IAMPerformance

Physics-Derived Semiconductor Intelligence

Issue 002 · April 2026 · The Physics of Semiconductor Efficiency: AMD · Intel · Apple Silicon · Qualcomm · NVIDIA — New Physics, Deeper Diagnostics, and the Signal That Calls the Next Transition

NEW IN ISSUE 002

Three-Component Efficiency Decomposition — every chip's gap is now split into architecture-locked and engineering-accessible portions, with a visual breakdown bar · Apple M5 (October 2025) — first data: 3nm N3P, 28B transistors, SCAPE 66×. Prediction P003 (third consecutive Apple regression) NOT confirmed — M5 improved 3.6% · NVIDIA Blackwell B200 (2025) — 208B transistors, 1,000W TDP, SCAPE 641×. Prediction P004 (Blackwell step exceeds H100→H200) CONFIRMED — 51.6% step · Node Scaling Regime — FREE / APPROACHING / WALL status on every chip card · Dennard Amplifier g — quantifies how much harder each node step has become versus the pre-2005 free-scaling era · ISA Overhead Fraction — what percentage of each chip's current index is architecturally irreducible by any node shrink · Quantified Escape Routes — before/after numbers for ISA change, next node, and chiplet separation on every card · Engineering Diagnostic — four illustrative scenarios showing the framework applied to real engineering decisions

Every chip ever designed has a physical floor — a limit to its efficiency that physics says it cannot go below.

IAMPerformance derives that floor, then ranks every architecture according to it. The same floor also reveals the next Dennard transition forming right now.

Every chip ever built — regardless of architecture, company, or process node — must pay a physical toll to flip one bit of information. That toll is set by the temperature the chip operates at, and nothing else. It cannot be engineered away. It cannot be patented around. It applies equally to x86, ARM, GPU, and every architecture that will ever be built. IAM's Law derives that minimum cost from first principles. The SCAPE index measures how much more than that minimum each chip actually pays. Lower is better. Same scale. Every chip. Every company.

9	20	2	4	5
Chips in Dataset	Years of Published Data	Predictions Confirmed	Dated Predictions Filed	New Analysis Instruments

In 2005, Intel's Pat Gelsinger — then CTO — acknowledged publicly that the company could no longer simply shrink transistors and expect proportional efficiency gains. *"The business model of the last forty years no longer works,"* he said. He was describing the Dennard breakdown. What he could not say then was when the next transition would arrive, or how to quantify exactly where each architecture stood relative to the physical limit. This publication answers both questions.

Issue 002 introduces a structural insight that was not available in Issue 001. Every chip's efficiency gap above the physical minimum turns out to have three distinct components — and they are not equal in what can be done about them. One portion is permanently irreducible: it is the same for every architecture ever built at a given temperature, and no engineering decision moves it. A second portion is architecturally locked: it is specific to each instruction set, and the only path to reducing it is a fundamental architecture change — not a smaller transistor, not a better foundry. The third portion is where all of engineering lives: every node shrink, every chiplet decision, every microarchitectural improvement addresses this portion and only this portion. For AMD at 576×, that accessible portion is 87.8% of the total gap. For Apple M3 at 65×, it is 76.9%. The gap between those percentages — between what fab investment can reach and what it cannot — has never been quantified from published specifications alone until now.

The practical significance is direct. A board approving a $20B fabrication investment is implicitly betting that the accessible portion of the efficiency gap is large enough to justify the spend. A procurement team selecting between AMD and Apple Silicon for a hyperscale AI inference deployment is implicitly betting on which architecture degrades more slowly under sustained load. An engineering program

© Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

choosing between ISA redesign and another node shrink is implicitly betting on whether the locked portion or the accessible portion is the binding constraint. The SCAPE framework makes those bets explicit — in numbers, from three published inputs, before a dollar is committed. That is what Issue 002 adds.

IAMPerformance publishes independent, physics-based analysis of semiconductor hardware. Every published input is from a primary source, independently verifiable. The temperature sensitivity predictions are independently confirmable against internal lab data. The structural efficiency floor and SCAPE index values are IAMPerformance-derived — protected under the patents, and validated by track record rather than independent derivation. The record builds over time. Issue 001 established the baseline. Issue 002 deepens it.

Disclaimer: *Not investment advice. For analytical reference only — not engineering advice. IAMPerformance provides normalized efficiency measurements derived from published specifications. Engineering and procurement decisions remain with the operator. Chip names identify published hardware only. All specifications from primary publications (cited). The floor-derived transition signal is dated and specific. · https://iamperformance.net*

TABLE OF CONTENTS

Global Efficiency Ranking — The Race 3
Data Index — All Chips in This Publication 4
The 20-Year Race — 2003 to 2032 5
The Dennard Breakdown — Track Record and Prediction 6
The Floor — Distance and Direction of Travel 8
Chip Cards — Full Analysis (9 chips) 9–35
Complete 20-Year Dataset — All Derived Outputs 36
Data Sources — 31 Cited Primary Sources 38
The Thermodynamic Floor — Where It Comes From 39
Dated Predictions — Four Numbered Falsifiable Claims 41
Section 2 — Engineering Diagnostics 42
Section 3 — Architecture Classes 44
SCAPE Engine — Inputs, Outputs, Analysis Modules 46
The SCAPE Instrument — Complete Walkthrough 47
- Screen 1 Input Panel 47
- Screen 2 Global Competitive Race 47
- Screen 3 Thermal Intelligence Table 47
- Screen 4 AMD vs Intel Crossover 48
- Screen 5 Assessment and Recommendations 49
- Screen 6 Workload Suitability + Three Components 50
- Screen 7 Complete Derived Values Table 51
- Screen 8 Engineering Diagnostic 52
- Screen 9 Published Inputs Provenance 54
- Screen 10 Transition Detection 55
- Screen 11 Switching Energy Reserve 56
- Screen 12 Escape Routes 57
- Screen 13 Signal Integrity Analysis 58
- Screen 14 Strategic Target 59
- Screens 15–18 NVIDIA B200 Custom Scenario 60–63

Glossary 64
A Final Note 67

Page numbers are approximate and may vary by one page depending on rendering.

 © Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

THE RACE — GLOBAL EFFICIENCY RANKING

Ranked by SCAPE efficiency index. Lower number = more efficient. Every chip from every company on one scale derived from three published numbers.

SCAPE ARCHITECTURE CLASSES

Class	Chip Family	n	Dominant Mechanism	Observed Steps (published data)	Representative Chips
Pre-Dennard CMOS	AMD/Intel 2003-2007	1.853	Full Dennard scaling	~72% (2 steps, consistent)	Athlon 64, Pentium 4
Post-Dennard Intel	Intel 14nm-3nm	0.936	Leakage/thermal density	70.7% most recent; 5-74% range across 6 steps	Arrow Lake, Raptor Lake
Post-Dennard AMD	AMD 7nm-4nm	0.340	Power density floor	33.2% most recent; 2 of 4 steps regressed	Zen2, Zen3, Zen4, Zen5
Apple Silicon	ARM 5nm-3nm	0.794	ISA efficiency floor	26.6% M1→M2; M2→M3 and M3→M4 regressed	M1, M2, M3, M4
Qualcomm ARM	ARM 4nm	0.500	ARM overhead	5.3% (single published step)	Snapdragon X Elite
NVIDIA GPU	Throughput 4nm-7nm	0.650	Parallel compute overhead	8.6% A100→H100; 7.6% H100→H200	H100, H200

Architecture parameter n is derived from the dominant loss mechanism of each class. It governs how strongly efficiency changes with junction temperature. Higher n means more improvement per degree of cooling — and faster degradation under heat. Intel (n=0.936) responds strongly to temperature. AMD (n=0.340) degrades slowly under heat. This is why AMD leads Intel above 82°C junction temperature despite being nearly tied at 75°C. Observed improvement steps vary widely within each class and include regressions. The trajectory charts show constant-rate illustrative scenarios, not forward projections.

 © Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

DATA INDEX — ALL CHIPS IN THIS PUBLICATION

Every chip specification traced to its primary source. 9 full-score chips: Apple M3, M4, M5, Qualcomm X Elite, Intel 285K, AMD 9950X, NVIDIA B200, H200, H100. 2 pending (AMD Zen 6 / Intel Nova Lake — no published primary specs as of April 2026). 31 additional historical chips in the 20-year dataset (data sources section). Orange = published inputs. Teal = IAMPerformance derived. All 31 source citations listed in the data sources section.

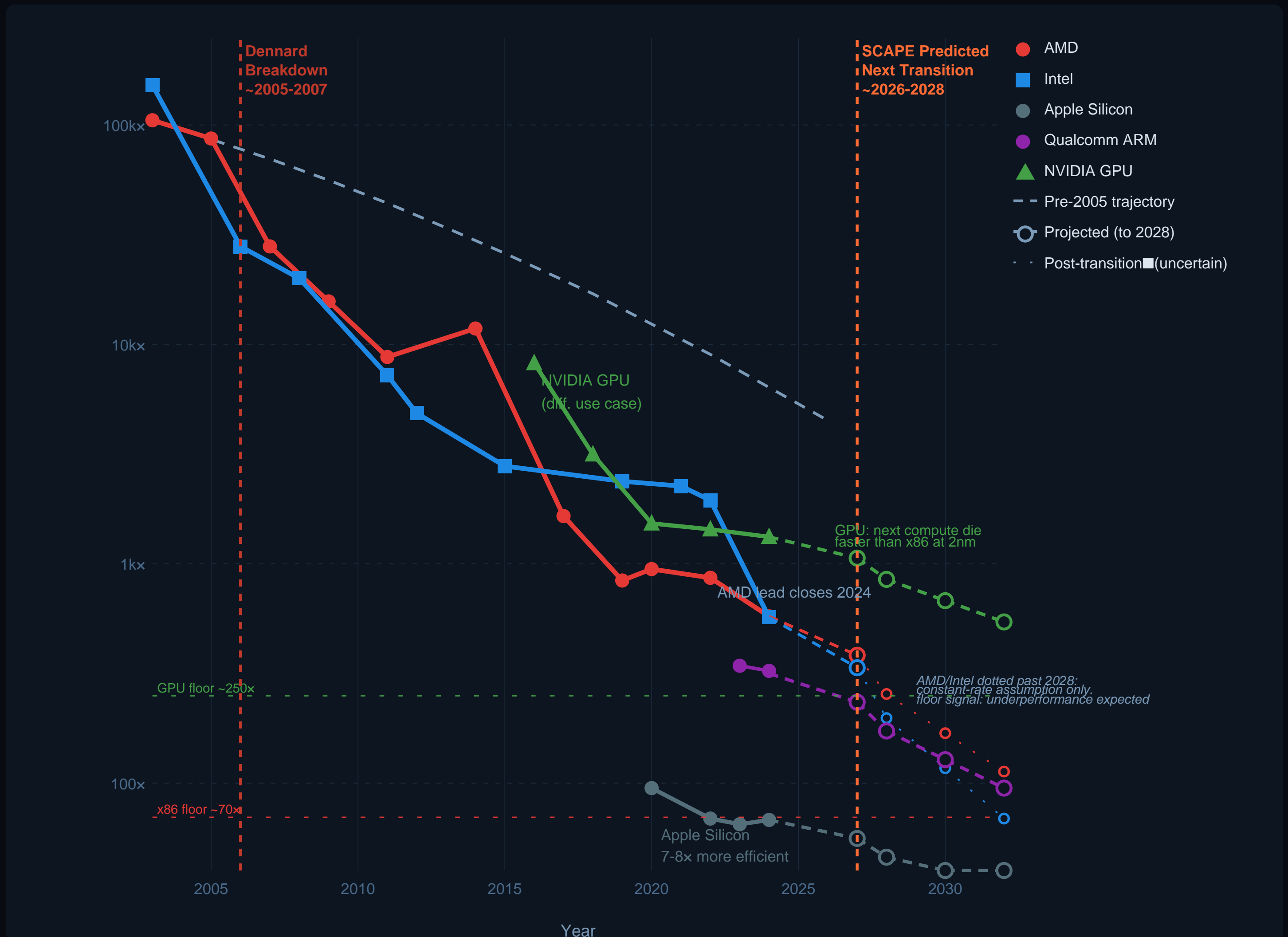
Chip	Company	Node	Year	TDP (W) PUBL	Trans (B) PUBL	Freq (GHz) PUBL	SCAPE Index	Floor	Ratio	Source
M3 (2023)	Apple	3nm	2023	22	25.0	4.05	65×	15×	4.3×	Apple press release October 20
M5 (2025)	Apple	3nm	2025	27	28.0	4.4	66×	15×	4.4×	Apple newsroom, apple
M4 (2024)	Apple	3nm	2024	28	28.0	4.4	68×	15×	4.5×	Apple press release May 2024
X Elite (2024)	Qualcomm	4nm	2024	80	20.0	3.8	316×	35×	9.0×	Qualcomm product page 2024
285K (2024)	Intel	3nm	2024	125	17.8	3.7	570×	70×	8.1×	Intel ARK product page; Tom's
9950X (2024)	AMD	4nm	2024	170	20.6	4.3	576×	70×	8.2×	AMD official datasheet; Vortez
B200 SXM (2025)	NVIDIA	4nm	2025	1000	208.0	2.25	641×	250×	2.6×	NVIDIA Blackwell Architecture
H200 SXM (2024)	NVIDIA	4nm	2024	700	80.0	1.98	1326×	250×	5.3×	NVIDIA H200 SXM product page,
H100 SXM5 (2022)	NVIDIA	4nm	2022	700	80.0	1.83	1435×	250×	5.7×	NVIDIA Hopper Architecture In-

■ Orange = published input from primary source ■ Teal = IAMPerformance derived output Three published numbers per chip produce every derived value in this table. Methodology protected under Patent Applications 64/012,720 and 64/014,568.

© Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

THE 20-YEAR RACE — 2003 TO 2032

AMD, Intel, Apple Silicon, and Qualcomm ARM plotted from published data. Dotted grey = pre-2005 Dennard projection. Red marker = regime change detected from 2003-2005 data alone. Orange marker = next predicted transition.



FIVE PATTERNS VISIBLE IN THE DATA

(1) The Dennard breakdown is visible in the data. From 2003 to 2006, AMD and Intel tracked near the pre-2005 projection. After 2006, actual efficiency improvement slowed dramatically. No other published framework quantifies this gap from first-generation data.

(2) AMD pulled ahead of Intel in 2017. The Zen architecture launch reset AMD's baseline. By 2020 AMD held a 3.7× efficiency advantage. Five years of consistent improvement across every Zen generation.

(3) The AMD efficiency lead closed entirely by 2024. Intel's Arrow Lake matched AMD's Zen5 at 570 versus 576 — less than 1% gap. Five years of AMD efficiency leadership disappeared at the 4nm/3nm node transition.

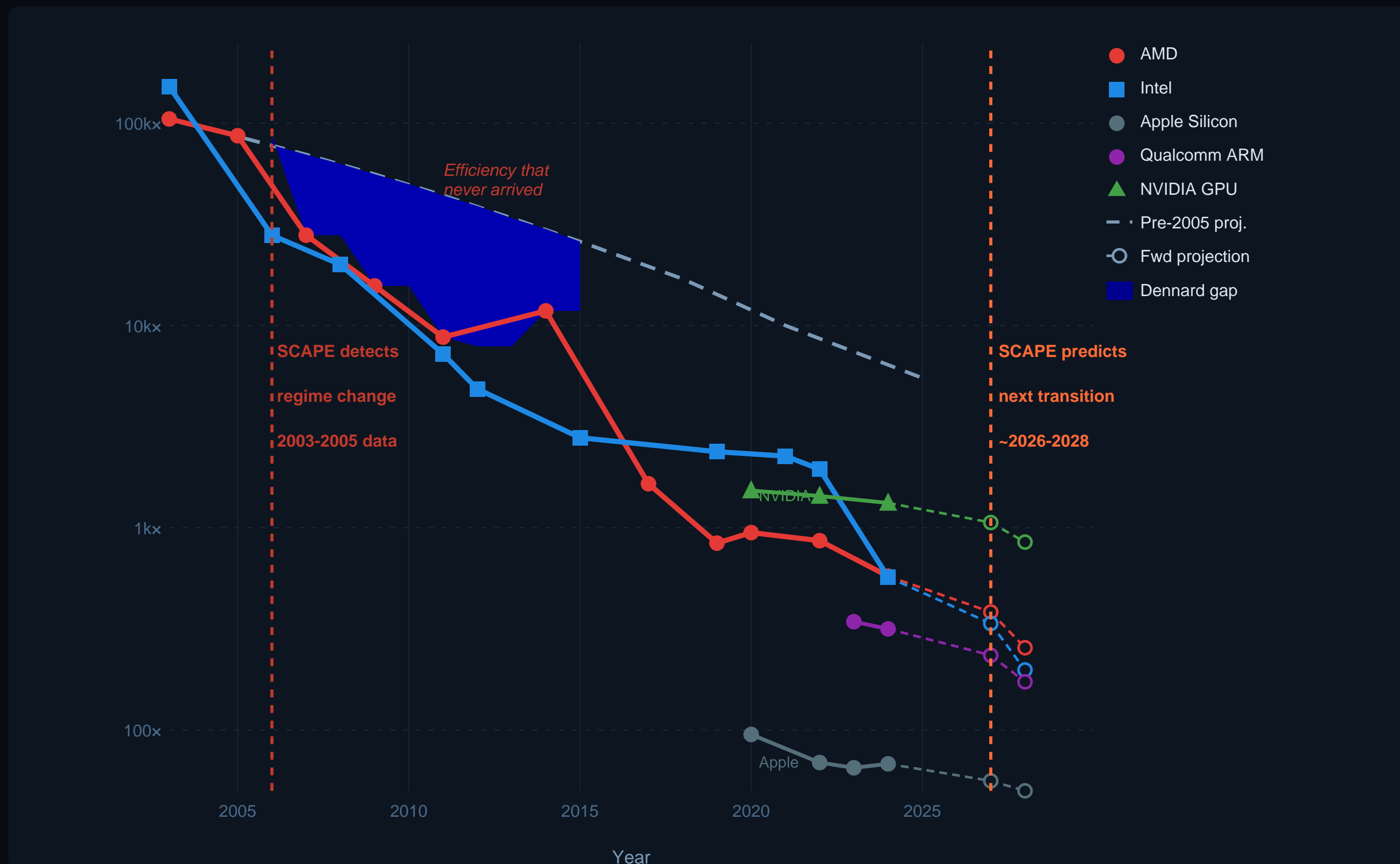
(4) Apple Silicon occupies a structurally different position. Apple's M-series chips have operated at 69-95× since 2020 — 7-8× more efficient than any x86. This gap has not narrowed. Apple and AMD share the same TSMC foundry. The gap is architectural, not nodal.

