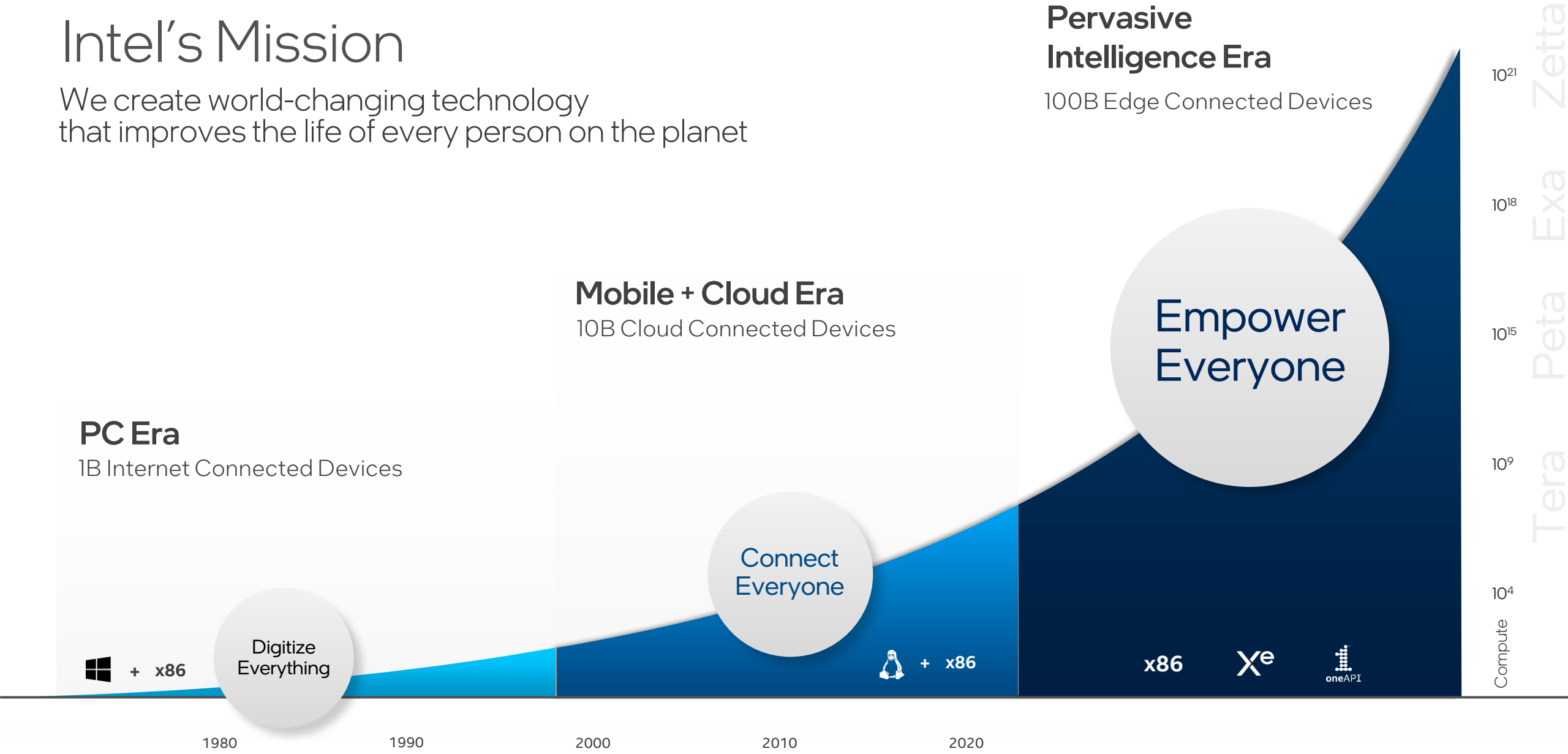
Intel Corporation

Meteor Lake and Arrow Lake Intel Next-Gen 3D Client Architecture Platform with Foveros

Wilfred Gomes, Slade Morgan, Boyd Phelps, Tim Wilson, Erik Hallnor

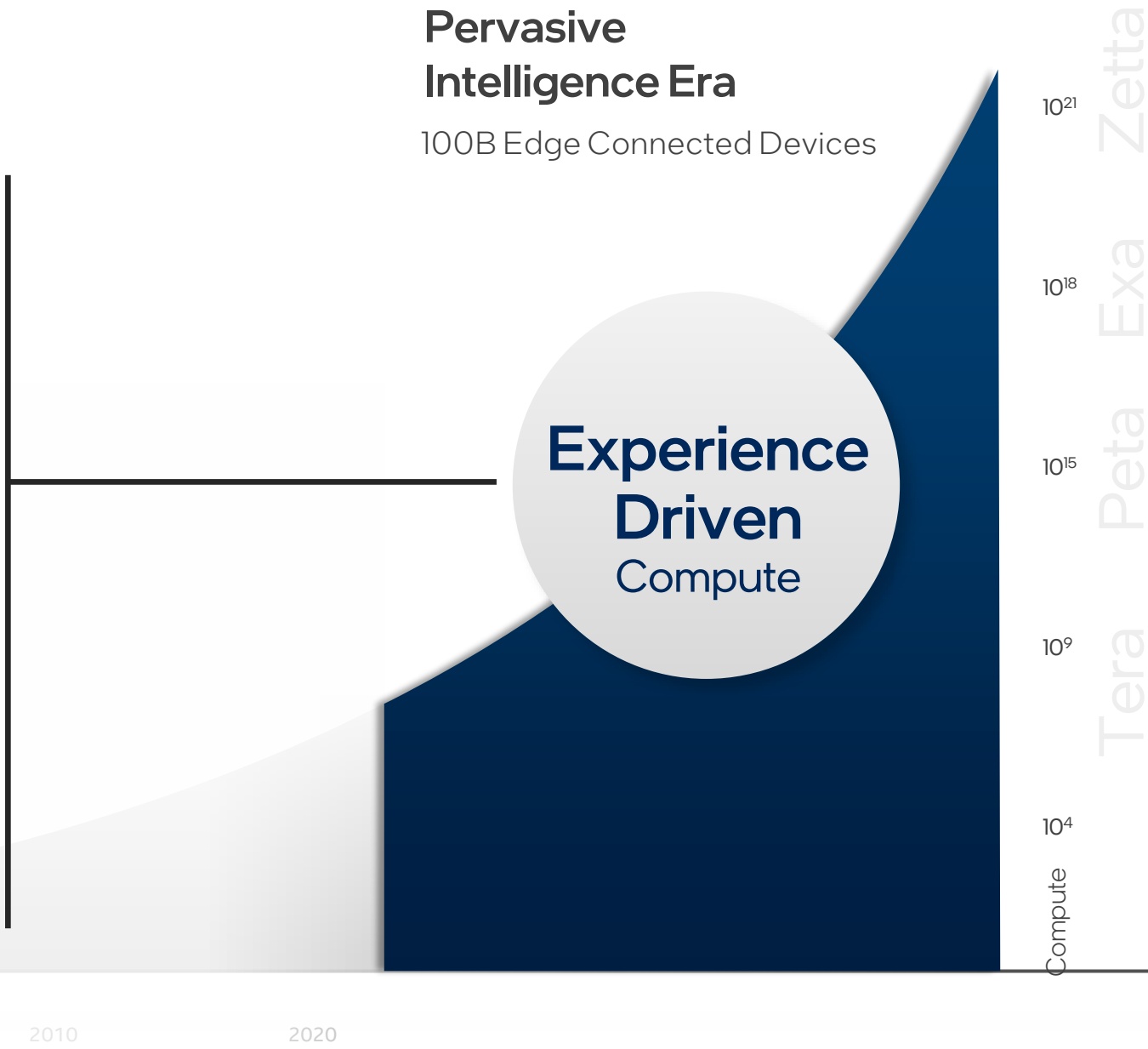
Intel's Mission

We create world-changing technology
that improves the life of every person on the planet



Experience Driven Client

- **Experience First**
- **Purposeful Performance**
- **Dynamic**
- **Scale**

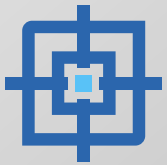


Experience Driven

Experience First

Purposeful Performance

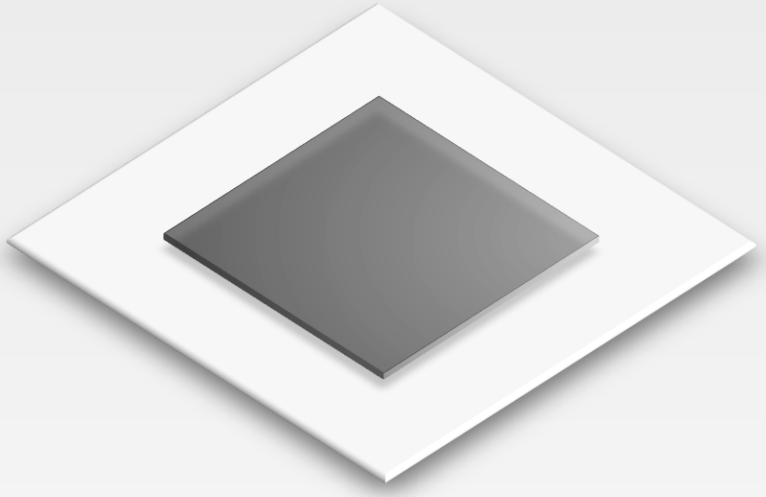
Dynamic

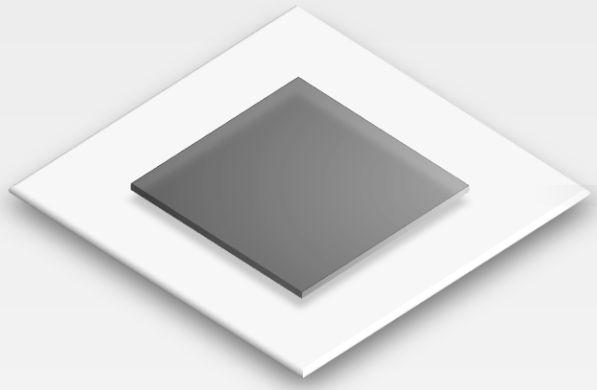
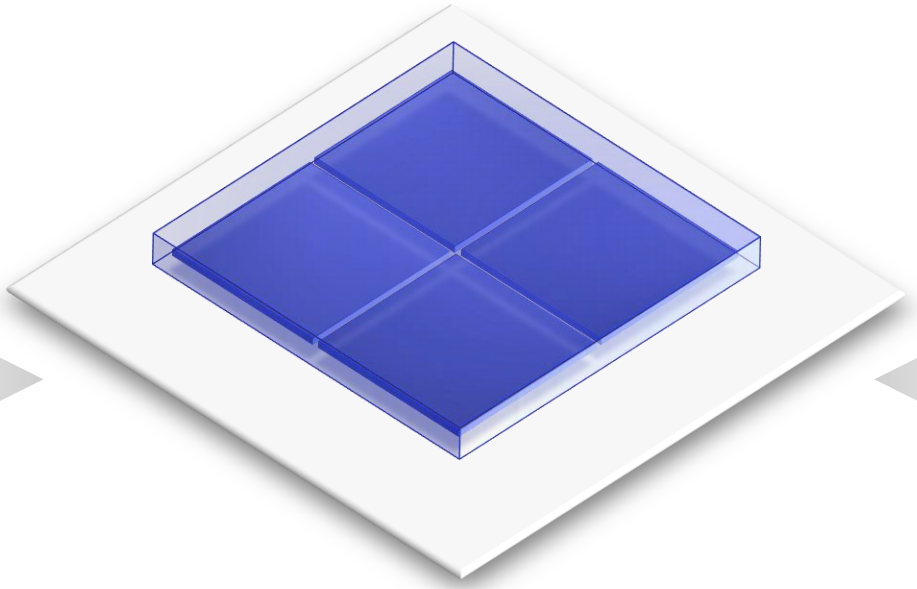
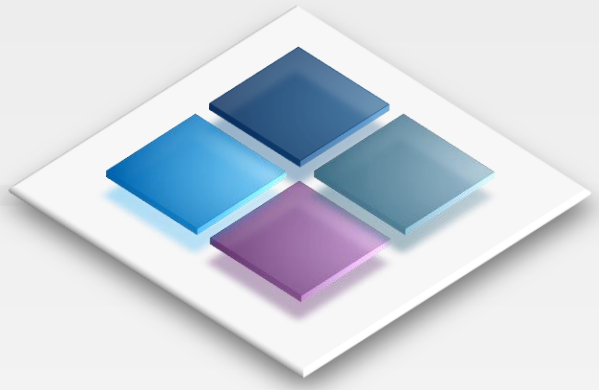


Scale

Implications for Client

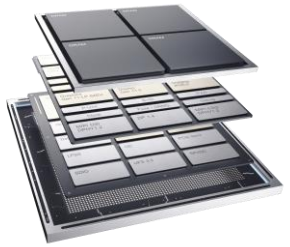
Performance	Performance, Perf/Watt
Flexibility	Mix & Match Blocks and Functions
Innovation Pace	Time-to-Market
Next Exponential	The next generation of devices

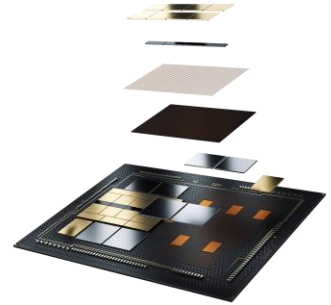
Monolithic		Disaggregated
		
Highest	**Performance**	Lower (tax on latency, power, B/W)
Very Limited	**Flexibility**	**Limited**
Slow (per SOC basis)	**Innovation Pace**	**Faster (release per new function)**
Low	**Scale**	**Higher**

Monolithic		Disaggregated
		
Highest		Lower (tax on latency, power, B/W)
Very Limited		**Limited**
Slow (per SOC basis)		**Faster (release per new function)**
Low		**Higher**

Can we get **monolithic** performance with **disaggregated** architecture benefits?

