2.1 SkyLake-SP: A 14nm 28-Core Xeon® Processor

Simon M. Tam, Harry Muljono, Min Huang, Sitaraman Iyer, Kalapi Royneogi, Nagmohan Satti, Rizwan Qureshi, Wei Chen, Tom Wang, Hubert Hsieh, Sujal Vora, Eddie Wang

Intel, Santa Clara, CA

SkyLake-SP (Scalable Performance), code name SKX, is the next generation Xeon® server processor fabricated on the Intel® 14nm tri-gate CMOS technology with 11-metal layers [1,2]. The SKX processor family has three core-count configurations. Each SKX core is accompanied by 1MB of dedicated L2 (2nd level cache) and 1.375MB of non-exclusive L3 (3rd level cache). At its maximum configuration of 28 cores, the SKX processor supports 6 DDR4 channels (2666MT/s), 3×20-lanes UPI processor-to-processor links (10.4GT/s) and x48+4 PCIE links (8GT/s). SKX supports per-core power-performance optimization enabled by on-die integrated voltage regulators (FIVR) [3, 4]. A new 2-dimensional synchronous on-die MESH fabric interconnects all the on-die components. Fig. 2.1.1 shows the overall architecture of the SKX processor.

Comparing to previous generation Xeon® server processors, SKX's higher maximum core-count, increased frequency, and improved IPC provided the generational performance improvements across all relevant server benchmarks. Aggressive core dynamic capacitance (Cdyn) reduction and frequency push were exercised to achieve the power and frequency targets. To facilitate high-volume production and silicon debug, SKX incorporated extensive DFT (design for test) features with high SCAN coverage and analog debug capabilities for critical analog circuits.

SKX deploys a highly configurable floorplan architecture to accommodate multiple products/sockets requirements. Key integration components are: (1) CORE-TILE and (2) on-die-fabric (MESH). The CORE-TILE unit integrates the core, LLC, and the core-to-MESH agent into one readily array-able modular object. The MESH is a 2D synchronous fabric architected to scale with core-tile count at higher data bandwidth and reduced latency relative to its RING predecessor [3] at lower frequencies. The 28-core SKX floorplan started from a 5-by-6 CORE-TILE array. Two CORE-TILEs, one each from the left and right columns were replaced with MEMORY CONTROLLER modules. At the top of the die, the "NORTHCAP" contains the IO agents, serial-IP ports, the clock generator unit (CGU), global power management and the fuse unit. Adding global chassis structures, the on-die-voltage-regulators (FIVR) and the DDR-IO complete the full-chip assembly. Fig. 2.1.2 shows the evolution of the 28-core floorplan.

Figure 2.1.3 shows the 9 primary VCC domains of the 28-core SKX processor physically partitioned into 35 VCC planes. Multiple FIVRs in conjunction with 5 mother board (MB) voltage regulators (MBVRs) serve these VCC partitions. FIVR vs. MBVR assignment was chosen based on die area, current compliance, VCC noise specifications and the power delivery efficiency attributed to MB and package IR loss. FIVR and a dedicated all-digital PLL for each core enables per-core voltage-frequency optimization to achieve the lowest "average" core power. Core-droop mitigation based on specific workload was added to help mitigate the core voltage droop. For the 28-core SKX, two FIVRs supply the un-CORE VCC (VCCCLM) serving the LLC and non-CORE components in the CORE-TILE. The entire un-CORE belongs to one clock domain and is served by a single ADPLL. LC filtered VCCs with the L realized on the package are also available for critical analog circuits (e.g. clock circuits in the high-speed IOs). Clock distribution schemes similar to that in [2] are adopted for the SKX processor. A new block to detect and enable throttling the peak current of FIVR input supply was added in SKX. This helped reduce the amount of VR caps needed to support current surge on that supply.