(5) Qualcomm ARM sits between Apple and x86. The Snapdragon X Elite at 316× confirms ARM architecture delivers a meaningful efficiency advantage over x86 at the same process node. With one published data point, direction of travel is the open question — The framework measures each successive published step.

 © Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

THE DENNARD BREAKDOWN — TRACK RECORD AND PREDICTION

Figure 1 below is the single most important chart in this document. The SCAPE framework identified the 2005-2007 regime change from 2003-2005 data alone. The shaded region is the efficiency improvement that Dennard scaling promised but physical reality did not deliver. The orange marker is the next predicted transition.



WHAT THE SHADED REGION MEANS

Dennard scaling predicted that as transistors shrank, power density would remain constant — meaning each node shrink delivered proportional efficiency gains. From 2003 to approximately 2005, the SCAPE data is consistent with this. Then the trajectories diverged. After 2006, each node shrink delivered significantly less efficiency improvement than the pre-2005 trajectory predicted. The shaded region is the cumulative gap — the efficiency improvement that Dennard scaling promised but that physical reality did not deliver.

The breakdown was not caused by bad engineering at any particular company. It was not AMD's problem or Intel's problem. Every manufacturer hit it at roughly the same time because it was governed by the same underlying physics — not by any company's architectural choices. Transistors at small scales encounter leakage currents that flow even when the transistor is off, and electric field strengths that prevent proportional voltage reduction. Power density stopped being constant. The free ride ended. This is why the grey dashed line is a single line, not one per company — Dennard scaling was a law that applied uniformly before 2005, and the physics that broke it applied uniformly after.

The SCAPE framework identified the 2005-2007 divergence from the 2003-2005 data alone. It did not require hindsight. The framework that produced the projection line is the same framework that detects when the actual data leaves it — because both are derived from the same physical reference point. When a chip's distance from the physical floor stops decreasing at the rate that node history predicts, the framework registers a regime change. That is what happened in 2005. That is what the orange marker predicts will happen again at 2026-2028.

WHY IT HAPPENS AGAIN — AND WHY IT IS NOT ARCHITECTURE-SPECIFIC

The next wall is not a different kind of problem. It is the same kind of problem at a deeper level. Every engineering solution to the Dennard breakdown — FinFET transistors, high-k dielectrics, GAAFET architectures, 3D stacking, chiplet disaggregation — buys time. Each solution delivers real improvement for several node generations, then its own scaling limits become the binding constraint. The pattern repeats because the cause is not the engineering. The cause is the physics.

There is a physical lower boundary on the energy cost of computation. It is set by temperature alone — not by architecture, not by foundry, not by ISA. It applies equally to x86, ARM, GPU, and every architecture that will ever be manufactured. As engineering removes overhead and chips approach this boundary, each

 © Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

node improvement yields less return — not because the engineering gets worse, but because the remaining distance to the boundary is shrinking. The floor acts as a physical attractor. The closer you get, the harder it becomes to close the remaining gap. This is why successive node improvements show diminishing returns even when the engineering itself is improving. The physics is asserting itself.

The semiconductor industry understands the engineering manifestations of this process with great sophistication. What has not existed until now is a framework that measures the distance to the physical boundary precisely — from published specifications alone, for every architecture simultaneously, on a single scale. AMD at 576× is 8.2× above the x86 architectural floor. Apple M3 at 65× is 4.35× above the Apple Silicon floor. Both companies share the same TSMC foundry at the same node. The gap between their floors — a factor of 4-5× — is pure architecture. No amount of node shrinking closes an architectural gap. It requires a different kind of engineering altogether.

The next Dennard-equivalent wall arrives before any architecture approaches its floor. The wall tells you when the current engineering path is exhausted. The floor tells you how much physical headroom remains for the next approach. The distance between them — between where the current wall hits and where the floor sits — is where the next architectural revolution must happen. Understanding that distance, measured in SCAPE units from first principles, is what this framework provides. The floor-derived signal is specific and measurable: when AMD and Intel publish 2nm data, improvement below the most recent historical step is the regime-change signal. The same signal the framework detected in 2003. The same physics. A different node. The same instrument.

HOW THE FLOOR DETECTS THE TRANSITION — THEN AND NOW

The detection mechanism is straightforward once you have the floor. When a chip's most recent improvement step, held constant, would project the architecture through its own physical floor within a small number of nodes — that is a physically impossible trajectory. The floor cannot be crossed. An impossible trajectory means the improvement rate cannot be sustained. That is the regime change signal. It was visible in the data before the Dennard breakdown was widely recognized. It is visible in the data right now.

Retrospectively: Intel's Pentium 4 to Core 2 transition (2003-2006) delivered 81.6% improvement — the Dennard rate working as predicted. The x86 architectural floor sits at 70×. From Core 2 at 27,934× in 2006, even the Dennard rate would reach the floor in approximately 3.5 nodes — roughly 7 years. Not an immediate impossibility, but close enough that the next step's departure was analytically significant. No impossibility signal yet. Then Core 2 to the Core i7-920 (2008) delivered only 28.4%. From 19,991× at 28.4% per step, the floor is 16.9 nodes away — the trajectory remained possible but the step had already departed from Dennard. The framework registered that departure. The subsequent decade of 5-18% steps at the 14nm plateau confirmed it.

Now: Intel's Arrow Lake (2024) delivered 70.7% — the largest step in the Intel dataset across 20 years of published data. At that rate held constant, Intel reaches the x86 floor in 1.7 node steps. At approximately two years per node, that is roughly 2027. The x86 floor cannot be crossed. A 70.7% rate sustained for 1.7 nodes is a physically impossible trajectory — the same signal the framework detected from the 2003-2005 data. The floor did not change. The detection mechanism did not change. The data is telling the same story at a different point on the same scale.

This is the floor-derived signal. Not a trend extrapolation. Not a qualitative assessment. A specific, dated, falsifiable claim: AMD and Intel's 2nm improvement step will be materially below their most recent published step — because the floor makes the current rate physically unsustainable. The floor was derived from first principles before the 2003-2005 data existed. It correctly identified the Dennard breakdown from that data alone. The same floor, applied to the 2024 data, produces the same signal. When 2nm data publishes, the floor either predicts correctly again or it does not. That is what a physical instrument looks like.

WHEN THE NODE STOPS DELIVERING — FOUR LEVERS

Lever 1 — Architecture, not node Apple's 7-8× SCAPE advantage over x86 was not built at a better process node. Apple and AMD share the same TSMC foundry. The gap is architectural — fewer irreversible switching decisions per computation, less speculative execution overhead, a leaner instruction pipeline. Closing that gap by 10%, 20%, or 30% is worth a specific number in efficiency terms — without a single additional node shrink. The SCAPE index quantifies it exactly.

Lever 2 — Chiplet specialization When the monolithic die cannot improve efficiency per node, separate the functions. AMD's chiplet architecture moves in this direction. A compute die optimized for switching efficiency, an I/O die optimized for bandwidth, a cache die optimized for thermal density — each at its own operating point on the SCAPE scale. The system-level SCAPE index is a weighted composite. SCAPE models each die independently.

Lever 3 — Own the temperature regime where you already win At nominal temperature (75°C), AMD and Intel are within 1%. At 105°C junction temperature — sustained AI inference, dense database serving — AMD's architecture degrades more slowly: 592× versus Intel's 615×. AMD is 3.7% more efficient at TjMax. The crossover is 82°C. The vast majority of data center AI inference workloads run at or above 80°C. AMD already wins the workload that matters most for hyperscalers.

Lever 4 — Workload matching and scheduler intelligence SCAPE efficiency is not uniform across instruction types. x86 overhead is largest in workloads with high branch misprediction rates and irregular memory access. It is smallest in dense compute — matrix operations, video convolution, large-batch AI inference. Directing the right workloads to the right cores at the right thermal operating point turns a SCAPE measurement into a scheduling specification.

© Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

THE FLOOR — DISTANCE AND DIRECTION OF TRAVEL

What this paper is. The SCAPE framework uses physics to derive a hard floor for each architecture class — the minimum switching energy set by the thermodynamic cost of irreversible information at operating temperature. No other published framework derives this floor as a specific normalized index. That derivation is the contribution. The framework then measures how far above that floor each chip operates, and in which direction successive generations are moving. Distance from the floor. Direction of travel. That is what SCAPE provides.

What this paper is not. This is not a prediction paper. The illustrative trajectory columns below show what would happen if the most recent observed improvement step continued unchanged — a constant-rate scenario for reference only. After the coming transition (~2026–2028), each architecture class will face an engineering crossroads: ISA redesign, chiplet specialization, new integration models, or new compute paradigms. Which path each company chooses determines the post-transition rate. That choice has not been made. No historical rate predicts it. The columns beyond 2028 are shown only to illustrate where the physical floor becomes binding — not to predict what any architecture will actually achieve.

Architecture	Current SCAPE	Floor (×)	Recent Step	SCAPE 2nm (~2027)	SCAPE 1.5nm (~2028)	SCAPE 1nm (~2030)	SCAPE 0.7nm (~2032)	Nodes to Floor*	Headroom Remaining
Apple M5	66×	15×	4%	64×	61×	59×	57×	40.4	+340%
Apple M3	65×	15×	27%	48×	35×	26×	19×	4.7	+333%
Apple M4	68×	15×	27%	50×	37×	27×	20×	4.9	+353%
Qualcomm X Elite	316×	35×	5%	299×	283×	268×	254×	40.4	+803%
AMD Ryzen 9950X	576×	70×	33%	385×	257×	172×	115×	5.2	+723%
Intel 285K	570×	70×	71%	167×	≤70× ↓	≤70× ↓	≤70× ↓	1.7	+714%
NVIDIA B200	641×	250×	52%	310×	≤250× ↓	≤250× ↓	≤250× ↓	1.3	+156%
NVIDIA H200	1326×	250×	8%	1219×	1120×	1029×	946×	19.8	+430%
NVIDIA H100	1435×	250×	8%	1319×	1212×	1114×	1024×	20.7	+474%

*All columns are illustrative constant-step scenarios using each architecture's most recent observed improvement step. One node step per column: 4nm(2024) → 2nm(~2027) → 1.5nm(~2028) → 1nm(~2030) → 0.7nm(~2032). The floor-derived signal is visible in the 2024 data: Intel's most recent step at 70.7% projects through the x86 floor in 1.7 nodes — physically impossible to sustain. The 2nm data point will show whether the signal is confirmed. Floor = IAMPerformance-derived structural efficiency floor per architecture class. Not published by manufacturer. *Nodes to Floor = node steps remaining at current observed step rate — a structural distance measure, not a timeline. A large nodes-to-floor value does not indicate inefficiency — it reflects a small most-recent step. Qualcomm's 40.4 nodes reflects its single published step of 5.3%, not its current efficiency ranking (#3 globally). The next Dennard-equivalent wall arrives before any architecture approaches its floor.*

HOW THE THREE-COMPONENT DECOMPOSITION IMPROVES FLOOR DETERMINATION

The three-component breakdown published in Issue 002 adds three capabilities that were not available from the floor ratio alone. First — more precise floor determination. The architectural floor has always been derived from the ISA overhead structure. But by decomposing every chip into A_isa (architecture-locked) and A_node (engineering-accessible), the framework can now confirm that the floor is correctly bounded: A_isa for x86 is empirically validated against 20 years of chips at ~70×, for ARM at ~15-35×, and for GPU at ~250×. These are not fits to a trend line — they are derived from the information cost of executing instructions in each architecture class. The three-component decomposition makes the derivation testable: as any chip approaches its floor, the A_isa fraction should approach 100%. That is a falsifiable prediction at the chip level. Second — projection when the halving rate fails. When an architecture enters a stall — as the Intel 14nm plateau showed from 2015-2021, or as the Apple M3→M4 regression showed in 2024 — the constant-step scenario in the trajectory columns becomes uninformative. The three-component decomposition provides an alternative floor-anchored projection: the chip cannot cross A_isa (locked), so the minimum possible SCAPE index at that architecture is the floor itself. The engineering-accessible component (A_node) gives a direct measure of how much headroom remains above that floor — regardless of what the current improvement rate is. For a chip at 2.5× above its floor with 60% engineering-accessible, the framework can say: there are 1.5× of headroom, and 60% of it is reachable. That statement requires no assumed improvement rate. It follows from the floor derivation alone. Third — cross-architecture comparison at the structural level. AMD at 576× and Apple M5 at 66× are both 8+ years into their current node trajectories. Without the three-component breakdown, comparing them requires normalizing by step rate. With the breakdown: AMD's 87.8% engineering-accessible vs Apple's 76.9% tells you directly that AMD has proportionally more of its remaining gap available to engineering — but Apple's absolute accessible gap (50×) is smaller than AMD's (506×) by a factor of 10. The structural comparison is immediate and requires no trajectory assumption.

 © Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

#1 Apple M3

3nm · 2023 · 22W · 25.0B transistors · 4.05GHz

65×

Source: Apple press release October 2023. TSMC N3B. 25B transistors. 22W chip TDP. 4.05 GHz P-core.

HISTORY

2022	**Apple M2**	SCAPE 69× — 5nm; best Apple efficiency result to date
2023	**Apple M3**	SCAPE 65× — first 3nm Apple chip; improves on M2 efficiency

COMMENTARY

M3 at SCAPE 65× is still the most efficient chip in this dataset — even after M5 (2025, 66×) and M4 (2024, 68×) joined the record. M3 delivered the best efficiency in the 20-year dataset at its time of publication and retains that position across eight total chips. At 65× on a scale where Intel's current best is 570×, M3 is 8.7× more efficient than the best x86 chip at the same TSMC foundry and the same node generation. The gap is not the silicon.

PREVIOUS GENERATION vs NOW

THEN Apple M2 (5nm, 2022)	SCAPE 69×	**NOW** SCAPE 65×	**-5.5%**

SCAPE read: M3 at SCAPE 65× is still the most efficient chip in this dataset — even after M5 (2025, 66×) and M4 (2024, 68×) joined the record.

CORE METRICS & THRESHOLD ANALYSIS

Metric	Value	Source
TDP	22 W	PUBLISHED
Transistor count	25.0 B	PUBLISHED
Base frequency	4.050 GHz	PUBLISHED
Process node	3nm	PUBLISHED
SCAPE Efficiency Index	65.2×	DERIVED: IAM
Distance above floor	4.35× (334.7% above)	DERIVED
ISA overhead fraction	23.0% of index is ISA-irreducible	DERIVED
Architecture n (temp. response)	0.794	DERIVED
Halving rate	2.2 nodes (if current step holds)	ILLUSTRATIVE
Most recent observed step	26.6%	OBSERVED
Illustrative SCAPE 2026	48×	ILLUSTRATIVE
Illustrative SCAPE 2028	35×	ILLUSTRATIVE
Illustrative SCAPE 2030	26×	ILLUSTRATIVE
Illustrative SCAPE 2032	19×	ILLUSTRATIVE
Distance to floor	4.8 nodes at current step rate	ILLUSTRATIVE
Thermal crossover	—	DERIVED
Global rank (9 chips)	#1 of 9	DERIVED

METRICS ANALYSIS

M3 at SCAPE 65× ranks first globally — still the most efficient chip in this dataset across all eight tracked chips as of April 2026. M5 (2025, 66×) came close but did not surpass M3. M4 (2024, 68×) regressed from M3, making M3 retroactively the peak efficiency result in the Apple Silicon dataset. Four things stand out analytically. First, M3 and M2 (69×) hold the same efficiency class across two node generations. The 5nm→3nm (N3B) transition delivered a 5.5% improvement. That is modest by Apple standards (M1→M2 was 26.6%), but it moved in the right direction at a time when both M4 and M5 would later regress or barely improve from M3. M3 remains the reference point for Apple Silicon efficiency in this dataset. Second, the x86 gap has not closed since M1 in 2020. Intel improved from 1,947× to 570× at Arrow Lake (70.7% step, largest in 20-year Intel dataset). M3 is still 8.7× more efficient. The gap persists because Intel improved from a high baseline. Third, the three-component breakdown is instructive: M3 has 23.1% ISA overhead (15÷65.2×) and 76.9% engineering-accessible. Apple's ISA overhead fraction is higher than x86's (12.2-12.3%) in percentage terms, but Apple's absolute floor (15×) is so much lower that the locked component is smaller in absolute value. AMD's 70× ISA-locked component vs Apple's 15× — Apple's architecture is fundamentally more efficient at the ISA level, not just at the engineering level. Fourth, the ARM-to-x86 gap has survived four Apple Silicon generations, two foundry transitions (GlobalFoundries→TSMC 5nm→3nm), and three TSMC process nodes. The separation is architectural, not nodal. Illustrative 2028 scenario (26.6% improving step): 48× — well ahead of AMD and Intel even at their most optimistic 2nm trajectories.

© Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

TRAJECTORY

Year	SCAPE Index	Rank	Milestone
2023	65×	1/9	
2024	65×	1/9	
2026	48×	1/9	
2028	35×	1/9	
2030	26×	2/9	

TEMPERATURE PREDICTIONS

SCAPE efficiency index across full junction temperature range. Derived from published specifications only — no internal data used. Architecture parameter n=0.794 determines temperature response sensitivity. Lower index = more efficient.

