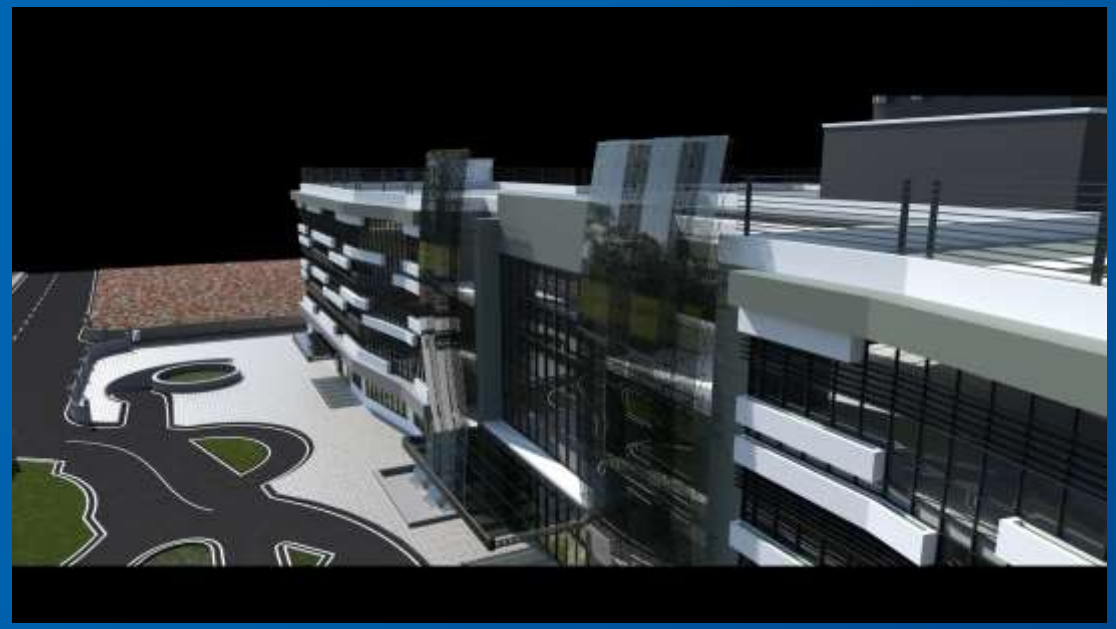
New Applications Enabled by Complex 3D Microsystems

Josh Fryman, PhD
Office of the CTO, Intel Fellow

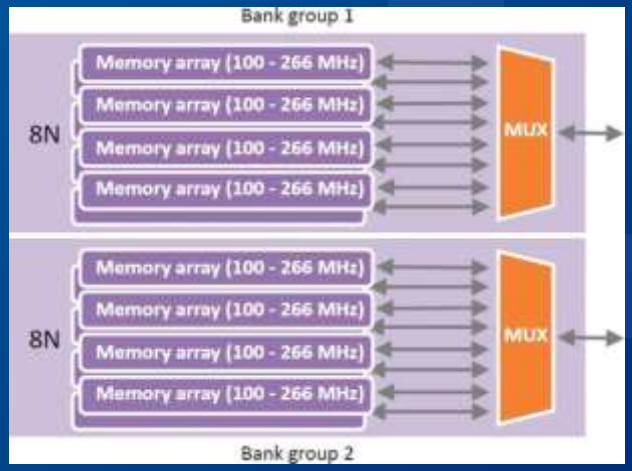
Aug 22, 2023

ELECTRONICS RESURGENCE INITIATIVE 2.0

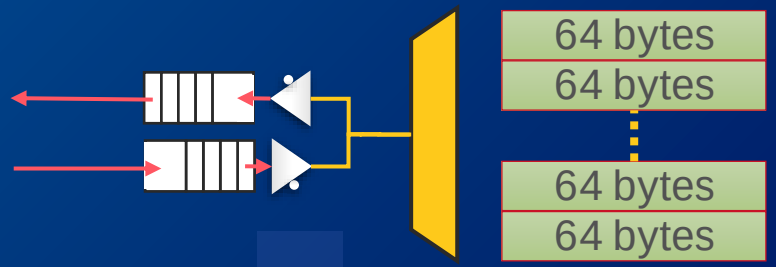
Let's talk about memory for a moment . . .



The real world is 3D (or higher)



Memory is 2D internally



Which we make look 1D

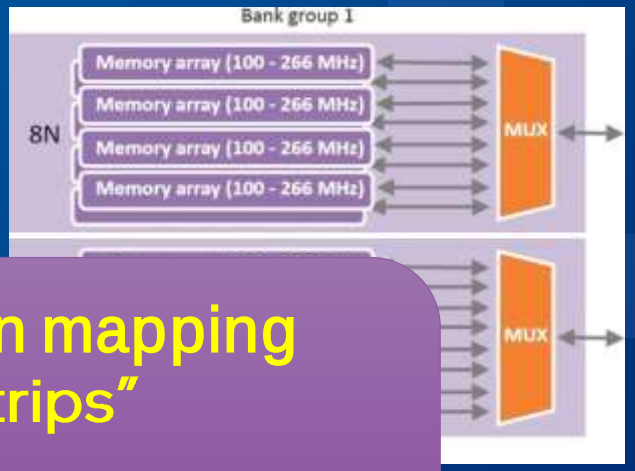
Image Credit:
https://www.cgtrader.com/3d-models/architectural/architectural-street/bilding-complex

Image Credit:
Intel Corporation

Image Credit:
https://www.semiconductor-digest.com/the-history-and-future-of-dram-architecture-in-different-application-domains-an-analysis/

Let's talk about memory for a moment . . .

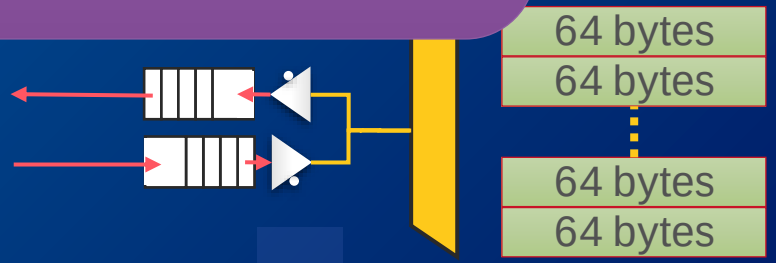
The real world is 3D (or higher)



ternally

We quickly bottleneck when mapping constructs onto 1D "strips"

Even though it has a multi-dimensional structure and potential to do more



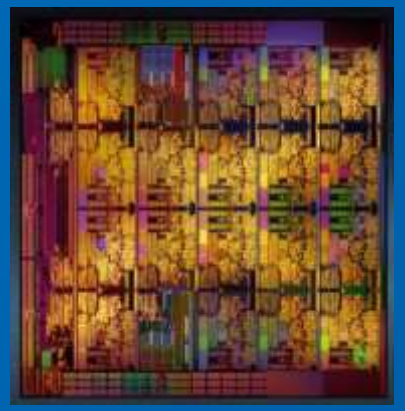
Which we make look 1D

Image Credit:
https://www.cgtrader.com/3d-models/architectural/architectural-street/bilding-complex

