interpretation of regression analysis is concerned. However, it seems that a careful reading is needed to remove the misunderstanding. The conclusion that 'irrespective of the arsenic content, 60% of the arsenic is removed by filtration using the clay-candle filter . . .', has not been drawn from the regression analysis. It has been drawn from the actual data shown in Table 2 of our article. The scatter plot with a best-fit line has been shown to emphasize the fact that there is, in fact, no statistically significant correlation. In other words, the 'per cent arsenic removal' has no linear dependency on arsenic concentration. As shown in Table 2, 100% removal has been achieved for samples with arsenic concentration as low as 0.02 to as high as 0.2 mg/l. On the other hand, less than 20% removal has been achieved for other samples with similar arsenic concentration (0.06 to 0.3 mg/l). The non-dependence of arsenic removal on concentration is evident from the data itself. Regression analysis was not used for reaching this conclusion. Similarly, 60% arsenic removal is also reflected in the data.

The central theme of our article is that if a set of factors is present in the environment, arsenic can be removed to an optimum extent by clay-candle filter, to benefit a large section of people. In the conclusion, it has been mentioned that these areas where favourable factors operate, can be identified by field studies. In these areas mitigation can be done by inexpensive clay-candle filter. For other areas, expensive filters can be used. We emphasize that our conclusions do not hinge upon any simple statistics as understood and expressed by Shashidhar.

S. SENGUPTA

*Geological Survey of India,*
*27, Jawaharlal Nehru Road,*
*Kolkata 700 013, India*
*e-mail: ssgpta@vsnl.net*

# Error in angles

There is an error in the angles shown in Figure 3 of the communication by Balasubramaniam and Dass[1].

According to the formula of spherical astronomy,

$$\sin\delta = \sin\phi \cos Z + \cos\phi \sin Z \cos A,$$

where $\delta$ is the declination of a celestial body, $\phi$ is the latitude of the place, $Z$ is the Zenith distance of that body and $A$ its azimuth measured from north. At sunrise or sunset $Z = 90°$, so $\cos A = \sin\delta/\cos\phi$.

On 21 June of AD 400, $\delta$(sun) = 23°39′, while $\phi$ (Udayagiri) = 23°31′. These give $A = \pm$ 64°4′ from north. Similarly, on 21 December of AD 400, $\delta$(sun) = – 23°39′ which gives $A = \pm$ 115°56′ from north. Consequently, with respect to EW line the angles will be 25°56′ and not 23.5°.

1. Balasubramaniam, R. and Dass, M. I., *Curr. Sci.*, 2004, **86**, 1134–1142.

K. D. ABHYANKAR

*'Akashaganga'*
*1-5-76, Vivekanandanagar,*
*Habsiguda, Street No. 8/26,*
*Hyderabad 500 007, India*
*e-mail: kda@ouastr.ernet.in*

*Response:*

We had provided the angle of 23.5° in Figure 3 in the general sense and did not calculate the exact angle based on spherical astronomy. We stand corrected and acknowledge Abhayankar for his valuable input. As was pointed out in the communication, the Udayagiri site needs to be studied in great detail by knowledgeable astronomers. The relationship of the early morning sun with the archaeological structures at Udayagiri must be carefully recorded over a period of at least a year, and critically analysed.

R. BALASUBRAMANIAM

*Department of Materials and*
*Metallurgical Engineering,*
*Indian Institute of Technology,*
*Kanpur 208 016, India*
*e-mail: bala@iitk.ac.in*

**NEWS**

NEWS FOCUS

# Grid computing: Technology propelling frontiers of research

*I believe grid computing will revolutionize the way we compute, in much the same way as the World Wide Web and Internet changed the way we communicate.*

