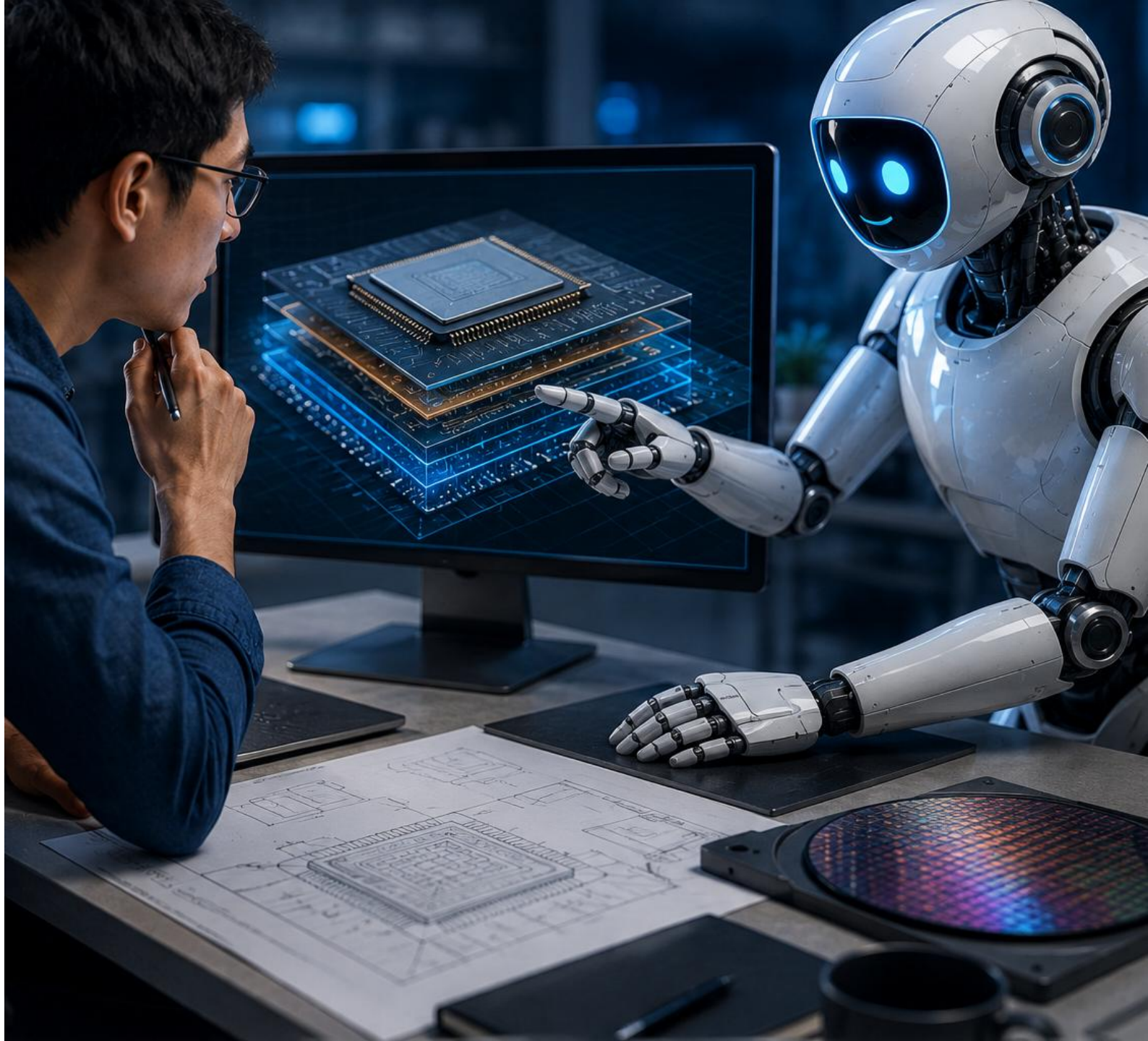
LOGIC FOLDING AND FINE-GRAINED LOGIC-ON-LOGIC 3D ICS: A KEY ENABLER OF TAU LAW

Sung Kyu Lim
University of Southern California
@ US-Korea Forum on Nanotechnology
September 14, 2026



TWO PHONES, TWO BRAINS

BOTH RELEASED IN SEPTEMBER 2026

DID YOU KNOW?

Apple used 7nm in A12-13 (2018-20).

iPhone Duo

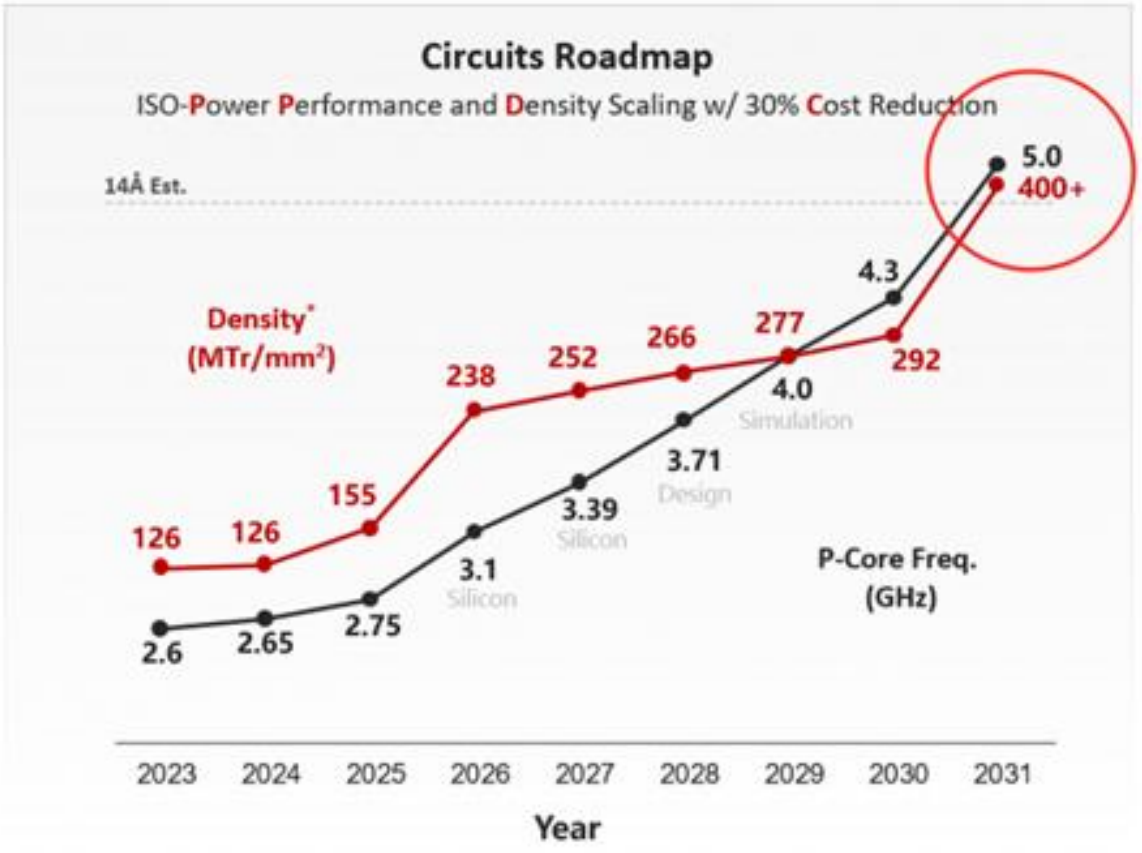
A20 Pro (TSMC **2nm**, 2D IC)

