Physical Dilemma of Large-Area Advanced-Node Chips: Irreversible Narrowing of Interconnect Channels

---

Ao Li

Independent Researcher

Abstract

The two pathways to increasing chip computing power—process scaling and chip area expansion—have both reached their physical and economic limits. This paper demonstrates that advanced-node devices face insurmountable bottlenecks (quantum tunneling, voltage scaling failure, statistical fluctuations) and that 2.5D/3D packaging and chiplet architectures merely shift rather than resolve the contradictions of interconnect narrowing. It argues that the industry's indiscriminate adoption of advanced nodes is physically unsustainable, especially as sub-1 nm nodes approach a point where controllability overtakes usability. The only physically self-consistent path forward is a system-integration paradigm centered on mature-node chiplets, where wide interconnects,

low thermal density, and traditional packaging together respect fundamental physical laws.

Keywords

Advanced Node; Interconnect Channel; Narrowing Effect; Resistance; Thermal Dissipation; Chiplet; Mature Process; Physical Limit; RC Delay; Controllability

---

0 The Only Two Paths to Performance Improvement: A Geometric Fact

Before any technical discussion, it is necessary to establish a premise dictated by geometry and basic physics: under current physical laws, there are only two fundamental paths to increasing chip computing power.

Path 1: Process scaling—increasing transistor density per unit area.

Accommodating more transistors within the same chip area to handle more simultaneous computation. This is the core logic of Moore's Law: continuously shrinking transistor feature size to integrate more computing

units into the same silicon piece [2].

Path 2: Chip area expansion.

Keeping transistor density constant and simply manufacturing a larger chip, trading total area for total transistor count.

There is no third path. The paths referred to here are specifically the means of transistor integration at the physical hardware level. Architectural innovations, instruction set optimizations, processing-in-memory, algorithmic co-design, and similar approaches improve the computational efficiency per transistor; they do not alter the upper bound on the total number of transistors a chip can accommodate, and therefore do not break the two-dimensional geometric constraint [2].

This conclusion is rigorously derived from geometry: a chip is a two-dimensional physical entity, and its total transistor count equals "area" multiplied by " transistor density per unit area." Any so-called "architectural innovation," " advanced packaging," or " heterogeneous integration" is ultimately a combination or compromise of these two paths and cannot escape this planar geometric constraint.

The following analysis will demonstrate: under advanced nodes, Path 1 encounters resistance surge, thermal management crises, and signal crosstalk caused by interconnect channel narrowing; the monolithic large-chip version of Path 2 encounters exponential yield degradation and edge-perimeter bottlenecks [8] [6]. Any attempt to forcibly splice the two paths through 2.5D/3D packaging can shift some contradictions, but it introduces more complex systemic challenges at new physical interfaces [3]. The key question is whether the industry has the courage to admit that advanced nodes are not the default optimum for all scenarios.

---

1 Device-Level Physical Bottlenecks of Advanced Nodes

Before discussing interconnects, it is necessary to acknowledge the physical boundaries that advanced-node transistors themselves have already reached.

1.1 Quantum Tunneling

When transistor channel length approaches the atomic scale, electrons can tunnel through potential barriers with significant probability even in the off-state [4]. The switch can no longer be fully turned off, and leakage current increases exponentially with size reduction. This is not a material defect; it is a fundamental prediction of quantum mechanics.

1.2 Voltage Scaling Limit

Dennard scaling once allowed voltage to be reduced proportionally with dimensions, maintaining constant power density as transistor density doubled [5]. However, the threshold voltage cannot be lowered infinitely due to thermodynamic limits—below a critical value, off-state leakage drowns the switching signal. When voltage stops scaling while density continues to double, power density doubles. This is a red line drawn jointly by Joule's law and thermodynamics.

1.3 Uncontrolled Statistical Fluctuations

When the number of dopant atoms in the channel drops to the order of a few dozen, random fluctuations cause each transistor's threshold voltage to differ. The slowest device on the chip drags down the overall frequency, while the fastest accelerates aging due to over-driving. Yield and reliability, at the atomic scale, become victims of statistics.

Three independent physical laws—quantum mechanics, thermodynamics, and statistical physics—each indicate that process scaling as a single path has reached its physical limits of applicability.

---

2 The Yield Calamity of Monolithic Large Chips

Directly fabricating large-area advanced-node chips leads to an exponential collapse of manufacturing yield. In semiconductor manufacturing, the defect density per unit area is approximately constant, and chip yield decreases exponentially with increasing area [8]. When the chip area exceeds a certain threshold, the per-die yield drops to an unacceptable level, and production cost skyrockets, potentially exceeding the chip's selling price and destroying economic viability. This is an unavoidable real-world constraint for large chips at advanced nodes [6].

