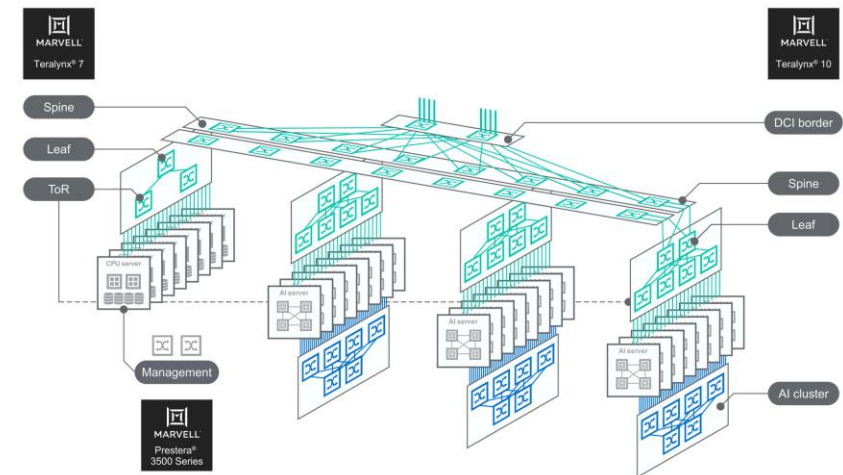
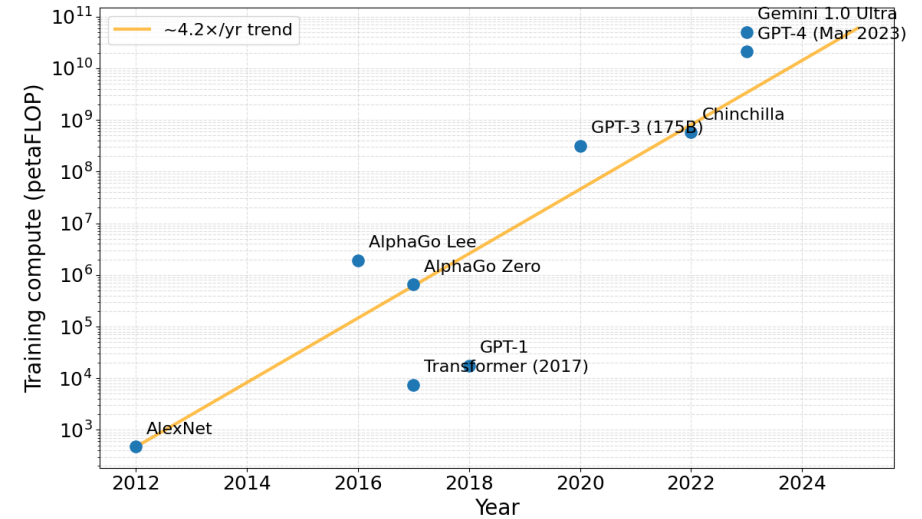


Die-to-Die Interconnect

Ahmed Gharbiya
Marvell Technology

Market Drivers

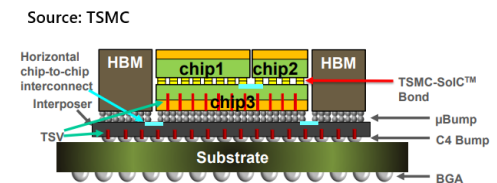
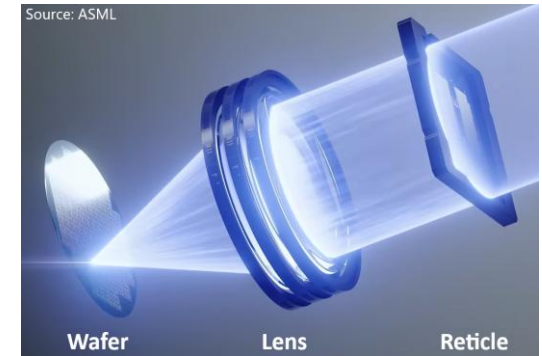
- Exponential growth in data-center demand
- Rapid advances in compute and connectivity requirements
- This talk focuses on chip-level scaling to meet future performance demands