Huawei Mate XT2

Kirin 9050 Pro (SMIC **7nm**, 3D IC)

HUAWEI'S TAU SCALING LAW

MAY 2026



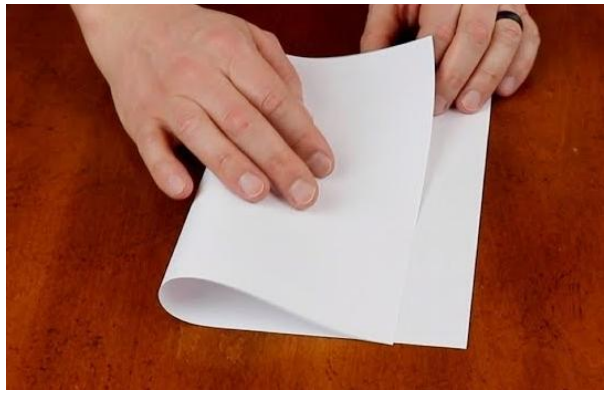
All of these

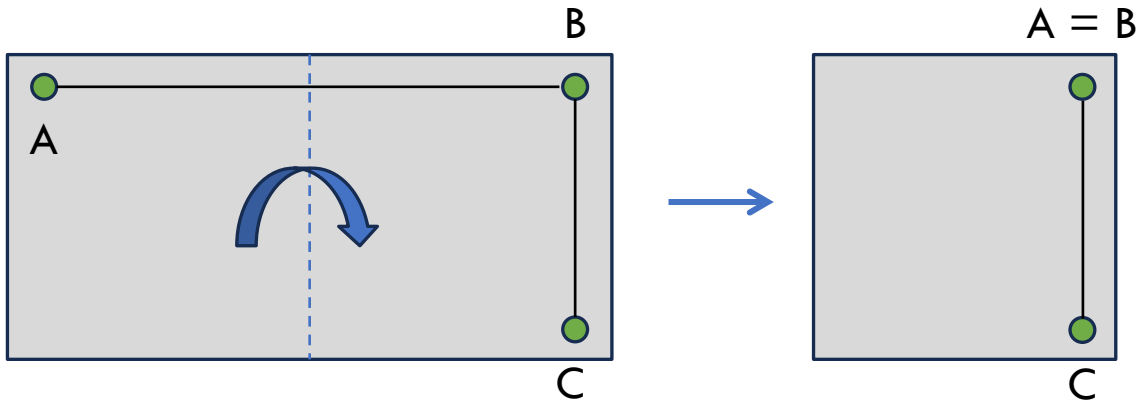
with 7nm transistors (no EUV)

&

LogicFolding (3D IC)

WHAT IS LOGIC FOLDING?





If I **fold** a 2x1 rectangle to 1x1

Wirelength savings:

- Best : A-B (100%)
- Corner-to-corner: A-C (67%)
- Ave saving: 35%

Wait, there is more:

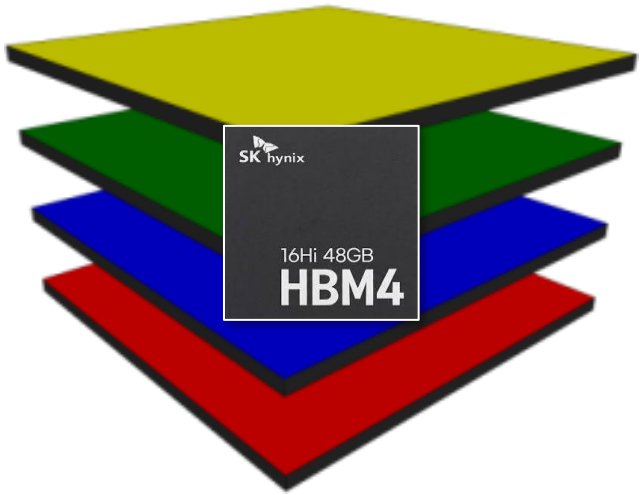
- We can **rearrange** points to maximize the saving even further