---

3 The Inherent Tension Between Chiplets and Advanced Nodes

Since monolithic large chips enjoy neither yield support nor economic feasibility, the industry has turned to the chiplet approach: splitting a large-chip function into multiple small dies and reassembling them into a complete system through advanced packaging [7]. This route circumvents the single-die yield trap, but it elevates the interconnect contradictions that were previously hidden inside the chip to the packaging interface.

Consider a chip of fixed area (e.g., 10 mm × 10 mm). At an advanced node (e.g., 5 nm), its transistor density is over 20 times that of a mature

node (e.g., 28 nm). The total number of signals that need to be processed inside the chip, and the number of I/O channels that need to enter or exit the chip, both scale roughly linearly with the transistor count. Yet the physical perimeter of the chip is fixed (40 mm). All interconnect channels that must pass through the chip edge are forced to crowd into this unchanged perimeter.

---

4 Mandatory Alignment Accuracy Imposed by Channel Narrowing

Advanced packaging and chiplet assembly do not eliminate the contradiction between the explosion in channel count and the limited interface area; they merely shift it to the packaging interface [3]. To accommodate a massive number of channels within that interface, interconnect pitch and line width must shrink further, directly forcing packaging processes to attain nanometer-scale alignment accuracy. Any misalignment immediately results in a sharp increase in contact resistance, signal attenuation, and uncontrolled crosstalk. Thus, the high-precision requirement for advanced packaging is not an optional technical choice; it is a mandatory physical constraint imposed by channel narrowing.

---

5 Channel Narrowing and Drastic Resistance Increase

Let the number of interconnect channels that must be accommodated within a fixed perimeter be N. Then the width of each channel $w \propto 1/N$. At advanced nodes, N increases dramatically, forcing w to shrink to the nanometer scale. According to the resistance law $R = \rho L/(w \cdot t)$, the resistance of a channel is inversely proportional to its width. When w is reduced to one-fifth of that at a mature node, the resistance increases by at least a factor of five. Additionally, electron surface scattering increases the effective resistivity ρ, so the actual resistance increase can exceed an order of magnitude [9] [11].

---

6 Signal Attenuation and the Thermal Challenge of Repeaters

High-resistance channels experience significantly increased RC delay and severe signal amplitude degradation. The traditional compensation method is inserting repeaters along the channel [10]. But this raises multiple issues: repeaters themselves consume dynamic power and generate Joule heat; advanced nodes already have extremely high power density; the massive number of channels necessitates a huge number of repeaters; and the dense distribution of repeaters creates localized

hotspots, where rising temperature increases leakage current, which in turn increases repeater power consumption, forming a positive feedback loop [9] [10] [1]. This thermal management challenge makes a scheme relying solely on repeaters face serious engineering feasibility problems.

---

7 Uncontrollable Signal Crosstalk

Narrow channels not only have high resistance; the spacing between adjacent channels also shrinks correspondingly. Coupling capacitance and mutual inductance between lines increase sharply, causing a significant rise in crosstalk noise [9]. In high-density interconnect regions, simultaneously switching signals can induce noise exceeding the logic threshold, causing data errors. Suppressing crosstalk requires increasing line spacing or inserting shield lines, but this further compresses the already crowded channel width, creating an engineering design cycle with no easy resolution.

---

8 2.5D Packaging: The Shifting and Transformation of Contradictions

2.5D packaging technologies, such as CoWoS, represent the mainstream industrial response to the above dilemmas. They shift the contradictions from within a single chip to the interposer and packaging interface. This shift is not without value—transforming the boundary of a problem is

itself a core engineering strategy. However, it also introduces new physical challenges.

8.1 Complication of the Heat Dissipation Path

The introduction of chip stacking and silicon interposers forces heat flow to traverse multiple layers of material interfaces with relatively low thermal conductivity. Advanced nodes already have extremely high power density, and the packaging structure adds further thermal resistance. The nominal peak performance that users pay a high cost to obtain is often difficult to sustain under real-world continuous workloads due to hitting the thermal wall [3].

8.2 The Step-Change Increase in Packaging Cost

To stack interconnects within constrained space, the additional process steps and ultra-high-precision equipment required push packaging cost from "negligible" to a magnitude comparable to wafer fabrication [6]. The marginal returns in thermal and signal integrity diminish even as costs escalate.

