Looking beyond 28nm CMOS

XIII FRONT-END ELECTRONICS

For Particle Physics, Photon Science and Related Applications

18 - 22 May 2026
Paris, France

Kostas Kloukinas (CERN)

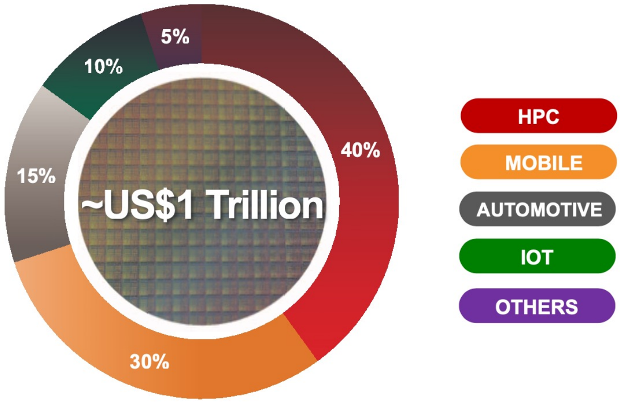
Outline

- Technology Landscape
- HEP considerations
- Outlook & sweet-spot question

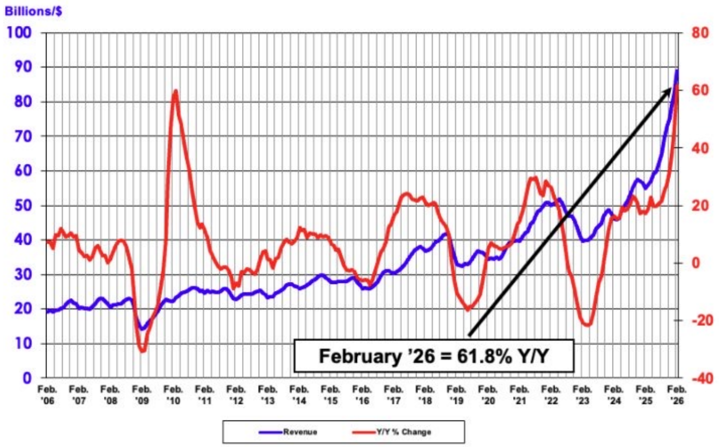
Kostas.Kloukinas@cern.ch

Semiconductor Market

- Will reach $1 trillion by 2026
- AI ignites record growth
  - Logic scaling
    - FinFET, nanosheet, GAA, CFET..
  - SRAM scaling
    - ~100 × bit cell area reduction from 130nm to 3nm
  - Compute architectures
    - From CPUs & GPUs to NPUs
  - Data rates
    - Silicon photonics, co-packaged optics
  - Power reduction
    - Also assisted by 3D interconnects



Source: TSMC Estimate



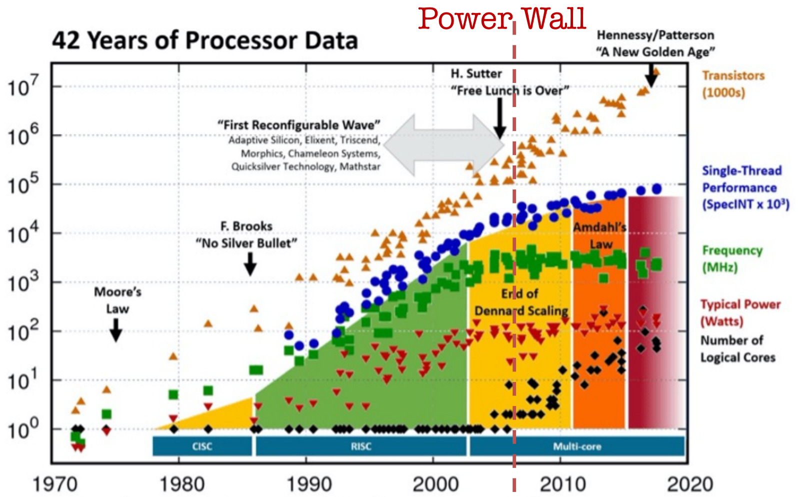
Source:
https://www.techi.com/semiconductor-industry-1-trillion-chip-boom-ai-demand/
https://bits-chips.com/article/chip-market-on-track-for-1-trillion-in-2026/

Kostas.Kloukinas@cern.ch

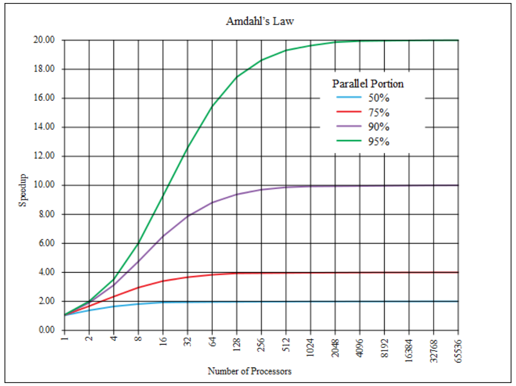
Scaling is Failing

- Moore’s Law
  - Number of transistors doubles approximately every two years.
  - Physics, costs and material constraints are slowing down the law
- Dennard’s Law
  - As transistors shrink, their power density remains constant — delivering more speed and density at no thermal cost
  - ~2005 onward (~65nm) Vt and I_{sub} leakage do not scale with size
  - Clock frequencies stopped scaling (the "frequency wall") and the industry pivoted to multicore processors
- Amdahl’s Law
  - The speedup of a system is fundamentally limited by its irreducible serial fraction, no matter how much parallelism is added
  - Multicore scaling is power limited

- Extending Moore’s Law

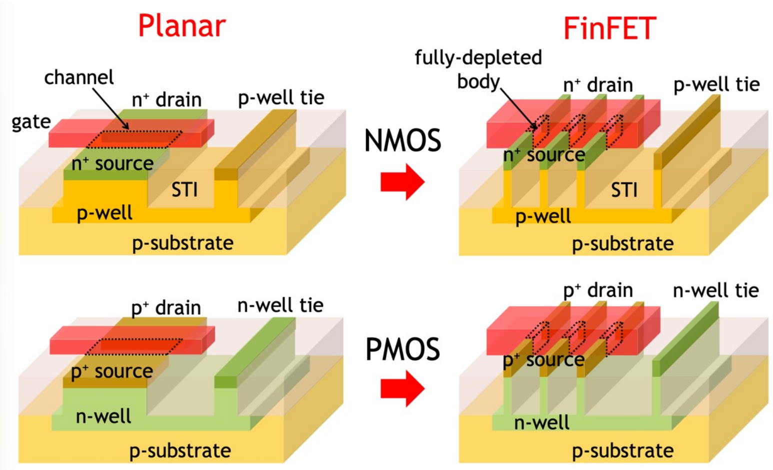


Hennessy and Patterson, Turing Lecture 2018, overlaid over “42 Years of Processors Data”
https://www.karlrupp.net/2018/02/42-years-of-microprocessor-trend-data/;“First Wave” added by Les Wilson, Frank Schirrmeister
Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten
New plot and data collected for 2010-2017 by K. Rupp

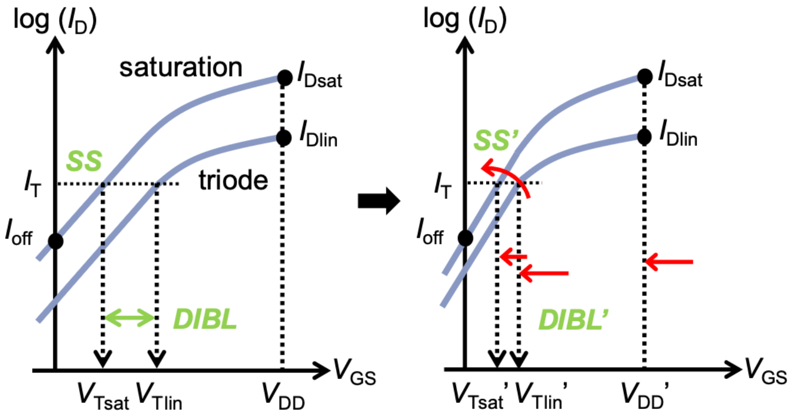


Migrating to Fully-Depleted finFET

- Subthreshold Swing (SS)
  - Planar MOSFETs: gate shares channel control with the bulk substrate → imperfect electrostatics, SS well above 60 mV/decade
  - FinFET gate wraps around three sides of the fin → near-perfect gate control → SS approaches the 60 mV/decade ideal
  - Steeper slope means the transistor switches off more abruptly → for the same I_off, threshold voltage V_T can be set lower
- DIBL (Drain-Induced Barrier Lowering)
  - Planar MOSFET: drain field penetrates deeply into the channel, lowering the source-channel barrier without gate action → large DIBL, premature conduction, poor V_T stability
  - FinFET geometry screens the channel from the drain field on three sides → DIBL reduced by 2–5 × compared to planar at the same node
  - Lower DIBL → V_T stable across operating conditions → predictable leakage and better I_off control
- Combined effect — enabling V_DD reduction
  - Together they allow V_DD to scale from ~0.9 V (28nm planar) to ~0.75 V (7nm FinFET) ~17% reduction in V_DD
  - Dynamic power scales as $P \propto CV^2f$ → ~31% reduction in dynamic power
  - Static leakage power ~5–10 × lower driven by improved SS and lower V_DD



Loke et al., Analog/Mixed-Signal Design in FinFET Technologies



Kostas.Kloukinas@cern.ch

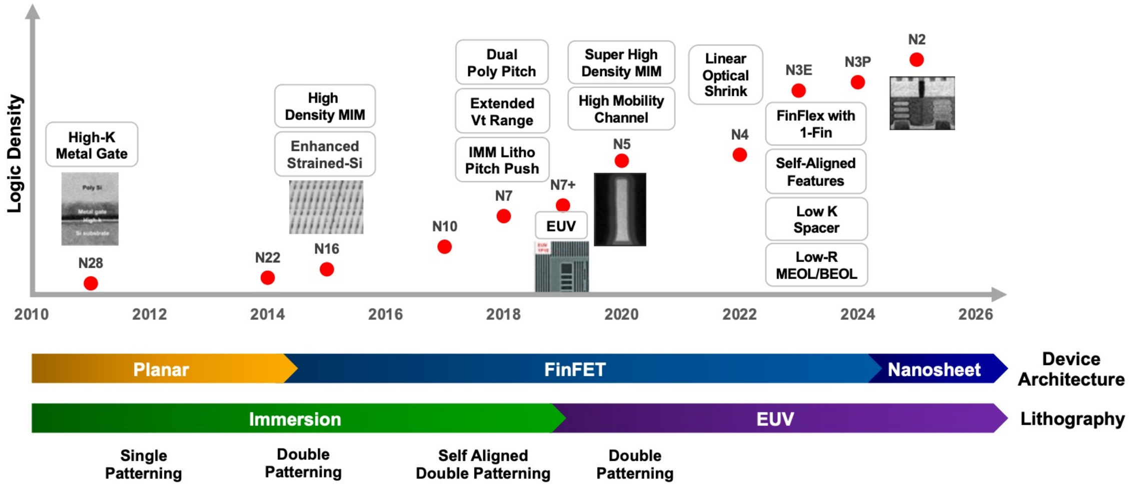
Extending Moore’s Law

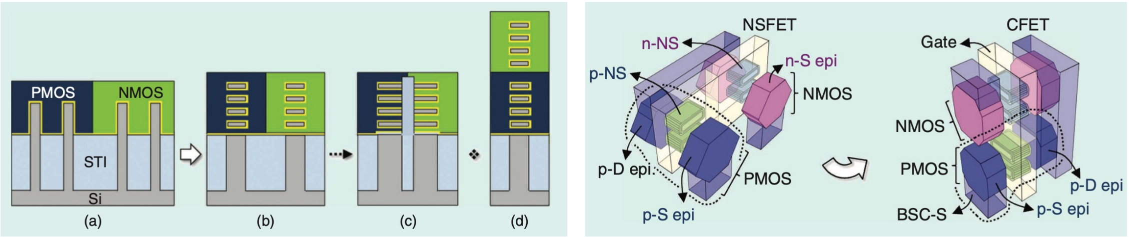
- Lithography evolution

  CD = $k_1 \times \lambda / NA$
  - UV (λ = 436 nm) -> DUV (λ = 193 nm)
  - Immersion -> Double Patterning
  - EUV (λ = 13.5 nm at 0.33 NA) single exposure
  - High-NA EUV (0.75 NA) -> Research

- Device Architecture evolution
  - Triple gate finFET
  - Gate-All-Around (GAA)
    - NSFET: Nanosheet FET
      - Enhanced electrostatic control of gate
    - FSFET: Forksheet FET
      - Minimizing gap between devices
    - CFET: Complementary FET
      - 2x the logic density
      - Monolithic or Sequential (bonded wafers)
      - ⚠ The last device architecture (?)

