our future through science

An Inclusive Vision for High Performance Computing at the CSIR

A GAZENDAM[1], P BOSSCHA, S DONOVAN, J HAARHOFF, U HORN, C MAHLASE, K MATTHEE[2], T NAIDOO, G SNEDDEN, H VAN NIEKERK, D VOGT AND J EKSTEEN[3]
[1-3]Meraka Institute, PO Box 395, Pretoria 0001
[1]Email: agazendam@csir.co.za

INTRODUCTION

Scientists doing research in fields ranging from genomics to astronomy face various challenges, including:

- Acquisition and processing of extremely large and valuable collections of primary data;
- The need for large *in-silico* or virtual experiments;
- The need to work and visualise results from the desk; and
- The globalisation of scientific collaborations.

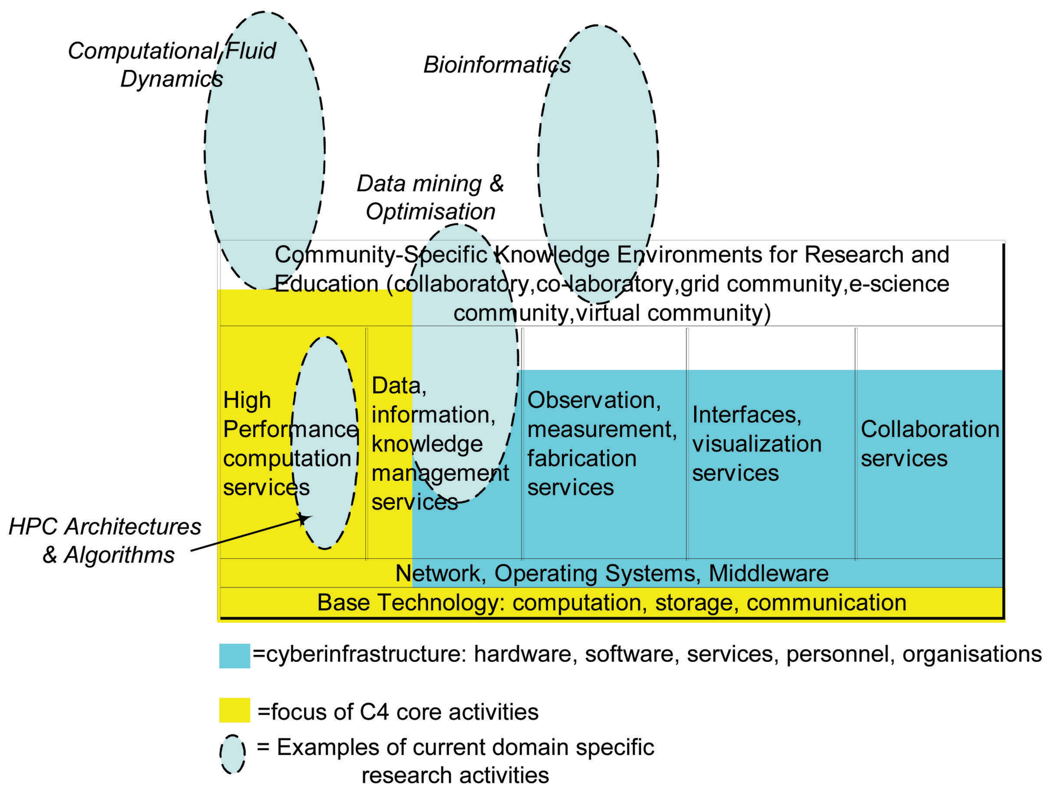

Figure 1: Cyberinfrastructure diagram indicating role of C4 and potential applications.

The use of digital technology to enhance existing research and enable new, previously unimaginable research fields is gaining support in South Africa. A reason for this relatively slow adoption, compared to that of other research communities, is the lack of appropriate scientific computing infrastructure. This includes dedicated instrumentation and equipment, commodity computing devices and communication networks, as well as the software frameworks that run on top of these. **Figure 1** shows how these components relate to the CSIR's high performance computing infrastructure project, C4, as well as examples of potential applications.