8.3 Thermomechanical Mismatch and Reliability Challenges

Silicon chips, silicon interposers, and organic substrates have different coefficients of thermal expansion. Under high-power operation, layers expand by different amounts, subjecting microbump connections to accelerated fatigue from cyclic thermal stress. Macro-level temperature control can regulate the average chip temperature, but localized hotspots generated by advanced-node power density produce heat fluxes approaching the thermal conductivity limits of silicon. It is micron-scale temperature gradients—rather than the average chip temperature—that are the root cause of thermomechanical mismatch failure [3]. This physical mechanism means that external cooling solutions alone cannot fundamentally eliminate the reliability risk.

Chip design is a systems discipline. When the single-point pursuit of extreme process nodes incurs disproportionate penalties in thermal, signal integrity, and mechanical reliability at the system level—penalties ultimately borne by end users in the form of throttled performance and inflated costs—a reexamination of the technology roadmap becomes necessary.

---

9 3D Packaging: The Thermal Boundary of Vertical Stacking

Vertically stacking chips (3D packaging) attempts to bypass edge-interconnect bottlenecks but encounters even more severe thermal constraints. Heat from middle layers in a stack must conduct laterally to the edges for extraction, and lateral thermal resistance increases linearly with distance. Under high power density, the temperature of inner layers can easily exceed allowable limits [3]. This makes 3D stacking, for the foreseeable future, more suitable for low-power memory or logic-plus-memory heterogeneous stacking, and unsuitable for vertically stacking high-power logic chips within the same layer, which would otherwise face an intractable thermal dead end.

---

10 Hybrid-Node Chiplet: Physical Negative Synergy and Channel Mismatch

The hybrid-node chiplet approach currently being promoted by the industry retains compute cores on advanced nodes while migrating only I/O and auxiliary functions to mature nodes. This approach is packaged in the industry narrative as an ingenious architecture that "combines the best of both worlds." However, when examined through the lens of physical laws, it not only fails to eliminate the contradictions demonstrated in this

paper, but causes the two process nodes to mutually sabotage each other's inherent advantages, while exposing a fundamental dimensional mismatch in their interconnect channels.

The high power density and localized hotspots generated by advanced-node compute cores transmit thermal stress directly to mature-node chips through the tightly coupled packaging structure, accelerating their aging and destroying the high reliability that mature nodes are supposed to provide. Mature nodes, constrained by their own I/O bandwidth and latency, force advanced-node compute cores to stall frequently and throttle their operating frequency, preventing their nominal performance from being realized in the system. The interconnect interface between the two demands extremely high alignment accuracy, yet the localized thermal expansion mismatch caused by advanced nodes precisely concentrates stress at these microbump connections, subjecting the entire system to accelerated fatigue failure under thermal cycling [3].

A more fundamental problem lies in the inherent dimensional mismatch between the interconnect channels of advanced and mature nodes. Advanced-node interconnect line widths have been compressed to tens of nanometers, with pin pitches of only a few microns, requiring sub-micron alignment accuracy [9]. Mature-node interconnects, in contrast, have line

widths in the micrometer to sub-micrometer range, with pin pitches several to over ten times larger than those of advanced nodes. When these two types of chips must complete signal connections at the same packaging interface, there are only three outcomes: the advanced-node signal lines are forcibly widened, forfeiting their density advantage; the mature-node pins are forcibly shrunk, increasing resistance and reliability risks; or a complex interposer adapter is introduced, further increasing cost and thermal resistance. This is not an engineering difficulty that can be optimized away—it is a fundamental physical scale mismatch between the two process nodes. Channel narrowing occurs not only within advanced-node chips, but also at the interface where advanced and mature nodes must meet.

Therefore, the hybrid-node chiplet is not a combination of complementary strengths, but a case of physical negative synergy: the performance of advanced nodes is diluted by the bandwidth limitations of mature nodes, the robustness of mature nodes is destroyed by the hotspots of advanced nodes, and the dimensional mismatch between the two makes the interconnect interface the most vulnerable failure point in the entire system. This is a patchwork solution that shifts physical contradictions from within a single die to the packaging interface at a higher cost [6]—not a genuine paradigm shift.

---

11 Re-examining the Applicability Boundaries of Advanced Nodes

What the above analysis reveals is not the absolute conclusion that "advanced nodes cannot be used under any scenario." On the contrary, there exists a class of scenarios—where the pursuit of computational density overrides all else, and the system can afford the corresponding thermal and cost penalties—in which advanced nodes are the necessary choice.

