Abstract
This paper first summarises the motivation of the international ITER project [1] to develop the International Tokamak Experimental Reactor and illustrates its present status. In a second part of the paper, the likely data handling requirements of ITER are discussed with a view to defining those parts which are conventional and those parts which will need more attention. The question of why the potentially enormous quantity of data should be archived is raised.

FUSION – A FUTURE ENERGY SOURCE
At the time of writing, the world is realising that the moment of reducing global production of oil is approaching, confirming the predictions of “Hubbert’s Peak”. Replacing the lost energy source with coal is not acceptable at a time of increasing awareness of the problems of global warming. Replacing the lost energy source with fission reactors is feasible but does not have popular acceptance. Replacing the high concentrations of power needed by modern society is implausible with many of the greener alternative or renewable energy sources. The quest for a clean and plentiful enough energy source for this century and beyond is therefore of extreme urgency, given the inevitably long lead-time for such a venture.

Fusion has long been considered as offering an attractive energy source, already noted in the 1920’s. Developing a weapon demonstrated that the underlying physics was understood. Developing an economically useful, controlled source of energy was predicted, correctly, to be a long uphill challenge. However, once mastered, fusion will provide a limitless supply of energy with acceptable environmental impact, no problems of proliferation and no danger of nuclear runaway reactions.

Over the last 20 years, the tokamak device has been the top performer for confining the thermal energy of a plasma. The requirement for the fuel being in the plasma state derives from the extremely high temperatures at which the fuel reacts, around 200,000,000 degrees, at which temperature all matter is in the plasma state. The energy needs to be confined long enough for the plasma to stay hot in a balance between the self-heating of the plasma through the fusion reaction products and the losses due to radiation and imperfect confinement by the magnetic field. The use of magnetic fields derives from the interaction between the charged plasma particles and the magnetic fields, which minimises the impact of high-energy particles on the surrounding structures.

In a nutshell, we have to obtain a minimal value of the plasma temperature times the plasma density times the plasma confinement time. Between the mid 60s and the mid 90s, the achieved “triple product” increased with a doubling time of just under 2 years, similar to semiconductor capacity and accelerator energy. Why? The simple answer is that we have built larger devices and have learned how to optimise their performance. This experience is embodied in empirical scaling laws for plasma confinement which describe existing devices with reasonable precision (~10%) and extrapolate to ITER.

WHY MUST WE BUILD ITER?
Extrapolating a set of experiments to a new device is conventional in many engineering fields, but a primary need is the certainty that the physics basis is continuous over the extrapolation. Boat hull design is a good example. Aircraft breaking the sound barrier is a bad example of physics continuity. A key mission of ITER is therefore to confirm that the underlying physics basis extrapolates to a burning plasma. Note that extrapolating from small tokamaks to the largest tokamak JET [2] had the same purpose and successfully demonstrated continuity of the physics, while discovering new phenomena. JET has already produced 16 MW of fusion power transiently and 4 MW for over 4 seconds. ITER’s job will be to increase the fusion power to power reactor levels, and to extend the burn time towards steady state.

A second mission is to provide a design base for the technology of a reactor. The most difficult task facing a fusion reactor designer is that of developing components that can withstand a high neutron fluence or can be replaced quickly if they should fail. The basic fusion reaction D+T→He+n generates a 14 MeV neutron, able to displace the lattice atoms in metal alloys to the level of 100’s of displacement per atom, an unprecedented challenge. In addition, the need to operate the plasma almost continuously with the best possible confinement leads to the adoption of high field superconducting magnets which have to work reliably under high stress for the reactor life.

These challenges cannot be met with the existing scaled down devices. The next step simply has to be taken.

1E-mail address: Jo.Lister@epfl.ch
THE ITER DESIGN BASIS

Using the empirical description of the performance of existing tokamaks, one can conceive a multitude of different devices and predict their performance. Each will have a different cost. Finding the lowest cost for the minimum required performance, compatible with reliable tested engineering has been the task of the central ITER design team, supported by Home Teams from the different partners, currently China, the EU plus Canada, Japan, Korea, the Russian Federation and the USA.

The result was a large device, finalised in 1998, which was rejected politically, despite the support of the partners, as being too expensive. The design team conceived a new version, maintaining the strategic objective of a single step between today’s experiments and the electrical-power-producing experiment that is to follow, while reducing the cost by 50%. Since the original design was well optimised, the final ITER design inevitably has a lower performance margin and lower technical margins on the different systems. The following list summarises the physical description of ITER:

- Fusion power: 500 MW
- Plasma major radius: 6.2 m
- Plasma minor radius: 2.0 m
- Plasma current: 15 MA
- Toroidal field: 5.3 T
- Plasma volume: 837 m$^3$
- Heating power: 73 MW
- Pulse lengths: 300-5000 seconds

More information and illustrations are available on the ITER web site [1] and background information can be found in [2] for JET and the EU fusion programme and [3] for fusion research in general.