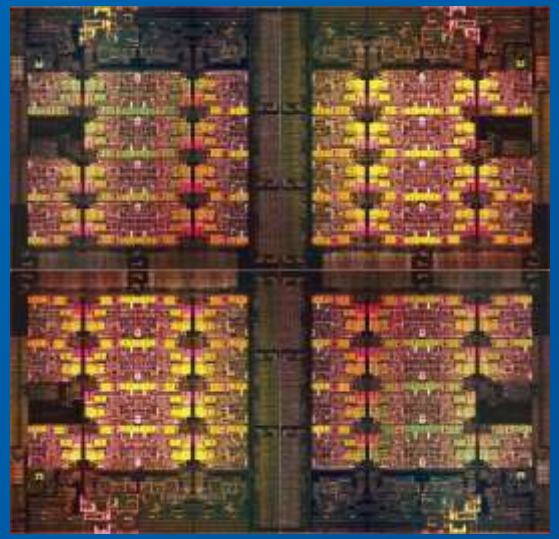
Image Credit:
Intel Corporation

Image Credit:
https://www.semiconductor-digest.com/the-history-and-future-of-dram-architecture-in-different-application-domains-an-analysis/

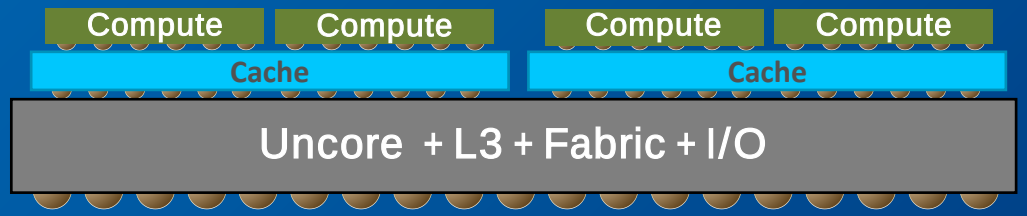
Chip design: are we 3D today?



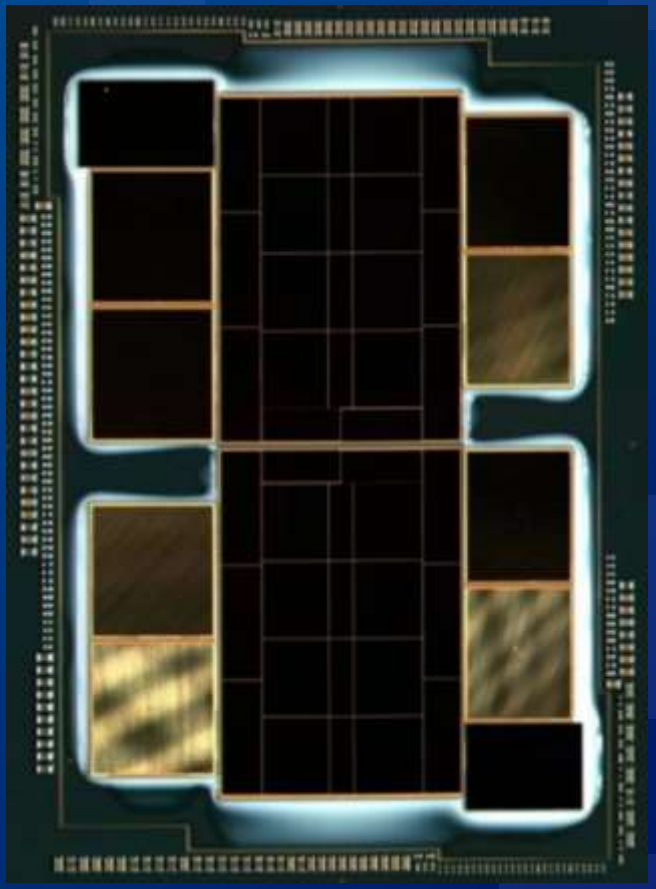
Monolithic



Polylithic
multi-chip



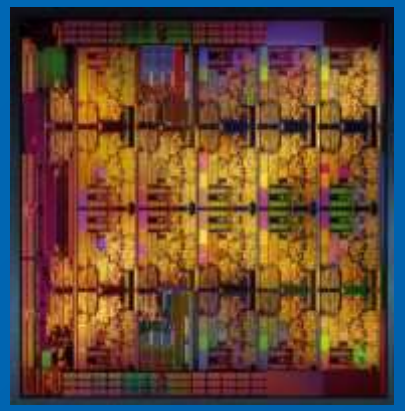
Polylithic tiled



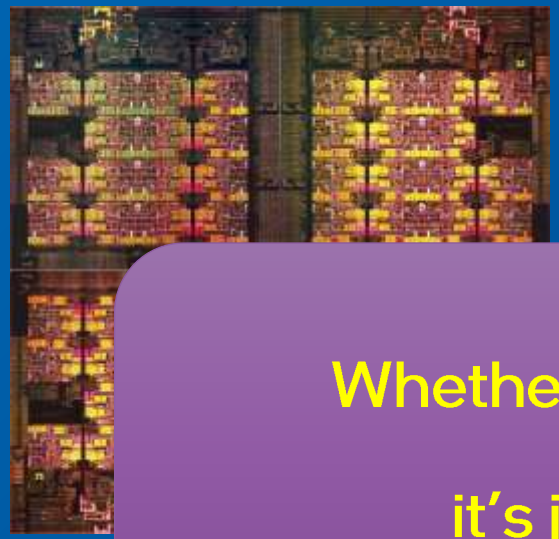
Heterogenous
tiled multi-chip

Image Credits: Intel Corp, TomsHardware

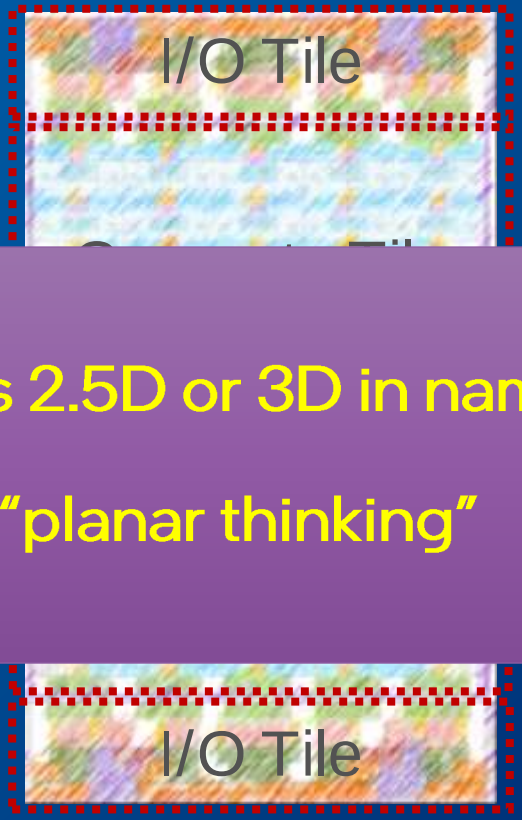
Chip design: are we 3D today?