What truly warrants caution is the industry's default reflex of treating advanced nodes as the standard option for every scenario. When scenarios with stringent power constraints, relaxed area requirements, or extremely high reliability demands are also blindly pushed onto the process-scaling race track, the system is forced to cover the warnings of physical laws with layer upon layer of engineering patches. Eventually, these warnings are passed on to end users in the form of soaring costs and discounted performance [6].

More worthy of reflection is the fact that as process nodes approach ~1 nm, the triple loss of control stemming from quantum tunneling, statistical fluctuation, and interconnect narrowing superimposes, causing

process "controllability" to overwhelm "usability." Chips can still be manufactured, but their distributions of yield, frequency, and power dissipation become statistically unpredictable [4]. To extract the last bit of nominal performance, manufacturers often pay a disproportionate price in power consumption [10]. In scenarios of sustained operation under a fixed power budget, a more power-efficient and physically robust "next-best" node may actually deliver higher practically usable total compute. This means advanced nodes are not only wasteful in misapplied scenarios; even within their supposed target domains, a divergence between nominal performance and practically deliverable performance has begun to set in. When controllability eclipses usability, a so-called "more advanced" node has fallen into a utility inversion for the majority of practical engineering contexts.

---

12 Mature-Node Chiplets: The Physical Return of System Integration

The above analysis also reveals a fact systematically obscured by the industry narrative: channel narrowing is a predicament unique to advanced nodes. Mature nodes have low transistor density, so the number of I/O channels required per unit area is naturally smaller, and channel widths need not be forced down to the nanometer scale [9].

Here it is necessary to highlight a fundamental physical advantage that mature nodes possess with respect to interconnect delay. The RC delay of an interconnect line is proportional to the product of its per-unit-length resistance and capacitance. At advanced nodes, interconnect lines are compressed to nanometer-scale cross-sections, causing the per-unit-length resistance to soar; dense repeater insertion is then mandatory to barely maintain delay, which in turn directly drives up power consumption and thermal density [9] [10]. In contrast, mature-node interconnects typically have widths in the micrometer to sub-micrometer range, with cross-sectional areas dozens of times larger than those of advanced-node thin lines. Their per-unit-length resistance is thus more than an order of magnitude lower [9] [11]. Such wide lines can maintain an acceptable RC delay over longer distances without any repeaters, and they do not introduce the extra power consumption and localized hotspots that repeaters bring [1]. This is precisely the irreplaceable physical capital of mature nodes in the context of system integration.

At the same time, mature nodes have low power density, no local hotspot hazards, and greatly reduced thermomechanical mismatch risks [3]. They require no sub-micron alignment and can be served by traditional packaging. Small-die assembly circumvents the single-die yield trap [8] [7], and the low resistance of individual interconnects keeps

signal-transmission power consumption manageable and economically viable [6].

This is not an argument for comprehensively replacing advanced nodes with mature ones. Rather, it points to a physically more self-consistent system integration path. The mature-node chiplet approach argued for in this paper requires that the compute cores themselves also adopt mature processes, so as to eliminate interconnect narrowing—the predicament unique to advanced nodes—at its physical root. Do not pursue extreme single-point density; pursue system-level physical self-consistency: every interconnect channel is sufficiently wide, RC delay is naturally manageable, every heat source is amply dispersed, and every packaging interface is spared the cyclic tearing of thermal expansion.

This is not a conservative retreat, but a renewed respect for physical laws. At a time when the physical penalties of advanced nodes are escalating rapidly and their practically deliverable performance is approaching a utility boundary, returning to system-level rational trade-offs is not only a response to engineering ethics, but also a responsibility to user interests and energy efficiency [6]. When one road becomes narrower and narrower, stepping back and choosing a broader path is not defeat; it is wisdom.

References

[1] J. C. Ku and Y. Ismail, "Thermal-Aware Methodology for Repeater Insertion in Low-Power VLSI Circuits," in IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 15, no. 8, pp. 963-970, Aug. 2007, doi: 10.1109/TVLSI.2007.900749.
keywords: {Repeaters;Very large scale integration;Delay;Temperature;Circuit simulation;Electrothermal effects;Coupling circuits;Analytical models;Integrated circuit interconnections;Logic;Low-power;repeater insertion;temperature},