THE MAJOR CHALLENGES

The ITER project presents many technical challenges. Among these, its size and technical complexity are foremost. The cylindrical cryostat housing the tokamak is 28m diameter by 24m high. The diameter of the largest horizontal coils, seen in Fig.1, are such that on-site winding may be necessary. The weight of the superconducting coils, 18*312 tons for the steady field and 925+6*130-390 for the pulsed fields means that new facilities must be constructed to produce the superconducting filaments. The peak electrical power required to start up the plasma and bring it to the required temperatures for burning the D+T fuel is several hundreds of MW. The peak power onto the surfaces which face the plasma can briefly reach 20MW/m$^2$ which is at the limit of conventional technology. The precision with which the whole structure must be assembled, so that the magnetic fields are correctly aligned, is at the limit of measurement accuracy.

Coupled with these engineering challenges, the plasma must be initiated, brought to high temperature and sustained for 100’s of seconds with carefully controlled contact with the surrounding structures. Once all this is achieved, there remains the scientific challenge of understanding the performance and finding routes to increasing it, the realm of the plasma physicist. To achieve this last aim, a battery of sophisticated measuring equipment, or diagnostics, surrounds the plasma. These record the radiation and particles emitted by it, and send beams of radiation or particles into the plasma to probe its properties.

Once ITER functions technically, understanding and optimising its performance will take time. During the final phase of the project, ITER must deliver an average neutron fluence of 0.3MW-year/m$^2$ for materials and blanket studies to refine the technology which will lead to the next step, a demonstration reactor.

R&D WORK FOR ITER

Given these new challenges, the ITER project team identified a number of technical challenges which should be validated by industry, requiring either unique specifications, unique precision, unique performance or often all at the same time. Seven long-term large R&D projects were launched between the founding partners of the ITER project and jointly financed to a total of approximately 660MS$. These seven R&D projects have now successfully reached fruition.

The challenges which these R&D projects had to meet are simplified in the following list:

- Central Solenoid model coil: Performance, construction
- Toroidal Field model coil: Performance, construction
- Full size vessel sectors: Scale, tolerance
- Full size blanket modules: Construction, tolerance
- Full size divertor module: Heat flux, construction
- Full size remote maintenance for blanket: Precision, leverage, speed
- Full size remote maintenance for divertor: Precision, mass, speed
THE PRESENT STATUS

At the time of writing, the ITER project is essentially ready for construction with detailed designs for industrial procurement of lead items. Costing has been confirmed by third parties. Japan, Canada, France and Spain have offered sites. Detailed financial agreements and the creation of a legal entity depend on the final site.

Once a site has been agreed, the construction should take around 10 years, including the establishment of the organisation which will own and construct ITER, as well as licensing, leading us to operation in 2014. The ITER project will naturally have a considerable effect on the fusion programme plans of all the participants.

ITER DATA HANDLING

As any high-technology research enterprise, ITER will have diverse data handling requirements. Some are conventional, within present experience and certainly within the technology which will be available during operation, such as:

- Data Source devices - diagnostic hardware
- Data Acquisition Front-ends - ADC’s, counters
- Data to generate a pulse scenario
- Data for Real-time Feedback Control methods
- Data for Slow Controls

Structuring the information and making it available to researchers worldwide with an adequate bandwidth is not treated in this paper, but is already implemented for LHD [4] and other magnetic fusion devices [6]. Other data handling requirements appear at first sight to be less easy to satisfy and are debated in this paper. These include:

- Data Transfer from the sources
- Data Filtering
- Data Archiving
- Data Validation
- Data Analysis
- Data Purpose

DATA SOURCE RATE AND VOLUME

How should we estimate the likely ITER data source rate and data volume? Three methods have been invoked in the past, which we summarise.

Extrapolation from existing devices

This approach was proposed [5] by noting the annual doubling of the acquired data on JET. Extrapolating this during the ITER construction period leads to 50-160GB/pulse and 180-600TB/year, considerably less than the data archiving for LHC [5]. This approach assumes that ITER’s data requirements are driven uniquely by the intervening timescale, which is questionable.

Physics reasoning

We consider that ITER has a pulse length typically 50 times JET. The physics phenomena occur at frequencies 1.0-0.1 times JET (100G), suggesting 250GB/pulse with similar diagnostics to exploit a similar project.

