that if we knew what the molecules looked like, we could predict more or less precisely how they would behave. This is validated beautifully in the science and technology of polyolefins. A central concept of chemistry, from the time of Kekule till today, has been the concept of a chemical 'bond'. A molecular level structural description of the catalysts used for making polyolefins illustrates the many mysteries of *metal to carbon* bonds. Hopefully for these reasons the article may be of some interest to all chemists, even the purists in the community.

A historical perspective: The discoveries and the innovations

The story of polyolefins spans more than 50 years and has its origin in the world of mega-industries and pristine academic laboratories. Polyethylene, the most important member of the polyolefin family, was discovered accidentally in the ICI laboratory in 1933. Ethylene in the presence of benzaldehyde was subjected to high temperature and pressure. The waxy material obtained this way was called low density polyethylene (LDPE). In the early 1950s two other industrial laboratories, Standard Oil (Indiana) and Philips Petroleum, reported manufacturing processes of a higher density polyethylene (HDPE). Another major technological breakthrough took place after more than two decades. A novel process called the Unipol process was reported by Union Carbide for the manufacture of another variant of polyethylene, the so-called linear low density polyethylene (LLDPE), in the late 1970s. Here ethylene is co-polymerized with a small quantity of an α-olefin, such as 1-octene or 1-butene. As shown in Figure 1, in terms of molecular structures, HDPE has very few branches, LDPE has many, and LLDPE is somewhere in between.

Around the time HDPE was discovered by Standard Oil and Philips Petroleum, Karl Ziegler, then the director of the Max Planck Institute for Coal Research in Mullheim reported a novel reaction of ethylene. He described this reaction as the 'Aufbau' (growth) reaction[1,2]. Ziegler's 'Aufbau' reaction is now of course commonly described as the polymerization reaction. Ziegler also studied the effects of different metals and reported what he called the 'nickel effect' which is now recognized to be a quick termination of the polymer growth. In fact in another highly successful industrial process (Shell higher olefin process), soluble nickel complexes are used to make ethylene oligomers of twenty to thirty carbon atoms.

Ziegler's discovery was soon followed by Giulio Natta's discovery in Italy[3]. Natta reported a catalytic system for making polypropylene of high crystallinity. As shown in Figure 1, the methyl groups in polypropylene can have different relative orientations. Natta realized that isotactic polypropylene, because of its regular molecular structure, had a high crystallinity and was a useful polymeric material. He was able to develop a catalytic system that gave predominantly isotactic polypropylene.

Ziegler and Natta shared the Nobel Prize in 1963,

Sumit Bhaduri and Virendra Kumar Gupta are with the Reliance Industries Limited, Maker Chambers IV, 9th Floor, 222 Nariman Point, Mumbai, 400 021, India.
*For correspondence. (e-mail: sumit_bhaduri@ril.com)

two years before R. B. Woodward and one year before Dorothy Hodgkin. In the citation, the Nobel Committee said that they were awarded the prize for 'their discoveries in the field of chemistry and *technology* of high polymers' (italics added). This probably is the only time when the word 'technology' found an explicit use in a Nobel citation. Zeigler in his Nobel lecture showed a world map indicating all the large polyethylene manufacturing facilities that were based on his discovery. The time gap between the discovery and the successful worldwide technology adoption was only ten years – a record even by today's standards.

The other interesting point in the present context is the attitude of Ziegler and Natta towards research. Ziegler had accepted the directorship of Kaiser Wilhelm Institute (later known as Max Plank Institute) for Coal Research in Mullheim on condition that he could work on any research project of his choice and not be limited to the chemistry of coal. Natta did his doctorate in chemical engineering rather than chemistry, as he wanted his research to be industrially useful[4]. His work, subsidized by Montecatini Company in Milan, involved among other things the use of X-ray and electron diffraction, highly sophisticated physical techniques of that time. This apparently esoteric work was, nonetheless, carried out in the Industrial Chemical Research Center of the Milan Polytechnic Institute.

