南开大学

Nankai University

南　　开　　大　　学

计　算　机　学　院

并行程序设计实验报告

体系调研报告

唐明昊

年级：2021 级

专业：计算机科学与技术

指导教师：王刚

2023 年 3 月 12 日

摘要

Fugaku 是由日本研究机构 RIKEN 和富士通公司合作开发的一款超级计算机，是目前世界上性能最强大的超级计算机之一，被广泛应用于各个领域。K computer 也是由 RIKEN 和富士通制造的超级计算机，是 Fugaku 的前身。Frontier 是由美国田纳西州橡树岭国家实验室制造的超级计算机，是当前超算 TOP500 第一名。

本调研报告旨在介绍超级计算机 Fugaku 的体系架构，并同其前身 K computer 和当前 TOP500 第一名 Frontier 的架构进行比较，分析并行体系结构的发展趋势与现状。

关键字：并行, 超算，Fugaku, K computer, Frontier

目录

一、 Fugaku

(一) 概述

Fugaku [3] 是日本研究机构 RIKEN 和富士通公司合作开发的一款超级计算机，它是目前世界上性能最强大的超级计算机之一。

Fugaku 的应用范围非常广泛，它可用于天气预报、地震模拟、流体力学、材料科学、基因组学等领域的高性能计算，同时也可以支持人工智能、机器学习和深度学习等应用。Fugaku 在 2020 年 6 月被列为 TOP500 全球超级计算机排名的第一位，获得了多个国际超级计算机性能和能效方面的奖项，至今仍是超算排名的第二位。

(二) 处理器架构

Fugaku 的节点层采用了高度定制化的 A64FX 处理器，具有高度并行的向量单元，可以在单个指令周期内执行多个数据操作。指令集架构采用 Armv8.2-A+SVE 521bit，主要设计用于高性能计算和人工智能应用。

A64FX 处理器拥有 48 个 CPU 核心，每个核心均可执行高效的 SIMD 向量计算操作，能够同时处理多个线程，每个核心的主频可以达到 2.2 GHz，性能非常强劲。采用了计算存储一体化的设计，在处理器内部集成存储器，减少了 CPU 和存储器之间的数据传输，提高了计算效率。另外 A64FX 处理器内置了性能强大的向量引擎和 Tensor Core 加速器，可支持 FP16 和 BF16 精度的计算，并且支持 INT8 和 I3NT4 量化计算，这使得它在人工智能计算方面非常出色。

内存容量和带宽上 A64FX 处理器采用了 HBM2 内存技术，内存带宽高达 1TB/s，可支持最大 1TB 的内存容量，这使得它非常适合处理大型数据集和深度学习模型。A64FX 处理器还支持 PCIe Gen 4 接口，可提供高达 16GB/s 的带宽，支持高速网络接口，如 100GbE 和 InfiniBand EDR，可满足高性能计算和大规模数据传输的需求。

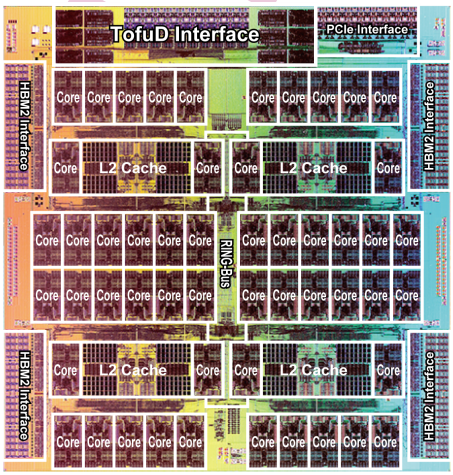

图 1: Fugaku 单节点

(三) 节点架构

Fugaku 采用了非常规的混合节点架构，它将超大规模的节点集成到一个庞大的全局交叉开关网络中，实现了高度的灵活性和可扩展性。

Fugaku 总共有 158976 个节点。单个 CPU 组成一个节点，两个 CPU（两个节点）安装在称为 CPU 内存单元（CPU Memory Unit）的板上。八个 CMU 组成一个“刀片堆（Bunch of

Blades)”，这意味着每个 BoB 有 16 个节点。三个 BoB 组成一个架子，因此每个架子有 48 个节点。八个磁盘架（384 个节点）安装在计算机机架中（某些机架有 192 个节点）。Fugaku 由 432 个机架组成，其中 396 个机架有 384 个节点，36 个机架有 192 个节点。

