MICROPROCESSOR *report*

Insightful Analysis of Processor Technology

JERICHO2C+ BRINGS 7NM TO ROUTERS

Broadcom Elevates Router Coopetition for Comms and Cloud Providers

By Bob Wheeler (September 14, 2020)

....

This month, Broadcom sampled its first 7nm router chip, the Jericho2c+ packet processor. The new line-card chip is compatible with the shipping 16nm Ramon fabric device, introduced with Jericho2 in 2018 (see *MPR 3/12/18,* "Broadcom Router Chips Aim at ASICs"). In 2019, the company unveiled Jericho2c with half of Jericho2's bandwidth, as Figure 1 shows, reducing power and cost for many service-provider designs. Jericho2c+ offers an upgrade path for both 16nm packet processors, reducing power per gigabit while increasing integration. It can also serve in data-center-interconnect (DCI) systems that link one data center to another. By integrating MACSec security, Jericho2c+ eliminates external MACSec PHYs from DCI systems.

The world of network-router chips makes for strange bedfellows. The leading network-equipment vendors develop in-house ASICs for their top-end routers, which target service-provider-core networks. Many also use merchant chips, however, in their higher-volume routers for metro aggregation, peering, and the edge. Broadcom offers the only merchant chips purpose-built for modular systems, by way of its decade-old Dune Networks acquisition.

In 2014, however, two Dune founders started Leaba Semiconductor, and Cisco acquired the stealth-mode startup in 2016 for $320 million. Leaba's design shipped in late 2019 as the Cisco Silicon One Q100, which serves in new 8000-Series routers. More startling was Cisco's newfound flexible business model, opening the possibility of selling chips rather than complete systems. At the same time, the company's NCS 5500 routers continue to use Broadcom's Jericho line, as do an increasing number of white-box systems.

Merchant vendors and OEMs have different development priorities, which are evident when comparing the Broadcom and Cisco approaches to routing silicon. The biggest difference lies in the tradeoff between optimizing silicon-area efficiency and developing multiple chips. Cisco's Silicon One name connotes one chip architecture spanning multiple market segments. By contrast, Broadcom offers three product lines to collectively handle data-center, enterprise, and service-provider applications.

Jericho2 Earns C+ in Power

Architecturally, Jericho2c+ is a straightforward evolution of Jericho2. Both generations employ 50Gbps PAM4 serdes, handling 400G Ethernet on the front panel and handling backplane connections to the Ramon cell-based fabric. The
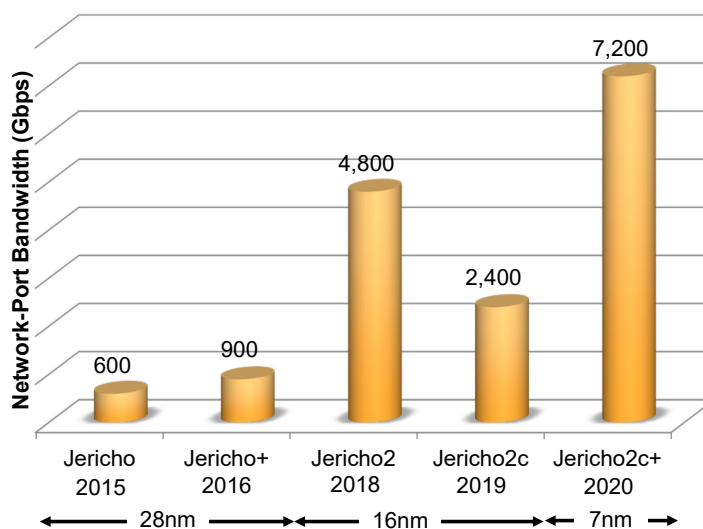
Figure 1. Jericho packet-processor evolution. Jericho2 adopted HBM2, enabling a giant bandwidth leap, and Jericho2c+ further increases bandwidth by 50% compared with Jericho2.