DESIGN PARADIGM SHIFT

NOT A MERE DESIGN TRICK

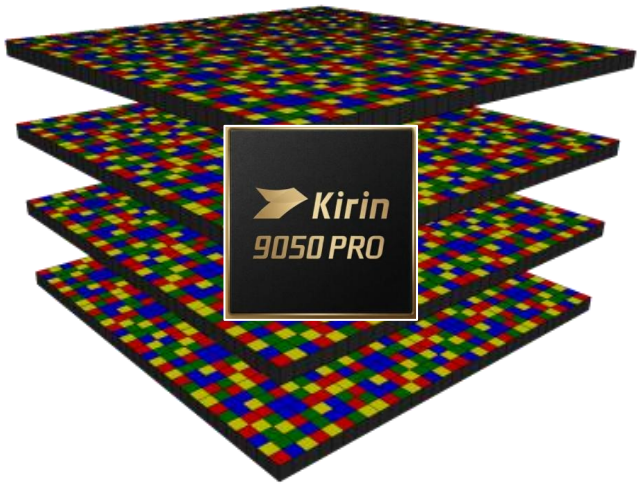
Today

stack 2D designs (coarse-grained)



Tomorrow

redesign directly in 3D (fine-grained)



WE DID A SIMILAR STUDY IN 2014

BUT DESIGN AND SIMULATION, NO CHIP DEMO

On Enhancing Power Benefits in 3D ICs: Block Folding and Bonding Styles Perspective

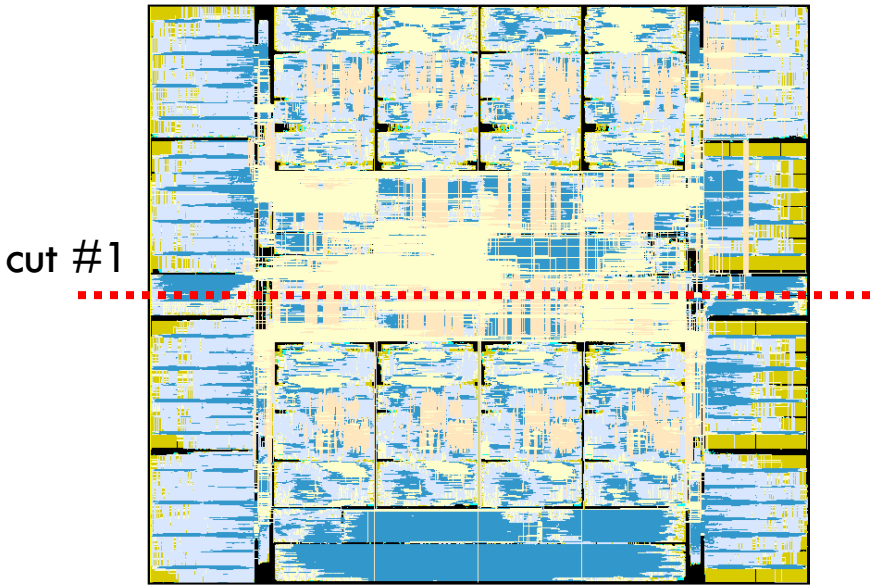
Moongon Jung, Taigon Song, Yang Wan, Yarui Peng, and Sung Kyu Lim
School of ECE, Georgia Institute of Technology, Atlanta, GA, USA
moongon@gatech.edu, limsk@ece.gatech.edu

ACM Design Automation Conference, 2014
Funded by Intel

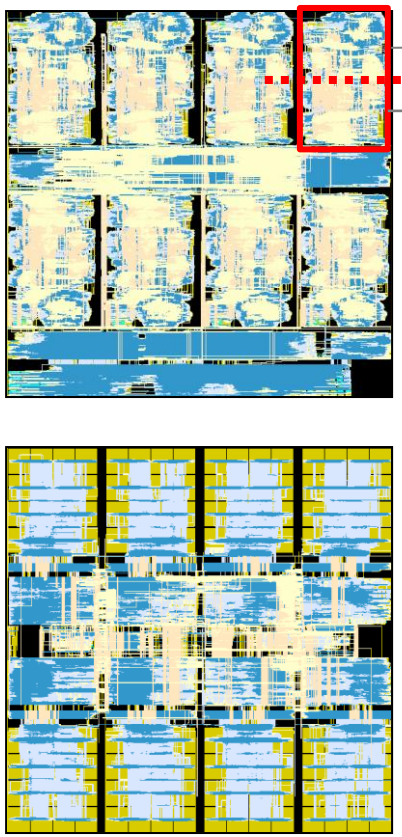
DESIGN RESULTS

WITH INTEL (2012-15)

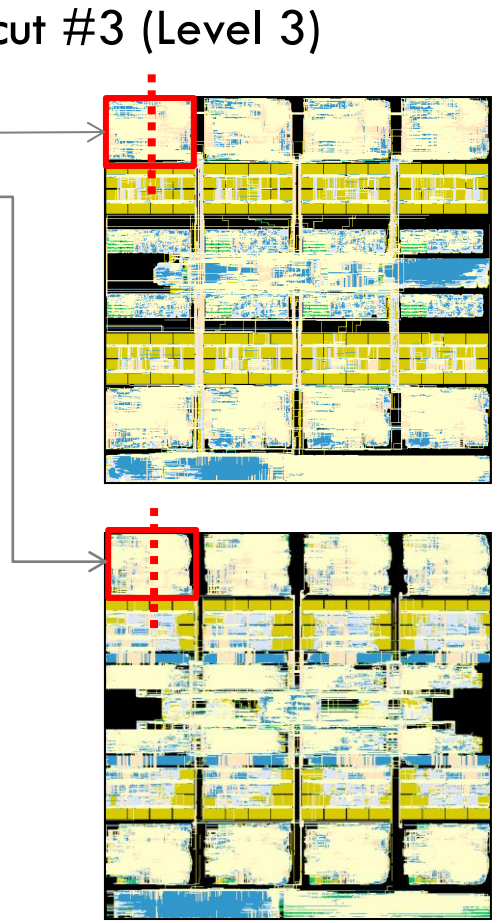
OpenSPARC T2 @ 28nm



2D

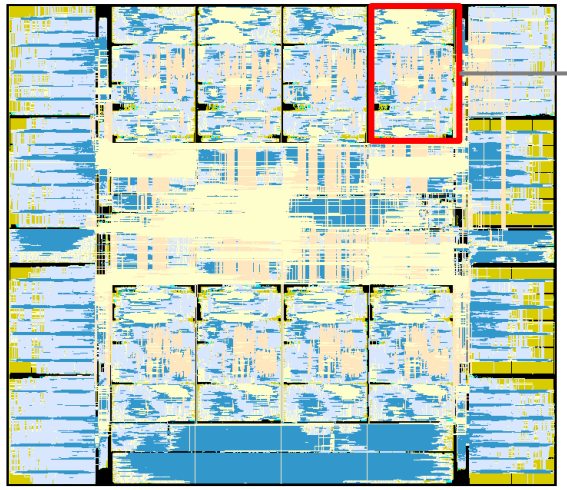


Die fold (Level 1)