The MESH forms the high-bandwidth on-die 2D global synchronous fabric interconnecting the AGENTs with the COREs and CACHEs (Fig. 2.1.4). The MESH eliminated the RING-to-RING-bridge logic from the previous RING architecture to provide a low latency 2D cross-point interconnecting every CORE-TILE with its 4 neighbors. It adopts a simple data routing algorithm comprising data being first routed in the vertical direction and then routed horizontally to the destination. This dataflow control, however, makes the vertical MESH latency the most critical factor in determining the overall MESH performance. At low VCCCLM for low un-CORE power, it is difficult to achieve single-cycle vertical MESH latency above ~2GHz due to the vertical TILE-to-TILE RC delay. The solution is to move the critical section of the V-MESH from the VCCCLM domain to a higher fixed voltage VCCIO supply. This technique enabled single-cycle vertical TILE to TILE latency without needing to raise the un-core VCC thus reducing the overall power.

The SKX has a performance enhanced and customized core. Specifically, the SKX core has two AVX (advanced vector extensions) processing units with AVX-512, a larger L2 cache (1MB, 1024 sets by 16 ways) achieving 2× the floating-point operation performance and higher cache bandwidth vs. its predecessor [3]. Additionally, each CORE-TILE has a 2048 sets by 11 ways L3, and a 2048 sets by 12-ways snoop filter (SF) to support the size of the L2. The SKX caches incorporated column and/or row repairs to achieve high manufacturability. Significant FUSE resources were allocated to provide a statistically robust cache repair capability to achieve low VCC in a 28-CORE-TILEs SKX embedded with 28MB L2 and 38.5MB L3. Silicon data showed the minimum VCC exceeding the product requirements. Fig. 2.1.4 shows the CORE-TILE containing the CORE, AVX, L2, LLC, and the CACHE-HOME AGENT (CHA) that interface the cache to the MESH.

The chip features 128 lanes of high-speed IOs including 48 PCIE, 4 DMI, 16 on-package PCIE lanes running at 2.5/5.0/8.0GT/s and 60 UPI lanes running at speeds up to 10.4GT/s. The TX architecture and circuits are based on the design reported in [5]. The RX architecture is an evolution of [5] and includes support for the PCIE Separate Refclk Independent SSC (SRIS) ECN. To achieve the higher speed of 10.4GT/s with minimal power impact relative to prior generations and to enable faster layout convergence, the RX front-end was re-architected to eliminate the variable gain amplifier (VGA) functionality from the CTLE. Instead, a front-end attenuator was introduced before the CTLE. The CTLE circuit topology was carefully constructed to completely avoid the need for a precision resistor. PVT variation was addressed by limiting the number of stages in the CTLE to 2 and by making key performance parameters ratios of device parameters. The topology of the CTLE is shown in Fig. 2.1.5. Simulations show higher than 10.8dB of AC peaking at Nyquist rate of 5.2GHz. The overall transceiver occupies 20% less area and 17% less power, (simulated) compared to the prior generation [3].

SKX has 6-channel DDR4 interfaces each capable of supporting 2-DIMM per channel and speed up to 2666MT/s achieving 128GB/s total memory bandwidth. The 6-channel interface is split into two independent and identical physical sections (3 channels each) residing to the left and right edges of the die. Data bytes are located on north and south sub-sections of the channel layout, while Command, Control, Clock signals and PVT compensation circuitry are located toward the middle section. This floorplan facilitates package routing escapes and pin-out order matching between the CPU and DIMM card resulting in shortened package and board routing length to improve signal integrity. Fig. 2.1.6 shows the SKX DDR4 receiver (RX) Architecture.

The SKX server processor fabricated on a 14nm CMOS process is fully functional and is performing to its specifications. Fig. 2.1.7 shows the die photograph of the 28-core SkyLake-SP processor.

References:

[1] S. Natarajan, et al., "A 14nm Logic Technology Featuring 2nd-Generation Finfet, Air-Gapped Interconnects, Self-Aligned Double Patterning And A 0.0588 μm² SRAM Cell Size," *IEDM*, pp. 3.7.1-3.7.3, 2014.