Possible roadmap





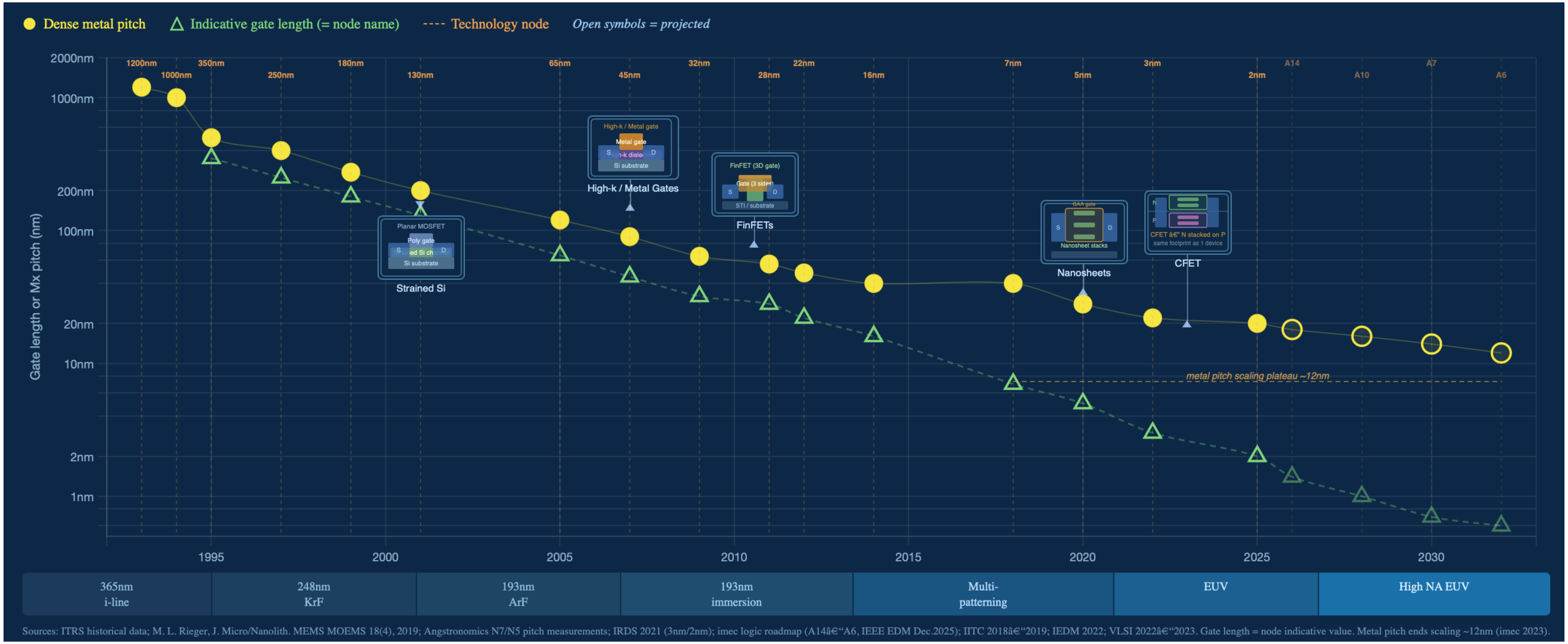
a) finFET -> b) NSFET -> c) FSFET -> d) CFET

Source: IEEE ELECTRON DEVICES MAGAZINE, DECEMBER 2025

Kostas.Kloukinas@cern.ch

Logic & Interconnect Scaling

Notable feature: the "metal pitch scaling slows/plateaus" annotation between 7nm and A6, reflecting the imec statement that metal pitch scaling ends at ~12–16nm around 2030.

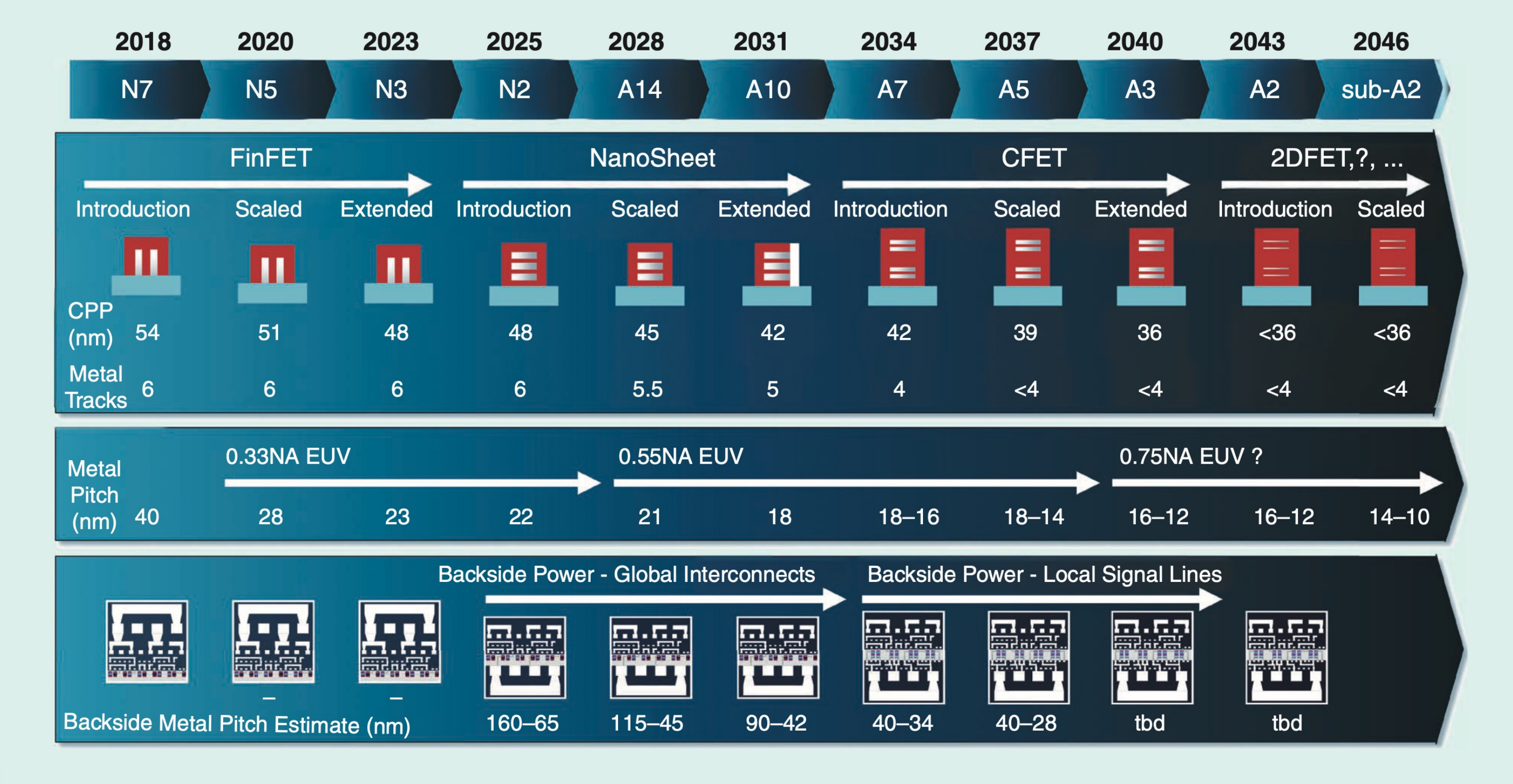


Sources: ITRS historical data; M. L. Rieger, J. Micro/Nanolith. MEMS MOEMS 18(4), 2019; Angstronomics N7/N5 pitch measurements; IRDS 2021 (3nm/2nm); imec logic roadmap (A14â€"A6, IEEE EDM Dec.2025); IITC 2018â€"2019; IEDM 2022; VLSI 2022â€"2023. Gate length = node indicative value. Metal pitch ends scaling ~12nm (imec 2023).

Kostas.Kloukinas@cern.ch

Logic Scaling Potential Roadmap

- Scaling Metrics
  - Device feature size
  - CPP: Contact Poly Pitch
  - Metal Pitch
  - Metals Tracks



	2018	2020	2023	2025	2028	2031	2034	2037	2040	2043	2046
	N7	N5	N3	N2	A14	A10	A7	A5	A3	A2	sub-A2
	FinFET			NanoSheet			CFET			2DFET,?, ...	
	Introduction	Scaled	Extended	Introduction	Scaled	Extended	Introduction	Scaled	Extended	Introduction	Scaled
CPP (nm)	54	51	48	48	45	42	42	39	36	<36	<36
Metal Tracks	6	6	6	6	5.5	5	4	<4	<4	<4	<4
		0.33NA EUV			0.55NA EUV				0.75NA EUV ?		
Metal Pitch (nm)	40	28	23	22	21	18	18–16	18–14	16–12	16–12	14–10
				Backside Power - Global Interconnects			Backside Power - Local Signal Lines				
Backside Metal Pitch Estimate (nm)		–	–	160–65	115–45	90–42	40–34	40–28	tbd	tbd	

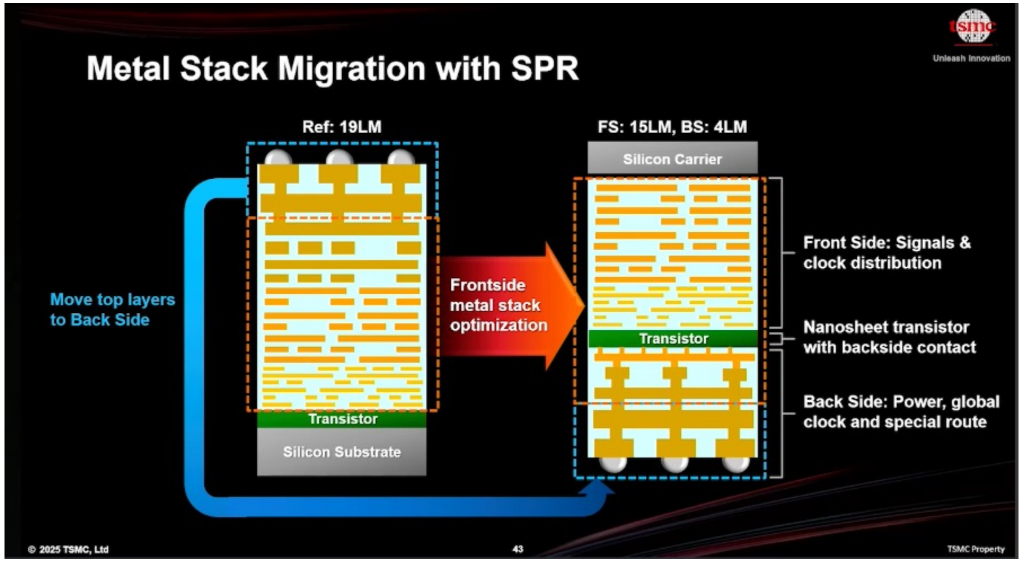
Source: IEEE ELECTRON DEVICES MAGAZINE, DECEMBER 2025

Back Side Interconnects

- BSPDN
  Back Side Power Delivery Network

- Buried Power Rail
  - ❑ Extreme wafer thinning (100nm)
  - ❑ Expect to reduce IR drop by 7x

- Global Clock & Global signals network
  - ❑ Nano TSVs to contact NSFET devices
  - ❑ Alignment precision ~10nm



Metal Stack Migration with SPR

⌂ / Imec launches the first design pathfinding process design kit for N2 node

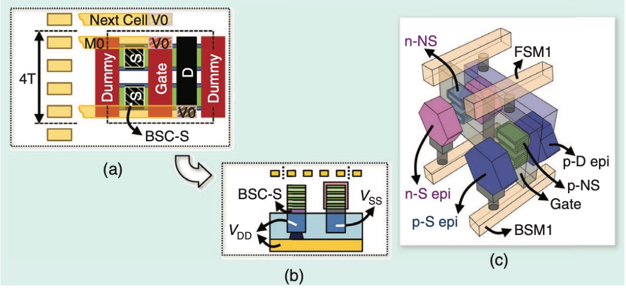
Press release

Imec launches the first design pathfinding process design kit for N2 node

The design pathfinding PDK lowers the threshold for academia and industry to access the most advanced semiconductor technologies

https://www.imec-int.com/en/press/imec-launches-first-design-pathfinding-process-design-kit-n2-node

Example of a standard cell layout (a) and a corresponding cross-section cut (b) to illustrate the BSPDN's scheme wherein each of the cell's transistors (e.g., NSFETs) is connected from the wafer's backside through a BS contact landing directly on the device's source (BSC-S). This enables more compact NSFET-based cells, as a simplified 3D schematics also shows in (c)

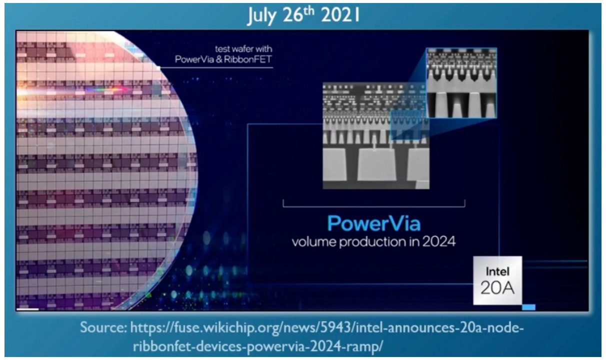


Source: IEEE ELECTRON DEVICES MAGAZINE, DECEMBER 2025
https://www.imec-int.com/en/articles/how-power-chips-backside

Kostas.Kloukinas@cern.ch

Industry’s adoption of BSPDN

- Samsung uses BSPDN technology for 2nm chips, improving performance by 44% and efficiency by 30%

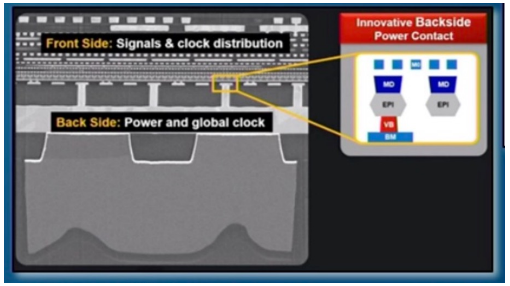


July 26th 2021

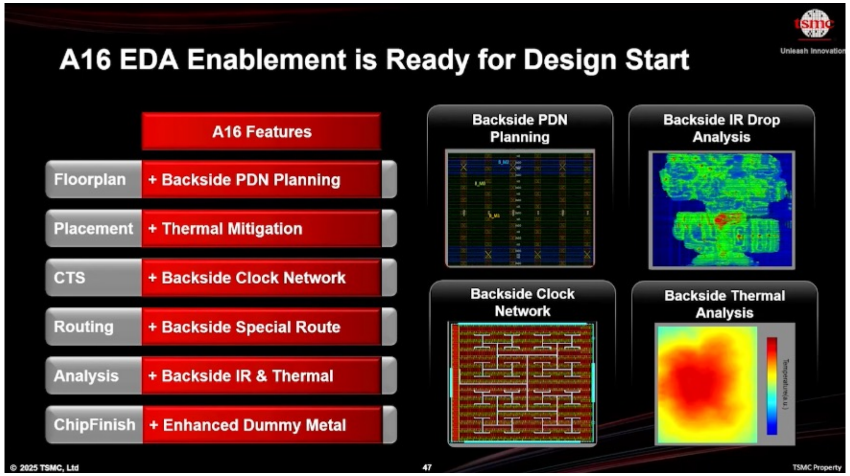
PowerVia
volume production in 2024

Source: https://fuse.wikichip.org/news/5943/intel-announces-20a-node-ribbonfet-devices-powervia-2024-ramp/

- TSMC Outlines 2nm Plans: N2P Brings Backside Power Delivery in 2026, N2X Added To Roadmap