Disaggregation Journey

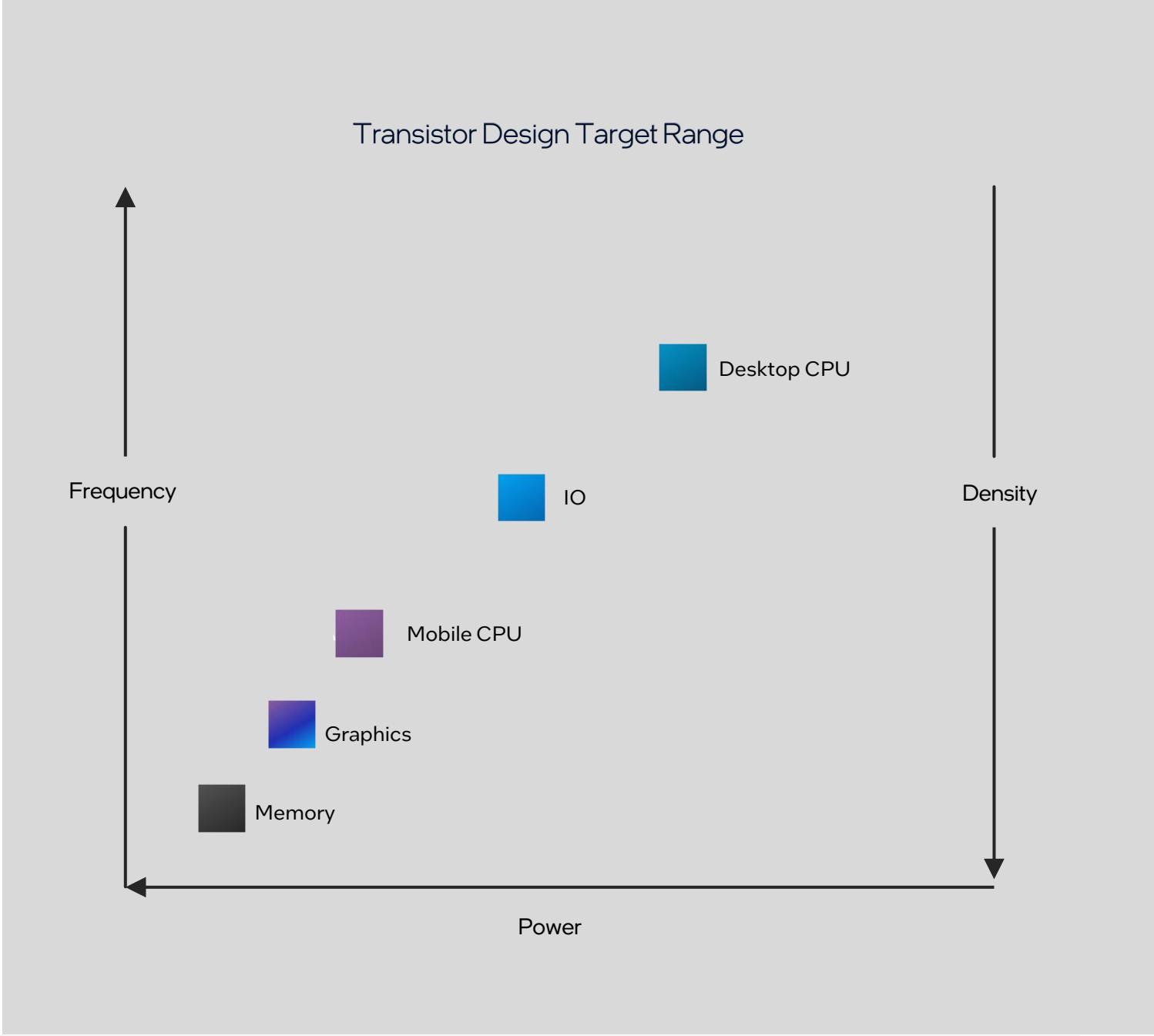




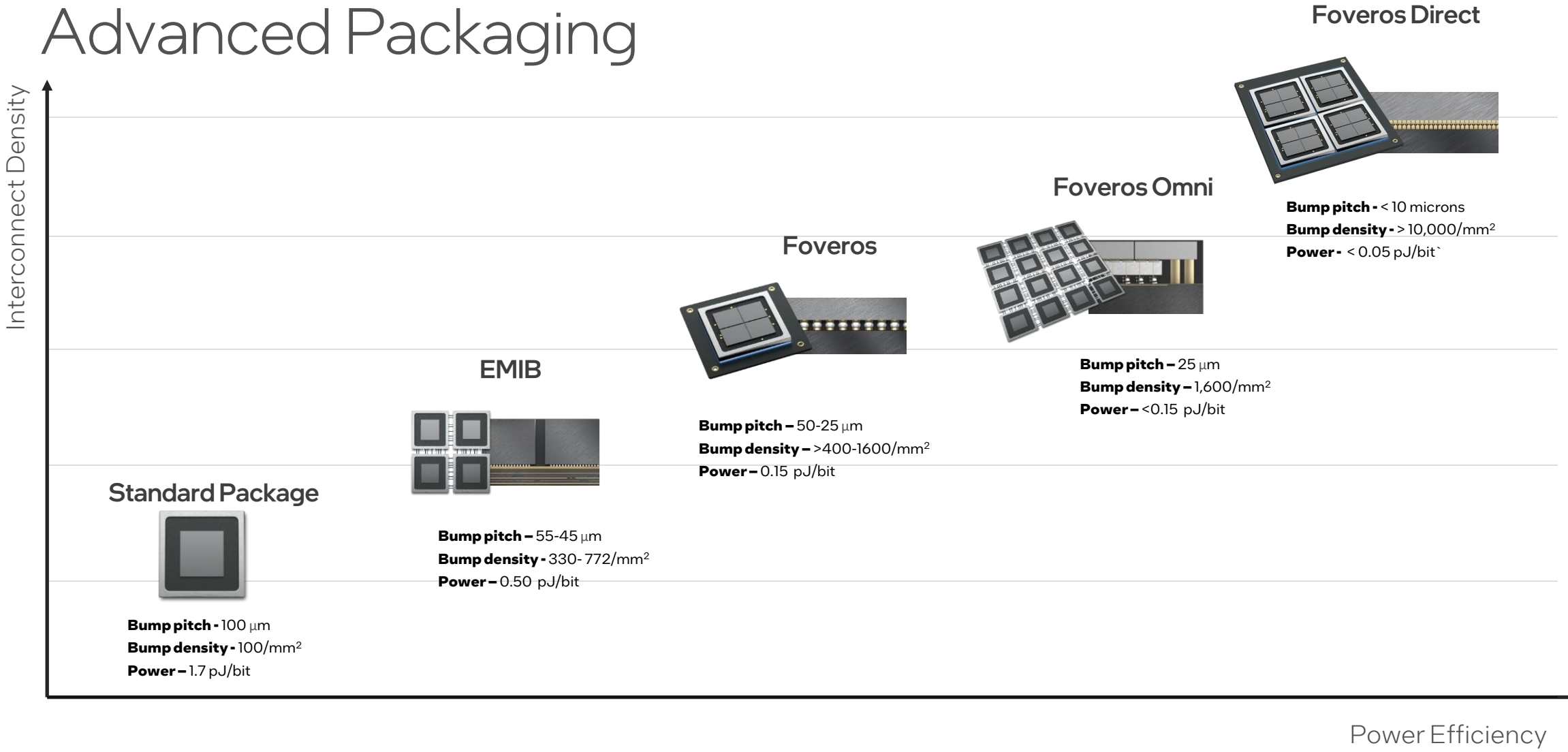
	Haswell	Kaby Lake G	Lakefield	Ponte Vecchio
	Ultra Thin & Light	Ultra Thin & Perf Graphics	Ultra Thin & Light	High Density & Performance
	2014	2017	2019	2022
Architecture	CPU/PCH/Memory partitioning	CPU/GFX partitioning	Hybrid Architecture CPU/PCH partitioning	47 Tiles Compute/mem/IO partitioning
Packaging	**2D** MCP	**2.5D + 2D** EMIB + MCP	**3D** 50µm Foveros	**2.5D + 3D** EMIB + 36µm Foveros
Process	Intel 22nm	GloFo 14nm Intel 14nm	Intel 22FFL Intel 10nm	TSMC N7 TSMC N5 Intel 7

Transistor Diversity Opportunity

- SOC Optimized
- Graphics Optimized
- Memory Optimized
- Mobile CPU Optimized
- Desktop CPU Optimized
- I/O Optimized



Advanced Packaging



Advanced Packaging

Interconnect Density

Power Efficiency

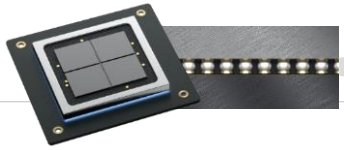
Foveros Direct

Bump pitch - < 10 microns
Bump density - > 10,000/mm²
Power - < 0.05 pJ/bit`

Foveros Omni

Bump pitch – 25 µm
Bump density – 1,600/mm²
Power – <0.15 pJ/bit

Foveros



Bump pitch – 50-25 µm
Bump density – >400-1600/mm²
Power – 0.15 pJ/bit

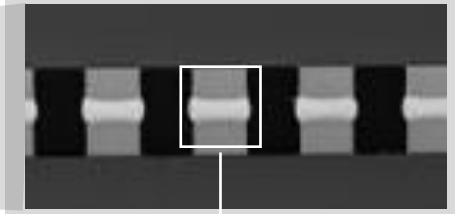
EMIB

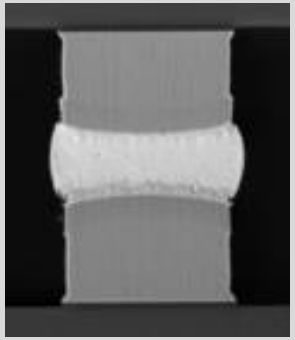
mp pitch – 55-45 µm
mp density - 330- 772/mm²
wer – 0.50 pJ/bit

High Yield & High Volume

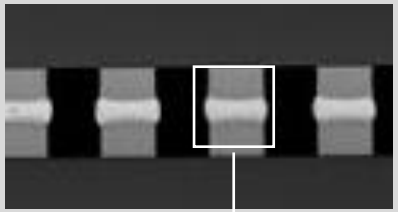
Manufacturing for large number of tiles

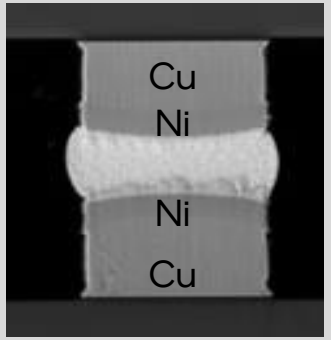
Die Center



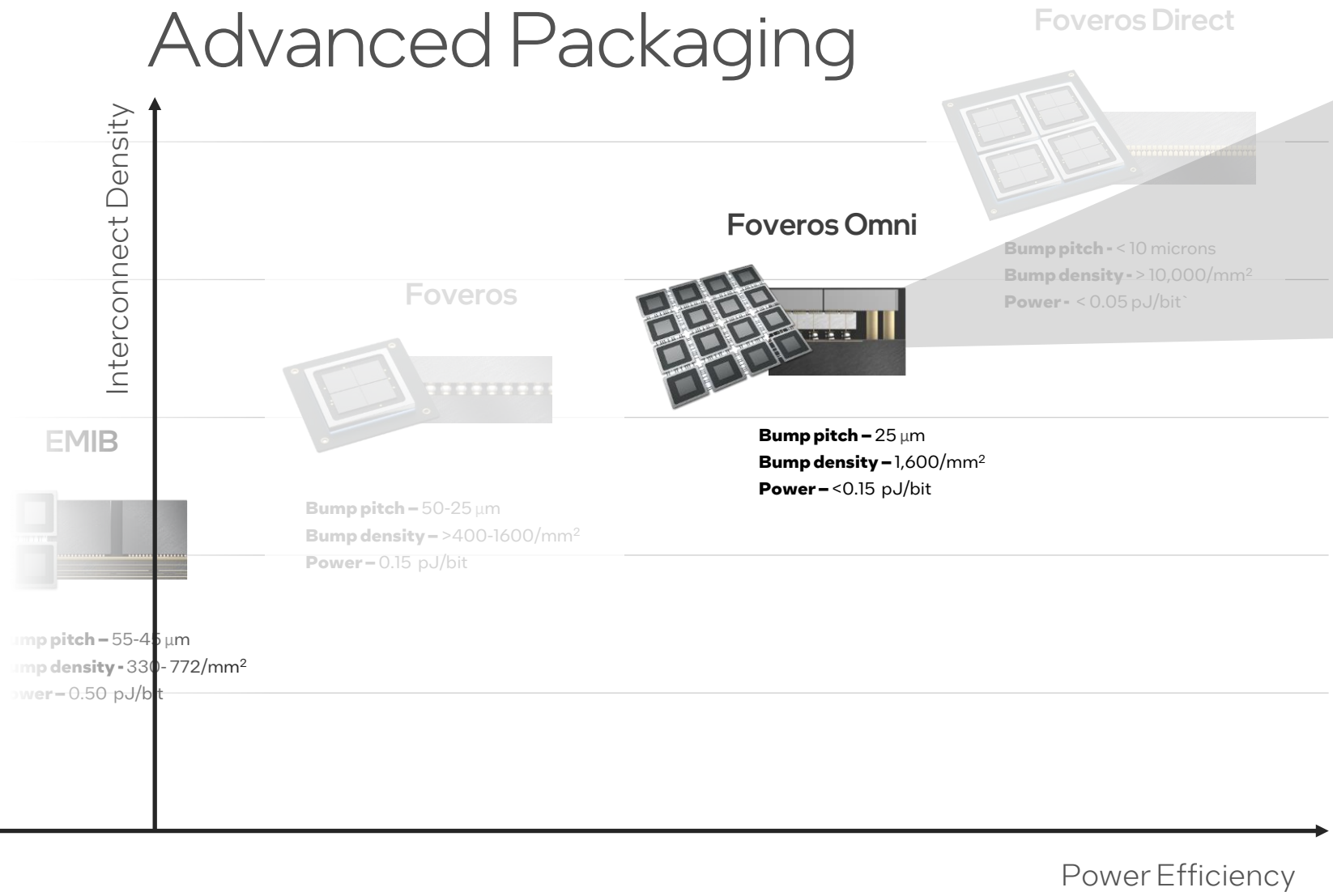


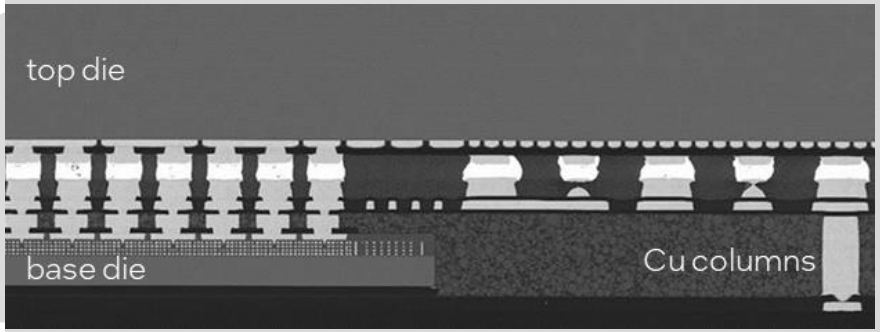
Die Edge





Advanced Packaging





‘Mix and match’ tiles in base die complex

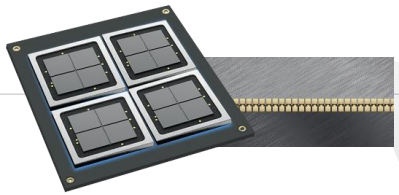
4x higher interconnect bump density vs EMIB

Advanced Packaging

Interconnect Density

Power Efficiency

Foveros Direct



Bump pitch - < 10 microns
Bump density - > 10,000/mm²
Power - < 0.05 pJ/bit`

Foveros Omni

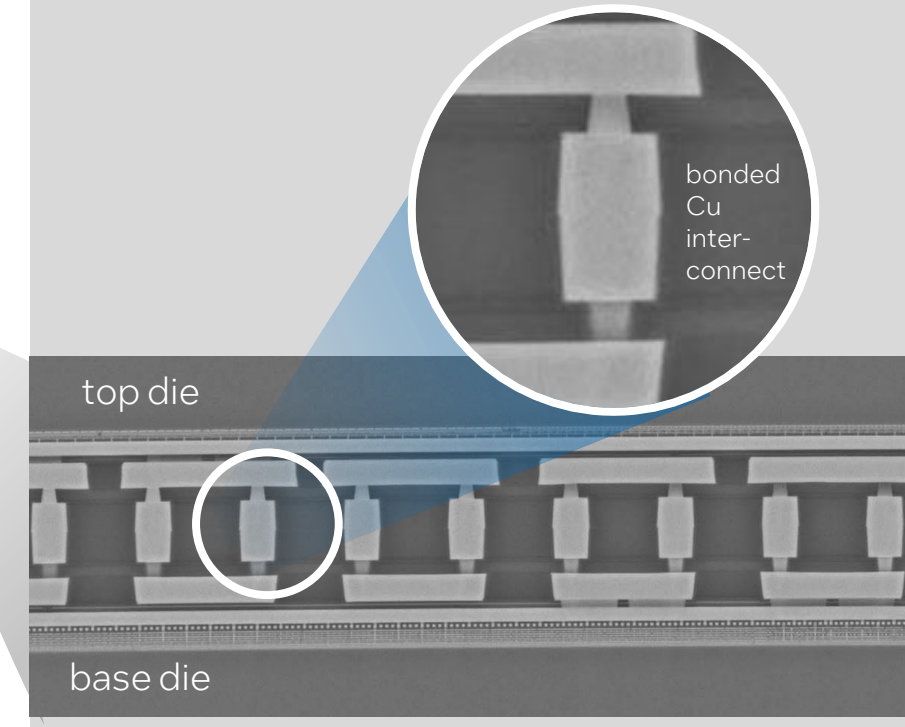
Bump pitch – 25 µm
Bump density – 1,600/mm²
Power – <0.15 pJ/bit

Foveros

Bump pitch – 50-25 µm
Bump density – >400-1600/mm²
Power – 0.15 pJ/bit

EMIB

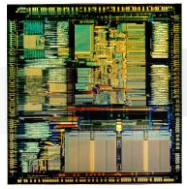
mp pitch – 55-45 µm
mp density - 330- 772/mm²
wer – 0.50 pJ/bit



16x higher interconnect bump density
vs Foveros @ 36 µm pitch

Higher B/W at lower latency, power, & die area

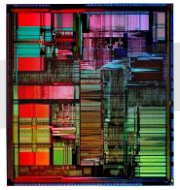
Architecture Evolution



Intel 386

1980

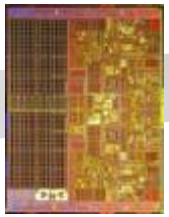
Design with
Schematics
Sea of transistors/ gates



