**CSCE 735 Parallel Computing**

**Minor Project**

Sai Namith Garapati

**Supercomputer Fugaku:**

The Fugaku supercomputer is a Japanese supercomputer that gets its name from the nation’s famous peak “Mt Fuji”. Fugaku was being manufactured jointly by Fujitsu and Riken since 2014 and it marked its completion in 2021. Fugaku runs on the custom-made Fujitsu Arm64FX microprocessor whereas the CPU is built upon the ARM version 8.2A instruction set architecture which also adapts the Scalable Vector Extensions for supercomputers. Fugaku contains a total of 158,976 nodes whereas a single node consists of 48 + 2 assistant cores. A single CPU makes up a node. Two CPUs which account for two nodes are mounted on a board called the CPU Memory Unit (CMU). Eight CPU Memory Units (CMUs) put together make up a “Bunch of Blades (BoB)’”, hence each BoB consists of 16 nodes. A shelf constitutes three BoBs which inherently make up 48 nodes. Eight shelves which amount to 384 nodes are put together in a computer rack where some racks can constitute 192 nodes. Fugaku consists of 432 such racks among which 396 racks consist of 384 nodes and 36 racks comprise 192 nodes which account for a total of 158,976 nodes. In this way, the system architecture of Fugaku is laid out as shown in Figure 1.

In terms of Performance, there are two modes Normal mode and boost mode in which Fugaku is run. CPU clocks up to 2GHz in normal mode whereas it clocks up to 2.2GHz in boost mode. It can be observed from the performance statistics that boost mode usually gives a 10% improvement in performance in comparison to the normal mode. The total memory of Fugaku is 4.85 PiB and the total memory bandwidth is 163 PB/s which is one of the prominent features of Fugaku. Fugaku’s power consumption lies in the range of 26248.36kW in optimal case to 29,899.2 kW on average. In Fugaku, Parallelism is achieved through the application of vector instructions used to run applications at high speed and also through optimizations.

The interconnect used in Fugaku is Tofu Interconnect D which is a 6D mesh /torus interconnect that helps in establishing communication between nodes. Remote Direct Memory Access (RDMA) engine helps us in achieving low latency and high throughput. The Latency achieved with 8B input is observed to be 0.49 – 0.54 microseconds whereas the throughput obtained with 1MiB input is 6.35 GB/s. The structure of the TofuD interconnect is given in Figure 2. It consists of six Tofu Network Interfaces (TNIs) each performing at 6.8GB/s accounting for a total of 40.8 GB/s. These six TNIs are connected to the Tofu network router which consists of 10 ports consisting of 2 lanes each. Fugaku runs on the Red Hat Enterprise Linux operating system. Fugaku supports Fujitsu MPI and MPICH-Tofu which are based on OpenMPI and MPICH respectively.

**Sierra:**

Sierra is an advanced High-Performance Computing (HPC) system developed for the CORAL (Collaboration of Oak Ridge, Argonne, and Livermore national laboratories). CORAL is a unique collaboration between the National Nuclear Security Administration (NNSA) and the Office of Science’s Advanced Scientific Computing Research (ASCR) program. Sierra is manufactured by IBM in collaboration with NVIDIA and Mellanox Technologies. Sierra leverages the IBM Power 9 CPUs along with the NVIDIA Tesla V100 (Volta) Tensor core GPUs. Sierra contains 4320 nodes and a total of 1,572,480 cores. It uses the Dual rail Mellanox EDR Infiniband as its Interconnect. Sierra’s heterogenous form of architecture takes huge advantage of the parallelism supported by GPUs along with providing enhanced energy efficiency. Sierra achieves a peak performance of 125.17-petaflops/second. The power consumption is 7,438.28 KW. Sierra is based on the Red Hat Enterprise Linux, and it supports IBM Spectrum MPI.