DESIGN CONSIDERATIONS

A myriad technologies and standards attempt to dictate the configuration of scientific computing infrastructure. In order to guide decision making, the following value system, depicted in **Figure 2**, was defined:

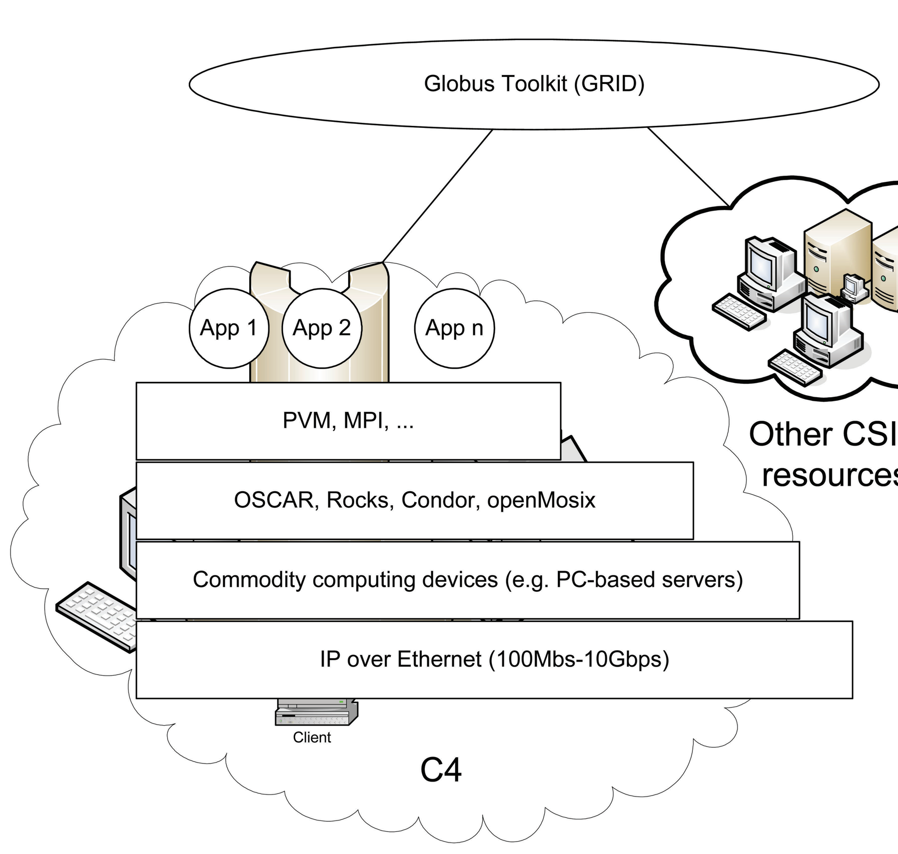

Figure 2: HPC technologies and standards within the CSIR context.

- The use of free and open source software saves money, builds skills, and contributes to the scientific community;
- Open standards and interfaces allow integration between technologies;
- A modular and scalable architecture allows growth and integration into larger national and international initiatives in the future; and
- Existing hardware and software platforms require an architecture capable of integrating heterogeneous resources.

PREPARING FOR THE CSIR CLUSTER COMPUTING CENTRE

The CSIR is currently investing in high-performance computing infrastructure called C4, see **Figure 4**. To prepare for its coming, the Mining competency area invested in a cluster, the DT-12 from Orion Multisystems – shown in **Figure 3**. While the individual nodes are modest, with 1.2 GHz x86 processors, each with 1 GB of RAM, there are twelve processors in a single box, at a cost roughly equal to that of four state-of-the-art workstations.

Figure 3: The Orion Multisystems DS-12.

THE CSIR CLUSTER COMPUTING CENTRE (C4)

C4 is a high performance computing resource of modest scale accessible to all researchers within the CSIR. Figure 3 shows a diagrammatic layout of the C4 system as it was presented to the vendor community for proposals. It is envisaged that C4 will be operational by the middle of 2006.