Pentium

1990

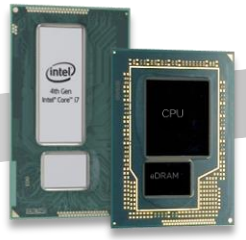
Design with
HDL
Sea of cells/blocks



Core Duo

2005

Design with
IP
functional, units/IPs



Haswell

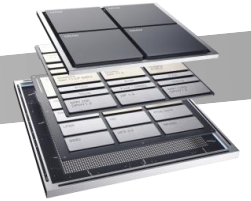
2013

Design with
2D Chiplets

Kaby Lake G

2017

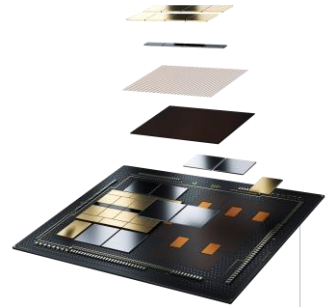
Design with
2D Chiplet and 2.5D



Lakefield

2019

First Commercial
3D Logic on Logic Stacking

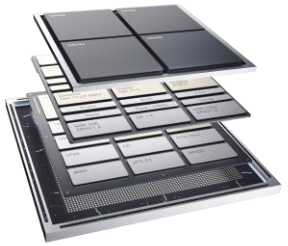


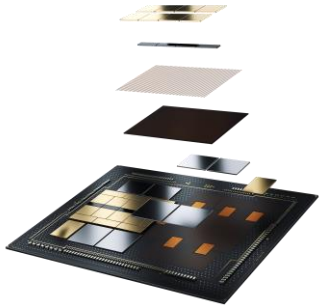
Ponte Vecchio

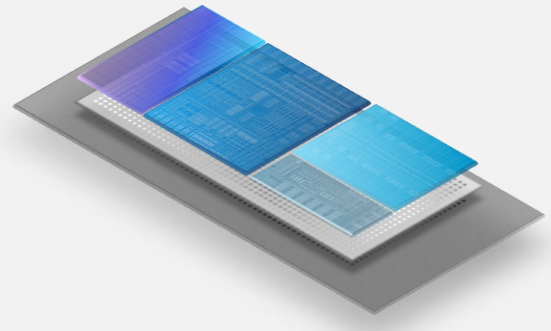
2022

2D+ 3D
Design Large Number of Tiles

Disaggregation Journey







Kaby Lake G
Ultra Thin & Perf Graphics

2017

Lakefield
Ultra Thin & Light

2019

Ponte Vecchio
High Density & Performance

2022

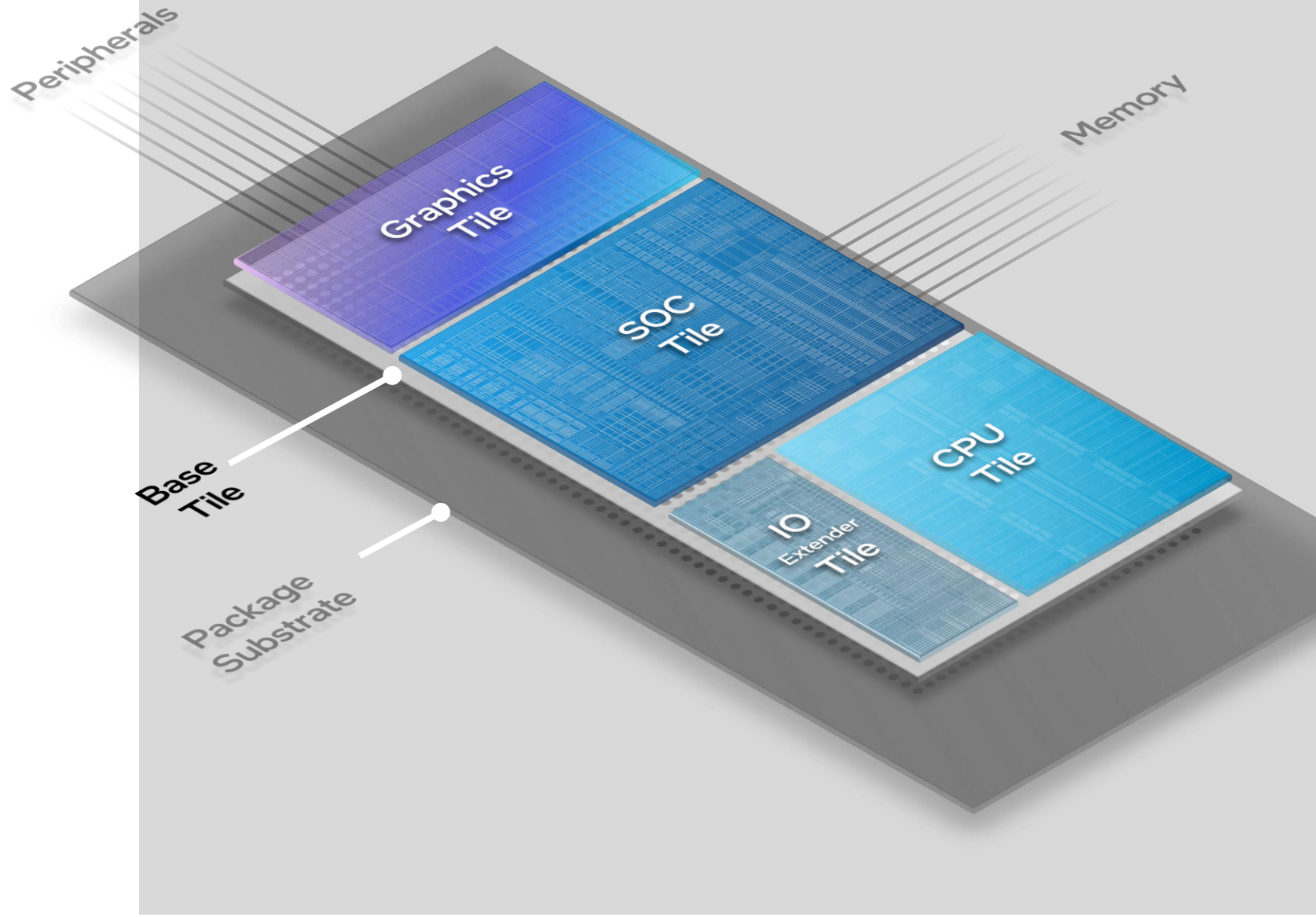
Architecture	CPU/GFX partitioning	Hybrid Architecture CPU/PCH partitioning	47 Tiles Compute/Memory/IO partitioning
Packaging	**2.5D + 2D** EMIB + MCP	**3D** 50µm Foveros	**2.5D + 3D** EMIB + 36µm Foveros
Process	GloFo 14nm Intel 14nm	Intel 22FFL Intel 10nm	TSMC N7 TSMC N5 Intel 7

Meteor Lake

Next Step in our Disaggregation Journey

Architecture and Design tradeoffs

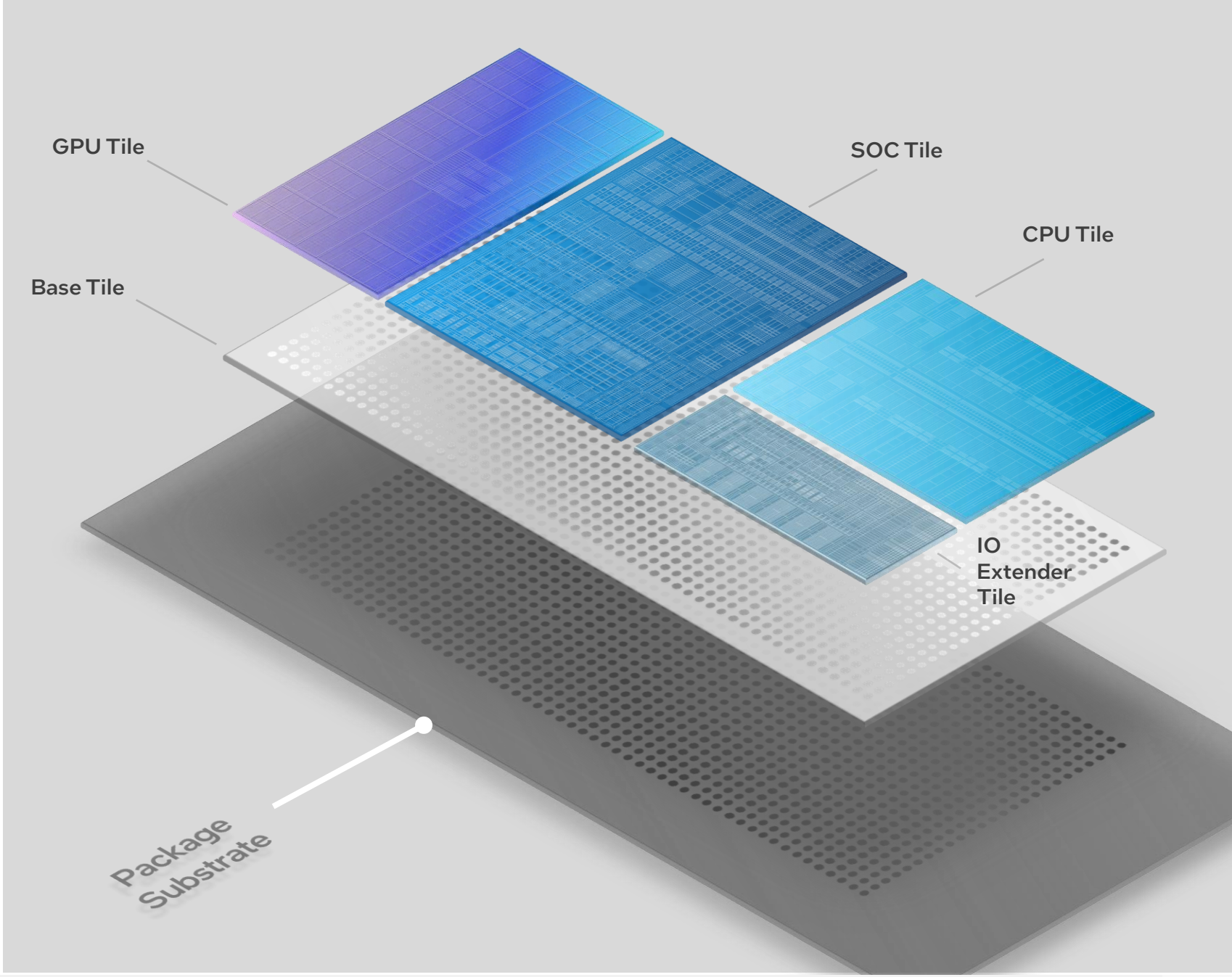
New Flexible Tiled Architecture



Scalable Architecture

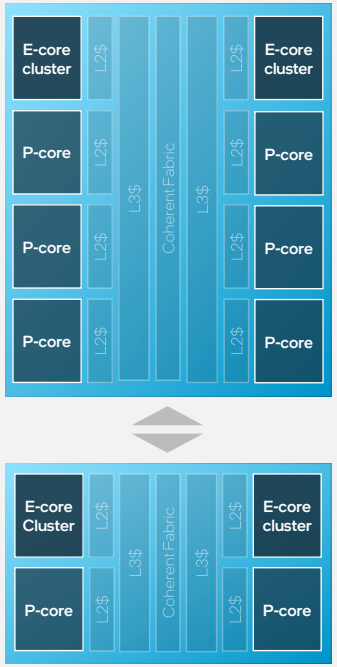
Goals

- Flexibility with core arch, count, process
- Flexibility with Graphics cores
- IO modularity
- Process node flexibility
- Ability to scale graphics and compute
- Low power to discrete graphics performance

