2.2 AMD Chiplet Architecture for High-Performance Server and Desktop Products

Samuel Naffziger[1], Kevin Lepak[2], Milam Paraschou[1], Mahesh Subramony[2]

[1]AMD, Fort Collins, CO
[2]AMD, Austin, TX

AMD's "Rome" and "Matisse" are second-generation AMD Infinity Fabric-based SoCs using 3 unique hybrid process technology chiplets to achieve leading performance, performance/$ and performance/W, targeting server and client markets, respectively (Fig. 2.2.1). The chiplet architecture enables leading edge 7nm [1] CPUs for multiple markets, while retaining backward compatibility to complex IO and memory subsystems in a scalable design with high reuse for improved time-to-market. A key benefit is the heterogeneous technology deployed between the CPUs and the IO/mixed-signal IP. It is well known that shrink factors in advanced nodes are much lower for analog circuitry than for digital logic and SRAM. By keeping the memory interfaces and SerDes in mature 12nm technology, costs are mitigated since those circuits see a very small shrink factor to 7nm and very little performance or power gain from advanced nodes. A low-cost 12nm IO die (IOD) with the high-yielding 8 "Zen2" core, 74mm^2 7nm CPU compute die (CCD) combine to provide very cost-effective performance.

The modularity of chiplets enables product configurations not achievable with monolithic designs. For both server/client, the top end core counts (64 and 16 resp.) would not be practical as a monolithically integrated chip, and in the case of the 64C server part, way beyond maximum reticle size (Fig. 2.2.2). This architecture required optimized on-package interconnect, engineering at a new level of package and silicon co-design, where each die floorplan is architected with the package routing constraints in mind, supporting product configurability, performance and latency requirements.

Infinity Fabric on-package (IFOP) SerDes provides a means for communication on short package routes (10-20mm) between the IOD and CCD. It was designed to maximize bandwidth and minimize power and latency. IFOP SerDes (Fig. 2.2.3) is the next evolutionary step in a low-power/high-speed interconnect building on the prior "Zeppelin" [2] design, which achieved 2 pJ/b power efficiency at a maximum bitrate of 6.4Gb/s per lane (not including external regulation power). This SerDes is asymmetric and receives a fabric clock from the IOD, which is distributed to the IOD TX/RX lanes, as well as the CCD PLL, which further distributes the clock to the CCD TX/RX lanes and CCD core. IOD and CCD transmitters utilize a forwarded clock from the core accompanying each lane data bundle to track the skew of the data bundles across the large phy area per TX lane. The TX uses a coarse bang-bang detector to train the local reference clock against the data-bundle-forwarded clock minimizing latency for the TX serialization cycle to ease the timing closure, which must cross 2000+μm routing across the phy. Single-ended signaling is used to save on power and package routing lanes, which terminate to VTT at each receiver. Leveraging balanced data patterns, TX currents driven from each transmitter are summed through each RX termination resistor to a common termination point requiring minimal power to actively maintain the termination voltage level. The VTT voltage is also used as reference for incoming data signal at the RX sampler with an added offset for random mismatch compensation.

To reduce power, active equalization is avoided within the transmitter or receiver; therefore, the incoming signal directly drives the receiver clocked samplers. T-Coils are used as passive equalization, on both the TX and RX. A local per-lane digitally controlled CDR is used for each lane to achieve bit lock in the receiver using a phase interpolator to lock and track to the phase of the incoming data pattern. A multiphase distributed bandwidth limited clock distribution provides the clock to the TX and RX allowing for CDR linear phase interpolation and lower clock distribution power. The total round-trip latency from CCD to IOD and back, including the digital control logic is 13 FCLKs, or less than 9ns, which represents a significant reduction over the prior generation. A key benefit of using organic package SerDes vs. 2.5D interposer [3] or EMiB [4] is longer reach, enabling 2-deep chiplet columns without which 8 CCDs would not be possible.

The "Rome" server product makes use of a 9-2-9 package for signal connectivity with 4 layers above the package core for signal routing while "Matisse" requires a 5-2-5 layer count. One of the signaling layers (others are similar) is shown in Fig. 2.2.4 along with the physical position of the CCD, IOD, as well as main external DRAM and SerDes interfaces. Each IFOP interface routes on 2 layers to connect each CCD and is routed to the center of the CCD to reduce memory latency from the L3 cache. The dense arrays of package vias to LGA signal pads must be outside of the physical locations of CCD and IOD due to high power-via-density underneath portions of CCD and IOD requiring detailed silicon + package co-design. Different "Rome" models do not populate all CCDs to optimize cost and performance profiles of the "Rome" product stack (Fig. 2.2.2). Each IFOP can supply ~55GB/s of effective bandwidth at 1.46GHz Fabric Clock (FCLK) with IFOP at 14.6GT/s leading to >= 4 CCDs populated in most configurations to balance overall bandwidth. In "Rome", all main uncore clocks can be decoupled, including FCLK, DDR MEMCLK, and IFIP (up to 18Gb/s) in contrast with the AMD EPYC™ 7001 Series (codenamed "Naples"), where these interface clocks were tied together. "Rome" (and "Matisse") have optimized clocking modes improving latency when FCLK = MEMCLK. Overall, the monolithic IOD, decoupled clocks and latency-optimized IFOP enabled a 19% average memory latency reduction for the socket (Fig. 2.2.6). The clocking flexibility, along with uncore DVFS and variable MEMCLK support, enables advanced power management for CPU vs uncore power within a fixed TDP budget. Infinity Fabric is optimized for CPU+IO access to DRAM, as well as peer-to-peer (P2P) PCIe® routing between devices. With 256B P2P read and write, and 512B DMA payload support, "Rome" is provisioned for PCIe-Gen4 (16Gbps) and beyond.

