

# MICROPROCESSOR *report*

Insightful Analysis of Processor Technology

## CEVA TELEGRAPHS 5G INTENT WITH XC12

*High-Performance DSP Targets Wireless Infrastructure*

By David Kanter (April 3, 2017)

The move to 5G and the arrival of myriad new networked devices will increase demand for bandwidth. To meet this demand, infrastructure will employ more antennas and higher bit rates, which all require greater computational throughput. At the same time, latency requirements will tighten to enable heterogeneous networks, such as a combination of licensed and unlicensed spectrum as well as responsive communication for automation.

To tackle these challenges, the new Ceva XC12 licensable DSP core takes a combined approach. For latency-sensitive tasks, such as coordinating multiple networks, it uses the scalar unit from its X4 DSP, which debuted last year (see [MPR 3/7/16](#), “Ceva’s New Gen-X DSPs Target 5G”). The XC12 pairs this unit with high-throughput vector units to handle the data plane. The vector units incorporate 128 multiply-accumulate (MAC) units that can perform 16-bit x 16-bit operations. The company supports floating-point data and uses a proprietary high-precision data format to increase the signal quality. Additionally, the XC12 includes a variety of new instructions that accelerate wireless algorithms.

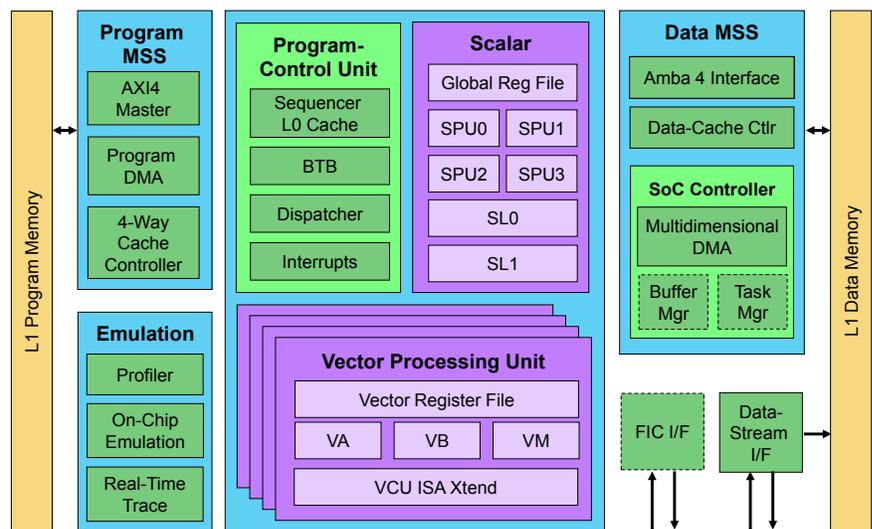
Historically, Ceva DSPs have been strongest in client devices (e.g., smartphone modems), but the company has also won a design in at least one base-station ASIC, from Ericsson. The XC12’s lineage traces back to the earlier XC4410 (see [MPR 3/5/12](#), “Ceva Exposes DSP Six Pack”). Both devices offer 128 MAC units, twice as many as the XC4500 (see [MPR 11/4/13](#), “Ceva Targets Wireless Infrastructure”). But infrastructure

customers opted for the smaller XC4210, so the XC4410 never entered production.

The XC12 supersedes the XC4500 by delivering a 2–8x performance boost for important 5G tasks such as channel estimation, making it attractive for new wireless base stations and other infrastructure. Ceva is wagering that the upcoming transition to 5G will spur demand for the XC12’s greater performance. Because the new DSP has already chalked up a base-station win and a client-device win for 2018, this expectation seems reasonable.

### XC12 Builds on X4

Ceva DSPs are programmable, and some can run an operating system, but the underlying hardware differs a bit



**Figure 1. Block diagram of Ceva-XC12.** The DSP couples the X4 scalar and program-control units with a vector processing unit that derives from the XC4410.

### Price and Availability

RTL for the Ceva-XC12 DSP is now available, but the company does not disclose pricing. For more information, visit [www.ceva-dsp.com/CEVA-XC12](http://www.ceva-dsp.com/CEVA-XC12).

from that of a general-purpose CPU. The company focuses tightly on cellular communications, where the workload is well understood and ripe for acceleration using dedicated hardware. An XC12 core can operate alone or in multicore clusters to serve in 5G base stations from femtocells to macrocells and even cloud radio access networks (see [MPR 2/16/15](#), “FPGAs Target Remote Radio Heads”).

