Australian Particle Physics Roadmap
Working Group 6 (Computing) contribution

December 2019 – April 2020

Contents

1 Australian HEP computing usage

1.1 ATLAS

The ATLAS experiment makes extensive use of computing grid technology. ATLAS employs the Panda Workload management system and the Rucio Data Distribution Model. ATLAS were one of the founders of the World Large Hadron collider Computing Grid (WLCG) and the LHCONE network collaboration. LHCONE is a virtual private network which provides guaranteed bandwidth at non-commercial rates. We are connected to LCHONE via AARNET which has provided outstanding service for our needs. The performance of the WLCG is recorded by the WLCG monitoring portal. The WLCG routinely provides 50 GB of data per second for 24 hours × 7 days per week. Over one year the WLCG processes over 1500 PB of data securely over a heterogeneous collection of computing clusters located around the world. ATLAS assumes a model of constant funding for computing into the future.

All participants in the ATLAS experiment are expected to provide their fair-share of resources to the computing grid. The Australian group has signed a Memorandum of Understanding pledging to supply our contribution on a best-effort basis. Since 2011, this was supplied by the ARC Centre of Excellence for Particle Physics, (CoEPP), which typically invested $500k per year in people and equipment into our Tier-2 grid site in Melbourne and Tier-3 sites in Sydney, Melbourne and Adelaide. In Melbourne we have a Tier-2 site which consists of a Compute Element comprising around 1500 CPU cores and a Storage Element of 1.2 PB.

We have small physical clusters in Sydney, Melbourne and Adelaide, and around 200 TB of storage, mostly built using out-of-warranty hardware and employed as a CEPH file system. ATLAS also makes use of the particle community's 700-core Tier-3 cluster based on a NeCTAR cloud allocation in Melbourne.

1.2 Belle II

The Belle II experiment makes extensive use of computing grid technology. Like LHCb, we employ the DIRAC interware as our primary grid computing grid tool. We are members of the WLCG, the LHCONE network collaboration, and report our usage statistics to the WLCG monitoring portal. Like ATLAS we are moving to employ Rucio as our Data Distribution Model. Belle II, through its Computing Steering Group, regularly forecasts its computing and storage needs. All countries are required to contribute in proportion to the number of Ph.D. physicists (academics and postdocs) working on the Belle II experiment. The Australian group has signed a Memorandum of Understanding pledging to supply our contribution to the Belle II experiment grid computing on a best-effort basis. We have a Belle II Compute Element comprising around 700 CPU cores (sourced via a NeCTAR cloud allocation) and Storage Elements at Melbourne (50 TB) and Adelaide (150 TB).

1.3 Theory

The Australian HEP theory community makes extensive use of computational methods. These range from analytical studies with a heavy reliance on symbolic packages such as Mathematica, to small and medium scale numerical calculations able to be carried out on local hardware or in batch mode on cloud hardware, to full-scale tightly-coupled parallel calculations requiring supercomputing facilities. The two largest users of compute time amongst Australian theorists are the GAMBIT (phenomenology) and lattice QCD communities; these will be discussed in under specific subheadings below.

Smaller-scale computing resources support a wide range of other activities. Collider phenomenology is an active area at nearly all institutes, with local institutional clusters regularly employed for the simulation of LHC collisions. The Monash group is a key contributor to one of the leading simulation suites (Pythia; `http://home.thep.lu.se/~torbjorn/Pythia.html`). Dark matter phenomenology is also pursued extensively, with numerical calculations of dark matter production in the Early Universe, direct and indirect signals carried out across the community, using both desktop and local cluster hardware. Groups at the Universities of Queensland and Melbourne also investigate the impacts of dark matter on stellar physics, sometimes requiring explicit stellar simulations. Institutional-level hardware also provides for the numerical requirements of activities in vacuum stability, phase transitions and other Early Universe physics (widespread), neutrino physics (predominantly UNSW and the University of Melbourne), cosmology (predominantly UNSW), gravitational waves (Monash), and flavour physics (widespread). All these activities, as well as investigations of theories of modified gravity (University of Sydney), are supported in their preparatory phases by extensive use of symbolic computation.

In many cases, theories for new physics predict observable signatures only in specific parts of their parameter spaces, with different signatures often appearing for different parameter combinations. Numerical exploration of the parameter spaces of such theories is often needed in these cases. Most of the phenomenology groups in Australia perform such studies to some extent; these range from basic random or grid scans done on local or cloud infrastructure, to more sophisticated one-off explorations using specific search algorithms on local clusters, to the orchestrated campaigns of the large ‘global fitting’ collaborations GAMBIT (Universities of Queensland and Adelaide, Monash University) and MasterCode (University of Melbourne).

