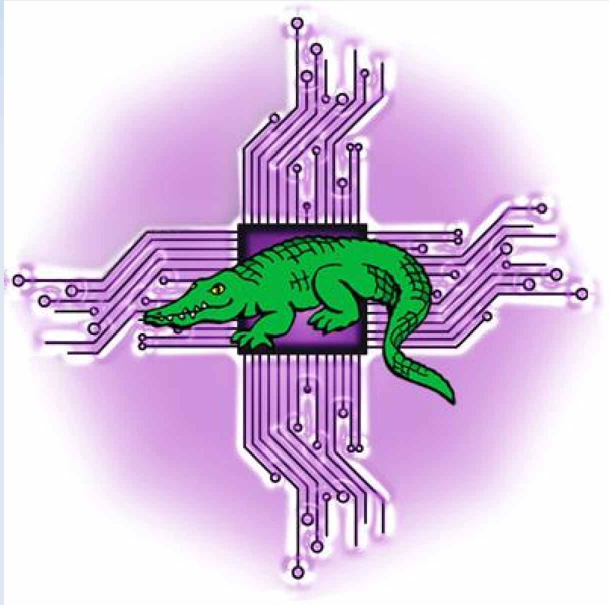
New Developments in Supercomputing

Presentation to Linux Users of Victoria



Melbourne, September 4, 2018

http://levlafayette.com

Supercomputing and Linux

- Supercomputers are defined as those systems which are at the peak of processing power at a particular point in time. For the past twenty years or so these have been entirely high performance compute clusters, systems where scheduler on a management node allocates resources and jobs to multiple redundant computer nodes.

- The typical measure in supercomputing is the Top 500 (systems), based on measured floating point operations per second. Linux continues to be absolutely dominant operating system in the supercomputing world, holding all 500 out of 500 systems (June 2018). Historical data includes 500 (Nov 2017), 498 (June 2017), 497 (June 2016), 489 (June 2015), 485 (June 2014), 476 (June 2013), 462 (June 2012), 457 (June 2011), 456 (June 2010), 442 (June 2009), 427 (June 2008), 367 (June 2007).

Supercomputing and Linux

- Reasons for this dominance are well established:

(a) The command line interface provides a great deal more power and is very resource efficient. The GNU/Linux operating system and utility suite scales and does so with stability and efficiency.
(b) Critical software such as the Message Parsing Interface (MPI) and nearly all scientific programs are designed to work with GNU/Linux. Linux is based on UNIX which is based on Multics means more than fifty years of software development.
(c) The operating system, utilities, and many applications are provided with "free and open source" licenses which are better placed to improve, optimize and maintain, and has replaced the proprietary UNIXes.

International Supercomputing Conference

International Supercomputing Conference

- The International Supercomputing Conference is the European equivalent of the ACM/IEEE Supercomputing Conference which has been held in the US. The precursor was the "Mannheim Supercomputer Seminar" (1986) which became the International Supercomputing Conference and Exhibition (ISC).

- Since 1993 the conference has been the venue for one of the twice yearly TOP500 announcements are named. Conference also hosts multiple award ceremonies; Hans Meuer Award (outstanding research paper), Gauss Award, PhD Forum Award, ISC Research Poster Award, and PRACE-ISC Research Poster Award.

International Supercomputing Conference

- First 15 ISC conferences were in Mannheim with 81 (1986) to 257 (2000) attendees. After that it was in Heidelberg ranging from 340 (2001) to 650 (2005) attendees, then Dresden with 915 (2006) to 1375 (2008), Hamburg 1670 (2009) to 2403 (2012), Leipzig 2423 (2013) to 2405 (2014), and Frankfurt 2846 (2015) to 3505 (2018) (up from 3253 last year).