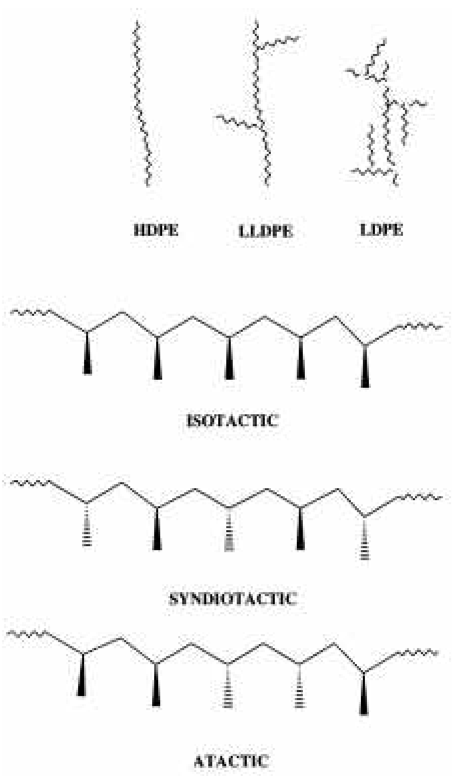

Figure 1. Schematic presentations of polymer molecules of different types of polythylene (upper left) and polypropylene. In isotactic and atactic polypropylene the orientations of the metal groups, with respect to the polymer chain, are all in the same direction and random respectively. In syndiotactic polypropylene every alternate methyl group has the same orientation.

In the early 1950s, while these exciting developments took place in the laboratories of Ziegler and Natta, another novel class of compounds, known as 'sandwich' complexes, were reported[5–7]. Two of the scientists, Wilkinson and Fischer, shared the Nobel Prize in 1973 for their pioneering work in this area. These sandwich complexes, typical examples of which are shown in Figure 2, are also known as *metallocenes*. They played a pivotal role in the subsequent growth and development of the organometallic chemistry of the transition elements. In the present context what is especially noteworthy is that in the 1990s metallocenes have emerged as the future industrial catalysts for polyolefins. Their origin however, like most important discoveries in science, was firmly rooted in pure curiosity-driven creativity that did not have any application whatsoever in mind. The basic question that Wilkinson, Fischer and others tried to answer dealt with the nature of the metal–carbon bond.

It is also interesting to note that, metallocenes did find an early, though limited, industrial application. A conventional *solid* catalyst for the manufacture of polyethylene is made by reacting chromocene with silica. However, as we will see, the present generation of *soluble* metallocene catalysts is qualitatively different from this early catalyst and is in a class by itself.

The 'market-pull' and 'technology-push' of the polyolefin industry

A generally accepted truism of research management is

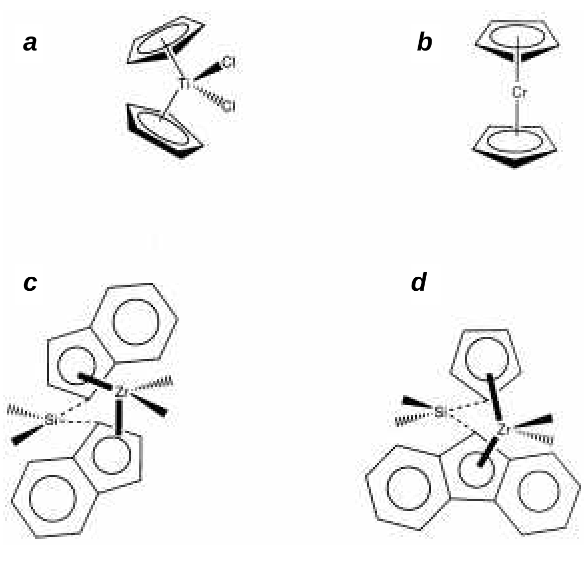

Figure 2. Examples of metallocene or 'sandwich' complexes. Strictly speaking chromocene (top right) is the only real 'sandwich' complex as it has a chromium atom sandwiched between two identical planar cyclopentadiene rings. Chromocene was used to make one of the early industrial catalysts by reacting it with silica. The two complexes at the bottom are recent metallocene catalysts. The one to the left has a C_2 symmetry axis and is chiral. It gives highly isotactic polypropylene. The one to the right has a plane of symmetry and gives highly syndiotactic polypropylene.

based on the 'push–pull' concept of organic reaction mechanisms[8]. Basically, it is thought that the successful commercialization of any new science is rapid and smooth only if there is a good synergy between the market-pull and the technology-push. This is clearly exemplified in the progressive improvements that have taken place over the last five decades in the polyolefin technology. The technologies of polyethylene and polypropylene manufacture, and the underlying science that makes it possible, have come a long way from that of the original Ziegler–Natta discoveries.