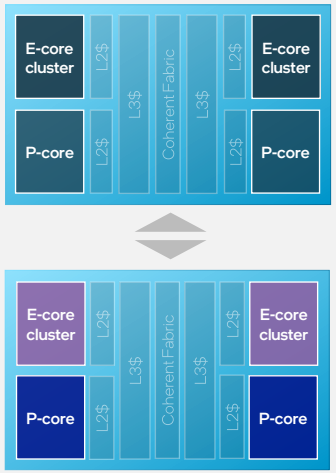


Compute Tile

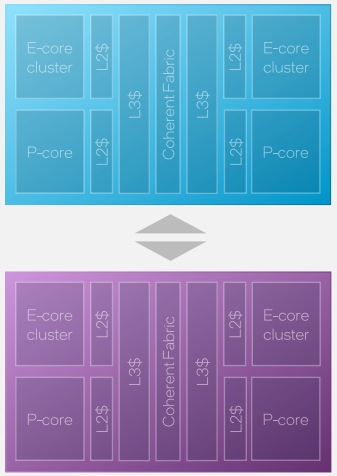
Core Count Scalability



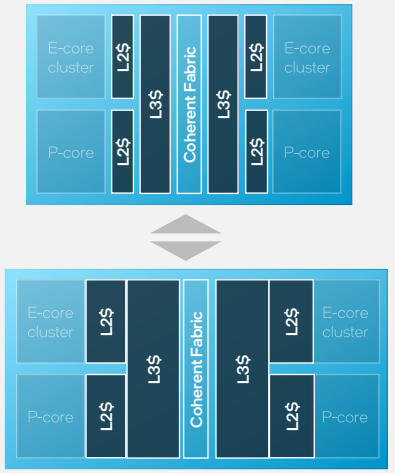
Core Generation Scalability

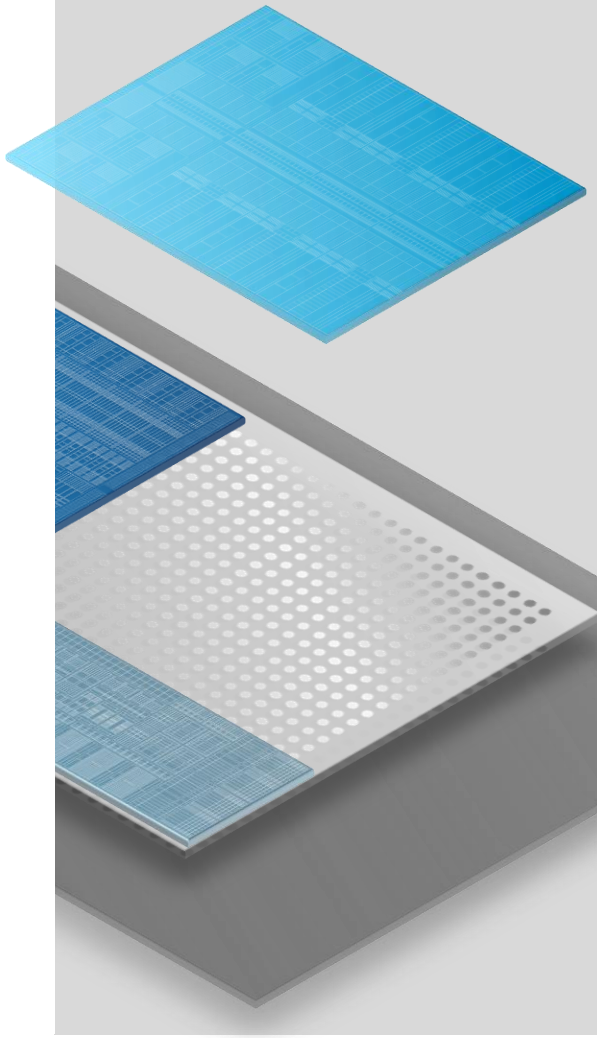


Node Scalability



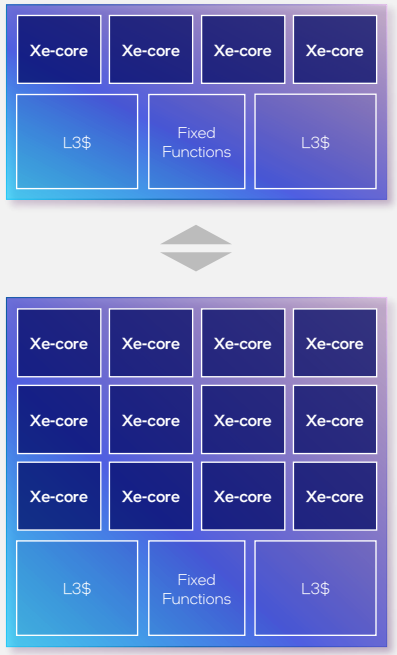
Cache Scalability



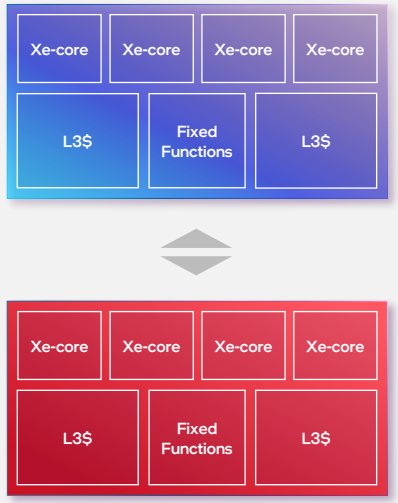


Graphics Tile

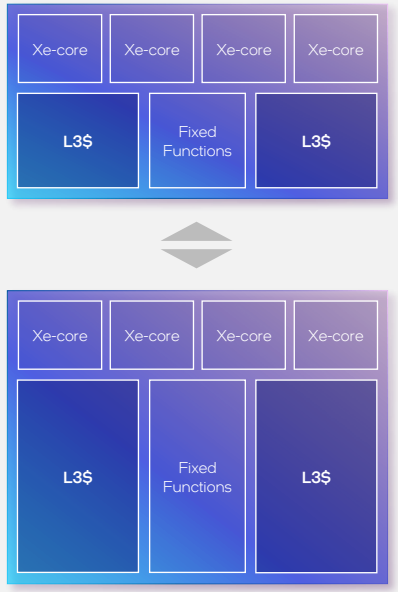
Core Count Scalability

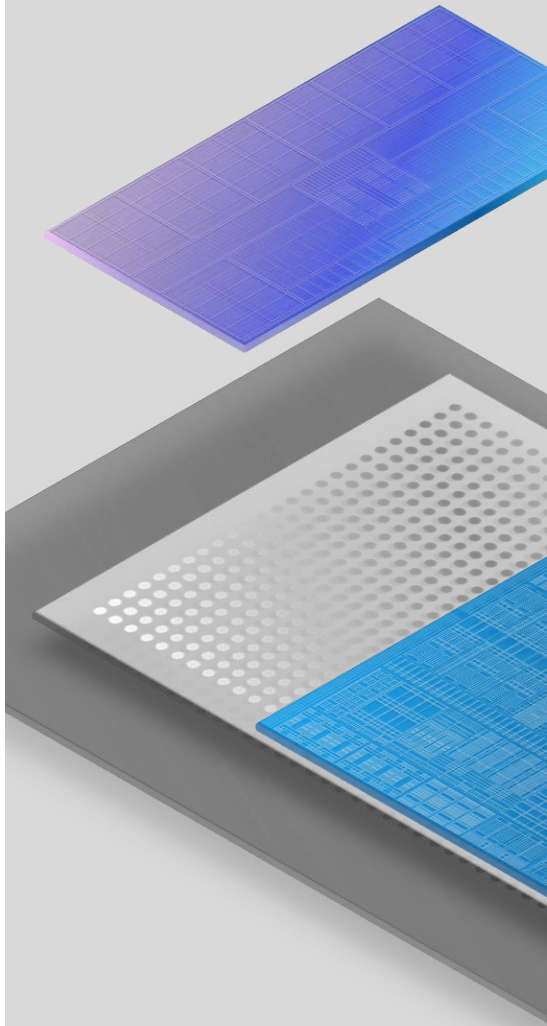


Node Scalability



Cache Scalability

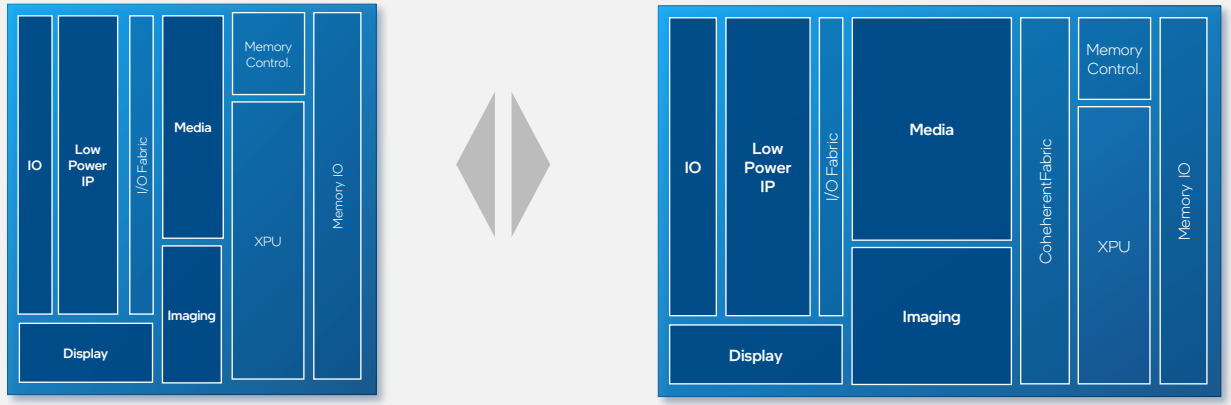


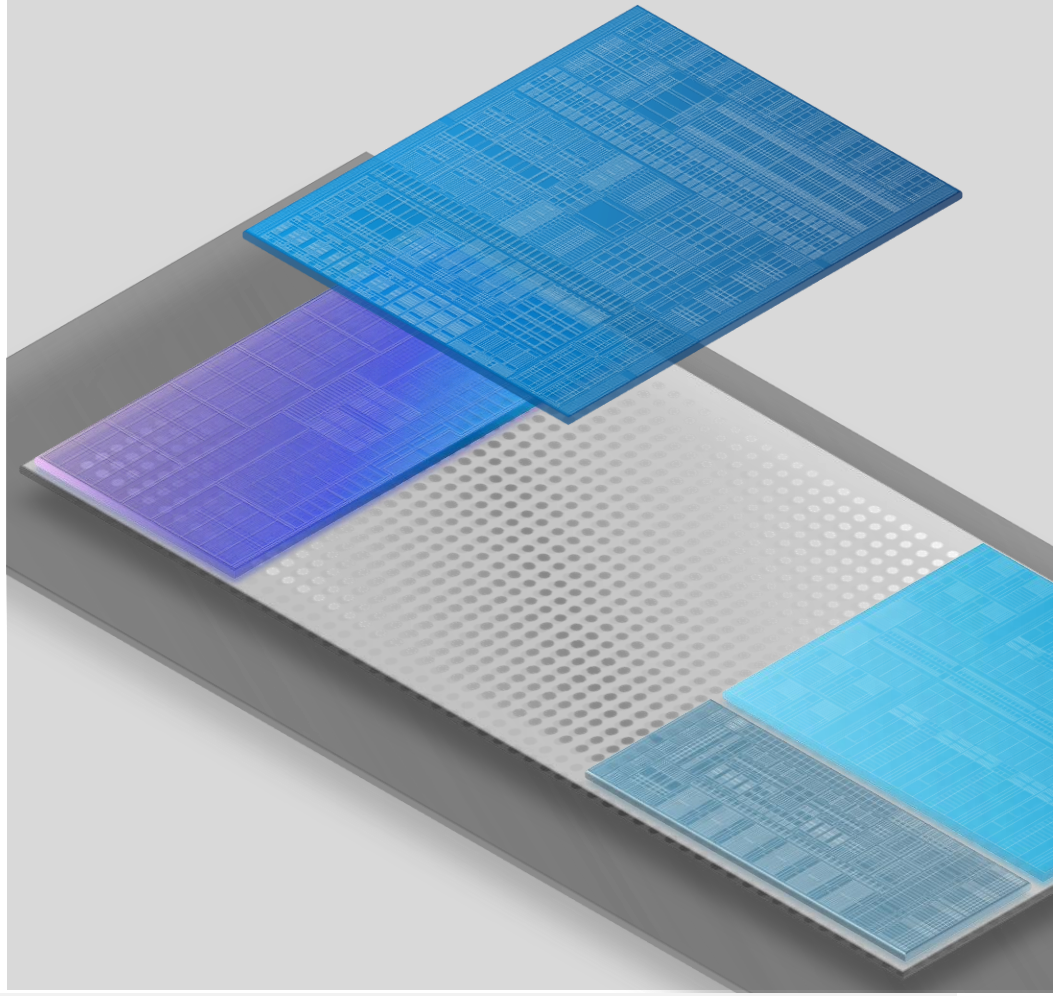


SOC Tile

Scalable IP Blocks

Low power IP	SRAM	High Voltage	IO

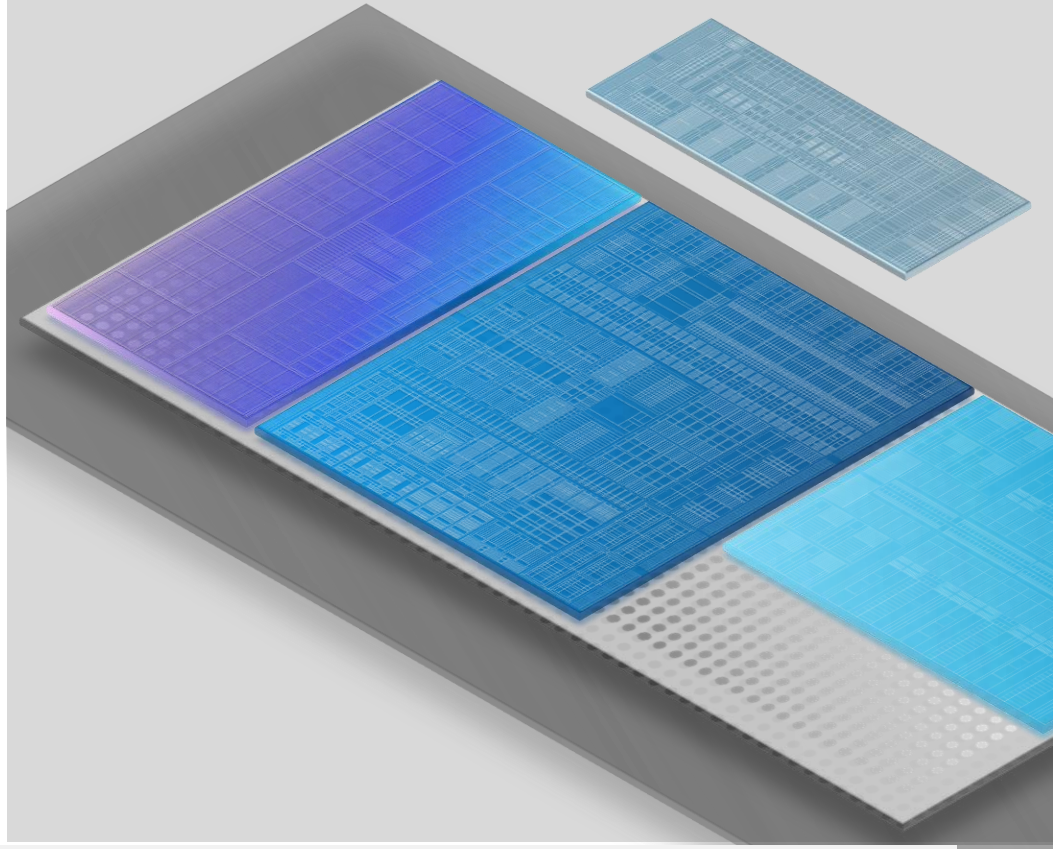




I/O Extender Tile

Scalable I/O Blocks

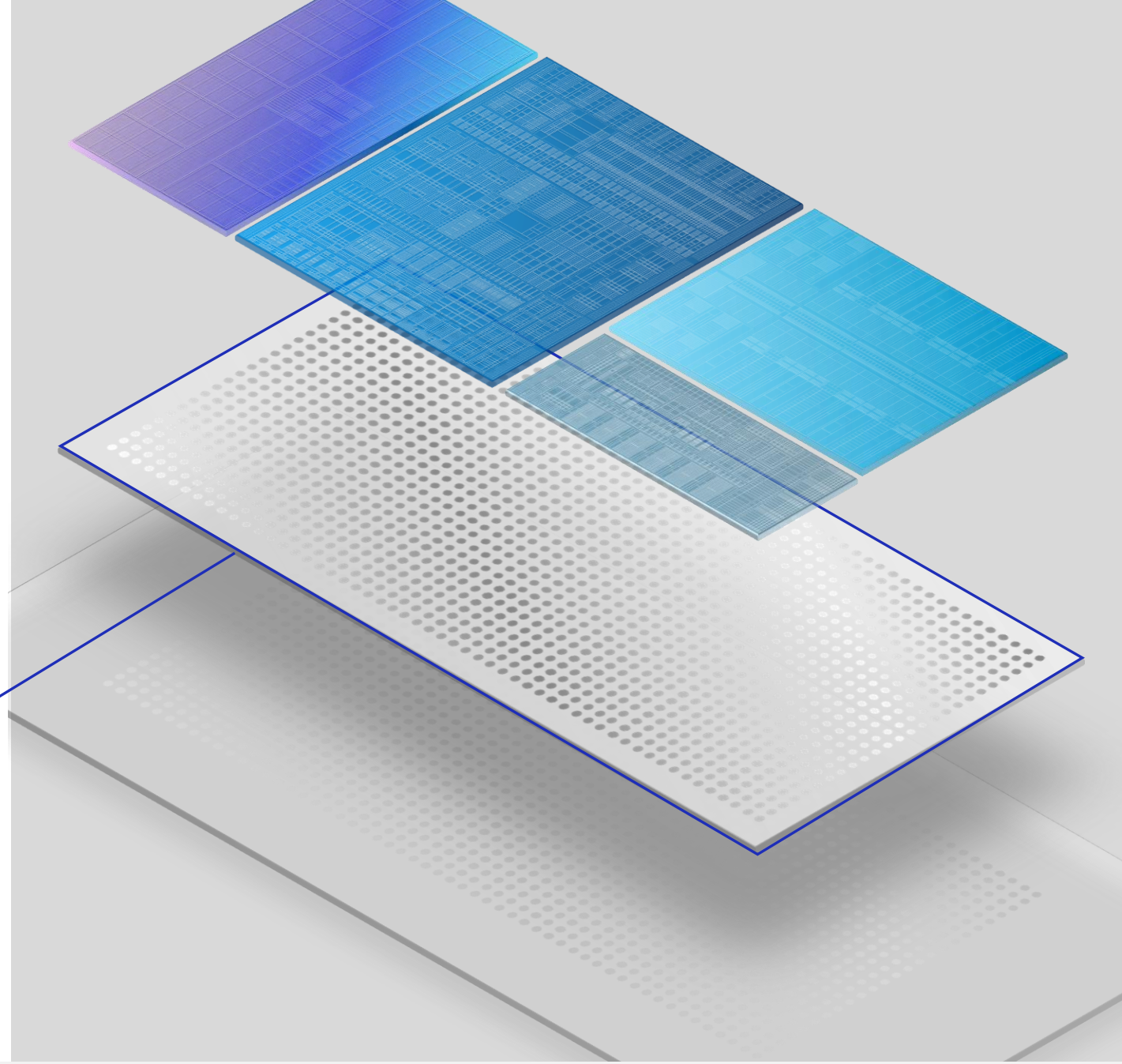
Number of Lanes	Bandwidth	Protocol	Speed



