

Multi-Die Design for HPC Applications

Author

Frank Malloy

Sr. Director of Marketing

Overview

High-performance computing (HPC) has evolved dramatically over the years to address the evolving complex requirements of today's workloads. Initially, the term referenced specialized supercomputers to solve complex scientific and engineering problems. In addition to AI, in recent years, HPC applications have increased to include:

- Automotive including advanced driver-assistance systems (ADAS) and in-vehicle infotainment systems (IVI)
- Data center networking
- Cloud data center applications
- Many others, from medical imaging, to genomics, to weather forecasting

This white paper explains how the semiconductor ecosystem is adopting multi-die designs to address the performance, power, area, and latency requirements of HPC chips, which are on the path to trillion-transistor designs.

HPC Design Challenges—Before Multi-Die Designs

To achieve the performance demands of HPC applications, along with greater usage of highly parallelized chip architectures (particularly critical to AI applications), die areas are increased to the point of greatly impacting yield. For this reason, manufacturing of such chips has reached reticle limits. Also, in a traditional package, the number of I/O connections between monolithic chips on a PCB are very limited, but multi-die designs can take advantage of high-density microbumps and hybrid bonding. And high-bandwidth memory (HBM) and Universal Chiplet Interface Express (UCIe) and other wide interfaces require high numbers of dense I/Os.

High-density and high-speed bus connections, along with long, high-capacitance signal paths between ASICs (typically logic chips to memory chips) on printed-circuit boards (PCBs) create performance and memory bottlenecks. These long paths create challenges in meeting timing as well as introducing signal integrity (SI) concerns. In addition, the cost of advanced process node fabrication combined with ever increasing die sizes became significant. Since monolithic SoC design for HPC applications are fixed at one selected process technology and foundry, taking advantage of block reuse is limited.

Benefits of Multi-Die Design for HPC

To meet the performance, memory bandwidth, latency, and other key performance indicators (KPIs), along with the scalability demands of HPC, traditional monolithic designs are stretched to the limit. The advent of multi-die design, which integrates heterogeneous and homogeneous dies (also called chiplets) into a single standard or advanced packaging, allows functional scalability, system performance and power efficiency, and reduced time-to-market.

Key benefits include:

- **Increased performance.** Through direct vertical die stacking or close die placements side-by-side on a common base die or interposer, the interconnect length for critical high-speed signals is dramatically reduced compared to package-to-package connections on a PCB. This increase in performance and decrease in latency are critical to AI and automotive hardware design applications that require real-time processing, especially for ADAS and IVIs.
- **Reduced system power consumption.** Multi-die design also reduces system power consumption since the shorter connections reduces the signal capacitance, requiring less powerful drivers.
- **Improved signal integrity.** Adding to the benefits of shorter die-to-die interconnect, signal integrity issues are also reduced. Crosstalk and other electromagnetic signal degradation effects are minimized.
- **Higher density and scalability.** Multi-die design, combined with advanced 2.5D and 3D packaging, enables a much higher transistor density per footprint area, thanks to high bandwidth and low latency UCIe 2.5D die to die connections and 3D vertical stacking of dies. The higher density allows for greater scalability that is not impacted by Moore's law limitations.
- **Potential cost savings.** Multi-die design can be initially costly, due to the extra design and verification steps along with the cost of advanced packaging. However, the possibility of combining less-expensive process technologies for less performance-critical functions along with IP design reuse and smaller die sizes (along with the associated higher yields) can lead to cost savings over the long term.
- **Enables specialized interconnect technologies.** HPC chips demand high-bandwidth, low-latency, and high-density interconnects to meet SI, power, and timing constraints. Key advantages of multi-die designs include HBM, UCIe, and stacked die-to-die connectivity design methodologies and IP, including bump-out and physical-layer (PHY) implementations.
- **Improved HPC design architectures.** Multi-die designs create greater opportunities for more innovative chip architectures that may not be feasible with monolithic designs using traditional packaging.