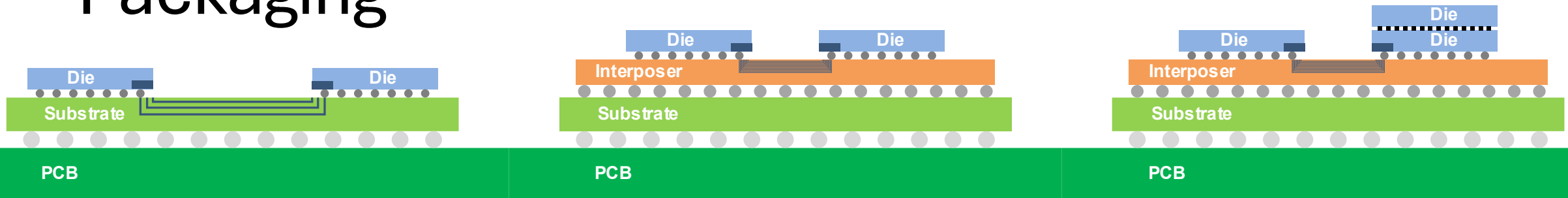
Chip Scaling and Chiplets

- System-on-Chip
 - Monolithic integration, the die size must increase to incorporate more functionality
 - The reticle limit is a technological constraint on the maximum die size that can be patterned in a single exposure*
 - Example, Marvell recently introduced the industry's first monolithic 100T fabric device
- Chiplet
 - Heterogeneous integration, multiple dies in a single package
 - Chiplets leverage advances in packaging [Ref 1] and die-to-die circuit innovations

*Larger-than-reticle designs are still possible using stitching or wafer-scale approaches



Packaging



2D

2.5D

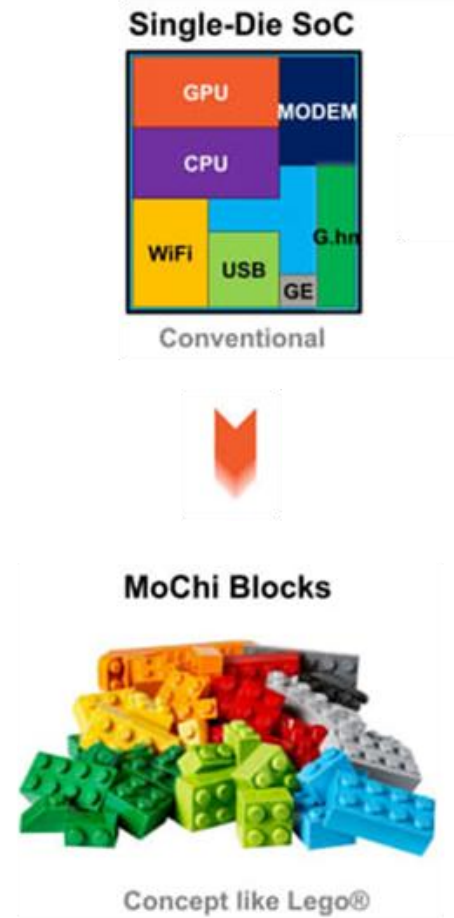
3D

Feature	2D Packaging	2.5D Packaging	3D Packaging
Die placement	Side-by-side on a package substrate	Side-by-side on an interposer	Stacked vertically
Integration Density	Low	Moderate	High
Area efficiency	Least space-efficient	More efficient than 2D	Most space-efficient
Performance	Limited	Improved bandwidth and power efficiency	Maximum performance
Cost	Most cost-effective	Higher than 2D, lower than 3D	Highest cost

Bonding Level	Interconnect Type	Typical Size
Package → PCB	BGA (Ball Grid Array)	0.4–1.0 mm
Interposer → Substrate	C4 (Controlled-Collapse Chip Connection)	90–150 μm
Die → Interposer (2.5D)	μbump (Micro-bump)	20–40 μm
Die → Die (3D stacking)	SoIC Hybrid Bond (Copper–Copper Direct Bond)	< 10 μm

