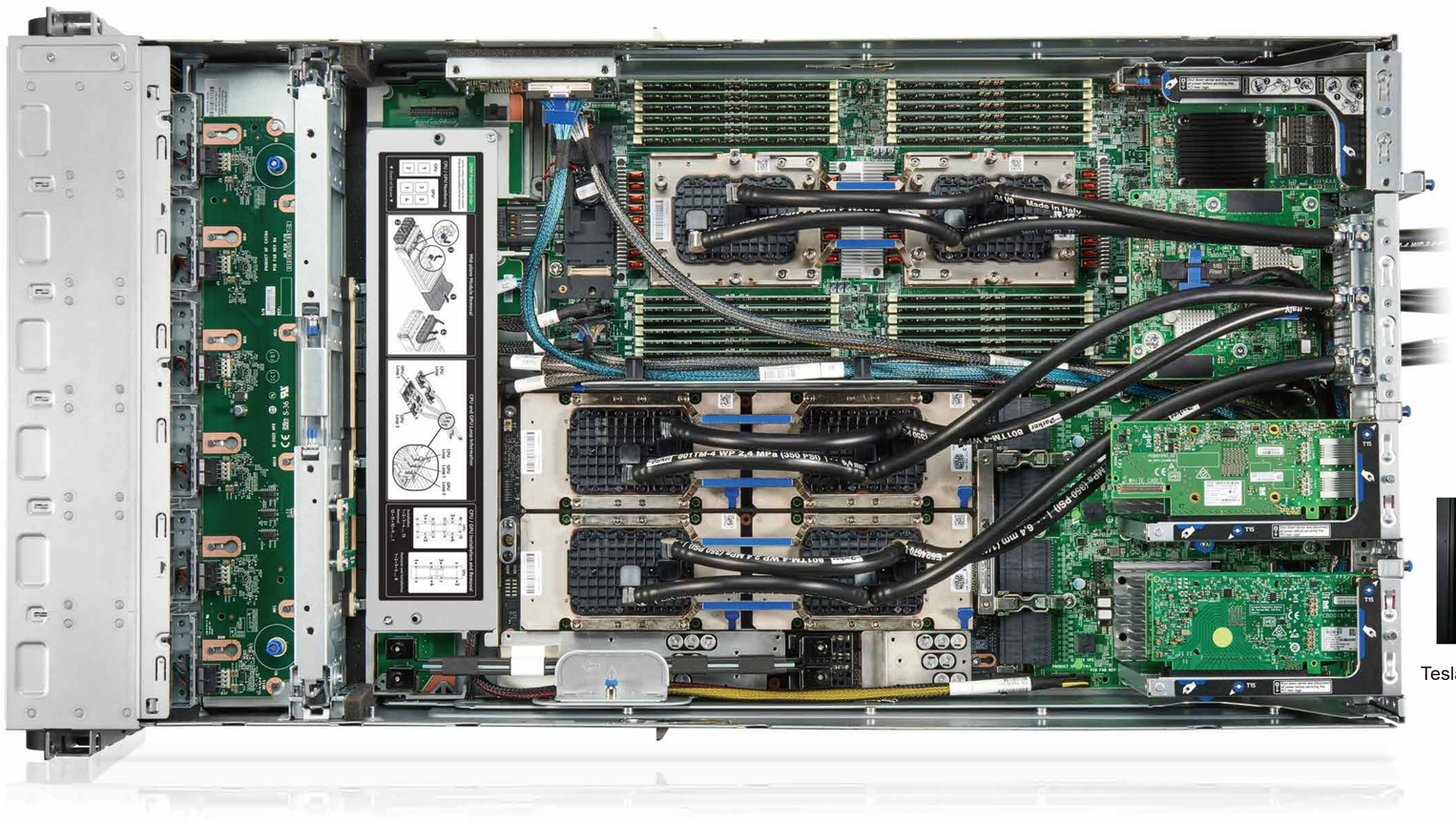
TSUBAME4.0 Compute Node
Water-cooled Node with 4 GPUs

Compute Node

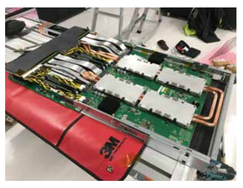


TSUBAME4.0 compute nodes are designed to accommodate all kinds of workloads in university research, including traditional HPC, Big Data, and AI, which require all of computing power, memory bandwidth, network interconnect, and local storage. These components are installed in a 4U rack-mounted node using water cooling.

Tesla S1070 (Tesla GT200) on TSUBAME1.2

Tesla M2050 (Fermi) on TSUBAME2.0

Tesla K20X (Kepler) on TSUBAME2.5



Tesla P100 (Pascal) on TSUBAME3.0

TSUBAME4.0 continues accelerated computing with GPUs.

