



UCle 3.0 Specification: Driving Innovation for Efficient, Scalable, and Reliable Chiplet Integration

Dr. Debendra Das Sharma
Intel Senior Fellow and co-GM Memory and I/O Technologies
Chair of UCle Consortium

The semiconductor industry is undergoing a transformative shift towards open chiplet-based architectures, driven by the need for increased performance, flexibility, cost-effectiveness, and open innovation. As the complexity and demands of modern computing applications continue to escalate, traditional monolithic chip designs are increasingly challenged to keep pace. In response, chiplet-based architectures have emerged as the solution, offering modularity, scalability, and the ability to integrate heterogeneous chiplets and technologies within a single package. UCle has emerged as the leading standard for chiplet interconnects, providing a unified framework for communication between diverse chiplets from multiple sources. With the release of UCle 3.0, the specification introduces several key enhancements to address the challenges faced by modern computing applications.

The Universal Chiplet Interconnect Express™ (UCle™) 3.0 specification marks a significant milestone in the evolution of chiplet interconnect technology, addressing the burgeoning demands of high-performance computing (HPC), artificial intelligence (AI), and other bandwidth-intensive applications. This white paper delves into the intricacies of UCle 3.0, highlighting its groundbreaking features including doubled data rates, support for continuous transmission protocols, enhanced operational and idle power savings, and enhanced manageability for seamless interoperability while maintaining full backwards compatibility with the prior generations of UCle. These advancements enhance the position of UCle as the pivotal technology in the evolution of system-in-package (SiP) designs, enabling more efficient, scalable, and reliable chiplet integration. We encourage you to read our white papers on [UCle 1.0](#), [UCle 1.1](#) and [UCle 2.0](#) as well as our [webinar recordings](#) for additional information on UCle 1.0, UCle 1.1 and UCle 2.0 for insight into previous specifications.

Doubling Data Rates for UCle-A (Advanced Package, 2.5D) and UCle-S (Standard Package, 2D)

The relentless demand for higher bandwidth is a defining characteristic of contemporary computing applications, particularly in fields such as AI, HPC, and data analytics. These applications have an insatiable demand for bandwidth within the tight shoreline constraints, necessitating improvements in shoreline bandwidth density. UCle 3.0 addresses this demand by increasing the maximum data rate from 32 GT/s to 64 GT/s (as well as adding support for 48GT/s). This enhancement ensures that UCle remains at the forefront of high-speed interconnect solutions, providing the necessary bandwidth to support the ever-growing data requirements of modern applications.

UCle 3.0 achieves this increase in data rates while maintaining full backward compatibility with previous versions of the specification. This is a critical consideration, as it ensures that existing

systems and infrastructure can seamlessly integrate with the new standard. The specification preserves existing sideband, valid, track, data, training, and signaling protocols, providing a smooth transition for system designers and developers while ensuring interoperability with older chiplets designed to prior generations of the specification.

The technical approach to achieving these higher data rates involves:

- Non-Return-to-Zero (NRZ) and unidirectional signaling, a proven technique that offers robust signal integrity and efficient data transmission, consistent with prior versions of the specification. Additionally, the specification mandates the existing quarter-rate clocking for 48/64 GT/s. The bit error rate (BER) is 10^{-15} for 48 GT/s and 10^{-12} for 64 GT/s. This ensures that data transmission remains reliable using existing CRC and replay mechanisms, even at increased speeds.
- Enhanced equalization techniques to support the increased data rates. These include a 3-tap TX feed-forward equalizer (FFE), which provides precise control over signal shaping and minimizes inter-symbol interference. The specification also incorporates a first-order passive RX continuous-time linear equalizer (CTLE), which further optimizes signal quality by compensating for channel impairments. An optional 1-tap RX decision feedback equalizer (DFE) can be deployed for applications that require additional signal conditioning.
- Enhancements to the Link Training State Machine to perform I/O corrections as well as EQ preset selections to establish the desired eye margins.