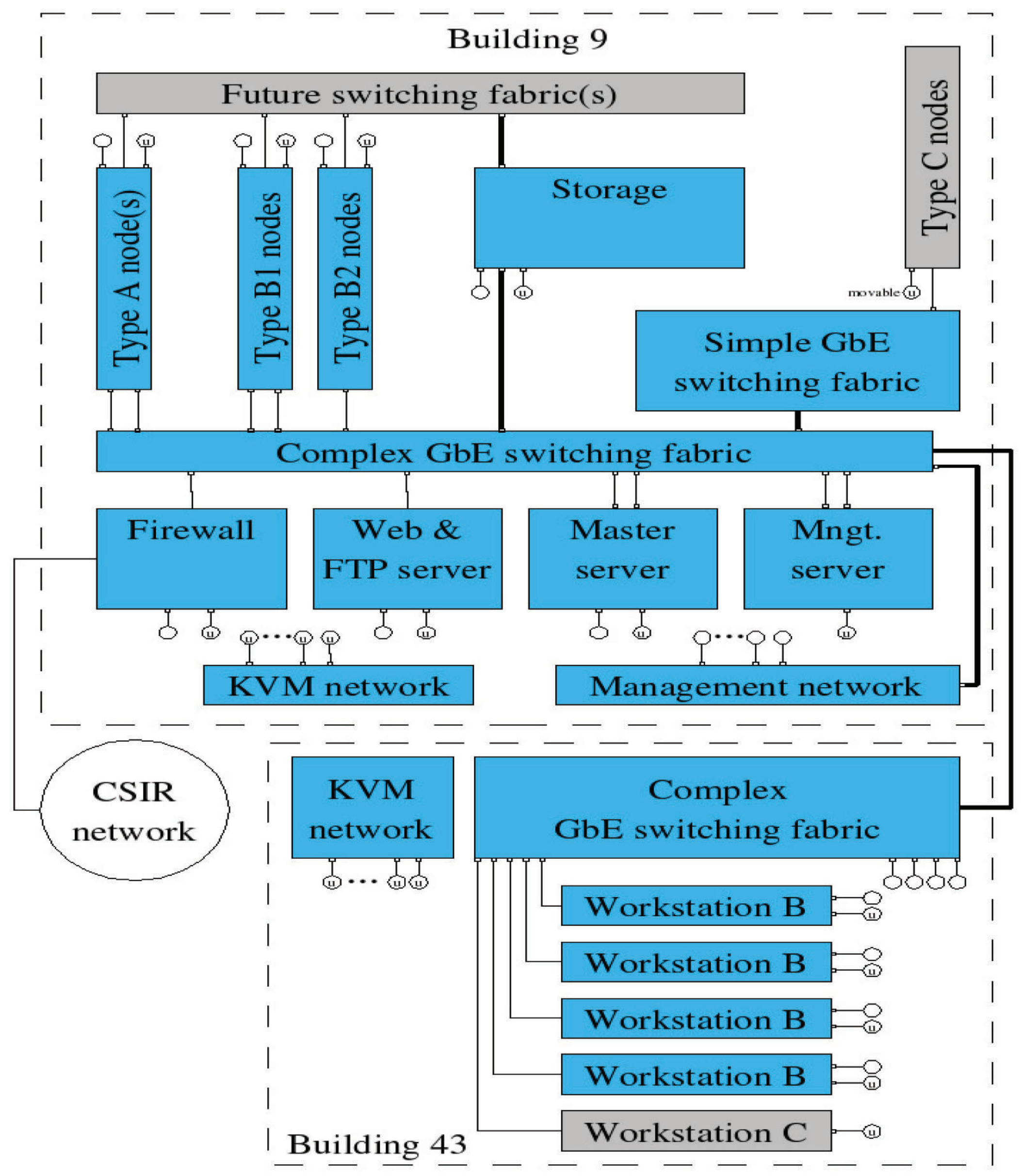

Figure 4: Diagram of the C4 system layout.

Unique factors of C4:

- Most powerful computing resource in the CSIR (~500 gigaflops);
- Probably one of the 10 most powerful computing resources in South Africa;
- Extremely flexible architecture capable of dynamic reconfiguration; and
- Designed around open source software and COTS hardware.

Technical aspects of C4:

- At least 32 dedicated high-end servers configured as cluster nodes;
- Mid-sized shared memory computing resource;
- Approximately 128 standard desktop PCs configured as additional cluster nodes;
- Gigabit Ethernet networking; and
- Several terabytes of storage space.

RELEVANCE TO R&D STRATEGY

The primary benefits of high performance computing and related applications lie in the enabling of new frontiers of research. Participation in post-genomic biology research, for example, has the potential to discover new medicines for enhanced quality of life, while discoveries could stimulate the creation of new industries with resultant economic benefits. Through supporting unprecedented collaboration between knowledge workers, bigger problems can be solved faster, thereby increasing the competitiveness of South African industry. The open nature of the suggested approach to high performance computing decreases dependence on expensive imported hardware and software, making South Africa more competitive with greater control over its destiny.