TSMC A16TM @ TSMC TechSymposium, USA, April 2024



A16 EDA Enablement is Ready for Design Start

	A16 Features
Floorplan	+ Backside PDN Planning
Placement	+ Thermal Mitigation
CTS	+ Backside Clock Network
Routing	+ Backside Special Route
Analysis	+ Backside IR & Thermal
ChipFinish	+ Enhanced Dummy Metal

© 2025 TSMC, Ltd

47

TSMC Property

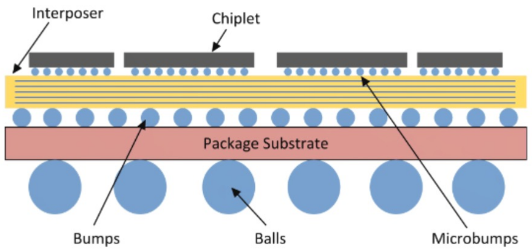
Kostas.Kloukinas@cern.ch

Chiplets: Reversing Integration

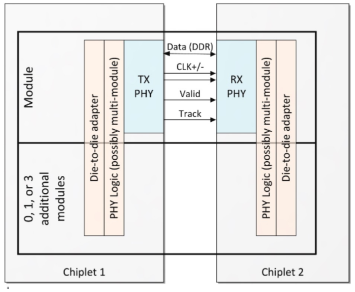
- **Chiplets** are the response to the end of monolithic scaling economics
  - Decompose the SoC into smaller, functionally-specialized dies ("chiplets")
  - Manufacture each chiplet in its optimal process node
  - Integrate them in a package via a high-bandwidth die-to-die interconnect

  - 2.5D interposer-based
    - Silicon interposer: passive Si with dense RDL, bump pitch ~45–55 µm,
    - Organic interposer: cheaper, lower density

  - 3D die stacking
    - Hybrid bonding: Cu-Cu thermocompression, pitch down to 1–3 µm, eliminates bumps entirely, bandwidth density ~10–100 × over flip-chip
    - Face-to-face or face-to-back configurations depending on power delivery constraints



UCIe Chiplet Interconnect protocol



Kostas.Kloukinas@cern.ch

Chiplets at HEP

- Disaggregate the analog front-end (noise-optimized) from the digital hit-processing fabric and the high-speed serializer PHY

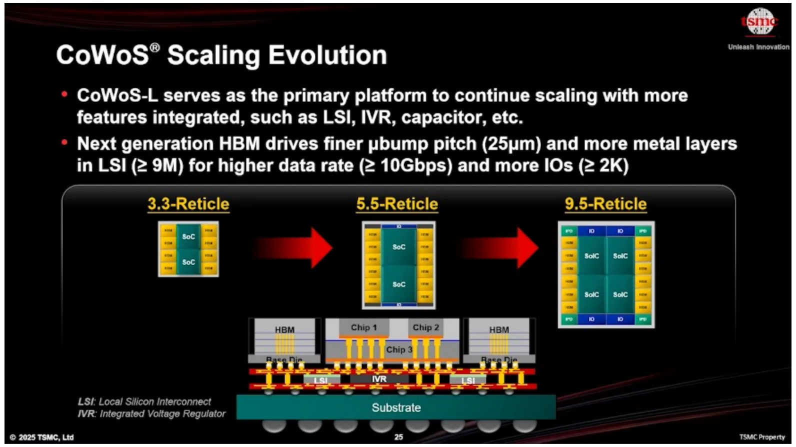
- Structural reconfigurability:
  - analog die is fixed and qualified
  - the digital processing chiplet is retargeted per experiment and application

- Alternative (non-chiplet approach):
  - Parameter-programmable not structurally reconfigurable
  - Configurable modes (ToT/ToA modes, per-pixel threshold trim, multiple readout modes. Still a fixed architecture underneath)
  - Analog front-end parameterization has limits

Kostas.Kloukinas@cern.ch

2.5 - 3D Interconnects

CoWoS: Chip-on-Wafer-on-Substrate



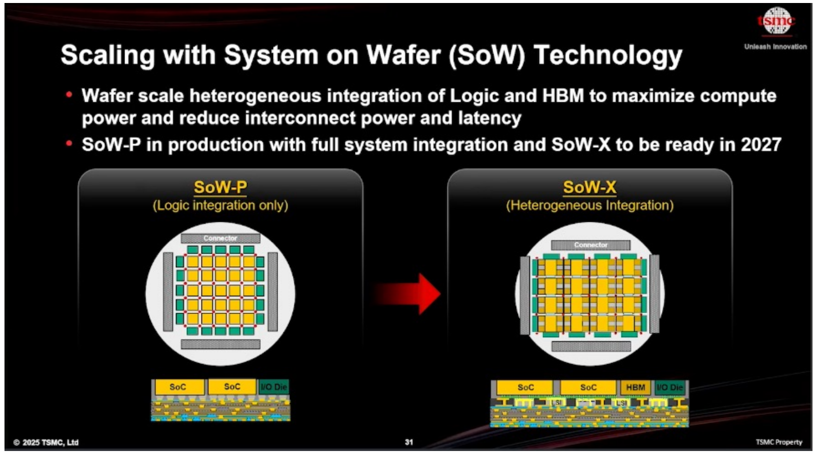
CoWoS® Scaling Evolution

- CoWoS-L serves as the primary platform to continue scaling with more features integrated, such as LSI, IVR, capacitor, etc.
- Next generation HBM drives finer µbump pitch (25µm) and more metal layers in LSI (≥ 9M) for higher data rate (≥ 10Gbps) and more IOs (≥ 2K)

LSI: *Local Silicon Interconnect*
IVR: *Integrated Voltage Regulator*

© 2025 TSMC, Ltd 25 TSMC Property

SoW: System-on-Wafer

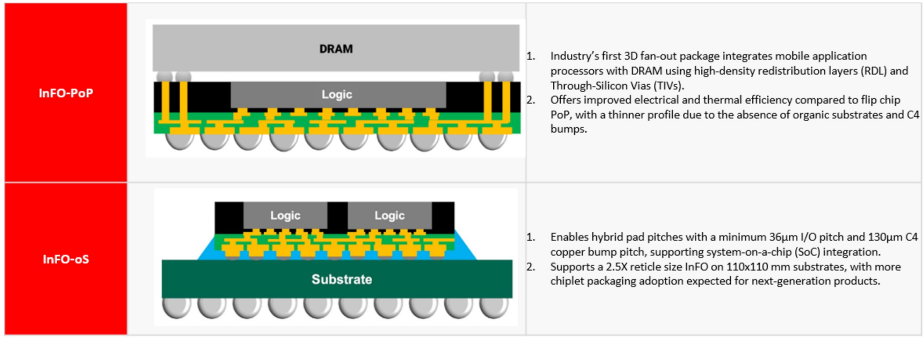


Scaling with System on Wafer (SoW) Technology

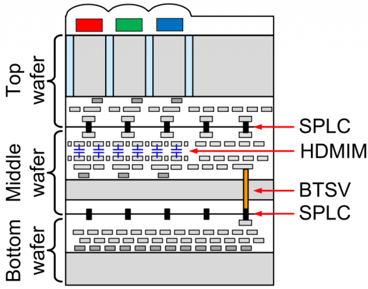
- Wafer scale heterogeneous integration of Logic and HBM to maximize compute power and reduce interconnect power and latency
- SoW-P in production with full system integration and SoW-X to be ready in 2027

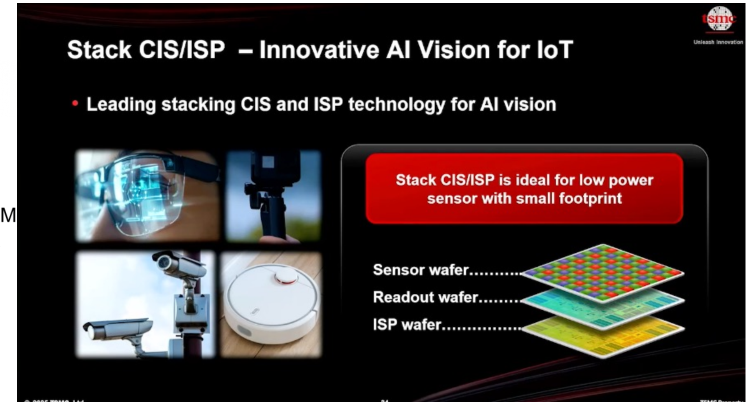
© 2025 TSMC, Ltd 31 TSMC Property

InFO: Integrated Fan-Out



InFO-PoP	1. Industry's first 3D fan-out package integrates mobile application processors with DRAM using high-density redistribution layers (RDL) and Through-Silicon Vias (TIVs). 2. Offers improved electrical and thermal efficiency compared to flip chip PoP, with a thinner profile due to the absence of organic substrates and C4 bumps.
InFO-oS	1. Enables hybrid pad pitches with a minimum 36µm I/O pitch and 130µm C4 copper bump pitch, supporting system-on-a-chip (SoC) integration. 2. Supports a 2.5X reticle size InFO on 110x110 mm substrates, with more chiplet packaging adoption expected for next-generation products.





Stack CIS/ISP – Innovative AI Vision for IoT

- Leading stacking CIS and ISP technology for AI vision

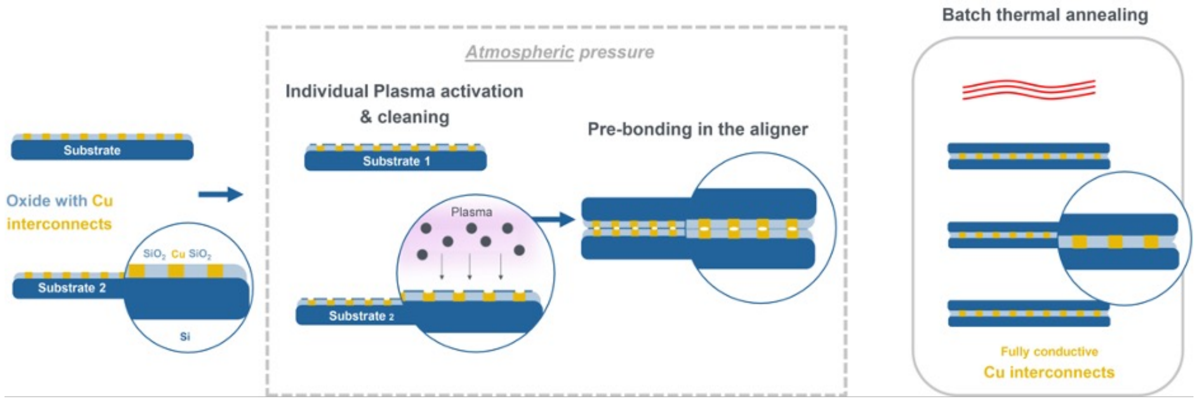
Stack CIS/ISP is ideal for low power sensor with small footprint

3D Interconnects

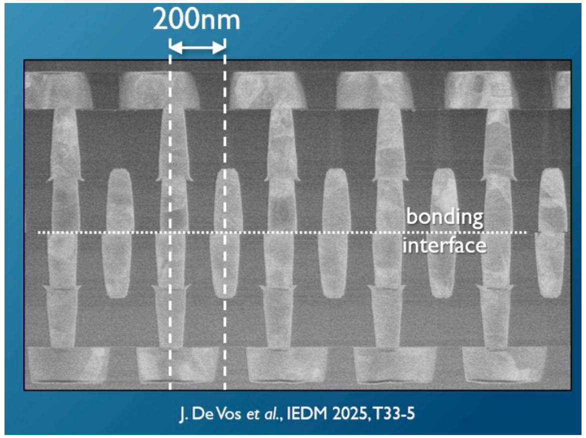
- Hybrid bonding: Wafer to Wafer hybrid bonding delivers unprecedented interlayer connectivity
- Tight 3D integration => shorter wires => Low Power

$$\text{Energy/op} \sim (C_{FET} + L_{WIRE} \bullet C_{WIRE}) \bullet V^2$$

- Low Noise for “sensor wafer” – ”readout ASIC wafer” bonding



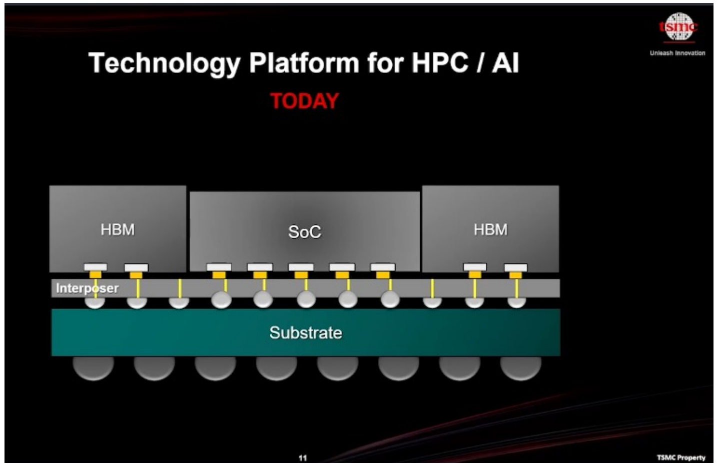
Wafer-to-wafer hybrid bonding process flow. Source: EV Group



J. De Vos *et al.*, IEDM 2025, T33-5

Kostas.Kloukinas@cern.ch

Interconnects & Photonics

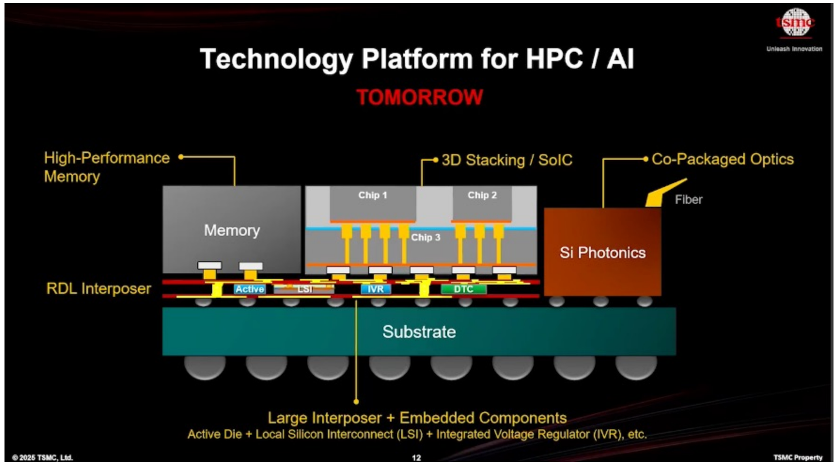


Technology Platform for HPC / AI

TODAY

11

TSMC Property



Technology Platform for HPC / AI

TOMORROW

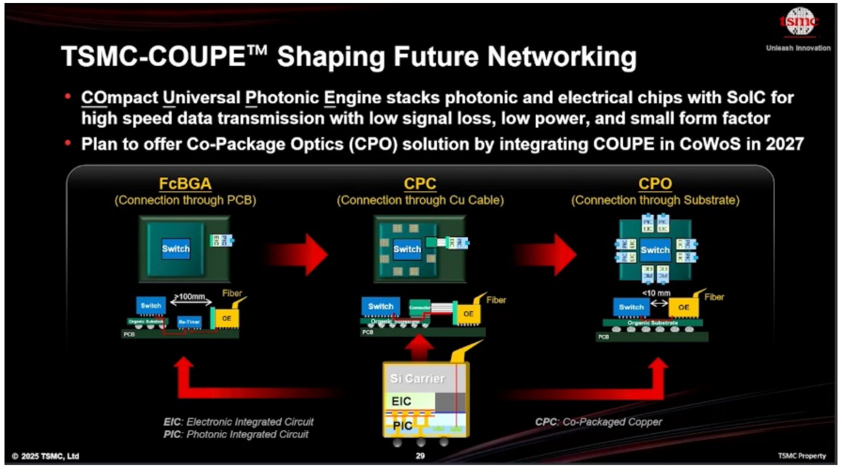
Large Interposer + Embedded Components

Active Die + Local Silicon Interconnect (LSI) + Integrated Voltage Regulator (IVR), etc.