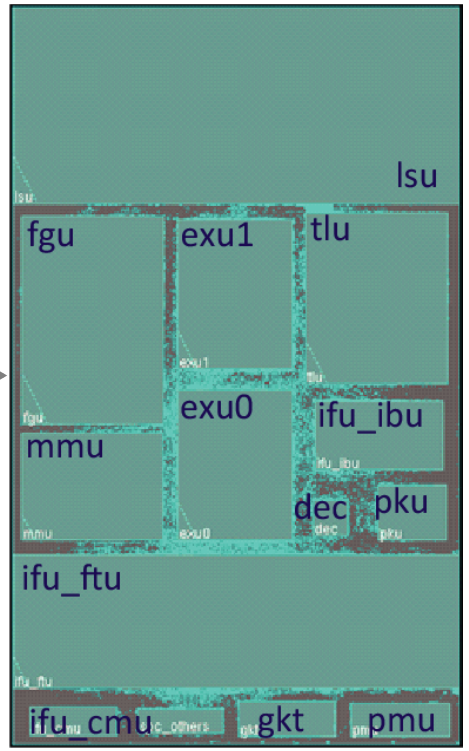


Module fold (Level 2)

MODULE FOLDING (LEVEL 2)

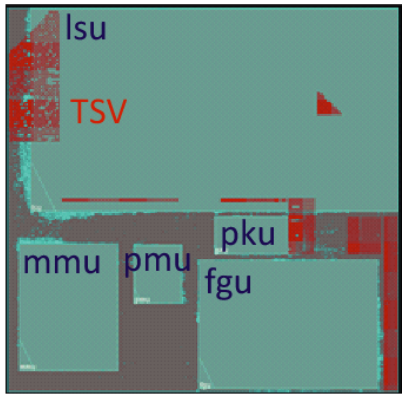


T2 fill chip

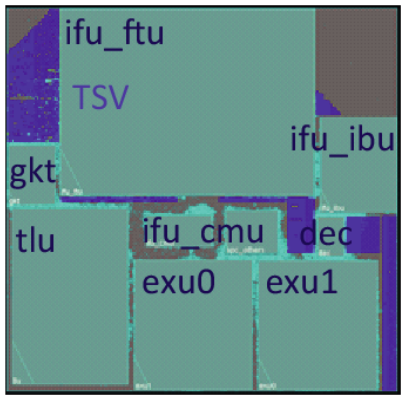


Unfolded core

Top tier



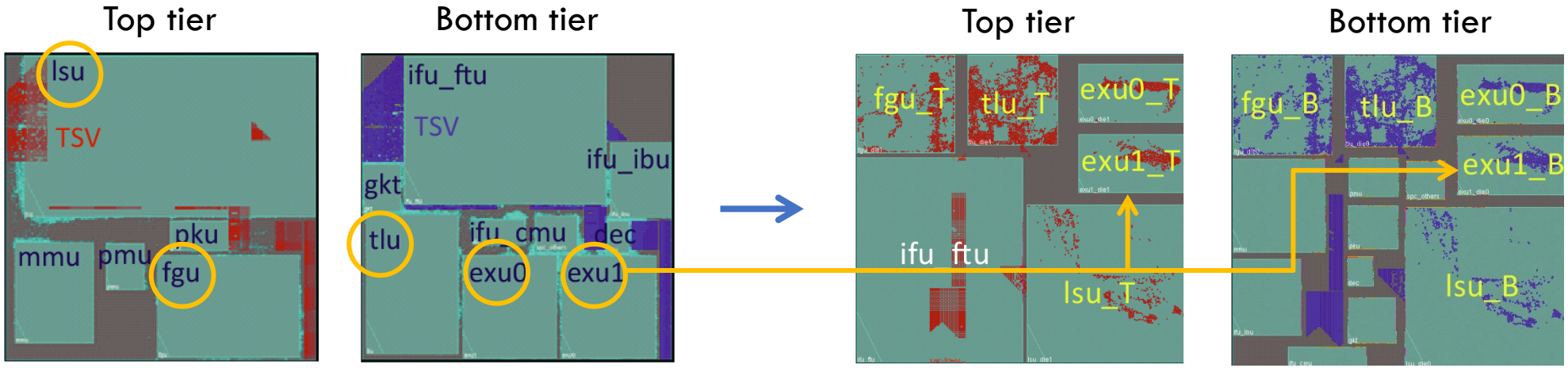
Bottom tier



Folded core

13.3% shorter wirelength
11.2% fewer buffers
14.2% lower power

SUB-MODULE FOLDING (LEVEL 3)

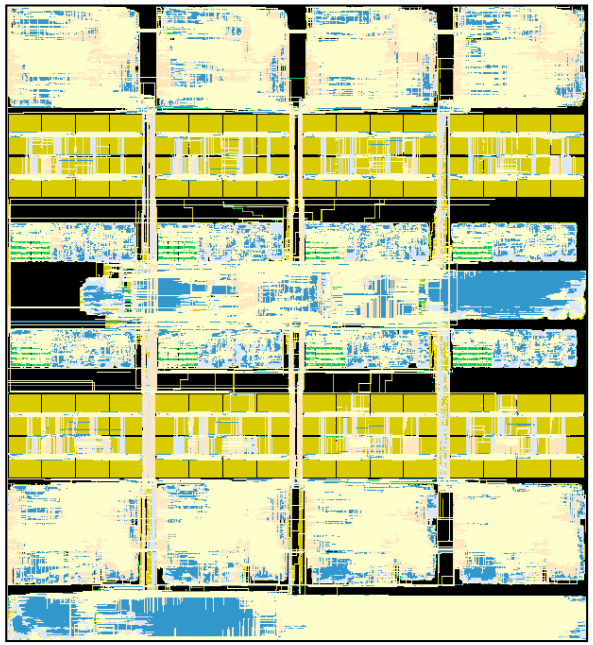


Folded core (Level 2)