| Characteristics / KPIs | UCIe-S (2D) | UCIe-A (2.5D) | UCIe 3D | Comments |
|------------------------------------|---|--|--|---|
| Characteristics | | | | |
| Data Rate (GT/s) | 4, 8, 12, 16, 24, 32, 48, 64 | | Up to 4 | UCIe 3D SoC Logic frequency – power efficiency is critical Added 48G and 64G with UCIe 3.0 |
| Width (each cluster) | 16 | 64 | 80 | UCIe 3D: Options or reduced width to 70, 60... |
| Bump Pitch (μm) | 100 – 130 | 25 – 55 | ≤ 10 (optimized) $> 10 - 25$ (functional) | Must scale so that UCIe fits within the bump area, UCIe-3D must support hybrid bonding |
| Channel Reach (mm) | ≤ 25 | ≤ 2 | 3D vertical | UCIe-3D: FtF, FtB, BtB, multi-stack possible |
| Target for Key Metrics | | | | |
| BW Shoreline (GB/s/mm) | 28 – 224 278, 370 | 165 – 1317 1975, 2634 | N/A (vertical) | For UCIe-S and UCIe-A: First row is for 4-32G. Second Row is for 48G and 64G respectively. Numbers are for 45u (UCIe-A) and 110u (UCIe-S) |
| BW Density (GB/s/mm ²) | 22 – 192 | 188 – 1646 | 4,000 (9 μm) – 300,000 (1 μm) | Numbers are for 45u (UCIe-A) and 110u (UCIe-S) |
| Power Efficiency Target (pJ/b) | 0.5 (≤ 16 G) 0.75 (≥ 32 G) | 0.25 (≤ 12 G) 0.3 (16G – 32G) 0.5 (≥ 48 G) | < 0.05 at 9 μm \rightarrow 0.01 at 1 μm | |
| Low-Power Entry/Exit | 0.5nS ≤ 16 G, 0.5-1nS ≥ 24 G | | 0nS | No preamble or post-amble |
| Reliability (FIT) | 0 < FIT (Failure in Time) $\ll 1$ | | 0 < FIT $\ll 1$ | |
| ESD | 30V CDM | | 5V CDM $\rightarrow \leq 3$ V | UCIe-3D: 5V CDM at introduction, no ESD for W2W hybrid bonding possible |

Table 1: Key performance indicators of UCIe 3.0

In terms of bandwidth density, UCIe 3.0 achieves a bandwidth density of 1.7-2x linear and 1.3-1.6x areal over UCIe 2.0, while keeping the power efficiency low, as shown in Table 1. This power efficiency is crucial for applications that require high performance without compromising energy consumption.

Support for Continuous Transmission Protocols (New Usage for UCIe)

Continuous transmission applications, such as digital signal processing (DSP), require seamless data flow between components like ADC chips and SoCs. These applications are characterized by

their need for uninterrupted data transmission, which is essential for maintaining the integrity and accuracy of signal processing tasks. Leading DSP companies have expressed interest in leveraging the UCle standard for these applications, recognizing its potential to streamline data communication and enhance system performance. Figure 1 shows an example of such an application where a separate ADC chiplet is connected to a SoC using UCle.

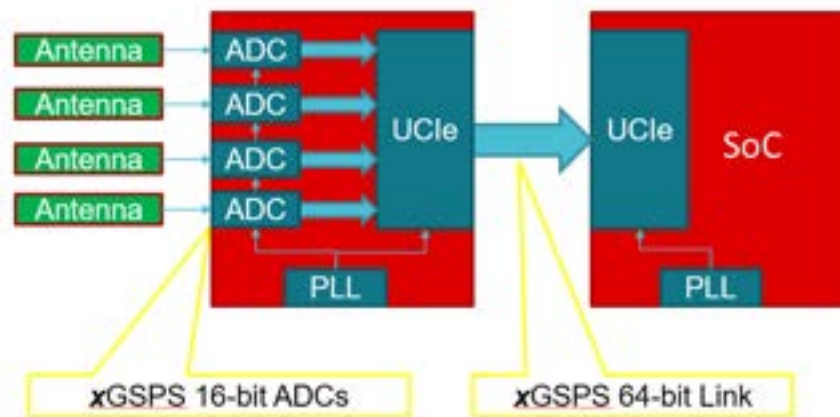


Figure 1: Example application for continuous transmission over UCle

UCle 3.0 supports continuous transmission protocols by running the link at the same data rate as data generation and consumption. This approach eliminates the need for separate phase-locked loops (PLLs), which can introduce frequency noise in sensitive analog circuits. By synchronizing the data rate with the rate of data generation and consumption, UCle 3.0 ensures that data flows smoothly and efficiently between components, reducing latency and improving overall system responsiveness. System designers can control the data rate by varying the REFCLK provided to the PLLs within the interoperability range, allowing for flexible and adaptive system configurations.

The specification enables the application to use the existing UCle retimer encodings, which are used for credit exchanges, to send periodic synchronization markers and/or parity, one bit per 8 UI, along with the necessary enhancements to the internal Raw Die-to-Die and Flit-Aware Die-to-Die (RDI/FDI) interfaces.

Operational Power Savings Enhancement

Power efficiency is a critical consideration in modern computing applications, particularly as systems become more complex and are constrained by available power. UCle 3.0 introduces a power-saving feature for the physical layer, allowing TX adjustment of clock-to-data skew during runtime recalibration of the link. This capability, previously available only on the RX side, repurposes the TX side's wider adjustment range (provided for link initialization flows) for runtime recalibration as well, resulting in significant power savings.