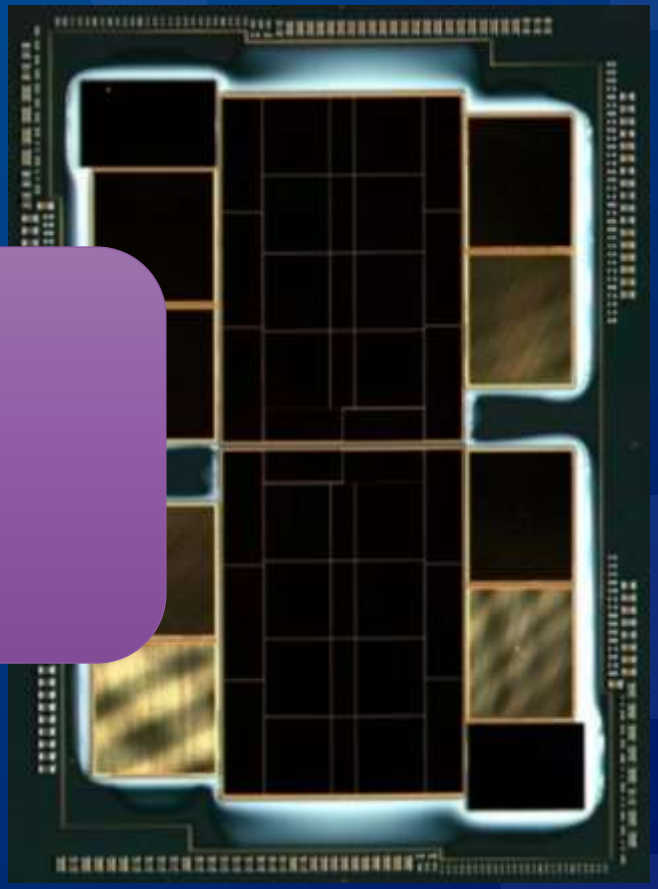
Monolithic



Multi-chip



Polylithic tiled



Heterogenous tiled multi-chip

Whether it's 2.5D or 3D in name,

it's just "planar thinking"

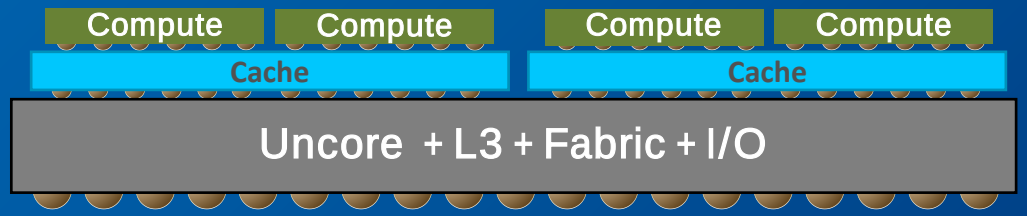
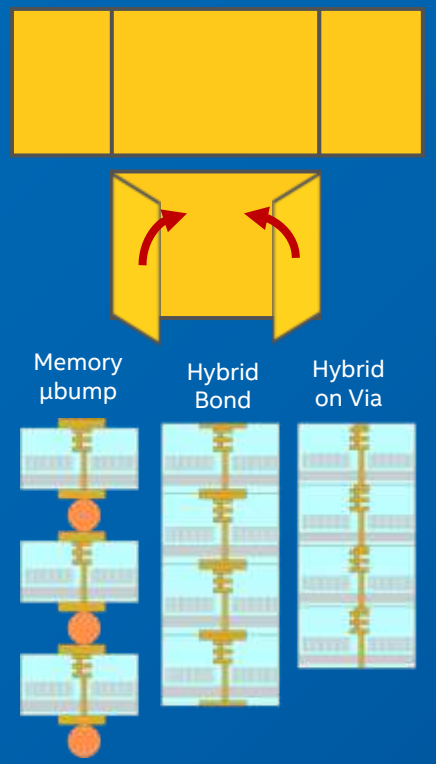
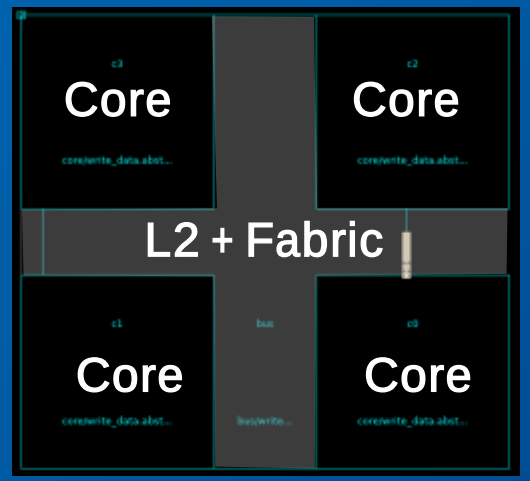


Image Credits: Intel Corp, TomsHardware

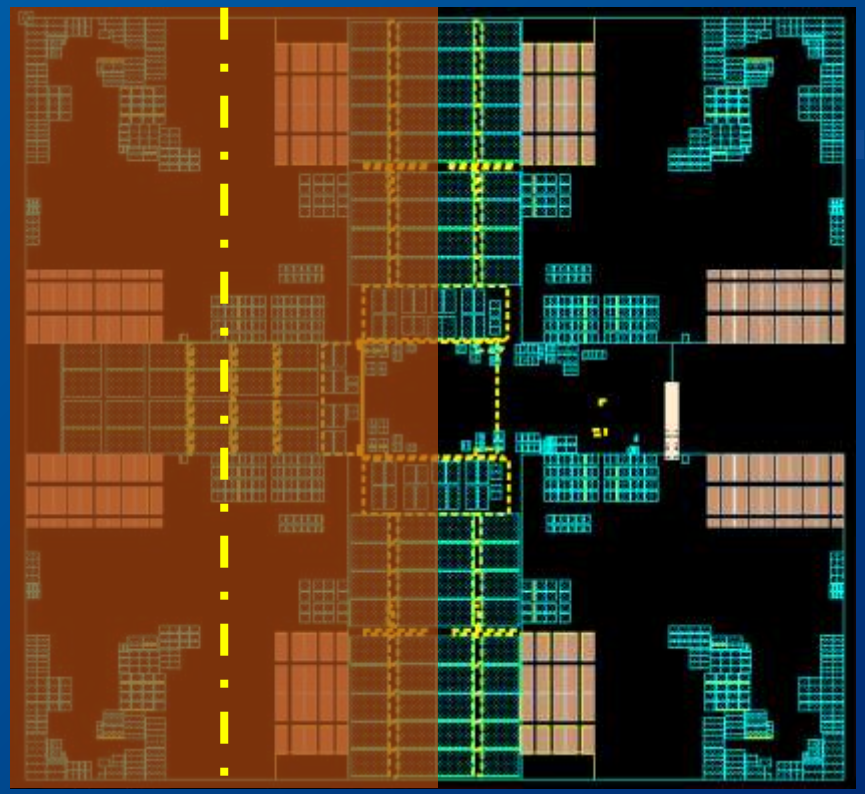
What about cutting-edge 3D?