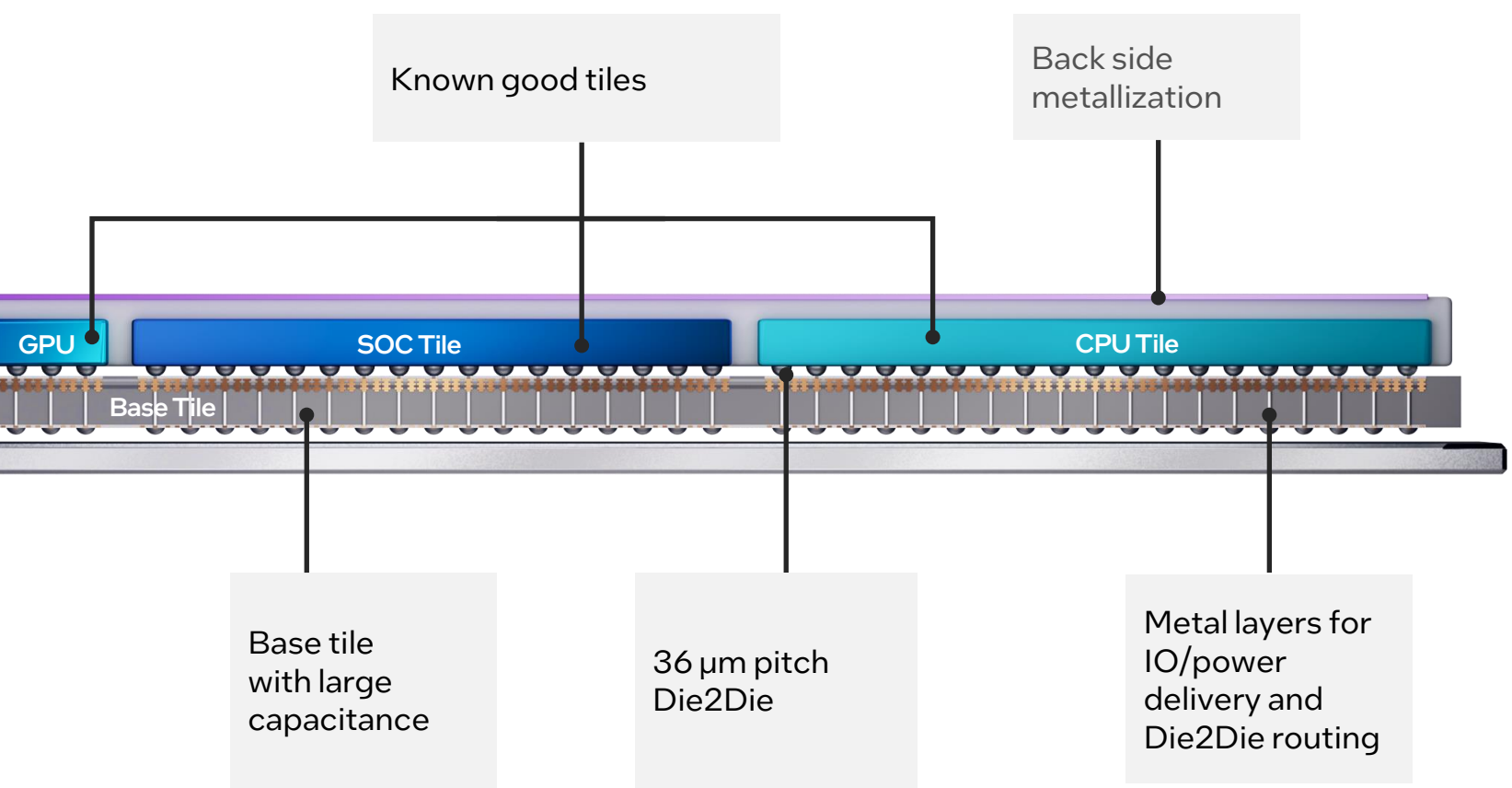
Meteor Lake

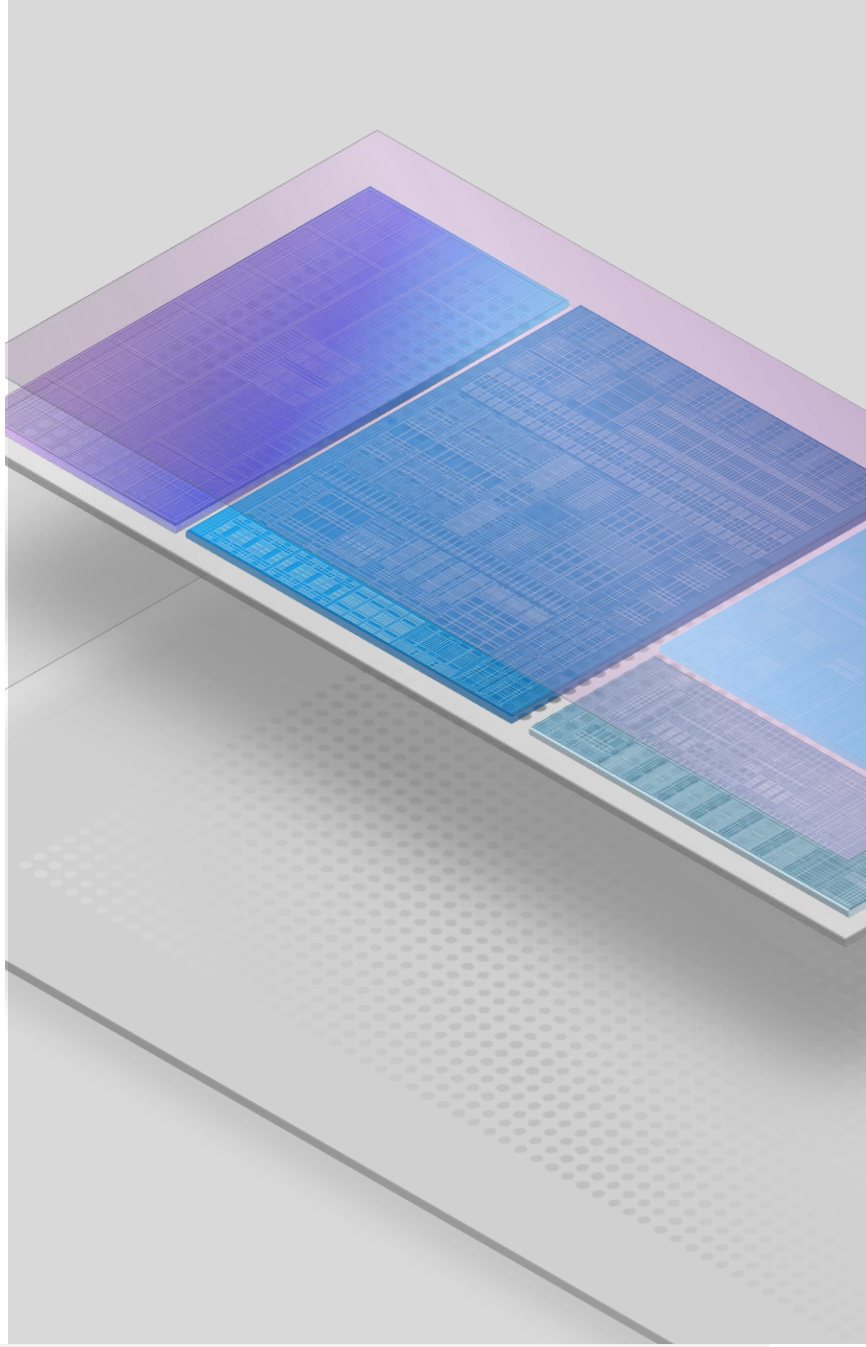
Construction

Base Tile



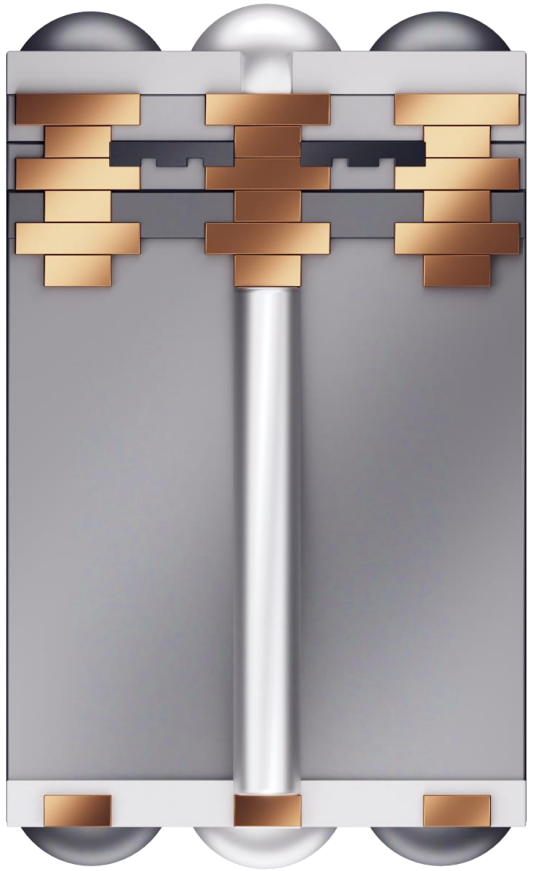
Meteor Lake


Known good tiles
Back side metallization
GPU
SOC Tile
CPU Tile
Base Tile
Base tile with large capacitance
36 µm pitch Die2Die
Metal layers for IO/power delivery and Die2Die routing



Meteor Lake

Base Tile



3D Capacitors

Package Bump

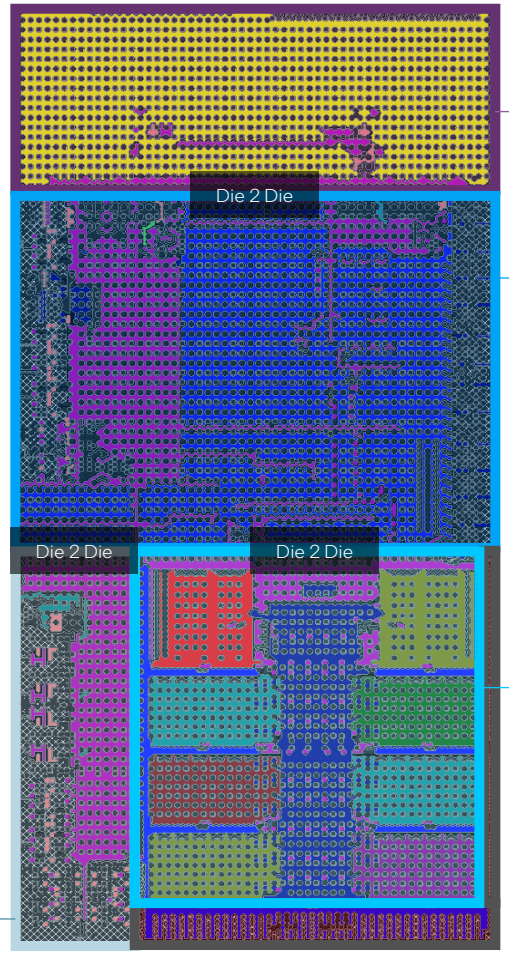
Die2Die power delivery, package IO routing