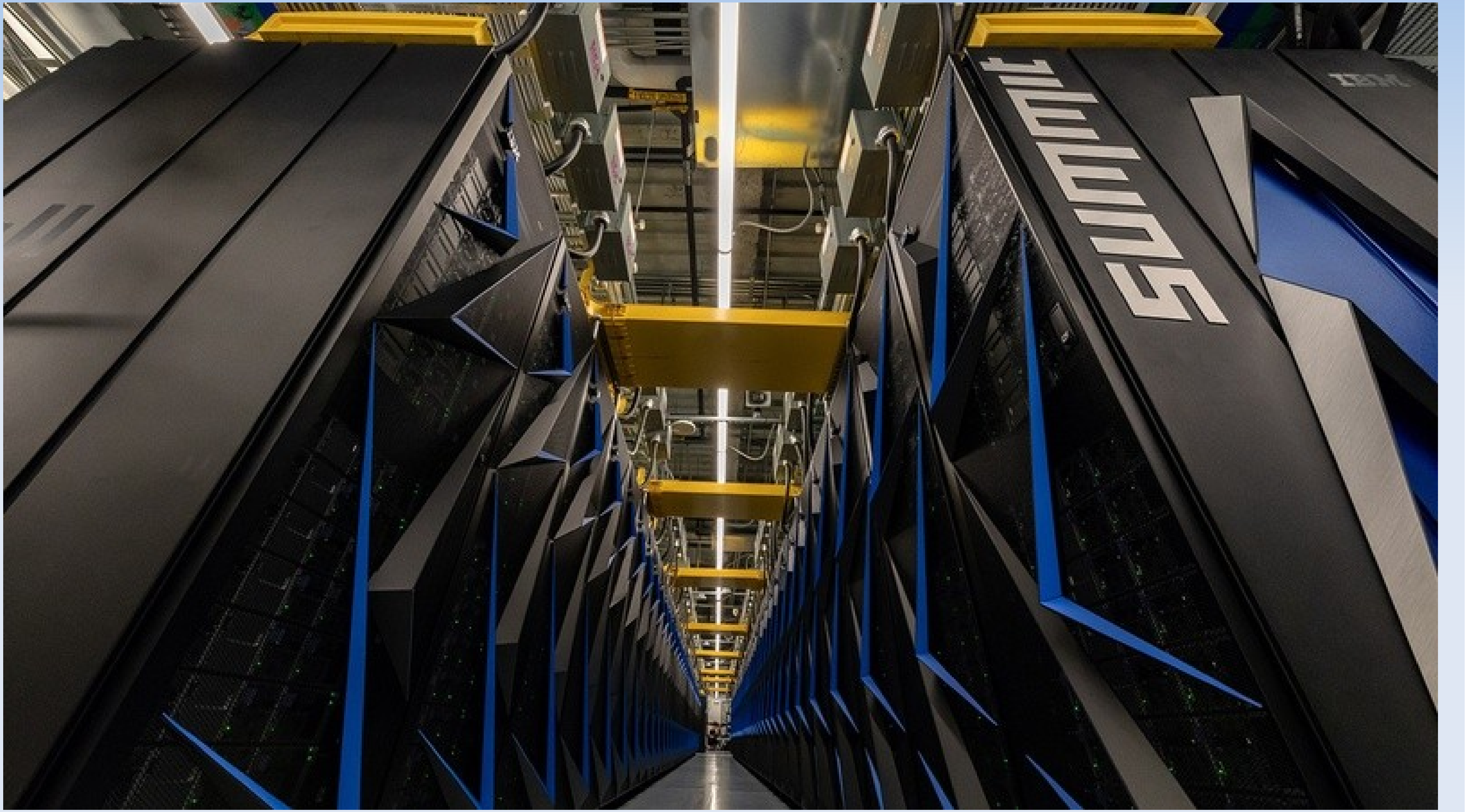
- Conference about 42% academia, 41% industry, 14% students, and 3% media. Main population groups are German (1149), USA (685), UK (315), Japan (186), France (150), China (130). Steering committee is now some 70 people, mainly Germany (25), USA (19), UK (5), Japan (5), Switzerland (4), China (PRC 2, ROC 1), including chair. Still nobody from Australia or New Zealand!