© 2025 TSMC, Ltd.

12

TSMC Property



TSMC-COUPE™ Shaping Future Networking

- COmpact Universal Photonic Engine stacks photonic and electrical chips with SoIC for high speed data transmission with low signal loss, low power, and small form factor
- Plan to offer Co-Package Optics (CPO) solution by integrating COUPE in CoWoS in 2027

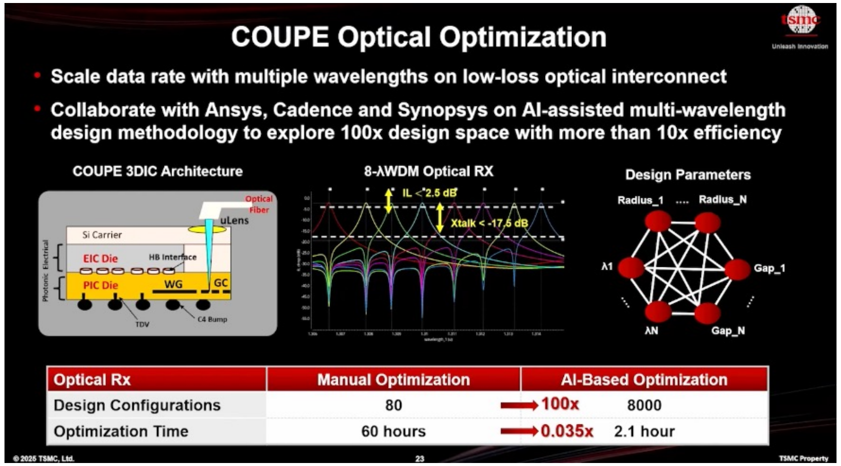
EIC: Electronic Integrated Circuit
PIC: Photonic Integrated Circuit

CPC: Co-Packaged Copper

© 2025 TSMC, Ltd

29

TSMC Property



COUPE Optical Optimization

- Scale data rate with multiple wavelengths on low-loss optical interconnect
- Collaborate with Ansys, Cadence and Synopsys on AI-assisted multi-wavelength design methodology to explore 100x design space with more than 10x efficiency

COUPE 3DIC Architecture

8-λWDM Optical RX

Design Parameters

Optical Rx	Manual Optimization	AI-Based Optimization
Design Configurations	80	→ 100x 8000
Optimization Time	60 hours	→ 0.035x 2.1 hour

© 2025 TSMC, Ltd.

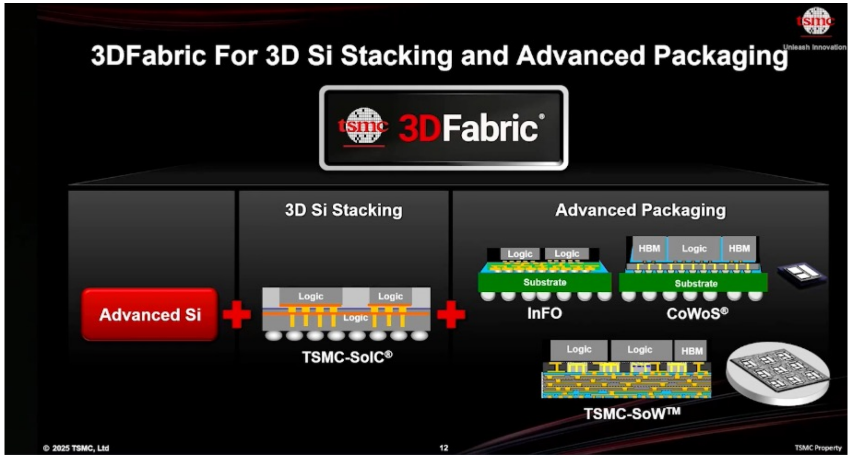
23

TSMC Property

Kostas.Kloukinas@cern.ch

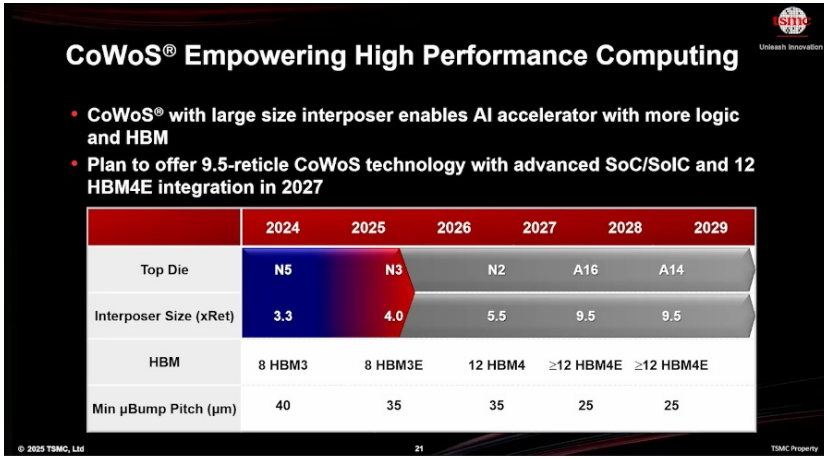
2.5 - 3D Interconnects

- Standardized ecosystem
- Available on “high-end” nodes
  - Minimum engagement volumes are calibrated to commercial customers, not HEP runs of a few hundred modules.
- Use of specialized EDA tools for electrothermal and mechanical analysis and optimization



3DFabric For 3D Si Stacking and Advanced Packaging

© 2025 TSMC, Ltd 12 TSMC Property

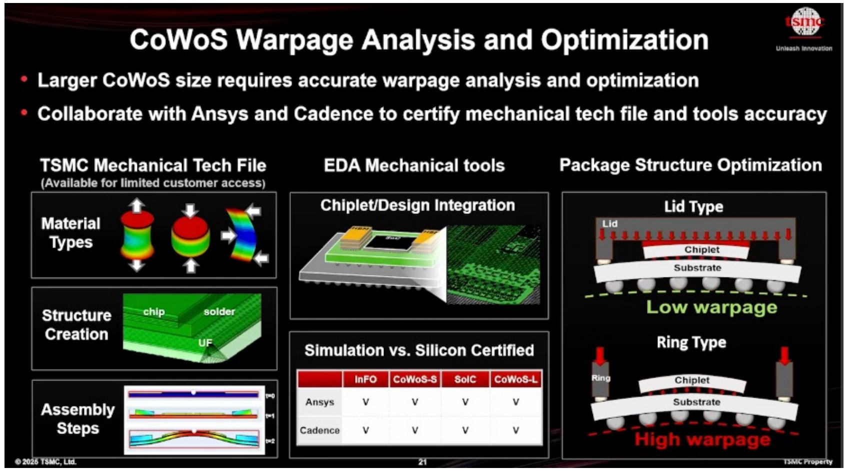


CoWoS® Empowering High Performance Computing

- CoWoS® with large size interposer enables AI accelerator with more logic and HBM
- Plan to offer 9.5-reticle CoWoS technology with advanced SoC/SoIC and 12 HBM4E integration in 2027

	2024	2025	2026	2027	2028	2029
Top Die	N5	N3	N2	A16	A14	
Interposer Size (xRet)	3.3	4.0	5.5	9.5	9.5	
HBM	8 HBM3	8 HBM3E	12 HBM4	≥12 HBM4E	≥12 HBM4E	
Min µBump Pitch (µm)	40	35	35	25	25	

© 2025 TSMC, Ltd 21 TSMC Property



CoWoS Warpage Analysis and Optimization

- Larger CoWoS size requires accurate warpage analysis and optimization
- Collaborate with Ansys and Cadence to certify mechanical tech file and tools accuracy

TSMC Mechanical Tech File
(Available for limited customer access)

EDA Mechanical tools

Simulation vs. Silicon Certified

	InFO	CoWoS-S	SoIC	CoWoS-L
Ansys	V	V	V	V
Cadence	V	V	V	V

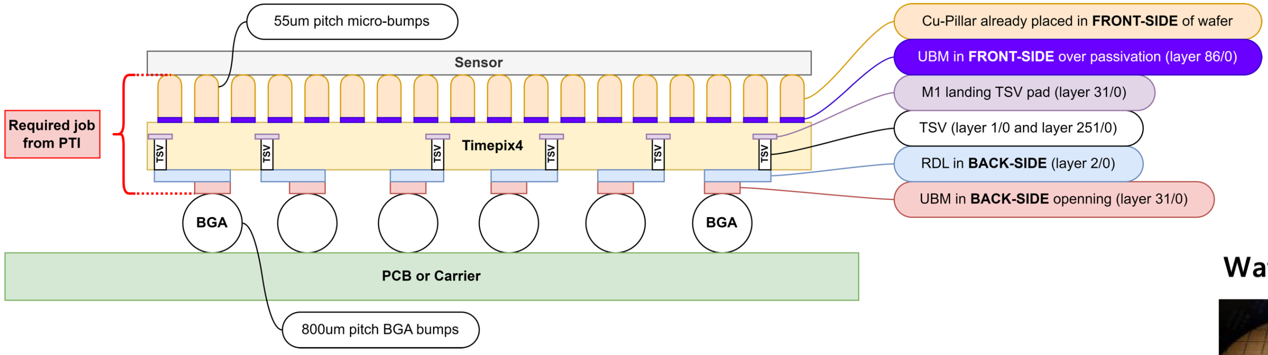
Package Structure Optimization

© 2025 TSMC, Ltd. 21 TSMC Property

Kostas.Kloukinas@cern.ch

Interconnects R&D at CERN using Timepix4 (65nm)

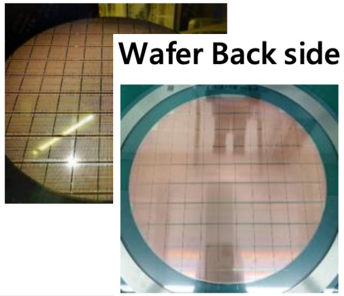
Requested Job option
1) TSV from M1 to back side
2) Back-side Cu RDL
3) UBM for BGA interconnect in Back-side
4) UBM for Cu-Pillars on Front-Side
5) Cu-Pillars on the Front-Side
6) Dicing
7) Delivery in GelPak or blue tape



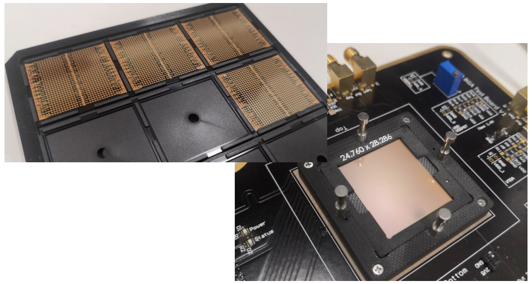
Cu-Pillar already placed in **FRONT-SIDE** of wafer
UBM in **FRONT-SIDE** over passivation (layer 86/0)
M1 landing TSV pad (layer 31/0)
TSV (layer 1/0 and layer 251/0)
RDL in **BACK-SIDE** (layer 2/0)
UBM in **BACK-SIDE** openning (layer 31/0)

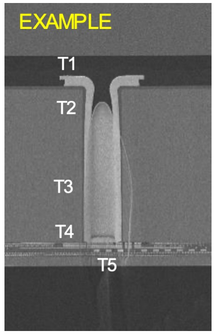
Wafer Front side

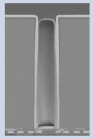
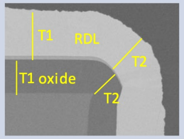
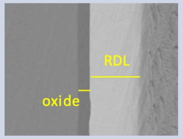
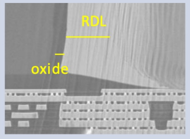
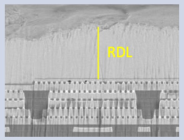
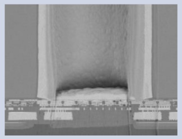
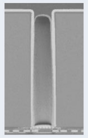
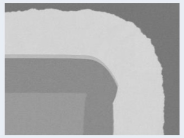
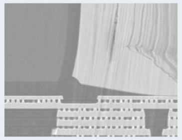
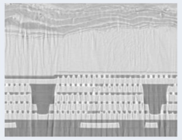
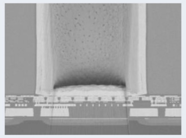
Wafer Back side



- Experience with several vendors for TSV Last processing
- Very promising results from a non-European vendor
  - For the first time we were able to process 300mm wafers
  - Just 0.1% dead pixels after processing. 5 wafers (73 chips / wafer) available @CERN. More on the way
  - Vendor announced EOL of TSV-last processing !!! 🥀 😟