The technological and commercial importance of these two leading members of the polyolefin family could be partly gauged from the global production figures. In 1997 total global production of polyolefin was approximately 75 million metric tons, and by the year 2015, the production of only polyethylene is projected to be of the order of 150 million metric tons[9]. The reason for this spectacular growth is primarily because of the enormous versatility of polyolefins, especially polyethylene and polypropylene, in terms of applications and economic processing. These are the 'market-pulls' that make new technology an essential component in the overall competitiveness of all large petrochemical industries. Second, molecular tailoring of the catalysts and innovative down stream processing provides the 'technology-push' that vastly extends the application horizon of these materials.

The crucial and probably the most important links in the innovation chains of all polyolefin technologies are the catalysts. From the time of Ziegler and Natta's original discovery, catalyst systems have undergone tremendous improvements. A standard reference book on chemical technology lists five distinct generations of catalysts with progressive and clear improvements over the previous generation[10]. To keep the discussion focused on how a molecular level understanding has led to these developments, we classify the catalyst systems in the following three categories:

1. Original Ziegler–Natta type systems which could be colloidal or solid catalysts. In this article we classify these as the first generation catalysts.
2. Solid catalysts consisting of magnesium chloride supported titanium chloride, and used in combination with additives. These we classify as the second generation catalysts.
3. Metal sandwich complexes or metallocenes.

In the rest of this article we concentrate on scientific findings that made the 'technology-push' a powerful and incessant one. In other words, we try to highlight how conceptual sophistication backed by hard scientific data, and technological improvements on a plant scale, have progressed hand in hand.

Evolution of the catalysts

The discovery and investigations of the first generation catalytic systems fairly soon led to the following conclusions: (1) A combination of titanium tetrachloride and an alkyl aluminum reagent (usually triethyl aluminium or diethyl aluminum chloride) gives a highly active polymerization catalyst for both ethylene and propylene. (2) The relative orientations of the methyl groups in polypropylene could be correlated to the solid state structures of the titanium trichloride crystals. The selectivity towards isotactic polypropylene is more if the crystal morphology is the right one. (3) In the case of propylene polymerization, additives like ethers, esters, etc. influence the activity and the selectivity of the catalyst.

Starting with the discovery of the sandwich complexes, academic organometallic chemistry in the next three decades made impressive progress in the rational synthesis, the elucidation of structure, and the explanation of observed reactivity of compounds with metal–carbon bonds. This was also the period when industrial laboratories achieved dramatic improvements in polyolefin manufacturing facilities both in terms of catalyst activity and reaction engineering[11–13]. This is the period that we have loosely classified as the phase of the second-generation catalysts.

The basic catalyst system of the second-generation catalysts consists of titanium chloride supported on magnesium chloride. The catalyst must be used in combination with an organoaluminum reagent. Additives such as phthalate esters or silicon atom containing diethers, commonly called 'electron donors', must be added for highly isotactic polypropylene. The inventions that relate to catalyst improvements both in terms of activity and selectivity have been the subject matter of many patents. However very little hard scientific data, that involves the use of sophisticated techniques for the study of solid surfaces, have been reported in the open literature.

This is not surprising for three reasons. First, industry will be happy as long as something works well, day in and day out, in a reproducible and predictable manner. Secondly, assuming that hard data is available, industry may not be keen to publish such data as open literature and make it available to the competitors. Thirdly (probably the most important reason is that) molecular level structural studies of the surface species of a highly air and moisture sensitive active catalyst is a difficult research problem even today. It is worth remembering that the plethora of surface microscopic and other techniques that may provide molecular level structural information is of fairly recent vintage, and not routinely available to or mastered by most synthetic chemists.

The third generation of catalysts belongs to the class of 'sandwich complexes'. While industry was busy perfecting the manufacturing process, a serendipitous innovation in the laboratory of Kaminsky at Hamburg paved the way for the technological exploitation of metallocenes as catalysts[14]. It was found that small quantities of water turned a mixture of Cp_2TiCl_2 (Figure 2) and Me_3Al into an active catalyst for the polymerization of olefins. Both industry and academia vigorously followed up this initial report.