new version grows to 144 front-panel serdes and 192 fabric-interface serdes, as Figure 2 shows, yielding 7.2Tbps and 9.6Tbps respective bandwidths. The fabric-interface over-speed precludes any blocking as packets traverse the fabric to the egress port on another line card. The packet processor supplies up to 18x400GbE and 72x100GbE ports, and the 100GbE ports also allow lesser 10/25/50GbE rates.
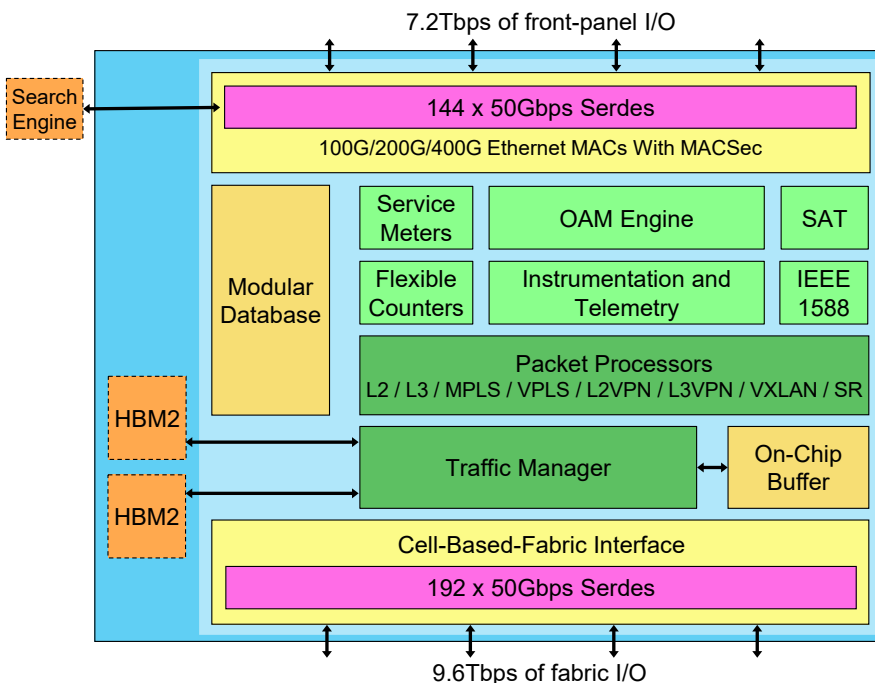
The primary new feature in Jericho2c+ is line-rate encryption for all network ports. Although principally designed for MACSec link encryption, the chip's packet processor is flexible enough to handle IPSec end-to-end encryption as well. Data-center operators demand the former, whereas communications-service providers may require the latter for their 5G networks. The move to 7nm also allowed Broadcom to expand on-chip tables, with the chip supporting up to four million IPv4 routes, 256K subscribers, 384K counters, and 256K meters. Like its predecessor, Jericho2c+ enables forwarding-table expansion using a Broadcom search engine (KBP) connected through an Interlaken Look-Aside interface. The optional KBP interface consumes 16 of the front-panel serdes, reducing maximum network-port bandwidth to 6.4Tbps.

The 7nm packet processor uses the same microcode-programmable ingress and egress pipelines as Jericho2. The 16nm chip has two pipelines, and we expect Jericho2c+ simply clocks them 50% faster. Although the pipelines have a fixed set of stages, Broadcom's intriguing Elastic Pipe carries over from Jericho2; it allows multiple general-purpose ALU stages to be inserted anywhere in the pipeline for feature extensions.

Like Jericho2, the new chip stores packets in an on-chip buffer (OCB), which doubles in size to a substantial 64MB. The OCB is backed by two HBM2 stacks, which deliver an 8GB deep buffer operating at 2.4GT/s for 614GB/s of memory bandwidth. The Jericho/Ramon architecture buffers packets on the ingress line card until the egress port is available. Uncongested flows remain in the OCB, so only congested flows move to HBM2. During short-lived congestion, the chip can use all 4.8Tbps of HBM2 bandwidth to store congested flows.