GAMBIT

The GAMBIT (Global and Modular Beyond-the-Standard-Model Inference Tool) Community is a group of approximately 60 theorists and experimentalists drawn from across particle and astroparticle physics, and 16 different countries. GAMBIT includes a software suite designed for carrying out global analyses of new physics models, by performing large-scale parameter sampling and calculation of combined likelihoods, accounting for all relevant experimental results and theoretical constraints. The job of the GAMBIT Community is to maintain and develop this software, and to make scientific use of it by analysing and comparing models of new physics.

The software is designed to run on essentially any hardware running OSX or Linux. It is parallelised at thread level with OpenMP, and at process level with MPI. It makes use of advanced statistical sampling algorithms such as differential evolution and nested sampling, which benefit from tight coupling across processes. Compute-intensive GAMBIT runs (e.g. those involving direct Monte Carlo simulation of LHC analyses) therefore typically require supercomputing rather than cloud infrastructure. Post-processing of samples obtained in tight-coupling configuration can however be done in a more loosely-coupled environment.

GAMBIT presently makes use of approximately 50–60M supercomputing core hours per year, running predominantly on the Marconi-KNL and Joliot-Cure Rome systems through the European PRACE Tier 0 initiative. It has also in the past made use of Tier 1 European access to MareNostrum (Spain), Cartesius (Netherlands) and Prometheus (Poland). Australian institutional clusters are sometimes used for code development work. Australian resources have to date played only a minimal role in production runs, due to the relatively small allocations available, even from NCI. This may change in the future, however.

GAMBIT analyses typically generate between 500 MB and 500 TB of data for a single paper. These mostly consist of parameter samples, corresponding observables and saved likelihood components. Speed of disk access is generally only a bottleneck when one comes to plot the final results from a GAMBIT run.

The GAMBIT code is publicly available via the open source 3-clause BSD licence (`https://gambit.hepforge.org`). All samples, benchmarks and plotting scripts arising from GAMBIT publications are made freely downloadable – and archived – at Zenodo (`https://zenodo.org/communities/gambit-official`).

Lattice

There is substantial evidence that the strong force of nature is described by a quantum field theory comprised of quarks and gluons — quantum chromodynamics (QCD). The most significant precision tests have been performed in high-energy collisions, many of which are discussed in this document, where established methods in quantum field theory are accurate. At low energies, however, the theory of QCD remains an unsolved mathematical problem. The best established theoretical tool to study the low-energy properties of quantum chromodynamics is through the numerical formulation of lattice-regularised QCD (lattice QCD).

Lattice QCD is a technique that discretises space-time into a four-dimensional grid, enabling numerical simulations to be carried out on large-scale computers to solve the equations of the fundamental quantum field theory. This method is recognised by the international community as essential in our theoretical understanding of the structure and interactions of hadrons from QCD. Building upon the seminal work of Ken Wilson in the 1970's, the field has now matured to the stage where rigorous comparison between experiment and theory can be made. In some instances, the results of lattice QCD are now used in conjuction with experiment to constrain fundamental interactions and thereby provide precision tests of fundamental symmetries in nature.

The maturity of lattice QCD is such that modern computations are able to efficiently evaluate integrals in spaces exceeding a billion (10^9) dimensions. These "computer experiments" are world renowned for their huge demand for supercomputer cycles, and have hence played a major role in driving advances in supercomputing for more than two decades. As a result, in the field of lattice QCD, the ability to write successful scientific proposals for large amounts of computer time on the world's fastest supercomputers, is just as important as successfully applying for grants. This is also true in Australia where the major source of computer time is the National Computational Infrastructure (NCI) National Facility in Canberra. Australian lattice QCD researchers have a strong track record of large grants on NCI and other Australian supercomputing resources through the National Computing Merit Allocation Scheme (NCMAS). These resources, however, are modest when compared to supercomputing resources available to the rest of the lattice community through international computing centres. In order to compete on the international stage, Australian researchers often access larger supercomputing allocations by collaborating with international collaborations such as QCDSF, which currently holds large allocations of computing time on the HRLN and JUWELS supercomputers in Germany, and the DiRAC supercomputer in the UK.

1.4 Medical physics

The Centre for Medical Radiation Physics (CMRP) has been a continuing research strength of the University of Wollongong for the last decade. It is the largest education and research multidisciplinary medical radiation physics institution in the Asia Pacific, with more than 70 HDR research students and a strong academic and research team. In 2015 and 2018, ERA evaluated CMRP with a score of 5 (above World class) under the 0299 FoR code. CMRP is an international leader in medical physics semiconductor radiation dosimetry/detector instrumentation, reflected in many patents, licensing and consultancies for industry. National and international recognition of CMRP in microdosimetry is reflected in the fact that it founded the biannual Mini-Micro-Nano-Dosimetry (MMND) and Innovation Technologies in Radiation Oncology (ITRO) international conference in 2000, and has hosted it ever since.