— John Ellis
Adviser to Director General, CERN

## The Internet and the World Wide Web

CERN, the European Organization for Nuclear Research near Geneva on the French/Swiss border, changed the way in which the world interacts with computers, through the work of the then CERN-based British researcher Tim Berners-Lee. He is credited with the creation of the World Wide Web a decade ago, as a means of allowing physicists to share documents. His vision created an 'information space of meaning built upon the World Wide Web'. In the words of Berners-Lee, 'The first breakthrough was the Internet, and I can't emphasize too often that I didn't invent the Internet! There were many networks, but they were of different types, some small, some large, and they used different sorts of connection. A computer could be on more than one network, and it was Vint Cerf and his colleagues who realized that a computer connected to more than one network could act as a kind of postal sorting office, and be used to forward information between the networks. Even though the little networks might use different numbering schemes for different computers, they imagined that each computer was on some global 'Inter-network' and gave each computer a number. To describe things simply, the information is passed around in little packets (rather, as Vint says, like postcards) and each has on it a number that is the address of the computer to which it has to be delivered. The forwarding computers just look at the address number on each packet to figure out which network to send it over next. In this way, all you have to do is send off a packet with the right address number on it, and sooner or later it will arrive at the right place. The Internet was invented around the 1970s. I was fortunate in that in 1989, when I was looking at the problems of networked information systems, it was deployed across the US and to a certain extent in Europe'[1].

## The grid concept

Like the Internet and the Web, a 'grid network' is 'computer collective, gaining its power from the clustering of many machines'. Whereas the Internet is used primarily as a huge communication network, a computing grid harnesses the processing power of all those linked PCs, creating a large virtual supercomputer. All those processors can be put to work on 'the constituent elements of very big computing problems, an approach known as distributed computing'.

Ian Foster, Argonne National Laboratory and Carl Kesselman, University of Southern California's Information Sciences Institute in Los Angeles were the first to envision a concept, namely the Globus Project, a quasi-OS for grid computing[2]. The term 'grid' was coined in mid-1990s. A grid has been described as 'An infrastructure that enables flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions and resources'. What Linus Torvalds, co-creator of the Linux operating system, is to the open-source movement, Foster, is said to be to the world of grid computing.

Grid technology is perhaps in its infancy at present. But, 'just as the Internet grew from a collection of small academic networks to a humongous octopus spreading its tentacles around the world, Foster predicts that "today's minigrids will grow into a huge global grid, a transcontinental processing pool engaged in all sorts of complex tasks, like designing and testing semiconductors and decoding the human genome. Tasks will be broken down, distributed to millions of connected processors, and then the reassembled results would be sent back to a single desktop".' The grid is somewhat utopian in its concept in the sense that it gives 'users at far-flung terminals access to supercomputer-grade power currently available to only a few researchers at chosen facilities'.

A good analogue of the grid is described by the following statement: 'Grid computing is like the national electric grid: A valuable resource, somewhere in a pool of such resources, is used, as needed, for a particular purpose. In this case, computing power, not electricity, is the utility. And instead of electric generators, it's idle, relatively idle, or overpowered workstations and servers that are combined into a single, federated computing grid'.

The company Gridware made a software that tied together many desktop computers into a minigrid that could perform like a supercomputer. It was such a big success, saving Ford as much as US $ 100 million. BMW, Boeing, Motorola, Novartis, Pacific Life, Saab and Synopsys are also using minigrids.

Unlike the Internet, which primarily is a network for communications, grids are networks for computation. Grids consist of high-end computers, servers, workstations, storage systems and databases that work in tandem across private and public networks. The crux of the project is to enable computers which are interconnected via the grid understand each other's language, protocols and the like. However, grids are not yet in the commercial domain because of concerns regarding security, privacy, resource sharing, etc. Grid users are expected 'to have at their disposal distributed high performance computers able to access and process terabytes of data stored in global databases, plus the appropriate tools to control these resources'. A suite of software tools known as 'Globus middleware' is designed for this purpose.

