

TPU and optical module



Overview

Google's new Ironwood TPU utilizes a 3D Torus network and Apollo optical circuit switch, addressing AI data center interconnect challenges.

Key Takeaways

Yes, Google TPUs use optical modules as part of their inter-chip communication network, specifically through optical circuit switches (OCS) for high-speed, low-latency data transfer.

Optical Modules in TPU Architecture

In TP

U pods, particularly TPUv4 and later generations, optical modules are used to connect TPU chips across cubes.

Each TP

U chip has multiple Inter-Chip Interconnect (ICI) links forming a 3D torus topology. While internal links within a cube rely on copper and PCB backplanes, the external links on the cube surfaces use optical connections that interface with OCS units for dynamic routing and large-scale scaling.

Role of Optical Circuit Switching

The optical circuit switches (OCS) act as a dynamic patch panel, creating direct optical paths between TPU chips. This design minimizes repeated optical-electrical-optical conversions, reducing latency and signal loss. For example, in a TPUv4 Pod, each cube has 96 optical links connected to OCS, and multiple OCS units are organized to handle traffic along the X, Y, and Z axes of the 3D torus.

MEM

S devices in the OCS physically redirect laser beams to establish these optical circuits.

Benefits of Optical Modules

Using optical modules in TPU pods provides several advantages: Low-latency communication: Signals remain in the optical domain for most inter-chip transfers, keeping latency under one microsecond. High bandwidth: Each ICI link can support terabits per second of bidirectional bandwidth, enabling efficient scaling across thousands of TPU chips. Fault tolerance: Optical links combined with ICI resiliency allow routing around faults without significant performance degradation.

Energy efficiency: Optical interconnects reduce the energy overhead compared to traditional electrical packet switching, contributing to TPU pods consuming 60–65% less energy than comparable GPU clusters.

Summary

In conclusion, TPUs do use optical modules, primarily for inter-cube communication in large TPU pods. These optical modules, integrated with optical circuit switches, enable high-speed, low-latency, and energy-efficient data transfer across thousands of TPU chips, which is critical for large-scale AI training and inference workloads.

Article Content

Google's TPU supply chain overview: Uncovering the "hidden ...

Optical modules and optical communication systems (data transmission "highways")
This is the core component of TPU cluster interconnection, with a single cluster requiring over 100,000

Google's High-Speed Interconnect Architecture to Push 800G+ Optical ...

Google's next-generation TPU, Ironwood, integrates a 3D Torus network topology with the Apollo optical circuit switch (OCS) all-optical network, marking a major step forward in AI data

Google TPU v8 Deep Dive: Precise Calculation of Optical, Copper,

This article will analyze in detail the Virgo Network architecture of TPU v8t, the Boardfly topology of TPU v8i, and the usage of optical transceivers, copper cables, PCBs, and OCS switches

Google's TPU v7: How 1.5 & 2.6 Optical Modules per Chi | BestHub

The article explains how Google's TPU v7 supercomputer uses a simple yet powerful networking scheme—1.5 optical modules per TPU for intra-rack communication and an additional 2.6

Google's Ironwood TPU and Apollo OCS to Drive 800G Optical

Google's new Ironwood TPU utilizes a 3D Torus network and Apollo optical circuit switch, addressing AI data center interconnect challenges. TrendForce forecasts that 800G+ optical

TPU7x (Ironwood) | Google Cloud Documentation

This page describes the architecture and available configurations for TPU7x, the latest TPU available on Google Cloud. TPU7x is the first release within the Ironwood family, Google Cloud's...

Unveiling Google's TPU Architecture: OCS Optical Circuit Switching ...

In this in-depth blog post, we dive deep into the evolution of Google's TPU intelligent computing clusters, focusing on the synergistic mechanisms of 3D Torus topologies and OCS

Google's Data Center Interconnect Architecture: Rise of 800G

Google is consolidating its proprietary TPUs, Ironwood racks, 3D Torus topology, and the Apollo OCS optical backbone into a unified high-speed interconnect architecture. As a result, the

Inside Google's Ironwood TPU v7 Supply Chain

This is a classic content-per-board story: every additional TPU node means more Celestica assembly, more VRM modules, and more high-current silicon on the board.

Tensor Processing Units (TPUs)

In this video, take a look at the components of the TPU system, including data center networking, optical circuit switches, water cooling systems, biometric security verification and more.

The Leading Player in 4-3-4 CPO: Google's Co-Packaged Optics

The Leading Player in 4-3-4 CPO: Google's Co-Packaged Optics Ambitions and Broadcom's Baily Platform Google pushes OCS & CPO tech for power saving in millions of TPU

Inside Google's Ironwood TPU v7 Supply Chain

Google Inc's Ironwood (TPU v7) is one of the least transparent but most important AI accelerators in production. The full bill of materials is not public, but if you stitch together Google

NVIDIA Strategic Investments in Optical Interconnects

Executive Summary Following newly announced investments, NVIDIA has committed over \$4 billion in direct strategic investments across two optical

Warning: The entire AI industry will likely be bottlenecked by two ...

And two vendors could freeze the global InP substrate market covering nearly all of: - Hyperscaler optics (TPU pods, etc) - Optical transceivers (5g, data) - LiDAR (robotaxis, drones,

Optical module design resources | TI

View the TI Optical module block diagram, product recommendations, reference designs and start designing.

ISCA 2023 Final Paper TPUv4

TPU v4 is the fifth Google domain specific architecture (DSA) and its third supercomputer for such ML models. Optical circuit switches (OCSes) dynamically reconfigure its interconnect topology to

[2304.01433] TPU v4: An Optically Reconfigurable Supercomputer for ...

In response to innovations in machine learning (ML) models, production workloads changed radically and rapidly. TPU v4 is the fifth Google domain specific architecture (DSA) and its

3 things to know about Ironwood, Google's latest TPU

General summary Ironwood our seventh-generation Tensor Processing Unit (TPU) is now available for Cloud customers. It's custom built for

Contact Us

Continue your research online:

Website: <https://bagyou.es>

Technical inquiry: <https://bagyou.es/project-enquiry.html>

This document is for preliminary research. Verify specifications against current standards, product documentation and project requirements.