Multi-Die Design and Analysis Requirements for HPC

Designing a complex HPC system using multi-die and advanced packaging requires special design and tool requirements and extra design steps. However, the benefits detailed above provide clear advantages that justify the extra design and analysis steps. For the design of HPC multi-die designs, advanced EDA tools are crucial to successful projects. The tools must be able to support the key technologies and design steps used by HPC designs:

- **Complex package support.** Multi-die designs require advanced packaging to achieve the chip and interconnect density. This packaging includes silicon interposers, organic fan-out wafer-level packaging (FOWLP), and organic substrates with embedded silicon bridges. This requires capabilities in the creation, planning, and routing for these specialized packages.

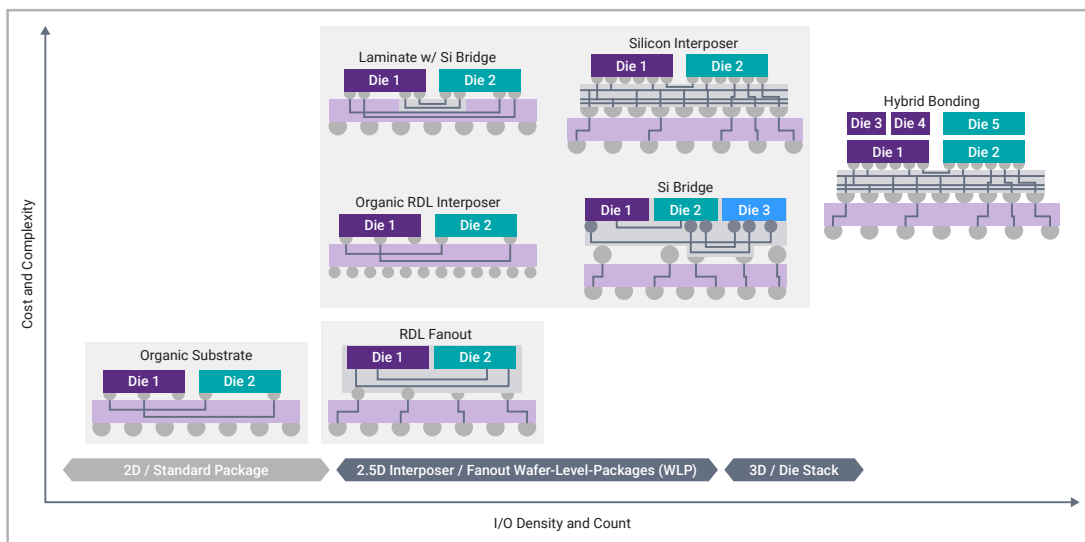


Figure 1: Examples of Complex Advanced Packages

- Feasibility and prototyping support.** HPC designs are built on complex architectures, which require architectural exploration capabilities to arrive at an optimal design. Exploring possible architectures and evaluating the optimal choice requires tools which can perform system pathfinding, software-based performance and workload modeling, and provide feasibility and prototyping capabilities, without having to go through the long process of detailed hardware design implementation.