History Trivia

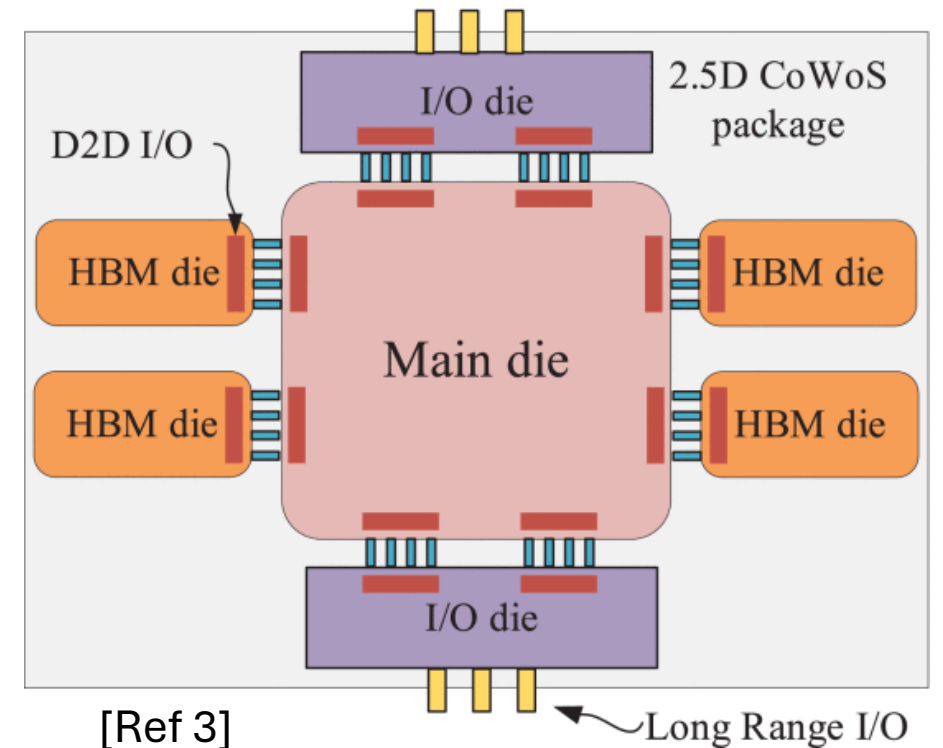
- Multi-die integration has existed for a while
- For example, in 2015
 - Marvell introduced one of the earliest commercial multi-die SoC architectures: MODular CHIp “MoChi” [Ref 2]
 - MoChi enabled multi-Gbps, low-latency die-to-die links between modular dies
- The concept can be traced back to the 1980s
 - Multichip modules (MCMs) laid the foundation for today’s modern chiplet architectures



[Ref 2]

Chipllets Advantages and Challenges

- Advantages
 - Scales integration beyond technological / economical limits
 - Enables mixing of process nodes
 - Improves yield through smaller die partitioning
 - Reduces development cost via IP reuse
 - Supports scalable product variants and roadmap flexibility
- Challenges
 - Requires tight co-optimization across design, technology, and packaging
 - Increases assembly and advanced packaging cost
 - Needs robust die-to-die interconnect circuits, architectures, and standards