Specifications

HPE Cray XD665 Server

CPU: AMD EPYC 9654 (2.4GHz, 360W) ×2 sockets
96 cores per socket, total 192 cores per node.

GPU: NVIDIA H100 SXM5 94GB HBM2e ×4
Hopper GPU, 94GB, 2395.87GB/s Memory,
33.5TFlops(FP64), 66.9TFlops(FP64 Tensor),
66.9TFlops(FP32), 494.7TFlops(TF32 Tensor),
989.4TFlops(FP16/BF16 Tensor), 1,978.9TOps(INT8 Tensor)

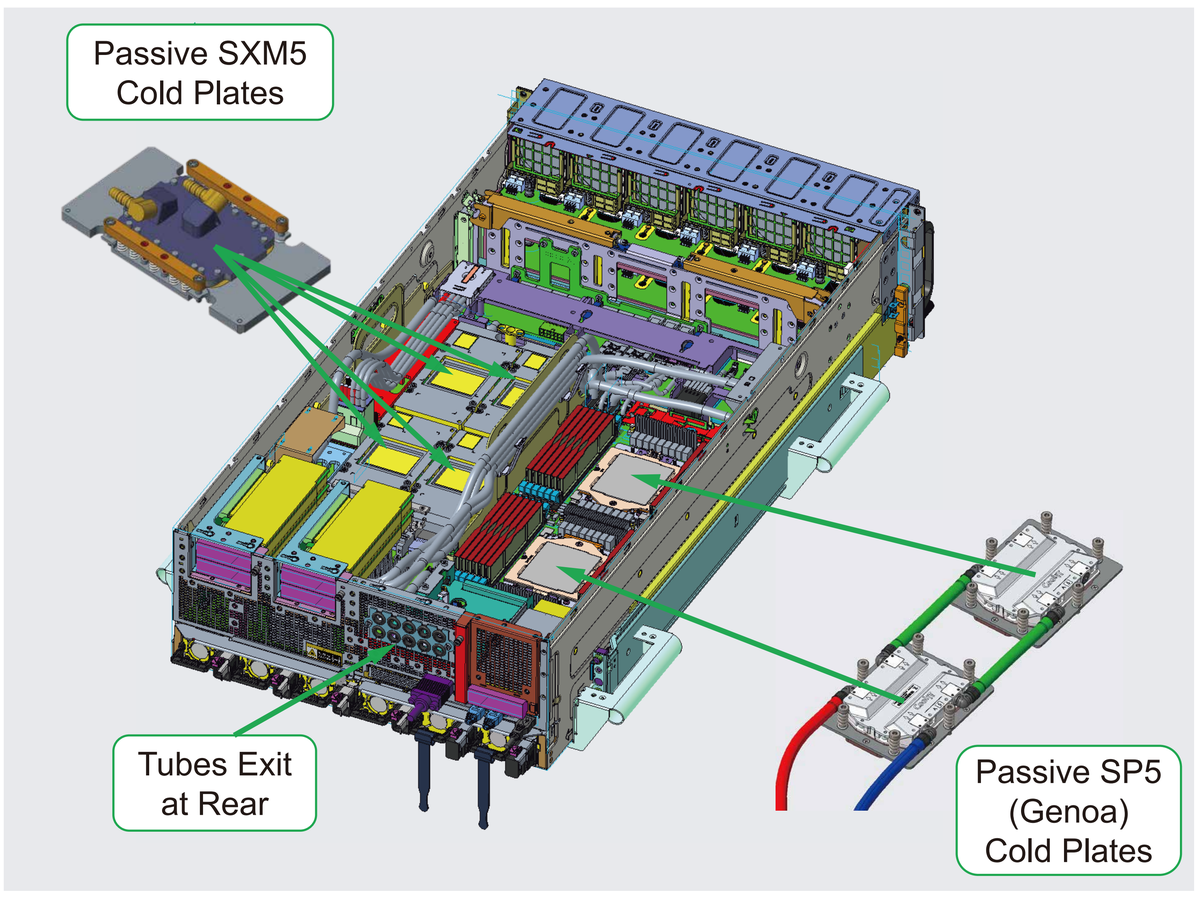
Mem: 768GiB (DDR5-4800 32GiB ECC Registered module ×24)

SSD: Samsung PM9A3 (1.92TB NVMe U.2)
sequential read 6.8GB/s, sequential write 2.7GB/s.

NW: InfiniBand NDR200 (200Gbps) ×4

Implementation and Cooling

CPUs and GPUs are directly water-cooled. Other components are cooled by air, which is generated at the rear door of the rack.
Inlet water temperature is expected to be ~20°C.



Block Diagram