The architecture of the Sierra system is laid out in the form of Frames/Racks, Nodes, File Systems, Networks, and HPSS Archival storage as shown in Figure 3. The nodes consist of IBM Power 9 CPUs running at a frequency of 3.1GHz along with the NVIDIA Volta GPUs. Compute Nodes such as Dual-socket IBM POWER9 (AC922) nodes are the heart of the system which ensures parallelism. Compute nodes contain 4 NVIDIA Tesla V100 GPUs per node. Sierra systems comprise the Mellanox 100Gb/s Enhanced Data Rate InfiniBand network which is the inter-node network for MPI communications. It consists of 2 power 9 processors for each node with dual-socket configuration. It consists of 22 cores per socket which amount to a total of 44 sockets per node. Each core supports 4 SMT threads and 176 SMT threads per node. A single GPU consists of 5120 CUDA cores which account for a total of 20480 cores per node. Each compute node consists of 256GB DDR4 memory with 170GBps peak bandwidth per socket. Additionally, a 16 HBM2 per GPU is present which provides a peak Bandwidth of 900GB/s.

Sierra’s interconnect consists of a Tapered Fat Tree along with Single Plane topology. The fat tree topology follows the pattern of more switches in the higher level in comparison to the lower level as shown in Figure 4. It is referred to as tapered since the connections for lower-level switches are incremented by a ratio of 2:1. Since all the nodes connect to a single fat tree network it is a single plane.

**Applications**:

The director of Riken said that Fugaku was born with an “application first philosophy” instead of striving for benchmark excellence. Fugaku has seen recent applications in the research related to coronavirus and global weather simulations. Fugaku also saw major application in providing simulation capabilities for data processing to simulate real-world objects like factories, climate, and weather systems etc. These kinds of applications are called digital twin applications which can produce data at terabytes/s, having the high-performance ability and large capacity for storage. The future motive remains to generate accurate global weather simulations to capture the ever-changing and challenging climatic trends in the world.

Sierra was mainly developed with a vision of providing simulation capabilities for challenging security modelling applications. Sierra helps in providing the necessary high-performance computing resources needed for nuclear weapon scientists to complete the stockpile stewardship mission which requires simulation along with underground testing. This supercomputer ensures the safety, and security of the country’s nuclear obstacle without the need for testing. Along with this Sierra is being used for key scientific and engineering calculations like material modelling problems, instabilities in turbulent flow systems, and laser-plasma calculations.

**Computational Capabilities, performance, and other metrics:**

The computational ability of supercomputers is measured in terms of Petaflops. Let us understand about petaflops before moving forward. Essentially, one Petaflop is equivalent to performing one quadrillion arithmetic operations.

**Fugaku:** Fugaku is the first ARM processor architecture-based supercomputer that stood in the top place with a peak performance of 537.21 petaflops in boost mode.

*Linpack Performance (Rmax)*: 442.01 PFlop/*s Theoretical peak*: 537.21 PFlops/s

*HPCG [TFlop/s]*: 16,004.5  *Power*: 29,899.23 KW

**Sierra:** Sierra’s computational capability peaks a performance of 125 Petaflops equivalent to 1015 floating point operations per second.