## Structure of the grid

One can talk of many types of grids:

A data grid is used for sharing information. For example, in the US, IBM and the University of Pennsylvania share a grid that connects hospitals at the University of Pennsylvania, University of Chicago, University of North Carolina, and the Sunnybrook and Women's College Hospital, Toronto. It is funded by the National Library of Medicine. Researchers can use the grid to share information and analytical tools about mammograms.

A computing grid is for fast, heavy number crunching. The smallpox and anthrax grids that IBM supports with grid.org are examples of this type of grid. As far as the smallpox grid is concerned, it works on the basis that 'there is a possible molecular target whose blockade would prevent the ravages of an infection (of smallpox). Grid computing is intended to be used to screen millions of potential anti-smallpox drugs against this target'.

The project can harness millions of computers belonging to people in over two hundred countries, all of whom will benefit from protection against smallpox. The anthrax project helped molecular screening of 3.57 billion molecules against an anthrax-related target protein in just 24 days. The challenge was to accelerate the process of identifying lead molecular candidates to ultimately develop a treatment for advanced-stage anthrax. According to Graham Richards, Director, Centre for Computational Drug Design, Oxford University, 'Had we done this using traditional methods, it would have taken years instead of less than four weeks'.

Although these two types of grids are identified based on the tasks that are basic to them, one also notes that conceptually, the grid is also thought of in terms of three layers. Underlying everything is the 'computational and data grid', the computer hardware and data networks upon which work will be conducted. Above this is the 'information grid': the databases of information to be accessed by the hardware, and systems for data manipulation. On top is the 'knowledge grid', where high-level applications will mine the data for knowledge that can form the basis of semantic understanding and intelligent decision-making.

There are other types of grids: enterprise grid, partner grid, service grid and so on. These are minigrids limited to an enterprise or a few partners in business and so on. What we are concerned with here is a wider concept referred as 'global grid', wherein one deals with global resources shared over the internet, taking into account globally distributed datasets.

Key to successful grid computing will be the 'middleware', the name given to a computer software 'that can organize and integrate disparate computational facilities into a coherent whole, automatically undertaking all the machine-to-machine negotiations required to weave the world's computational and informational resources into a seamless fabric for e-research, e-learning and e-commerce, while permitting individual users customized access or "portals" to those facilities required for the tasks at hand'. In other words middleware ties together the computing elements on the grid, as well as grid software that keeps track of availability of resources to schedule processing time accordingly.

In operation, grid-computing software divides a computational job into multiple, smaller, discrete tasks that are then distributed to the various computing nodes on the grid. As individual computing tasks finish, the computing nodes send their results back to the grid server. The server, in turn, compiles the results and, if necessary, sends new tasks to the computing nodes. When a job is complete, the results are made available as in any conventional client/server environment.

There are three key infrastructure requirements for enabling and implementing a grid: a solid network, accessible data and standard system images.

In relation to the middleware, one identifies a few salient components. When a job is submitted, the 'resource broker negotiates with computers signed up to its network to book the computing power the job needs' from among the various resources available (including supercomputers and PCs at various locations). 'It also scours data centres worldwide for the files and data that the job needs, and, if need be, converts them into compatible formats'. It will enable even the software needed to perform a particular analysis.

The 'information service' keeps track on-line 'information on where and how much computing power is available, and how fast various data links are running'. The 'replica catalogue' keeps track 'of the type and whereabouts of all data and files'. 'The "logging and bookkeeping" server tracks the status and history of all submitted jobs'. So in totality, the middleware acts to interface among the users, the data they wish to process, and the computational resources required for this processing. 'Central to these interactions is metadata, that is to say descriptive information about these three types of entity, organized in a systematic manner that makes automated interactions possible'. Then there are 'agents' to represent the users, the data and the resources on the grid by presenting their metadata, and 'brokers' to 'undertake machine-to-machine negotiations required for user authentication and authorization, and strike "deals" for access to and payment for specific data and resources, and then schedule the computational activities and oversee the data transfers required for the particular task to be undertaken, all this occurring in a fraction of the time that it would take the user to do the same manually'.