[2] G. E. Moore, "Cramming more components onto integrated circuits, Reprinted from Electronics, volume 38, number 8, April 19, 1965, pp.114 ff.," in IEEE Solid-State Circuits Society Newsletter, vol. 11, no. 3, pp. 33-35, Sept. 2006, doi: 10.1109/N-SSC.2006.4785860.
keywords: {Integrated circuits;Computers;Silicon;Films;Heating;Microwave amplifiers;Data mining},

[3] H. Kim, J. Y. Hwang, S. E. Kim, Y. -C. Joo and H. Jang, "Thermomechanical Challenges of 2.5-D Packaging: A Review of Warpage and Interconnect Reliability," in IEEE Transactions on

Components, Packaging and Manufacturing Technology, vol. 13, no. 10, pp. 1624-1641, Oct. 2023, doi: 10.1109/TCPMT.2023.3317383. keywords: {Packaging;Substrates;Silicon;Reliability;Three-dimensional displays;Thermomechanical processes;Temperature measurement;25-D packaging;board-level reliability;interposer;solder fatigue life;warpage},

[4] M. Salmani Jelodar et al., "Tunneling: The major issue in ultra-scaled MOSFETs," 2015 IEEE 15th International Conference on Nanotechnology (IEEE-NANO), Rome, Italy, 2015, pp. 670-673, doi: 10.1109/NANO.2015.7388694. keywords: {Tunneling;Logic gates;High K dielectric materials;Performance evaluation;Leakage currents;MOSFET;Transistor scaling;quantum tunneling;high-k dielectric;source to drain tunneling},

[5] R. H. Dennard, F. H. Gaensslen, H. -N. Yu, V. L. Rideout, E. Bassous and A. R. LeBlanc, "Design of ion-implanted MOSFET's with very small physical dimensions," in IEEE Journal of Solid-State Circuits, vol. 9, no. 5, pp. 256-268, Oct. 1974, doi: 10.1109/JSSC.1974.1050511. keywords: {Doping profiles;Semiconductor process modeling;Predictive models;Threshold voltage;Fabrication;Switching circuits;Digital integrated circuits;MOSFET circuits;Ion implantation;Length measurement},

[6] A. Mallik et al., "Economics of semiconductor scaling - a cost analysis for advanced technology node," 2019 Symposium on VLSI Technology, Kyoto, Japan, 2019, pp. T202-T203, doi: 10.23919/VLSIT.2019.8776521. keywords: {Computer architecture;Cost benefit analysis;Nanoscale devices;Microprocessors;Industries;Tools},

[7] S. Naffziger et al., "Pioneering Chiplet Technology and Design for the AMD EPYC™ and Ryzen™ Processor Families : Industrial Product," 2021 ACM/IEEE 48th Annual International Symposium on Computer Architecture (ISCA), Valencia, Spain, 2021, pp. 57-70, doi: 10.1109/ISCA52012.2021.00014. keywords: {Industries;Program processors;Computer architecture;Packaging;Routing;Silicon;System-on-chip;Chiplets;Moore's Law;Processors;Modular;Industry},

[8] C. H. Stapper, "On Murphy's yield integral (IC manufacture)," in IEEE Transactions on Semiconductor Manufacturing, vol. 4, no. 4, pp. 294-297, Nov. 1991, doi: 10.1109/66.97812.
keywords: {Gaussian distribution;Integrated circuit yield;Integrated circuit modeling;Integrated circuit manufacture;Semiconductor device manufacture;Yield estimation;Manufacturing industries;Statistical

distributions;Manufacturing processes;Costs},

[9] R. Brain, "Interconnect scaling: Challenges and opportunities," 2016 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 2016, pp. 9.3.1-9.3.4, doi: 10.1109/IEDM.2016.7838381. keywords: {Integrated circuit interconnections;Metals;Delays;Resistance;Transistors;Wires;Capacitance},

[10] P. Kapur, G. Chandra and K. C. Saraswat, "Power estimation in global interconnects and its reduction using a novel repeater optimization methodology," Proceedings 2002 Design Automation Conference (IEEE Cat. No.02CH37324), New Orleans, LA, USA, 2002, pp. 461-466, doi: 10.1109/DAC.2002.1012669. keywords: {Repeaters;Integrated circuit interconnections;Power dissipation;Wires;Energy consumption;Delay;Computational Intelligence Society;Clocks;Permission;Optimization methods},

[11] W. Wu, S. H. Brongersma, M. Van Hove, K. Maex; Influence of surface and grain-boundary scattering on the resistivity of copper in reduced dimensions. Appl. Phys. Lett. 12 April 2004; 84 (15): 2838–2840. https://doi.org/10.1063/1.1703844