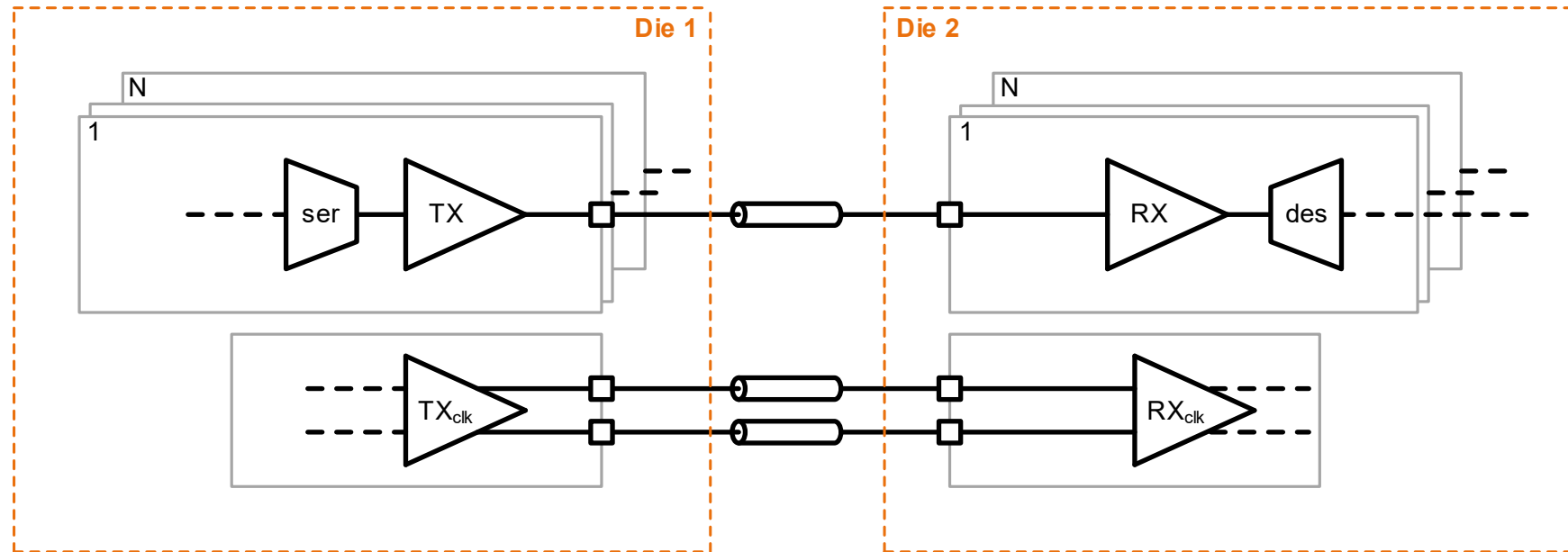


D2D Interconnect Metrics

- D2D key performance metrics

Parameter	Example [JSSC 25, Ref 3]	Unit
Data Rate per Lane (Speed)	32	Gbps
Power Efficiency (Power)	0.36	pJ/bit
BER (Accuracy)	1e-15	
Latency	1.4	ns
Bandwidth Perimeter Density (Data Rate/Beachfront)	7.68	Tbps/mm
Bandwidth Area Density (Data Rate/Area)		Tbps/mm ²

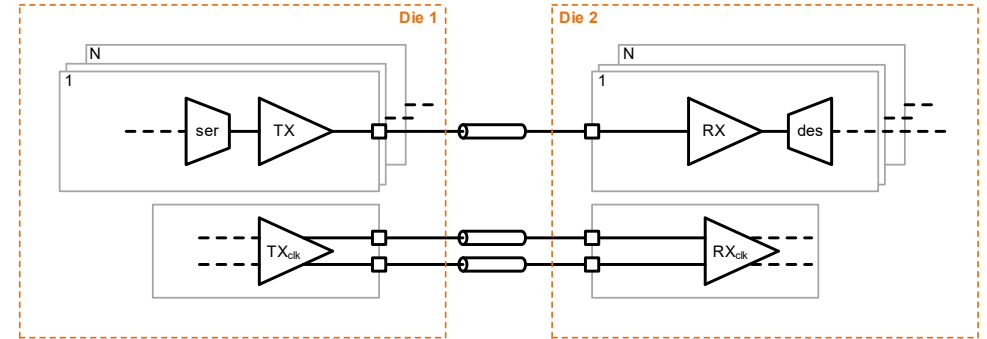
D2D Physical Layer



- The dominant D2D Phy:
 - Source-synchronous architecture
 - Parallel single-ended data interface
 - Differential clock forwarding

D2D Phy

Advantages and Challenges



Source-synchronous

- ✓ Avoids complex clock recovery
- ✓ Jitter tracking

Parallel single-ended Data

- ✓ At lower data rate the insertion loss and cross-talk are lower
- ✓ Single-ended for high bandwidth density