## Grids in operation

'Historically, grid computing has been associated with solving very specific, generally large, complex problems'. The SETI@Home (Search for Extra-Terrestrial Intelligence) project is likely the best known for its use of millions of Internet-connected PCs via a clever screensaver. In fact, the screensaver is simple to install and get going by downloading the screensaver from http://setiathome.ssl.berkeley.edu/. When the PC (or computer) is unattended, the PC/computer 'will be analysing the data specially captured by Arecibo radio telescope, the world's largest radio telescope in the quest for extraterrestrial intelligence'. But it shuts down as soon as the user resumes his work on the PC. SETI@home has nearly 3 million users across the world, making use of the most powerful virtual supercomputer on the planet. Because these projects are distributed throughout the world – like the Internet – taking a few thousand machines off-line, during any blackouts there will be almost no impact on the performance of the grid. SETI@home completed five years of operation on 17 May 2004. As of March 2003, 'for nearly four years, the SETI@home receiver atop the giant Arecibo radio telescope has been surveying the skies in search of an alien signal. For nearly four years, millions around the world have been processing the raw data from Arecibo on their personal computers in search of unique patterns that might represent an intelligent transmission. No less than 5 billion(!) different candidate signals have now accumulated at SETI@home headquarters in Berkeley. Each of these just might be that one true signal from an alien civilization'. For the first time, SETI@home scientists put this vast amount of data to a test and that was 'for eight hours each day, on 18 March through the 20 March 2003, SETI@home scientists had the full use of the Arecibo radio telescope. They used it to target between 100 and 200 locations in the sky where the strongest, clearest, and most promising candidate signals have previously been detected. Only a candidate signal that has been revisited and confirmed in this manner can be considered to be a potential intelligent transmission from the stars'.

Apart from SETI@home, the impetus to develop the grid has come from the scientific community, for example, to process the vast amounts of bioinformatics data that have been produced by the Human Genome Project, and the terabytes of particle physics data that will soon flow from the Large Hadron Collider at CERN. This has been supported by significant investment of public funds. Coordinating the worldwide efforts to develop grid technology is the Global Grid Forum (www.gridforum.org/).

The success of SETI@home encouraged other grid applications such as the (protein) Folding@Home project, which investigates potential cures for Alzheimer's, AIDS and Parkinson's diseases. The latest 'personal' grid computing project is ClimatePrediction.net, which hopes to facilitate better understanding of the probability and impacts of global warming'. There have been other application middlewares to gene sequencing, numerical analysis, simulation and modeling, etc.

Folding@Home, run by Vijay Pande, simulates the process of self-assembly of a protein, making use of spare computing cycles of ordinary home computers. The project was the first to successfully model a complete protein fold, a feat that had eluded supercomputers. It has also succeeded in simulating how proteins misfold.

Why does one have to understand how proteins fold or misfold? 'Proteins, which control all cellular functions in the human body, fold into highly complex, three-dimensional shapes that determine their function. Misfolds can change the shape of the protein, turning a desirable protein into a disease'. Certain small-chain proteins, for example, peptides are known to misfold in patients suffering from Alzheimer's disease and Type 2 diabetes. How is the grid going to help in the case of diseases like the Alzheimer's? NSF's middleware initiative is helping in the growth of Biomedical Informatics Research Network to large-scale data sharing and analysis. This enables to increase 'the ability to share and compare massive datasets such as MRI brain scans or high-resolution electron microscopy images, essential to participants' research into Alzheimer's disease, depression, schizophrenia, multiple sclerosis and other disorders'.