Fold 5 sub-modules (Level 3)

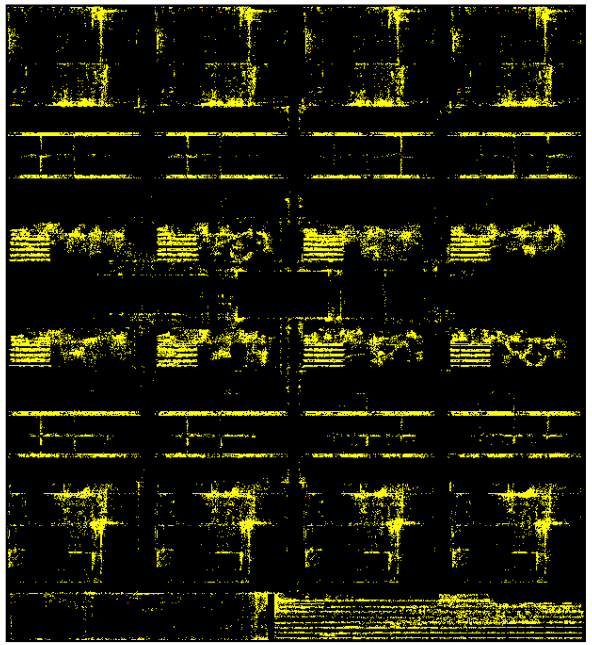
9.2% shorter wirelength
10.8% fewer buffers
5.1% lower power

F2F PAD USAGE



Bottom die

Level 1 folding
F2F pads: 7,263
(1um pitch, 0.1Ω, 0.2fF)



Level 3 folding
F2F pads: 101,555
(1um pitch, 0.1Ω, 0.2fF)

2D VS. 3D PPA COMPARISON

Goal is **low power**

- Under iso-performance

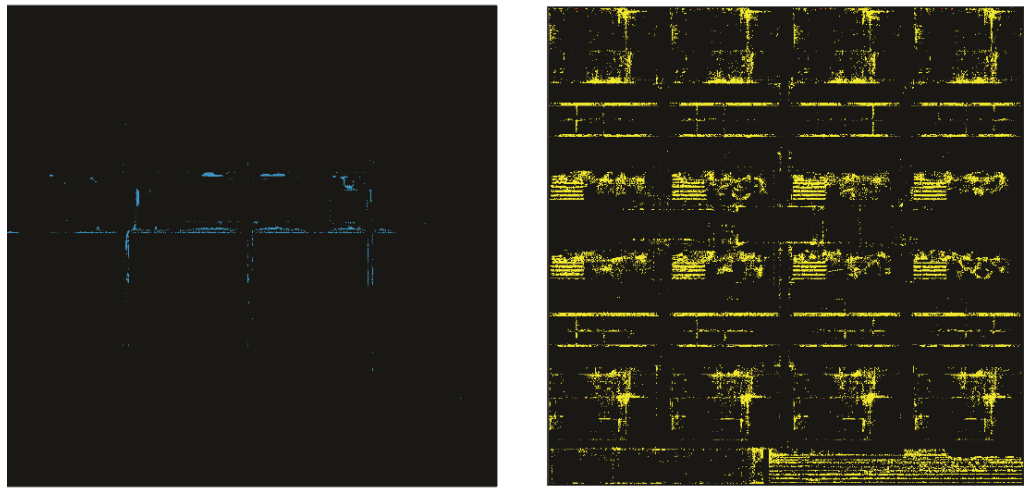
	Encounter 2D	Level 1 Folding	Level 3 Folding
Fmax (MHz)	500	500	500
Power (W)	8.2	7.1	6.6
Footprint (mm)	8x9	6x6.4	6x6.4
Tmax (°C)	45.4	61.7	57.2

This is a commercial 2D IC!

CHIPS GOT COLDER WITH FOLDING

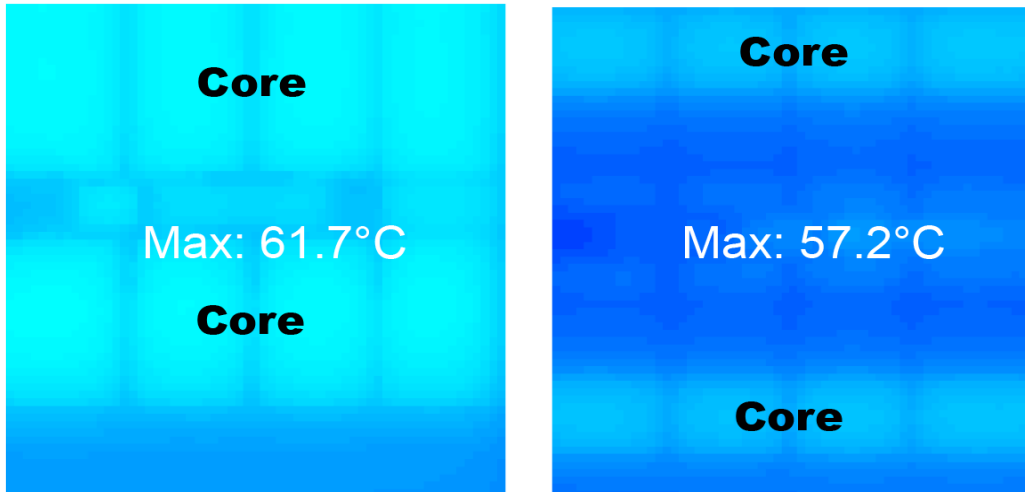
Better thermal conductivity + power consumption: colder chip