"Matisse" products leverage the same CPU chiplets with a unique but leveraged IOD, while maintaining backward compatibility to the current socket/platform infrastructure (AMD Socket AM4). Multiple floorplan experiments were conducted to reuse tiles as-is. while solving for IO placement requirements, maintaining data flow connectivity and minimizing memory latency. Maintaining IO socket compatibility with reused PHYs was accomplished by a system of firmware that is distributed across the SoC and executed by on-die micro controllers. Retaining voltage rail compatibility to the prior generation was achieved using low dropout linear regulators for the necessary logic. Managing mixed processes in a single package was solved by transitioning the 12nm IOD to copper pillars from the traditional solder bumps. Copper pillar is more compact, conducive to tighter bump pitches and enables common die height after assembly. In addition, using robust harvesting techniques, specialized IO-traps and secure firmware load flows, the IOD in a standalone package operates as the X570 chipset for the AMD AM4 socket, enabling the industry's first PCIe® 4.0 ready desktop platform (Fig. 2.2.5). Each CCD is fully tested for defects prior to assembly. If damaged during assembly, firmware takes action to 'down-bin' the CCD, putting CCD/related interconnect logic in a known low-power state.

The heterogeneous chiplet architecture required significant advances in engineering the package, interconnect, test and power management infrastructure, but enabled product configurations across multiple markets that are far more performant and cost effective than otherwise possible. Technology trends are such that chiplet approaches will become more prevalent as they provide a means to overcome the slowing of Moore's Law, while meeting growing demands for cost-effective computation.

References:

[1] S.-Y. Wu et al., "A 7nm CMOS Platform Technology Featuring 4th Generation FinFET Transistors with a 0.027μm^2 High Density 6-T SRAM Cell for Mobile SoC Applications," *IEDM*, 2.6.1-2.6.4, 2016.
[2] N. Beck et al., "Zeppelin: An SoC for Multichip Architectures," *ISSCC*, pp. 40-42, Feb. 2018.
[3] J. Macri, "AMD's Next Generation GPU and High Bandwidth Memory Architecture: FURY," *IEEE Hot Chips Symp.*, 2015.
[4] D. Greenhill et al., "A 14nm 1GHz FPGA with 2.5D Transceiver Integration," *ISSCC*, pp 54-55, Feb. 2017.

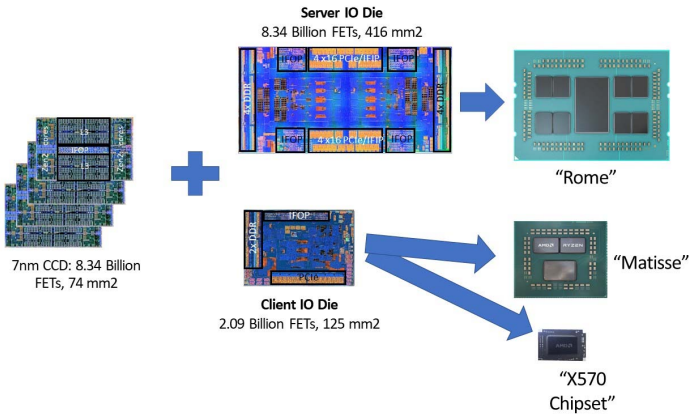


Figure 2.2.1: Three heterogeneous technology chiplets leveraged to many products and markets.

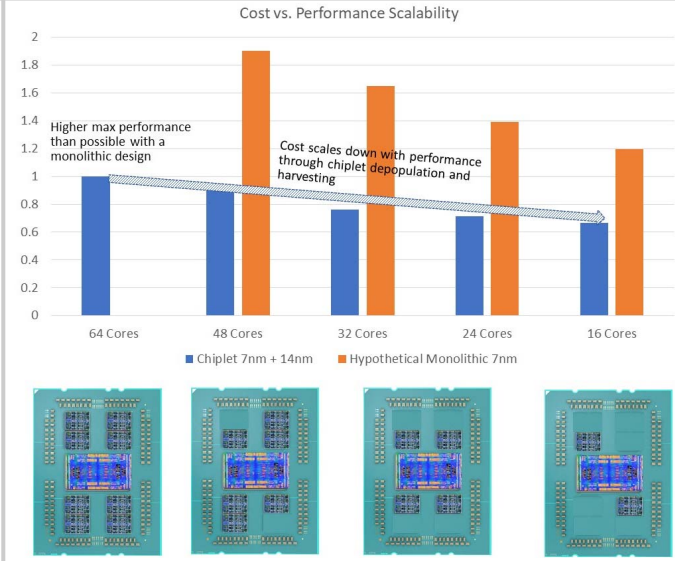


Figure 2.2.2: Cost-performance scalability with chiplet design.