[2] E. Fayneh, et al., "14nm 6th-Generation Core Processor SoC with Low Power Consumption and Improved Performance," *ISSCC*, pp.72-73, 2016.

[3] B. Bowhill, et al., "The Xeon® Processor E5-2600 v3: A 22nm 18-Core Product Family," *ISSCC*, pp. 78-79, 2015.

[4] A. Nalamalpu, "Design Optimization of Computing Systems - from the Transistor to the Data Center," *ISSCC*, 2017.

[5] F. Spagna, et al., "A 78mW 11.8Gb/s Serial Link Transceiver with Adaptive RX Equalization and Baud-Rate CDR in 32nm CMOS," *ISSCC*, pp. 366-377, 2010.

 978-1-5090-4940-0/18/$31.00

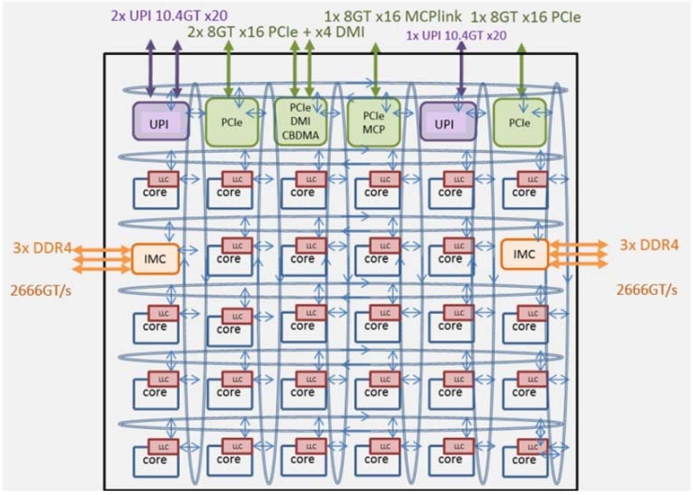

Figure 2.1.1: SKX processor architecture.

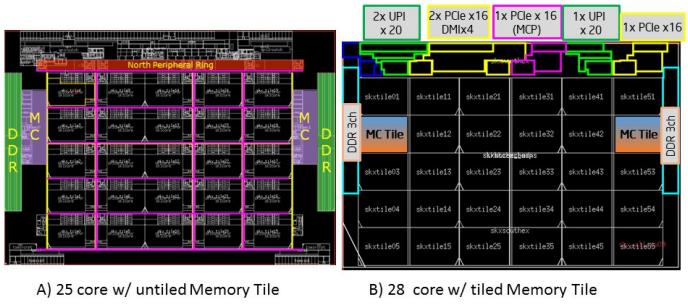

Figure 2.1.2: SKX core-tile-based floorplan evolution.

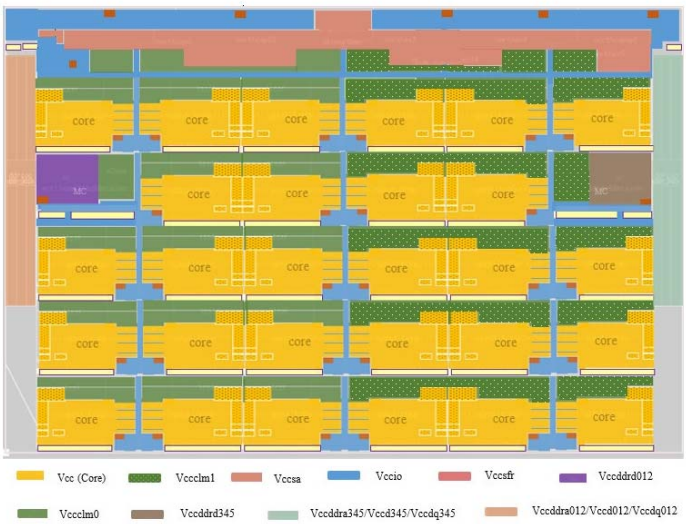

Figure 2.1.3: VCC domains of a 28-core SKX processor.