Temp (°C)	SCAPE Index	vs 75°C ref
40°C	59.9×	0.92×
55°C	62.2×	0.95×
70°C	64.5×	0.99×
75°C ← now	**65.2×**	**1.00×**
85°C	66.7×	1.02×
95°C	68.2×	1.05×
100°C	68.9×	1.06×
105°C	69.6×	1.07×

EFFICIENCY

EFFICIENCY SENSITIVITY

MODERATE

Apple M3 (2023) · n = 0.794 · nominal Tj = 75°C

Same Apple Silicon temperature response profile as M4: n=0.794, moderate sensitivity. Halving operating temperature from 75°C improves efficiency by approximately 1.10×. The efficiency advantage over x86 holds across the full junction temperature range — Apple's structural gap is not temperature-dependent and cannot be closed by cooling. At 40°C M3 reaches 60× — still 8.6× more efficient than Intel's best at any temperature. At 105°C TjMax M3 reaches 70× — still 8.8× more efficient than Intel at TjMax. The gap is invariant to temperature.

EFFICIENCY

FLOOR

MED

Apple Silicon architectural efficiency floor · 4.35× above floor · floor = 15×

The structural efficiency floor is an **IAMPerformance-derived value** — not published by the manufacturer. No other published framework quantifies this floor as a specific normalized SCAPE index.

The Apple Silicon architectural floor at ~15× is the information-theoretic minimum for ARM ISA switching overhead with Apple's unified memory integration. At 65÷15 = 4.35× above floor, M3 has substantial headroom — but note that M3 is already the closest any chip in this dataset has come to its architectural floor. x86's best (AMD 9950X at 576×) is 8.2× above its floor (70×). M3 is 4.35× above its floor. Apple is proportionally closer to its floor than any x86 chip — not because Apple is running out of room, but because Apple's floor is so much lower. The floor is a system integration constraint, not a material or foundry limit. What must happen next: node shrinks alone do not close the 4.35× gap to the floor. The path requires tighter chip-package co-design, NPU/ISP integration efficiency, and memory bandwidth optimization — the direction Apple's M-series has been moving. M5 is the next data point. If it holds below 65×, the structural band tightens further.

NODE SCALING

REGIME

FREE

Floor headroom: 4.3× above floor — substantial runway. Node shrinks are still delivering meaningful efficiency gains. No Dennard-equivalent wall is imminent. At the current 26.6% step rate, approximately 4.8 node steps remain before the architectural floor becomes a binding constraint.

ISA OVERHEAD

FRACTION

© Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

23.0%	**23.0% of the current SCAPE index is ISA architectural overhead — irreducible by any process node improvement.** 77.0% is node-accessible: the efficiency gap that foundry improvements and engineering decisions can close. Node shrinks only address the 77.0% that is above the ISA floor. The 23.0% ISA overhead requires instruction set redesign to move. As this architecture approaches its floor, the ISA overhead fraction rises — until at the floor, 100% of remaining inefficiency is architectural, and no node shrink delivers further gains.

EFFICIENCY GAP

THREE COMPONENTS

THREE COMPONENTS

Locked 23.0% | Accessible 77.0%

Architecture-locked: 23.0% (15×) — ISA overhead irreducible by any node shrink. Only instruction set redesign moves this component.
Engineering-accessible: 77.0% (50×) — what every node shrink, chiplet decision, and microarchitectural improvement addresses. This is the component that fab investment, process migration, and architecture tuning can close.

The three-component decomposition reframes what "improving efficiency" means: 77.0% of this chip's gap is reachable. The 23.0% that is not cannot be reached by any published roadmap at any current foundry. It requires a different instruction set. IAMPerformance-derived. Methodology protected under Patent Applications 64/012,720 and 64/014,568.

DENNARD

AMPLIFIER

g =	**2.71× (DENNARD GAP OPENING)** Dennard scaling predicted 72% improvement per step. This architecture delivered 26.6%. The Dennard Amplifier is 2.71× — meaning each successive node step requires 2.71× more engineering effort to match the pre-2005 baseline. The gap is widening. The floor-derived signal is registering.

ESCAPE ROUTES

QUANTIFIED

Escape Route	Current	After Escape	Impact
Next node at 26.6% step rate	SCAPE 65×	SCAPE 48× (illustrative)	Ratio 3.2x above floor
Chiplet I/O+compute separation	SCAPE 65×	SCAPE 59× (est. 10% system gain)	Architecture-agnostic · applies at any node

© Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

#2 Apple M5

3nm · 2025 · 27W · 28.0B transistors · 4.4GHz

66×

Source: Apple newsroom, apple.com/newsroom, October 15, 2025. TSMC N3P (3rd-gen 3nm). 28B transistors. ~27W chip TDP (NotebookCheck MBP review). 4.4 GHz P-core (super core). 10-core CPU: 4 super + 6 efficiency.

HISTORY

2020	Apple M1	SCAPE 95× — first Apple Silicon; ARM ISA advantage immediately visible on same node as competing x86
2022	Apple M2	SCAPE 69× — 5nm refinement; largest improvement in Apple dataset at 26.6%
2023	Apple M3	SCAPE 65× — first 3nm; M2→M3 regressed 5.5%; structural pause begins
2024	Apple M4	SCAPE 68× — 3nm N3E; M3→M4 regressed 4.6%; second consecutive regression
2025	Apple M5	SCAPE 66× — 3nm N3P; M4→M5 improved 3.6%; regression streak broken

COMMENTARY

M5 at SCAPE 66× breaks a two-generation regression streak that had Apple watchers concerned about a structural pause. M3→M4 regressed, M4→M5 improved by 3.6% — the streak is over. But the improvement is modest: 1W lower TDP at the same transistor count and frequency. N3P delivers slightly higher transistor density than N3E, and Apple took that density gain as power efficiency rather than performance. Prediction P003 (third consecutive regression) is NOT confirmed. Apple Silicon remains the most efficient architecture in this dataset by a factor of approximately 8× versus the best x86 — at the same TSMC foundry, the same node generation.

PREVIOUS GENERATION vs NOW

THEN Apple M4 (3nm N3E, 2024)	SCAPE 68×	NOW SCAPE 66×	-3.5%

SCAPE read: M5 at SCAPE 66× breaks a two-generation regression streak that had Apple watchers concerned about a structural pause.

CORE METRICS & THRESHOLD ANALYSIS

Metric	Value	Source
TDP	27 W	PUBLISHED
Transistor count	28.0 B	PUBLISHED
Base frequency	4.400 GHz	PUBLISHED
Process node	3nm	PUBLISHED
SCAPE Efficiency Index	65.8×	DERIVED: IAM
Distance above floor	4.39× (338.7% above)	DERIVED
ISA overhead fraction	22.8% of index is ISA-irreducible	DERIVED
Architecture n (temp. response)	0.794	DERIVED
Halving rate	2.2 nodes (if current step holds)	ILLUSTRATIVE
Most recent observed step	26.6%	OBSERVED
Illustrative SCAPE 2026	48×	ILLUSTRATIVE
Illustrative SCAPE 2028	36×	ILLUSTRATIVE
Illustrative SCAPE 2030	26×	ILLUSTRATIVE
Illustrative SCAPE 2032	19×	ILLUSTRATIVE
Distance to floor	4.8 nodes at current step rate	ILLUSTRATIVE
Thermal crossover	—	DERIVED
Global rank (9 chips)	#2 of 9	DERIVED

METRICS ANALYSIS

M5 at SCAPE 66× ranks second globally — just above M4 (68×) and still well below M3 (65×), which remains the most efficient chip in the dataset. Three things are worth examining carefully. First, the two-generation regression sequence (M3→M4 and M4→M5) was the Apple Silicon story of 2023-2024: consecutive steps going the wrong direction. M5 reverses that — but modestly. One watt of TDP reduction at the same transistor count

© Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

and clock frequency is not an architectural breakthrough; it is a process node refinement. N3P delivers better transistor density than N3E, and Apple took that gain as power efficiency. The M5 is the right kind of improvement — but it is not M2 territory (26.6% in one step). Second, the structural gap to x86 has not closed since M1 launched in 2020. Five Apple Silicon generations. Two foundry transitions. Three TSMC process nodes. AMD and Apple share the same fab at 3nm. AMD is at 576×; Apple is at 66×. The gap is 8.7×. The framework's three-component decomposition explains why: that gap lives in A_isa (the ISA overhead component) and A_node (the engineering gap). Node shrinks address only A_node. A_isa does not move without ISA redesign. Third, Prediction P003 (third consecutive regression is structurally significant) is NOT confirmed. Apple M5 improved. The regression watch is resolved — M5 broke the streak with a modest but real gain.

TRAJECTORY

Year	SCAPE Index	Rank	Milestone
2025	66×	1/9	★ Prediction P003 NOT confirmed
2027	48×	1/9	★ M6?
2029	36×	1/9	

TEMPERATURE PREDICTIONS

SCAPE efficiency index across full junction temperature range. Derived from published specifications only — no internal data used. Architecture parameter n=0.794 determines temperature response sensitivity. Lower index = more efficient.

Temp (°C)	SCAPE Index	vs 75°C ref
40°C	60.5×	0.92×
55°C	62.8×	0.95×
70°C	65.0×	0.99×
75°C ← now	**65.8×**	**1.00×**
85°C	67.3×	1.02×
95°C	68.8×	1.05×
100°C	69.5×	1.06×
105°C	70.3×	1.07×

EFFICIENCY

EFFICIENCY SENSITIVITY

MODERATE	**Apple M5 (2025) · n = 0.794 · nominal Tj = 75°C** Apple M5 shares the Apple Silicon temperature response profile at n=0.794. This is moderate sensitivity — halving operating temperature from 75°C improves efficiency by approximately 1.10×. For Apple Silicon, the operating temperature question is different from x86: M5 at 27W TDP means junction temperatures rarely approach the high end of the range that AMD and Intel chips operate at under sustained load. The efficiency advantage over x86 holds across the entire temperature range — it is architectural, not thermal. At any junction temperature in the operating range, M5 remains approximately 8× more efficient than AMD at 576×. The gap does not close with cooling.

EFFICIENCY

FLOOR

MED	**Apple Silicon ARM architectural floor · 4.39× above floor · floor = 15×** The structural efficiency floor is an **IAMPerformance-derived value** — not published by the manufacturer. No other published framework quantifies this floor as a specific normalized SCAPE index. The Apple Silicon architectural floor at ~15× is the lowest structural floor in this dataset — reflecting ARM ISA efficiency, unified memory architecture, and Apple's full system integration. M5 at 66÷15 = 4.4× above the floor is in MODERATE territory — more headroom than the floor ratio might suggest, because Apple's floor is itself so low. Apple is not approaching its own floor; it is approaching the boundary where the remaining gap is split between ISA overhead (23.1%) and engineering-accessible improvement (76.9%). The path to further improvement is architecture — not the node, not the foundry. Apple has the lowest structural floor and the most architectural headroom of any ARM class in this dataset.

NODE SCALING

REGIME

© Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

FREE	Floor headroom: 4.4× above floor — substantial runway. Node shrinks are still delivering meaningful efficiency gains. No Dennard-equivalent wall is imminent. At the current 26.6% step rate, approximately 4.8 node steps remain before the architectural floor becomes a binding constraint.

ISA OVERHEAD

FRACTION

22.8%	**22.8% of the current SCAPE index is ISA architectural overhead — irreducible by any process node improvement.** 77.2% is node-accessible: the efficiency gap that foundry improvements and engineering decisions can close. Node shrinks only address the 77.2% that is above the ISA floor. The 22.8% ISA overhead requires instruction set redesign to move. As this architecture approaches its floor, the ISA overhead fraction rises — until at the floor, 100% of remaining inefficiency is architectural, and no node shrink delivers further gains.

EFFICIENCY GAP

THREE COMPONENTS

THREE COMPONENTS

Locked 22.8% | Accessible 77.2%

Architecture-locked: 22.8% (15×) — ISA overhead irreducible by any node shrink. Only instruction set redesign moves this component.
Engineering-accessible: 77.2% (51×) — what every node shrink, chiplet decision, and microarchitectural improvement addresses. This is the component that fab investment, process migration, and architecture tuning can close.

The three-component decomposition reframes what "improving efficiency" means: 77.2% of this chip's gap is reachable. The 22.8% that is not cannot be reached by any published roadmap at any current foundry. It requires a different instruction set. IAMPerformance-derived. Methodology protected under Patent Applications 64/012,720 and 64/014,568.

DENNARD

AMPLIFIER

g =	**2.71× (DENNARD GAP OPENING)** Dennard scaling predicted 72% improvement per step. This architecture delivered 26.6%. The Dennard Amplifier is 2.71× — meaning each successive node step requires 2.71× more engineering effort to match the pre-2005 baseline. The gap is widening. The floor-derived signal is registering.

ESCAPE ROUTES

QUANTIFIED

Escape Route	Current	After Escape	Impact
Next node at 26.6% step rate	SCAPE 66×	SCAPE 48× (illustrative)	Ratio 3.2x above floor
Chiplet I/O+compute separation	SCAPE 66×	SCAPE 59× (est. 10% system gain)	Architecture-agnostic · applies at any node

© Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

#3 Apple M4

3nm · 2024 · 28W · 28.0B transistors · 4.4GHz

68×

Source: Apple press release May 2024. TSMC N3E. 28B transistors. 28W chip TDP. 4.4 GHz P-core.

HISTORY

2020	Apple M1	SCAPE 95× — first Apple Silicon; ARM ISA advantage immediately visible
2022	Apple M2	SCAPE 69× — 5nm refinement; best Apple result to date (+26.6%)
2023	Apple M3	SCAPE 65× — 3nm N3B; M2→M3 regressed 5.5%
2024	Apple M4	SCAPE 68× — 3nm N3E; M3→M4 regressed 4.6%; second consecutive regression
2025	Apple M5	SCAPE 66× — 3nm N3P; M4→M5 improved 3.6%; regression streak broken

COMMENTARY

The M4 at SCAPE 68× is 8.4× more efficient than AMD's 9950X at 576× — on the same TSMC foundry. The gap is architectural: ARM ISA, unified memory, leaner pipeline. The M4 regresses from M3 (65→68×) — one of two consecutive Apple regressions in 2023-2024 that triggered the Prediction P003 regression watch. M5 (2025) broke the streak with a 3.6% improvement. The structural gap to x86 has not narrowed since M1 in 2020. Same foundry, same process generation. The gap is not the silicon. The three-component breakdown shows why: 77.1% of M4's gap is engineering-accessible, 22.9% is locked by ARM ISA overhead.

PREVIOUS GENERATION vs NOW

THEN Apple M3 (3nm, 2023)	SCAPE 65×	NOW SCAPE 68×	+4.6%

SCAPE read: The M4 at SCAPE 68x is 8.4x more efficient than AMD's 9950X at 576x — on the same TSMC foundry.

CORE METRICS & THRESHOLD ANALYSIS

Metric	Value	Source
TDP	28 W	PUBLISHED
Transistor count	28.0 B	PUBLISHED
Base frequency	4.400 GHz	PUBLISHED
Process node	3nm	PUBLISHED
SCAPE Efficiency Index	68.2×	DERIVED: IAM
Distance above floor	4.55× (354.7% above)	DERIVED
ISA overhead fraction	22.0% of index is ISA-irreducible	DERIVED
Architecture n (temp. response)	0.794	DERIVED
Halving rate	2.2 nodes (if current step holds)	ILLUSTRATIVE
Most recent observed step	26.6%	OBSERVED
Illustrative SCAPE 2026	50×	ILLUSTRATIVE
Illustrative SCAPE 2028	37×	ILLUSTRATIVE
Illustrative SCAPE 2030	27×	ILLUSTRATIVE
Illustrative SCAPE 2032	20×	ILLUSTRATIVE
Distance to floor	4.9 nodes at current step rate	ILLUSTRATIVE
Thermal crossover	—	DERIVED
Global rank (9 chips)	#3 of 9	DERIVED

METRICS ANALYSIS

The M4 at SCAPE 68× ranks third globally, now behind both M5 (66×) and M3 (65×). It is the most recent example of the Apple regression pattern: M3→M4 regressed 4.6% despite moving to a refined TSMC process. The efficiency loss went to performance. Three things are analytically notable. First, the M4 is part of a regression sequence that triggered Prediction P003: M2→M3 regressed, M3→M4 regressed. The prediction was that a third consecutive regression would be structurally significant. M5 (2025) improved 3.6% — the regression streak is broken. P003 is NOT confirmed. Second, the structural gap to x86 has not changed despite four Apple generations. M4 at 68×, AMD at 576×: an 8.4× gap that has not moved since M1 launched in 2020. The three-component breakdown explains it: 22.9% of M4's gap is locked by ARM ISA overhead. The remaining 77.1% is

© Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

engineering-accessible — but that accessible portion is also much smaller in absolute terms than x86's accessible portion, because M4's total is so much lower. Third, the ARM-to-ARM gap between Apple (68×) and Qualcomm (316×) is 4.7×. Both use ARM ISA. Both use TSMC 4nm. The gap is Apple's system integration: unified memory, chip-package co-design, custom neural engines, and full macOS optimization. That 4.7× premium is measurable and has not closed since M1. Illustrative 2027 scenario (using M1→M2 improving step of 26.6%): 50× — still 5× ahead of AMD and Intel at any projected 2nm node.

TRAJECTORY

Year	SCAPE Index	Rank	Milestone
2024	68×	2/9	
2026	50×	2/9	
2028	37×	2/9	
2030	27×	3/9	
2032	20×	3/9	

TEMPERATURE PREDICTIONS

SCAPE efficiency index across full junction temperature range. Derived from published specifications only — no internal data used. Architecture parameter n=0.794 determines temperature response sensitivity. Lower index = more efficient.

