Supercomputer “Fugaku”
Formerly known as Post-K
RIKEN and Fujitsu are currently developing Japan's next-generation flagship supercomputer, the successor to the K computer, as the most advanced general-purpose supercomputer in the world.

- RIKEN and Fujitsu announced that manufacturing started in March 2019.
- RIKEN announced on May 23, 2019 that the supercomputer is named “Fugaku”
Goals and Approaches for Fugaku

**Goals**

- High application performance
- Good usability and wide range of uses
- Keeping application compatibility

RIKEN announced predicted performance:

- More than 100x+ faster than K computer for GENESIS and NICAM+LETKF
- Geometric mean of speedup over K computer in 9 Priority Issues is greater than 37x+

https://postk-web.r-ccs.riken.jp/perf.html
Goals and Approaches for Fugaku

**Goals**

- High application performance
- Good usability and wide range of uses
- Keeping application compatibility

**Approaches**

**Develop**
1. High-performance Arm CPU A64FX in HPC and AI areas
2. Cutting-edge hardware design
3. System software stack

**Achieve**
- High performance in real applications
- High efficiency in key features for AI applications
1. High-Performance Arm CPU A64FX in HPC and AI Areas

- **Architecture features**
  - **ISA**: Armv8.2-A (AArch64 only) SVE (Scalable Vector Extension)
  - **SIMD width**: 512-bit
  - **Precision**: FP64/32/16, INT64/32/16/8
  - **Cores**: 48 computing cores + 4 assistant cores (4 CMGs)
  - **Memory**: HBM2: Peak B/W 1,024 GB/s
  - **Interconnect**: TofuD: 28 Gbps x 2 lanes x 10 ports

- **Peak performance (Chip level)**
  - | Element size | 64 bits | 32 bits | 16 bits | 8 bits |
  - | HPC | A64FX (Fugaku) | 0.128 | 0.128 | N/A | N/A |
  - | AI | SPARC64 VIIIIfx (K computer) | 2.7+ | 5.4+ | 10.8+ | 21.6+ |
1. High-Performance Arm CPU A64FX in HPC and AI Areas

Architecture features

- **ISA**: Armv8.2-A (AArch64 only) SVE (Scalable Vector Extension)
- **SIMD width**: 512-bit
- **Precision**: FP64/32/16, INT64/32/16/8
- **Cores**: 48 computing cores + 4 assistant cores (4 CMGs)
- **Memory**: HBM2: Peak B/W 1,024 GB/s
- **Interconnect**: TofuD: 28 Gbps x 2 lanes x 10 ports

Peak performance (Chip level)

<table>
<thead>
<tr>
<th></th>
<th>HPC</th>
<th>AI</th>
</tr>
</thead>
<tbody>
<tr>
<td>64 bits</td>
<td>A64FX (Fugaku)</td>
<td>0.128</td>
</tr>
<tr>
<td>32 bits</td>
<td>2.7+</td>
<td>5.4+</td>
</tr>
<tr>
<td>16 bits</td>
<td>N/A</td>
<td>10.8+</td>
</tr>
<tr>
<td>8 bits</td>
<td>N/A</td>
<td>21.6+</td>
</tr>
</tbody>
</table>

(TOPS)
2. Cutting-edge Hardware Design

**1 PFlops by Fugaku and K computer**

<table>
<thead>
<tr>
<th></th>
<th>Fugaku</th>
<th>K computer</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Configuration</strong></td>
<td>1x rack including SSDs</td>
<td>80x compute racks &amp; 20x disk racks</td>
</tr>
<tr>
<td><strong>Nodes</strong></td>
<td>384</td>
<td>8,160</td>
</tr>
<tr>
<td><strong>Footprint</strong></td>
<td>1.1 m² (0.8 m x 1.4 m)</td>
<td>128 m² (4 m x 32 m)</td>
</tr>
</tbody>
</table>

**Scalable design**

- 100% direct water cooling
- 3x QSFP for AOC (Active Optical Cables)
- Single-sided blind mate connectors for electrical signals and water