Modularity with active silicon for memory and logic

Redistribution layers with active silicon

Meteor Lake

Colors Represent
3D Capacitors,
Voltage Islands

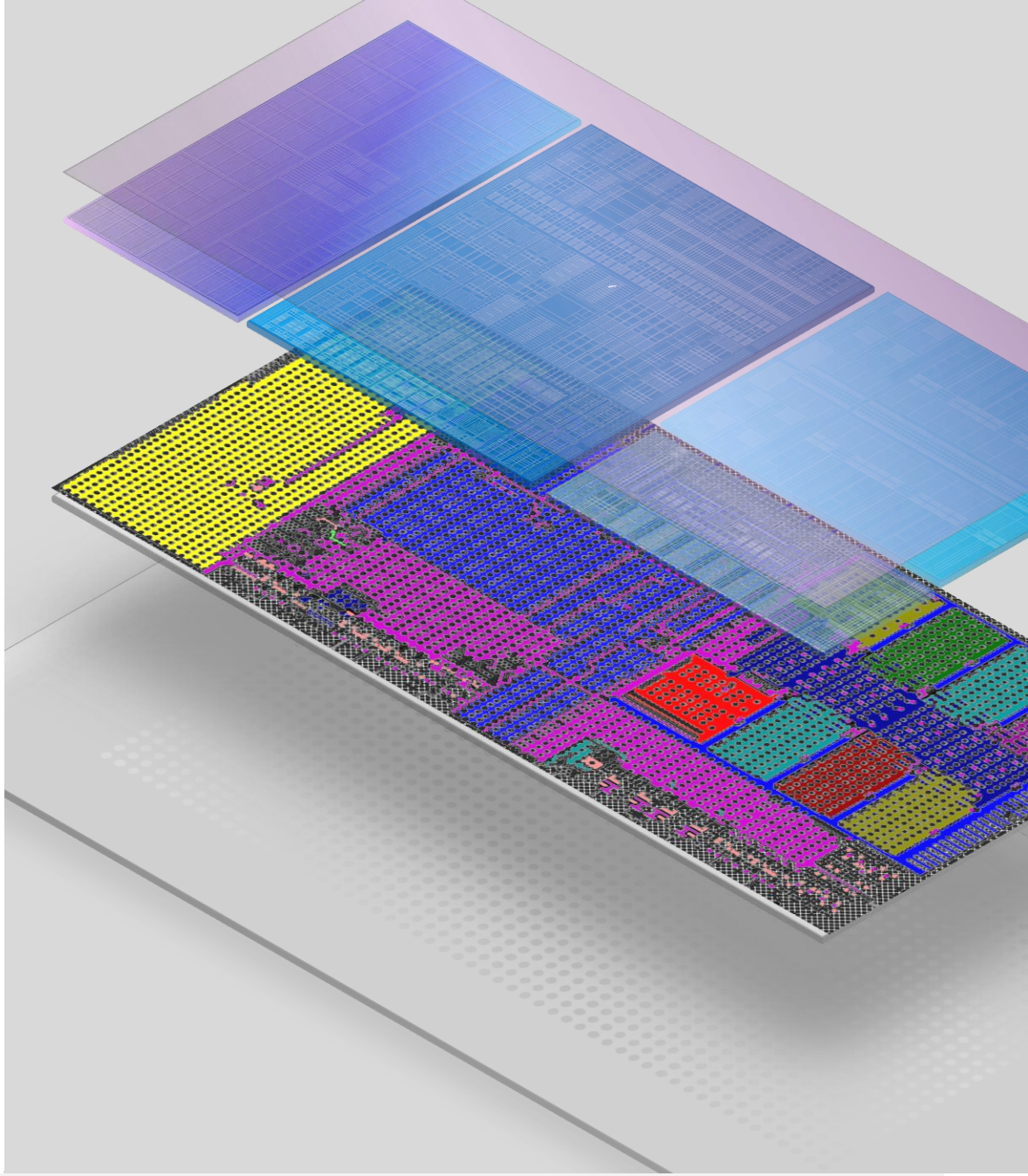


Graphics Tile

SOC Tile

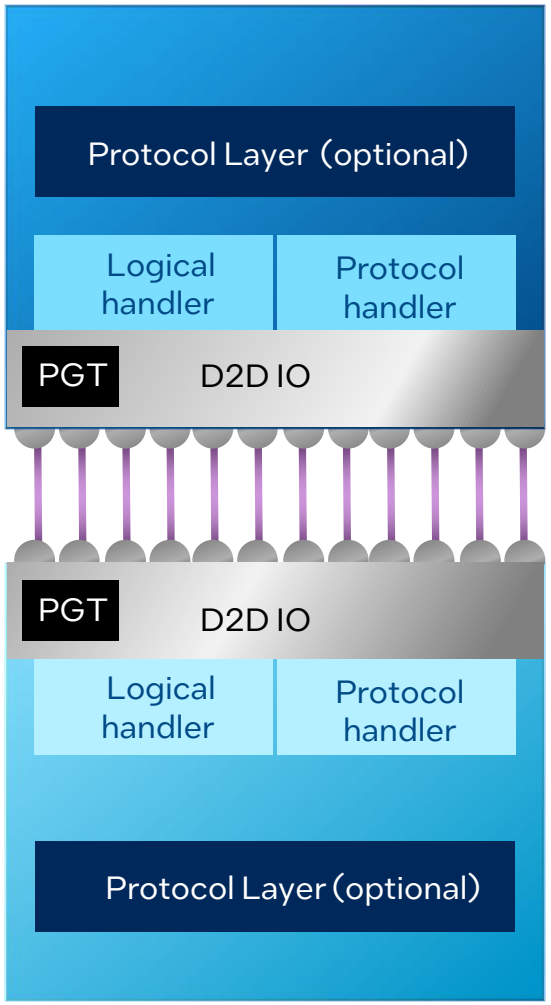
CPU Tile

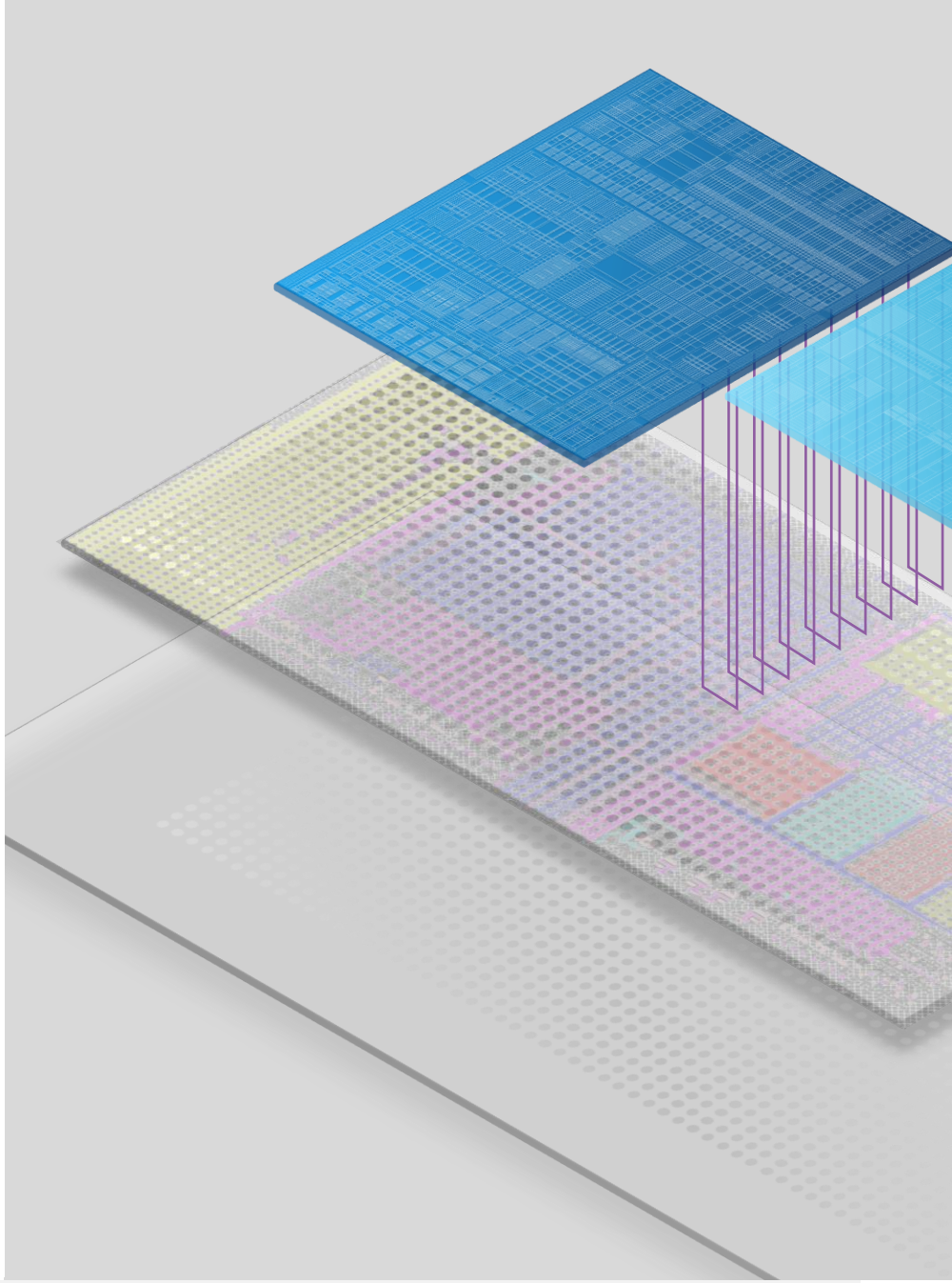
IOE Tile



FDI - Foveros Die Interconnect

- Low Voltage CMOS interface
- High Bandwidth, Low Latency
- Synchronous and asynchronous signaling
- Low area overhead
- Operation @ 2 Ghz, 0.15 – 0.3 pJ/bit

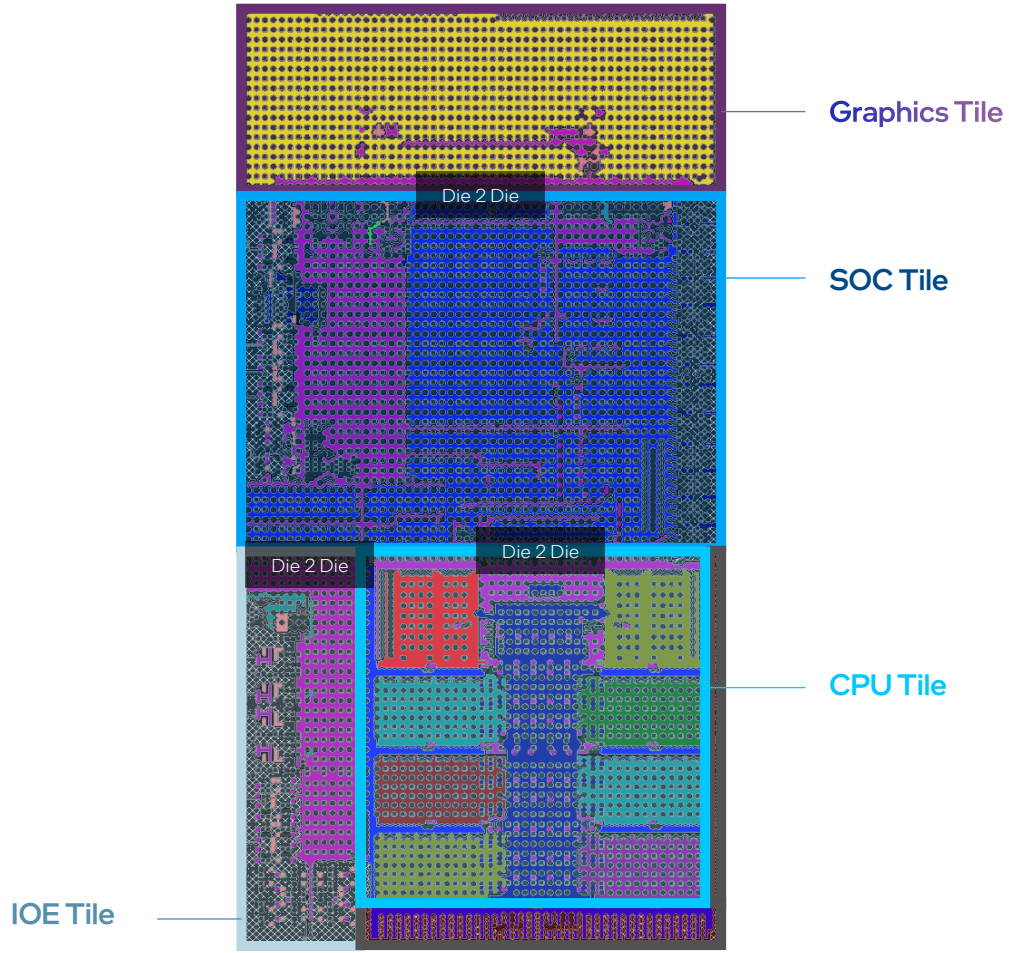




Meteor Lake

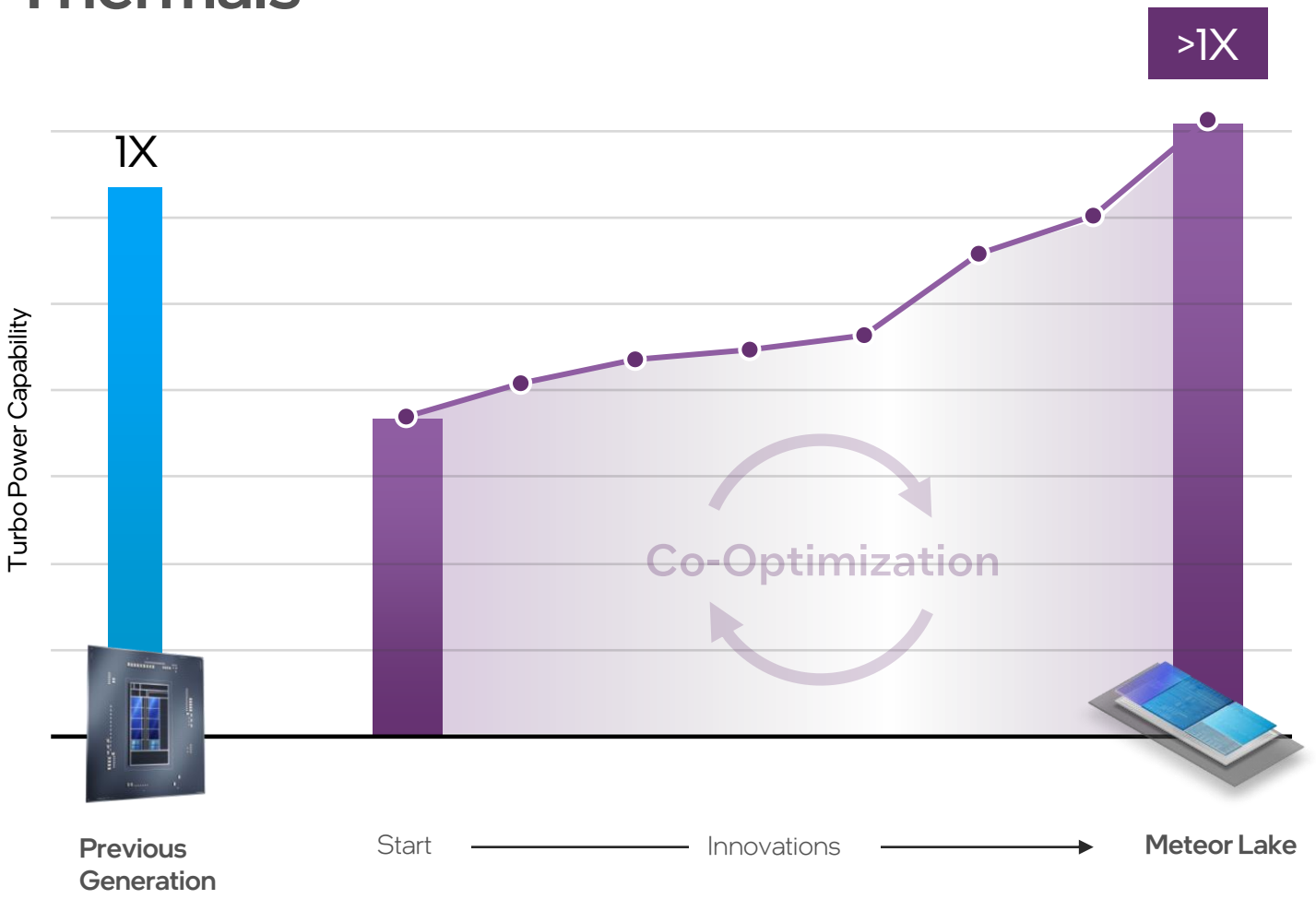
Interconnect

Link	Mainband width	Mainband Protocol
CPU - SoC	~2K	2x IDI
Graphics - SoC	~2K	2x iCXL
SoC - IOE	~1K	IOSF, 4x Display Port



Meteor Lake

Thermals



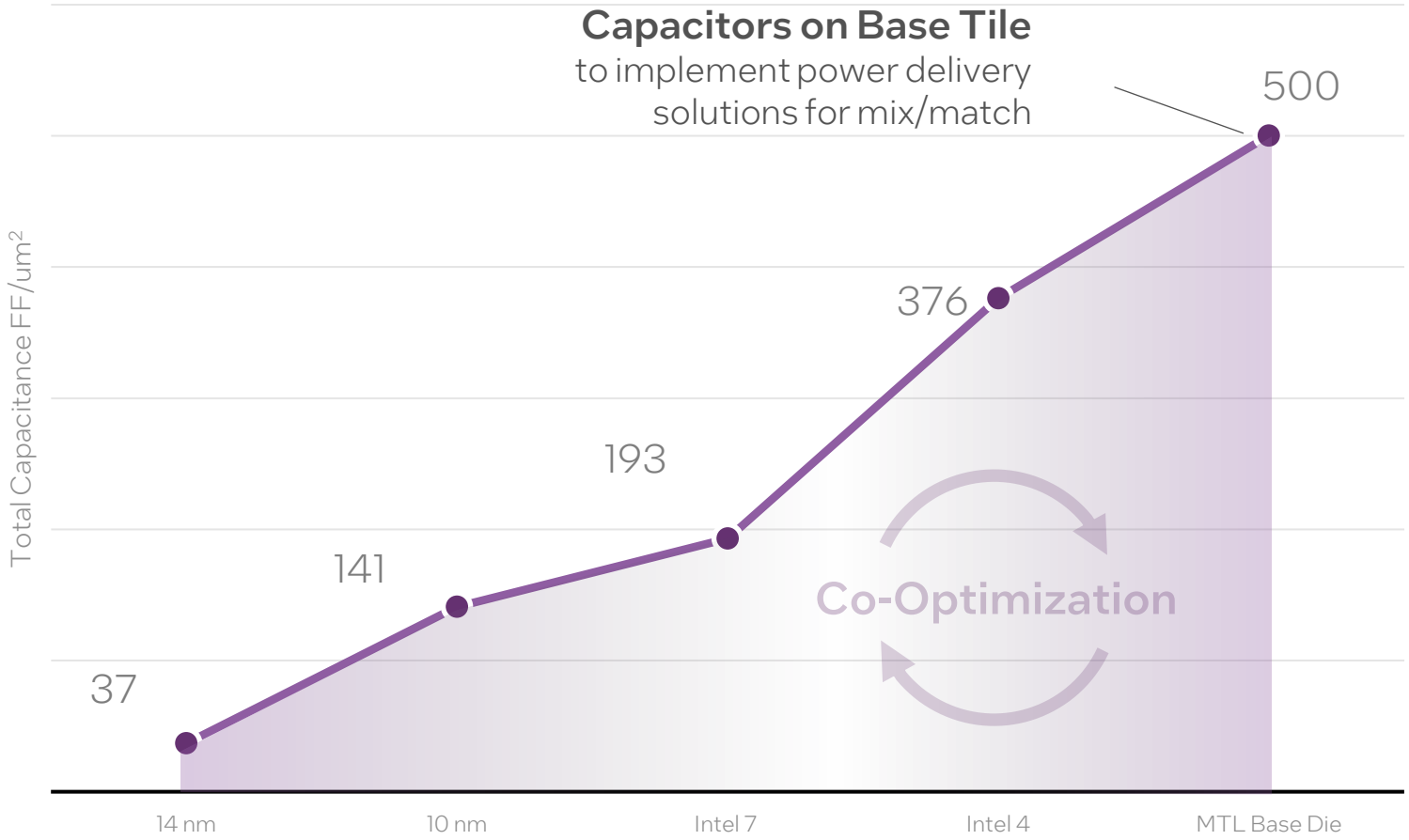
System, Software, Silicon

Co-optimization

- Floorplan
- Materials
- System
- Process
- Architecture

Meteor Lake

Power Delivery



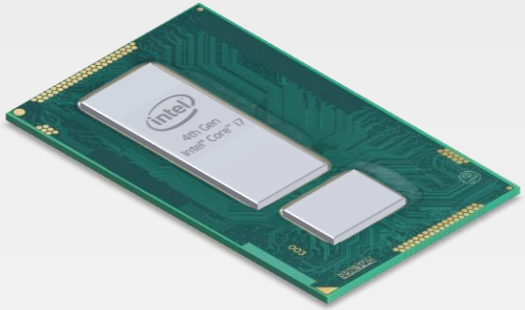
System, Software, Silicon

Co-optimization

- Decoupling/Noise
- Voltage Regulator
- IP Mix and Match
- Form Factor
- Package Optimization

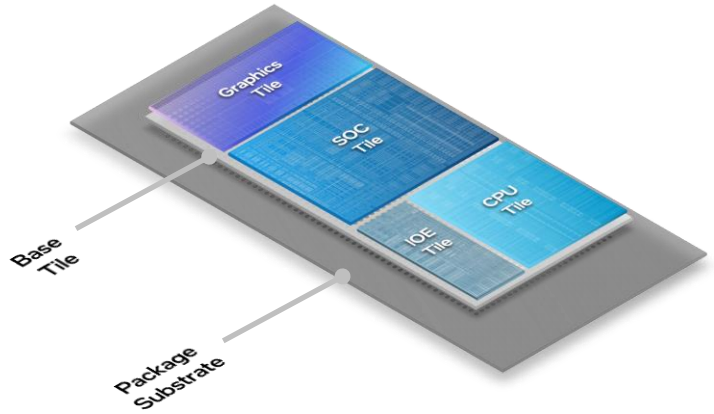
Client (Haswell)