Diff Clock

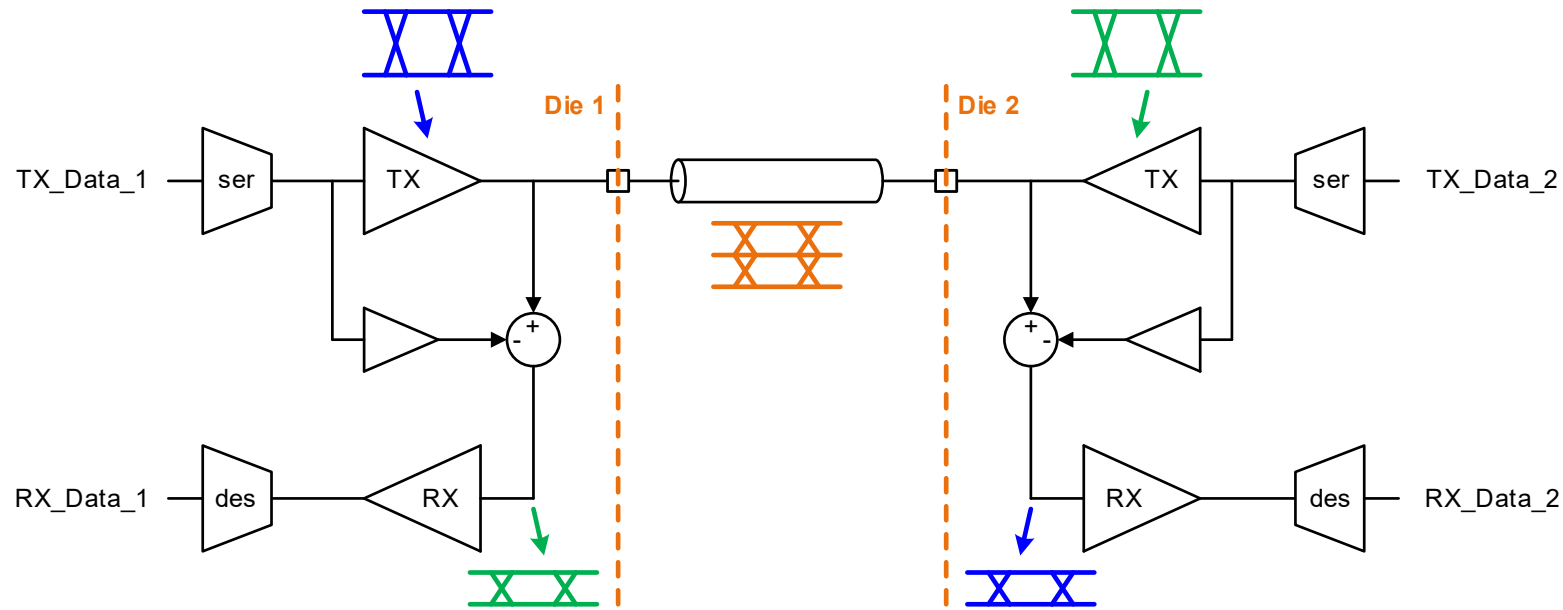
- ✓ Better immunity to interferences

- ✗ Requires tight channel matching
- ✗ Requires per-lane deskew logic

- ✗ Lower data rate requires high pin count
- ✗ More sensitive to interferences
- ✗ Single-ended signaling is a worse supply aggressor and more sensitive to supply disturbance