The XC12 architecture has a 14-stage pipeline that executes VLIW bundles containing up to eight instructions. As Figure 1 illustrates, it contains a mix of scalar units for low-latency execution and vector units for massive throughput. To boost throughput and reduce latency for 5G services, the company enhanced nearly every portion of the DSP relative to the prior generation. The front-end and scalar units are nearly identical to those of the X4 family, but many of the vector-unit enhancements are new.

The program-memory controller fetches bundles from the instruction memory. Each bundle contains up to eight instructions that are dispatched to the scalar and vector execution units. Each instruction can be 16–96 bits long, and a bundle contains up to 256 bits. Designers can configure the L1 program memory to hold 0–256KB in an explicitly addressed SRAM, which is often called a tightly coupled memory (TCM). For conventional-OS support, the program-memory controller can include a 0–128KB L1 instruction cache. Compared with prior generations, this cache doubles the associativity to four ways in the XC12. The core also has an L0 instruction cache, which is a four-cache-line fetch buffer that’s nonblocking and has automatic next-line prefetching.

The program-memory controller reduces latency through hardware and software prefetching and contains a branch target buffer (BTB) to reduce branch latency. As in the X4, the BTB is two- or four-way set associative with 64–256 entries. Each BTB entry contains the most recent target address and a 2-bit branch-history counter.

The scalar unit includes four scalar processing units (SPUs) and an orthogonal 64-entry register file that’s 32 bits wide. The SPUs can each perform two 16x16-bit MACs or a single 32x32-bit MAC per cycle. All have optional floating-point units, although that option increases die area and power.

The scalar unit also has two independent load/store pipelines (LS0 and LS1). Each pipeline can calculate addresses on behalf of the scalar or vector units and perform a load or a store. Ceva claims the XC12 scalar unit achieves 4.4 CoreMarks per megahertz, similar to ARM’s Cortex-R7. That score, however, is 10% higher than the

4.0 rating of the X4 and likely reflects compiler optimizations rather than microarchitectural changes (see [MPR 7/22/13](#), “Compilers Catapult CoreMark Scores”).

### Aiming for Higher Standards

The vector processing units, which provide most of the XC12’s computational power, are thoroughly enhanced to deliver greater raw throughput and better performance on major 5G algorithms. The DSP supports 16- and 32-bit integers, two proprietary floating-point (FP) formats, and (with the optional FPU) single- and half-precision FP.

The previous generation uses a proprietary 40-bit pseudo-FP format that derives from a 32-bit integer format. It has a 32-bit mantissa and 8-bit exponent for multiplication, but it has a 40-bit mantissa for addition. The pseudo-FP capability allowed Ceva to achieve the greater precision and dynamic range of FP with the throughput of integer computations. The XC12 extends this concept to a 20-bit pseudo-FP format that is based on 16-bit integers and enables additional precision. Additionally, because the Ceva pseudo-FP approach doesn’t check all IEEE 754 error and exception conditions (e.g., division by zero), the application code must mitigate or avoid these problems.

Feeding each VPU is a set of sixteen 320-bit registers, so the XC12 has a total of 64 vector registers. The VPUs contain up to four execution-unit types. The first (VA) is a MAC unit—a staple of DSP algorithms. The XC12 doubles the throughput of the prior generation and can sustain 128 MACs using 20-bit pseudo-FP data. The additional throughput supports larger matrices (e.g., 256x256) for 5G signal processing, boosting performance for interference-rejection combining. The company claims 4x greater performance; half of this gain is from doubling the number of MAC units relative to the XC4500, and the other half is from implementing shift, saturation, and other dynamic post-processing in line rather than using separate instructions.

The second execution unit (VB) performs time-consuming operations such as division, square root, inverse square root, and trigonometric operations. The previous generation approximated these nonlinear operations using lookup tables; the new native implementation is more accurate. The third execution unit (VM) handles bit-stream instructions such as de-interleaving.

Ceva’s instruction set is extensible to allow custom hardware acceleration of wireless algorithms (see [MPR 3/4/13](#), “Ceva Extends XC DSPs”). The company has added a number of specialized instructions (e.g., for QAM demodulation), but it has withheld details. Customers can instantiate their own instructions, boosting performance and efficiency on specific algorithms. The fourth (optional) execution unit handles these custom instructions.

### Flexible Memory Serves Many Masters

The data-memory subsystem accesses memory on behalf of all requesting devices in the XC12. It handles accesses that

originate from the scalar unit's load/store pipelines, such as loading a 320-bit vector from memory, as well as requests from hardware accelerators.