At a rough estimate, by the year 2003, at least 5% of the total global production of polypropylene (approximately half a million metric ton) is expected to be based on metallocene catalysts. Apart from polypropylene, metallocene catalysts are being used for making a variety of specialty polymers. This is because unlike the catalysts of the previous generations, metallocenes can polymerize a very wide range of alkenes and mixtures of alkenes[15,16]. The polymers obtained by using metallocene catalysts also have a very narrow range of molecular weight distribution. Most importantly, with these catalysts the tacticity of the polymers can be controlled with high precision. We now discuss the mechanistic and structural details of the catalysts at a molecular level. For the metallocene catalysts such a description ties up beautifully with the structural details of the resultant polymer molecule.

Molecules and mechanisms

The basic questions that must be addressed are as follows.

1. What is the fundamental difference between the three generations of catalysts at a molecular level?
2. Since polymer formation means the formation of new carbon to carbon bonds, what role does the metal play in that?
3. If the formation of a metal to carbon bond is involved at any stage of the mechanism, what sort of a bond is it? In other words, is it a single bond or a double bond or not quite a bond, but an interaction?
4. Why was it necessary to have trace quantities of water for the metallocene catalysts activated with trimethyl aluminum?
5. How is it possible to control the molecular structures of the resultant polymer molecules by tinkering with the molecular structures of the catalysts?

The first question deals with the difference between *homogeneous* and *heterogeneous* catalysts. The second-generation catalysts, i.e. $TiCl_4$ with or without other additives supported on $MgCl_2$ are *heterogeneous* catalysts. The coordination environments around the titanium ions on the surface are not identical. Many such environments are possible and are actually present. Of these various molecular structures and environments only a few will be catalytically active. These are called the active sites. Even within the class of active sites there are structural and environmental differences. The titanium ions may be at the edge of the crystal, or at the corner, or may have different numbers of anion vacancies around them and so on. Because of this inequivalence the polymer molecules that grow on each of these sites need not have identical structure, length, and molecular weight.

Although Ziegler catalysts are often made in an organic solvent by the treatment of $TiCl_4$ with trialkylaluminum, in terms of molecular level structures they are similar to the heterogeneous catalysts. This is because these are colloidal systems and the molecular structures and environments at and around the titanium ions on the surface of these colloidal assemblies are obviously not identical.

Unlike the first and second generation catalysts, metallocenes are *homogeneous* catalysts. For a given metallocene catalyst, all molecules have the same structure and behave in an identical manner. Therefore the polymer molecules that grow on each of these sites (molecules) should have nearly identical structure and molecular weight. This in fact is what is observed in practice. Consequently the metallocene catalysts are often referred to as *single site* catalysts.

The answers to the second and third questions are based on some of the fundamental and recent concepts of organometallic chemistry. It may be worth remembering that the direct evidence for any sort of a bond in most cases is based on single crystal X-ray study, a technique that became more or less routine only in the early 1970s. During the development of the first generation catalysts it was clear that, among the first row transition metals, titanium showed maximum polymerization activity. Good polymerization activity was also observed with other early transition metals such as vanadium and chromium.

Although the formal charge on the titanium ion at the active site, was assumed to be 4 +, direct evidence was not available. Indeed until the early 1980s there was some speculation on whether or not paramagnetic metal ions (e.g. Ti^{3+}) have a special role to play in the polymerization process. Some elegant and careful work in Du Pont and other laboratories showed that paramagnetic metal ions do not have any special role[17]. Such work also provided direct spectroscopic evidence for all the proposed mechanistic steps of a polymerization reaction. A simplified mechanistic scheme, which is based on standard well-established organometallic reactions, is shown in Figure 3. Here it should be mentioned that recent studies[18] also suggest that in the commercial silica supported chromocene-derived catalyst the active sites are most probably Cr^{3+} rather than Cr^{2+}.

As already mentioned, sophisticated spectroscopic and microscopic studies on $MgCl_2$ supported $TiCl_4$ plus additive based active catalytic systems, have not been reported in the open literature. Cossee and Arlman proposed the original mechanism of polymerization with the first generation catalysts[19]. The mechanistic scheme is virtually the same as that shown in Figure 3. The important point is that the formation of the new carbon to carbon bond was assumed to involve *direct* insertion of the olefin into the metal–carbon bond.