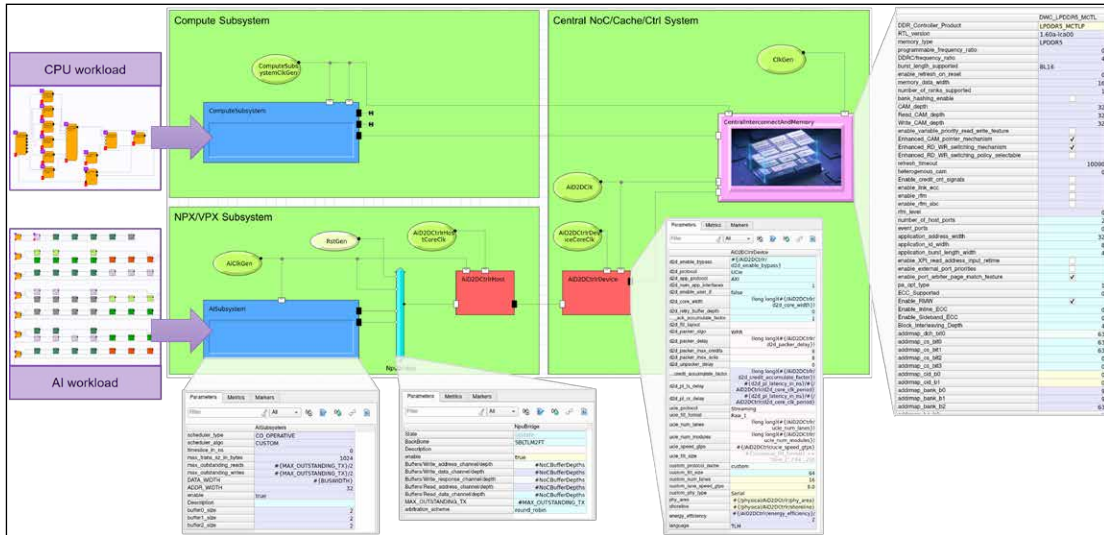


Figure 2: Architecture Modeling in Synopsys Platform Architect for Multi-Die

- Multi-die design for test support.** While all EDA tools provide design for test (DFT) capabilities, multi-die designs require additional test insertion, pattern generation, etc. to support assembly and manufacturing testability across multiple chips in a complex HPC design. Support for the IEEE 1838 Standard for Test Access Architecture for Three-Dimensional Stacked Integrated Circuits is a key requirement for EDA tools. Lane Test and Repair (LTR) provides system-level built-in self-test (BIST) for high-volume lanes in die-to-die interconnections. LTR and monitor, test, and repair (MTR) can be used to recover from faults over time, as well as recover yield from assembly and test failures. In UCIe designs, the standardized bump pitches are very tight, which is challenging for traditional test probing. Here, a built-in on-die approach using dedicated monitor, test, and repair (MTR) IP is needed to provide high interconnect defect and fault coverage.
- HBM support.** HPC designs often require access to fast, high-bandwidth memories (HBMs) placed close to the compute die(s) for shortest wirelength and minimum possible wire resistance. An HBM memory bus is very wide in comparison to other DRAM memories such as DDR4. The larger number of connections from the compute die(s) to the memory die(s), relative to DDR4, benefits greatly from advanced, dedicated bump planning and specialized auto-routing capabilities in EDA design implementation tools. The bump pitch is very tight, the bus signals require shielding for SI purposes, and the routing utilization is very high (nearing 100%) and requires 45-degree routing over only one or two layers. To perform such routing using traditional, manual techniques is error prone and time consuming.