The L1 data memory is a TCM that's configurable from 256KB to 2MB. A sophisticated DMA controller accesses the data TCM. An optional L1 data cache (0–64KB) is also available. Most memory systems focus on linear DMA commands, which are typically described as an address and a data length (e.g., read 4KB starting at address 0x4). Since 5G supports multidimensional beam forming, the XC12's DMA controller can fetch a 2D or 3D data structure as well. Accesses that come from accelerators go into management queues, and data to or from the accelerators goes into special buffers, reducing the hardware-offloading overhead.

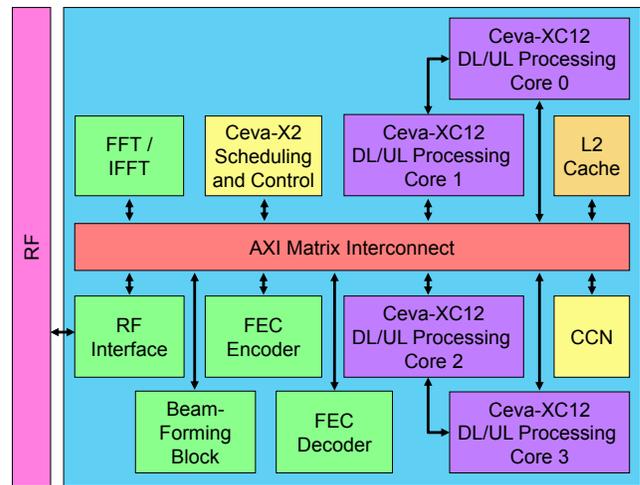
Because minimizing latency is crucial and the requesting clients are heterogeneous, the memory subsystem can arbitrate among different requesting devices (e.g., scalar units, vector units, and dedicated hardware) and ensure quality-of-service guarantees. It also has a proprietary 512-bit-wide fast interconnect (FIC) for multicore implementations. The FIC includes both master and slave ports to connect to other XC12 instances.

### Paving the Way for 5G

Ceva expects the XC12 to reach 1.8GHz in a 10nm FinFET process; at this speed, the core generates a maximum of 490 billion 16-bit integer calculations per second. The company claims the XC12 reduces power by 50% compared with the XC4500 when running an LTE baseband workload with both control and DSP code. For this comparison, it configured both cores with a 256KB data TCM, 32KB data cache, 64KB program TCM, and 32KB instruction cache. Both use the same TSMC 16nm FinFET process, and the XC12 is twice the size of the XC4500. Ceva declined to specify the XC12's absolute active and idle power consumption.

The 50% power reduction applies only to a fixed-size radio workload moved from the XC4500. In practice, Ceva's target infrastructure customers will likely use the XC12 to generate greater performance at the same power, rather than reduce power. This comparison indicates a theoretical 2x performance improvement at the same power, plus additional gains if the design moves from 16nm to 10nm. These gains, however, come at the cost of greater die area.

The XC12 is more compelling for new 5G reference architectures than existing LTE infrastructure. For 5G systems, the XC12's additional performance can reduce system cost. Figure 2 illustrates a Ceva reference design targeting a 5G base station that can handle 128x8 MIMO, QAM-256, and a 125-microsecond subframe over a 160MHz TDD



**Figure 2. Ceva reference design for wireless processing.** DL/UL=download/upload. The clustered architecture uses a variety of the company's hardware accelerators and programmable DSPs.

frequency range. The design can receive and transmit one 160MHz component carrier at an 8Gbps data rate using four XC12s. The FIC links between XC12 pair enable the cores to work on the same task while sharing data. The reference design also includes a variety of system-level accelerators outside of the XC12. For example, the beam-forming and forward-error-correction (FEC) accelerators all sit on the system fabric; Ceva supplies many of these blocks.

The company estimates that implementing the same 5G base station using the older XC4500 would require 16 cores. Although the XC12 is twice the size of the XC4500, it delivers much greater performance, reducing the core count. Ceva claims that using the XC12 reduces the compute-cluster size by 40% and power consumption by 60%, validating the benefits of a higher-performance DSP.

In Ceva's case, persistence could pay off. Initially, customers were indifferent to the company's high-performance XC4410, which never reached production. The XC12 is a spiritual successor to that product, but this time, 5G has ignited interest in high-performance DSPs, helping Ceva win customers for both client devices and infrastructure. The company has licensed the XC12 to an OEM for a 5G modem in a client that's scheduled for deployment at the February 2018 Olympics in South Korea. Additionally, another unnamed OEM has licensed the DSP for 5G base stations. Volume deployments of 5G base stations will likely start in 2019, but Ceva has a clear stake in the future, and the XC12 could ride the 5G wave to new customers and markets. ♦

To subscribe to *Microprocessor Report*, access [www.linleygroup.com/mpr](http://www.linleygroup.com/mpr) or phone us at 408-270-3772.