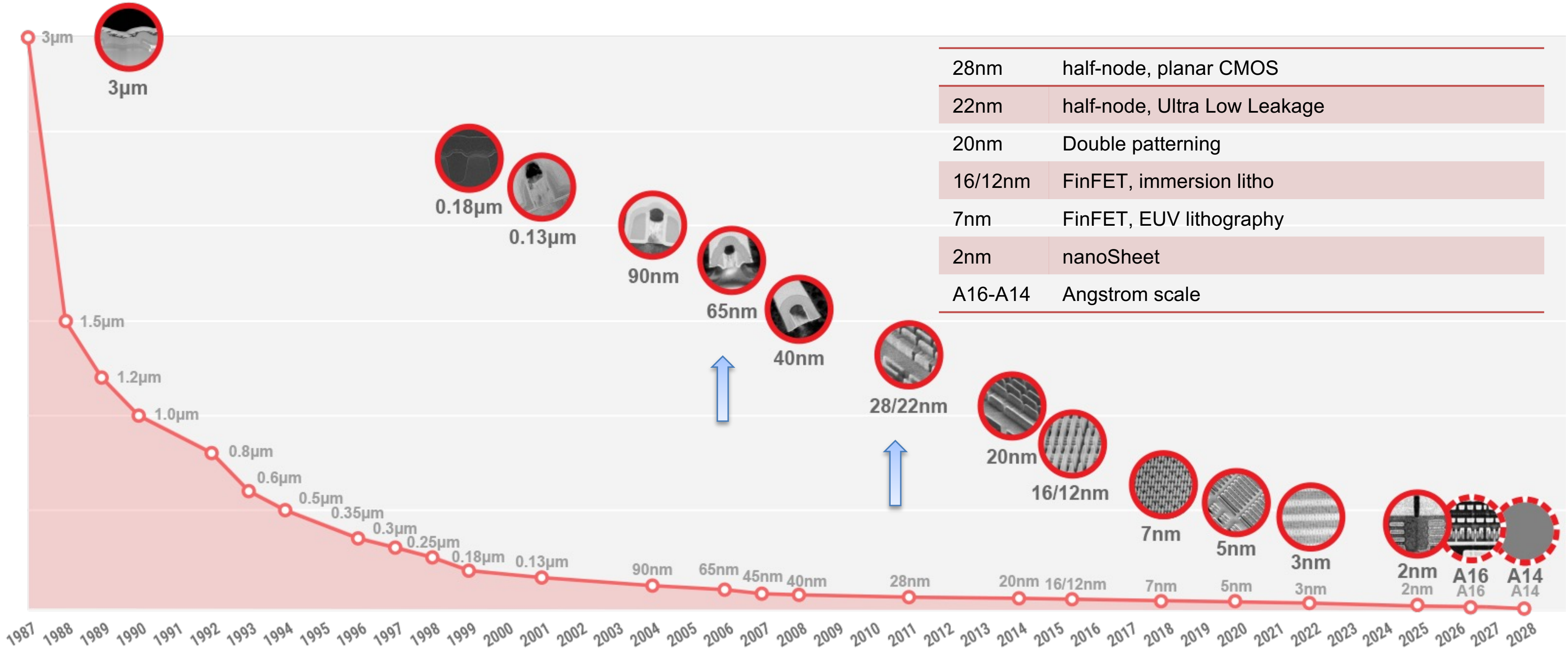




	Full view	T1, T2	T3	T4	T5	Bottom check
Center						
Edge						

Kostas.Kloukinas@cern.ch

Logic Scaling



28nm	half-node, planar CMOS
22nm	half-node, Ultra Low Leakage
20nm	Double patterning
16/12nm	FinFET, immersion litho
7nm	FinFET, EUV lithography
2nm	nanoSheet
A16-A14	Angstrom scale

PERFORMANCE COMPARED TO 65 NM:

x 4-5 GATE DENSITY INCREASE

> x2 FASTER

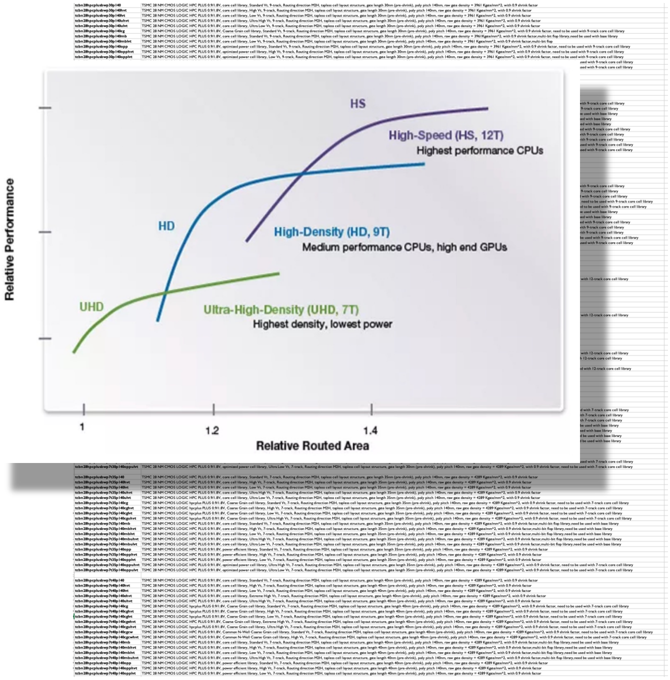
➢ x50 LEAKAGE INCREASE – CAN BE REDUCED EXPLOITING MULTI-VT and MULTI-GL DESIGNS

TSMC foundry process technologies:
https://www.tsmc.com/english/dedicatedFoundry/technology/logic

Kostas.Kloukinas@cern.ch

Exploiting performance at 28nm

- 120 different digital library sets
- Optimized implementation for Performance and Power



tcbn28hpcbwp35p140

30 nm
35 nm
40 nm

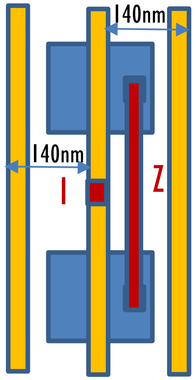
Faster

Less Power

Gate length

HPC+ 12T	Raw Gate Density Kgates/mm2	Speed GHz	Active Power nW/MHz	Static Power nW/MHz
30nm	2,971	1	1	1
35nm	2,971	0.9	0.992	0.446
40nm	2,971	0.806	1.003	0.232

Poly pitch



7, 9, 12

Tracks

HPC+ 30P	Raw Gate Density Kgates/mm2	Speed GHz	Active Power nW/MHz	Static Power nW/MHz
7T	4,289	0.742	0.601	0.485
9T	3,961	0.886	0.781	0.714
12T	2,971	1	1	1

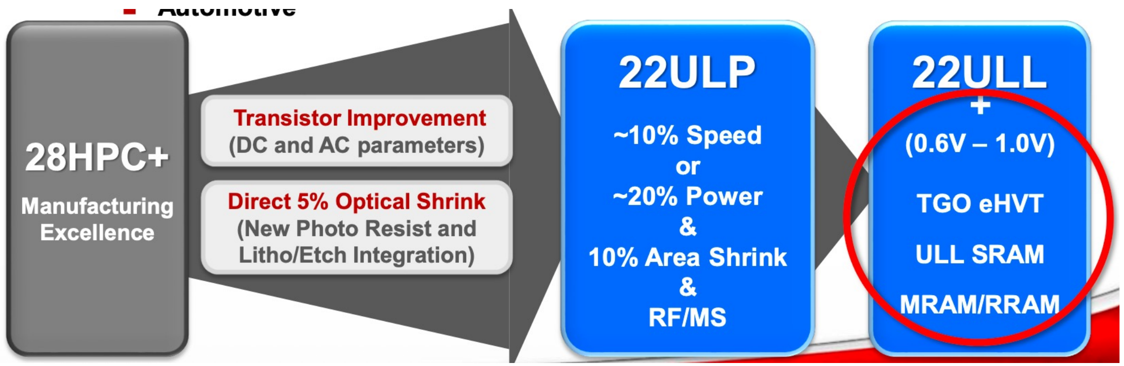
SVT, LVT, HVT, UHVT, ULVHT

Voltage threshold

HPC+ 12T 30P	Raw Gate Density Kgates/mm2	Speed GHz	Active Power nW/MHz	Static Power nW/MHz
LVT	2,971	1.351	1.235	5.305
SVT	2,971	1	1	1
HVT	2,971	0.654	0.911	0.127
EHVT	2,971	0.229	0.903	0.019

Kostas.Kloukinas@cern.ch

Jumping to 22nm ULL & ULP



- Automotive

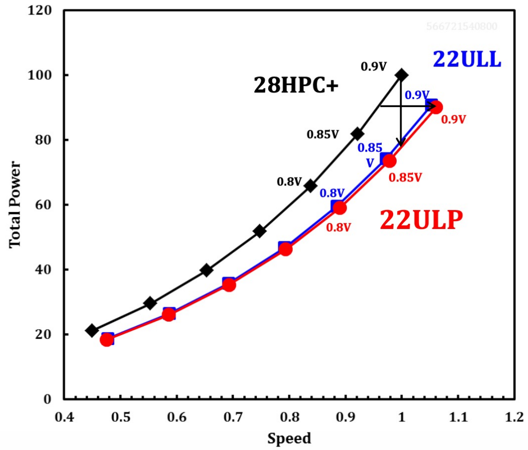
28HPC+
Manufacturing Excellence

Transistor Improvement
(DC and AC parameters)

Direct 5% Optical Shrink
(New Photo Resist and Litho/Etch Integration)

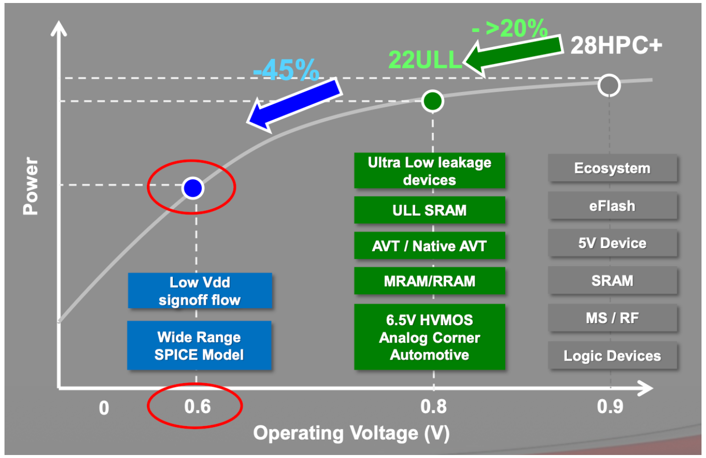
22ULP
~10% Speed
or
~20% Power
&
10% Area Shrink
&
RF/MS

22ULL
+
(0.6V – 1.0V)
TGO eHVT
ULL SRAM
MRAM/RRAM



- 22 ULP: Ultra Low Power
  - uLVT, LVT, SVT, HVT, UHVT
- 22 ULL: Ultra Low Leakage
  - uLVT, LVT, SVT, HVT, UHVT, eHVT (+Thick Gate)
  - SRAM ULL
  - Low Vdd solution, down to 0.6V
  - Foundry Low-Vdd libraries
  - MRAM/RRAM offering on MPWs

TID response ?



Low Vdd signoff flow
Wide Range SPICE Model

Ultra Low leakage devices
ULL SRAM
AVT / Native AVT
MRAM/RRAM
6.5V HVMOS Analog Corner Automotive

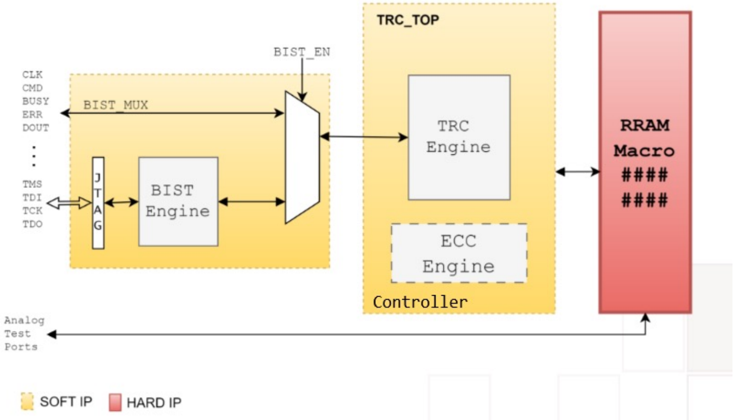
Ecosystem
eFlash
5V Device
SRAM
MS / RF
Logic Devices

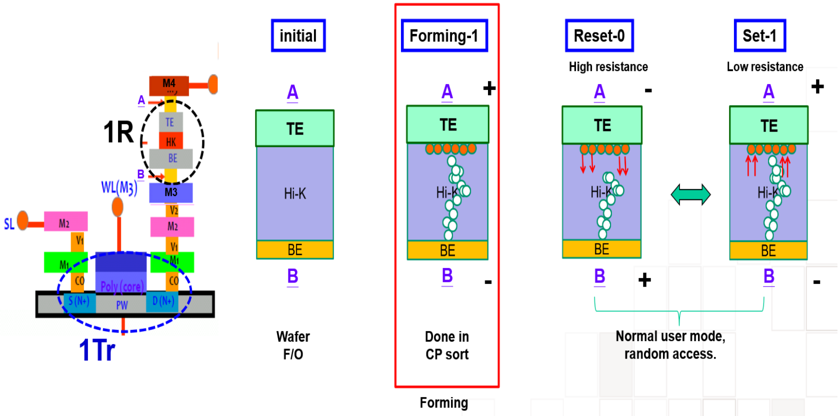
Kostas.Kloukinas@cern.ch

22nm RRAM

- Advanced RRAM Integration
  - Resistive RAM is embedded into TSMC 22nm ULL manufacturing process, combining advanced non-volatile memory with cutting-edge planar semiconductor technology

- Low Power and Fast Access
  - RRAM technology is fully logic compatible with low process complexity. It is an excellent replacement for eFlash with comparable write speed, endurance and retention





FinFET Design Challenges

⚙ DEVICE & PROCESS

Quantized Fin Geometry Width is discrete (N fins), not continuous. Complicates analog sizing, current matching, and W-stepping that planar CMOS handles freely.

Layout-Dependent Effects (LDE) Vth shifts 10s of mV from neighbour fins, STI stress, and MEOL contacts. Geometry-aware SPICE models mandatory.