In another project Genome@home, Pande and tens of thousands of coworkers attempt to design new proteins from existing genomes. 'The goal of Genome@home is to design new genes that can form working proteins in the cell. Genome@home uses a computer algorithm (SPA), based on the physical and biochemical rules by which genes and proteins behave, to design new proteins (and hence new genes) that have not been found in nature. By comparing these 'virtual genomes' to those found in nature, we can gain a much better understanding of how natural genomes have evolved and how natural genes and proteins work'. In fact, in the write-up entitled 'How you can help' to be found at htpp://www.standford.edu/group/pandegroup/genome/using.html#help, the promoters of this activity state, 'To design these large numbers of protein sequences, we need lots of computers. By running the Genome@Home protein sequence design client, you can lend us your computer while you're not using it, for as long or as little as you like. It simply runs alongside your other programs and does its calculations in the unused cpu time while you're away from your desk, or even while you're working on your computer. You won't notice a loss of speed, and your computer will work as usual. All you see is a small window that shows you the protein sequences you're designing. If you don't want to look at it, just minimize the window and move it to a corner of your desktop. A day or two's worth of running Genome@Home is enough to design new protein sequences that the world has never seen before. All the sequences get added to the Genome@Home database, so every little bit helps'. Here we have given two examples, namely those of Seti@Home and Genome@Home to indicate how simple it is for anyone from anywhere to 'partake' in these and perhaps other projects which are already on. 'The best way to get first hand experience with Grid Computing is to do it yourself. . . . Grid projects that you can lend CPU time to, include fighting Aids, doing genome research, participating in stock market forecasting using neural networks or finding the 5 largest prime numbers. . . . Users can even set-up their own Grids by downloading from small, Windows only InfoDesign's DeskGrid which works on Linux and Solaris'.

In Intel's Cancer Project, about 556,000 users participate. As if to appeal to the social conscience, Scott Griffin, the program manager of Cancer Project noted 'We want to increase the value of the PC; the PC is there when people aren't at it, like when they are in meetings. A great thing about this is you get every-day users involved in research that they care about. Not only do they get to help out, but they get to help cure these terrible diseases'.

## Other developments

According to an announcement in 2001 from the National Science Foundation, four US supercomputer centres were to be linked together into one massive 'grid'-style computer. The facility would help researchers understand the origins of the universe, cure cancer, unlock secrets of the brain, predict tornadoes, and save lives in an earthquake. The project costing nearly $50 million covered over 1000 servers from IBM, Intel's second-generation Itanium processor, communication links among server clusters, etc. together into a virtual computer. The core software is the set of open source tools, namely the Globus. The facility, called the 'TerraGrid', was aimed at processing over 11 trillion commands per second. The TerraGrid is based on 'the open source development model, meaning no one will own the software it develops'.

According to Dan Reed, Director of the National Center for Supercomputing Applications, Illinois: 'It will allow scientists to collaborate in a way they haven't been able to before, to eliminate the tyranny of time and space limitations'.

Similarly, IBM was involved in a contract to connect four research facilities in Britain together into a grid network.

## CERN and the data grid

### *Background*

Physicists are never satisfied by the number-crunching power that they have at their fingertips at any time, more so now when demands of experimental elementary particle physics to detect signatures of exotic subatomic particles such as the Higgs boson, in the midst of large 'noise' from the detectors is on the crescendo.

Computer networks have been of importance at CERN since the early 1970s, when the first links were set-up between experiments and the computer centre. The first external links were established during the late 1970s and had limited purposes, such as remote job submission and output file retrieval. The first international 2 Mbps (megabits per second) link was installed by INFN during the summer of 1989, just in time for the start-up of CERN's large electron positron collider. However, there was still no Europe-wide consensus on a common protocol, and as a consequence multiple backbones had to be maintained, e.g. DECnet, SNA, X25 and TCP/IP (Transmission Control Protocol/Internet Protocol).