F2F pads



Level 1 Level 3

Thermal maps

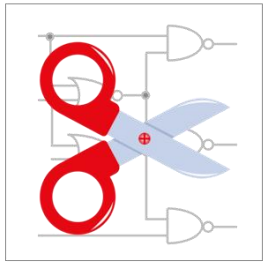


Level 1 Level 3

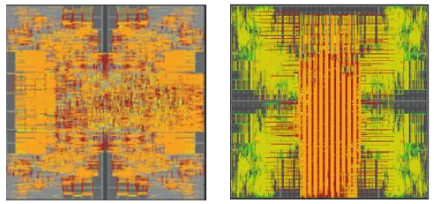
EDA READINESS

Divide-and-conquer
EDA vendors

1. Cut

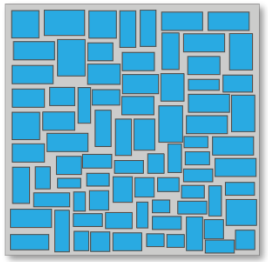


2. Design Separately

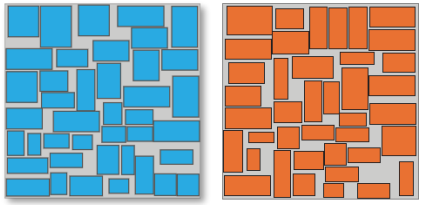


Pseudo-3D
Our work

1. Design 2D

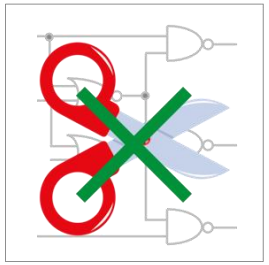


2. Convert to 3D

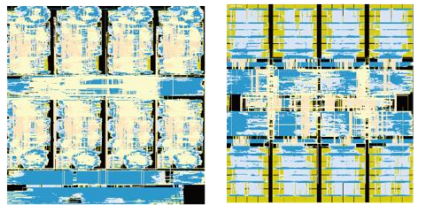


True-3D
Research

1. Do Not Cut



2. Design Together



COMPACT-2D: PRACTICAL FRONT-END

Intermediate 2D
- Place all cells together -

→

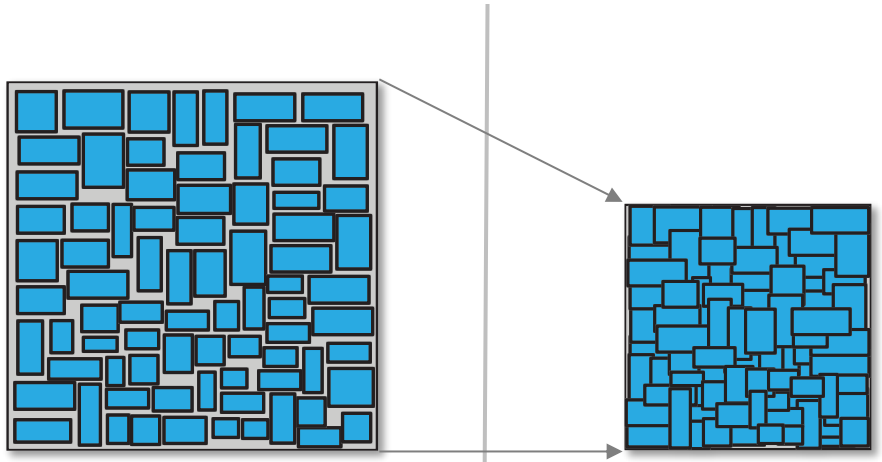
Compact
- Into 50% area -

→

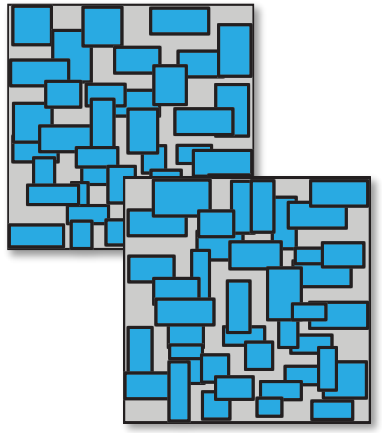
Partition
- Optimize PPA -

→

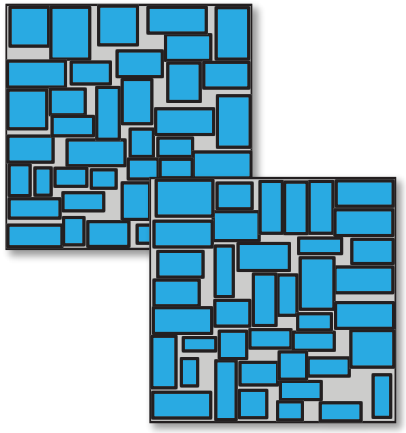
Cleanup
- Both tiers together -



We use 2D IC commercial tools.