Self-Heating Narrow Si body surrounded by low-κ isolation traps heat. Thermal-aware electrothermal co-simulation required, especially for RF/high-current blocks.

Gate Length Variability RDF suppressed but LER/LWR dominates at 7/5nm. Pelgrom law holds with different physical roots — process control is critical.

~ ANALOG & MIXED-SIGNAL

Constrained Analog Sizing — FinFET electrostatics preserve output resistance (ro) well at moderate L.
The real analog burden is discrete W quantization: fin count replaces continuous W/L trade-offs, forcing coarser gm and bias point control than planar designers are accustomed to.

Reduced Supply Headroom VDD = 0.7–0.8 V at 7nm squeezes dynamic range (DR $\propto$ VDD^2/noise) and limits cascode stacking depth.

Discrete-Width Mismatch Classical Pelgrom matching ($\sigma\_\Delta Vth \propto 1/\sqrt{WL}$) restricted to discrete W options. Common-centroid layouts need identical fin-neighbour environments.

ESD Protection Thick-oxide I/O devices harder to implement; fin geometry elevates snapback current density and limits ballasting strategies.

Kostas.Kloukinas@cern.ch

FinFET Design Challenges

▦ DIGITAL DESIGN

Std-Cell Library Complexity Fin quantization + multi-patterning-aware routing rules produce tight track heights (e.g. 6T at 5nm). PnR must respect coloring constraints throughout.

Multi-Patterning (SADP/SAQP) Below 16nm, no single exposure covers critical layers. LELE/SADP introduce overlay errors and forbidden pitches. EUV from ~7nm alleviates this but is not universal.

Clock & Power Grid Tight pitch raises metal resistance; PDN IR drop and EM limits tighten. Power/clock gating must be co-optimised with physical implementation from day one.

Timing Closure & OCV LDE/stress/proximity grow OCV budgets. Flat SPEF-based timing insufficient; AOCV/POCV (path-based) models required for reliable signoff.

☢ RADIATION HARDNESS

Much better TID tolerance than planar CMOS

Reasons:

- No thick field oxide near channel
- Reduced charge trapping impact
- Better electrostatic control.

Single Event Upset (SEU) Mixed behavior

Smaller sensitive volume → less charge collection

Lower node capacitance → less charge needed to flip a bit

Net result: SEU cross-section ≈ similar to 28 nm

Single Event Transient (SET) Still a major issue

Fast transients due to high drive current and fast edges

Can propagate further in deep pipelines

Single Event Latchup (SEL) Major improvement

FinFETs are typically: Latchup-immune

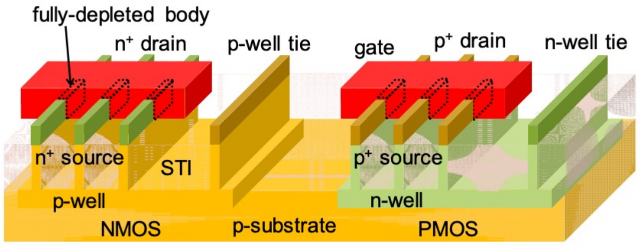
No continuous bulk substrate path

No classical parasitic SCR structure

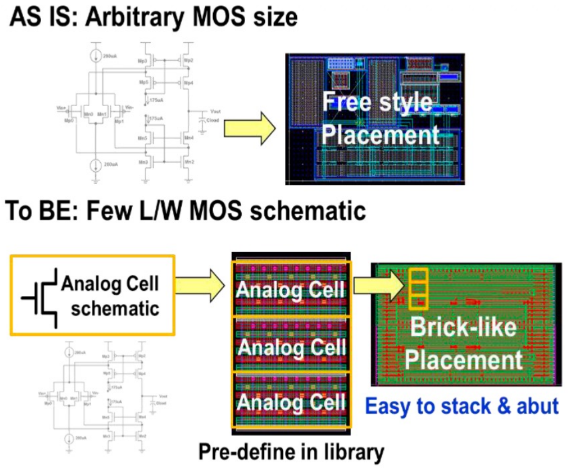
Kostas.Kloukinas@cern.ch

Analog Design in finFETs

- PROBLEM
  - Arbitrary transistor sizing → complex, non-uniform hand-drawn layouts
  - Increased layout dependent effects (LDE): fin profile variation, stress, well proximity — all harder to model accurately
  - Growing simulation-to-silicon gap at advanced nodes
- ANALOG BLOCK APPROACH
  - Few selected W/L options with track-based, pre-drawn cell heights
  - Brick-like abutment enables structured, uniform array placement
  - Guard ring cells share the same layout structure as active cells, enabling seamless abutment between arrays
  - This co-design eliminates the area overhead of conventional guard ring utilities.



Source: 978-4-86348-815-1 ©2025 JSAP 2025 Symposium on VLSI Technology and Circuits Digest of Technical Papers



Guard Ring _C	Guard Ring _H	Guard Ring _H	Guard Ring _H	Guard Ring _H	Guard Ring _H	Guard Ring _C
Guard Ring _V	Active Analog Cell_A	Active Analog Cell_A	Guard Ring _I	Active Analog Cell_A	Active Analog Cell_A	Guard Ring _V
Guard Ring _V	Active Analog Cell_A	Active Analog Cell_A	Guard Ring _I	Active Analog Cell_A	Active Analog Cell_A	Guard Ring _V
Guard Ring _V	Guard Ring _I	Guard Ring _I	Guard Ring _I	Guard Ring _I	Guard Ring _I	Guard Ring _V
Guard Ring _V	Active Analog Cell_A	Active Analog Cell_A	Guard Ring _I	Active Analog Cell_A	Active Analog Cell_A	Guard Ring _V
Guard Ring _V	Active Analog Cell_A	Active Analog Cell_A	Guard Ring _I	Active Analog Cell_A	Active Analog Cell_A	Guard Ring _V
Guard Ring _C	Guard Ring _H	Guard Ring _H	Guard Ring _H	Guard Ring _H	Guard Ring _H	Guard Ring _C

High Energy Physics Motivation

Why Advanced Nodes Help

Hit Rate & Pixel Pitch

- Smaller technology nodes enable much higher logic and memory density per mm²
- Allows integration of Larger pixel matrices without increasing chip area and More complex per-pixel logic (zero suppression, buffering, clustering)
- Enables finer pixel pitch by reducing routing congestion and area per channel
- Critical for high-luminosity environments where occupancy is extreme

Data Bandwidth

- Hit rates in the tens of MHz per pixel require on-pixel zero-suppression, time-stamping, and buffering — tasks that demand dense digital logic only achievable in advanced nodes
- High-speed serial I/O drivers (multi-Gbps per link) benefit from the higher ft/fmax of short-channel transistors, enabling the aggregate bandwidth (Tbps scale for large detectors) needed for trigger-less or data-driven readout architectures
- Advanced nodes offer more metal layers and tighter pitch interconnect, easing the routing of high-density data buses within the pixel matrix

Timing Resolution

- Sub-100 ps (and increasingly sub-10 ps for 4D tracking) timing requires fast slewing front-end circuits and low-jitter discriminators, both of which benefit from the higher transconductance-to-current ratio (gm/ID) and reduced Cgate of advanced nodes
- Time-to-Digital Converters (TDCs) embedded per pixel or per column require fine delay cells; the shorter gate delays in FinFETs/FD-SOI allow higher TDC resolution without excessive area or power
- Reduced threshold voltage mismatch lowers timing walk and improves hit-time uniformity across large pixel arrays

Power Budget

- Power-per-channel reduction despite increasing functionality
- Advanced nodes operate at lower supply voltages reducing dynamic power quadratically ($P \propto CV^2f$)
- The improved Ion/Ioff ratio in FinFETs reduces leakage in large arrays of idle pixels, which dominate static power at high pixel counts

Radiation Tolerance

- Total ionizing dose (TID) >1 Grad generally robust but node-to-node variations to be evaluated.
- SEU, SET and SEL robustness

Kostas.Kloukinas@cern.ch

High Energy Physics Implications

HEP is NOT the Consumer Market — Key Differences

Long Design Cycle 5-10 Years

Specification → tape-out: 3–5 yrs
System integration → physics: 3–5 yrs
Node must remain accessible & supported

Radiation Environment

TID: 1 Mrad–1 Grad (inner trackers)
SEU, SET, SEL robust
RADHARD-by-design (RHBD) techniques

System-on-Chip Complexity

Analog front-end + ADC + DSP + I/O
Multi-function mixed-signal SoC
Full custom + standard cell co-design

Volume & Economics

Prototype: 100 ~1k die / MPW
Production: 10k–100k die
Product Cost is driven by Foundry NRE: $1.5M–$10M at (28nm to 7nm)

EDA Tool Complexity

PDK complexity grows with node
Mixed-signal design closure challenges
IP reuse: limited — often full-custom
Formal verification mandatory

Foundry Access

Foundry long term engagement
Regular affordable MPWs is essential

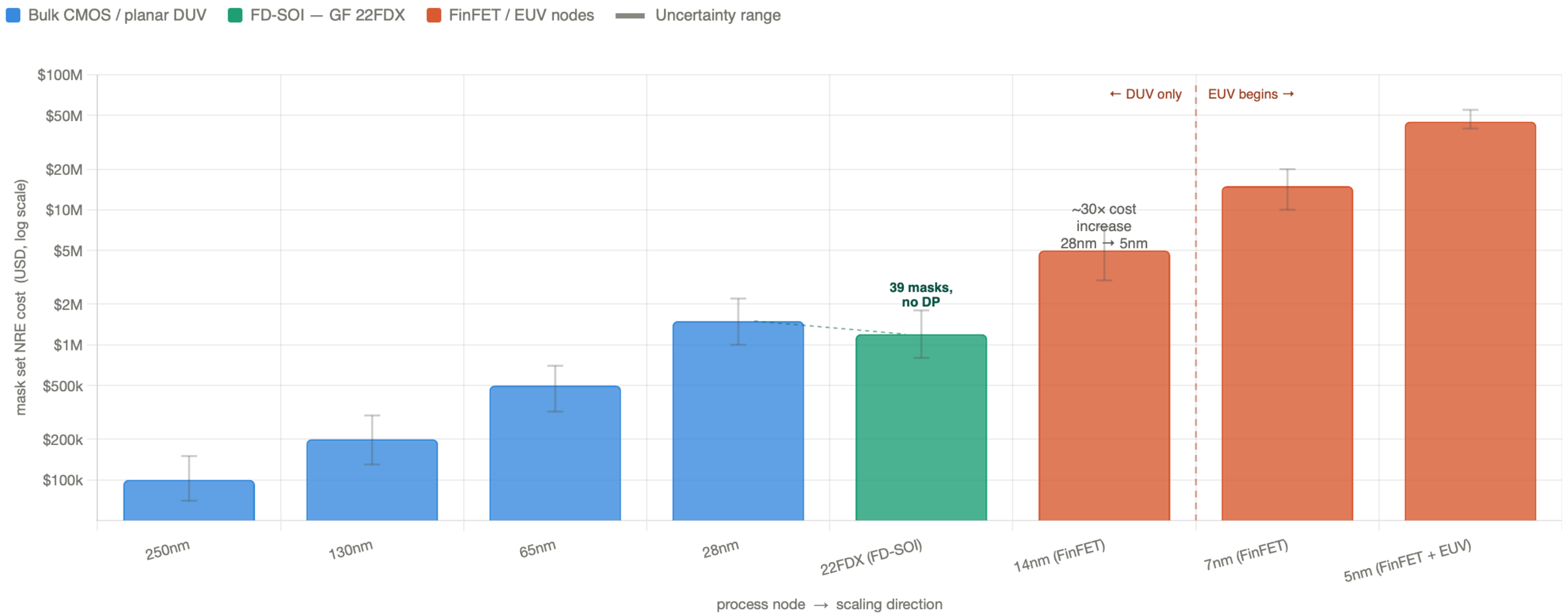
Kostas.Kloukinas@cern.ch

Mask set NRE cost

log scale

Mask set NRE cost by representative process node

Selected nodes · log scale · mask set fabrication cost only — excludes design, EDA, IP, and verification

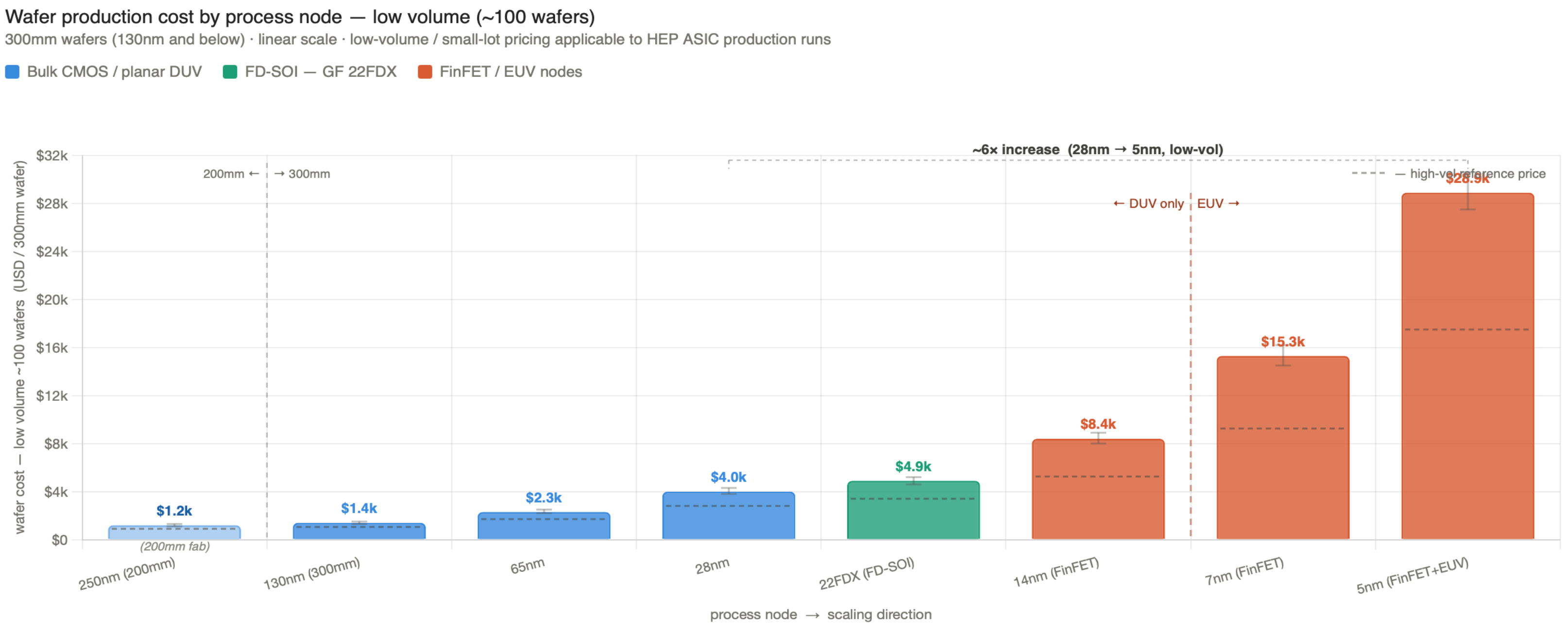


Data sources & anchors:

- 28nm: IBS/SemiEngineering consensus ~$1–2M mask set NRE (widely cited)
- 22FDX: GF confirmed 39 mask layers — 10% fewer than 28nm bulk, no double patterning → mask set cost slightly *below* 28nm (~$0.8–1.8M)
- 14nm: industry estimates ~$3–7M (multi-patterning, no EUV); Quora/EDA board reports $2–3M at 16/14nm for smaller designs
- 7nm: IBS ~$15M; WikiChip / SemiAnalysis consensus $10–20M; first EUV layers
- 5nm: TSMC tapeout mask set $40–50M (confirmed industry reports, 2022–2023); heavy EUV use (up to 14 EUV layers)
- 250nm–65nm: ITRS-era data, AnySilicon mask cost reference, industry consensus

Kostas.Kloukinas@cern.ch

Wafer Production cost by process node



Wafer production cost by process node — low volume (~100 wafers)

300mm wafers (130nm and below) · linear scale · low-volume / small-lot pricing applicable to HEP ASIC production runs

Low-volume pricing basis (~100 wafers): mature nodes +30–40% above HVM contract price; leading FinFET/EUV nodes +50–80% (EUV tool reservation, strict minimum lot sizes).