International leadership of CMRP in Monte Carlo simulations for medical physics is reflected in the international roles of A/Prof S. Guatelli as member of the Geant4 Steering Board (`http://www.geant4.org`), Coordinator of the Geant4 Advanced Examples (`https://geant4.web.cern.ch/collaboration/working_groups/advanced_examples`) and Coordinator of the G4-Med effort (`https://twiki.cern.ch/twiki/bin/view/Geant4/G4MSBG`), which counts more than 50 members in Australia, Japan, Europe and the US. The national leadership of CMRP in this field in reflected in the foundation of the CMRP Geant4 International Workshop and Schools since 2011.

The CMRP uses a mixture of local of national computing infrastructures for the following projects, spanning from development of physics models in the Monte Carlo code Geant4 to the application of Geant4 in medical physics applications:

- Development of AI-based treatment planning for radiotherapy.
- Regression testing of Geant4 on the CERN computing facility for medical physics application within the G4-Med effort.
- Development of a new data library in Geant4 for Proton Induced X-ray Emission, in collaborating with ANSTO, Lucas Heights.
- Development of new physics models in the Geant4-DNA Package to describe electron interactions in gold nanoparticles (of interest for nanomedicine).
- Modelling of the interaction of radiation with matter from macro- to nano- scales. This includes modelling of detector responses for dosimetry/microdosimetry and quality assurance (QA) in radiation therapy, modelling of the interaction of radiation with matter at the level of cells (microdosimetry) and DNA (nanodosimetry), and evaluation of treatment planning systems based on the radiobiological effectiveness (RBE) of different radiation treatments.
- Use of Geant4 to support the development of QA instrumentation for conventional X-ray radiotherapy, microbeam radiation therapy, proton and heavy ion therapy.
- Accurate study of the fragmentation process in order to determine more accurately the RBE in heavy ion therapy, which is then input to the treatment planning system.
- Modelling of medical accelerators for particle therapy (proton therapy, heavy ion therapy and neutron beamlines for boron neutron capture therapy).
- Monte Carlo simulations for radiation protection studies (aviation and space missions).

- Monte Carlo simulations to study shielding solutions in particle therapy facilities.
- Study of enhancement of radiotherapy outcome by means of nanoparticles, in conventional X-ray radiotherapy, microbeam radiation therapy and particle therapy.
- Development of in-vivo imaging solutions for proton and heavy ion therapy, in collaboration with ANSTO, Lucas Heights.

1.5 Accelerator physics

Computing requirements for accelerator physics development at the Australian Synchroton are rather modest. These are are mainly to do with simulations for design and particle-tracking studies. These are not particularly difficult requirements to meet, but do require cluster-level computing. Intensive design studies make use of the local on-site Synchrotron cluster ASCI, which is nominally used mostly for computationally-intensive beamline techniques, to do image reconstruction and analysis.

2 Current facilities

2.1 High-Performance Computing (HPC)

HPC facilities are available nationally from the National Compute Infrastructure (NCI; `https://nci.org.au/`) and the Pawsey Supercomputing Centre (`https://pawsey.org.au`). NCI hosts Gadi, a 144 000-core, 3000-node supercomputer based on the Intel Cascade Lake architecture. Phase 1 of Gadi (approximately half its final configuration) was placed number 47 in the Top500 list of the world's most powerful supercomputers (`https://www.top500.org/list/2019/11`). Pawsey hosts Magnus, a mid-size facility (1488 nodes, 35 712 cores) based on the older Intel Haswell architecture, along with a number of smaller peripheral clusters.

Particle physics groups receive time on both NCI and Pawsey facilities through the competitive National Computational Merit Allocation Scheme (NCMAS), as well as via partner shares held by their respective institutions. Together, these typically offer a maximum of approximately 15M core hours per group per annum. The Wollongong group will receive approximately 1 MSU[1] on Gadi via NCMAS in 2020, and 0.4–1 MSU via the partner program. The Adelaide lattice group have been heavy users of the NCI facilities for nearly 2 decades. In the recent 2020 NCMAS round they received 10 MSU on Gadi and 5 MSU on Pawsey. Through the partner program, they will receive a further 21 MSU on Gadi in 2020. **Still needs info on any time awarded to others via NCMAS or NCI/Pawsey partner share.**

---

[1] MSU = million service units, with 2 SU corresponding to a single core hour on Gadi.

Working together with international collaborators, Australian particle physics groups also been successful in gaining access to a significantly larger number of core hours on international HPC facilities through participation in foreign competitive merit allocation schemes (e.g. PRACE: https://prace-ri.eu/). These include MareNostrum (Spain), Marconi (Italy), Joliot-Curie (France), Prometheus (Poland), HRLN (Germany), JUWELS (Germany) and DiRAC (UK).