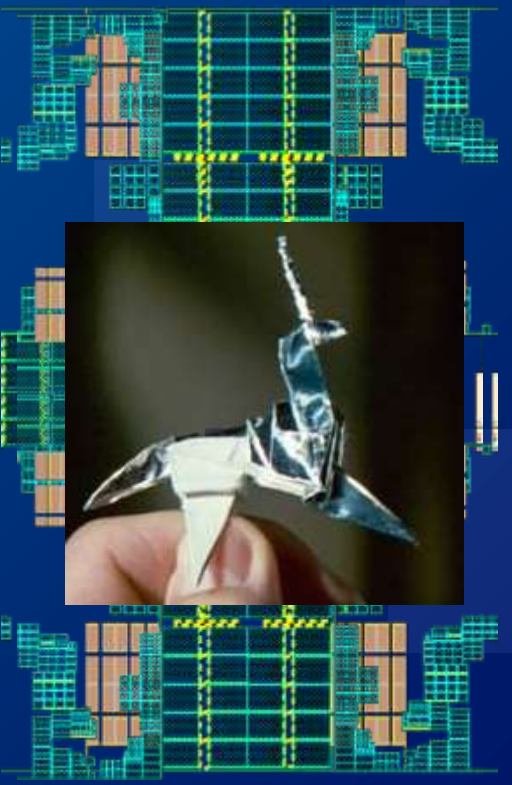
Core Folding



4-Core Tile Floorplan



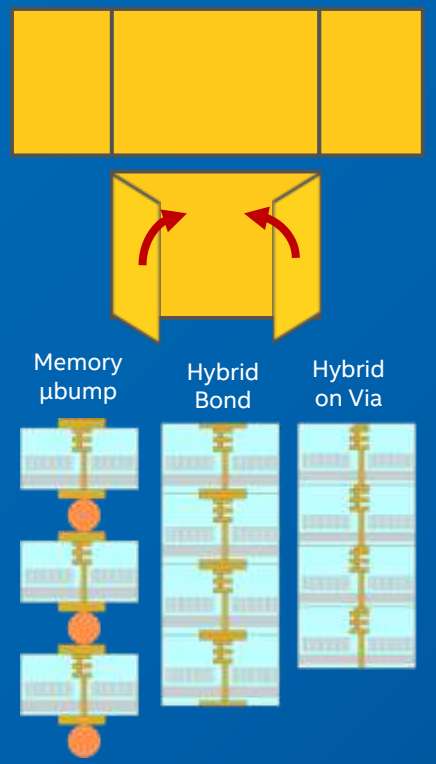
Fold Line Concept with Hybrid Bonding



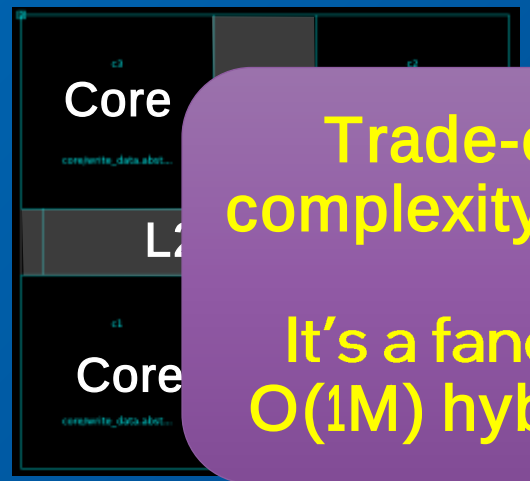
"Shrink"

Image Credits: Intel Corp, Warner Bros (Blade Runner)

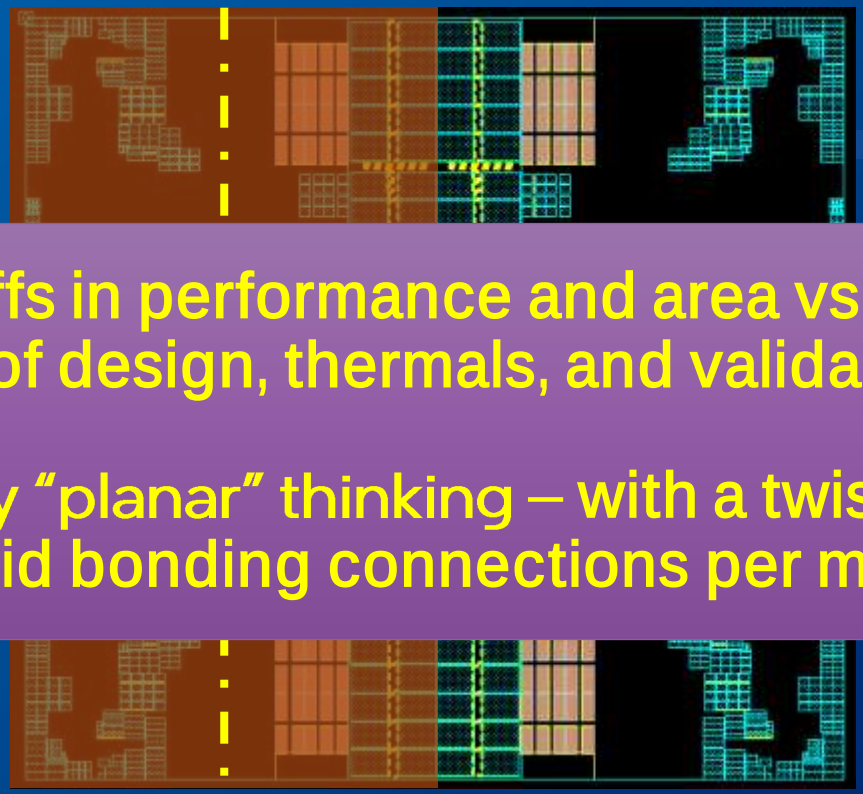
What about cutting-edge 3D?