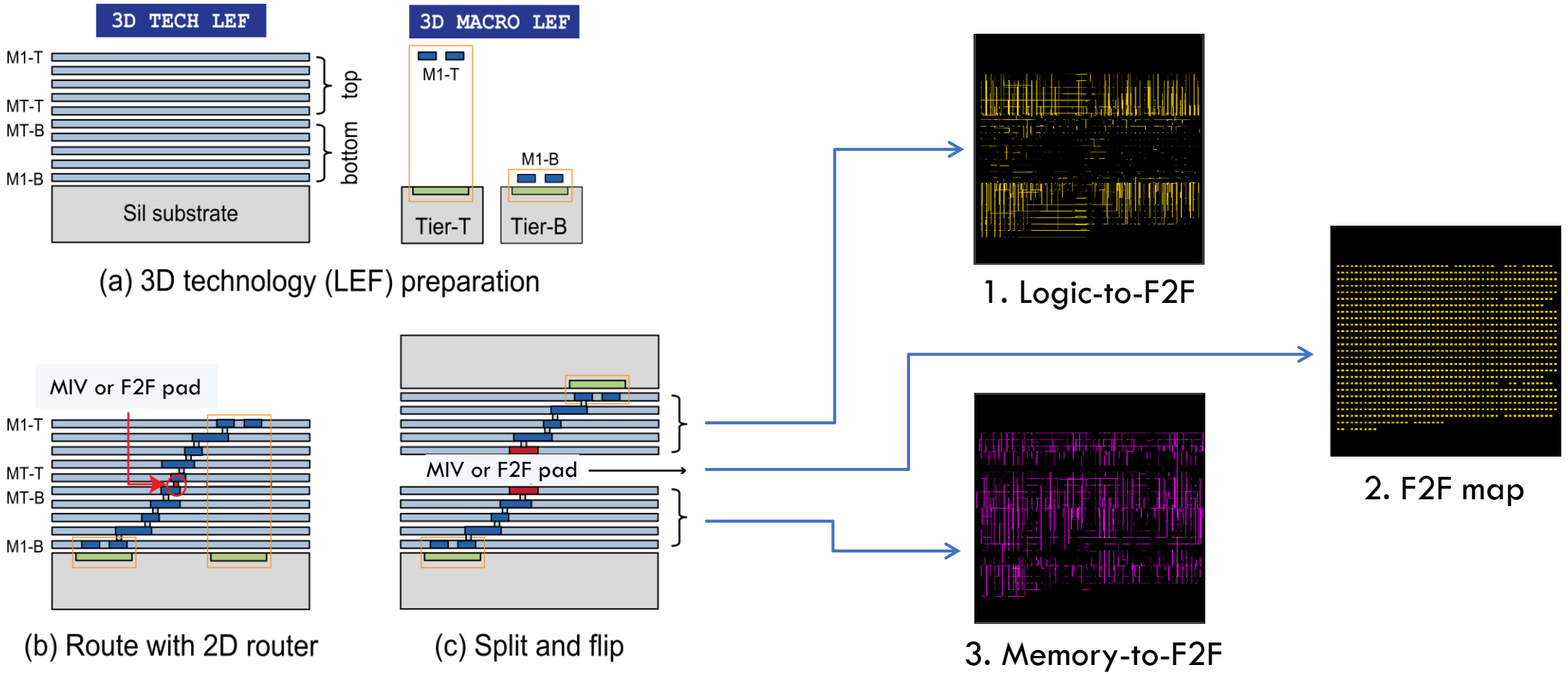
We develop new 3D IC tools.



We develop new 3D IC tools.

Bon Woong Ku, et al, "Compact-2D: A Physical Design Methodology to Build Two-Tier Gate-level 3D ICs," IEEE Trans. on CAD Best Paper Award (2022).