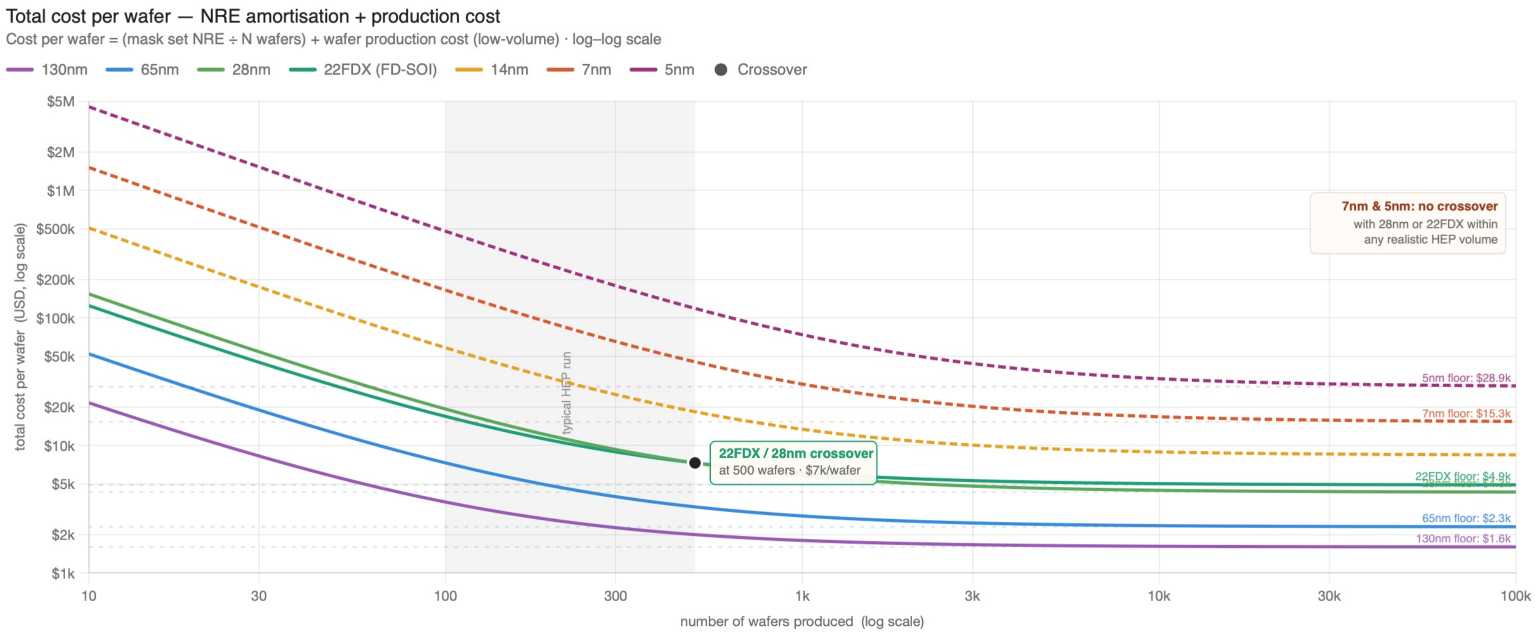
High-volume pricing & threshold — selected nodes:

NODE	HVM PRICE / WAFER	LOW-VOL PREMIUM	HVM THRESHOLD	HEP RUN STATUS
65nm	~$1,400–2,000	+30–40%	≥ 2,000 wafers / year	Always low-vol
28nm	~$2,500–3,000	+35–50%	≥ 5,000 wafers / year	Always low-vol
7nm	~$8,500–10,000	+50–80%	≥ 10,000–25,000 wafers / year	Always low-vol

HEP ASIC production runs (typically 100–500 wafers) never qualify for high-volume rates at any node. Sources: SiliconAnalysts (April 2026), TrendForce, AnySilicon, TSMC public statements.

Kostas.Kloukinas@cern.ch

NRE amortization



Total cost per wafer — NRE amortisation + production cost

Cost per wafer = (mask set NRE ÷ N wafers) + wafer production cost (low-volume) · log–log scale

Formula: Total cost / wafer = NRE ÷ N + C_{wafer}. Asymptote at large N = wafer production cost floor.

NRE (mid): 130nm $200k · 65nm $500k · 28nm $1.5M · 22FDX $1.2M · 14nm $5M · 7nm $15M · 5nm $45M.

Wafer cost (low-vol ~100 wafers): 130nm $1.6k · 65nm $2.3k · 28nm $4.3k · 22FDX $4.9k · 14nm $8.4k · 7nm $15.3k · 5nm $28.9k.

Crossovers within plot range: 65nm crosses below 130nm at ~230 wafers · 28nm crosses below 65nm — never (28nm wafer floor > 65nm floor) · 22FDX crosses below 28nm at ~500 wafers (lower NRE, higher wafer floor).

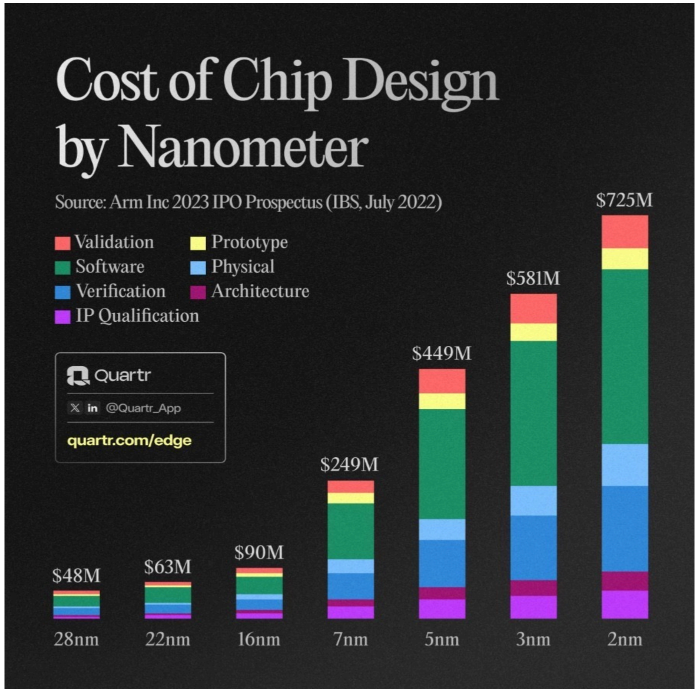
7nm and 5nm never cross 28nm or 22FDX at any realistic volume: wafer production floors are 3–7× higher and no NRE amortisation can bridge this gap. Their use in HEP must be justified by physics performance, not economics.

Shaded band: typical HEP ASIC production range (100–500 wafers). Sources: SiliconAnalysts (April 2026), AnySilicon, TrendForce, GF/industry NRE consensus.

Kostas.Kloukinas@cern.ch

Product development cost in Industry

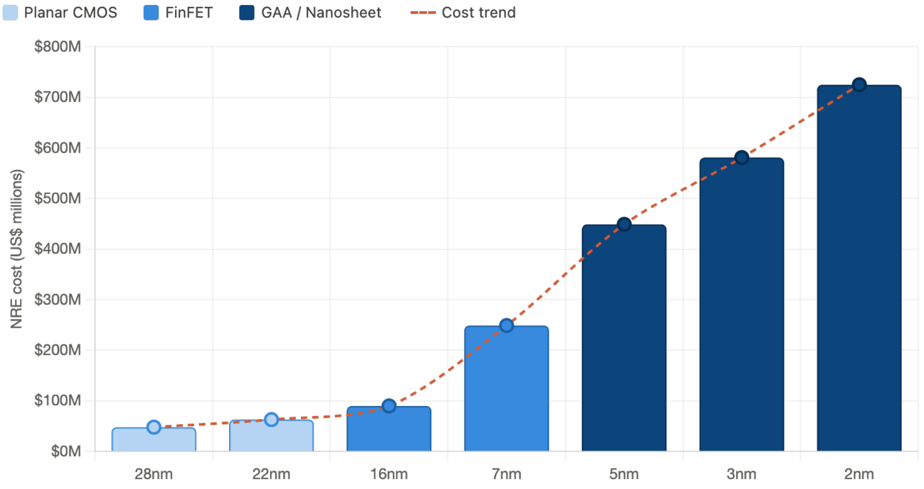
Moore's Law isn't just about transistor density—it's about economics.



Cost of Chip Design by Nanometer

Source: Arm Inc 2023 IPO Prospectus (IBS, July 2022)

https://testflowinc.com/blog/2nm-chip-development-cost-725-million-analysis



28nm → 2nm
15×
total cost increase

Steepest jump
+177%
16nm → 7nm (FinFET + EUV)

Validation at 2nm
$159M
22% of total — up 300% since 7nm

Min. units to break even
50M+
at 2nm node

Source: TestFlow Inc., "How Much Does a 2nm Chip Really Cost? $725M Development Breakdown", Jan. 2026 — testflowinc.com. Figures represent full NRE for a complex SoC (architecture, IP, physical design, validation, EDA tools). Simpler designs at the same node cost considerably less.

Kostas.Kloukinas@cern.ch

The 16nm → 7nm Cost Cliff

The 16nm → 7nm step is not a single increase — it is the simultaneous collision of multiple independent cost drivers.

Mask set cost

16nm: ~$12M **7nm: ~$42M**

EUV masks need reflective multilayer optics
Features at 7nm have far more OPC polygons — up to 100× more per layer. Inverse Lithography Technology is computational intensive. E-beam inspection of masks (slower than optical)
HEP: one failed prototype iteration can consume an entire collaboration's annual hardware budget.

Wafer cost

16nm: ~$5K/wafer **7nm: ~$10K/wafer**

Lithography complexity. More mask layers. A 7nm process can require 80–100+ mask layers vs. ~50–60 for 16nm.. Lower yield, especially at project ramp. Defect density scales unfavorably with feature size.

DRC complexity

16nm: ~5,000 rules **7nm: 30,000+ rules**

Full DRC on a large ASICs takes days.
HEP: small design teams cannot absorb the specialist effort to debug tens of thousands of interacting coloring and patterning rules.

Design methodology

16nm: schematic-driven **7nm: gridded / RDR**

Free-form custom layout replaced by mandatory manufacturing grids and restricted design rules (RDR). Parasitic-aware layout in the loop is mandatory — open-loop design then extract gives wrong answers at 7nm pitches. Mixed-signal co-simulation requires FastSPICE at scale.
HEP: no existing IP at 16nm or 7nm. The methodology itself must be learnt from scratch.

Metal stack changes

16nm: relaxed / familiar **7nm: gridded / coloured**

A new local interconnect layer (M0/LIG) between contacts and M1 has no equivalent at 16nm. Via rules forbid many combinations; mandatory via arrays replace single vias. M1–M3 are fully unidirectional and coloring-constrained at 28–36 nm pitch. Need powerfull PnR..
HEP: coloring-constrained routing and local interconnect are skills absent in our community.

Signoff, closure & EDA

16nm: standard flow **7nm: specialist infrastr.**

Full 3D parasitic extraction on every layer. IR-drop, EM and thermal co-simulated. PVT corners double. Multi-patterning coloring inside Calibre — a new discipline entirely. LVS needs updated device decks for FinFET fin counts and well proximity. Tool licences at 7nm cost significantly more; much higher compute runtimes.
HEP: no trained staff nor compute infrastructure for this flow exists in our groups.

! **Bottom line for HEP**

Companies that absorbed this cliff — Apple, Qualcomm, NVIDIA — sell hundreds of millions of units, amortising $249M NRE across a product lifetime. A HEP collaboration producing 10–100k ASICs cannot. 28nm remains the highest-volume node in the industry a decade after introduction — the pragmatic ceiling for our community unless a clear physics mandate and industrial partnership justify the step to 7 nm.