Propylene and ethylene are both planar molecules. Unlike ethylene, in propylene the two faces of the plane have opposite handedness. In chemical terminology this is called a ‘prochiral molecule’. Cossee and Arlman also proposed that for isotactic polypropylene, the titanium centres where the polymer chain grew were accessible to propylene only if it approached the metal centre through one particular face. In other words the coordination environment around titanium had an in-built asymmetry that led to isotacticity.

We mentioned earlier that addition of various bases or the so-called 'electron donors' has a beneficial effect on the stereoregularity of polypropylene. On the basis of kinetic analysis, it is by and large agreed that the 'donors' coordinate to the more active but less selective sites. This increases the isotacticity of the polymer but often at the cost of the overall rate of polymerization.

During the 1970s and 1980s, academic research tried to integrate the mechanism of polymerization with the rapidly advancing frontiers of organometallic chemistry. The evidences for the existence of an intermediate with a metal to carbon *single* bond were many. A large number of transition metal alkyl complexes were synthesized and structurally characterized. One of them reported by Wilkinson in the early 1970s is shown in Figure 4 (structure **4.1**).

The chemistry of complexes with metal to carbon *double* bonds, the so-called metal–carbene complexes[20], was also fully developed by Fischer (structure **4.2**, Figure 4). Instead of the Cossee–Arlman mechanism that involves metal–alkyl intermediates, an alternative mechanism involving a carbene intermediate was proposed by Green *et al*.[21]. Carefully-designed experiments to differentiate between these two mechanisms by using isotopically labelled compounds were subsequently reported by others. No evidence was found for discrete metal–carbene type intermediates.

At present, on the basis of a variety of studies that include single crystal X-ray, multinuclear NMR, kinetic isotope effects, theoretical calculations, etc. we have a reasonably clear molecular level understanding of the mechanism[16,22,23]. As mentioned earlier, the metal to carbon bond is a standard metal to alkyl single bond, and the polymer chain grows by the continuous insertion of the olefin into the metal–carbon bond. Furthermore, a short-lived *interaction* between the metal and the hydrogen atom attached to the carbon nearest to the metal, the a carbon, seems to be important for the tacticity of polypropylene. In the case of polyethylene such an interaction may or may not be present and does not have any practical consequence.

This interaction where the hydrogen atom flirts with the metal atom but does not divorce the carbon and marry the metal is called an 'agostic' (a Greek word) interaction. It has been observed by X-ray and NMR in a large number of organometallic complexes. The currently accepted mechanism for propylene polymerization with metallocene complexes is shown in Figure 5. The basic idea is that in the *transition state* an agostic interaction prevents the rotation around the metal–carbon bond. In this way the relative stereochemical arrangements of the methyl groups in the polypropylene chain are retained.

The answer to question number 4 mentioned earlier is that partial hydrolysis of methyl aluminum generates a mixture of compounds. All these compounds contain methyl groups and many Al–O–Al bonds. These compounds are called methyl aluminoxanes (MAO). The suggested structure of one of the many possible MAO

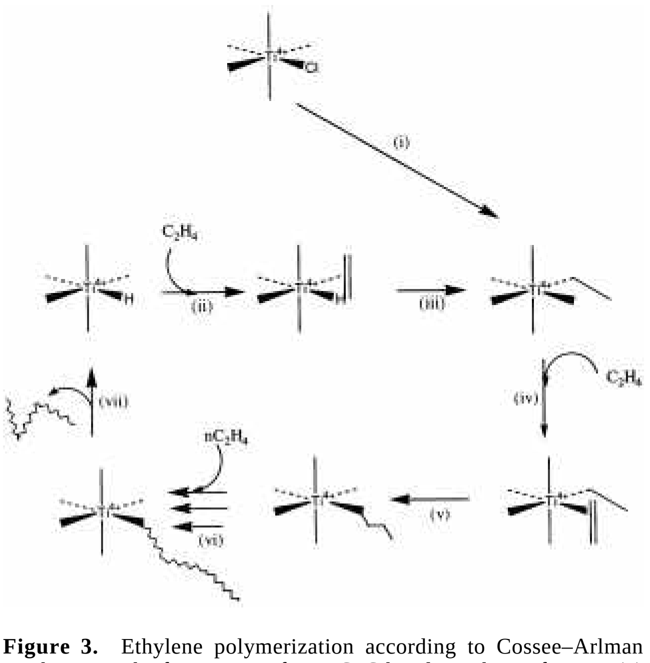

Figure 3. Ethylene polymerization according to Cossee–Arlman mechanism. The formations of new C–C bonds as shown for step (v) is thought to be direct (no intermediates). In step (i) the Et_3Al or Et_2AlCl converts the chloride precursor to a titanium ethyl species.