Number of channels and acquisition rates

We take a 2003 delivery product used in tokamak research. This offers 100 channels, 10MSample/s generating 2GB/s during a 1000 second pulse with a 9GB cyclic buffer. 50 front-end boxes like this should satisfy the proposed requirements for ITER. However, they create a potential source rate of 100TB/pulse for a total box cost of only around 1MS. Investment cost will therefore not naturally limit the data rate or data volume as in the past.

The most sobering remark is that the maximum data quantity predicted for JET at the time of construction was 0.1-1.0MSamples per pulse, forcing us to question any prediction whatsoever. The only approach at this early stage is therefore to design for the maximum capacity and to reduce the real capacity used as far as possible. Overcapacity will not be particularly penalising financially.

We therefore have taken the position that a huge data source rate must be envisaged for very long pulses and that this source rate will have to be dramatically reduced for the data to become useable. It is clear that the bottleneck is not the data storage but the data access for local and remote analysis.

SCHEMATIC DATA REDUCTION

In this section, we present an illustrative sketch of data reduction using available technology. This is not a design, but simply serves to show that a straightforward solution exists. Since the typical front-end devices can source data at 100GB/s, enormous data reduction is essential as the experimental information percolates to the level at which it can be used to operate and understand ITER. We consider a 3 layer structure, similar to existing architectures, in which front-ends communicate with a sub-system and sub-systems communicate with a top-level, Fig. 2.

Our typical front-end (Level –2) has a source rate potential of 2GB/s, but is assumed to have an average over the pulse of 50MB/s. This data source can be fibrechannel streamed with loss-free compression to two archiving targets, one for nuclear operation and the other for research, since access authority will inevitably be very different. This rate and was already achieved on the LHD device with existing technology [4].

The front-ends have to reduce the data by a factor 100 to deliver 20MB/s peak up to Level –1 for which conventional communication is adequate. Units which under-sample at 50kHz already satisfy this Level -2 data reduction target, even before physics processing.

Level -1 receives the reduced data stream from typically 7 front-ends. It streams compressed data for archiving in the same way. It has to reduce the incoming data by a factor of 100 to transmit 1-3MB/s peak up to Level 0. This time, it can correlate the physics information from 7 data sources while performing the reduction, using information circulating at Level 0 as well.

Proceedings of ICALEPCS2003, Gyeongju, Korea

591
Level 0 is where the operation of ITER takes place. It has an inflow of 10-20MB/s and distributes all the information to operate the plant as well as making it available to Levels –1 and –2 for data reduction. We are therefore assuming that 10-20MB/s of reduced data into Level 0 should be adequate for full control and surveillance of ITER using specific intelligent processing.

Figure 2: Illustrative schematic of the data flow reduction, compatible with today’s technology. Data are streamed to storage at all levels. Storage of 800GB per pulse for 50 front-ends would be feasible.

There is a physics requirement for this reduction to be acceptable at any level, namely that it only requires Level 0 data to be circulated. Analysis will not require raw data from other sources. Diagnostics with more than 100 channels are compatible, so long as the average data rate can be streamed to storage.

Will ITER be a unique step forwards? ITER will be the first experimental reactor with full nuclear licensing and will therefore have new constraints. However LHD [4] already operates 150 second pulses, 4 minutes pulses are performed at Tore Supra [6]. W7-X will operate long pulses before ITER. The ITER technology test beds will also offer an environment for testing data handling concepts for ITER.

The conclusion of this section is that the range of estimates of the ITER data flow and the data volume can be technically handled with today’s conventional technology [7]. LHC will source at 40TB/s by way of comparison. The raw data can be archived at all levels, although this will be expensive rather than difficult. The 3 major issues to be debated therefore appear to be “why archive all that data”, “how can you use all that data” and “how do you achieve the required level of data reduction”.

WHY ARCHIVE THE RAW DATA?
Having suggested that there exist and will exist solutions for sinking and archiving all of the considerable front-end data flow, we should ask the question why we should wish to do this. Typical answers to this question are:

- Because each pulse is expensive compared with the cost of archiving – true
- Because if you make irreversible compression, the data can never be recovered – true
- Because even if you do not see any systematic trends in the data while it is flowing in real time, you could analyse your data in detail later – true
- Because you do not have to select the data to reject.

These arguments do not necessarily justify the cost and effort of archiving at the rate of 800GB/pulse. It is possible that a portion of the raw data will have to be archived for legal reasons, in case of unforeseen events – satisfying such a requirement would not be open to debate. Occasional burst recording of everything would be wise.