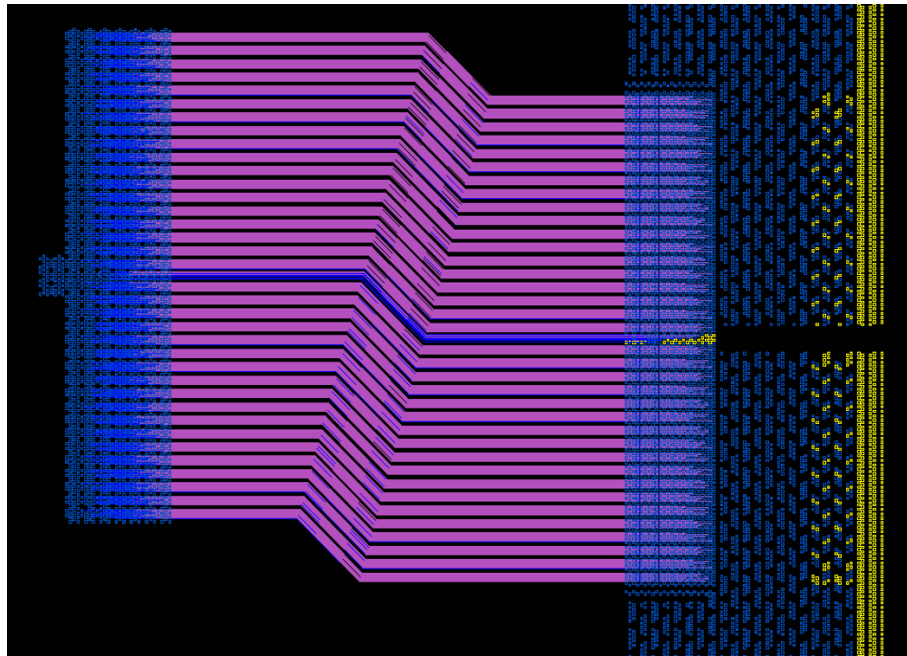


Figure 3: 45-Degree HBM Routing

- **Photonics capabilities.** Photonics designs have a wide range of applications, and HPC systems can benefit from this technology. One important application is to achieve high-speed data transmission between servers and storage systems, and in automotive light detection and ranging (LIDAR) systems used in ADAS. Optical interconnects and silicon photonics help reduce latency and increase bandwidth. Combining a photonics integrated circuit (PIC) die with logic/memory dies in a multi-die design creates a more powerful HPC system with maximum bandwidth.
- **Silicon IP.** Heterogeneous HPC systems require a portfolio of best-in-class IPs to accelerate time to market and reduce risk. The right IP is needed on the process node best suited for the functionality required. For example, the compute die may be on most advanced node and the IO functions like 224G, UAL link, or PCIe 7.0 may be implemented in an older node to save cost. Important consideration is also needed for IPs such as CPUs for control and management, PVT monitors and control functions for silicon life cycle management, monitor, test and repair features for high-speed links, and security features for preventing counterfeit parts during manufacturing and to protect user keys when encrypting and decrypting data. An extensive and proven portfolio of IP designs is a significant benefit to time-to-market, PPA, and latency. For HPC designs in general, examples of key IP include 112G/224G Ethernet, UALink, Ultra Ethernet, PCIe 7.0, digital signal processors (DSPs), HBM, UCIe, and Foundation IP. For 3D integration, IPs will need to have their PHYs modified for a wide range of 3D topologies. For example, PHY must be reoriented so signals from the bumps pass through a die's substrate using through silicon vias (TSVs) to connect to the metal layers and route to the diffusion layer silicon devices. IOs also may need to account for TSVs and routing to connect signals and power to the top die.
- **Multi-die design analysis and signoff capabilities.** HPC multi-die designs require not only die-level analysis and signoff but also support for additional analysis requirements that impact a disaggregated design. HPC designs utilize direct die stacking, and dense placement of large numbers of high-performance compute dies on an interposer or substrate. This causes thermal/heat dissipation issues. Thermal analysis of a multi-die design in three dimensions is required to reveal thermal hotspots which may require changes in die placement, or in heat sink and thermal mitigation. Electromigration (EM) and IR-drop (IR) analyses are key, since in any die stack the IR drop of a top die is impacted by the IR drop of the die below. This must be accounted for in the analysis.

Also, high-speed die-to-die signal paths on an interposer or substrate can be subject to SI issues such as crosstalk, reflections, and signal degradation. Electromagnetic solver-based extraction, along with SI simulation/analysis is a must to guarantee signal integrity.

Static-timing analysis (STA) of a multi-die design can involve cross-die paths of dies using different process technologies (even from different foundries) with different timing libraries. In designs containing large numbers of dies, significant compute resources are required for timing analysis, so efficient timing models are required for the dies to reduce memory usage and analysis runtime. Also, each die in the design can have its own set of timing corners (process, voltage, and temperature) and modes, and the designer needs to be able to setup desired valid combinations and perform analyses efficiently.

Physical verification—design-rule checking (DRC) and layout-vs-schematic (LVS) checking—is also significantly more complex for HPC multi-die designs. In addition to traditional in-die DRC/LVS checks, new constraints are required which are unique to multi-die designs. Examples include die-to-die minimum spacing and coverage, die enclosure, bump alignment, and cross-die logical-physical consistency.

The Synopsys Solution for HPC Multi-Die Designs

Creating and analyzing HPC multi-die designs requires a comprehensive and scalable solution that offers a complete path from early architecture exploration, to design feasibility and prototyping, through simulation and detailed design implementation, to analysis and verification. The Synopsys Multi-Die Solution provides a comprehensive and scalable solution for fast heterogeneous integration. The solution, including EDA and IP products, enables early architecture exploration, rapid software development and system validation, efficient die/package co-design and optimization, robust die-to-die and chip-to-chip connectivity, and improved manufacturing and reliability. The solution helps designers reduce time-to-market (TTM) and increase productivity, which are critical in the fast-moving HPC design development.