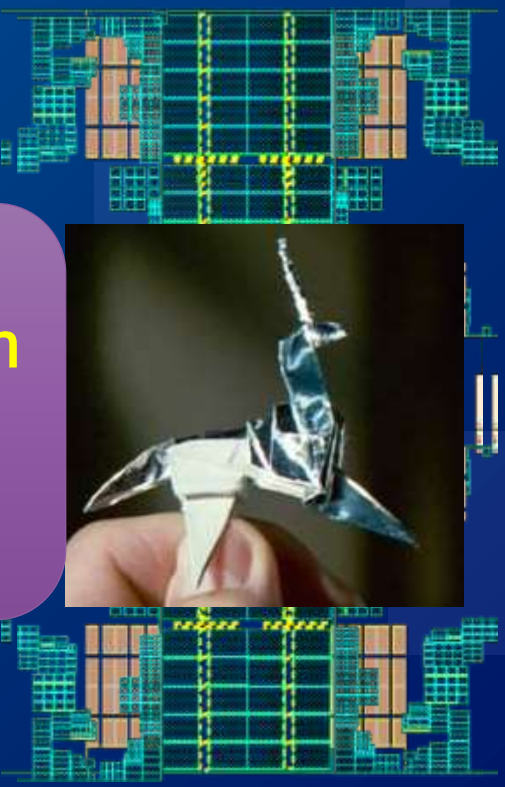
Core Folding



4-Core Tile Floorplan



Fold Line Concept with Hybrid Bonding



"Shrink"

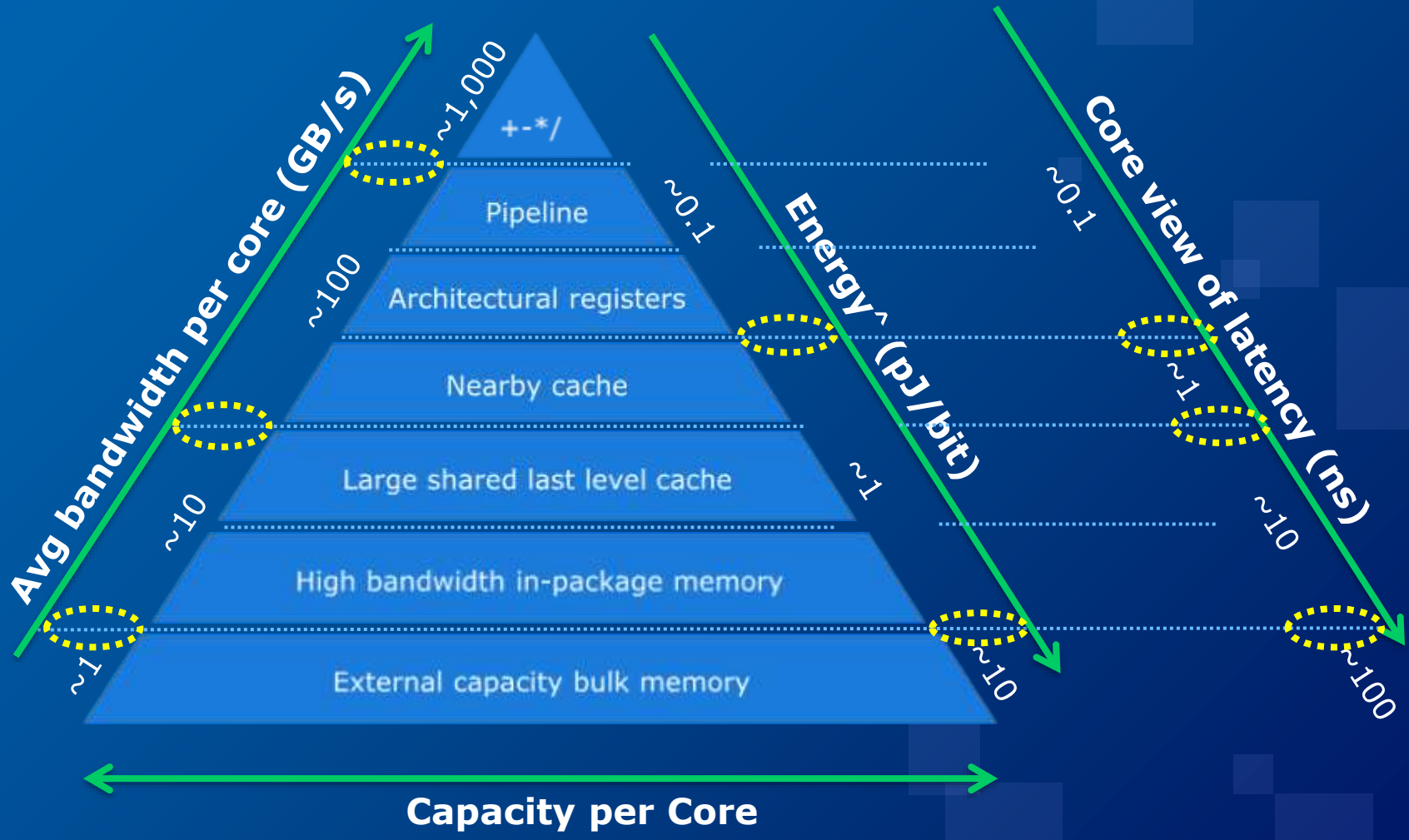
Trade-offs in performance and area vs. complexity of design, thermals, and validation

It's a fancy "planar" thinking – with a twist: O(1M) hybrid bonding connections per mm²

Image Credits: Intel Corp, Warner Bros (Blade Runner)

But are we exploiting the right things?

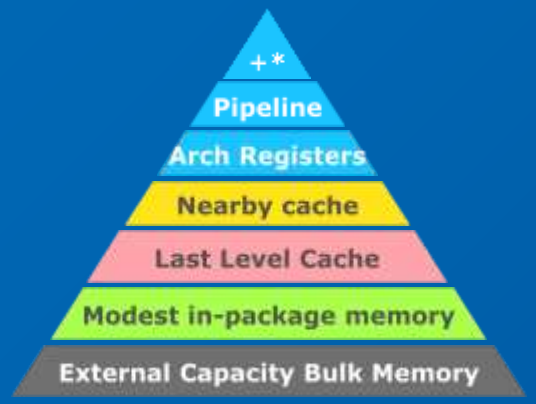
Tapering divisions in bandwidth, energy, and latency are opportunities to optimize