Once we abandon archiving all of the raw data, we must reject some of the data. This should be based on identifying which part of the archived data might ever be retrieved. Ideally, only that part should have been archived.

The large quantity of raw data and the limited number of talented people to interrogate raw data from a particular front-end will mean that only a small fraction of the data will ever be retrieved. Leaving all of the raw data in an archival system will in fact defeat the purpose of archiving, namely to make usable data available. How can we develop the criteria for keeping only a very small fraction of the information? How do we create a filter which does not irreversibly delete potentially useful information?

A KNOWLEDGE MODEL FILTER
In order to address these questions, we step back a long way and we ask why we are performing the ITER pulse in the first place. The answer is: a) to learn more about the device and to test our understanding or b) to deliver a pulse of known performance. Initial exploitation and physics research falls under category a) and accumulating the neutron fluence falls under b).

The aim of learning something from a particular pulse suggests a filtering approach which rejects all information from which we will learn little or nothing. This information is equal to all observations which correspond to our existing understanding of the device, our preconceptions, or more formally our model of ITER. If any part of the acquired or analysed information is equal to our model of ITER, then it has little or no additional learning content and can safely be rejected. As ITER evolves, and as our model of ITER becomes more complete and more accurate, then the fraction of data falling into the “no new information” trap will naturally increase. We would have to retain a small fraction for comparison.

Does this sound science fiction and impossible to quantify? In fact, this model represents what the experienced analyst does with his mountain of data. He scans through the data, normally visually (if he has the time) and identifies periods during which the behaviour of his information appears to be different from his preconceptions – he then explores this more fully, trying
to understand it. Once he understands the phenomenon, his model has become more precise and the next time he sees this it is regarded as “understood”. What we are considering is to formalise this to cover all of the data, not simply the small fraction of the data an experimentalist has time to consider.

Is this simply an “event trigger”? The answer is negative, since we would require unreasonably complete knowledge to construct a description of all unpredicted events. If this knowledge were badly expressed, then we might miss significant data from which we could learn something new. We are rather constructing a “non-event” filter to identify data which correspond to a definitely unneeded category, which are then rejected.

AN EXAMPLE

Generating such a knowledge filter sounds rather abstract. We have examined the idea for several diagnostic sources and find that it should be feasible. We explain a single example of Thomson Scattering data reduction to illustrate this. This diagnostic exists on all tokamaks and is a candidate for ITER.

A laser beam is sent through the plasma and the light scatters off electrons. The thermal velocity of these electrons causes a Doppler shift of the scattered photons. Analysing the spectrum of scattered light therefore gives us information on the electron velocity distribution, from which the electron temperature can be extracted on the basis of a physics model.

The raw data on scattered photons are calibrated and fitted to a physics model to provide an electron temperature and some residuals. The knowledge model has to contain a range of “plausible” temperatures and distributions of the residuals. If the residuals are within the error distribution and if the deviations are not biased, then the experimentalist will be satisfied and the fitted temperature can be validated and the raw data classed as modelled. The residuals will not teach us anything new and our rejection filter is behaving like our expert. Insight into such an approach can be found in [8].

The electron temperatures are then mapped onto a reconstructed plasma equilibrium, obtained from Level 0 after reduction from a different Level –1 sub-system, and the spatial distribution is regressed as a function of the magnetic flux coordinate. Our knowledge model is now a magnetic flux coordinate. Our knowledge model is now a

Developing the knowledge model filters is specific to each data source, to embody the understanding of the experimentalist. In today’s experiments, the data flow is already too high for regular human appreciation and only a small fraction is analysed in depth. We suggest that this approach could usefully be developed on our present-day experiments for optimising their use, as well as learning how to implement such a procedure as the proposed “knowledge model filter” for ITER data reduction.

CONCLUSION

The ITER project status has been summarised. Although the technology available for all aspects of data handling will have evolved significantly when ITER operates, we can already sense a solution using existing technology. We consider that ITER data handling, although requiring a substantial effort, should not present any novel technical challenges beyond those already being met by the participants at this conference.

The major challenge will be in reducing the predicted data flow to its useful minimum, developing methods to assimilate this information by a research team. We propose the concept of “learnable information” for keeping or discarding raw or analysed data. Developing this concept during the next few years on existing experiments could lead to a data reduction approach to speed up the learning necessary to exploit the ITER device rapidly.

ACKNOWLEDGEMENTS

It is a pleasure to acknowledge discussions with Peter Milne of D-TACQ on the technical advances in high volume data acquisition. The work was partly supported by the Fonds National Suisse de la Recherche Scientifique.

REFERENCES