2.2 CERN Computing Infrastructure

CERN offers a range of essential computing infrastructure to member experiments, projects and states, as well as a number of community services. These include LXPLUS, a large cloud facility for interactive jobs, the CERN HTCondor Batch Service, a 235 000-core grid facility serving the full range of CERN users and providing the main underpinning of the WLCG, continuous integration facilities for experimental analysis frameworks and other crucial software projects such as Geant4, and data serving facilities such as Zenodo and the CERN Open Data project. These facilities are essential for specialised tasks such as regression testing of Geant4 for medical physics applications, and serving large datasets to the community from GAMBIT studies. LXPLUS and HTCondor are also widely used by Australian experimentalists and theorists, both through the WLCG and as an explicit supplement to their institutional hardware.

2.3 ATLAS Tier 2

The University of Melbourne hosts an ATLAS Tier-2 site, which consists of a Compute Element comprising around 1500 CPU cores and a Storage Element of 1.2 PB. One of the features of grid computing is our use of non-POSIX network file systems. This enables us to build clusters of thousands of CPUs to analyze hundred-petascale scale data sets at around 10% of the cost of a commercial file system. For example, storage space on the new University of Melbourne Spectrum-Scale file system costs ten times as much as a Tier-2 Storage Element based on the Disk Pool Manager (DPM) file system.

2.4 Australian Research Data Commons (ARDC)

The Australian physics community has an allocation 1400 CPU cores in the NeCTAR (National eResearch Collaboration Tools and Resources) OpenStack cloud. Of these, nominally 700 are dedicated to the Belle II Tier-2 facility. The remaining 700 cores are allocated as either large memory 16-core interactive servers (4 of these), various other useful servers (a DIRAC development server, test-airflow server and others), or as a general-purpose Tier-3 cluster for the particle physics community. The Tier-3 cluster has been widely used by the particle physics community. It has seen extensive use by Belle and Belle II physicists, for ATLAS analyses, for simulations of the SUPL laboratory, the

SABRE experiment and for numerous theory projects. Access was provided via CoEPP with very liberal policies. Essentially anyone with an interest in particle physics was granted full rights to use the facility. Our experience as been that only gentle pressure has been required to prevent abuse of the resources.

These cloud-based services were built with middleware developed by the NeCTAR RT-07 eResearch Tools project and maintained by the CoEPP Research Computing professionals. Of these only Sean Crosby is still in the University system. He has moved on to to a senior role the University of Melbourne Research Computing services.

2.5 University resources

The University of Queensland

The University of Queensland hosts two major supercomputers: Tinaroo is a 244-node SGI HPC system featuring approximately 6 000 CPU cores, 30 TB of main memory, and a fast interconnect. It is complemented by FlashLite, a 68-node machine with 34 TB of main memory and over 500 TB of solid-state storage, optimised for memory-intensive computing. These are further supported by Awoonga, a more traditional cluster without fast interconnect, and Wiener, a dedicated imaging cluster. The University of Queensland is also host to QRIScloud, the 4000-core Queensland node of the national Research Data Storage Infrastructure (RDSI) and cloud computing service NeCTAR.

Adelaide University

The University of Adelaide hosts the Phoenix supercomputer, a Lenovo NeXtScale system consisting of 260 nodes, capable of 450 teraflops of performance. Along with an addition GPU cluster named Volta, there are 8892 CPU cores and 352 GPU cores with 56 TB of memory. The University of Adelaide is currently engaged in preparing a technology roadmap that will expand this in future.

The University of Melbourne

The University of Melbourne is 2 years into a 5-year "Petascale Campus Initiative" (PCI) designed to make a substantial uplift of its research computing infrastructure. The total funding envelope is $20 million per year to be invested in both hardware and people. Particle physics has provided substantial input and leadership into the project. The base storage technology will be an IBM Spectrum-scale network file system with at least 5 PB of storage. This will support a general-purpose CPU cluster of at least 10 000 cores and another GPU cluster equivalent to hundreds of NVIDIA P1000 GPU's. There will in addition be provision made for direct capture of data from high-bandwidth instruments, of which next-generation genome-sequencers are the prime example.

In addition to this hardware, the University plans to employ 10–20 FTE of academic specialists in the fields of research computing to provide general uplift to the University workforce.

Our challenge will be to ensure that particle physics makes effective use of this combination of hardware and people.

In addition the School of Physics hosts a substantial legacy of computing infrastructure developed by CoEPP. This is no longer maintained and will become increasingly unreliable over the next year.

Monash University

Monash University is part of MASSIVE, which is a collaboration between Monash University, CSIRO, ANSTO and The University of Wollongong. It hosts the M3 facility, which consists of 4 000 cores, 168 NVIDIA coprocessors, and a 3 PB file system. Access works through a fair-share protocol using the Slurm batch system. Applications can be made for enhanced access to the system if resources are provided for further investment.