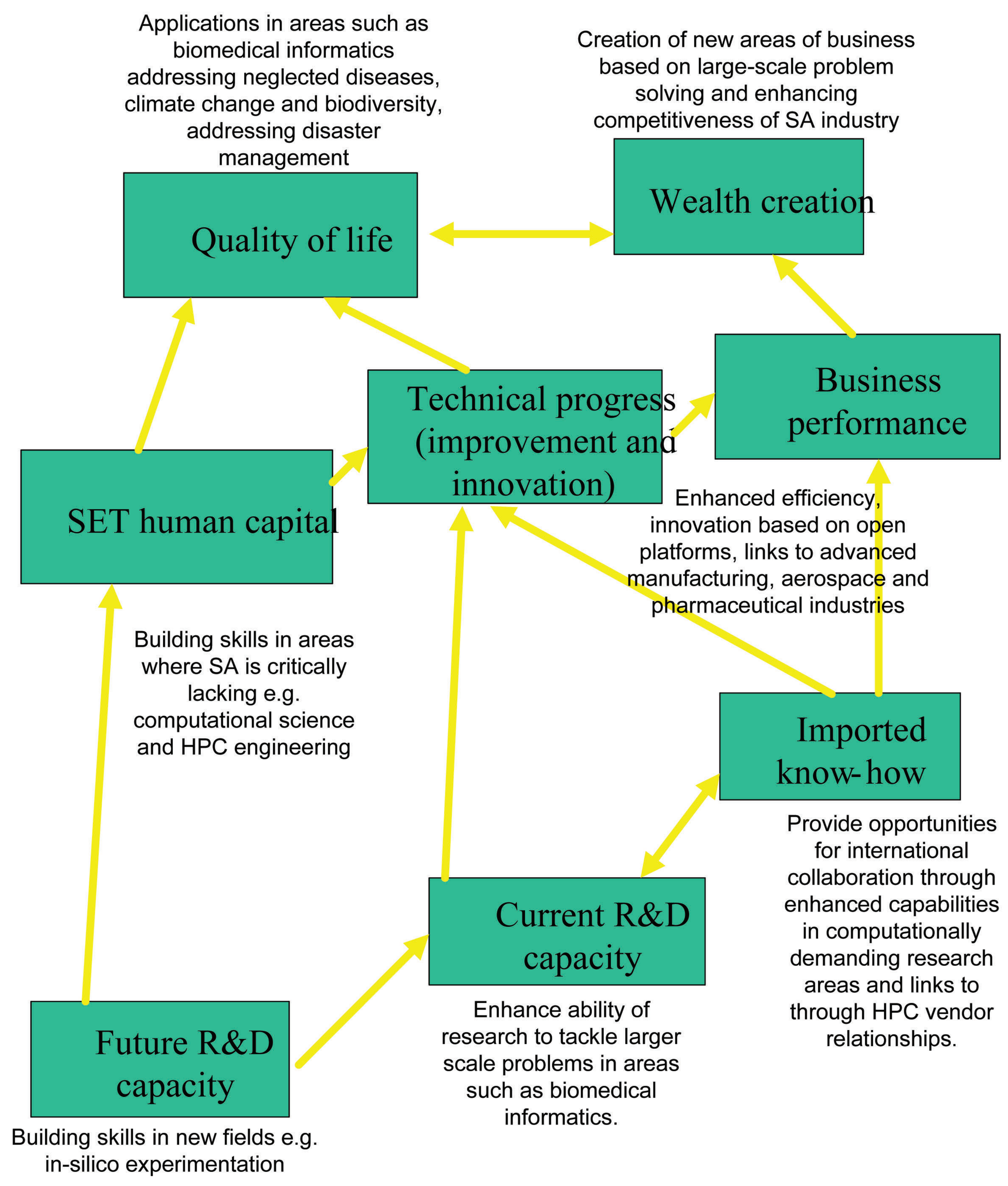

Figure 5: Diagram showing role of HPC in the national strategic context.

CONCLUSION

The Department of Science and Technology (DST) has identified high performance computing as one of its focus technology areas and is financing the establishment of the Centre for High Performance Computing (CHPC). Linkages to the Square Kilometre Array (SKA) and South African National Research and Education Network (SANREN) initiatives confirm the importance of CHPC to the national science infrastructure agenda.

meraka INSTITUTE
African Advanced Institute for Information & Communications Technology