- ✗ Requires extra bump and routing

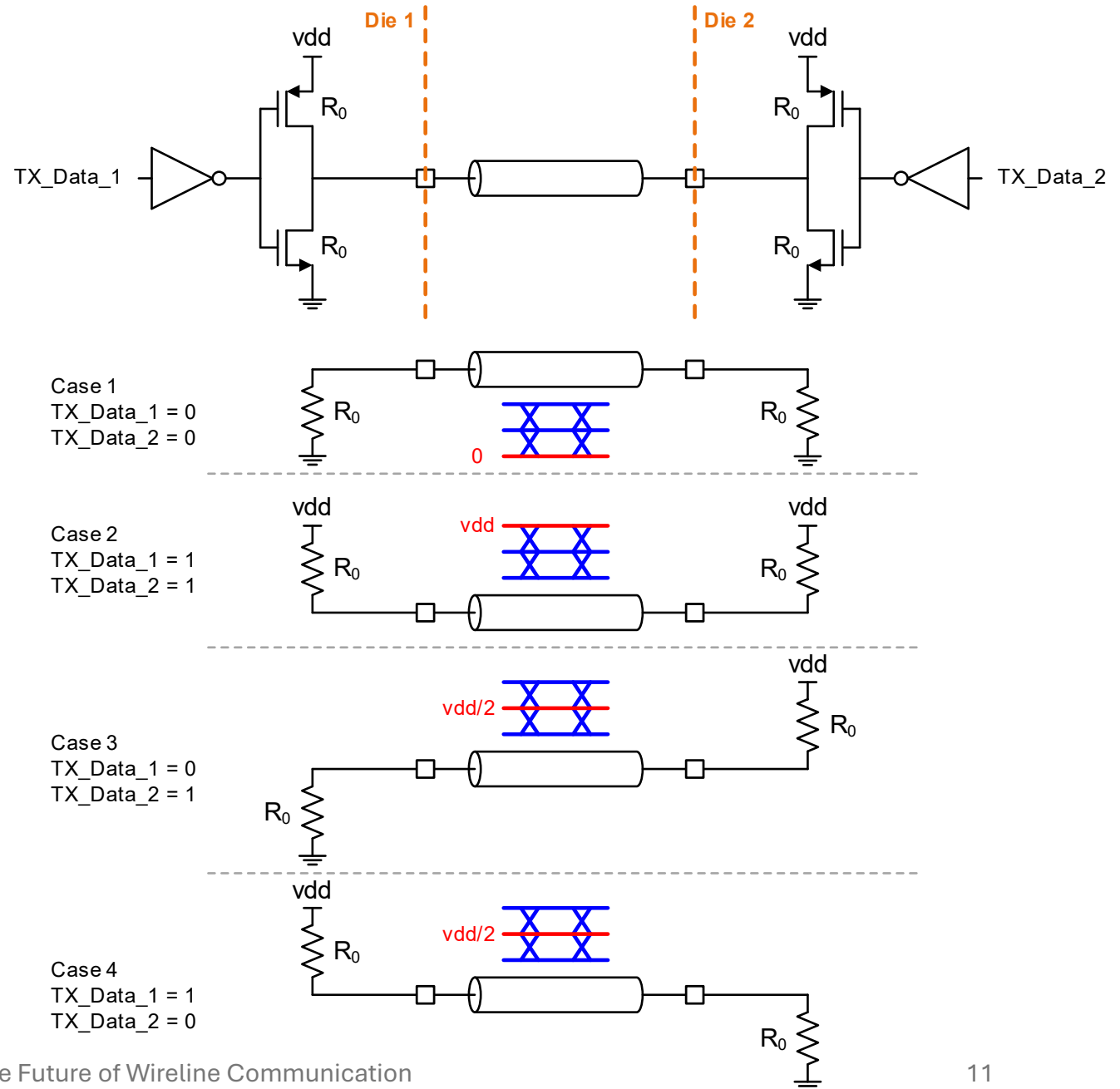
Current Trends: Simultaneous bidirectional



- To achieve higher data rates, simultaneous bidirectional (SBD) signaling **doubles the bandwidth** per wire at the same baud rate at the cost of halving the amplitude
- It works by driving NRZ data from both ends of the channel simultaneously
- The channel carries the sum of the two NRZ data streams, creating a PAM-3-like waveform
- Each receiver must subtract its own transmitted signal to recover the remote data