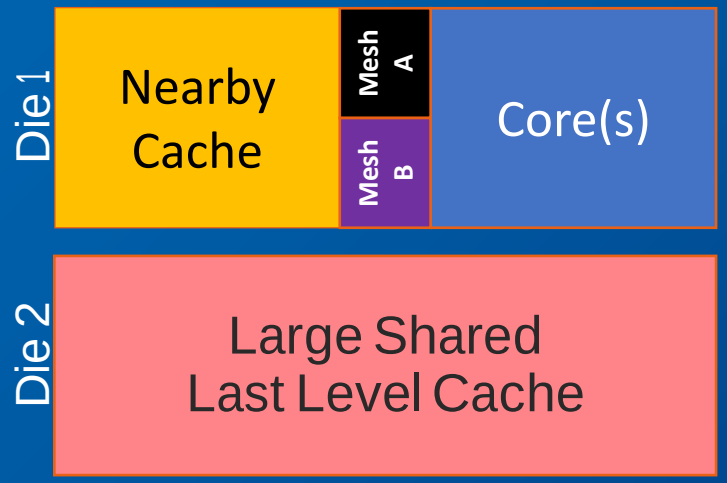


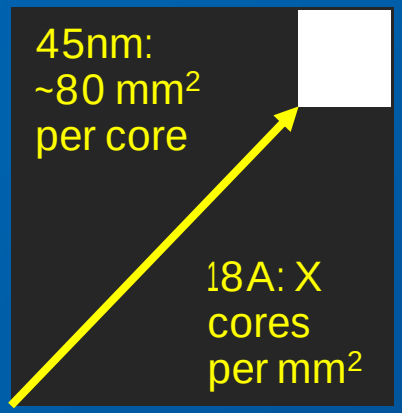
^Not counting overheads for access

Image Credits: Intel Corp

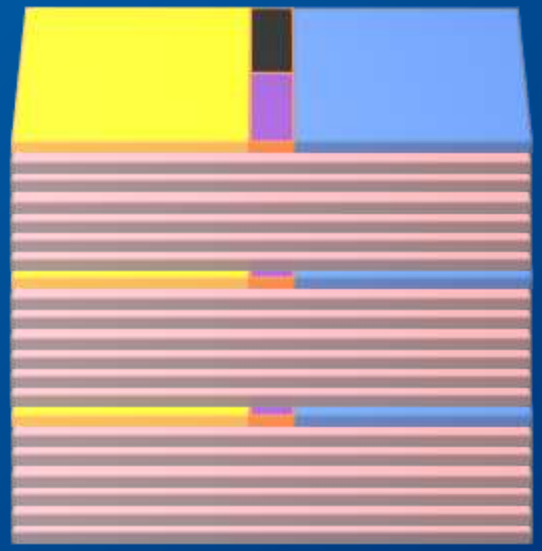
3D as a spatial concept, not planar







Fab Shrink



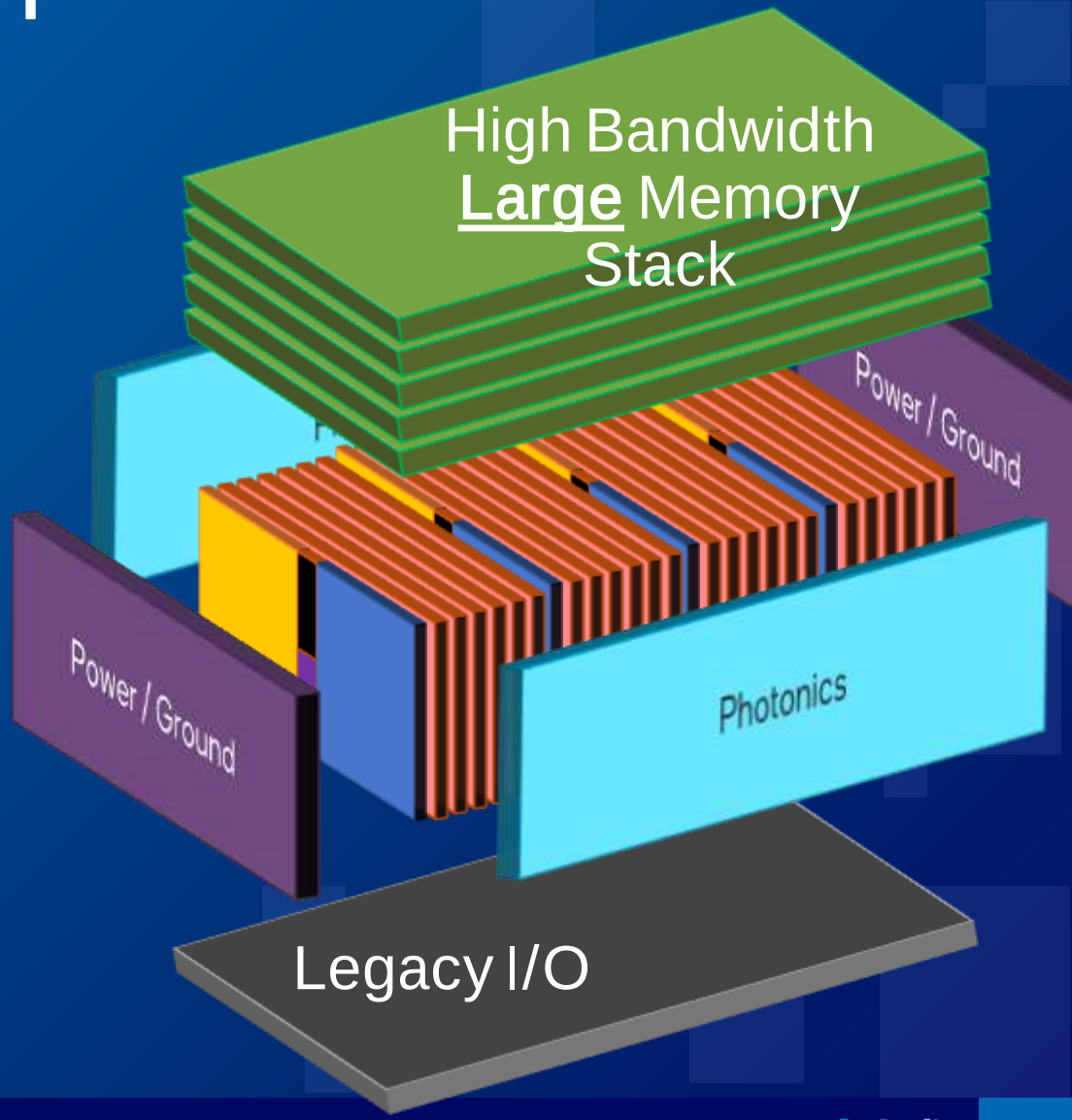
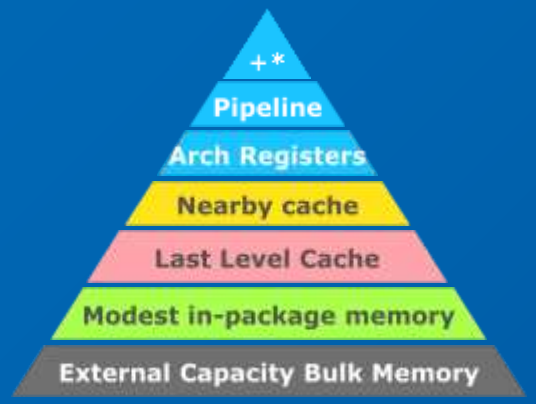
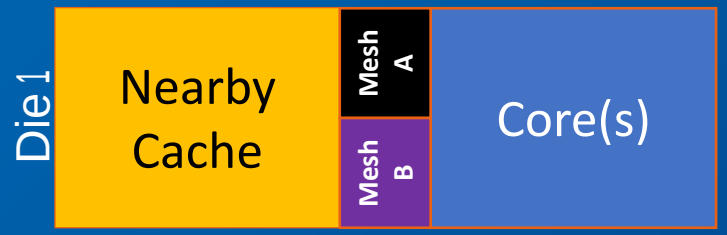
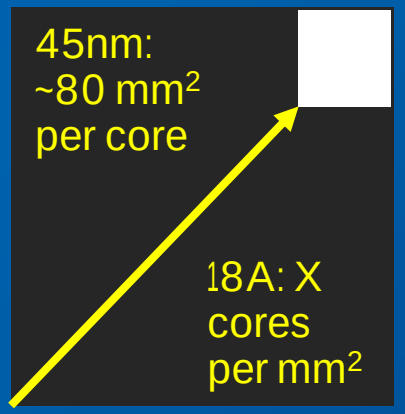


Image Credits: Intel Corp

3D as a spatial concept, not planar



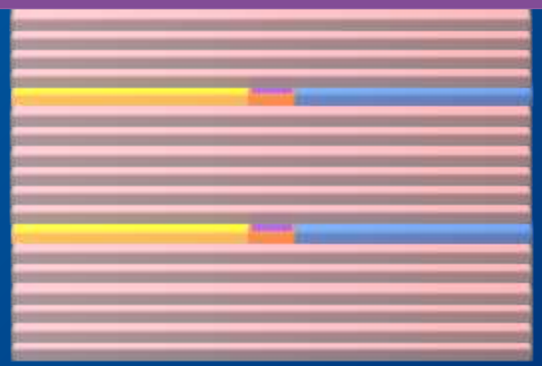




Fab Shrink

Match trade-offs to natural architecture inflection points – speeds and feeds.