（四） 互连网络架构

Fugaku 的全局交叉开关网络采用了 Tofu D 网络，它是一个自主设计的高性能互连网络。

Tofu D 网络采用了三维网格拓扑结构，包括一个内层的 28x28 二维拓扑结构和一个外层的 6x6 三维拓扑结构。网络中的节点分布在三个维度上，并且每个节点连接着周围的 6 个节点。这种拓扑结构可以提供高效的通信和低延迟的数据传输。同时，多层拓扑结构还可以支持大规模并行计算。

Tofu D 网络采用了高速光纤传输技术，支持高达 4TB/s 的总带宽和每个节点高达 512GB/s 的带宽。Fugaku 的光纤互连网络使用了自适应路由和预取技术，可以避免网络拥塞和数据丢失。自适应路由技术可以根据网络拥塞情况选择最优的路径，从而减少数据传输的延迟和丢失。预取技术可以提前将需要的数据存储在本地缓存中，减少数据传输的时间和延迟。

它还支持低延迟、高吞吐量的全局广播和全局归约操作，可以加速大规模并行计算和通信密集型应用程序的执行。另外，网络采用了冗余设计，可以保证节点之间的通信不中断，它使用多条路径传输机制和虚拟通道技术，可以保证数据传输的可靠性和完整性。

（五） I/O 存储架构

Fugaku 的 I/O 存储系统提供了高速、低延迟、大容量和高可靠性的数据传输，以满足 Fugaku 系统中大规模计算应用的需求。Fugaku 的 I/O 和存储系统由多个组件组成，包括：

- SSD 存储设备：Fugaku 使用 NVMe（非易失性存储）SSD（固态硬盘）作为主要的存储设备。这些 SSD 可以提供高达 3TB 的存储容量和超过 10GB/s 的传输速度，可以满足大规模数据的快速读写需求。
- 文件系统：Fugaku 采用了 Lustre 并行文件系统作为主要的文件系统。Lustre 具有高并发、高吞吐量和高扩展性的特点，能够支持大规模并行计算应用的高性能 I/O 需求。
- 数据传输组件：Fugaku 的 I/O 存储系统还包括高速数据传输组件，如 PCIe Gen4 总线和高速网络。这些组件能够提供高达 200GB/s 的数据传输速度，以确保快速、高效的数据传输。
- 数据备份和恢复：Fugaku 的 I/O 存储系统有着备份和恢复机制，以确保数据的安全和可靠性。备份机制可以在出现数据丢失或硬件故障时快速恢复数据，以保证计算应用的连续性和可靠性。

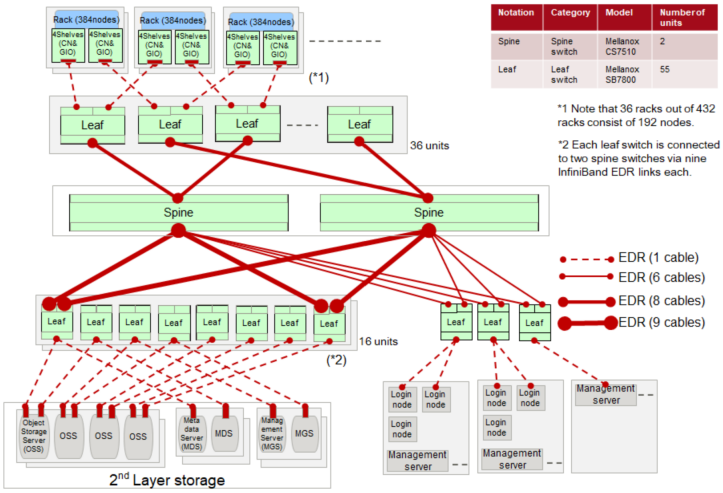

图 2: Fugaku I/O

二、 K computer

(一) 概述

K computer(京) [2] 也是由 RIKEN 和富士通制造的超级计算机，于 2011 年 6 月被 TOP500 列为世界上最快的超级计算机，计算速度超过 8 petaflops。K computer 基于分布式内存架构，拥有超过 80000 个计算节点。K computer 的操作系统基于 Linux 内核，并设计了额外的驱动程序以利用计算机的硬件。