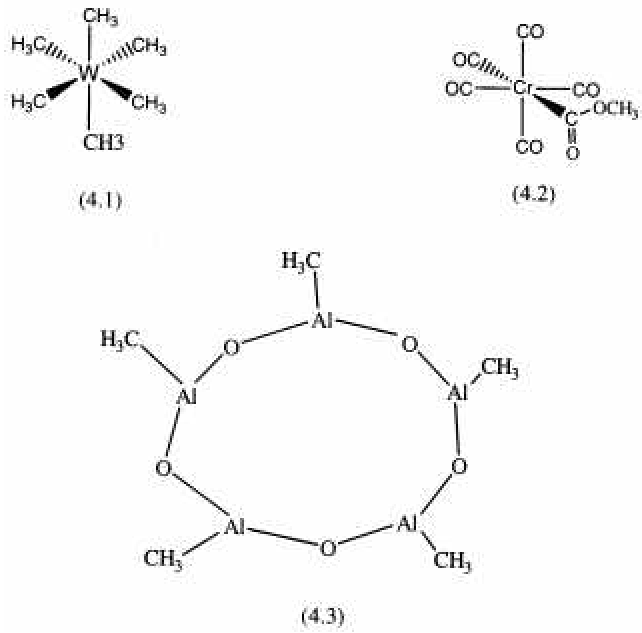

Figure 4. Hexamethyl tungsten (**4.1**) is the first all alkyl transition metal complex made by Wilkinson with all metal to carbon single bonds. The Fischer carbene complex (**4.2**) is the first of its kind with chromium to carbon double bond. To retain tetravalent carbon, the metal to carbon double bond is formally not shown but the bond distance is indicative of such a bond. Structure **4.3** is one of the early MAO molecules characterized by mass spectrometry.

molecules is shown by structure **4.3** in Figure 4. The role of MAO is to pull away a methyl group from the metal centre so that there is room for an olefin to come and bond to the metal. This after all is one of the essential requirements for polymerization. In other words MAO produces what in the homogeneous catalysis parlance is called coordinative unsaturation. The sequence of reactions that lead to polymerization is shown in Figure 5.

The answer to the last question is obviously of tremendous practical importance. A large number of sandwich complexes of titanium and zirconium have been found to be efficient polymerization catalysts. The good catalytic activity of the zirconium complexes is not surprising. After all titanium and zirconium belong to the same group in the periodic table. What olefin and how efficiently it may be polymerized would obviously depend on the specific ligand environment used in the metal complex.

Like many fundamental chemical properties, the symmetry of the catalyst molecule, or more precisely the point group to which the catalyst molecule belongs, has turned out to be the controlling factor in determining the tacticity of polypropylene. Thus by using molecules that are very similar but have different symmetries as catalysts (see structures 2 *c* and *d*) *isotactic* and *syndiotactic* polypropylene could be made in a predictable manner. Chemists would recognize that structures 2 *c* and *d* belong to C_2 and C_s point groups, respectively.

The space available near the metal forces the growing polymer chain and the methyl group of each incoming propylene molecule to adopt a specific relative position. As already mentioned, due to agostic interaction, rotation around the metal to carbon bond is prevented, i.e. the specific relative position of methyl and the polymer chain is retained. In situations where the path of approach of the olefin to the metal centre cannot be random basically because of the bulk of the ligand, symmetry considerations predict that C_2 and C_s symmetry should lead to the formation of isotactic and syndiotactic polymers. This indeed is what is observed in practice. A simple but aesthetically pleasing and practically useful correlation indeed[16,24].