Independently in the US, during late 1988, the National Science Foundation (NSF) established NSFnet, the first TCP/IP-based nationwide 1.5 Mbps backbone. This was initially used to connect the NSF-sponsored supercomputer centres and was later extended to serve regional networks, which themselves connected universities. The NSFnet served as the commercial Internet backbone until 1990 when CERN with IBM and other academic partners in Europe developed the use of EASInet (European Academic Supercomputer Initiative Network), a multi-protocol backbone at 2 Mbps TCP/IP for the European researchers and had a 1.5 Mbps link to NSFnet through Cornell University. This was the beginning of European Internet. These developments established TCP/IP as the major protocol for Internet backbones around the world.

The 'large hadron collider' (LHC) at CERN will be the most powerful atom-smashing machine that is expected to generate several petabytes (10,000 gigabytes) of data annually, which is 'more than any existing supercomputer or grid can cope with'. 'Even the most basic analysis of these data would require 20 teraflops of computing power (a teraflop is a trillion floating-point operations per second and is a measure of a computer's speed). Yet the world's most powerful supercomputer can do only 3 teraflops per second'. Hence CERN initiated a major European project to create a vast 'grid' research network of computers across

Europe. When completed, the 10 million euro, Linux-based endeavour called DataGrid, will become a principal European computing resource for researchers of many disciplines. DataGrid is equivalent in size and importance to US grid networks, such as the Grid Physics Network. [The initiative called the Grid Physics Network or GriPhyN (pronounced 'griffin'), is funded by NSF. The GriPhyN project is developing grid technologies for scientific and engineering projects that must collect and analyse distributed, petabyte-scale datasets.] Unlike SETI@home, DataGrid is in real time and on-line.

CERN is relying on the network being able to process the data streams by 2006, when LHC is scheduled to be completed. Getting the LHC grid on-line requires formal collaboration between the IT service departments of the institutions taking part, to agree on joint standards and best practices.

### *Recent achievements*

The DataGrid project is not just of benefit to CERN. Webs of computers that form the overall grid, spread across Europe, will benefit a host of other user communities. European astronomers, meteorologists studying global weather patterns, biologists engaged in genetics research and organizations such as the European Space Agency and many others are involved and will be using the DataGrid. According to Ellis, 'grid computing is an explosive moment in the development of computing brainpower'.

On five occasions during 2003, a team led by Harvey Newman, Caltech and Olivier Martin, CERN established new records for long-distance data transfer. This year, new records are expected to be set as the performance of single-stream TCP is pushed closer towards 10 Gbps (gigabits per second). In 1980, 'high speed' meant data transfers of 9.6 kbps (kilobits per second), using analogue transmission lines. So the achievement of 10 Gbps in 2004 corresponds to an increase by a factor of 1 million in 25 years – an advance that is more impressive than the classic 'Moore's law' of computer processing, in which the number of transistors per integrated circuit (i.e. the processing power) increases by a factor of two every 18 months, or 1000 every 15 years.

Newman has pointed out, 'these records are a major milestone towards the goal of providing on-demand access to high-energy physics data from around the world, using servers that are affordable to physicists from all regions. Indeed, for the first time in the history of wide-area networking, performance has been limited only by the end systems and not by the network: servers side by side have the same TCP performance as servers separated by 10,000 km'.

## Is grid-computing only a 'vision'?

There are already tangible results for one to see and appreciate. 'In a three-month project centred on the Supercomputing 2003 Conference (SC 2003) last November in Phoenix Arizona, a Boston–UK team of scientists linked more than 6000 processors and 17 teraflops of computing at six different facilities on two continents.' 'Their grid-based effort has led to significant new scientific understanding and represents probably the most impressive example to date', say scientists, 'of how grid computing gives a powerful boost to large-scale scientific computation'.