The University of Sydney

The University of Sydney hosts an HPC cluster called Artemis, which is provided to support research at the University, and is a partner in the National Computational Infrastructure (NCI). Artemis consists of 4264 cores of compute capacity consisting of 136 standard memory (128 GB) compute nodes, two high memory (512 GB) nodes, three very high memory (6 TB) nodes and five GPU compute nodes. The School of Physics also hosts infrastructure that has been purchased by research groups within the School. This is housed locally in a controlled environment and maintained by School of Physics IT staff (approximately 3 FTE) who support the computing needs of the School. The Particle Physics Group's Tier-3 facility, which supports ATLAS, Belle II and some theory work, resides in this infrastructure.

The University of New South Wales

The University of New South Wales operates the Katana shared computational cluster, which provides a total of more than 3 000 CPU cores spread over almost 200 nodes, with up to 1 TB memory each. General use of Katana is available to anyone at UNSW, but by contributing additional nodes, a group's level of access to Katana's resources can be increased, proportionally to their investment in the system.

The University of Wollongong

The Centre for Medical Radiation Physics runs its own local cluster, consisting of 200 cores shared by $\sim$40 users per year. Centre staff also make extensive use of NCI and CERN computing resources (Sec. 2.1).

3 Impacts of HEP computing on Australian society

3.1 Improving quality of life with medical physics

Cancer is one of the greatest economic health burdens in the world, costing the global economy two trillion dollars annually and growing rapidly. Twenty percent of all death is due to cancer worldwide. Radiation therapy is a cornerstone of cancer treatment, essential for half of all patients. In Australia over 75 000 people with cancer will need radiation therapy in 2020. Advanced computing covers an important role in improving radiotherapy and the quality of life of patients with cancer (QOL). In particular Monte Carlo particle transport codes (such as Geant4; see Sec. 1.4), recognised to be the most accurate method to calculate the dose in the patient and often referred as "gold standard" in medical physics, are widely used

1. to verify patient-specific clinical treatment planning systems,
2. as engines of some treatment planning systems,
3. to study novel radiotherapy treatments and to optimise existing ones.

At CMRP, the use of Geant4 was pivotal to theoretically calculate the radiobiological effectiveness in heavy ion therapy, which is an input parameter to patient treatment planning.

Geant4 has been extensively used to characterise novel silicon microdosimeters designed and developed at CMRP for radiation protection of astronauts. These are currently on the path towards commercialisation (licensed to the Norwegian company SINTEF). Geant4 was also adopted by CMRP to characterise the MOSkin detector, which is also in the pathway of commercialisation (licensed to a start-up company), to measure in vivo dosage of patients having radiotherapy treatment.

All these projects have the ultimate goals of improving QOL, which translates to an economic benefit of billons of dollars just in Australia.

HEP computing infrastructure is needed to develop Monte Carlo simulation tools for medical physics, because although these software tools are recognised as the gold standard in medical physics, they are too computationally intensive to be introduced to routine clinical practice.

3.2 Advanced workforce training in computing, statistics and data science

The nature of Australian employment is changing rapidly. The oncoming era of automation will remove many jobs in society, but create many more for those who are able to develop machine learning systems. At the same time, the huge amount of data collected by public and private organisations has led to a global data science industry that is expected to be worth $400 billion by 2022, and which is poised to deliver revolutions in healthcare, agriculture, manufacturing, mining, finance and myriad other sectors. Data science involves a combination of expertise in a particular discipline (e.g. physics, biology), with mathematics, computer science and statistics, and the first of these is best supplied by conventional scientific training.

The data science industry employs large numbers of high energy physicists for a number of reasons. Firstly, the massive datasets collected by collider experiments mean that high energy physicists acquire novel skills in *big data* processing that are well beyond the norm. Secondly, machine learning in a science context is quite different from a conventional industry environment. Unlike the industry setting, where practitioners frequently use off-the-shelf frameworks, frontier research in physics forces the development of state-of-the-art solutions. Such new solutions can then feed back into industrial applications – both in the form of domain-specific advances, and in a workforce that is not only trained to use machine learning tools, but to innovate at the cutting edge. Finally, the signals in HEP that generate Nobel-prize winning discoveries are typically dwarfed by background processes that are a million times larger. This requires the development of ingenious approaches to anomaly detection that can be re-purposed for a range of problems, from credit card fraud detection and cybersecurity applications to medical diagnosis problems.

The Australian particle physics community is leading the training of this new Australian workforce in several ways:

- The HEP community attracts a large number of graduate researchers from both Australia and abroad, most of whom leave the field to enter industry in Australia.
- The HEP community attracts world-leading international researchers to Australia at the postdoctoral and faculty level, who are active in transferring their skills to Australian industry.
- HEP researchers are leading the charge in developing novel data science teaching initiatives for undergraduates, even for those outside of physics. Examples include the new University of Adelaide program in Applied Data Analytics, which combines studies in public health, economics, physics, biology, chemistry or geology with bespoke machine learning training. The program was proposed and designed by a high energy physicist.