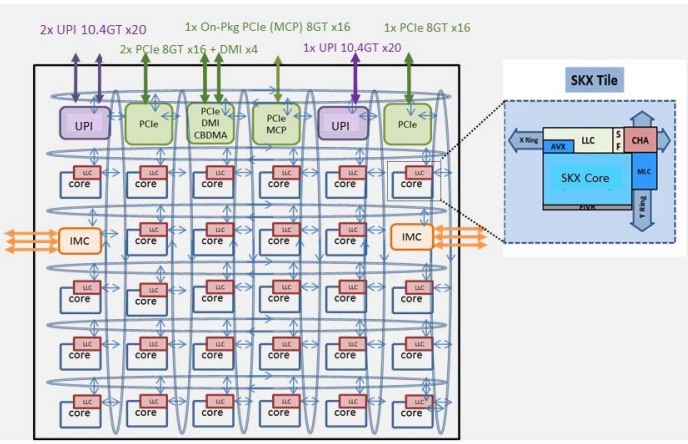

Figure 2.1.4: MESH architecture connecting cores and caches.

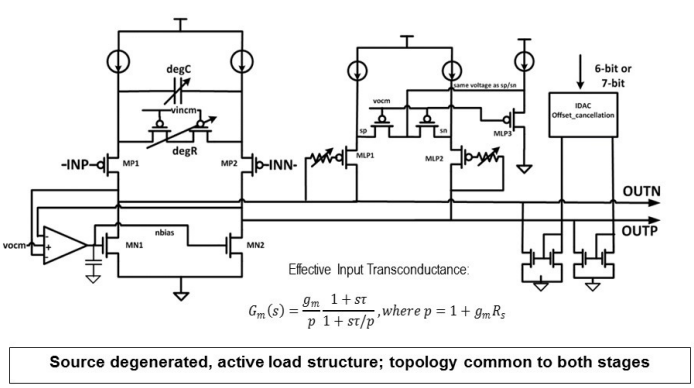

Figure 2.1.5: SKX SERDES CTLE circuit topology.

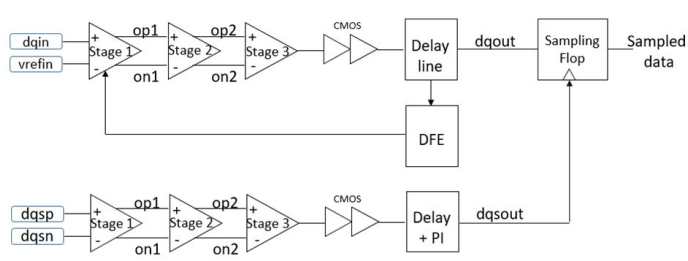

Figure 2.1.6: SKX DDR4 receiver (RX) architecture.

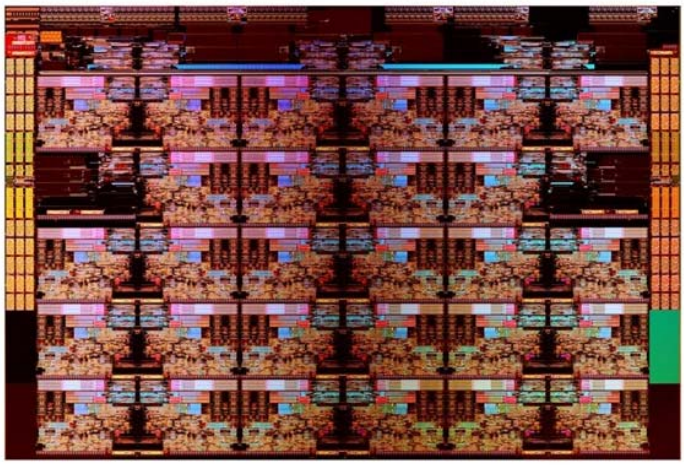

Figure 2.1.7: SKX die photograph.

978-1-5090-4940-0/18/$31.00