Idle Power Savings Optimization during L2

In addition to operational power savings, UCle 3.0 introduces additional idle power savings through L2 optimizations. Deeper power savings are achievable by turning off power and clock to the sideband infrastructure in L2, which is the deepest power-saving state.

UCle 3.0 utilizes existing sideband clock and data pins to indicate L2 exit using DC signal levels. A small amount of logic remains active to detect the toggle in the sideband clock and sideband data on exit from L2 and wake up the rest of the sideband. This approach ensures that systems can aggressively power gate circuits during L2 and still have a robust transition from a low-power state to full operation. Rules are provided to ensure symmetric or one-sided exit, allowing for flexible system configurations and ensuring compatibility with a wide range of applications.

Manageability Enhancements for Seamless Interoperability

The need for enhanced manageability and interoperability becomes increasingly important with the proliferation of UCle chiplets. UCle 3.0 introduces several features designed to streamline system management and ensure seamless communication between chiplets. UCle 3.0 introduces five new features listed below to help build out an open chiplet-based ecosystem: firmware download, priority packets in sideband, extending sideband reach for star topology, support for open drain pin at the SiP level, and fast throttle and shutdown.

Firmware Download

Modern chiplets may require firmware to function properly, supporting external interfaces and internal structures, management features, and more. Avoiding the need for each chiplet to have its own external flash or firmware loading mechanisms is crucial, as it simplifies system design and reduces costs. UCle 3.0 introduces a simple register-based mechanism for firmware download which can be used via the sideband or mainband and can be implemented with a simple hardware finite state machine (FSM). The Director Chiplet plays a crucial role in managing firmware updates within a system-in-package (SiP) by loading mutable firmware from external sources and initializing the side-band management network. This facilitates the subsequent download of the initial mutable firmware to individual chiplets, which then boot using the downloaded firmware. Once operational, chiplets can request further firmware updates through either the sideband or mainband, utilizing protocols such as Management Component Transport Protocol (MCTP) and Platform Level Data Model (PLDM). To ensure seamless interoperability between chiplets, data structures like circular buffers are defined, allowing efficient communication and coordination within the SiP. Figure 2 shows an example topology with Director Chiplet and other chiplets within a SiP.

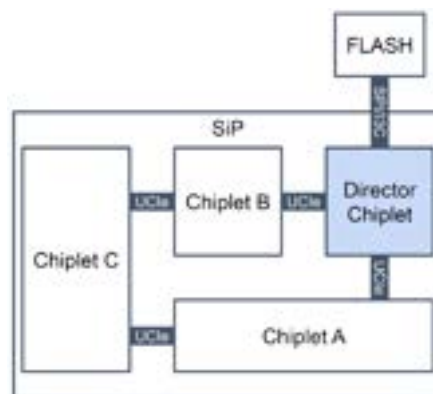


Figure 2: Example SiP topology with Director Chiplet

Priority Sideband Packets

Certain events, such as power down, wake-up, and low-latency telemetry data, require high-priority notification over others, such as debug dump, which can be large bulk transfers. These events

should not be delayed by normal traffic, as timely communication is essential for maintaining system stability and performance. UCle 3.0 creates a mechanism to interrupt regular sideband packets to deliver priority packets with an upper bound on hop-to-hop transfer of 48UI (8UI to get to the boundary + 8UI SB clock low to switch to high priority + 32UI of transfer) at 800MHz or 60ns. Regular sideband packets are interrupted at the next aligned 8UI boundary. This is followed by the sideband clock going low for 8UI to indicate the priority switch. After that, the priority packet is sent, inserting priority vectors for transport to remote link partners. The priority packet consists of 32UI of information, with 23 bits for the priority vector, 5 bits for opcode, 3 reserved bits, and 1 bit for parity. Figure 3 shows an example of this transfer.

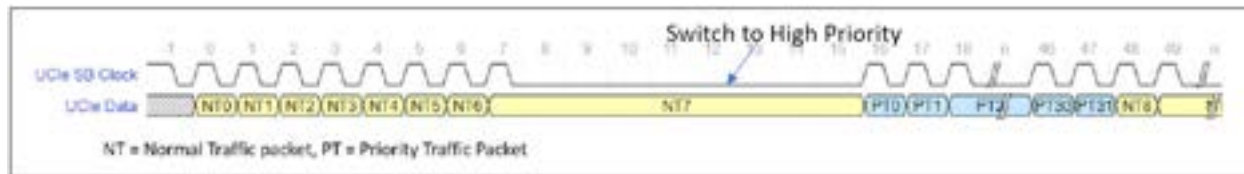


Figure 3: Example of interrupting a sideband packet in order to transmit a priority packet

Extended Reach Sideband (UCle-S Only)

Extending the sideband channel reach to 100mm minimizes hops and daisy chaining in SiP designs, enabling practical usage of a star topology. This topology has the added benefit of higher bandwidth, lower latency, and better security characteristics since the director chiplet can communicate directly with other chiplets instead of going through hops. This extension is particularly valuable in complex systems where multiple chiplets must communicate over longer distances. UCle 3.0 provides guidelines for extended reach, specifying Input high voltage (V_{ih}), Input low voltage (V_{il}) levels and emphasizing Txron as a meaningful measurement for eye height and width. Driver Ron (On Resistance) is limited to 60 Ohm in the worst case, ensuring reliable signal transmission over extended distances.