(二) 处理器架构对比

处理器架构方面 K computer 处理器采用的是 SPARC64 VIIIfx 处理器，每个处理器有 8 个核心。这些处理器采用的是 64 位 SPARC V9 指令集，可以支持高效的浮点运算和向量计算。Fugaku 采用的是自主研发的 Arm A64FX 处理器，每个处理器有 48 个核心。这些处理器采用的是 ARMv8.2-A 指令集，支持 512 位 SIMD 向量计算和 128 位 FP16/Half 精度浮点运算。

在内存架构上 K computer 采用的是 NUMA 结构，每个处理器有其本地存储和内存，可以直接访问。不同处理器之间通过高速互连网络进行通信和数据传输。Fugaku 同样也采用了 NUMA 结构，但是在内存管理方面更加灵活和智能。它采用了新型的分布式内存架构，将存储器分为多个分区，并动态地分配和管理内存资源，以最大化内存利用率和性能。

(三) 存储系统对比

K computer 的存储系统采用的是自主研发的高速文件系统和并行文件系统。高速文件系统主要用于存储短期数据，提供快速的读写速度和低延迟；并行文件系统主要用于存储大规模的科学计算数据，提供高带宽和高吞吐量。

Fugaku 的存储系统采用的是新型的 SSD 和内存层次结构，既可以满足高速存储需求，又可以提供大容量存储。Fugaku 的存储系统容量可以达到 150PB，是前者的 15 倍，文件系统带宽可以达到 4TB/s，是前者的 4 倍。

(四) 互连网络对比

K computer 采用的 Tofu 互连网络结构，采用 6 个方向的 4x4 自适应路由交换网络，可以支持不同规模的系统。Fugaku 的 Tofu Interconnect D 互连网络结构，是一种三层结构的网络拓扑结构，由芯片内部和外部互连两个层次组成，可以支持高效的点对点和全局通信。

前者的互连网络带宽为 5.2TB/s，延迟为 3.5us。后者网络带宽可以达到 6.0TB/s，延迟为 1.5us。

Fugaku 的互连网络不仅支持异步通信 RDMA 技术，还支持更复杂的通信模式，如收集、分配、归约和广播等。它还可以自适应地调整网络拓扑结构以满足不同应用程序的需求，并支持任务调度和负载平衡等功能，可以提高系统的灵活性和性能。另外 Fugaku 的互联网络更加节能环保，可以在高温环境下运行，具有更高的可靠性和稳定性。还采用了多种故障预测和纠正技术，系统的容错能力也更高。

(五) 小结

总的来说，从并行体系结构的角度来看，Fugaku 相对于 K computer 具有了更好的能效优势和更高的性能表现。这也表明了 Fugaku 在设计上更加注重系统整体的能效优化和性能提升。未来的超级计算机也必将会在探索更加高效、能耗更低的并行体系结构持续发力。

三、 Frontier

(一) 概述

Frontier [1] 是由美国田纳西州橡树岭国家实验室制造的超级计算机，也是全球首台百亿亿次级计算机，每秒可执行百亿亿次运算。根据 NERSC 的数据，Frontier 的总浮点运算性能将达到 1.5 exaflops，也就是每秒能够执行 10^{18} 个浮点运算。2022 年 6 月，Frontier 超过 Fugaku，获得 TOP500 全球排名第一名，其性能代表了整个榜单计算能力的四分之一。

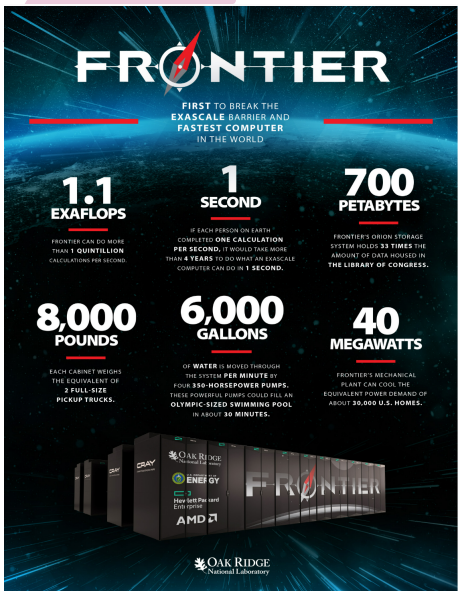

图 3: Frontier

(二) 计算节点比较

Frontier 采用了一种混合的体系架构，将 AMD EPYC CPU 和 Nvidia A100 GPU 结合在一起，以提高系统的性能和效率。

每个 Frontier 节点都配备两个 AMD EPYC CPU，最多可配备 8 个 Nvidia A100 GPU。每个 CPU 拥有 64 个核心和 128 个线程，这些 CPU 将使用 AMD Infinity Architecture 技术连接，以提供高速的内部通信。CPU 节点还配备 1.3 PB 的内存，可在大规模数据集上进行分布式计算。每个 GPU 具有 6,912 个 CUDA 核心，以及 1.6 TB/s 的内存带宽。这些 GPU 使用第二代 Nvidia NVLink 互连技术连接，从而实现高效的 GPU 到 GPU 通信和 GPU 到 CPU 通信。