The other major finding is very recent and truly exciting[9,25]. It has been reported that many other ligand and metal combinations, and *not necessarily only sandwich type complexes of titanium or zirconium*, may do the job. What is required is that the ligand forms more than one bond with the metal ion, i.e. it acts as a bi- or tridentate chelate, and is reasonably bulky. An optimum bulk of the ligand is very important for best activity. With these easily synthesized ligands, a cheap innocuous metal like iron gives a very efficient catalyst for the polymerization of ethylene. Like the metallocene catalysts these catalysts are also activated with MAO, and the basic chemistry of polymerization remains the same.

The future and unanswered questions

There is no doubt that metallocenes and similar catalysts will be increasingly used for making special tailor-made polymers. However, there are many reasons that make *total* replacement of the second generation catalysts by metallo-

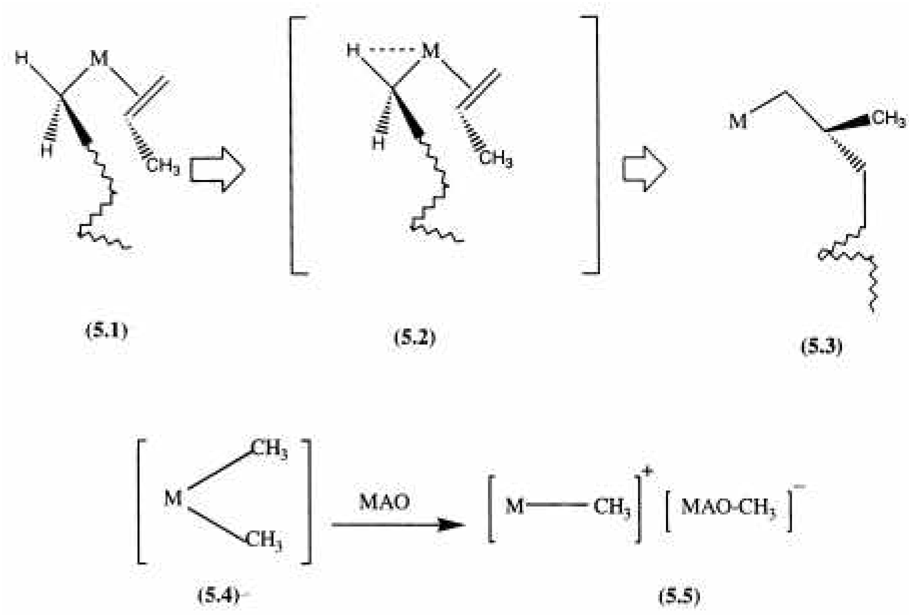

Figure 5. Currently accepted mechanism of polymerization with metallocene catalysts. (Top) Control of tacticity; and (bottom) generation of co-ordinative unsaturation by MAO. The relative *trans* orientation of the methyl group and the growing polymer chain in **5.1** is retained in the transition state **5.2** due to agostic interaction. A broken line represents the agostic interaction between the metal atom and the hydrogen. One of the methyl groups of **5.4** is abstracted by MAO to give the weakly bound ion-pair (**5.5**). The other ligands on the metal are not shown for clarity.

cenes unlikely in the foreseeable future. There are two chemistry-related issues that may be of special interest to the readers. First of all, for polyethylene and polypropylene, the activities of these new catalysts must be significantly greater than that of the conventional catalysts. Otherwise any plant or hardware modification would obviously be uneconomical.

Many approaches have been adopted to home in on the optimum ligand environment that gives best activity. Potentially a promising one is the so-called 'combinatorial approach' reported by Symx technologies where multiple ligand environments can be simultaneously and quickly evaluated[26]. It is interesting to note that a 'combinatorial approach' was originally reported for solid phase peptide synthesis in the area of pharmaceutical research. The second problem with the new generation catalysts is that these catalysts are activated only in the presence of rather large quantities of MAO. A less expensive and effective way of activating these catalysts is required. While a fair amount of success has been achieved in finding other molecules that do the same job as MAO at a much lower concentration, more needs to be done.

In the Indian context the following long-term techno-commercial trends are certain to prevail. The yearly rate of growth in polyolefin consumption in India, shows a robust 'market-pull' for existing and new technology. The demand for new technology will also become urgent as, with globalization and a level playing field, the market for tailor-made special polymers and differentiated products within the country grows.