Use all 6 faces of "cubic 3D" and simplify: O(1k) hybrid bonding connections per mm²



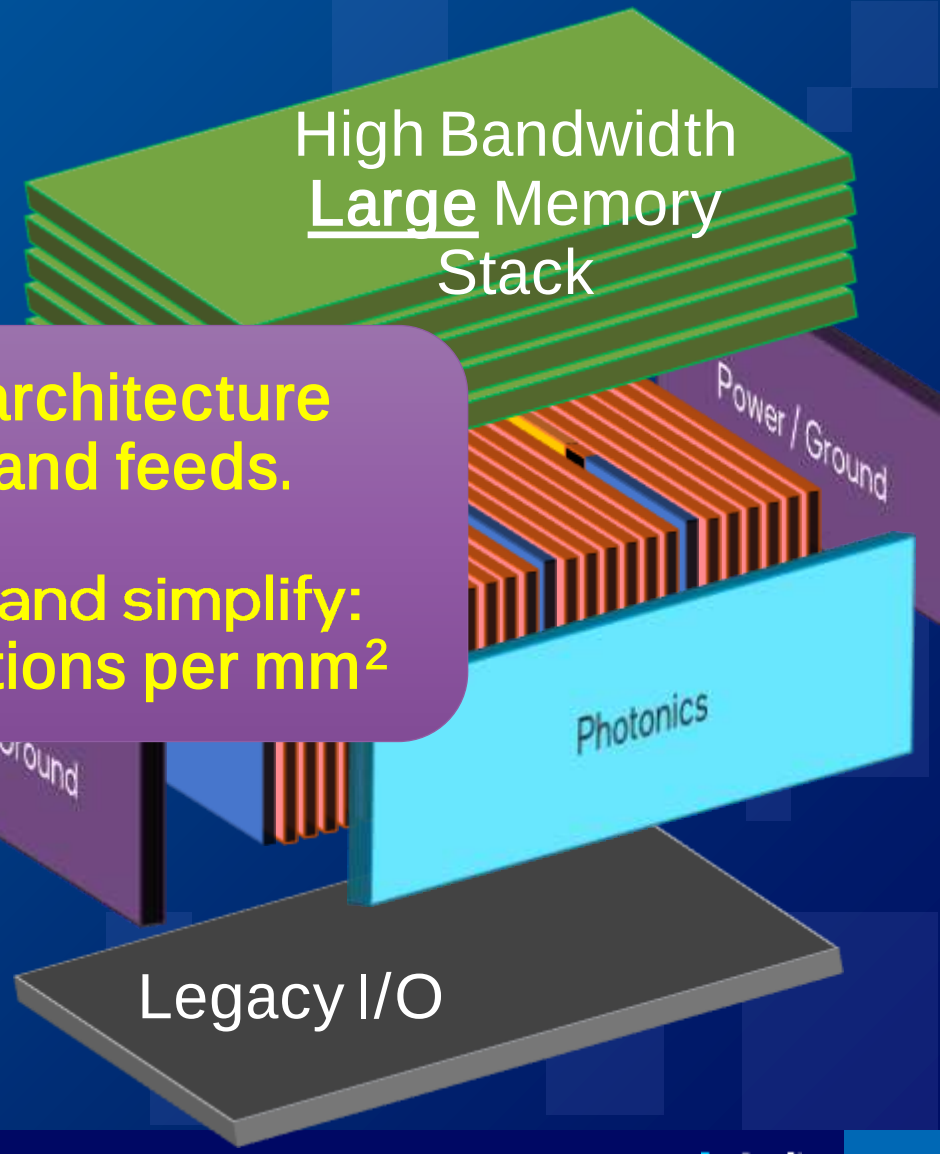


Image Credits: Intel Corp

Application breakthroughs of dense 3DHI

Aurora 2.0 ExaFLOP System

- 1 Rack ≈ 9.8 m^3
  - 128 Xeons + 384 X^e + Slingshot fabric
- Fits in 3D cube at 40 mm per side
  - >100,000x smaller volume

Apple's VisionPro AR Headset

- Multiple types of silicon, packages
  - Size, weight, and power are limiters
- Rice ≈ 2 mm x 2.75 mm x 6 mm per grain [1]
  - Fits Apple silicon and ~0.5 TB memory