2013

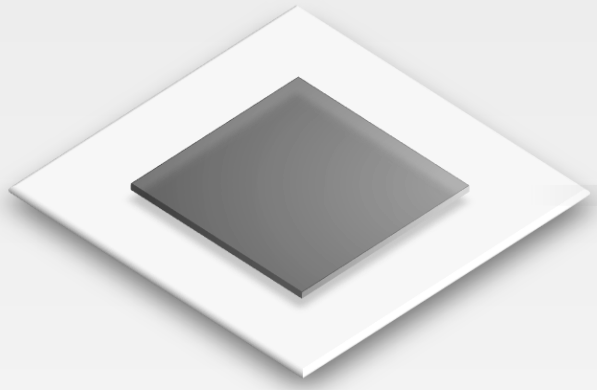
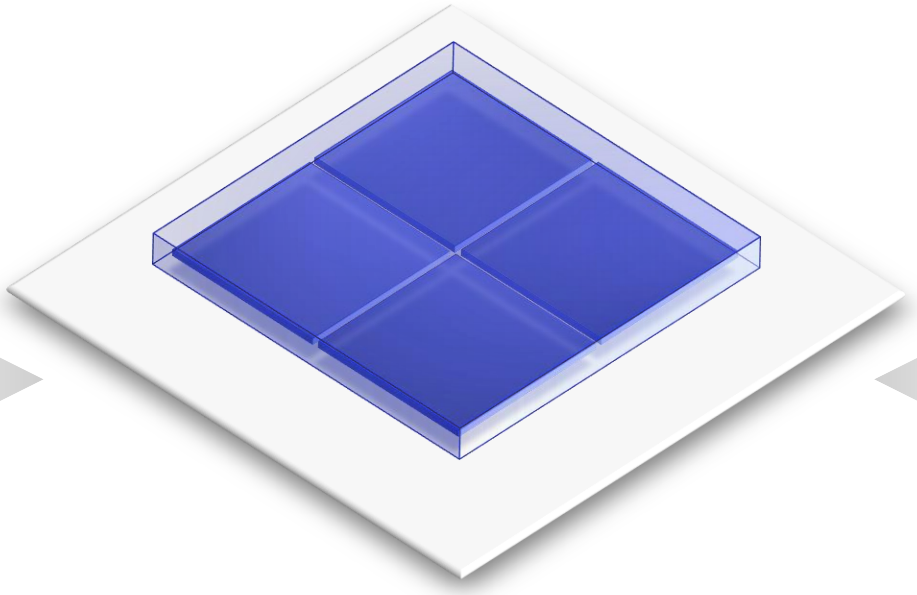
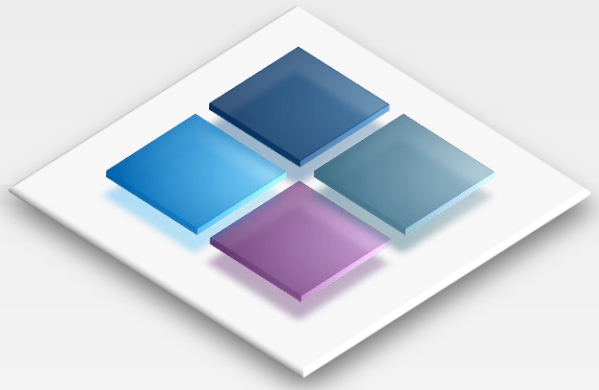


Meteor Lake

2023

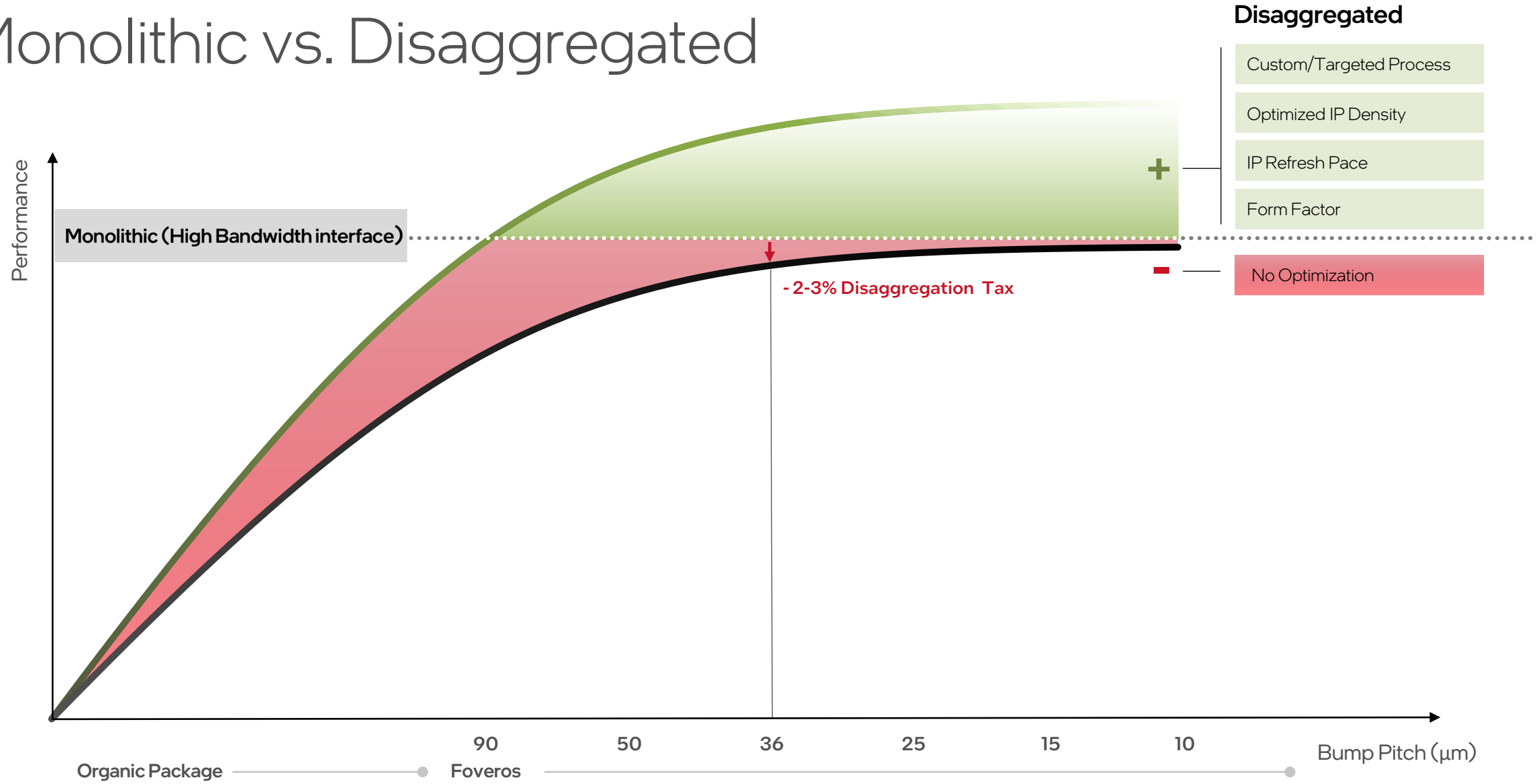


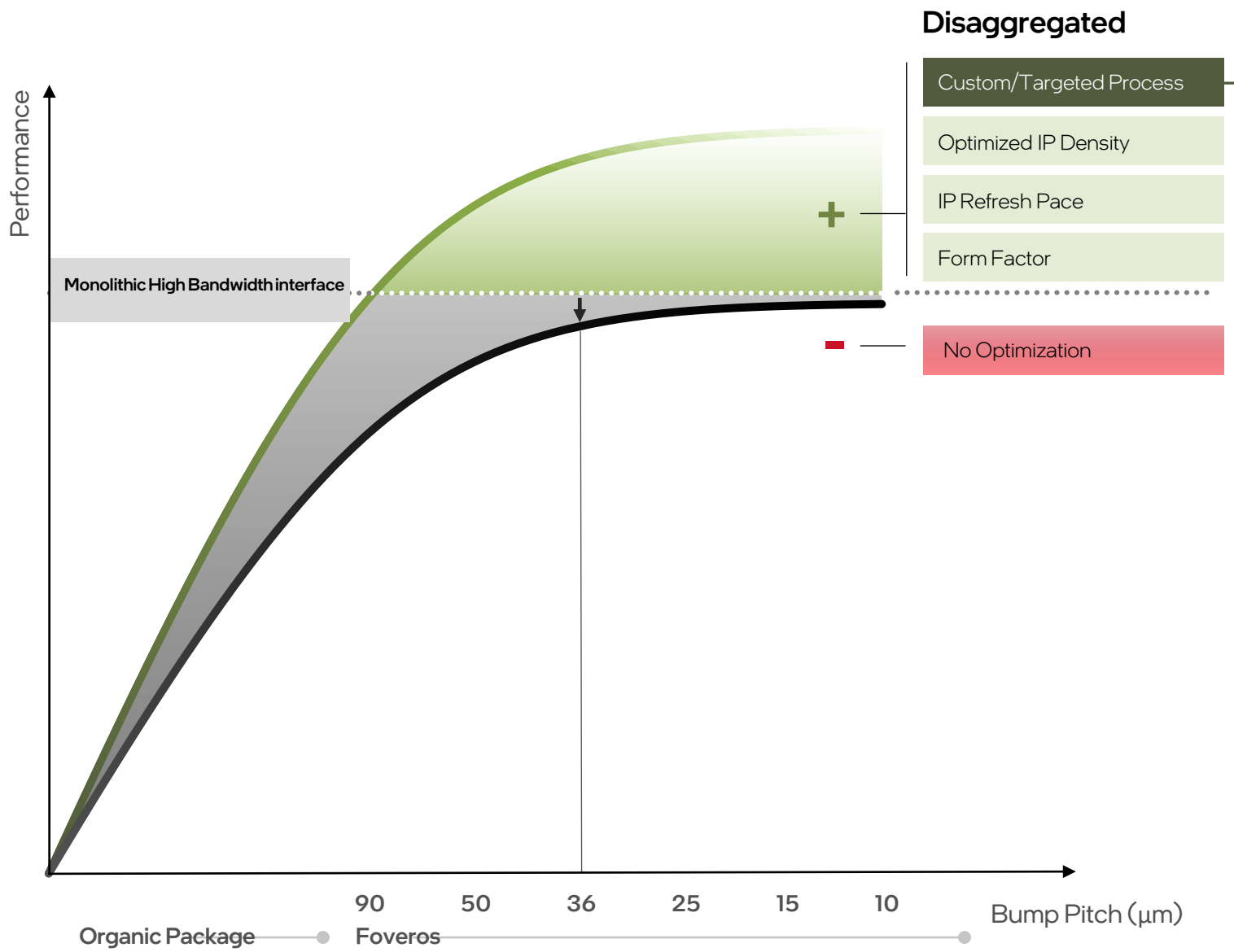
Interface	**OPIO (On Package IO)**		**FDI (Foveros Die Interconnect)**
Speed	2- 8GT/s	→	**2 GT/s**
IO/mm^2	1X (110 um)	→	**10X (36 µm)**
Latency	10- 20 ns	→	**< 10 ns**
Power	1 pJ/bit	→	**0.2 - 0.3 pJ/bit**
Number of Tiles	2	→	**5**

Monolithic		Disaggregated
		
Highest		Lower (tax on latency, power, B/W)
Very Limited		**Limited**
Slow (per SOC basis)		**Faster (release per new function)**
Low		**Higher**

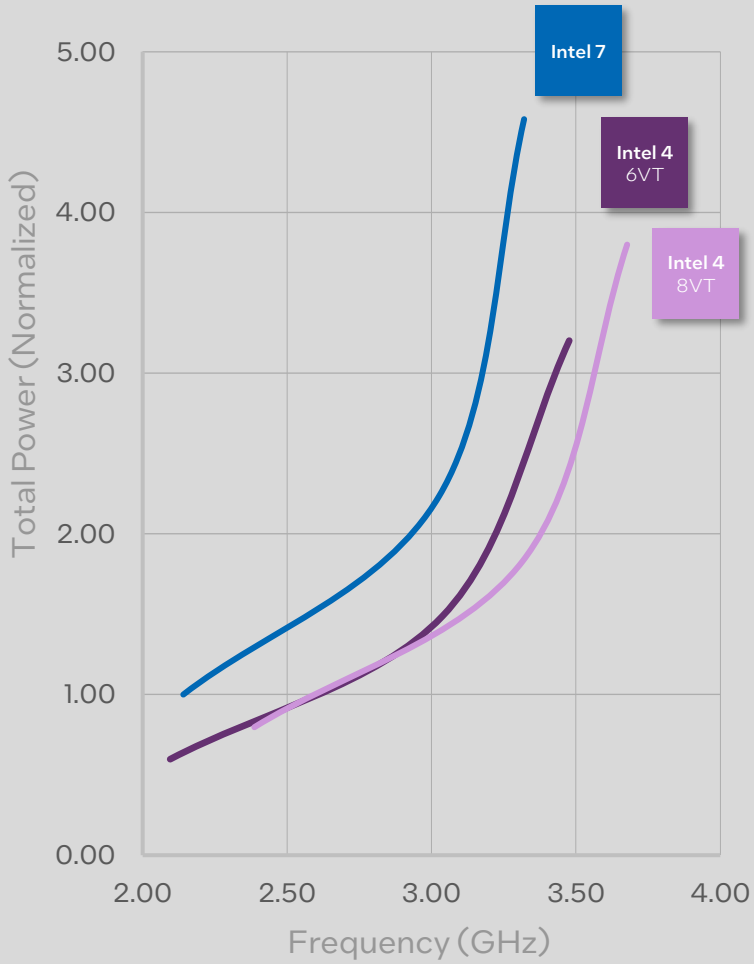
Can we get **monolithic** performance with **disaggregated** architecture benefits?