Referred to as the TeraGyroid Project, with an approach called 'computational steering', 'the researchers linked the National Science Foundation's TeraGrid with the UK E-Science Grid via dedicated trans-Atlantic fiber. They tied together resources at Daresbury Lab, UK and at four TeraGrid sites: Pittsburgh Supercomputing Center, the National Center for Supercomputing Applications (NCSA), San Diego Supercomputing Center and Argonne National Laboratory along with others'. They carried out the largest simulation of its type (lattice-Boltzmann method) to date, with resolution of more than a billion lattice-sites. Their work focused on the shape of complex materials, known as gyroids, with properties between those of a liquid and those of a solid.

These researchers from the US and UK 'cocreated a demonstration at the SC 2003, in which they simulated a phase-separating amphiphilic fluid – a mixture of oil, water, and a surfactant, separating due to the surface tension of the oil/water interface, but ultimately stabilized due to the surfactant'. Such mixtures can form fascinating 'liquid crystals', including an elusive 'gyroid' phase that has been seen in the laboratory, but has been notoriously difficult to simulate. They were able to evolve a homogenized ternary mixture of oil, water and surfactant to a stable gyroid phase, and study the evolution of dislocations–imperfections in the liquid crystalline structure. Such a simulation must necessarily be done on a large grid, since small grid sizes will suppress the formation of the dislocations. Boghosian's team ran simulations on grids of various sizes on a variety of machines, including the Pittsburgh Supercomputer Center, and NCSA. They generated graphical output from these simulations on machines in London and Manchester. As the output was rendered, . . . was able to alter parameters in response to what appeared on the screen, modifying the simulation 'on the fly' – a technique called 'computational steering'. All of this was broadcast over the AccessGrid to the Supercomputing meeting in Arizona, and numerous other sites throughout the world. The presentation was awarded the HPC Challenge Award for the 'Most Innovative Data-Intensive Application'. More recently, the project received a 2004 ISC Award, the major supercomputing award in Europe, for 'Integrated Data and Information Management'.

According to Rick Stevens, Argonne National Laboratory and the University of Chicago, TeraGrid project director, 'The TeraGyroid Project exemplifies what's possible with Grid technologies', 'It's a major success for the NSF vision of integrated national cyber infrastructure, and it helps us to appreciate that–just as the economy is global – with the grid, science too has become global'.

Boghosian, Tufts University and Coveney, UK emphasize that the vision of grid computing – 'transparent' access to resources without regard to location – is not yet a reality, and it was a large task to coordinate this ambitious project, which involved about 30 organizations and more than 100 individuals.

The TeraGrid is a multi-year effort to build and deploy the world's largest, most comprehensive distributed infrastructure for open scientific research. The TeraGrid also offers storage, visualization, database and data-collection capabilities. Hardware at multiple sites across the country is networked through a 40 Gbps second backplane – the fastest research network on the planet.

### Indian scenario

In the Indian context, it is realized that 'stand-alone supercomputers are passé. The decade-old massively parallel processing systems are increasingly considered old hat. Cluster computing is "in" and grids are "hot".' C-DAC, a scientific society under Ministry of Information and Communications Technology, was set-up in 1988 to design and develop high-performance computing systems. C-DAC has been developing the PARAM series of computers for this purpose from Pune. It has begun work on the India-Grid or I-Grid to create 10 supercomputing sites in the country, contributing totally ten teraflops of computing power. The I-Grid is expected to help Indian scientists initiate/carry out a variety of scientific and engineering applications like *N*-body simulations, finite element and finite difference algorithms on large-scale problems at a faster rate. The applications may cover various disciplines like computational-atmospherics, -chemistry, -structural mechanics, -fluid dynamics, evolutionary computing and seismic data processing. The I-Grid project is expected to be fully operational by 2007, with Pune and Bangalore scheduled as the hubs of grid activity. It may eventually become a part of the Global Grid initiated during SC2003.

To conclude, 'The vision for the future using grid computing is that we will never have to concern ourselves again about where our files are stored, or on which continent the computer processing our data is located. The grid will transparently give us seamless access to the globally distributed computational resources at our disposal, so that we can get our work done in the most efficient manner possible'. However some opine that 'a great many challenges have to be overcome before grid computing may be a common reality'.