This presents many opportunities to academic researchers. The possibility of developing new ligand systems, the applications of computational chemistry and sophisticated surface techniques are some of the most obvious directions. As we have seen with the commercial second generation solid catalysts, very little structural information at a molecular level is available. The explanation of the very important roles played by the additives still remains in the realm of hand-waiving arguments and educated conjectures. Unravelling their mysteries no doubt presents difficult research problems. Only a creative multidisciplinary approach may provide answers to some of the 'know why' questions. The traditional boundaries between the different branches of chemistry, and between chemistry and materials science must be crossed if we want our contributions to be recognized and rewarded at the international level.

1. Eisch, J. J., *Chem. Educ.*, 1983, **60**, 1009.
2. Ziegler, K., *Advances in Organometallic Chemistry* (eds Stone, F. G. A. and West, R.), Academic Press, vol. 6, p. 1; Ziegler, K., Holzcamp, E., Martin, H. and Breil, H., *Angew. Chem.*, 1955, **67**, 426; 541.
3. Natta, G., *Angew. Chem.*, 1956, **68**, 393–403; *Sci. Am.*, 1957, **197**, 98; Natta, G., Pino, P., Coradini, P., Danusso, F., Mantica, E., Mazzanti, G. and Moraglio, G., *J. Am. Chem. Soc.*, 1955, **77**, 1708.
4. *Notable Twentieth Century Scientists* (eds McMurray, E. J., Koesk, J. K. and ValadeIII, R. M.), Gale Research Inc, ITP, New York, 1997.
5. Kealy, T. J. and Pauson, P. L., *Nature*, 1951, **168**, 1039; Miller, S. A., Tebboth, J. A. and Termaine, J. F., *J. Chem. Soc.*, 1952, **632**.
6. Wilkinson, G., Rosenblum, M., Whiting, M. C. and Woodward, R. B., *J. Am. Chem. Soc.*, 1952, **74**, 2125.
7. Fischer, E. O. and Pfab, W. A., *Z. Naturforsch. B*, 1952, **7**, 377.
8. Whitesides, G., *Chemtech*, 1992, **22**, 17, 144.
9. Bennett, A. M. A., *Chemtech*, 1999, **29**, 24–28.
10. Kirk, R. E. and Othmer, D. F., *Encyclopedia of Chemical Technology*, John Wiley, New York, 1996, 4th edn, vol. 17, p. 792.
11. Boor, J., Jr., *Ziegler–Natta Catalysts and Polymerization*, Academic Press, 1979.
12. Goodall, B. L., *J. Chem. Educ.*, 1986, **63**, 191.
13. Arxoumanidis, G. G. and Karyannis, N. M., *Chemtech*, 1993, 43–48.
14. Sinn, H. and Kaminsky, W., in *Advances in Organometallic Chemistry* (eds Stone, F. G. A. and West, R.), Academic Press, 1980, vol. 18, p. 99.
15. Richardson, K., *Chem. Br.*, 1994, **30**, 87–88.
16. Britzinger, H. H. *et al.*, *Angew. Chem. Int. Ed. Engl.*, 1995, **34**, 1143.
17. Watson, P. L. and Parshall, G. W., *Acc. Chem. Res.*, 1985, **18**, 51.
18. Theopold, K. H., *Chemtech*, 1997, **27**, 26.
19. Cossee, P. and Arlman, E. J., *J. Catal.*, 1964, **3**, 99.
20. Fischer, E. O., in *Advances in Organometallic Chemistry* (eds Stone, F. G. A. and West, R.), Academic Press, 1976, **14**, 1.
21. Ivin, K. J., Rooney, J. J., Stewart, C. D., Green, M. L. H. and Mahatab, R., *J. Chem. Soc. Chem. Commun.*, 1978, 604.
22. Reddy, S. S. and Sivaram, S., *Progr. Polym. Sci.*, 1995, **20**, 309.
23. Grubbs, R. H. and Coates, G. W., *Acc. Chem. Res.*, 1996, **29**, 85.
24. Coates, G. W. and Waymouth, R. M., *Science*, 1995, **67**, 217.
25. Britovsek, G. J. P., Gibson, V. C. and Wass, D. F., *Angew. Chem. Int. Ed. Engl.*, 1999, **38**, 428.
26. Boussie, T. R., Coutard, C., Turner, H. W., Murphy, V. and Powers, T. S., *Angew. Chem. Int. Ed. Engl.*, 1998, **37**, 3272.

Received 1 December 1999; revised accepted 28 January 2000