3.3 Driving computing innovation

Particle physics has been directly responsible for a number of the most significant innovations in the history of computing. It is widely known that the World Wide Web was invented at CERN in the late 80s and early 90s as a by-product of the particle research taking place there. More recently, the HEP community's development of grid technology for distributed computing has paved the way for modern cloud computing. These two innovations underpin the operation of the vast majority of modern businesses, throughout every sector of the economy.

The particle physics community have developed the World Large Hadron collider Computing Grid (WLCG), which routinely processes 50 GB of data per second over a collection of computing resources located throughout the world. This presents an annual throughput of 1500 PB of data per year. By way of scale, the SKA has a requirement of 300 PB of data processing per year. The WLCG achieves this securely and far more cheaply than commercial offerings. All of our middleware and software solutions are open-source and available free of charge. As such, the WLCG is an outstanding example of cybersecurity done right, and offers many lessons to other organisations as they move to truly large-scale scale data processing.

The field of lattice QCD is recognised as having playing a major role in the rapid advances in supercomputing in particular. Lattice QCD spurred the early development of the QCDSP and QCDOC machines, through to the development of the IBM BlueGene series, culminating in the BlueGene/Q, which topped the world rankings in 2012 when it achieved a peak performance of more than 20 PetaFlops. The lattice community is also recognised for pioneering unconventional supercomputing architectures relevant for scientific calculations. A prime demonstration of this is the fact that they were one of the first scientific communities to harness the power of graphics-card based systems, first through an original paper in 2006 titled *"Lattice QCD as a video game"*, and then through direct collaboration with NVidia. Another example is the development of the QPACE system within the lattice community, which was constructed using the CeLL processor, more commonly found in Sony's PlayStation 3 gaming console.

3.4 Public engagement

Computing plays a crucial role in the HEP community's extensive public outreach and engagement activities. For example, . . .

[needs specific examples of contributions of HEP computing to outreach, including LHC@home projects] Skands: overlap / interconnection with WG5?

The LHC@home project "Test4Theory" offers a unique opportunity for members of the public to engage with particle physics research. As an integrated project in the Berkeley Open Infrastructure for Network Computing (BOINC), the LHC@home project enables volunteers to donate "spare" CPU cycles to accelerator and particle physics applications. Test4Theory was the first such cloud project in the world to be based on virtualisation technology; since 2014 its Science Project Leader has been based at Monash U with CERN hosting the LHC@home computing infrastructure. Test4Theory currently provides an average of about 1,000 cores for vetting and validation of detailed particle physics simulations, such as the Pythia program in which Monash is also a key collaborator. Results are publicly available at `mcplots.cern.ch`, and there are lively discussions on the forums of the project.

4 Future requirements and opportunities

4.1 ATLAS

The ATLAS computing model assumes constant funding (with no inflation). Additional growth in capability to meet future computing needs will occur through technology and software improvements. This is a significant effort on behalf the world community of particle physics computing professionals.

Accordingly, the experiment is pushed to employ new technologies and nationally-provided computing services. Within the USA, the HEP community has been encouraged to deploy grid computing within very high capacity HPC systems. This has been a challenging endeavour, but has proven successful. It may be worth considering approaching NCI and Pawsey with a similar scheme, although in the case of the USA, the scale of the interaction was such that it was worth dedicated manpower to make it happen. This will almost certainly be the case for NCI and Pawsey in Australia; however, it is not clear that valuable manpower should be used this way.

The HEP community has also made extensive use of cloud resources in both nationally funded research-only facilities and also in commercial clouds. We in Australia have made substantial use of the NeCTAR cloud for our Belle II Tier-2 facility, and may also be able to re-use additional out-of-warranty compute hardware, seeing as grid computing is inherently fault-tolerant.

For Tier-2 storage we should continue to take advantage of WLCG-compliant, non-POSIX-compliant network file systems, which are substantially cheaper than commercial offerings for the same scale. To this end we have requested funding from a LIEF grant to provide 3.8 PB of data storage for ATLAS and Belle II Tier-2 facilities.

Our Tier-2 facilities are now in a very precarious state, as funding for their ongoing maintenance in both hardware and people is no longer available. We have written LIEF grants to support the hardware required but the ongoing human resources required to support the grid facilities are not in place.

4.2 Belle II

Belle II's use of grid computing resources will grow as more data are collected. Our current best estimates for the Belle II grid resources over the years 2021–2025 are:

tape	4.2–35 PB
disk	12–52 PB
CPU	600–1100 KHEP-SPEC

(1 KHEP-SPEC of computing resource is approximately 100 CPU cores.)

Australia provides around 2% of the Ph.D. physicists working on Belle II. The tape storage will be provided by large regional data centres, typically associated with existing Tier-1 grid sites. Since we do not have one of these in Australia, we are not required to make a contribution to tape storage. Applying the 2% contribution to the other resources our contribution should be:

tape	0 PB
disk	0.24 PB - 1.0 PB
CPU	12 KHEP-SPEC - 22 KHEP-SPEC

In addition to this we require access to a general-purpose cluster of 1000 CPU's and several hundred TB of data storage for the final stage of analysis. This can easily be shared with ATLAS, LHCb, COMET, medical and accelerator physics and theory groups. An enabling technology for this feature is the use of CVMFS for distributing the latest iterations of the scientific software to the compute nodes.