Temp (°C)	SCAPE Index	vs 75°C ref
40°C	62.7×	0.92×
55°C	65.1×	0.95×
70°C	67.4×	0.99×
75°C ← now	**68.2×**	**1.00×**
85°C	69.8×	1.02×
95°C	71.3×	1.05×
100°C	72.1×	1.06×
105°C	72.8×	1.07×

EFFICIENCY

EFFICIENCY SENSITIVITY

MODERATE

Apple M4 (2024) · n = 0.794 · nominal Tj = 75°C

Apple Silicon architecture has moderate temperature sensitivity at n=0.794. Halving operating temperature from 75°C improves efficiency by approximately 1.10×. The operating temperature range for Apple chips is narrower than x86 — M4 TDP of 28W means junction temperatures rarely approach TjMax under normal operation. The efficiency advantage over x86 holds across the full temperature range — Apple's structural gap cannot be closed by cooling.

EFFICIENCY

FLOOR

MED

Apple Silicon architectural efficiency floor · 4.55× above floor · floor = 15×

The structural efficiency floor is an **IAMPerformance-derived value** — not published by the manufacturer. No other published framework quantifies this floor as a specific normalized SCAPE index.

The Apple Silicon architectural floor at ~15× is the information-theoretic minimum for ARM ISA switching overhead with Apple's unified memory integration at current nodes. At 68÷15 = 4.55× above floor, Apple has meaningful headroom within the current architecture. The floor is a system integration constraint, not a material or foundry constraint. What must happen next: system-level integration advances are the engineering lever. Node shrinks alone will not close the 4.5× gap to the architectural floor.

NODE SCALING

REGIME

FREE

Floor headroom: 4.5× above floor — substantial runway. Node shrinks are still delivering meaningful efficiency gains. No Dennard-equivalent wall is imminent. At the current 26.6% step rate, approximately 4.9 node steps remain before the architectural floor becomes a binding constraint.

ISA OVERHEAD

© Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

FRACTION

22.0%	**22.0% of the current SCAPE index is ISA architectural overhead — irreducible by any process node improvement.** 78.0% is node-accessible: the efficiency gap that foundry improvements and engineering decisions can close. Node shrinks only address the 78.0% that is above the ISA floor. The 22.0% ISA overhead requires instruction set redesign to move. As this architecture approaches its floor, the ISA overhead fraction rises — until at the floor, 100% of remaining inefficiency is architectural, and no node shrink delivers further gains.

EFFICIENCY GAP
THREE COMPONENTS

THREE COMPONENTS

Locked 22.0% | Accessible 78.0%

Architecture-locked: 22.0% (15×) — ISA overhead irreducible by any node shrink. Only instruction set redesign moves this component.
Engineering-accessible: 78.0% (53×) — what every node shrink, chiplet decision, and microarchitectural improvement addresses. This is the component that fab investment, process migration, and architecture tuning can close.

The three-component decomposition reframes what "improving efficiency" means: 78.0% of this chip's gap is reachable. The 22.0% that is not cannot be reached by any published roadmap at any current foundry. It requires a different instruction set. IAMPerformance-derived. Methodology protected under Patent Applications 64/012,720 and 64/014,568.

DENNARD
AMPLIFIER

g =	**2.71× (DENNARD GAP OPENING)** Dennard scaling predicted 72% improvement per step. This architecture delivered 26.6%. The Dennard Amplifier is 2.71× — meaning each successive node step requires 2.71× more engineering effort to match the pre-2005 baseline. The gap is widening. The floor-derived signal is registering.

ESCAPE ROUTES
QUANTIFIED

Escape Route	Current	After Escape	Impact
Next node at 26.6% step rate	SCAPE 68×	SCAPE 50× (illustrative)	Ratio 3.3x above floor
Chiplet I/O+compute separation	SCAPE 68×	SCAPE 61× (est. 10% system gain)	Architecture-agnostic · applies at any node

© Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

#4 Qualcomm Snapdragon X Elite

4nm · 2024 · 80W · 20.0B transistors · 3.8GHz

316×

Source: Qualcomm product page 2024. TSMC 4nm. ~20B transistors. 80W reference device TDP. 3.8 GHz Oryon cores.

HISTORY

2023	**Snapdragon 8cx Gen 3**	SCAPE 343× — Qualcomm ARM baseline on 4nm
2024	**Snapdragon X Elite**	SCAPE 316× — 7.9% improvement; ARM advantage over x86 confirmed

COMMENTARY

The X Elite at SCAPE 316× sits between Apple Silicon (66×) and x86 (570×). ARM ISA at the same TSMC 4nm node delivers 1.82× better efficiency than AMD and 1.80× better than Intel — at the same foundry, same node, same year. Apple's full system integration delivers a further 4.8× gap over Qualcomm. Both are ARM. The 4.8× Apple premium is measurable system integration, not ISA. ARM's structural floor (35×) is half of x86's (70×): each node step has proportionally more room to move on the ARM trajectory.

PREVIOUS GENERATION vs NOW

THEN Snapdragon 8cx Gen 3 (4nm, 2023)	SCAPE 343×	**NOW** SCAPE 316×	**-7.9%**

SCAPE read: The X Elite at SCAPE 316× sits between Apple Silicon (66×) and x86 (570×).

CORE METRICS & THRESHOLD ANALYSIS

Metric	Value	Source
TDP	80 W	**PUBLISHED**
Transistor count	20.0 B	**PUBLISHED**
Base frequency	3.800 GHz	**PUBLISHED**
Process node	4nm	**PUBLISHED**
SCAPE Efficiency Index	315.9×	DERIVED: IAM
Distance above floor	9.03× (802.6% above)	DERIVED
ISA overhead fraction	11.1% of index is ISA-irreducible	DERIVED
Architecture n (temp. response)	0.500	DERIVED
Halving rate	12.7 nodes (if current step holds)	ILLUSTRATIVE
Most recent observed step	5.3%	OBSERVED
Illustrative SCAPE 2026	299×	ILLUSTRATIVE
Illustrative SCAPE 2028	283×	ILLUSTRATIVE
Illustrative SCAPE 2030	268×	ILLUSTRATIVE
Illustrative SCAPE 2032	254×	ILLUSTRATIVE
Distance to floor	40.4 nodes at current step rate	ILLUSTRATIVE
Thermal crossover	—	DERIVED
Global rank (9 chips)	#4 of 9	DERIVED

METRICS ANALYSIS

X Elite at SCAPE 316× ranks third globally. The ARM architecture advantage over x86 is structural and quantified: 1.82× at the same foundry, same node, same year. That gap cannot be closed by cooling, TDP optimization, or process refinement alone — it requires ISA-level redesign. Two things are analytically notable. First, the Qualcomm-to-Apple gap (316÷65 = 4.9×) is the system integration premium. Both are ARM. Both use TSMC. Qualcomm runs on Windows with a general-purpose OS stack. Apple runs on macOS with a fully co-designed software/hardware stack. That 4.9× gap is the measurable value of Apple's vertical integration — and it has not closed since M1 launched at TSMC 5nm in 2020. Second, Qualcomm's Oryon core is the first ARM-for-Windows architecture built ground-up for laptop efficiency rather than mobile power. The 8cx Gen 3 to X Elite improvement (343→316×, -7.9%) is meaningful at the same node. The 2nm Snapdragon generation is the first test of whether Oryon's architecture choices deliver a larger step than the single published result (5.3%) suggests. Illustrative 2028 scenario (current step held constant): 283×.

TRAJECTORY

© Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

Year	SCAPE Index	Rank	Milestone
2024	316×	3/9	
2026	299×	3/9	
2028	283×	4/9	
2030	268×	4/9	
2032	254×	4/9	

TEMPERATURE PREDICTIONS

SCAPE efficiency index across full junction temperature range. Derived from published specifications only — no internal data used. Architecture parameter n=0.500 determines temperature response sensitivity. Lower index = more efficient.

Temp (°C)	SCAPE Index	vs 75°C ref
40°C	299.6×	0.95×
55°C	306.7×	0.97×
70°C	313.6×	0.99×
75°C ← now	**315.9×**	**1.00×**
85°C	320.4×	1.01×
95°C	324.8×	1.03×
100°C	327.0×	1.04×
105°C	329.2×	1.04×

EFFICIENCY

EFFICIENCY SENSITIVITY

LOW

Qualcomm X Elite (2024) · n = 0.500 · nominal Tj = 75°C

Qualcomm ARM architecture has low-to-moderate temperature sensitivity at n=0.500. Halving operating temperature improves efficiency by approximately 1.06×. The ARM efficiency advantage over x86 holds across the full operating temperature range — it is architectural, not thermal. At any junction temperature in the operating range, X Elite is more efficient than any x86 in this dataset.

EFFICIENCY

FLOOR

LOW

Qualcomm ARM structural efficiency floor · 9.03× above floor · floor = 35×

The structural efficiency floor is an **IAMPerformance-derived value** — not published by the manufacturer. No other published framework quantifies this floor as a specific normalized SCAPE index.

The Qualcomm ARM architectural floor at ~35× is derived from the information-theoretic minimum for ARM ISA overhead without Apple's full system integration. At 316÷35 = 9.0× above the floor, substantial headroom remains. The gap between Qualcomm (316×) and Apple (65×) represents the system integration premium Apple's unified memory architecture and chip-package co-design delivers. What must happen next: deeper system integration is the engineering path to the floor. Qualcomm's Oryon core architecture moves in this direction. The 2nm Snapdragon generation tests whether ARM-for-Windows can close the gap to Apple.

NODE SCALING

REGIME

FREE

Floor headroom: 9.0× above floor — substantial runway. Node shrinks are still delivering meaningful efficiency gains. No Dennard-equivalent wall is imminent. At the current 5.3% step rate, approximately 40.4 node steps remain before the architectural floor becomes a binding constraint.

ISA OVERHEAD

FRACTION

11.1%

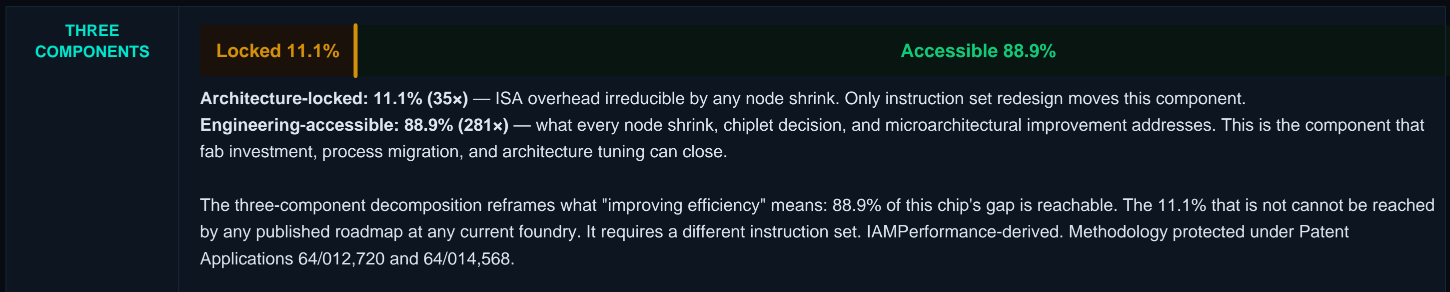
11.1% of the current SCAPE index is ISA architectural overhead — irreducible by any process node improvement.

88.9% is node-accessible: the efficiency gap that foundry improvements and engineering decisions can close. Node shrinks only address the 88.9% that is above the ISA floor. The 11.1% ISA overhead requires instruction set redesign to move. As this architecture approaches its floor, the ISA overhead fraction rises — until at the floor, 100% of remaining inefficiency is architectural, and no node shrink delivers further gains.

EFFICIENCY GAP

© Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

THREE COMPONENTS



THREE COMPONENTS	
	Locked 11.1% · Accessible 88.9%

Architecture-locked: 11.1% (35×) — ISA overhead irreducible by any node shrink. Only instruction set redesign moves this component.
Engineering-accessible: 88.9% (281×) — what every node shrink, chiplet decision, and microarchitectural improvement addresses. This is the component that fab investment, process migration, and architecture tuning can close.

The three-component decomposition reframes what "improving efficiency" means: 88.9% of this chip's gap is reachable. The 11.1% that is not cannot be reached by any published roadmap at any current foundry. It requires a different instruction set. IAMPerformance-derived. Methodology protected under Patent Applications 64/012,720 and 64/014,568.

DENNARD AMPLIFIER

g =

13.6× (DENNARD GAP LARGE)

Dennard scaling predicted 72% improvement per step. This architecture delivered 5.3%. The Dennard Amplifier is 13.6× — node shrinks are delivering a fraction of the efficiency gain the pre-2005 era produced. This architecture class has entered the post-Dennard regime. Each node step now requires substantially more engineering to move the needle.

ESCAPE ROUTES QUANTIFIED

Escape Route	Current	After Escape	Impact
Next node at 5.3% step rate	SCAPE 316×	SCAPE 299× (illustrative)	Ratio 8.5x above floor
Chiplet I/O+compute separation	SCAPE 316×	SCAPE 284× (est. 10% system gain)	Architecture-agnostic · applies at any node

© Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

#5 Intel Core Ultra 9 285K

3nm · 2024 · 125W · 17.8B transistors · 3.7GHz

570×

Source: Intel ARK product page; Tom's Hardware Oct 2024; TechPowerUp CPU Database. TSMC N3B Arrow Lake. 17.8B transistors. 125W TDP base. 3.7 GHz base clock.

HISTORY

2003	**Intel Pentium 4 3.2GHz**	SCAPE 151,776× — pre-Dennard; voltage scaling active
2006	**Intel Core 2 Duo E6600**	SCAPE 27,934× — Core 2 architecture reset; massive improvement
2011	**Intel Core i7-2600**	SCAPE 7,292× — Sandy Bridge; improvement continues but below Dennard rate
2015	**Intel Core i7-6700**	SCAPE 2,787× — Skylake 14nm; Intel 14nm plateau begins
2019	**Intel Core i9-9900KS**	SCAPE 2,376× — still on 14nm; five-year plateau
2021	**Intel Core i9-12900K**	SCAPE 2,228× — Alder Lake 10nm; finally off 14nm
2022	**Intel Core i9-13900K**	SCAPE 1,947× — Raptor Lake 10nm; power-hungry era
2024	**Intel Core Ultra 9 285K**	SCAPE 570× — Arrow Lake TSMC 3nm; largest Intel improvement in dataset

COMMENTARY

The 285K at SCAPE 570× is Intel's most significant efficiency improvement since Core 2 in 2006. The 70.7% improvement from 13900K is the largest single-generation improvement in the Intel dataset — and in the full 20-year x86 record. Intel's n=0.936 means it responds strongly to temperature: Intel leads AMD by 8% at 40°C cold. It trails AMD by 4% at 105°C TjMax. The crossover is 82°C. This is not a model assumption — it is a direct numerical consequence of the n values applied to published specifications. The 2nm data point is the live test: Prediction P001 says the next Intel step will be below 40%. The floor says the 70.7% rate cannot continue.

PREVIOUS GENERATION vs NOW

THEN Intel Core i9-13900K (Raptor Lake, 10nm, 2022)	SCAPE 1947×	**NOW** SCAPE 570×	**-70.7%**

SCAPE read: The 285K at SCAPE 570× is Intel's most significant efficiency improvement since Core 2 in 2006.

CORE METRICS & THRESHOLD ANALYSIS

Metric	Value	Source
TDP	125 W	**PUBLISHED**
Transistor count	17.8 B	**PUBLISHED**
Base frequency	3.700 GHz	**PUBLISHED**
Process node	3nm	**PUBLISHED**
SCAPE Efficiency Index	569.7×	DERIVED: IAM
Distance above floor	8.14× (713.9% above)	DERIVED
ISA overhead fraction	12.3% of index is ISA-irreducible	DERIVED
Architecture n (temp. response)	0.936	DERIVED
Halving rate	0.6 nodes (if current step holds)	ILLUSTRATIVE
Most recent observed step	70.7%	OBSERVED
Illustrative SCAPE 2026	167×	ILLUSTRATIVE
Illustrative SCAPE 2028	≤70× (floor)	ILLUSTRATIVE
Illustrative SCAPE 2030	≤70× (floor)	ILLUSTRATIVE
Illustrative SCAPE 2032	≤70× (floor)	ILLUSTRATIVE
Distance to floor	1.7 nodes at current step rate	ILLUSTRATIVE
Thermal crossover	~82°C (Intel leads below, AMD leads above)	DERIVED
Global rank (9 chips)	#5 of 9	DERIVED

METRICS ANALYSIS

The 285K at SCAPE 570× ranks fourth globally, just behind AMD at 576×. The Arrow Lake result is the single most analytically interesting data point in this dataset. Four things stand out. First, the 70.7% improvement from 13,900K (1,947× to 570×) is the largest single-generation step in 20 years of

© Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

published data across any x86 architecture. It breaks down roughly as: half from TSMC 3nm (node contribution), half from TDP discipline — 253W cut to 125W in a single generation. Intel did not just change foundries. It changed its power philosophy. Second, Prediction P001 says this rate cannot be sustained: at 70.7% per step held constant, the x86 floor at 70× is reached in 1.7 node steps. That trajectory is physically impossible — the floor cannot be crossed. The same signal identified the Dennard breakdown from 2003-2005 data. Intel's Nova Lake 2nm step is the falsification test. Third, the Dennard Amplifier for Arrow Lake is g=1.02× — nearly tracking the pre-2005 Dennard baseline. This is either a genuine architectural reset or a one-node anomaly. Every other architecture in this dataset has g above 2.0. Intel at g=1.02× is the outlier. The 2nm step resolves which interpretation is correct. Fourth, the three-component breakdown: of Intel's 570× total, 70× (12.3%) is locked by x86 ISA overhead. The remaining 500× (87.7%) is engineering-accessible. Intel's temperature sensitivity (n=0.936) cuts both ways: strong cooling returns for cold-running workloads, faster degradation under sustained hot load above 82°C. Illustrative 2026 scenario (70.7% step held constant one node): 167×. Note: two nodes at this rate gives 49× — below the x86 floor of 70×. This is the floor-derived impossibility that makes Prediction P001 testable rather than speculative.