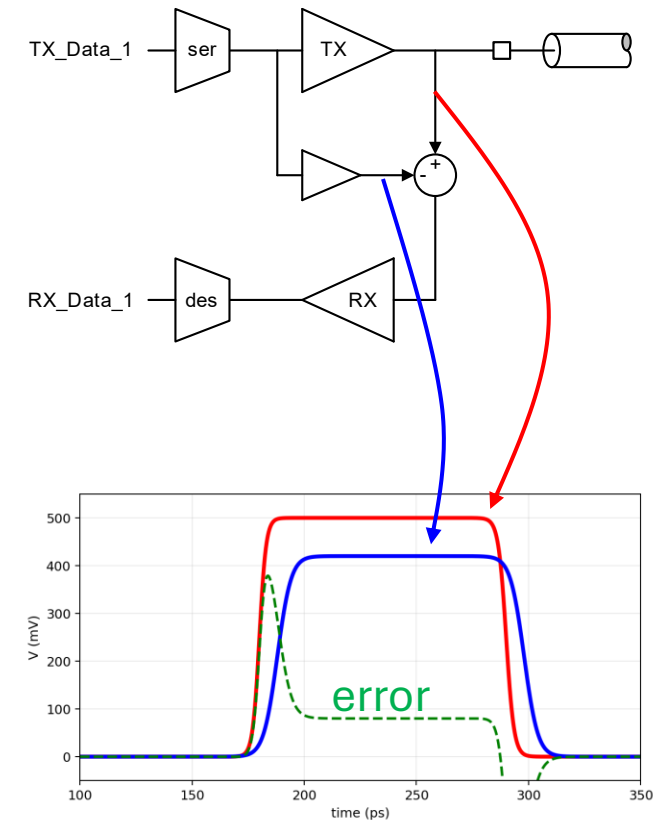
Current Trends: SBD

- Shows only the transmitters on both ends of the channel, other components are ignored for simplicity
- Each transmitted bit pair (00, 01, 10, 11) maps to one of three amplitude levels, forming a PAM-3-like signal



SBD: Local TX Subtraction

- If we look at the pulse response of the local transmitted signal and the replica signal, they will not match
- Imperfect subtraction at the receiver input gives rise to an error signal which degrades eye-opening
- Therefore, the replica path gain, delay, and bandwidth must be calibrated to minimize error



*Along with local TX subtraction, a feature unique to SBD, the receiver can additionally incorporate near-end and far-end echo and crosstalk.

D2D Interface at ISSCC

- The ISSCC 2026 wireline session highlights the rapidly growing interest in die-to-die interconnects across both industry and academia

**A 48Gb/s/lane 1.24Tb/s/mm UCle-Compliant
Die-to-Die Link Over 30mm Standard Package**

Susnata Mondal
Junyi Qiu, Soumya Bc
**A 32Gb/s 12.35Tb/s/mm² 0.36pJ/b UCle-Like
Die-to-Die Interface Featuring Edge-Triggered
Transceivers in 3nm with Active LSI Packaging**

**Wei-Chih Chen¹, H
Jie Huar**
Chin-Hua Wen¹,
Sheng-Tsun,
Tsong-Hsien Tsi
**8.3 A 0.23pJ/b 24Gb/s Modular D2D
Interface With Zero Wake Penalty Clock
Gating in 3nm**

Ravi Shivnar,
Socrates Vamv,
Liu¹, Marina Sali
Aws Shallal², Kai

**A 112Gb/s/wire Single-Ended
Simultaneous Bi-Directional
Transceiver with Dynamic Equalizer
for Die-to-Die Interface in 28nm CMOS**

Zhiwen Huang^{*1}, Zhifei Wang^{*1}, Bingyi Ye², Tianchen Ye¹,
Dunshan Yu¹, Wei Wang^{1,3}, Weixin Gai^{1,3}

SBD D2D Interface at Marvell®

- “Marvell Unveils Industry’s First 64 Gbps/wire Bi-Directional Die-to-Die Interface IP in 2nm to Power Next Generation XPU’s”
 - 32 Gbps per direction of true full-duplex connectivity over a single wire
 - Bandwidth density exceeding 30 Tbps/mm
 - Minimal depth PHY architecture that reduces compute-die area overhead
 - Advanced adaptive power-management techniques for optimal performance-per-watt
 - Redundant lanes with automatic lane repair to enhance yield and reduce BER

Conclusion

- Chiplet architectures are essential to scaling compute performance
- Advances in packaging and die-to-die interconnects unlock higher bandwidth, lower power, and greater integration density
- Industry momentum is accelerating, with both academia and companies pushing next-generation high-speed D2D IP
- The future of compute is heterogeneous integration, and die-to-die interconnects are the glue holding it together

References

1. OJSSCS 2024, High-Bandwidth Chiplet Interconnects for Advanced Packaging Technologies in AI ML Applications: Challenges and Solutions
2. ISSCC 2015, The future of IC design innovation
3. JSSC 2025, A 32 Gb/s 0.36 pJ/bit 3 nm Chiplet IO Using 2.5-D CoWoS Package With Real-Time and Per-Lane CDR and Bathtub Monitoring

Thank you

