

BCM78919

102.4-Tb/s Multilayer CPO Switch with 200G SerDes

Overview

The Broadcom® BCM78919 102.4-Tb/s co-packaged optics switch is a high-radix, high-bandwidth network switching device supporting up to 64 × 1.6TbE, 128 × 800GbE, 256 × 400GbE, 512 × 200GbE, 512 × 100GbE, or 512 × 50GbE ports. The device family features a maximum of 64 integrated Condor 3-nm SerDes cores, each with eight integrated 212.5-Gb/s PAM4 SerDes transceivers and associated PCS.

The integrated optical engines enable direct high-speed I/O to the BCM78919 using 512 duplex single-mode optical fibers to the front panel in a variety of different port configurations. Each optical engine is directly driven from the Condor SerDes, providing the lowest-power and lowest-cost optical interconnect solution. Each duplex fiber is capable of supporting 200GbE, 100GbE, or 50GbE using a single wavelength at 1310 nm. Co-packaging with optics eliminates the need to source and manage high-speed pluggable transceivers in deployment, dramatically lowers system power consumption, and simplifies system design leading to improved network reliability.

The BCM78919 is a high-performance and high-capacity device designed to meet the requirements for next-generation data center, artificial intelligence (AI), machine learning (ML), and cloud computing environments. The BCM78919 architecture delivers complete L2 and L3 switching, routing, and tunneling capabilities at line rate and maximum port density, with low power and latency. Software compatibility is maintained across the StrataXGS® product portfolio to simplify customer designs and reduce customer time to market.

The BCM78919 has extensive features to address the rapidly increasing scale of data center network deployments and distributed computing applications. These include the following features:

- Large forwarding databases
- Extensive support for in-band telemetry
- Robust buffer performance including many-to-one burst absorption capabilities that assist in TCP Incast scenarios

With the BCM78919 device, customers can build data centers with much higher server node counts while simultaneously achieving the highest per-port power efficiency. The BCM78919 is built using state-of-the-art 3-nm silicon process and co-packaged optics silicon photonics technology.

Key Features

The BCM78919 has extensive features to address the rapidly increasing scale of data center network deployments and artificial intelligence/machine learning applications:

- 512 × 200G PAM4 all-optical I/O
- Flexible port configurations
- Low pin-to-pin latency in cut-through and store-and-forward modes
- Enhanced load balancing
- Network monitoring and congestion detection

Applications

- AI-connected fabrics
- Front-end and back-end networks of large-scale generative AI compute clusters
- Data center spine, leaf, ToR, blade, and aggregation switching
- Cloud computing and ML applications
- Service provider core switching

Data Center Features

- Hardware-based encapsulation
- Tail-latency improvements
- 512 port radix
- Low-latency OBM queue added for high-priority traffic
- Per-virtual machine traffic shaping
- Support for VXLAN overlays

L2 and L3 Packet Processing

- Full IPv4 and IPv6 routing support
- Hardware-based encapsulations, including Multiprotocol Label Switching (MPLS), generic routing encapsulation (GRE)
- Unified forwarding table for flexible allocation of L3 host, longest prefix match (LPM) entry, IFP key compression entry, and IFP Exact Match entry
- Flexible ingress and egress counter pools

Buffering and Traffic Management

- Full support for RDMA over Converged Ethernet v2 (RoCEv2)
- Full Quality of Service (QoS) support, which includes the following features:
 - Priority flow control (PFC)
 - Weighted random early discard (WRED)
 - Single-rate Three Color Marking (srTCM) and two-rate Three Color Marking (trTCM) for color marking and metering
- Packetized MMU statistics
- Network monitoring
- Network congestion detection
- sFlow redirect
- Enhanced load balancing

- Enhanced trunk hashing capabilities: Versatile Hash (VH), RTAG7, symmetric hash, flex hash, and resilient hash
- Support for jumbo frames up to 9416 bytes
- Support for 12 queues per port and the following scheduling algorithms: strict priority (SP), round-robin (RR), weighted round-robin (WRR), and weighted deficit round-robin (WDRR)

Cloud-Optimized PP and MMU for AI Networks

- Updated congestion management features: FastCNP, PFC-aware DLB, DCN, and PFC-aware ECN
- 267-MB unified packet buffer
- 324B streaming frame size (SFS)

Additional Features

- Segment routing V6 initiation, transit, and termination for SID or GSID
- Advanced statistics support, which includes the following features:
 - Enhanced stateful accounting capabilities (flex state)
 - Added significant number of dedicated counters, including Port TX, Port RX, VFP, IFP, and EFP
- Comprehensive time synchronization support, which includes the following features:
 - Integrated IEEE 1588v2 processor
 - IEEE 802.1AS support
 - Synchronous Ethernet (SyncE)
 - Ingress and egress per packet time stamping
- sFlow support including ingress, egress, and flex sampler, with an option to forward truncated sFlow packets to a remote agent
- Four dedicated 50G auxiliary Ethernet ports that can be used for management
- In-band flow analyzer record flow across networks
- x4 PCI Express (PCIe) Gen5 interface to support a local CPU
- Egress (EP) recirculate
- Stateful Mirror on Drop (MoD)
- Postcard mirror
- True egress mirror
- Adjustable voltage scaling (AVS) for reduced average and peak power
- Multi-die design for lower heat density and enabling air cooling
- Support for network load balancing and congestion control:
 - Generalized congestion notification message (CNM)
 - Generalized packet drop notification (PDN)
 - Generalized trim flexible trimming includes Broadcom-proprietary DCN
 - Cognitive routing
 - Dynamic load balancing for equal cost multipath (ECMP) groups
 - Accelerates AI workloads with cognitive routing and advanced telemetry
 - Global load balancing (GLB)
 - Reactive path rebalancing
 - Fast link failover
 - Drop congestion notification (DCN) packet trimming, which enables the highest performance when using edge-queued datagram service (EQDS)
 - L2 and L3 instrumentation (in-band telemetry using fixed-size metadata)