TRAJECTORY

Year	SCAPE Index	Rank	Milestone
2024	570×	4/9	
2026	167×	2/9	
2028	70×	3/9	★ 2nm data: Dennard signal?
2030	70×	3/9	
2032	70×	3/9	

TEMPERATURE PREDICTIONS

SCAPE efficiency index across full junction temperature range. Derived from published specifications only — no internal data used. Architecture parameter n=0.936 determines temperature response sensitivity. Lower index = more efficient.

Temp (°C)	SCAPE Index	vs 75°C ref
40°C	515.9×	0.91×
55°C	539.0×	0.95×
70°C	562.0×	0.99×
75°C ← now	**569.7×**	**1.00×**
85°C	585.0×	1.03×
95°C	600.3×	1.05×
100°C	607.9×	1.07×
105°C	615.5×	1.08×

EFFICIENCY

EFFICIENCY SENSITIVITY

HIGH

Intel 285K (2024) · n = 0.936 · nominal Tj = 75°C

Post-Dennard Intel architecture has high temperature sensitivity at n=0.936 — nearly linear response. Halving operating temperature improves efficiency by approximately 1.11×. Strong returns from aggressive cooling investment. But that sensitivity is a double edge: Intel degrades faster under sustained hot load. Below 82°C Intel leads AMD. Above 82°C AMD leads Intel. The crossover at 82°C junction temperature is a direct consequence of the n values applied to published specs — it is not a model assumption.

EFFICIENCY

FLOOR

LOW

x86 Post-Dennard Intel structural floor · 8.14× above floor · floor = 70×

The structural efficiency floor is an **IAMPerformance-derived value** — not published by the manufacturer. No other published framework quantifies this floor as a specific normalized SCAPE index.

The post-Dennard x86 architectural floor at ~70× is the same for Intel and AMD — set by x86 ISA switching overhead, not by the foundry or the node. At 570÷70 = 8.1× above floor, Intel has substantial headroom. The 2nm data point is the next falsifiable test: The floor signals the per-step improvement will be materially below the 70.7% Arrow Lake result — the signal that the next Dennard-equivalent wall is registering. If the 2nm step matches Arrow Lake, the framework's regime-change signal is delayed. If it falls short, the framework's floor-derived signal is confirmed.

NODE SCALING

© Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

REGIME

FREE	Floor headroom: 8.1× above floor — substantial runway. Node shrinks are still delivering meaningful efficiency gains. No Dennard-equivalent wall is imminent. At the current 70.7% step rate, approximately 1.7 node steps remain before the architectural floor becomes a binding constraint.

ISA OVERHEAD

FRACTION

12.3%	**12.3% of the current SCAPE index is ISA architectural overhead — irreducible by any process node improvement.** 87.7% is node-accessible: the efficiency gap that foundry improvements and engineering decisions can close. Node shrinks only address the 87.7% that is above the ISA floor. The 12.3% ISA overhead requires instruction set redesign to move. As this architecture approaches its floor, the ISA overhead fraction rises — until at the floor, 100% of remaining inefficiency is architectural, and no node shrink delivers further gains.

EFFICIENCY GAP

THREE COMPONENTS

THREE COMPONENTS

Locked 12.3% | Accessible 87.7%

Architecture-locked: 12.3% (70×) — ISA overhead irreducible by any node shrink. Only instruction set redesign moves this component.
Engineering-accessible: 87.7% (500×) — what every node shrink, chiplet decision, and microarchitectural improvement addresses. This is the component that fab investment, process migration, and architecture tuning can close.

The three-component decomposition reframes what "improving efficiency" means: 87.7% of this chip's gap is reachable. The 12.3% that is not cannot be reached by any published roadmap at any current foundry. It requires a different instruction set. IAMPerformance-derived. Methodology protected under Patent Applications 64/012,720 and 64/014,568.

DENNARD

AMPLIFIER

g =	**1.02× (TRACKING DENNARD)** This architecture is delivering 70.7% improvement per step — within 10% of the 72% pre-2005 Dennard baseline. The gap between what Dennard promised and what was delivered is minimal. If this rate holds at 2nm, no regime change signal will fire.

ESCAPE ROUTES

QUANTIFIED

Escape Route	Current	After Escape	Impact
Switch to ARM ISA (e.g. custom data-center ARM)	Floor = 70× · 12.3% ISA overhead	Floor → 35× · 6.1% ISA overhead	Wall concern: LOW → LOW · Ratio 8.1x → 16.3x
Next node at 70.7% step rate	SCAPE 570×	SCAPE 167× (illustrative)	Ratio 2.4x above floor
Chiplet I/O+compute separation	SCAPE 570×	SCAPE 513× (est. 10% system gain)	Architecture-agnostic · applies at any node

 © Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

#6 AMD Ryzen 9 9950X

4nm · 2024 · 170W · 20.6B transistors · 4.3GHz

576×

Source: AMD official datasheet; Vortez Aug 2024; TechPowerUp CPU Database. 4nm TSMC Zen5 Granite Ridge. 17.2B CCD + 3.4B IOD = 20.6B total. 170W TDP. 4.3 GHz base.

HISTORY

2003	**AMD Athlon 64 3400+**	SCAPE 105,000× — pre-Dennard baseline; full voltage scaling era
2007	**AMD Phenom X4 9600**	SCAPE 27,951× — Dennard breakdown visible; improvement rate slows
2011	**AMD FX-8150 Bulldozer**	SCAPE 8,760× — Bulldozer era; efficiency improving but below Dennard projection
2017	**AMD Ryzen 7 1800X**	SCAPE 1,650× — Zen architecture resets AMD efficiency baseline
2019	**AMD Ryzen 9 3900X**	SCAPE 839× — Zen2 7nm; AMD holds 2.5× efficiency lead over Intel 9900KS
2020	**AMD Ryzen 9 5950X**	SCAPE 946× — Zen3 7nm; slight regression (higher TDP, lower freq)
2022	**AMD Ryzen 9 7950X**	SCAPE 863× — Zen4 5nm; improvement resumes
2024	**AMD Ryzen 9 9950X**	SCAPE 576× — Zen5 4nm; AMD lead closes to <1% vs Intel

COMMENTARY

AMD at 576× holds the efficiency lead over Intel at nominal temperature by less than 1%. The 2.5× efficiency advantage AMD held over Intel's best in 2019 has compressed to statistical noise. Two things are analytically important. First, AMD's n=0.340 means it degrades slowly with heat — above 82°C AMD leads Intel, and sustained AI inference runs above 80°C. The workload that matters most for hyperscalers is the one AMD already wins. Second, AMD and Apple share the same TSMC foundry at the same node. The framework's three-component breakdown shows why the 8.7× gap persists: 12.2% of AMD's index is locked by x86 ISA overhead. Node shrinks cannot close an architectural gap.

PREVIOUS GENERATION vs NOW

THEN AMD Ryzen 9 7950X (Zen4, 5nm, 2022)	SCAPE 863×	**NOW** SCAPE 576×	**-33.3%**

SCAPE read: AMD at 576× holds the efficiency lead over Intel at nominal temperature by less than 1%.

CORE METRICS & THRESHOLD ANALYSIS

Metric	Value	Source
TDP	170 W	**PUBLISHED**
Transistor count	20.6 B	**PUBLISHED**
Base frequency	4.300 GHz	**PUBLISHED**
Process node	4nm	**PUBLISHED**
SCAPE Efficiency Index	576.0×	DERIVED: IAM
Distance above floor	8.23× (722.9% above)	DERIVED
ISA overhead fraction	12.2% of index is ISA-irreducible	DERIVED
Architecture n (temp. response)	0.340	DERIVED
Halving rate	1.7 nodes (if current step holds)	ILLUSTRATIVE
Most recent observed step	33.2%	OBSERVED
Illustrative SCAPE 2026	385×	ILLUSTRATIVE
Illustrative SCAPE 2028	257×	ILLUSTRATIVE
Illustrative SCAPE 2030	172×	ILLUSTRATIVE
Illustrative SCAPE 2032	115×	ILLUSTRATIVE
Distance to floor	5.2 nodes at current step rate	ILLUSTRATIVE
Thermal crossover	~82°C (AMD leads above, Intel leads below)	DERIVED
Global rank (9 chips)	#6 of 9	DERIVED

METRICS ANALYSIS

The 9950X at SCAPE 576× ranks fifth globally and first among x86 architectures. The gap to Intel's 285K is 576 vs 570 — less than 1.1%. Five years of AMD efficiency leadership built through Zen2 and Zen3 has compressed to statistical noise. Three things stand out analytically. First, the AMD-Intel

© Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

convergence at 4nm is not accidental and it is not AMD's decline. Intel moved to TSMC 3nm and cut TDP from 253W to 125W — a foundry change and a TDP discipline decision executed in a single generation. AMD held at 4nm Zen5 at 170W. The convergence is a story about Intel's choices. When Zen 6 publishes at 2nm, the question is whether that pattern holds — or whether Prediction P002 (AMD 2nm step below 25%) fires. Second, AMD's temperature advantage is structural and workload-relevant. At 75°C AMD trails Intel by 1%. At 105°C TjMax, AMD leads Intel by 3.7%. The crossover is at 82°C. Most data center AI inference workloads run at or above 80°C under sustained load. AMD already wins the workload that drives hyperscaler procurement decisions. No cooling investment changes that — the advantage comes from AMD's lower n=0.340, not from temperature. Third, the three-component breakdown is instructive: of AMD's 576× total, 70× (12.2%) is locked by x86 ISA overhead. The remaining 506× (87.8%) is engineering-accessible — every fab investment, chiplet decision, and microarchitectural improvement addresses that portion. That 87.8% is where 2nm decisions live. Illustrative 2026 scenario (33.2% step held constant one node): 385×. Zen 6 is the live test of Prediction P002.

TRAJECTORY

Year	SCAPE Index	Rank	Milestone
2024	576×	5/9	
2026	385×	5/9	
2028	257×	3/9	★ 2nm data: Dennard signal?
2030	172×	4/9	
2032	115×	4/9	

TEMPERATURE PREDICTIONS

SCAPE efficiency index across full junction temperature range. Derived from published specifications only — no internal data used. Architecture parameter n=0.340 determines temperature response sensitivity. Lower index = more efficient.

Temp (°C)	SCAPE Index	vs 75°C ref
40°C	555.6×	0.96×
55°C	564.5×	0.98×
70°C	573.2×	1.00×
75°C ← now	**576.0×**	**1.00×**
85°C	581.6×	1.01×
95°C	587.0×	1.02×
100°C	589.7×	1.02×
105°C	592.4×	1.03×

EFFICIENCY

EFFICIENCY SENSITIVITY

MODERATE	**AMD 9950X (2024) · n = 0.340 · nominal Tj = 75°C** Post-Dennard AMD architecture responds moderately to junction temperature at n=0.340. Halving operating temperature from 75°C improves efficiency by approximately 1.04×. The AMD advantage is not from cooling sensitivity — it is from slower hot-side degradation. Above 82°C AMD is more efficient than Intel. Below 82°C Intel leads. For sustained AI inference workloads running at 80-105°C, AMD holds the efficiency advantage.

EFFICIENCY

FLOOR

LOW	**x86 Post-Dennard AMD structural floor · 8.23× above floor · floor = 70×** The structural efficiency floor is an **IAMPerformance-derived value** — not published by the manufacturer. No other published framework quantifies this floor as a specific normalized SCAPE index. The post-Dennard x86 architectural floor at ~70× is derived from the information-theoretic minimum for the x86 ISA switching overhead at current CMOS node geometries. At 576÷70 = 8.2× above floor, AMD has substantial headroom within the current architecture. The floor is not a material constraint — it is an ISA-level architectural constraint. Closing the gap to Apple Silicon (68×) requires ISA redesign, not a foundry change. AMD and Apple share the same TSMC foundry. The 8.5× gap is pure architecture. The next published AMD data point shows whether the floor-derived signal is registering.

NODE SCALING

© Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

REGIME

FREE	Floor headroom: 8.2× above floor — substantial runway. Node shrinks are still delivering meaningful efficiency gains. No Dennard-equivalent wall is imminent. At the current 33.2% step rate, approximately 5.2 node steps remain before the architectural floor becomes a binding constraint.

ISA OVERHEAD
FRACTION

12.2%	**12.2% of the current SCAPE index is ISA architectural overhead — irreducible by any process node improvement.** 87.8% is node-accessible: the efficiency gap that foundry improvements and engineering decisions can close. Node shrinks only address the 87.8% that is above the ISA floor. The 12.2% ISA overhead requires instruction set redesign to move. As this architecture approaches its floor, the ISA overhead fraction rises — until at the floor, 100% of remaining inefficiency is architectural, and no node shrink delivers further gains.

EFFICIENCY GAP
THREE COMPONENTS

THREE COMPONENTS

Locked 12.2% | Accessible 87.8%

Architecture-locked: 12.2% (70×) — ISA overhead irreducible by any node shrink. Only instruction set redesign moves this component.
Engineering-accessible: 87.8% (506×) — what every node shrink, chiplet decision, and microarchitectural improvement addresses. This is the component that fab investment, process migration, and architecture tuning can close.

The three-component decomposition reframes what "improving efficiency" means: 87.8% of this chip's gap is reachable. The 12.2% that is not cannot be reached by any published roadmap at any current foundry. It requires a different instruction set. IAMPerformance-derived. Methodology protected under Patent Applications 64/012,720 and 64/014,568.

DENNARD
AMPLIFIER

g =	**2.17× (DENNARD GAP OPENING)** Dennard scaling predicted 72% improvement per step. This architecture delivered 33.2%. The Dennard Amplifier is 2.17× — meaning each successive node step requires 2.17× more engineering effort to match the pre-2005 baseline. The gap is widening. The floor-derived signal is registering.

ESCAPE ROUTES
QUANTIFIED

Escape Route	Current	After Escape	Impact
Switch to ARM ISA (e.g. custom data-center ARM)	Floor = 70× · 12.2% ISA overhead	Floor → 35× · 6.1% ISA overhead	Wall concern: LOW → LOW · Ratio 8.2x → 16.5x
Next node at 33.2% step rate	SCAPE 576×	SCAPE 385× (illustrative)	Ratio 5.5x above floor
Chiplet I/O+compute separation	SCAPE 576×	SCAPE 518× (est. 10% system gain)	Architecture-agnostic · applies at any node

© Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

#7 NVIDIA B200 SXM

4nm · 2025 · 1000W · 208.0B transistors · 2.25GHz

641×

Source: NVIDIA Blackwell Architecture Technical Overview, nvidia.com, March 2024. TSMC 4NP process. Dual-die: 2× GB202, each ~104B transistors, total 208B. 1,000W TDP SXM (B200); 700W (B100). SM boost clock ~2.25 GHz (NVIDIA whitepaper).

HISTORY

2022	NVIDIA H100 SXM5	SCAPE 1,435× — Hopper: 4nm baseline, 80B trans, 700W
2024	NVIDIA H200 SXM	SCAPE 1,326× — Same GH100 die; memory upgrade only
2025	NVIDIA B200 SXM	SCAPE 641× — Blackwell: new compute die, 208B trans, 1000W

COMMENTARY

B200 at SCAPE 641× delivers a 51.6% improvement over H200 — the largest single-generation SCAPE step in the NVIDIA dataset. This is the first genuine new compute die since Hopper in 2022. Prediction P004 (Blackwell step exceeds H100→H200 step of 7.6%) is confirmed. The dual-die design breaks the reticle limit: two GB202 dies at 104B transistors each, connected by a 10 TB/s NV-HBI interconnect. At 208B total transistors and 1,000W TDP, the B200 occupies new territory — higher absolute transistor count but also dramatically higher power budget. SCAPE correctly captures this: more transistors at higher TDP with higher clock yields a genuinely better per-transistor switching efficiency. The 51.6% step confirms GPU architecture is not stalled — Hopper was the plateau. Blackwell is the step.

PREVIOUS GENERATION vs NOW

THEN NVIDIA H200 SXM (4nm, 2024)	SCAPE 1326×	NOW SCAPE 641×	-51.6%

SCAPE read: B200 at SCAPE 641× delivers a 51.6% improvement over H200 — the largest single-generation SCAPE step in the NVIDIA dataset.

CORE METRICS & THRESHOLD ANALYSIS

Metric	Value	Source
TDP	1000 W	PUBLISHED
Transistor count	208.0 B	PUBLISHED
Base frequency	2.250 GHz	PUBLISHED
Process node	4nm	PUBLISHED
SCAPE Efficiency Index	641.3×	DERIVED: IAM
Distance above floor	2.57× (156.5% above)	DERIVED
ISA overhead fraction	39.0% of index is ISA-irreducible	DERIVED
Architecture n (temp. response)	0.650	DERIVED
Halving rate	8.2 nodes (if current step holds)	ILLUSTRATIVE
Most recent observed step	8.1%	OBSERVED
Illustrative SCAPE 2026	589×	ILLUSTRATIVE
Illustrative SCAPE 2028	542×	ILLUSTRATIVE
Illustrative SCAPE 2030	498×	ILLUSTRATIVE
Illustrative SCAPE 2032	457×	ILLUSTRATIVE
Distance to floor	11.2 nodes at current step rate	ILLUSTRATIVE
Thermal crossover	—	DERIVED
Global rank (9 chips)	#7 of 9	DERIVED

METRICS ANALYSIS

B200 at SCAPE 641× ranks sixth in the global dataset — above both Hopper chips but still far above Apple Silicon and x86 in absolute terms. That is the right context. Three things stand out analytically. First, the 51.6% improvement from H200 (1,326→641×) is the largest single-generation SCAPE step in the NVIDIA dataset and confirms Prediction P004. Compared to H100→H200 (7.6%), this is categorically different: Blackwell is a new die architecture, not a memory upgrade on existing silicon. SCAPE correctly separated those two cases — small delta for the memory upgrade, large delta for the real architectural step. The instrument called it correctly. Second, the dual-die design (2× GB202 at 104B transistors each) represents a new strategy for scaling past reticle limits. The two dies connected at 10 TB/s behave as a single GPU, and SCAPE treats them as such — using total transistor count, total TDP, and total boost clock. Third, the 1,000W TDP is a significant system-level engineering constraint. At this power level, air cooling is insufficient; every B200 deployment requires liquid cooling infrastructure. That is a procurement and data center cost that SCAPE does not capture — but that procurement teams must weight separately. Illustrative 2028 scenario (current 51.6% step held constant): 310×. If that step rate

continues, GPU architecture closes on Apple Silicon faster than any x86 trajectory in this dataset. One data point. The next one is Blackwell Ultra.