[1] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6145214

Image Credits: DOE ANL, CNET

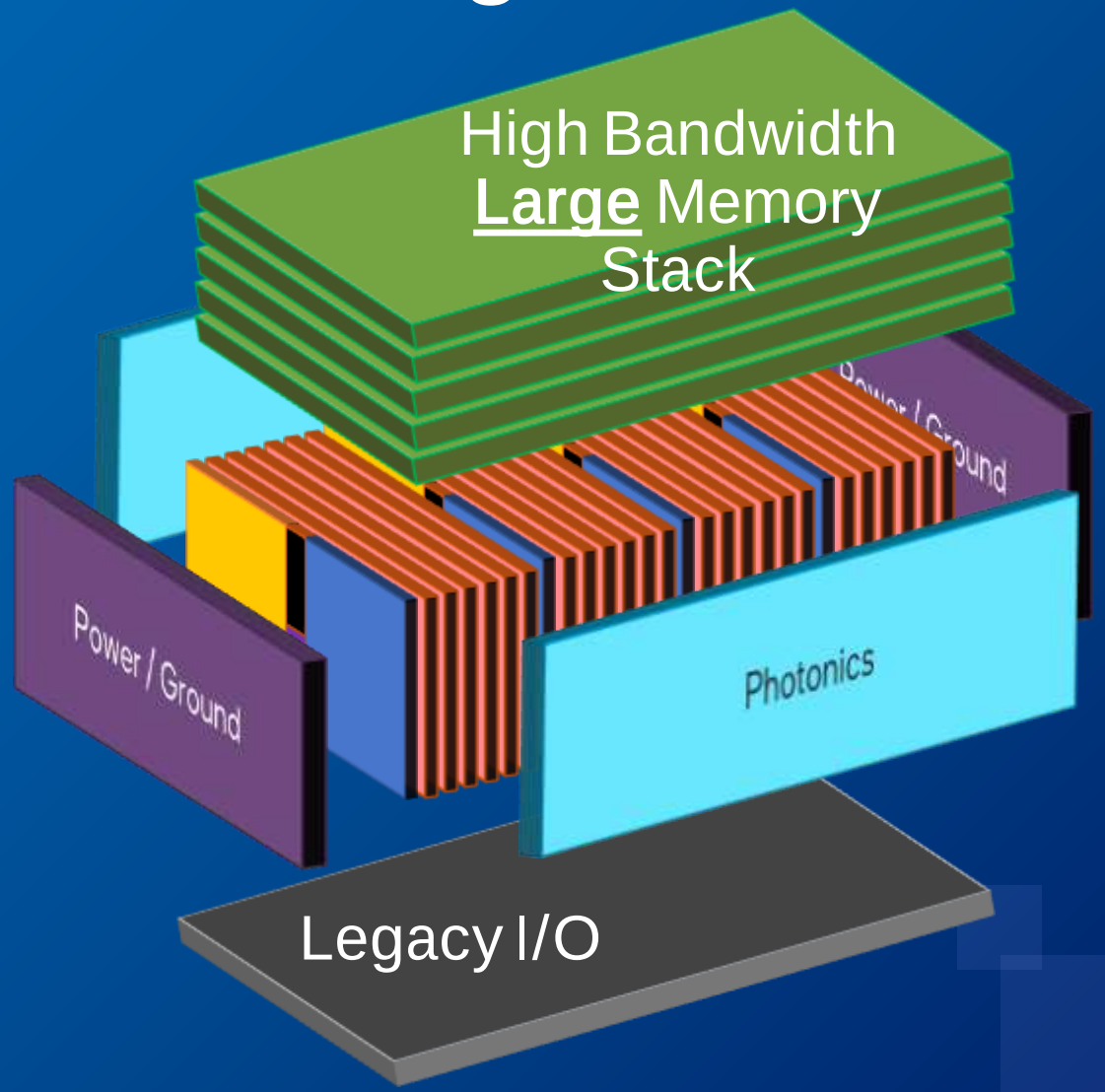
Challenges to achieving dense 3D

Manufacturing:

- edge polish
- right-angle attach
- via density
- bonding speed
- rework support
- redundancy
- tooling Z-height
- wafer thinning

Power and Thermals:

- thermal density
- power delivery
- cooling layers
- heterogeneous material



EDA:

- rotated die
- taper point isolation
- non-planar libraries
- formal verification
- DF<all> 1,000 layers
- scan chain time
- tests and coverage

Design:

- graceful degradation
- extreme interop
- built-in redundancy
- pluggable modules
- abstractions for all

Image Credits: Intel Corp

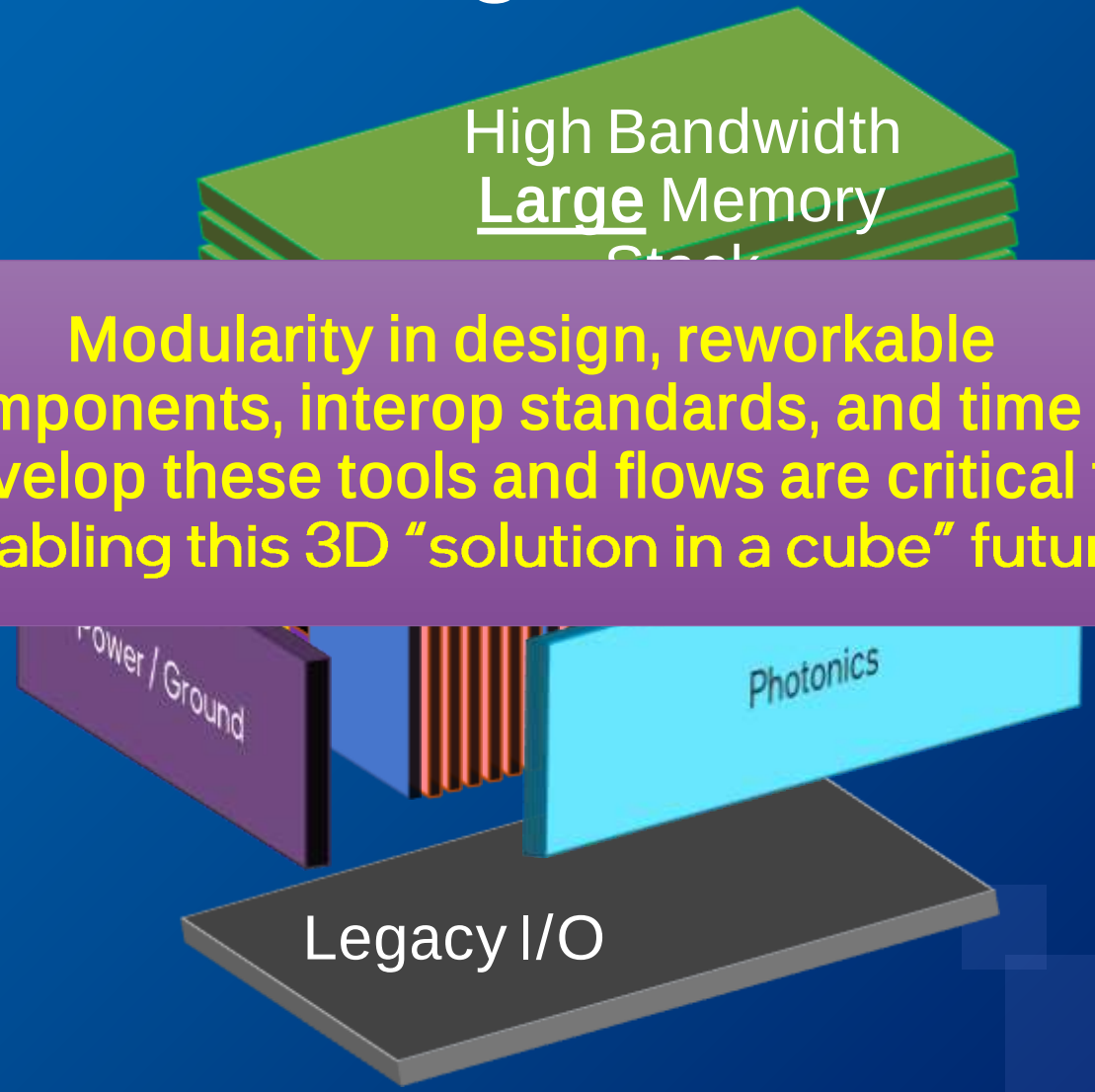
Challenges to achieving dense 3D

Manufacturing:

- edge polish
- right-angle attach
- via density
- bonding speed
- rework support
- redundancy
- tooling Z-height
- wafer thinning

Power and Thermals:

- thermal density
- power delivery
- cooling layers
- heterogeneous material



Modularity in design, reworkable components, interop standards, and time to develop these tools and flows are critical to enabling this 3D "solution in a cube" future

EDA:

- rotated die
- taper point isolation
- non-planar libraries
- formal verification
- DF<all> 1,000 layers
- scan chain time
- tests and coverage

Design:

- graceful degradation
- extreme interop
- built-in redundancy
- pluggable modules
- abstractions for all

Image Credits: Intel Corp


intel