Top 500 June 2018 and Beyond

- **Big changes in the Top 500 announced at ISC included the U.S. taking the number 1 place for the first time since 2012. "Summit", an IBM system used at Department of Energy's (DOE) Oak Ridge National Laboratory (ORNL), has a peak performance of 122.3 petaflops. "Summit" has 4,356 nodes, each with two 22-core Power9 CPUs, and six NVIDIA Tesla V100 GPUs using Mellanox dual-rail EDR InfiniBand network. The number 2 system was "Sunway TaihuLight", from China's National Research Center of Parallel Computer Engineering & Technology (NRCPC). The DOE also took the number 3 place with "Sierra" at 71.6 petaflops with a similar architecture to "Summit". Overall, US systems contribute 38% to aggregate performance, China has 29%.**

- **Petaflop increases over the years for number one system: Summit 122.3 petaflops (June 2018), Sunway TaihuLight 93.0 (June 2017), Sunway TaihuLight 93.0 (June 2016), Tianhe-2 33.9 (June 2015), Tianhe-2 33.9 (June 2014), Tianhe-2 33.9 (June 2013), Sequoia, 16.3 (June 2012), K Computer, 8.2 (June 2011).**

Top 500 June 2018 and Beyond



Top 500 June 2018 and Beyond

- There has been long recognition that the Top 500 use of HPLinpack to measure performance is quite limited; it concentrates on floating point operations whereas HPC computational tasks will also include memory bandwidth, communication capacity, I/O etc. Also the Top500 has a high degree of anonymity and repeat submissions.

- The High Performance Gradients (HPCG) benchmark is emerging as strong and disruptive alternative with a 50% crossover with Linpack. HPCG uses a preconditioned conjugate gradient algorithm, global collective operations, and sparse data structures, aimed to stress the memory subsystem, as this is what most systems do and are more limited in memory than compute.

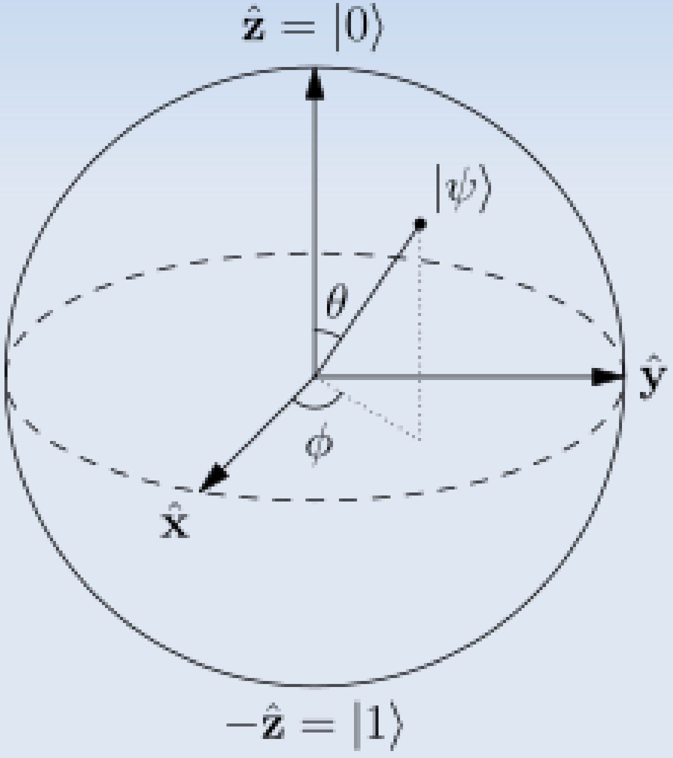
Architecture and Application Debates

- Architecture debates face limitations of nanoscale technologies which provide the second serious challenge to Moore's Law (in addition to heat/power issues in clockspeed). In recent years there has been a significant shift towards GPUs in HPC, which correlates to Linpack performance metrics, but also to the sort of processing common for machine learning and data analytics; NVidia's V100 (launched May 2017) is the industry leader. Current Top500 shows extreme heterogeneous coarse grain systems, cpu+gpu+fpga (e.g., Summit) and extreme homogeneous fine grained, myriad lightweight cores (e.g., Taihulight) systems.