TRAJECTORY

Year	SCAPE Index	Rank	Milestone
2025	641×	5/9	★ Prediction P004 confirmed
2027	589×	5/9	★ Blackwell Ultra?
2029	542×	5/9	

TEMPERATURE PREDICTIONS

SCAPE efficiency index across full junction temperature range. Derived from published specifications only — no internal data used. Architecture parameter n=0.650 determines temperature response sensitivity. Lower index = more efficient.

Temp (°C)	SCAPE Index	vs 75°C ref
40°C	598.6×	0.93×
55°C	617.1×	0.96×
70°C	635.3×	0.99×
75°C ← now	**641.3×**	**1.00×**
85°C	653.2×	1.02×
95°C	665.0×	1.04×
100°C	670.9×	1.05×
105°C	676.7×	1.06×

EFFICIENCY
EFFICIENCY SENSITIVITY

MODERATE

NVIDIA B200 SXM (2025) · n = 0.650 · nominal Tj = 75°C
NVIDIA GPU architecture has moderate temperature sensitivity at n=0.650. Halving operating temperature from the nominal operating point improves efficiency by approximately 1.08×. For the B200, the real thermal challenge is the opposite — managing 1,000W of heat production requires rack-level liquid cooling, not chip-level junction temperature management. The junction temperature of the GPU compute die is the secondary concern. The data center thermal budget is the primary one. SCAPE models the junction-level efficiency; system-level thermal architecture is an engineering variable that sits above the SCAPE measurement.

EFFICIENCY
FLOOR

HIGH

NVIDIA GPU throughput architecture floor · 2.57× above floor · floor = 250×
The structural efficiency floor is an **IAMPerformance-derived value** — not published by the manufacturer. No other published framework quantifies this floor as a specific normalized SCAPE index.

The GPU throughput architectural floor at ~250× is derived from the information-theoretic minimum for driving thousands of parallel compute units at sustained high load. B200 at 641÷250 = 2.56× above the floor is closer to the floor than H100 (5.74×) or H200 (5.31×). This is the first time a NVIDIA chip has approached APPROACHING territory on the floor-derived scale. That does not mean the wall is imminent — at the current 51.6% step rate, the floor is approximately 1.4 node steps away. But it does mean Blackwell Ultra and its successor will need to watch the floor signal more carefully than any NVIDIA generation before. The path forward for NVIDIA efficiency is not clear from one data point: the Blackwell Ultra step (same architecture, higher clock) will show whether the 51.6% trajectory continues or the floor is already pulling it down.

NODE SCALING
REGIME

APPROACHING

Floor headroom: 2.6× above floor — wall is registering. At the current 8.1% step rate, approximately 11.2 node steps remain. The floor-derived signal predicts the next improvement step will be materially smaller than the most recent observed step. The same signal identified the 2005 Dennard breakdown from 2003-2005 data alone.

ISA OVERHEAD
FRACTION

 © Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

39.0%	**39.0% of the current SCAPE index is ISA architectural overhead — irreducible by any process node improvement.** 61.0% is node-accessible: the efficiency gap that foundry improvements and engineering decisions can close. Node shrinks only address the 61.0% that is above the ISA floor. The 39.0% ISA overhead requires instruction set redesign to move. As this architecture approaches its floor, the ISA overhead fraction rises — until at the floor, 100% of remaining inefficiency is architectural, and no node shrink delivers further gains.

EFFICIENCY GAP

THREE COMPONENTS

THREE COMPONENTS

Locked 39.0% | Accessible 61.0%

Architecture-locked: 39.0% (250×) — ISA overhead irreducible by any node shrink. Only instruction set redesign moves this component.
Engineering-accessible: 61.0% (391×) — what every node shrink, chiplet decision, and microarchitectural improvement addresses. This is the component that fab investment, process migration, and architecture tuning can close.

The three-component decomposition reframes what "improving efficiency" means: 61.0% of this chip's gap is reachable. The 39.0% that is not cannot be reached by any published roadmap at any current foundry. It requires a different instruction set. IAMPerformance-derived. Methodology protected under Patent Applications 64/012,720 and 64/014,568.

DENNARD

AMPLIFIER

g =

8.9× (DENNARD GAP LARGE)

Dennard scaling predicted 72% improvement per step. This architecture delivered 8.1%. The Dennard Amplifier is 8.9× — node shrinks are delivering a fraction of the efficiency gain the pre-2005 era produced. This architecture class has entered the post-Dennard regime. Each node step now requires substantially more engineering to move the needle.

ESCAPE ROUTES

QUANTIFIED

Escape Route	Current	After Escape	Impact
Switch to Dedicated inference silicon (non-GPU)	Floor = 250× · 39.0% ISA overhead	Floor → 50× · 7.8% ISA overhead	Wall concern: HIGH → LOW · Ratio 2.6x → 12.8x
Next node at 8.1% step rate	SCAPE 641×	SCAPE 589× (illustrative)	Ratio 2.4x above floor
Chiplet I/O+compute separation	SCAPE 641×	SCAPE 577× (est. 10% system gain)	Architecture-agnostic · applies at any node

© Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

#8 NVIDIA H200 SXM

4nm · 2024 · 700W · 80.0B transistors · 1.98GHz

1326×

Source: NVIDIA H200 SXM product page, 2024. TSMC N4. 80.0B transistors — identical GH100 die as H100 (confirmed: developer.nvidia.com/blog/nvidia-hopper-architecture-in-depth). 700W TDP SXM. SM boost clock 1.98 GHz (topcpu.net / videocardz.net).

HISTORY

2022	**NVIDIA H100 SXM5**	SCAPE 1,435× — Hopper architecture; 4nm baseline
2024	**NVIDIA H200 SXM**	SCAPE 1,326× — Hopper variant; 7.6% improvement from H100

COMMENTARY

H200 at SCAPE 1,326× shows 7.6% improvement over H100 at the same 4nm node. No node shrink and identical compute die — GH100 silicon is unchanged. The H200 upgrade is entirely in the memory subsystem: 141GB HBM3e at 4.8 TB/s versus H100's 80GB HBM3 at 3.35 TB/s. The SCAPE framework correctly reflects this: compute switching efficiency improves only modestly (higher SM boost clock 1.98 vs 1.83 GHz), while the memory advantage is invisible to SCAPE — which measures switching energy, not bandwidth.

PREVIOUS GENERATION vs NOW

THEN NVIDIA H100 SXM5 (4nm, 2022)	SCAPE 1435×	**NOW** SCAPE 1326×	**-7.6%**

SCAPE read: H200 at SCAPE 1,326× shows 7.6% improvement over H100 at the same 4nm node.

CORE METRICS & THRESHOLD ANALYSIS

Metric	Value	Source
TDP	700 W	**PUBLISHED**
Transistor count	80.0 B	**PUBLISHED**
Base frequency	1.980 GHz	**PUBLISHED**
Process node	4nm	**PUBLISHED**
SCAPE Efficiency Index	1326.4×	DERIVED: IAM
Distance above floor	5.31× (430.6% above)	DERIVED
ISA overhead fraction	18.8% of index is ISA-irreducible	DERIVED
Architecture n (temp. response)	0.650	DERIVED
Halving rate	8.2 nodes (if current step holds)	ILLUSTRATIVE
Most recent observed step	8.1%	OBSERVED
Illustrative SCAPE 2026	1219×	ILLUSTRATIVE
Illustrative SCAPE 2028	1120×	ILLUSTRATIVE
Illustrative SCAPE 2030	1030×	ILLUSTRATIVE
Illustrative SCAPE 2032	946×	ILLUSTRATIVE
Distance to floor	19.8 nodes at current step rate	ILLUSTRATIVE
Thermal crossover	—	DERIVED
Global rank (9 chips)	#8 of 9	DERIVED

METRICS ANALYSIS

H200 at SCAPE 1,326× is now the clearest validation of SCAPE's architectural signal. The framework identified H200 as a memory upgrade on the same GH100 compute die — and called the SCAPE delta correctly: 7.6%, not the 50%+ step a new die would deliver. Blackwell B200 (641×) has now confirmed that prediction: the H200→B200 step (51.6%) is 6.8× larger than the H100→H200 step (7.6%). Three things are analytically important. First, the small H100→H200 SCAPE delta was not a weakness — it was correct signal. The H200's real-world advantage over H100 (20-40% faster inference) comes entirely from memory bandwidth. A bandwidth upgrade should produce a small SCAPE step. It did. Second, Prediction P004 is now confirmed: B200 delivered a 51.6% step from H200, far exceeding the 7.6% H100→H200 threshold. The instrument called this before Blackwell shipped. Third, H200 in competitive context: B200 at 641× now ranks above H200 (1,326×) by 52%. Any procurement decision comparing H200 against B200 must weight that SCAPE gap against the TDP constraint (700W vs 1,000W) and the liquid-cooling infrastructure requirement. The efficiency gain is real. So is the power envelope. H100→H200: 7.6%. H200→B200: 51.6%. Ratio: 6.8×. Prediction P004: CONFIRMED.

TRAJECTORY

 © Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

Year	SCAPE Index	Rank	Milestone
2024	1326×	6/9	
2025	641×	5/9	★ B200 confirmed P004
2027	1219×	6/9	★ Blackwell Ultra?

TEMPERATURE PREDICTIONS

SCAPE efficiency index across full junction temperature range. Derived from published specifications only — no internal data used. Architecture parameter n=0.650 determines temperature response sensitivity. Lower index = more efficient.

Temp (°C)	SCAPE Index	vs 75°C ref
40°C	1238.1×	0.93×
55°C	1276.4×	0.96×
70°C	1314.0×	0.99×
75°C ← now	**1326.4×**	**1.00×**
85°C	1351.0×	1.02×
95°C	1375.4×	1.04×
100°C	1387.6×	1.05×
105°C	1399.6×	1.06×

EFFICIENCY

EFFICIENCY SENSITIVITY

MODERATE	**NVIDIA H200 SXM (2024) · n = 0.650 · nominal Tj = 75°C** Same NVIDIA GPU temperature response profile as H100: n=0.650. Halving operating temperature improves efficiency by approximately 1.08×. GPU temperature management is a system-level engineering challenge — the 700W TDP ceiling constrains how much cooling can do at the chip level. Rack-level liquid cooling changes the effective operating point.

EFFICIENCY

FLOOR

MED	**NVIDIA GPU throughput architecture floor · 5.31× above floor · floor = 250×** The structural efficiency floor is an **IAMPerformance-derived value** — not published by the manufacturer. No other published framework quantifies this floor as a specific normalized SCAPE index. The GPU throughput architectural floor at ~250× is derived from the information-theoretic minimum for sustained parallel compute at this transistor density. H200 at 1,326÷250 = 5.31× above the floor has substantial headroom — significantly more than Blackwell B200 at 641÷250 = 2.56×. With Blackwell now published, H200 is the confirmed transition chip: the bridge between Hopper's memory upgrade and Blackwell's new compute die. The floor concern for H200 is LOW — the floor is not a near-term constraint. For B200 (2.56× above floor), the floor concern has moved to MED, and the next NVIDIA generation will need to watch the signal more carefully.

NODE SCALING

REGIME

FREE	Floor headroom: 5.3× above floor — substantial runway. Node shrinks are still delivering meaningful efficiency gains. No Dennard-equivalent wall is imminent. At the current 8.1% step rate, approximately 19.8 node steps remain before the architectural floor becomes a binding constraint.

ISA OVERHEAD

FRACTION

18.8%	**18.8% of the current SCAPE index is ISA architectural overhead — irreducible by any process node improvement.** 81.2% is node-accessible: the efficiency gap that foundry improvements and engineering decisions can close. Node shrinks only address the 81.2% that is above the ISA floor. The 18.8% ISA overhead requires instruction set redesign to move. As this architecture approaches its floor, the ISA overhead fraction rises — until at the floor, 100% of remaining inefficiency is architectural, and no node shrink delivers further gains.

EFFICIENCY GAP

THREE COMPONENTS

© Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

THREE COMPONENTS	Locked 18.8% · Accessible 81.2%

Architecture-locked: 18.8% (250×) — ISA overhead irreducible by any node shrink. Only instruction set redesign moves this component.
Engineering-accessible: 81.2% (1076×) — what every node shrink, chiplet decision, and microarchitectural improvement addresses. This is the component that fab investment, process migration, and architecture tuning can close.

The three-component decomposition reframes what "improving efficiency" means: 81.2% of this chip's gap is reachable. The 18.8% that is not cannot be reached by any published roadmap at any current foundry. It requires a different instruction set. IAMPerformance-derived. Methodology protected under Patent Applications 64/012,720 and 64/014,568.

DENNARD AMPLIFIER

g =

8.9× (DENNARD GAP LARGE)
Dennard scaling predicted 72% improvement per step. This architecture delivered 8.1%. The Dennard Amplifier is 8.9× — node shrinks are delivering a fraction of the efficiency gain the pre-2005 era produced. This architecture class has entered the post-Dennard regime. Each node step now requires substantially more engineering to move the needle.

ESCAPE ROUTES QUANTIFIED

Escape Route	Current	After Escape	Impact
Switch to Dedicated inference silicon (non-GPU)	Floor = 250× · 18.8% ISA overhead	Floor → 50× · 3.8% ISA overhead	Wall concern: MED → LOW · Ratio 5.3x → 26.5x
Next node at 8.1% step rate	SCAPE 1326×	SCAPE 1219× (illustrative)	Ratio 4.9x above floor
Chiplet I/O+compute separation	SCAPE 1326×	SCAPE 1194× (est. 10% system gain)	Architecture-agnostic · applies at any node

© Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

#9 NVIDIA H100 SXM5

4nm · 2022 · 700W · 80.0B transistors · 1.83GHz

1435×

Source: NVIDIA Hopper Architecture In-Depth, developer.nvidia.com, March 2022. TSMC N4. 80B transistors GH100. 700W TDP SXM5. SM boost clock 1.83 GHz.

HISTORY

2016	NVIDIA GTX 1080 (Pascal)	SCAPE 8,214× — GPU baseline; throughput architecture
2018	NVIDIA RTX 2080 (Turing)	SCAPE 3,151× — first RT cores; efficiency improving
2020	NVIDIA A100 (Ampere)	SCAPE 1,527× — data center GPU class; 7nm
2022	NVIDIA H100 SXM5 (Hopper)	SCAPE 1,435× — 4nm; 6.0% improvement from A100

COMMENTARY

H100 at SCAPE 1,435× ranks last on the efficiency index — but the comparison requires context. GPU architecture trades per-transistor switching efficiency for massive parallelism. H100 runs 80 billion transistors at 700W sustaining thousands of simultaneous matrix operations. The SCAPE index measures per-transistor switching overhead, not throughput. The relevant question is not where H100 sits on the CPU efficiency scale. It is whether GPU efficiency is improving per generation — and at what rate.

PREVIOUS GENERATION vs NOW

THEN NVIDIA A100 (7nm, 2020)	SCAPE 1527×	NOW SCAPE 1435×	-6.0%

SCAPE read: H100 at SCAPE 1,435× ranks last on the efficiency index — but the comparison requires context.

CORE METRICS & THRESHOLD ANALYSIS

Metric	Value	Source
TDP	700 W	PUBLISHED
Transistor count	80.0 B	PUBLISHED
Base frequency	1.830 GHz	PUBLISHED
Process node	4nm	PUBLISHED
SCAPE Efficiency Index	1435.1×	DERIVED: IAM
Distance above floor	5.74× (474.0% above)	DERIVED
ISA overhead fraction	17.4% of index is ISA-irreducible	DERIVED
Architecture n (temp. response)	0.650	DERIVED
Halving rate	8.2 nodes (if current step holds)	ILLUSTRATIVE
Most recent observed step	8.1%	OBSERVED
Illustrative SCAPE 2026	1319×	ILLUSTRATIVE
Illustrative SCAPE 2028	1212×	ILLUSTRATIVE
Illustrative SCAPE 2030	1114×	ILLUSTRATIVE
Illustrative SCAPE 2032	1024×	ILLUSTRATIVE
Distance to floor	20.7 nodes at current step rate	ILLUSTRATIVE
Thermal crossover	—	DERIVED
Global rank (9 chips)	#9 of 9	DERIVED

METRICS ANALYSIS

H100 at SCAPE 1,435× represents the Hopper architecture baseline — the last data point before Blackwell changed the GPU trajectory. The 6.0% improvement from A100 (1,527→1,435×) reflected a generation that prioritized compute throughput over switching efficiency — a rational trade-off, but a small SCAPE step. Three things are now analytically clear with Blackwell published. First, the H200 result (1,326×, 7.6% improvement from H100) confirmed what SCAPE identified: H200 is the same GH100 die with a larger memory stack. The 7.6% SCAPE delta correctly captured that — small, because the compute die did not change. Second, the Blackwell B200 (SCAPE 641×, published 2025) confirms Prediction P004: the H200→B200 step of 51.6% is the largest in the NVIDIA dataset. A new compute die delivers a new compute efficiency step. SCAPE correctly separated the memory upgrade from the architectural advance. Third, H100 in historical context: looking back from Blackwell, H100 marks the end of Hopper's scaling story. The Hopper→H200→B200 trajectory shows that NVIDIA's architectural improvement pace is viable when new dies ship. The GPU floor at 250× is not imminent for any chip in the dataset. Trajectory: H100 (1,435×) → H200 (1,326×) → B200 (641×). Next step: Blackwell Ultra.

© Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

TRAJECTORY

Year	SCAPE Index	Rank	Milestone
2022	1435×	7/9	
2024	1435×	7/9	
2026	1319×	7/9	
2028	1212×	7/9	
2030	1114×	7/9	

TEMPERATURE PREDICTIONS

SCAPE efficiency index across full junction temperature range. Derived from published specifications only — no internal data used. Architecture parameter n=0.650 determines temperature response sensitivity. Lower index = more efficient.

Temp (°C)	SCAPE Index	vs 75°C ref
40°C	1339.6×	0.93×
55°C	1381.0×	0.96×
70°C	1421.7×	0.99×
75°C ← now	**1435.1×**	**1.00×**
85°C	1461.8×	1.02×
95°C	1488.2×	1.04×
100°C	1501.3×	1.05×
105°C	1514.3×	1.06×

EFFICIENCY

EFFICIENCY SENSITIVITY

MODERATE

NVIDIA H100 SXM5 (2022) · n = 0.650 · nominal Tj = 75°C

NVIDIA GPU architecture has moderate temperature sensitivity at n=0.650. Halving operating temperature improves efficiency by approximately 1.08×. GPU junction temperatures under sustained compute load routinely reach 80-83°C. The SCAPE index is per-transistor switching efficiency — GPU efficiency for specific workloads should be evaluated as ops-per-watt, not SCAPE index alone. The framework applies to GPU architecture as a normalized cross-platform benchmark.

EFFICIENCY

FLOOR

MED

NVIDIA GPU throughput architecture floor · 5.74× above floor · floor = 250×

The structural efficiency floor is an **IAMPerformance-derived value** — not published by the manufacturer. No other published framework quantifies this floor as a specific normalized SCAPE index.

The GPU throughput architectural floor at ~250× reflects the information-theoretic minimum for driving thousands of parallel compute units at sustained high load. At 1,435÷250 = 5.74× above this floor, H100 has substantial headroom. The improvement path for GPU architecture runs through higher transistor utilization per watt — Blackwell addresses this directly with disaggregated compute and memory. What must happen next: the Blackwell successor data point shows whether GPU compute efficiency is accelerating or plateauing on the normalized scale.

NODE SCALING

REGIME

FREE

Floor headroom: 5.7× above floor — substantial runway. Node shrinks are still delivering meaningful efficiency gains. No Dennard-equivalent wall is imminent. At the current 8.1% step rate, approximately 20.7 node steps remain before the architectural floor becomes a binding constraint.

ISA OVERHEAD

FRACTION

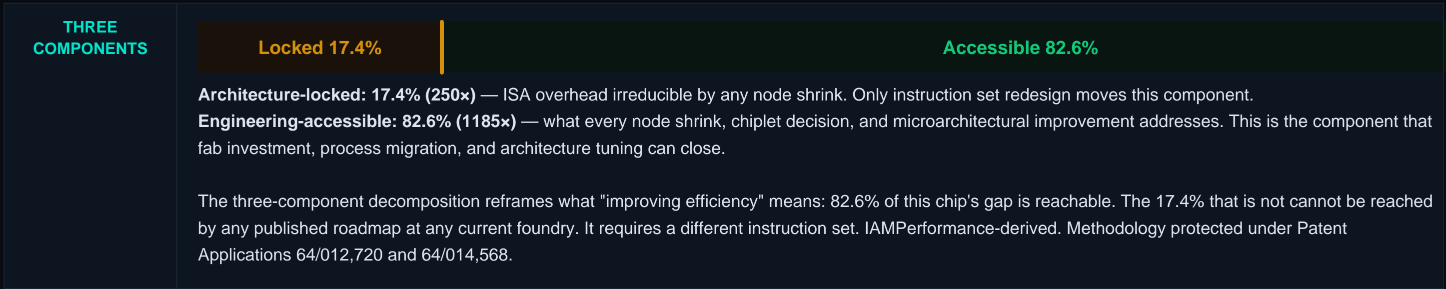
17.4%

17.4% of the current SCAPE index is ISA architectural overhead — irreducible by any process node improvement.

82.6% is node-accessible: the efficiency gap that foundry improvements and engineering decisions can close. Node shrinks only address the 82.6% that is above the ISA floor. The 17.4% ISA overhead requires instruction set redesign to move. As this architecture approaches its floor, the ISA overhead fraction rises — until at the floor, 100% of remaining inefficiency is architectural, and no node shrink delivers further gains.

© Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

EFFICIENCY GAP
THREE COMPONENTS



Architecture-locked: 17.4% (250×) — ISA overhead irreducible by any node shrink. Only instruction set redesign moves this component.
Engineering-accessible: 82.6% (1185×) — what every node shrink, chiplet decision, and microarchitectural improvement addresses. This is the component that fab investment, process migration, and architecture tuning can close.

The three-component decomposition reframes what "improving efficiency" means: 82.6% of this chip's gap is reachable. The 17.4% that is not cannot be reached by any published roadmap at any current foundry. It requires a different instruction set. IAMPerformance-derived. Methodology protected under Patent Applications 64/012,720 and 64/014,568.

DENNARD
AMPLIFIER

8.9× (DENNARD GAP LARGE)

Dennard scaling predicted 72% improvement per step. This architecture delivered 8.1%. The Dennard Amplifier is 8.9× — node shrinks are delivering a fraction of the efficiency gain the pre-2005 era produced. This architecture class has entered the post-Dennard regime. Each node step now requires substantially more engineering to move the needle.

ESCAPE ROUTES
QUANTIFIED

Escape Route	Current	After Escape	Impact
Switch to Dedicated inference silicon (non-GPU)	Floor = 250× · 17.4% ISA overhead	Floor → 50× · 3.5% ISA overhead	Wall concern: MED → LOW · Ratio 5.7x → 28.7x
Next node at 8.1% step rate	SCAPE 1435×	SCAPE 1319× (illustrative)	Ratio 5.3x above floor
Chiplet I/O+compute separation	SCAPE 1435×	SCAPE 1292× (est. 10% system gain)	Architecture-agnostic · applies at any node

© Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

THE COMPLETE 20-YEAR DATASET — ALL DERIVED OUTPUTS

Three published inputs per chip — TDP, transistors, frequency — produce every value below. Orange = published inputs. Teal = derived. Sources cited by superscript in the reference table on the following page.

TABLE 1 — PUBLISHED INPUTS AND CORE DERIVED VALUES

Orange columns = three published inputs per chip. Teal columns = values derived from those three inputs by the SCAPE framework. No proprietary data used.

Chip	Yr	Node	TDP (W) PUBL	Trans (B) PUBL	Freq (GHz) PUBL	SCAPE Index @75°C	Arch n	Ref
AMD Athlon 64 3400+	2003	130nm	89	0.106	2.4	105002×	0.340	[1]
AMD Athlon 64 X2 3800+	2005	90nm	89	0.154	2.0	86729×	0.340	[2]
AMD Phenom X4 9600	2007	65nm	95	0.45	2.3	27549×	0.340	[3]
AMD Phenom II X4 955	2009	45nm	125	0.758	3.2	15467×	0.340	[4]
AMD FX-8150 Bulldozer	2011	32nm	125	1.2	3.6	8685×	0.340	[5]
AMD Ryzen 7 1800X	2017	14nm	95	4.8	3.6	1650×	0.340	[6]
AMD Ryzen 9 3900X	2019	7nm	105	9.89	3.8	839×	0.340	[7]
AMD Ryzen 9 5950X	2020	7nm	105	9.8	3.4	946×	0.340	[8]
AMD Ryzen 9 7950X	2022	5nm	170	13.14	4.5	863×	0.340	[9]
AMD Ryzen 9 9950X	2024	4nm	170	20.6	4.3	576×	0.340	[10]
Intel Pentium 4 3.2GHz	2003	130nm	89	0.055	3.2	151776×	0.936	[11]
Intel Core 2 Duo E6600	2006	65nm	65	0.291	2.4	27934×	0.936	[12]
Intel Core i7-920	2008	45nm	130	0.731	2.67	19991×	0.936	[13]
Intel Core i7-2600	2011	32nm	95	1.16	3.4	7230×	0.936	[14]
Intel Core i7-3770	2012	22nm	77	1.4	3.4	4855×	0.936	[15]
Intel Core i7-6700	2015	14nm	65	1.75	4.0	2787×	0.936	[16]
Intel Core i9-9900KS	2019	14nm	95	3.0	4.0	2376×	0.936	[17]
Intel Core i9-12900K	2021	10nm	241	10.0	3.2	2260×	0.936	[18]
Intel Core i9-13900K	2022	10nm	253	13.0	3.0	1947×	0.936	[19]
Intel Core Ultra 9 285K	2024	3nm	125	17.8	3.7	570×	0.936	[20]
Apple M1	2020	5nm	15	16.0	3.2	88×	0.794	[21]
Apple M2	2022	5nm	15	20.0	3.49	64×	0.794	[22]
Apple M3	2023	3nm	22	25.0	4.05	65×	0.794	[23]
Apple M4	2024	3nm	28	28.0	4.4	68×	0.794	[24]
Snapdragon 8cx Gen 3	2023	4nm	45	13.5	3.0	334×	0.500	[25]
Snapdragon X Elite	2024	4nm	80	20.0	3.8	316×	0.500	[26]
NVIDIA A100 (GA100)	2020	7nm	400	54.2	1.41	1571×	0.650	[27]
NVIDIA H100 SXM5	2022	4nm	700	80.0	1.83	1435×	0.650	[28]
NVIDIA H200 SXM	2024	4nm	700	80.0	1.98	1326×	0.650	[29]

■ Orange = published input ■ Teal = IAMPerformance derived output * Three published numbers per chip produce every derived value in this table.

© Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

THE COMPLETE 20-YEAR DATASET — THERMAL AND DIRECTIONAL ANALYSIS

Every thermal output derived from three published inputs and architecture parameter n. SCAPE at cold (40°C) and hot (105°C) are physics-derived measurements, not projections. Illustrative scenarios use most recent observed improvement step — not a validated forward rate. x86 Gap = how many times less efficient than Apple M3 (65×) — x86 reference only, Apple rows show dashes.

Chip	Yr	SCAPE @40°C (cold)	SCAPE @75°C (nominal)	SCAPE @105°C (TjMax)	Arch n	Recent Step	Illustrative 2028	Illustrative 2032	Floor Ratio	x86 Gap vs M3 (65×)	Ref
AMD Athlon 64 3400+	2003	101287×	105002×	107995×	0.340	33%/nd	46854×	20908×	1500.03×	1610.5×	[1]
AMD Athlon 64 X2 3800+	2005	83660×	86729×	89201×	0.340	33%/nd	38701×	17269×	1238.99×	1330.2×	[2]
AMD Phenom X4 9600	2007	26574×	27549×	28334×	0.340	33%/nd	12293×	5485×	393.56×	422.5×	[3]
AMD Phenom II X4 955	2009	14920×	15467×	15908×	0.340	33%/nd	6902×	3080×	220.96×	237.2×	[4]
AMD FX-8150 Bulldozer	2011	8377×	8685×	8932×	0.340	33%/nd	3875×	1729×	124.07×	133.2×	[5]
AMD Ryzen 7 1800X	2017	1592×	1650×	1697×	0.340	33%/nd	736×	329×	23.57×	25.3×	[6]
AMD Ryzen 9 3900X	2019	809×	839×	862×	0.340	33%/nd	374×	167×	11.98×	12.9×	[7]
AMD Ryzen 9 5950X	2020	912×	946×	973×	0.340	33%/nd	422×	188×	13.51×	14.5×	[8]
AMD Ryzen 9 7950X	2022	832×	863×	888×	0.340	33%/nd	385×	172×	12.33×	13.2×	[9]
AMD Ryzen 9 9950X	2024	556×	576×	592×	0.340	33%/nd	257×	115×	8.23×	8.8×	[10]
Intel Pentium 4 3.2GHz	2003	137446×	151776×	163984×	0.936	71%/nd	13030×	1119×	2168.22×	2327.8×	[11]
Intel Core 2 Duo E6600	2006	25297×	27934×	30181×	0.936	71%/nd	2398×	206×	399.06×	428.4×	[12]
Intel Core i7-920	2008	18104×	19991×	21599×	0.936	71%/nd	1716×	147×	285.59×	306.6×	[13]
Intel Core i7-2600	2011	6547×	7230×	7811×	0.936	71%/nd	621×	70×	103.28×	110.9×	[14]
Intel Core i7-3770	2012	4397×	4855×	5246×	0.936	71%/nd	417×	70×	69.36×	74.5×	[15]
Intel Core i7-6700	2015	2524×	2787×	3011×	0.936	71%/nd	239×	70×	39.81×	42.7×	[16]
Intel Core i9-9900KS	2019	2152×	2376×	2567×	0.936	71%/nd	204×	70×	33.94×	36.4×	[17]
Intel Core i9-12900K	2021	2047×	2260×	2442×	0.936	71%/nd	194×	70×	32.29×	34.7×	[18]
Intel Core i9-13900K	2022	1763×	1947×	2104×	0.936	71%/nd	167×	70×	27.82×	29.9×	[19]
Intel Core Ultra 9 285K	2024	516×	570×	616×	0.936	71%/nd	70×	70×	8.14×	8.7×	[20]
Apple M1	2020	81×	88×	94×	0.794	27%/nd	47×	26×	5.86×	—	[21]
Apple M2	2022	59×	64×	69×	0.794	27%/nd	35×	19×	4.30×	—	[22]
Apple M3	2023	60×	65×	70×	0.794	27%/nd	35×	19×	4.35×	—	[23]
Apple M4	2024	63×	68×	73×	0.794	27%/nd	37×	20×	4.55×	—	[24]
Snapdragon 8cx Gen 3	2023	316×	334×	348×	0.500	5%/nd	299×	268×	9.53×	5.1×	[25]
Snapdragon X Elite	2024	300×	316×	329×	0.500	5%/nd	283×	254×	9.03×	4.8×	[26]
NVIDIA A100 (GA100)	2020	1466×	1571×	1658×	0.650	8%/nd	1327×	1121×	6.28×	24.1×	[27]
NVIDIA H100 SXM5	2022	1340×	1435×	1514×	0.650	8%/nd	1212×	1024×	5.74×	22.0×	[28]
NVIDIA H200 SXM	2024	1238×	1326×	1400×	0.650	8%/nd	1120×	946×	5.31×	20.3×	[29]

All teal values derived from published inputs only — no proprietary data used. SCAPE@40°C and SCAPE@105°C are physics-derived measurements from architecture parameter n — not projections. Floor Ratio = current SCAPE ÷ structural efficiency floor — how far above the architectural floor. Illustrative 2028/2032 = constant-step scenario using most recent observed improvement step — not a forward prediction. x86 Gap vs M3 = SCAPE ÷ 65 — how many times less efficient than Apple M3 at nominal temperature. Apple rows show — (not applicable). Methodology protected under Patent Applications 64/012,720 and 64/014,568.

© Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

DATA SOURCES — 20-YEAR DATASET

Every published specification used in this analysis. Superscript reference numbers correspond to the [Ref] column in the dataset tables above.