### Bibliography


1. Berners-Lee, T., 'Japan Prize 2002 Commemorative Lecture'.
2. Foster, I. and Kesselman, C., *The Grid: Blueprint for a New Computing Infrastructure* (Morgan Kaufmann, now part of Elsevier), 1999.
3. Joseph, J. and Fellenstein, C., *Grid Computing*, Prentice Hall PTR Series; IBM Press, 2003.
4. Foster, I. *et al.*, Anatomy of the grid. http://www.globus.org/research/papers/anatomy.pdf
5. Foster, I., *Phys. Today*, 2002, **55**, 42–47.
6. NLANR, Beginner's Guide to Network-Distributed Resource Usage.
7. Other articles, especially from www.gridtoday.com


**K. R. Rao,** 29/2, 11th Cross Road, Off 3rd Main (Margosa) Road, Malleswaram, Bangalore 560 003, India. e-mail: krrias@yahoo.com

## RESEARCH NEWS

# ATP synthase and the torsional mechanism: Resolving a 50-year-old mystery

*S. Jain, R. Murugavel and L. D. Hansen*

As the energy currency and regulator of cellular metabolism, adenosine triphosphate (ATP) participates in all the pathways of metabolism. ATP is synthesized from adenosine diphosphate (ADP) and inorganic phosphate ($P_i$) by the enzyme $F_1F_0$-ATP synthase which is located in the cytoplasmic membranes of prokaryotes and in the membranes of mitochondria and chloroplasts of eukaryotes. ATP synthase has two domains, the $F_0$ domain is a transmembrane domain and the $F_1$ domain, which is joined to the $F_0$ portion by the $\gamma$-subunit of the enzyme, is the site of ATP synthesis (Figure 1). The $F_1$ part of ATP synthase was first isolated in 1960 by Ephraim Racker and colleagues[1] and the first high-resolution structure of the $F_1$ portion of the enzyme was solved by Abrahams *et al.*[2] in 1994.

The mechanism of ATP synthesis was a major question even before the structure was known. The chemiosmotic hypothesis of oxidative phosphorylation, proposed by Peter Mitchell[3] in the 1960s, postulated a delocalized, protonmotive force, with both its components ($\Delta$pH and $\Delta\psi$) being created by a proton gradient, as the driving force for ATP synthesis. That is, the organelle was seen as a delocalized electric cell that used a proton current as the sole energy source for ATP synthesis. Mitchell was awarded the Nobel Prize in Chemistry in 1978. The binding change mechanism was postulated by Paul Boyer[4] in 1973, to explain how the proton current and ATP synthesis were coupled. In the binding change mechanism, energy stored as ion gradients across the membrane containing the $F_0$ domain is used for free rotation of the c-rotor, the $\gamma$-shaft and the $\varepsilon$-subunit attached to the rotor. This free rotation gets translated into binding changes in the catalytic sites in the $\beta$-subunits of the $F_1$ domain causing ADP and $P_i$ to combine spontaneously to form ATP, followed by endergonic release of the product. Boyer shared the Chemistry Nobel Prize for this work in 1997.

Both the chemiosmotic and binding change mechanisms have been modified over time in efforts to make them consistent with more recent experimental data. However, there are basic flaws in both the Mitchell and Boyer mechanisms that cannot be corrected by minor modifications. A fundamental problem that is not addressed correctly by either mechanism is the thermodynamics of energy coupling between transmembrane ion currents and ATP synthesis. Both mechanisms were developed from a static equilibrium viewpoint, and both can be shown to violate the laws of thermodynamics for a dynamic system (Box 1). The novel element in Boyer's mechanism is that the energy of the proton gradient is not used in the synthesis step, but only to release the ATP from the ATP synthase enzyme. At no step in the binding change mechanism does energy get locked into the