Following customer disclosures, we now know the 16nm Jericho2 pushes air-cooling limits by consuming nearly 400W (maximum). Jericho2c reduced power and cost by implementing a single pipeline and HBM2 stack (see *MPR 3/18/19,* "Broadcom Fills Out 5G-Network Line"). Broadcom says Jericho2c+ cuts power per bit by 50% compared with Jericho2c, but with three times the port bandwidth, it remains power hungry. For a 36x400GbE DCI line card requiring MACSec, two Jericho2c+ chips replace three Jericho2 chips as well as 18 PHYs with MACSec. The company pegs the absolute power reduction for such a design at 50%, from which we estimate Jericho2c+ dissipates about 350W (maximum), or 11% less than Jericho2.

Jericho2c+ (BCM88850) is Broadcom's third 7nm switch chip, with Trident 4 being the first and already in production (see *MPR 6/24/19,* "Broadcom Samples Trident 4 Switch"). Because they share the same 50Gbps PAM4 serdes, that function is well proven. Likewise, Jericho2 is shipping with HBM2, although the interface comes in 16nm technology. As a result, the company expects to qualify Jericho2c+ for production by year-end, leading to system availability in 1H21.

Service Providers Embrace White Boxes

Although leading OEMs remain committed to developing ASICs, end customers are increasingly embracing white-box systems built on merchant silicon. This trend started with hyperscale-data-center operators, but communications-service providers are following suit. Although not strictly a prerequisite, white-box adoption has followed network virtualization. In both cases, AT&T was a trailblazer, but many other providers have pursued similar plans. Recent provider announcements herald fully virtualized 5G networks, in particular.

To create blueprints for carrier-network systems, the Open Compute Project (OCP) in 2016 formed a Telco Project working group, and it has since approved nearly 20 designs contributed by seven companies. AT&T is the most active carrier, and among its early submissions is a cell-site gateway based on Broadcom's Qumran AX, which derives from Jericho+. For 100GbE aggregation, ODM Edgecore

Figure 2. Jericho2c+ block diagram. SAT=Service Activation Testing (MEF 48). The chip packs a total of 336 serdes, and 8GB of in-package High Bandwidth Memory (HBM2) enables deep buffers.

contributed a Jericho2-based system with up to 80 network ports.

Last year, working with Broadcom, ODM UfiSpace, and network-operating-system (NOS) vendor DriveNets, AT&T developed the Distributed Disaggregated Chassis (DDC) routing system based on Jericho2 and Ramon. The DDC essentially takes a chassis-based router and disaggregates it into "line-card" boxes (NCPs) and fabric boxes (NCFs). Instead of a backplane, the boxes connect using copper or optical cabling. Because Jericho2c+ works with Ramon, customers can upgrade deployed DDC systems with a new NCP3 box employing that chip. An NCP3 has 36x400GbE network ports and 40x400Gbps fabric ports in a 2U form factor, enabling customers to build what Broadcom touts as the world's biggest router with 18,432x400GbE ports (7.4Pbps). A more realistic three-rack DDC example offers 864x400GbE ports.

Service providers can specify and acquire systems directly from ODMs and OEMs, so OCP submissions don't capture all white-box deployments. Still, every accepted Telco Project router and access system contains Jericho/Qumran products. The other chip vendor that dominates these designs is Intel, as they all employ Xeon D control-plane processors.

Cisco's Shifting Sands

Although Leaba never emerged from stealth, its vision appears in Silicon One. Apparently, its bold goal was to develop a single chip architecture that could span multiple end markets. In addition to one chip serving multiple router functions, the startup also intended its design for data-center switching. Given its desire to serve multiple markets, it identified programmability as a baseline requirement.

After initially developing C-like code, the startup found the P4 network programming language was a better fit for its target applications. P4's match+action semantics were maturing around the time Cisco acquired Leaba (see *MPR 8/8/16,* "Barefoot's Tofino Gives P4 a Test Spin"). The Q100 follows a run-to-completion model, however, whereas most P4 implementations have pipelines with a fixed set of match+action stages. The chip has one ingress stage, a shared-memory traffic manager, and one egress stage. Cisco withheld other architecture details, such as how many threads the stages handle.