*Linpack Performance (Rmax)*: 94.64 PFlop/*s Theoretical peak*: 125.71 PFlops/s

*HPCG [TFlop/s]*: 1,795.67  *Power*: 7,438.28 KW

**Key Comparisons:**

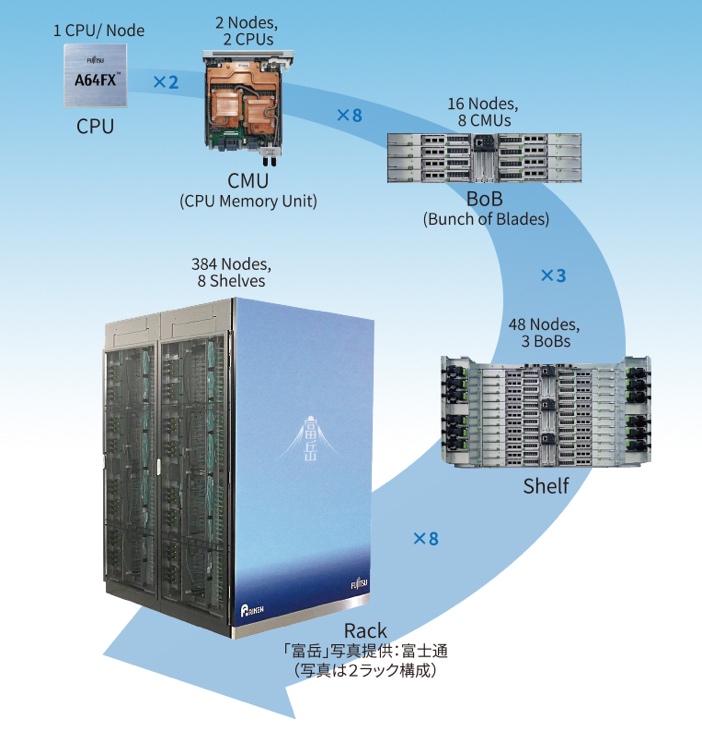
* Fugaku achieves a peak performance of 537.21 Petaflops per second with a total of 7,630,848 cores whereas Sierra showed a peak performance of 125.71 Petaflops/s with a total of 1,572,480 cores which clearly indicates the higher computational ability of Fugaku
* Since the number of cores is more, the power consumption in Fugaku also is high relatively in comparison to Sierra.
* Fugaku has a relatively higher memory of 4.85 PiB in comparison to the 2 – 2.4 PiB memory of the Sierra.
* It can be observed that Sierra clocks a higher clock speed of 3.4 GHz than the boost mode clock speed of 2.2 GHz for Fugaku
* Fugaku uses the Tofu-D interconnect which is a 6D Mesh/torus topology to establish a connection between the nodes whereas Sierra uses Dual rail Mellanox EDR Infiniband which is a Tapered fat-tree topology.

**References:**

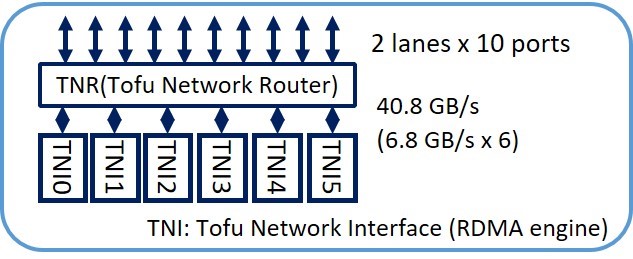
1. <https://top500.org/system/179807/>
2. <https://top500.org/system/179398/>
3. <https://www.r-ccs.riken.jp/en/fugaku/about/>
4. <https://hpc.llnl.gov/hardware/compute-platforms/sierra>
5. <https://www.llnl.gov/news/sierra-honored-top-supercomputing-achievement-hpcwire>
6. <https://en.wikipedia.org/wiki/Sierra>

**Appendix:**

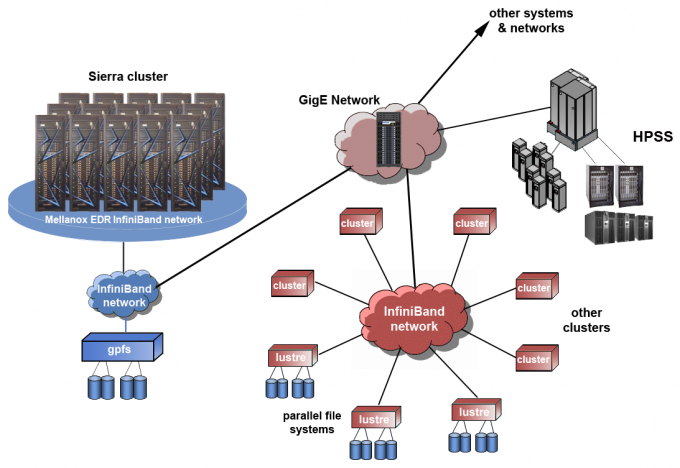
**Figures:**

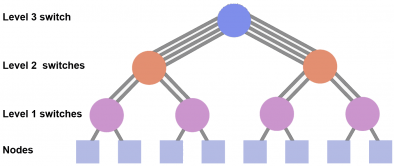


**Fig 2: TofuD Interconnect Structure**



**Fig 1: System Architecture of Supercomputer Fugaku**





**Fig 4: Fat Tree structure**

**Fig 3: System Architecture of Sierra**