As described in the ARDC section of the document, we plan to build a platform to supply this via a combination of internal, University, ARC LIEF and ARDC funding.

4.3 LHCb

The main purpose of the LHCb experiment is to search for phenomena in the Universe that are not described by the current Standard Model of particle physics. This is done by analysing the decays of billions of beauty and charm hadrons for indirect signatures of much heavier virtual particles mediating the decays. The experiment completed its first phase in 2018 and an upgraded experiment will start to take data at a much higher rate in 2021, and continue until the end of the LHC in 2035 or beyond. To perform this kind of precision measurements involves data retention rates of 10 GB/s, resulting in storage requirements for the experiment as a whole of around 500 PB by 2025. The requirements for simulation of the experiment and the physics to match this are very large as well, and are in fact projected to take up 90% of the offline computing CPU resources. In total, by 2025 the offline CPU requirements will correspond to about 3 000 kHS06 (about 120 000 cores as of 2020) operating continuously.

Computing in the LHCb experiment is carried out within the framework of the *LHC Computing Grid* (LCG). All countries that have institutes in the experiment are expected to contribute to this. To make entry easier for countries with smaller contributions, the data simulation is distributed world wide while the main part of the data analysis is located at six centres in Europe and the USA. The requirements for computing in Australia are threefold:

- Support for data analysis carried out by Australian research staff and students on highly processed data. The requirements are for quick turnarounds and interactive access to modest amounts of data. The data storage requirements will not exceed 1 PB for this. Data access should be easy, through both grid computing and on an interactive basis from desktops.
- Full scale simulations for the LHCb experiment. The requirements in terms of storage are very limited, but it is mandatory that the computing system is fully integrated into the LCG and has high-speed network access. The DIRAC workload management system will be used for the creation of the simulation requests in a fully automated way. The support of virtualisation is mandatory, as simulations must be carried out with the software precisely as it was configured in the years where the corresponding data was taken. Based on number of collaborators in the experiment, Australia would be expected to contribute to this with around 30 kHS06.
- Development of new data analysis paradigms. This typically requires the creation of small-scale simulation followed by extensive fitting of the data. The storage requirements are very small, but CPU requirements are large. Over the next decade, the use of accelerator technologies, in the form of CPU resources with associated GPUs, will become mandatory in this area.

Ulrik: While this is most relevant for LHCb, it could go into some wider software development section that justify the need for manpower as well. Pat: good point; we should probably have a section for software development in the summary that collects needs such as this and presents them compactly (in addition to keeping the following text here). In addition to hardware resources, it is necessary to maintain and develop the software stack associated with the LHCb experiment. Australia has a commitment to support and develop the *Ganga* user interface to distributed computing resources for analysis. The entire LHCb collaboration as well as many other science projects depends on this open source project for managing the access to large and heterogeneous computing resources by individual scientists. In total about 0.5 FTE is required for this.

4.4 COMET

The COMET experiment is currently under construction. Its main purpose is to search muon conversion into an electron through interaction with a nucleus, but with no emission of neutrinos. As the process is forbidden within the Standard Model, a discovery, even at an extremely rare level, would be a signature of New Physics. The main challenge of the experiment is to reduce the background level from beam-induced processes and cosmic rays by many orders of magnitude. To understand this background, huge simulation samples are required which, as in the case of LHCb, translates into requirements on CPU power but very limited requirements on storage.

4.5 SABRE & ARC Centre of Excellence for Dark Matter Particle Physics

Three significant developments for Australia in dark matter research have occurred over recent times. The first is the funding and currently ongoing construction of the Stawell Underground Physics Laboratory (SUPL) in the Stawell Gold Mine in regional Victoria. The funding of $10M was provided equally by the federal and Victorian governments. The initial motivation for SUPL was to host dark matter direct detection searches and to provide an ultralow background radiation facility for ANSTO. SUPL is expected to become available in early 2021 and the aspiration is that it become a long-term national underground facility. The second development is the SABRE South experiment, which has been funded by several ARC LIEF grants over the last few years. SABRE South is currently undergoing construction, and is expected to be ready for installation in SUPL as soon as the lab becomes available. SABRE South is anticipated to be one of a pair of identical experiments, with the southern detector at SUPL and SABRE North at Gran Sasso in Italy. The purpose of the dual experiments is to eliminate unaccounted-for seasonal backgrounds as the possible source of the enigmatic DAMA-LIBRA annual modulation signal. The third significant development is the recent funding of the ARC Centre of Excellence for Dark Matter Particle Physics (ARC funds of $35M over 7 years), involving the University of Melbourne, the University of Adelaide, Australian National University, Swinburne University of Technology, the University of Sydney and the University of Western Australia.