Kostas.Kloukinas@cern.ch

Node Comparison for HEP ASICs

Key candidate nodes — pros, cons, and HEP relevance

Node	Architecture	Strengths for HEP	Challenges	HEP use
28nm	Planar + HKMG shrink node	Mature PDK, MPW readily available, affordable NRE, excellent radiation behavior. HEP Design Kit and design methodology are validated.	Power and density advantages vs. 65nm, limited analog IP ecosystem.	PicoPIX
22nm	Planar + HKMG shrink node	Ultra Low Leakage devices. 10% area reduction, >10% speed gain, or 20% power reduction compared to 28nm. Heavily promoted node by the foundry.	Build HEP Design Kit and design methodology. Build analog IP ecosystem. Validate TID performance.	Backup to 28nm
22nm FD-SOI	FD-SOI (buried oxide)	Body biasing → dynamic VT tuning, excellent low-power, good analog	(GF, STmicro), cost ↑ vs. bulk, TID<10MRad, 10x more robust for SEE	Recommended: for low power, high frequency, application in low TID radiation field, Monolithic sensors (?)
16/14nm	FinFET (3D gate)	High logic density, >200 GHz fT, ML-edge possible.	High NRE ($2–5M), EDA complexity, design methodology gap	Recommended: adopt only with clear physics mandate. Regular MPWs & mini@sics
7nm	FinFET (dense)	Maximum density & performance, AI on-chip	Very high NRE (>$5M), EDA complexity, severe design methodology gap	Recommended: adopt only with clear physics mandate and budgetary considerations Regular MPWs

Kostas.Kloukinas@cern.ch

Power Performance Area per selected nodes

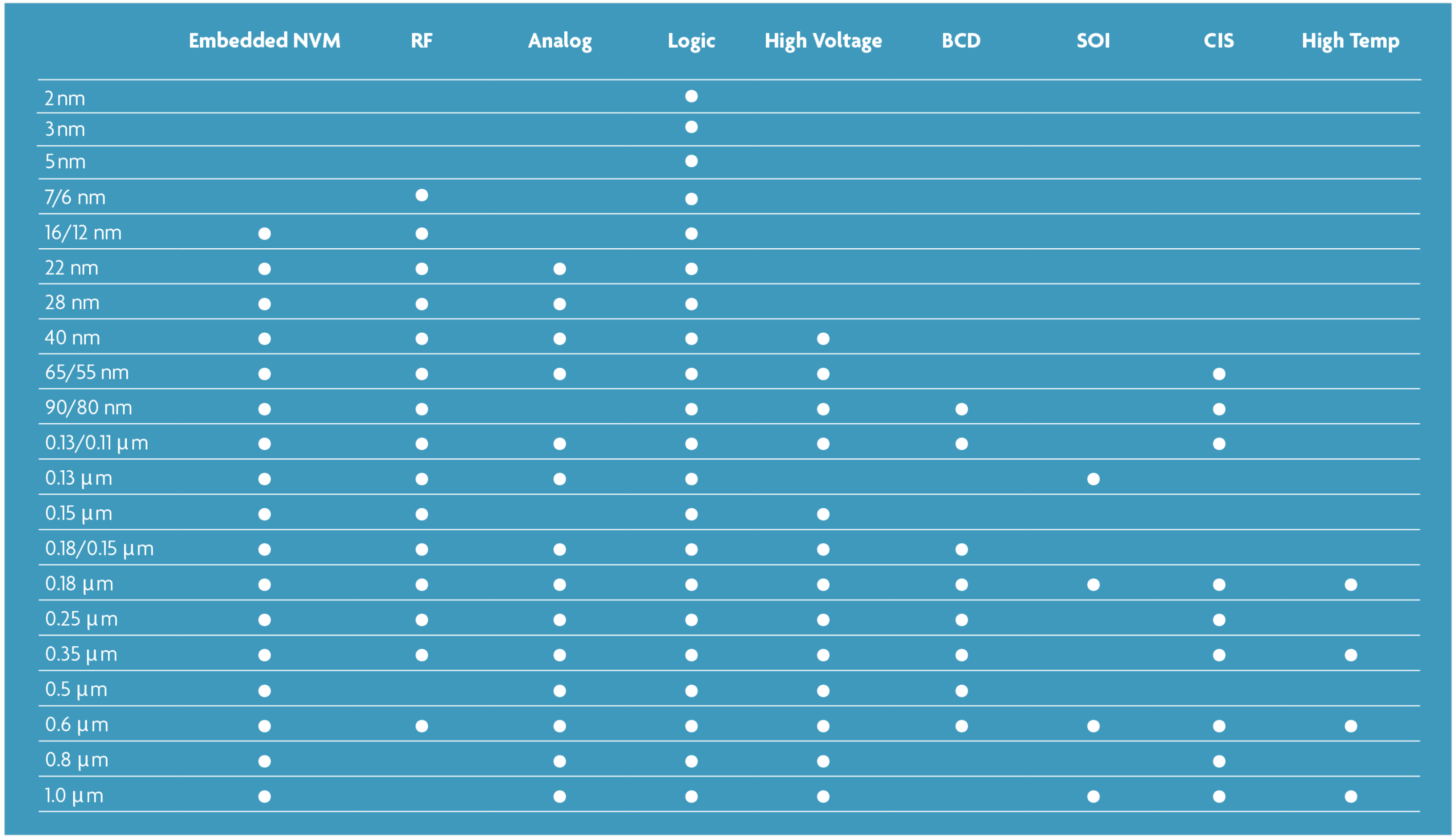
Parameter	130nm	65nm	28nm	22nm ULL	16nm	7nm
Process type	Bulk planar	Bulk planar	Bulk HKMG	*Bulk HKMG*	FinFET	FinFET
Vdd nominal (V)	1.2 – 1.5	1.0 – 1.2	0.9 – 1.0	0.8 – 0.9	0.8 – 0.9	0.7 – 0.8
Vdd low-power (V)	1.0	0.9	0.8	0.7 – 0.75	0.72	0.65
Fmax typical (GHz)	0.3 – 0.5	0.8 – 1.2	1.5 – 2.5	*~1.0 – 1.5*	2.5 – 4.0	4.0 – 6.0
Dynamic power (norm.)	16×	8×	2×	*~1.5×*	1.3×	1×
Leakage power (norm.)	1×	10× – 20x	50× – 100×	*~5× – 15×*	~20× – 40×	~30× – 60×
Logic density (MTx/mm^2)	~0.5	~2	~10	*~15*	~30	~90
SRAM cell size (μm^2)	~1.0	~0.42	~0.127	*~0.10*	~0.070	~0.027
Metal layers	6 – 8	7 – 9	8 – 10	*8 – 10*	10 – 13	13 – 17
M1 pitch (nm)	320	180	90	*90*	64	40

Kostas.Kloukinas@cern.ch

finFET offerings within reach

- Via Europractice/IMEC
  - Regular MPW runs for **16 nm** and **7 nm**
  - Regular mini@sics runs for **16 nm**
  - **Special University Program**
    - 16nm: 60kEur/4mm^2
    - 7nm: 50kEur/2mm^2
    - https://europractice-ic.com/tsmc-university-finfet-program/#PricesTSMCfinfet

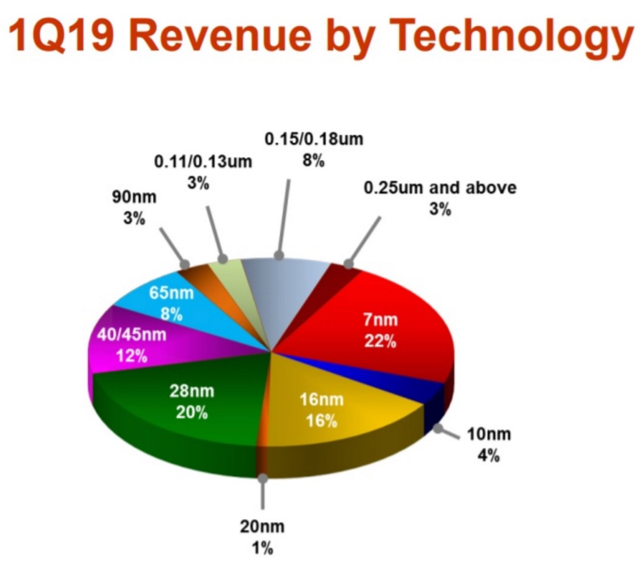
Technologies and IP availability

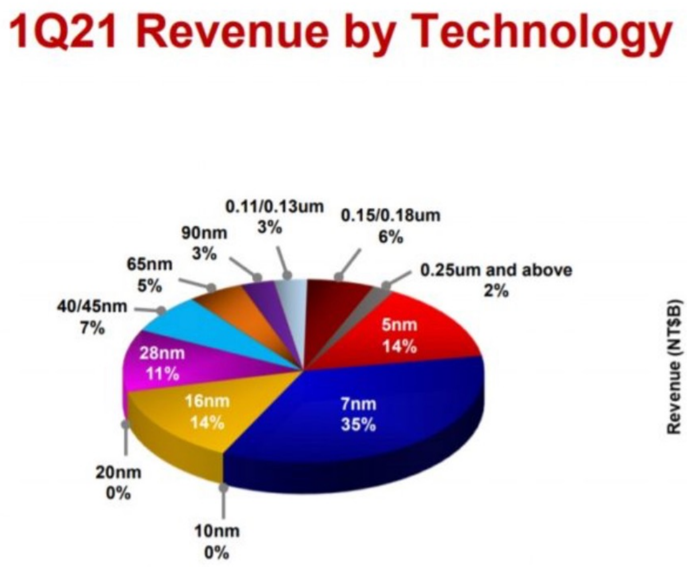


	Embedded NVM	RF	Analog	Logic	High Voltage	BCD	SOI	CIS	High Temp
2 nm				●					
3 nm				●					
5 nm				●					
7/6 nm		●		●					
16/12 nm	●	●		●					
22 nm	●	●	●	●					
28 nm	●	●	●	●					
40 nm	●	●	●	●	●				
65/55 nm	●	●	●	●	●			●	
90/80 nm	●	●		●	●	●		●	
0.13/0.11 µm	●	●	●	●	●	●		●	
0.13 µm	●	●	●	●			●		
0.15 µm	●	●		●	●				
0.18/0.15 µm	●	●	●	●	●	●			
0.18 µm	●	●	●	●	●	●	●	●	●
0.25 µm	●	●	●	●	●	●		●	
0.35 µm	●	●	●	●	●	●		●	●
0.5 µm	●		●	●	●	●			
0.6 µm	●	●	●	●	●	●	●	●	●
0.8 µm	●		●	●	●			●	
1.0 µm	●		●	●	●		●	●	●

Kostas.Kloukinas@cern.ch

Foundry economics





- **Bottom line:** The foundry is pushing their 500+ customers hard into the FinFET era and that will again change the foundry landscape.

- **The trillion dollar question is:** What will happen to the mature (non-FinFET) nodes in the not too distant future? And more importantly, what will do the foundries that did not make the jump to FinFETs?

Kostas.Kloukinas@cern.ch

Conclusions & Take-Aways

- The technology landscape is driven by AI/ML, IoT, 5G and automotive
  HEP benefits as a passenger on this roadmap and must be selective about which nodes to ride.
- Advanced CMOS below 28nm is justified for HEP by hit-rate, pixel pitch, bandwidth and timing demands — particularly for HL-LHC and beyond.
- HEP specifics (long design cycles, radiation, economics, foundry access) strongly constrain the candidate set — 7nm is currently out of scope for most collaborations.
- Sweet spots: 28nm bulk (primary), 22nm (backup), 14nm FinFET (high-density/high performance).
  Engage foundries; build expertise.

Kostas.Kloukinas@cern.ch

Kostas.Kloukinas@cern.ch

Power Performance Area per node (28nm baseline)

🟦 Planar CMOS 🟩 FinFET 🟧 GAA Nanosheet ☐ ● Foundry-announced ☐ ○ Estimated / ITRS

Node	Architecture	Lithography	Density (×28nm)	Perf (×28nm iso-pwr)	Power (÷28nm iso-freq)	Perf/W/mm² (×28nm)	Key innovation
130nm **2002**	**Planar**	248nm KrF	**● 0.06×**	**● 0.48×**	**● 0.31×**	**● 0.01×**	Cu interconnect, low-k dielectric — pre-baseline reference
90nm **2004**	**Planar**	193nm ArF	**○ 0.13×**	**○ 0.57×**	**○ 0.41×**	**○ 0.03×**	Strained Si channel; first 193nm ArF generation
65nm **2006**	**Planar**	193nm ArF	**● 0.25×**	**● 0.67×**	**● 0.53×**	**● 0.09×**	2× gate density vs 90nm; 30–50% speed gain (TSMC datasheet)
45/40nm **2008**	**Planar**	193nm immersion	**○ 0.50×**	**○ 0.81×**	**○ 0.69×**	**○ 0.28×**	Immersion litho; HKMG introduced (Intel 45nm); leakage reduction
28nm ☆ **2011**	**Planar + HKMG**	193nm immersion	● 1.0×	● 1.0×	● 1.0×	● 1×	Full HKMG universally adopted; VDD scaling resumes — BASELINE
16/14nm **2014**	**FinFET**	193nm immersion+SADP	**● 2.0×**	**● 1.5×**	**● 2.5×**	**● 7.6×**	FinFET: ~50% speed or ~60% power vs 28nm (TSMC N16)
10nm **2017**	**FinFET**	193nm SAQP	**● 4.0×**	**● 1.8×**	**● 3.8×**	**● 26×**	~15% speed or ~35% power vs 16nm; 2× density (TSMC N10)
7nm **2018**	**FinFET**	EUV (partial)	**● 6.3×**	**● 2.1×**	**● 6.3×**	**● 81×**	~20% speed or ~40% power vs 10nm; EUV for critical layers (TSMC N7)
5nm **2020**	**FinFET**	EUV	**● 11×**	**● 2.4×**	**● 9.1×**	**● 246×**	~15% speed or ~30% power vs 7nm; 1.8× density (TSMC N5)
3nm **2022**	**FinFET / GAA***	EUV	**● 19×**	**● 2.8×**	**● 13×**	**● 675×**	~12% speed or ~28% power vs 5nm; 1.7× density (TSMC N3E)
2nm **2025**	**GAA Nanosheet**	EUV + High-NA	**● 22×**	**● 3.2×**	**● 18×**	**● 1,267×**	~15% speed or ~30% power vs 3nm; GAA nanosheet (TSMC N2, IEDM 2024)

*● Foundry-announced ○ Estimated / ITRS roadmap interpolation *GAA at Samsung 3nm; TSMC N3 remains FinFET*

Sources: TSMC process announcements (N16–N2); ITRS (130nm–28nm); StatNano / NPD (65nm); Wikipedia; SemiEngineering; IC Knowledge / Handel Jones

Kostas.Kloukinas@cern.ch