**CMU (CPU Memory Unit)**

- 1 PFlops by Fugaku and K computer
- 2.7 T+ 5.4 T+ 43 T+ 129 T+ 1 P+ 400 P+

---

*Copyright 2019 FUJITSU LIMITED*
3. System Software Stack

- Fujitsu developing system software in collaboration with RIKEN
  - Fujitsu Technical Computing Suite implementing development and execution environments with great usability on large-scale system

<table>
<thead>
<tr>
<th>Management software</th>
<th>File system</th>
<th>Programming environment</th>
</tr>
</thead>
<tbody>
<tr>
<td>System management for high availability &amp; power saving operation</td>
<td>FEFS Lustre-based distributed file system</td>
<td>XcalableMP</td>
</tr>
<tr>
<td>Job management for higher system utilization &amp; power efficiency</td>
<td>LLIO NVM-based file I/O accelerator</td>
<td>OpenMP, Coarray</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Compilers (C, C++, Fortran)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Debugging and tuning tools</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Math. libs.</td>
</tr>
</tbody>
</table>

Linux OS / McKernel (Lightweight kernel)

Fugaku system hardware
3. System Software Stack

- Fujitsu developing system software in collaboration with RIKEN
- Fujitsu Technical Computing Suite implementing development and execution environments with great usability on large-scale system

<table>
<thead>
<tr>
<th>Fugaku applications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fujitsu Technical Computing Suite / RIKEN developing system software</td>
</tr>
<tr>
<td>Programming environment</td>
</tr>
<tr>
<td>Exploits hardware performance by compiler optimizations such as SVE vectorization</td>
</tr>
<tr>
<td>Supports new programming language standards and data type FP16</td>
</tr>
<tr>
<td>Linux OS / McKernel (Lightweight kernel)</td>
</tr>
<tr>
<td>Fugaku system hardware</td>
</tr>
</tbody>
</table>

- Exploits hardware performance by compiler optimizations such as SVE vectorization
- Supports new programming language standards and data type FP16
- Exploits hardware performance by compiler optimizations such as SVE vectorization
- Supports new programming language standards and data type FP16
Achieve High Performance in Real Application

- **WRF:** Weather Research and Forecasting model (v3.8.1)
  - Vectorizing loops including IF statements is key optimization
- **Himeno Benchmark (Fortran90, size: XL)**
  - Stencil calculation to solve Poisson’s equation by Jacobi method

---

**WRF v3.8.1 (48-hour, 12km, CONUS) on 48 cores**

- **Skylake (Xeon Platinum 8168) 2 CPUs**
  - Performance Ratio: 1.00
- **A64FX 1 CPU with source tuning**
  - Performance Ratio: 1.56

**Himeno Benchmark (Fortran90)**

- **Skylake (Xeon Platinum 8168) 2CPUs**
  - GFlops: 85
- **FX100 1CPU**
  - GFlops: 103
- **A64FX 1CPU**
  - GFlops: 346
- **SX-Aurora TSUBASA 1VE†**
  - GFlops: 286
- **Tesla V100 1GPU†**
  - GFlops: 305

*Normalized by the average elapsed time for timestep of Skylake

†Performance evaluation of a vector supercomputer SX-aurora TSUBASA
https://dl.acm.org/citation.cfm?id=3291728

High memory B/W and long SIMD length work effectively
High Efficiency in Key Features for AI Applications

High FP16 & INT8 peak performance and high memory peak B/W

FP16 performance: 10.8+ TOPS, > 90%@HGEMM
INT8 performance: 21.6+ TOPS in partial dot product
Memory B/W: 1,024 GB/s, > 80%@STREAM Triad

Functions contributing to key features in AI fields

A64FX CPU
- 2x 512-bit wide SIMD pipelines per core for FP16 and INT8
- High memory B/W and calculation throughput

Compilers & libraries
- Vectorization and software pipelining
- FP16 as data type of programming language (e.g., real (kind=2) in Fortran)
- Mathematical Library for HGEMM
Future Plans

Supercomputer Fugaku

- Operations starting around CY2021

Fujitsu HPC Products

- Fujitsu will begin global sales of supercomputers based on the Supercomputer Fugaku technology in the 2nd half of FY2019
FUJITSU

shaping tomorrow with you