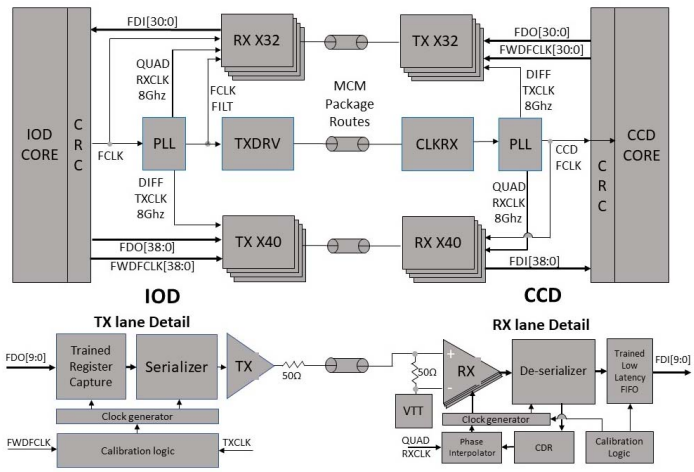


Figure 2.2.3: Infinity Fabric On-Package (IFOP) SerDes architecture.

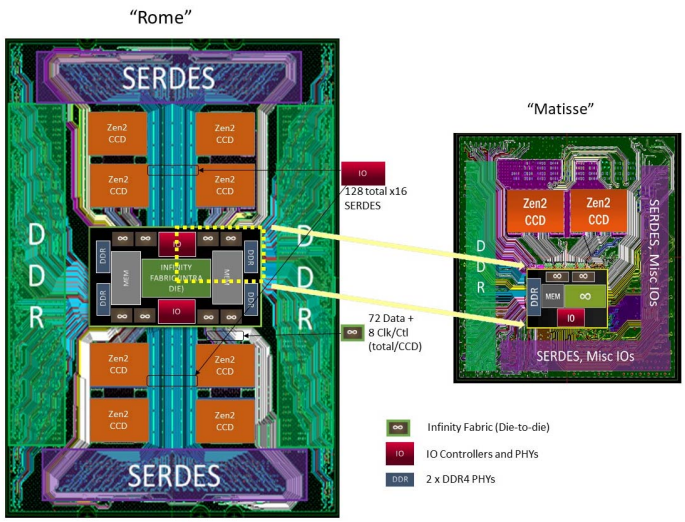


Figure 2.2.4: 'Rome' and 'Matisse' package design and IOD leverage.

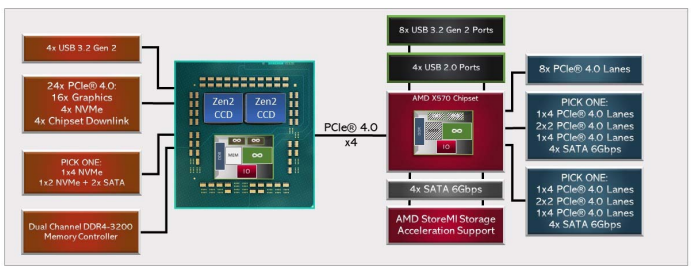
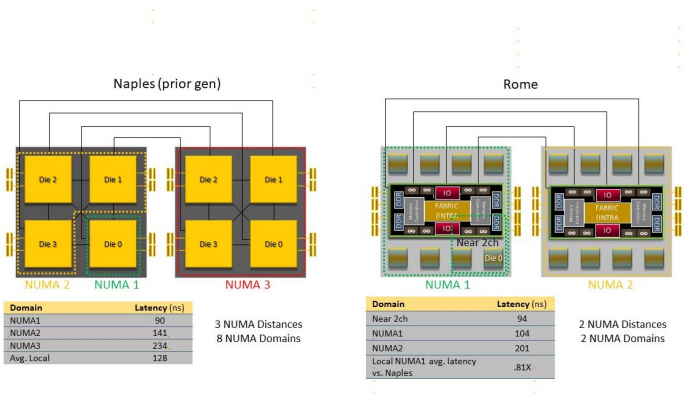


Figure 2.2.5: 'Matisse' IO connectivity with IOD-based chipset expansion.



Domain	Latency (ns)
NUMA1	90
NUMA2	141
NUMA3	234
Avg. Local	128

Domain	Latency (ns)
Near 2ch	94
NUMA1	104
NUMA2	201
Local NUMA1 avg. latency vs. Naples	.81X

Figure 2.2.6: 'Rome' central IOD reduces the number of NUMA domains and distances for much improved memory latency attributes relative to it's predecessor.