The computing and data storage needs of SABRE are primarily in the form of data storage and analysis, detector design and simulation and the modelling of backgrounds. In addition to direct detection with SABRE, the Centre will have an ATLAS component associated specifically with dark matter searches, and will be doing R&D on advanced detector development and simulation. The theory needs of the CoE will have a GAMBIT component, general theory model-building, and some aspects of lattice gauge theory calculations relevant to dark matter.

4.6 Theory

A few dot points for now, to be expanded with inputs from across the community and then turned into something more cohesive after the April 17 meeting:

- GAMBIT HPC requirements are only going to grow. A third of the GAMBIT Community is now based in Australia, but much less than an third of the compute load is filled from Australian resources; a disproportionate amount is filled by European HPC resources. This reduces the impact and visibility of the Australian groups within GAMBIT relative to the European groups, and is not a secure access route to HPC resources in general (access rules may change to exclude non-European applicants or users at any time). The inauguration of Gadi will help, but NCMAS and partner shares are still unlikely to be able to provide more than 10–20% of the GAMBIT HPC requirements going forward. Large-scale allocation schemes such the new Australasian Leadership Computing Grants (administered by NCI) constitute a very welcome addition to this landscape, and will likely be the only means by which the Australian GAMBIT contingent can make a proportionate HPC contribution in coming years.

- CDMPP theory requirements: Some HPC will be necessary. However what will mostly be needed are central university-supplied resources and purpose-built clusters that could be joint with the pheno community. A common operating system and suite of tools across the clusters might be desirable but this capability could also be delivered as a cloud resource.
- Pythia development in Australia is currently done mostly on personal computers, occasionally supplemented by Monash cloud resources. The standard in the field is however quickly moving towards precision matrix elements, in particular calculations at NLO accuracy and beyond, and this is likely to significantly increase future resource requirements not only for users but also on the author / developer side. There is no formulated strategy yet for how to deal with that, but it is a looming issue.
- Upcoming high-precision measurements of the large scale structure of the Universe will lead to increased requirements for numerically demanding calculations of non-linear structure growth. These will be useful for testing theories with new light bosons, non-standard neutrino properties and/or interacting dark matter.
- Lattice HPC requirements
- MasterCode computing requirements
- Other small-scale compute requirements of pheno community (cloud?)
- more?

4.7 Medical physics

The extensive use of machine learning (ML) algorithms in medical imaging will significantly expand the horizon of opportunities, including:

- Radiomics projects. A large number of features are extracted from radiographic medical images using ML algorithms, which will contribute to defining patient-specific cancer treatments. This research will then translate into an improvement of the quality of life of cancer patients.
- AI for advanced analysis of medical imaging (MRI) in the pathway of personalised medicine.
- With the boosting of space mission programs in Australia, space medicine represents an opportunity for expansion of scientific computing. In this context, HEP computing infrastructure will be needed to study the effect

of radiation on astronauts (at a macro scale, but also at cellular level), and to develop novel radiation shielding solutions by means of Monte Carlo simulations; the latter will prove a stimulus for new materials science.

- A future pathway will also include the analysis of large data sets.

These opportunities are very computationally demanding, and would receive a substantial advantage from the establishment of a HEP computing infrastructure in Australia to be shared with the medical physics community. To expand HEP computing is crucial for future knowledge gain, to explore and better understand nature and to improve the quality of life of the Australian people by improving our ability to treat cancer.

4.8 Community-wide initiatives

The HEP community have a developed a proposal for a computing platform to support Australian particle physics, and applied to ARDC to fund the development of the platform in the first funding round last year. Whilst that application was not successful, we plan to submit a revised proposal to the next funding round.

This year we developed a LIEF grant (LE210100072) to fund the hardware required for the platform. That proposal would provide the equipment required for a scalable platform to enable Australian particle physicists to lead global scientific endeavours. It would be comprised of (i) international Computing Grid services closely coupled with (ii) a suite of integrated compute and data processing services. By distributing the facilities across the Research Computing facilities of the partner institutes, we aim to enable our researchers to make optimum use of local connectivity, infrastructure and expertise. In addition, we would enable easy collaboration by providing resources owned by the particle physics community and accessible via a community authorisation and authentication system, based on the Open ID connect framework.

It is our intention to make another application to the ARDC to fund the FTE required to build the platform. Separately, we have made representations to the University of Melbourne Research Computing services team to work collaboratively on the development of the authorisation and authentication system.

4.9 Cross-community initiatives

Initiatives of possible relevance:

- National Data Centres
- CERN-SKA Computing Collaboration Agreement

Anyone with any current info about either of these, or with ideas for other things to add to this list, please let us know!

5 Summary

- Discussion of interplay with other parts of roadmap
- executive summary of needs, broken into hardware purchases + support personnel and software development + related FTE support
- Discussion of committed/desired/possible/aspirational status