Open Drain Pins

Critical events like emergency shutdown or fast throttle require SiP-wide simultaneous broadcast to all chiplets. UCle 3.0 introduces open drain pins for low-latency, bi-directional events, used for specified events and vendor-defined events. These pins are intended for package-level routing, providing a reliable and efficient means of communication across the SiP.

Fast Throttle and Emergency Shutdown

In systems-in-package (SiP) designs featuring multiple chiplets from different vendors, managing thermal limits is a complex challenge due to the varied technology nodes and temperature reliability limits of each chiplet. Each chiplet has its own temperature sensing capabilities and associated error margins, resulting in a specific maximum junction temperature (T_j) limit, beyond which transistor function is compromised. To prevent exceeding these limits, SiPs must implement effective mitigation mechanisms, both reactive and proactive, to ensure the system remains within safe operating temperatures. This necessitates a standardized approach across chiplet vendors to maintain critical function interoperability at the SiP level.

The solution provided in UCle 3.0 involves two key mechanisms: fast throttle and emergency shutdown. Fast throttle is implemented by introducing a common dedicated open drain bidirectional pin for all chiplets in a thermal zone on the package. This setup allows any chiplet to assert the pin

when its internal fail-safe temperature limit is reached or when signaled by an external platform experiencing high temperatures. Upon assertion, all participating chiplets throttle their operations to a pre-negotiated level at a defined rate, effectively reducing heat generation and maintaining safe temperatures. For emergency shutdown, a similar open drain bidirectional pin is used, along with an off-package driver. When a chiplet's maximum temperature limit is exceeded, it asserts the pin, triggering a shutdown signal to both internal and external power sources. This coordinated response ensures that power supplies are cut off to prevent potential damage to the SiP or the broader system, safeguarding against catastrophic thermal events.

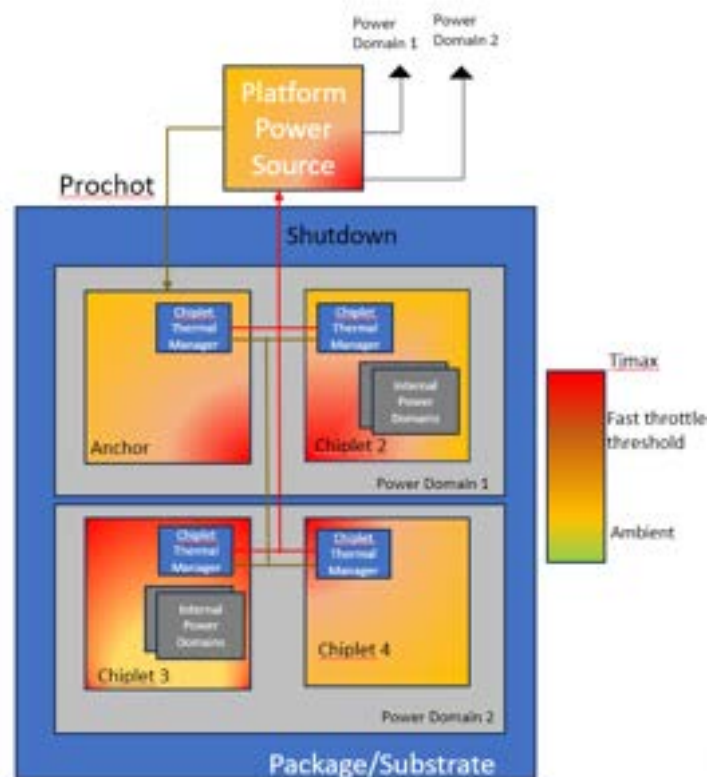


Figure 4: Example connectivity and settings for Fast Throttle and Emergency Shutdown

Conclusion

UCle 3.0 represents a transformative advancement in chiplet interconnect technology, addressing the critical needs of modern computing applications. By doubling data rates, supporting continuous transmission protocols, optimizing power savings, and enhancing manageability, UCle 3.0 sets a new standard for efficient, scalable, and reliable chiplet integration. As the semiconductor industry continues to evolve, UCle 3.0 will play a pivotal role in enabling the next generation of high-performance SiP designs, providing the necessary infrastructure to support the ever-growing demands of modern computing applications.