- Intel is facing challenges following dropping Knights Hill Xeon Phi processors in 2017, and being confronted by Spectre and Meltdown in 2018, whilst at the same time ARM and AMD are making new headroads. AMD now has EPYC an alternative to Intel's Xeon with more cores, better I/O and memory connectivity etc than Skylake Xeon. Fujitsu have announced A64FX, an ARM CPU to Japan's first exascale supercomputer, Post-K, planned for 2021.

Architecture and Application Debates

- **Potential designs under consideration discussed in Thomas Sterling's keynote; quantum computing, neuromorphic computing, optical computing, superconducting computing, and non-von Neumann architectures, all of which are in nascent levels of development. Quantum computing uses quantum bits or qubits, which can be in superpositions of states. Neuromorphic computing has diverse approaches to mimic the human brain. Optical computing has been pursued in active data storage and logical data. Superconducted computing is cryogenic computing using the unique properties of superconductors (e.g., zero-resistance wires). Non-von Neumann architectures include cellular automata and data flow.**



HPC Education

- ISC included the poster exhibition of the International HPC Certification program (https://2018.isc-program.com/?page_id=10&id=proj129&sess=sess144), and the inaugural face-to-face meeting of the group. Seventeen people attended, with Julian Kunkel taking up the role as program chairperson; Kai Himstedt as the curriculum chair, Anja Gerbes as topic chair for performance engineering, Jean-Thomas Acquaviva as the topic chair for HPC use, and Lev Lafayette as the topic chair for HPC knowledge. The topic chair for software development is currently vacant. The publicity chair is Weronika Filinger.

- Project does not specify teaching methods or curriculum, but will rather concentrate on outcomes. Individual organisations will determine how they deliver content. Current development includes a draft high-level curriculum on Github, a website (https://www.hpc-certification.org/), active mailing list, regular monthly meetings (alas, at 1:00 am in the morning for AEST), a planned BoF at SC, and coordination with PRACE.

ISC Workshops

- ISC also has a day of tutorials and a day of workshops. The latter represents mini-conferences in their own right. I attended two (in adjacent rooms), the "Workshop on Performance and Scalability of Storage Systems" (https://hps.vi4io.org/events/2018/iodc) and "Workshop on Performance and Scalability of Performance Systems" (https://wopsss.org/).

- These were mostly examples of tools for observation and policies; for example Argonne National Laboratory uses Darshen for observation with policies so not to overload system, based on existing libraries for non-volatile memory (e.g., libbpmem, libpmemblk etc).

- From the latter, a performance study on non-volatile memories; latency test on GPFS, SSD, NVME (slower, faster, fastest) - all unsurprising. However NVM Express has much higher R/W to SSDs. NVME are really good for reads, and scale, writes are usually slower than reads.

HPC Advisory Council

- The HPC Advisory Council held a small conference in Fremantle, Western Australia (28-29 August) covering various designs and practises mainly within the local community. Particularly useful presentations included the software usage metric tool, XALT which an extension to the LMOD environment modules system, which itself is integrated into the EasyBuild software build system. Significant discussion on the development of systems at Pawsey (who hosted the event), and the International Centre for Radio Astronomy Research (ICRAR) for data management, with a presentation on the use of HPC for gravitational wave discovery.

- On a broader level (also a topic at ISC) was the European Processor Initiative (EPI), a program to develop a processors domestic supercomputers, based on ARM (as the CPU) and RISC-V (for the accelerator). This was launched in March 2018 by the European Commission. Currently planned to for deployment across the EU in 2020 and 2021 and to power systems for 2023-2024 for both HPC systems and automotive industry.

THANKS FOR WATCHING
& LISTENING PATIENTLY
makeameme.org