PIN-3D: PRACTICAL BACK-END

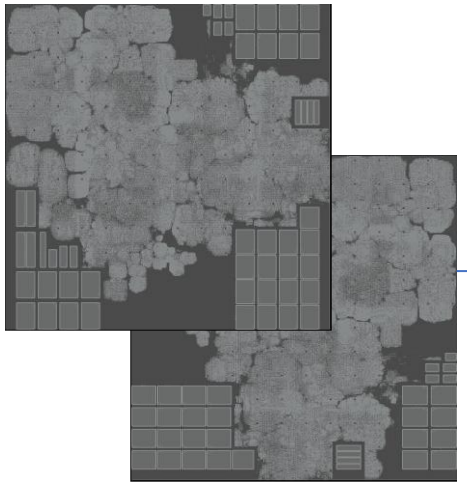


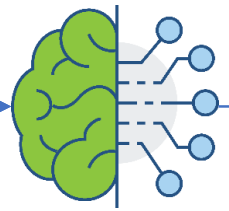
TRUE-3D WITH AI

Initial placement
= Pseudo-3D

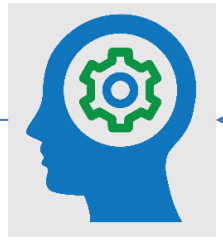
Refine placement

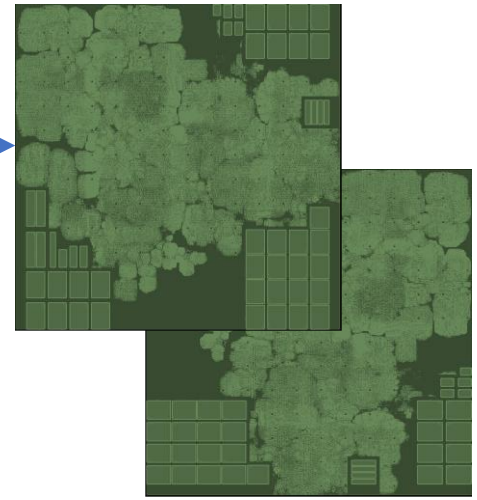
Final placement
= True-3D





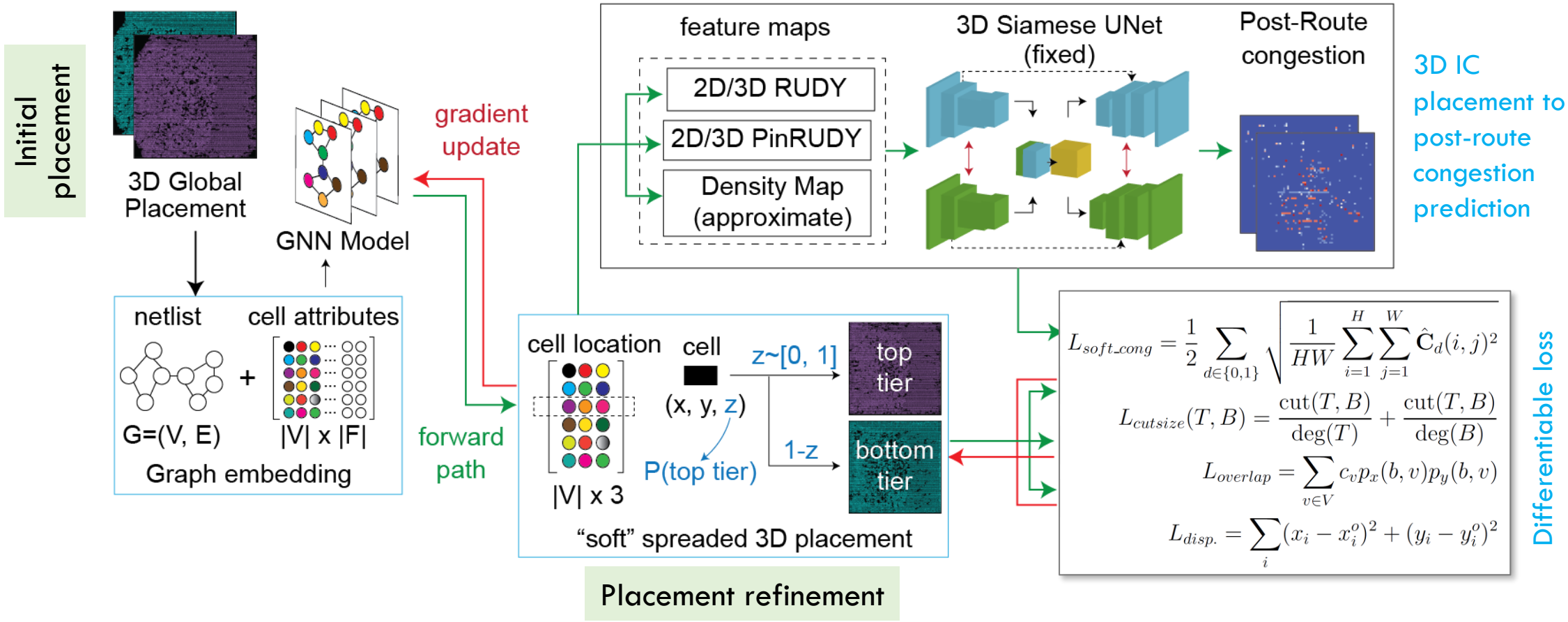
Predict QoR

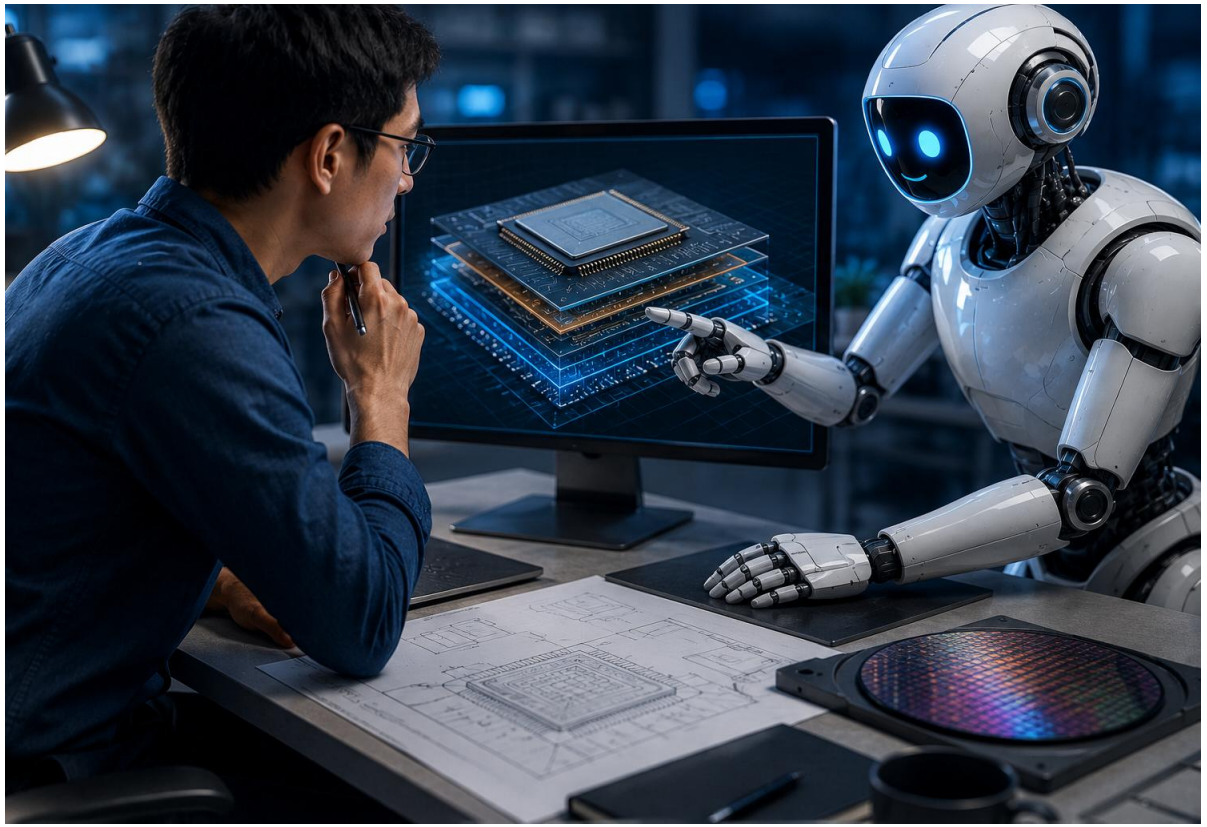




Update (x, y, z) using U-net + GNN

U-NET + GNN + DIFFERENTIABLE OPTIMIZATION





CONCLUSION

Tau Law

- Demonstrating that delay can continue to scale through 3D integration.

Folding-Based Designs

- Achieved substantial power savings through effective vertical partitioning.

EDA Tool Evolution

- Pseudo-3D design methodologies proved practical and deployable.
- More work needed: > 2 dies, “true-3D”, etc

AI-Assisted Design

- Emerging AI techniques further enhanced optimization and automation.