虽然 AMD EPYC CPU 的核心数比 Fugaku 的 SPARC64 XVIfx 多，但 SPARC64 XVIfx 拥有更高的时钟速度和更大的缓存，可以在某些应用中提供更好的性能。Nvidia A100 GPU 的 CUDA 核心数也比 A64FX 处理器多，但 A64FX 处理器具有更高的矢量计算能力和更高的内存带宽，可以更好地支持科学模拟和机器学习等计算密集型任务。

二者在不同场景各具优势，Frontier 的 GPU 配置使得它在机器学习等计算密集型任务中表现出色，而 Fugaku 的矢量计算能力和内存带宽使得它在科学模拟等任务中更为优秀。

(三) 网络架构比较

网络架构方面，Frontier 采用了全局高速互连网络（GHPI），它是一种由 Cray 设计的自定义网络拓扑，可提供高效的通信性能和低延迟。GHPI 使用自适应路由和非阻塞交换机设计，支持高度并行的通信和数据传输。

Fugaku 采用自主研发的 Tofu 拓扑网络，它采用了类似于二维网格的结构，具有高带宽、低延迟和高可扩展性。Tofu 网络采用了自适应路由和虚拟通道技术，完成高效的通信和数据传输。

对比来看 Frontier 的网络带宽为 4.4 TB/s，网络延迟为 2 微秒。Fugaku 的网络带宽为 6 TB/s，网络延迟为 1.2 微秒。

(四) 小结

总的来说，Frontier 和 Fugaku 都是世界上最快的超级计算机之一，它们在并行体系架构方面都采用了先进的技术和设计理念，具有出色的性能和可扩展性。虽然它们在一些方面有所不同，例如处理器架构、互连网络技术和并行编程模型等，但它们都为超级计算和数据通信领域的发展做出了重要贡献，并在不同的领域实现了极高的性能和能效。

四、 总结

经过以上的调研，我们可以做个简单的总结，当今的高性能计算机系统所采用的并行体系结构的发展趋势主要有以下几个方面:

1. 处理器架构：为了在单个芯片上提供更多的计算能力, 现代计算机处理器的核心数量越来越多。多核架构主要以增加核心数量、提高核心频率、增加缓存容量等，从而提高单个节点的性能。未来会涌现越来越多 ARM，GPU，FPGA 等架构的新型处理器。

2. 内存架构: K computer 采用的是共享式内存架构, 为了提高计算机系统的整体性能, Fugaku 采用分布式架构，通过多个节点之间的网络连接协同工作。分布式架构的发展趋势是提高网络带宽、降低网络延迟、优化网络拓扑结构等，以提高节点之间的通信性能。

3. 互连网络：K computer 采用了 Infiniband 互连网络，而 Fugaku 则采用了 Tofu D 拓扑的高性能互连网络。在超级计算机领域，互连网络的变化可以带来更低的延迟和更高的带宽。未来超级计算机的互连网络也可能会不断变化，比如采用光互连等新型互连技术。

4. 高可靠性和容错性：对于高性能计算机系统，故障率和停机时间的影响非常重要。增加系统的冗余度、提高故障检测和修复能力、提高节点之间的通信鲁棒性等，是常用到来提高系统的可靠性和容错性的手段。

5. 能效优化：高性能计算机系统的能源消耗是一个非常重要的问题。减少系统的功耗、提高计算效率、优化系统的能源管理等，以提高系统的能效是未来发展的目标。

参考文献

[1] Oak Ridge National Laboratory. Frontier, 2022. https://www.ornl.gov/news/frontier-supercomputer-debuts-worlds-fastest-breaking-exascale-barrier.

[2] RIKEN. The k computer. https://www.riken.jp/en/collab/resources/kcomputer/.

[3] RIKEN. About fugaku, 2021. https://www.r-ccs.riken.jp/en/fugaku/about/.