Monolithic vs. Disaggregated

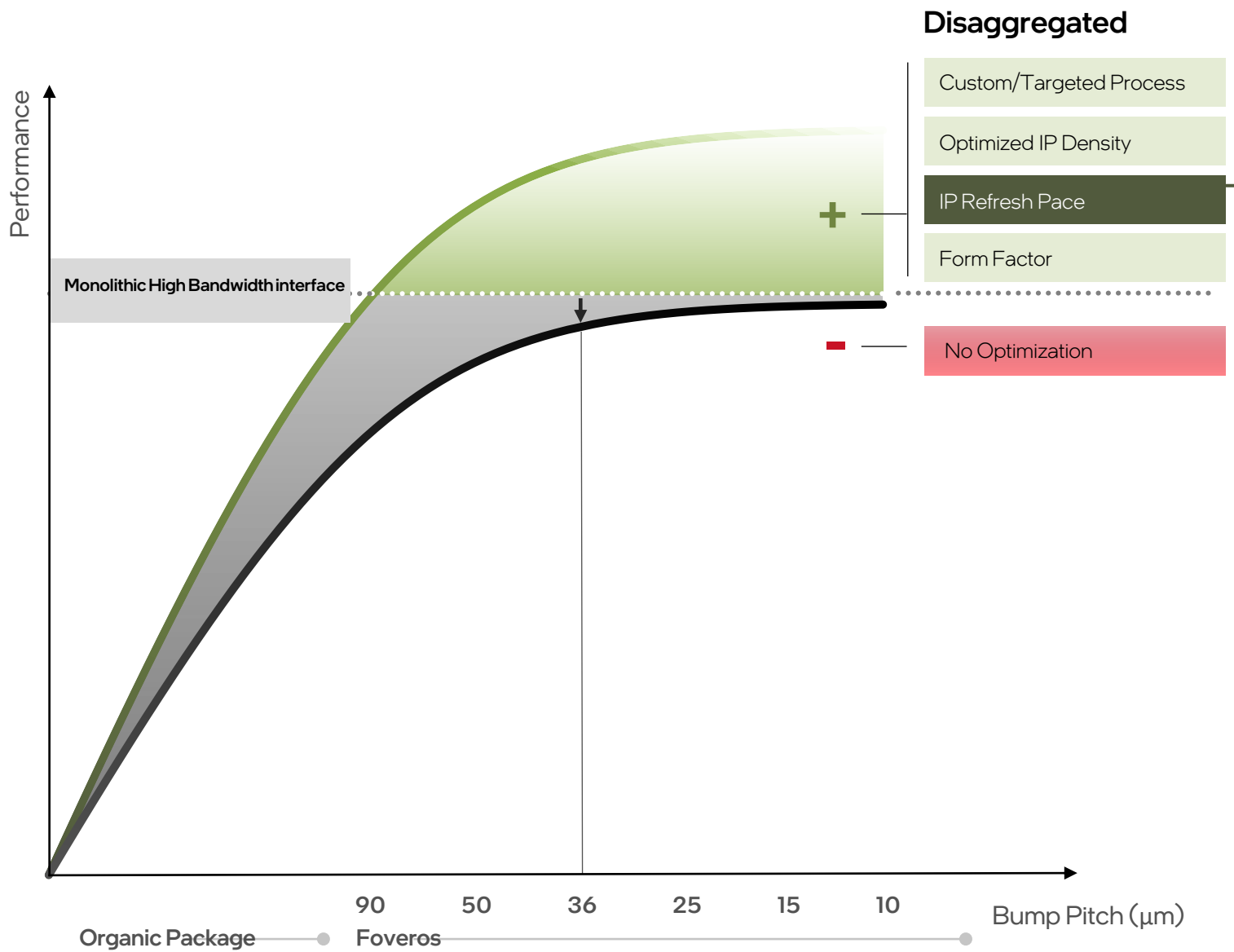




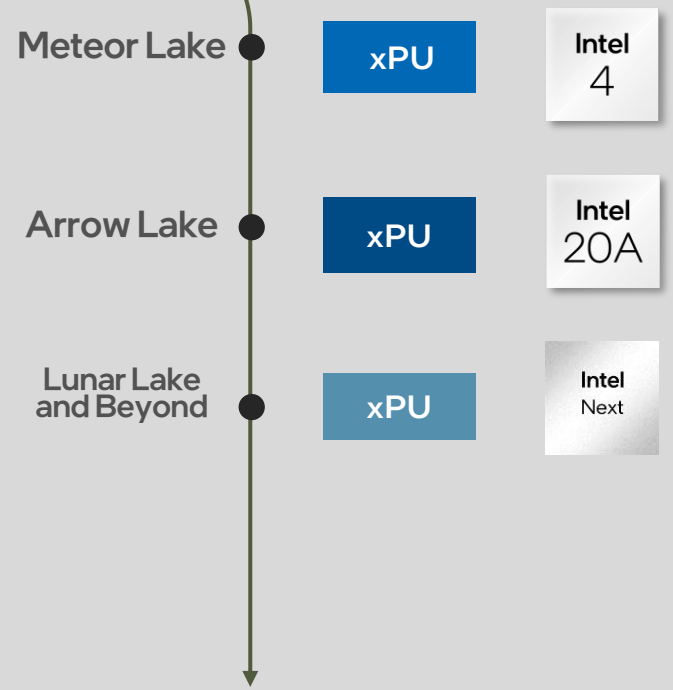
Transistor Performance Uplift[1]



1. B. Sell et al., "Intel 4 CMOS Technology Featuring Advanced FinFET Transistors optimized for High Density and High-Performance Computing," 2022 IEEE Symposium on VLSI Technology and Circuits (VLSI Technology and Circuits), 2022, pp. 282-283, doi: 10.1109/VLSITechnologyandCir46769.2022.9830194.

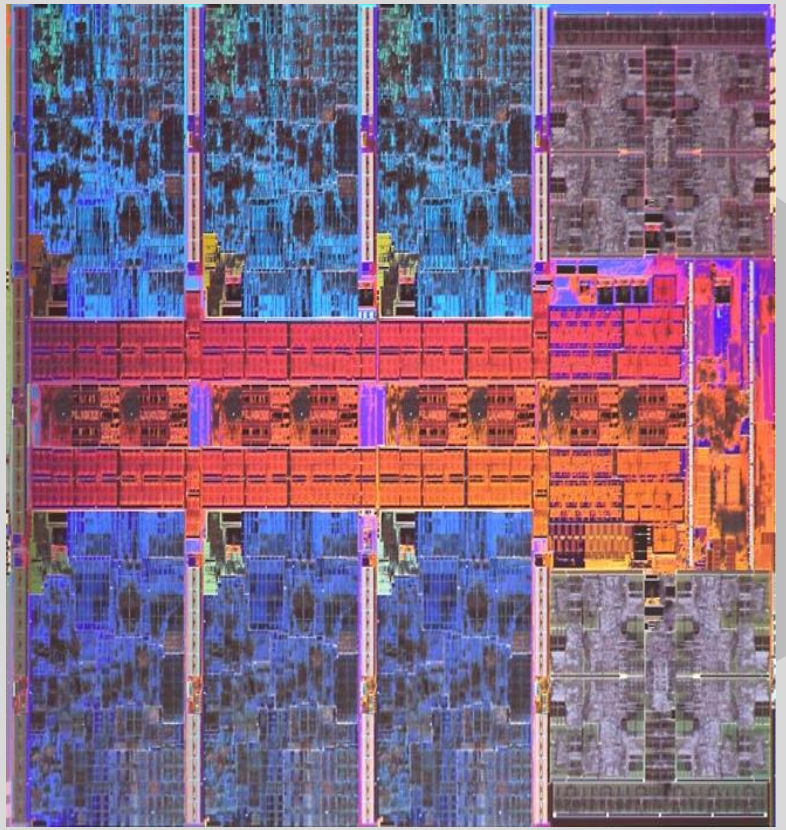


IP refresh, Process node



Meteor Lake Status

Meteor Lake booted and in the lab



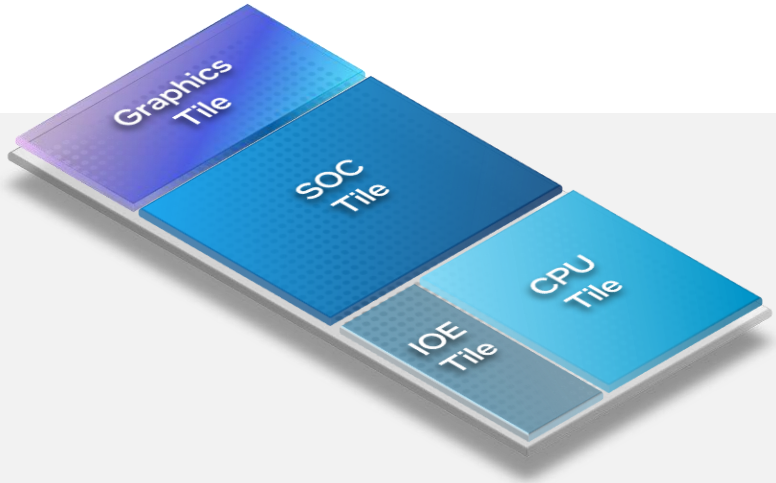
*Graphics for illustrative purposes only and not to scale.

One Architecture – Multiple Performance Points

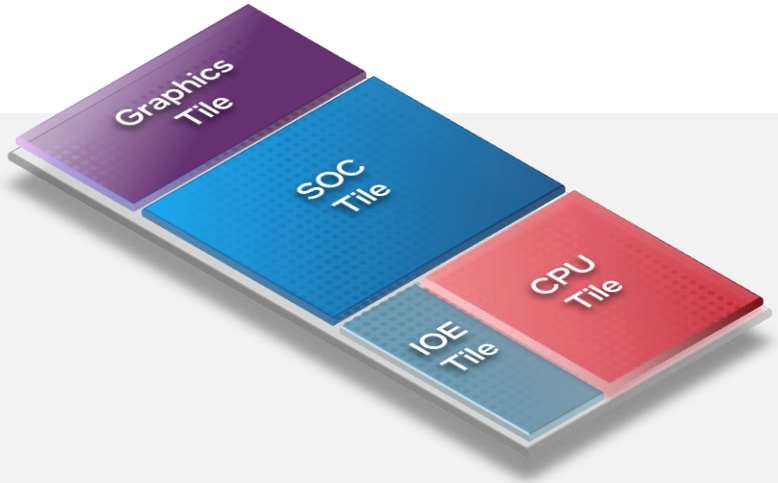
- Small Graphics Tile
- Efficient CPU Tile
- Reduced IO Tile

- Max Config GPU Tile
- Max Config CPU Tile
- Expanded IO Tile

Scalable Architecture across Multiple Generations



Meteor Lake



Arrow Lake

Lunar Lake & Beyond

	Meteor Lake	Arrow Lake	Lunar Lake & Beyond
Packaging	▪ Foveros ▪ 36 µm pitch	▪ Foveros ▪ 36 µm pitch	▪ Foveros ▪ 25 µm pitch
Process	Intel 4 · External	Intel 20A · External	Intel Next · External

Future of Client

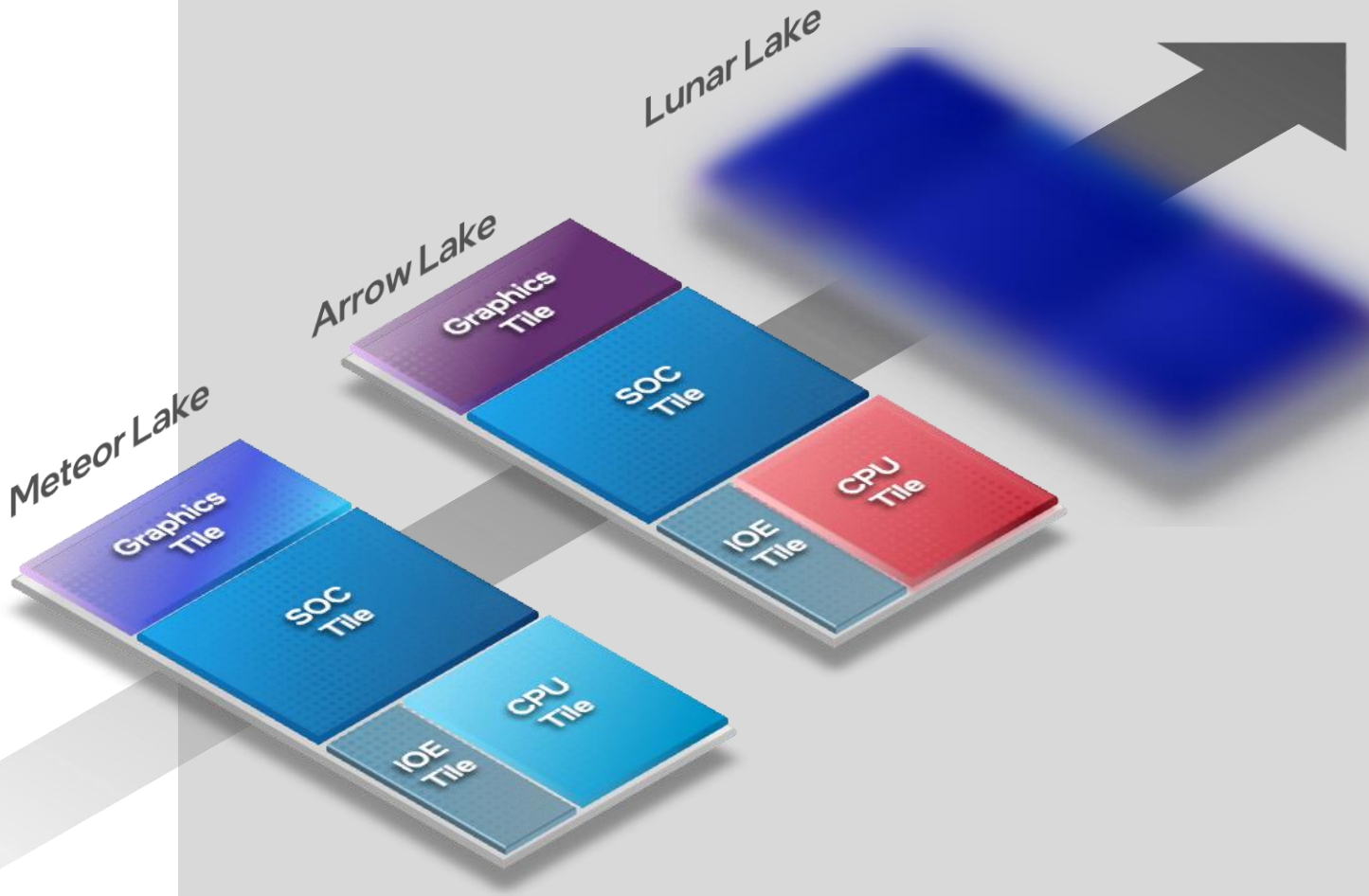
"Experience First" Client
drives New Era of **System level integration**

Monolithic performance with
disaggregated benefits

Process, packaging and architecture
together make this possible

Meteor Lake first tiled disaggregated architecture on Intel 4

Architecture is extremely flexible and
scales across design points and **into future**



Future of Compute

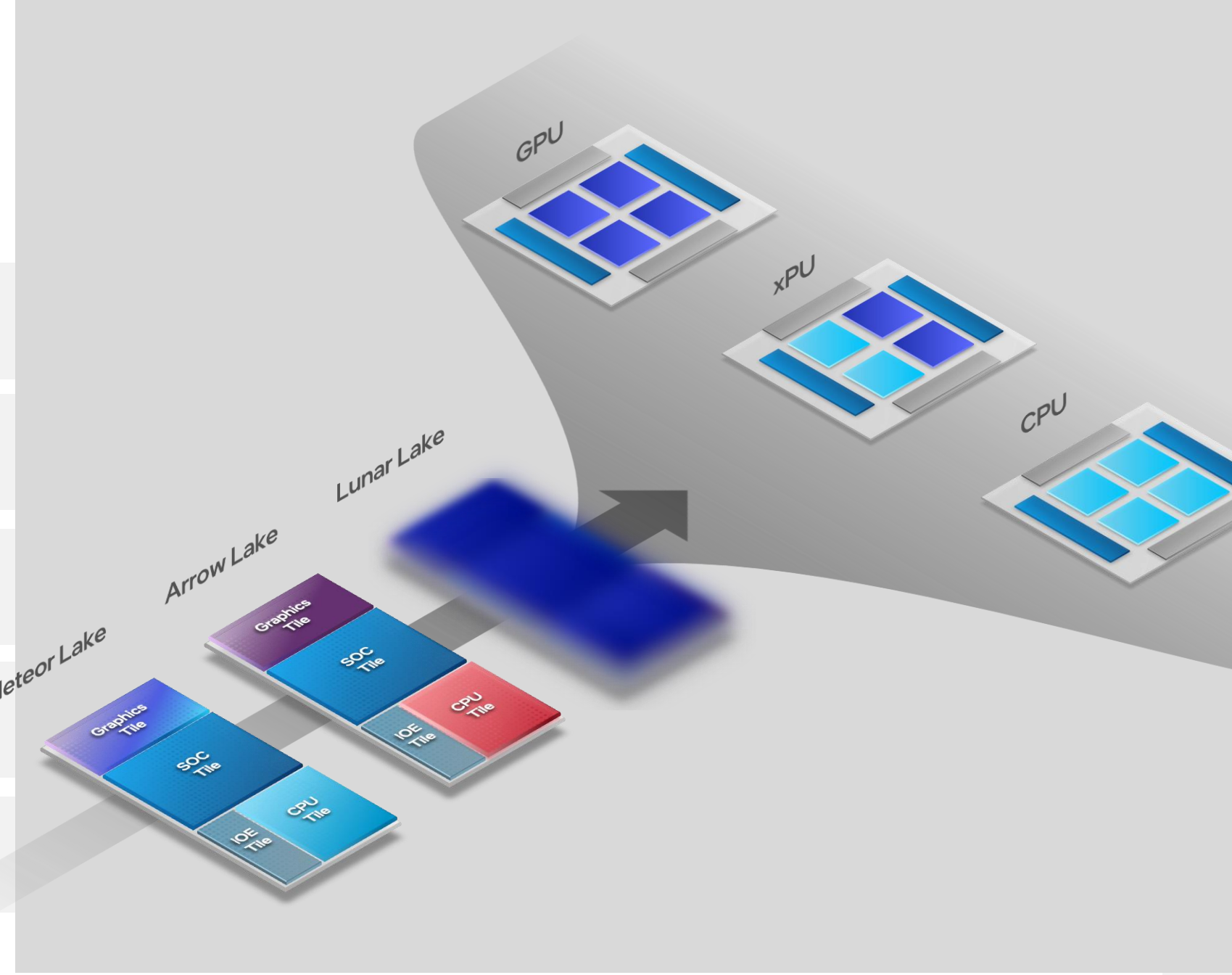
Scalable Architecture,
Construction and Packaging

Large range of Thermal and
Performance Envelope

Flexible across process nodes,
Monolithic benefits

3D Monolithic
Multi-layer, Logic on Logic

Manufacturing at Scale for
the next Billion Devices



Q&A