The company rates the Q100 at 10.8Tbps and more than six billion packets per second (Bpps) while handling an advanced feature set comparable to that of Jericho2. The chip uses HBM for both forwarding-table expansion and deep packet buffers. It has 50Gbps PAM4 serdes, needing at least 216 to match its rated throughput. Cisco highlights the 415W (typical) power efficiency (excluding optics) of the Q100-based 8200, a fixed-configuration 10.8Tbps router; we estimate the chip dissipates about 230W (typical) in this design. The company withholds maximum system power, however, which could be up to twice the typical rating.

The 8800 chassis router uses the Q100 on line cards as well as fabric cards. The 36x400GbE line card has four Q100 chips, meaning each handles only 3.6Tbps of network-port bandwidth. Although the Q100 could handle 4.8Tbps on the line side, Cisco chose to handle smaller packets at line rate as well as to increase the ratio of HBM2-to-network bandwidth for buffering.

Many traditional router customers will purchase the 8000-Series complete with Cisco's IOS XR7 NOS, but the company is decoupling its software and hardware for customers that prefer another NOS. For the first time, Cisco is enabling customers to install the Sonic open-source NOS on its router, providing an alternative to ODM-supplied white boxes. Its willingness to sell Silicon One chips appears to indicate a focus on hyperscale-data-center operators, with which it currently enjoys little business.

Broadcom Raises 7nm Chip Ante

Comparing the Q100 with Jericho2c+ is unfair, as the former is about two years ahead in reaching the market. Instead, the Q100 lines up with Broadcom's 16nm generation, which is also shipping. In a line-card design, we estimate Jericho2 and the Q100 provide similar power efficiency. On the other hand, Broadcom's 9.6Tbps Ramon fabric chip dissipates only 170W (maximum), which we estimate is significantly less power per gigabit relative to the Q100 acting in its fabric role. Overall, a chassis router employing the 16nm Broadcom chipset should have the power-efficiency edge.

Performance is more difficult to compare, given the limited vendor disclosures. Cisco is using Jericho2 in new NCS 5700 line cards, and it rates the chip at 2Bpps, or about two-thirds the rating of the Q100 in a line card. Broadcom's pipelined design is deterministic, however, whereas the Q100's packet rate will vary on the basis of traffic patterns and features. IOS XR7 supports both NCS 5500 and 8000-Series routers, but the latter's initial features are a subset of the former's. Although the Q100 may close that gap in future releases, Cisco hasn't revealed any planned features that exceed Jericho2's capabilities.

Cisco's foray into merchant silicon remains a wild-card, and it could reduce Broadcom's sales into DCI systems at hyperscale operators. The Q100 stands out in fixed-configuration (pizza-box) systems with 10.8Tbps of network-port bandwidth, which require two Jericho2 chips placed back to back. Broadcom should have lower manu-

facturing cost thanks to unit volumes from multiple customers combined with the lower wafer pricing associated with being a top foundry patron. The Jericho roadmap also demonstrates a record unmatched by Cisco, a fledgling merchant-chip supplier. Still, Cisco's embrace of flexible business models could unsettle the incumbent merchant-chip and ODM white-box supply chains, if it aggressively pursues these opportunities.

By sampling Jericho2c+, Broadcom demonstrated its ongoing commitment to three switch-chip lines, with Trident 4 and Tomahawk 4 preceding that device to 7nm. As the incumbent merchant line, Jericho faces new competition from Cisco, but we expect it will be mostly indirect competition through white-box sales. Competing with your customers can be uncomfortable, but so is competing with your suppliers. Meanwhile, Broadcom entered AT&T's Internet Protocol (IP) backbone with the DDC, broadening its footprint. Jericho2c+ builds on that success with improved efficiency and security, undermining OEM claims of ASIC superiority. ♦

To subscribe to *Microprocessor Report,* access *www.linleygroup.com/mpr* or phone us at 408-270-3772.