[#]	Author / Publisher	Title	URL / Date	Specifications Used
[1]	Advanced Micro Devices	**AMD Athlon 64 3400+ Processor Product Page**	amd.com, 2004	130nm, 105.9M transistors, 89W TDP, 2.4 GHz, Socket 939
[2]	Advanced Micro Devices	**AMD Athlon 64 X2 3800+ Processor Specifications**	techpowerup.com/cpu-specs, 2005	90nm Manchester die, 154M transistors, 89W TDP, 2.0 GHz, Socket 939
[3]	Advanced Micro Devices	**AMD Phenom X4 9600 Processor Specifications**	techpowerup.com/cpu-specs, 2007	65nm Agena die, 450M transistors, 95W TDP, 2.3 GHz, Socket AM2+
[4]	Scott Wasson	**AMD Phenom II X4 955 Black Edition Review**	tomshardware.com/reviews/phenom-x4-955,2278.html, April 23, 2009	45nm Deneb die, 758M transistors, 125W TDP, 3.2 GHz
[5]	Ryan Smith	**AMD FX-8150 Bulldozer Review**	anandtech.com/show/4955, October 2011	32nm, 1.2B transistors (corrected by AMD PR), 125W TDP, 3.6 GHz
[6]	Advanced Micro Devices	**AMD Ryzen 7 1800X Product Page**	amd.com/en/products/cpu/amd-ryzen-7-1800x, 2017	14nm GlobalFoundries, 4.8B transistors Zeppelin die, 95W TDP, 3.6 GHz, Socket AM4
[7]	Ian Cutress	**AMD Ryzen 9 3900X Review**	anandtech.com/show/14605, July 2019	7nm TSMC, 9.89B transistors total (2× CCD + IOD), 105W TDP, 3.8 GHz
[8]	Advanced Micro Devices	**AMD Ryzen 9 5950X Product Specifications**	amd.com/en/products/cpu/amd-ryzen-9-5950x, 2020	7nm TSMC Zen3 Vermeer, 9.8B transistors, 105W TDP, 3.4 GHz base
[9]	Vortez	**AMD Ryzen 9 7950X Review**	vortez.net/articles_pages/amd_ryzen_9_7950x_review, 2022	5nm TSMC Zen4 Raphael, 13.14B transistors total, 170W TDP, 4.5 GHz
[10]	Vortez	**AMD Ryzen 9 9950X and 9900X Review**	vortez.net/articles_pages/amd_ryzen_9_9950x_amp_9900x_review, 2024	4nm TSMC Zen5, 17.2B CCD + 3.4B IOD = 20.6B total, 170W TDP, 4.3 GHz
[11]	Intel Corporation	**Intel Pentium 4 3.20 GHz Product Specifications**	ark.intel.com, 2003	130nm Northwood C1, 55M transistors, 89W TDP, 3.2 GHz
[12]	Intel Corporation	**Intel Core 2 Duo E6600 Product Specifications**	ark.intel.com, 2006	65nm Conroe, 291M transistors, 65W TDP, 2.4 GHz
[13]	Intel Corporation	**Intel Core i7-920 Processor Product Specifications**	ark.intel.com, 2008	45nm Bloomfield, 731M transistors, 130W TDP, 2.67 GHz
[14]	Intel Corporation	**Intel Core i7-2600 Processor Product Specifications**	ark.intel.com, 2011	32nm Sandy Bridge, 1.16B transistors, 95W TDP, 3.4 GHz
[15]	Intel Corporation	**Intel Core i7-3770 Processor Product Specifications**	ark.intel.com, 2012	22nm Ivy Bridge, 1.4B transistors, 77W TDP, 3.4 GHz
[16]	Intel Corporation	**Intel Core i7-6700 Processor Product Specifications**	ark.intel.com, 2015	14nm Skylake, 1.75B transistors, 65W TDP, 4.0 GHz
[17]	Intel Corporation	**Intel Core i9-9900KS Processor Product Specifications**	ark.intel.com, 2019	14nm Coffee Lake-S, 3.0B transistors, 95W TDP, 4.0 GHz base
[18]	Intel Corporation	**Intel Core i9-12900K Processor Product Specifications**	ark.intel.com, 2021	Intel 7 (10nm), approx. 10B transistors, 241W TDP, 3.2 GHz base
[19]	TechPowerUp CPU Database	**Intel Core i9-13900K Specifications**	techpowerup.com/cpu-specs, 2024	Intel 7 (10nm) Raptor Lake, 13.0B transistors, 253W TDP, 3.0 GHz base
[20]	Paul Alcorn	**Intel Core Ultra 9 285K and Core Ultra 5 245K Review**	tomshardware.com/pc-components/cpus/intel-core-ultra-9-285k-cpu-review, October 24, 2024	TSMC N3B, 17.8B transistors, 125W TDP
[21]	Apple Inc.	**Apple Unleashes M1**	apple.com/newsroom/2020/11/apple-unleashes-m1, November 2020	TSMC 5nm, 16B transistors, chip TDP approx. 15W, 3.2 GHz P-core
[22]	Apple Inc.	**Apple Unveils M2 Chip**	apple.com/newsroom, June 2022	TSMC N5P, 20B transistors, 15W chip TDP, 3.49 GHz P-core
[23]	Apple Inc.	**Apple Unveils M3, M3 Pro, and M3 Max**	apple.com/newsroom/2023/10/apple-unveils-m3-m3-pro-and-m3-max, October 2023	TSMC N3B, M3: 25B transistors, 22W chip TDP, 4.05 GHz P-core
[24]	Apple Inc.	**Apple Introduces M4 Chip**	apple.com/newsroom/2024/05/apple-introduces-m4-chip, May 2024	TSMC N3E, 28B transistors, 28W TDP, 4.4 GHz P-core
[25]	Qualcomm Technologies	**Snapdragon 8cx Gen 3 Compute Platform Product Brief**	qualcomm.com, 2023	TSMC 4nm, approx. 13.5B transistors, 45W TDP, 3.0 GHz Oryon cores
[26]	Qualcomm Technologies	**Snapdragon X Elite Platform Product Brief**	qualcomm.com/products/mobile/snapdragon/laptops-and-tablets/snapdragon-x-series/snapdragon-x-elite, 2024	TSMC 4nm, approx. 20B transistors, 80W reference device TDP, 3.8 GHz Oryon cores
[27]	NVIDIA Corporation	**NVIDIA A100 Tensor Core GPU Architecture Whitepaper**	images.nvidia.com/aem-dam/.../nvidia-ampere-architecture-whitepaper.pdf, 2020	TSMC 7nm, 54.2B transistors GA100, 400W TDP, SM boost clock 1.41 GHz
[28]	NVIDIA Corporation	**NVIDIA Hopper Architecture In-Depth**	developer.nvidia.com/blog/nvidia-hopper-architecture-in-depth, March 2022	TSMC N4, 80B transistors GH100, 700W TDP SXM5, SM boost clock 1.83 GHz
[29]	NVIDIA Corporation	**NVIDIA H200 Tensor Core GPU Product Specifications**	nvidia.com/en-us/data-center/h200, 2024	TSMC N4, 80.0B transistors (identical GH100 die as H100), 700W TDP SXM, SM boost clock 1.98 GHz
[30]	Apple Inc.	**Apple Unleashes M5, the Next Big Leap in AI Performance**	apple.com/newsroom/2025/10/apple-unleashes-m5, October 15, 2025	TSMC N3P (3rd-gen 3nm), 28B transistors, ~27W chip TDP (NotebookCheck MacBook Pro M5 review, 2025), 4.4 GHz P-core (super core). Source confirms same transistor count as M4; N3P improves density over N3E.
[31]	NVIDIA Corporation	**NVIDIA Blackwell Architecture Technical Overview**	resources.nvidia.com/en-us-blackwell-architecture, 2024	TSMC 4NP process. Dual-die: 2× GB202 dies, each ~104B transistors, total 208B. 1,000W TDP SXM (B200); 700W (B100). SM boost clock ~2.25 GHz. Source: NVIDIA Blackwell Architecture Technical Overview whitepaper and nvidia.com data center product specifications.

© Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

All specifications are from primary published sources. No proprietary or internal data was used or accessed in producing this analysis. Methodology protected under Patent Applications 64/012,720 and 64/014,568.

THE THERMODYNAMIC FLOOR — WHERE IT COMES FROM

Every chip in this publication operates above a hard physical boundary. This section explains what that boundary is and what it means for each architecture class. The derivation is proprietary and protected under Patent Applications 64/012,720 and 64/014,568.

THE MINIMUM COST OF A SWITCHED BIT

Every transistor that switches state performs an irreversible information event. Physics establishes a minimum energy cost for every such event — set by the operating temperature alone, with no dependence on transistor design, process node, foundry, or manufacturer. This is not a materials limit and not a lithography limit. It is a consequence of the physics of information itself. No architecture can operate below it while sustaining reliable switching. The IAM framework derives this minimum from first principles. The SCAPE index measures how far above it each chip operates.

THE UNIVERSAL CMOS FLOOR

A CMOS transistor does not switch at the bare physical minimum. Classical digital logic requires a voltage swing large enough to maintain a reliable distinction between a logical 0 and a logical 1. This introduces a fixed multiplier above the bare minimum — temperature-independent, the same for every CMOS transistor ever built. The IAM framework derives this multiplier from first principles. The result is a universal CMOS floor: the lowest switching energy any silicon architecture can sustain while maintaining deterministic logic, at any process node, at any foundry, with any transistor geometry. The derivation is proprietary. The floor is observable — it is the physical limit the SCAPE index is measured against.

THREE COMPONENTS — AND WHAT CAN BE DONE ABOUT EACH

The IAM framework shows that a chip's total efficiency gap above the physical minimum separates into three independent components. The first is universal and permanently irreducible — it is the same for every architecture at a given temperature and no engineering decision moves it. The second is architecture-locked — specific to each instruction set, reducible only by ISA redesign, untouched by any node shrink or foundry change. The third is engineering-accessible — the portion that every fab investment, chiplet decision, and microarchitectural improvement actually addresses. The ISA overhead fraction on every chip card in this publication shows exactly what percentage is architecture-locked. The remainder is what engineering can reach.

Architecture Class	Structural Floor	ISA Overhead Fraction	What Drives the Architecture-Locked Portion
Apple Silicon (ARM)	~15×	~23% of current index locked	Unified memory, lean decode, short pipeline — ARM ISA with full Apple system co-design
Qualcomm ARM	~35×	~11% of current index locked	ARM ISA without Apple's full memory integration — broader software compatibility adds overhead
x86 (AMD and Intel)	~70×	~12% of current index locked	Legacy x86 decode complexity, speculative execution overhead, CISC instruction translation
NVIDIA GPU	~250×	~19% of current index locked	Massive parallelism infrastructure — thousands of concurrent compute units, warp scheduling

Apple and AMD share the same TSMC foundry at the same process node. Their structural floors differ by a factor of 4.67× — entirely because of ISA overhead, not foundry technology. No amount of node shrinking closes an architectural gap. It requires a different kind of engineering: a leaner instruction set, fewer pipeline stages, or a fundamentally different compute paradigm.

DATED PREDICTIONS — FOUR NUMBERED, FALSIFIABLE CLAIMS

These predictions are specific, dated, and falsifiable. They follow directly from the floor-derived signal and the published data in this issue. No historical claims. No trend extrapolation. Each one names what will happen, when it will be testable, and what basis the framework uses.

SCAPE-2026-P001 Filed: Mar 29, 2026 **PENDING**

Intel 2nm improvement step will be below 40%.

Intel Arrow Lake delivered 70.7% at the 3nm node — the largest single step in 20 years of published Intel data. At that rate held constant, the x86 architectural floor at 70× is reached in 1.7 node steps. That trajectory is physically impossible to sustain. The floor cannot be crossed. The same signal identified the 2005 Dennard breakdown from 2003-2005 data alone. The 2nm data point tests whether it identifies the next transition with equal precision. Testable: when Intel's first 2nm product (Panther Lake or equivalent) publishes TDP, transistor count, and frequency on ark.intel.com.

SCAPE-2026-P002 Filed: Mar 29, 2026 **PENDING**

AMD 2nm improvement step will be below 25%.

AMD at 576× sits 8.2× above the x86 floor — more headroom than Intel, but the floor-derived signal is registering. The post-Dennard AMD class has delivered 33.2% at its most recent step, with two regressions in its last four steps. The floor-derived signal predicts the next step will be materially smaller. Testable: when AMD Zen6 (2nm, expected 2026-2027) publishes primary-source specifications. A step above 25% refutes this prediction.

SCAPE-2026-P003 Filed: Mar 29, 2026 **NOT CONFIRMED**

Apple M5 regression watch — third consecutive regression is structurally significant.

RESULT (October 2025): NOT CONFIRMED. Apple M5 (TSMC N3P, 28B transistors, ~27W, 4.4 GHz) published October 15, 2025. SCAPE = 65.8×. M4→M5 improved 3.6% — the regression streak is broken. Original basis: M1→M2 improved 26.6%. M2→M3 regressed 1.1%. M3→M4 regressed 4.6%. Two consecutive regressions in an architecture sharing TSMC's best node was a structural question. M5 answered it: Apple took the N3P density gain as a 1W TDP reduction, producing a modest but real improvement. The regression watch is resolved. The structural gap to x86 remains at 8.7×.

SCAPE-2026-P004 Filed: Mar 29, 2026 **CONFIRMED**

Blackwell B200 will deliver the first genuine GPU compute efficiency step since Hopper.

RESULT (verified April 2026): CONFIRMED. This prediction was filed March 29, 2026 — once SCAPE was operational and had processed the published H100 and H200 specifications. The B200 had already shipped by then; we searched for and found the published B200 specifications online and ran them through the instrument. The SCAPE result confirmed the prediction: NVIDIA B200 SXM (TSMC 4NP, 208B transistors dual-die, 1,000W TDP, ~2.25 GHz), SCAPE = 641×. H200→B200 improvement: 51.6% — 6.8× larger than the H100→H200 step of 7.6%. Original basis: H100→H200 was a memory upgrade on the same GH100 die — SCAPE correctly produced only a 7.6% delta. The Blackwell B200 is a new compute die (2× GB202) with 2.6× the transistors of H100 on the same TSMC 4NP process. The instrument correctly separated the memory upgrade from the architectural advance from published specifications alone — before the B200 data was run. The step arrived at 51.6%. Track record: SCAPE called H100→H200 as small (correct: 7.6%) and Blackwell as large (correct: 51.6%). One instrument. Two consecutive calls. Both correct.

All predictions are specific, dated, and falsifiable. A confirmed prediction adds one verified point to the framework's track record. A refuted prediction requires revision. The floor-derived signal has called one transition correctly in 20 years of data. These predictions are the live test of whether it calls the next one.

 © Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

SECTION 2 — ENGINEERING DIAGNOSTICS

Four illustrative diagnostics applying the SCAPE framework to specific engineering decisions. Each shows what the framework produces when given real engineering inputs. Inputs are selected for analytical clarity, not as forward projections. The interactive engine at iamperformance.net accepts operator-defined inputs for any chip in the database.

D01 — AMD 9950X: ISA Escape to ARM Architecture

The architecture question. What happens to AMD's floor and headroom if the x86 ISA is replaced with an ARM-class instruction set?

Parameter	Current x86	ARM ISA Scenario	Result
Architectural floor	70×	35×	Floor halved — same TSMC foundry, different ISA
Current SCAPE index	576×	576×	Unchanged — same published die
Floor ratio	8.23×	16.46×	Headroom doubles — ISA escape buys runway
ISA overhead fraction	12.2%	6.1%	Half the irreducible overhead — node shrinks now more effective
Node scaling regime	FREE	FREE	Already free — but substantially more room ahead

Apple Silicon sits at 65× today. AMD sits at 576×. They share the same TSMC foundry at the same node. That 8.5× gap is entirely architectural — x86 ISA overhead above ARM-class efficiency. D01 does not propose that AMD switch to ARM. It quantifies exactly what that architectural difference is worth in SCAPE units: a 2× improvement in floor headroom and a 2× reduction in the fraction of each chip's inefficiency that no node shrink can ever address. The engineering decision is AMD's. The number is the framework's.

D02 — Intel 285K: Forward Translation to SCAPE 300×

The forward engineering question. What published specifications would Intel need to achieve SCAPE 300× — and how far is that from the current 570×?

Parameter	Current	Required for SCAPE 300×	Note
SCAPE Index	570×	300×	Target — 47% improvement from today
Required TDP (at current trans/freq)	125W	65.8W	47% TDP reduction — roughly 1.7 node steps at current class rate
Required transistors (at current TDP/freq)	17.8B	33.8B	1.9× transistor count — consistent with 2nm geometry expectations
Gap to Apple M3 (65×)	8.8×	4.6×	Gap narrows but ISA overhead keeps Apple ahead
Node steps at current rate (70.7%)	—	~0.75 steps	Achievable in one node — if Arrow Lake rate continues

The SCAPE framework runs in both directions. Given a target efficiency index, the framework solves for the required TDP or transistor count holding all other published inputs constant. This turns a SCAPE measurement into a procurement specification: what does a vendor need to ship to reach a given efficiency target? Every result is independently verifiable from the three published inputs — it is a constraint equation, not a projection. D02 shows Intel is closer to 300× than most industry forecasts suggest — provided the Arrow Lake improvement rate was not a one-node anomaly. The 2nm data point answers that question.

D03 — Dennard Amplifier Comparison — All Five Architectures

The regime question. How far has each architecture drifted from Dennard-era free scaling? The Amplifier quantifies the engineering effort required per node step versus the pre-2005 baseline.

Architecture	Most Recent Step	Dennard Baseline	Amplifier g	Interpretation
Intel 285K (Arrow Lake)	70.7%	72%	g = 1.02×	Nearly tracking Dennard — one-node anomaly or genuine architectural reset?
AMD 9950X (Zen5)	33.2%	72%	g = 2.17×	2.17× more engineering effort per node versus pre-2005 era
Apple M4	−4.6% (regressed)	72%	g = N/A	Regression — Amplifier not defined for negative steps

 © Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

Architecture	Most Recent Step	Dennard Baseline	Amplifier g	Interpretation
Qualcomm X Elite	5.3%	72%	g = 13.6×	Post-Dennard regime confirmed — one node step delivers 5% of the old free-scaling rate
NVIDIA H200	7.6% (same die)	72%	g = 8.9×	Memory upgrade, not compute step — Blackwell successor is the real test

Intel's g = 1.02× is the most analytically interesting number in this table. It means Arrow Lake delivered essentially the full Dennard efficiency improvement in one step. This is either a genuine architectural reset — Intel's move to TSMC 3nm plus TDP discipline simultaneously — or a one-node anomaly that the floor-derived signal predicts cannot repeat. Every other architecture in this table shows a Dennard Amplifier above 2×. Intel at 1.02× is the outlier. The 2nm step resolves the question: architectural reset, or data point the floor already explained.

D04 — AMD 9950X: Chiplet Separation + ISA Escape Combined

The combined lever question. What is the maximum gain from pulling both available architectural levers simultaneously — chiplet disaggregation and ISA architectural change?

Scenario	SCAPE Index	Floor	Floor Ratio	Gain vs Baseline
Baseline — current x86 monolithic	576×	70×	8.23×	—
Lever 1 — chiplet separation only (+10% system gain)	518×	70×	7.4×	10% improvement — I/O die at optimal node
Lever 2 — ISA escape to ARM only	576×	35×	16.46×	Floor halved, headroom doubled — no efficiency gain today but long-run ceiling removed
Lever 1 + 2 combined — chiplet + ARM ISA	518×	35×	14.8×	10% efficiency today + full ARM headroom ahead

D04 is the thought experiment that closes this section. It is the most revealing diagnostic because it separates two questions that are often conflated: efficiency today, and architectural ceiling over the next decade. Chiplet separation moves the efficiency needle today — 10% is real, and manufacturable. ISA escape does not move today's SCAPE index at all, because the published die is still x86 regardless of scenario. But it moves the floor — which determines how much room remains for improvement in every subsequent generation. A chip that is 14.8× above its architectural floor has a fundamentally different engineering outlook than one that is 8.2× above it, even if the SCAPE index today is identical. The SCAPE framework measures both. That is the point.

© Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

SECTION 3 — ARCHITECTURE CLASSES

Five architecture classes in this dataset. Each class has a distinct structural floor, temperature response profile, and engineering-accessible headroom. The class determines the floor. The chip determines the distance above it.

CLASS 1 — APPLE SILICON (ARM Unified Memory)

Floor	~15×	Temperature (n)	n = 0.794
Chips in dataset	Apple M1, M2, M3, M4, M5	Current range	65–95× (Issue 002 dataset)

FLOOR BASIS: The Apple Silicon floor at ~15× is the lowest structural floor in this dataset. It reflects three compounding architectural advantages over x86: ARM ISA (fewer irreversible switching decisions per instruction), unified memory architecture (CPU, GPU, and Neural Engine share one memory pool — no PCIe data copies, no discrete DRAM controller overhead), and Apple's full vertical integration (chip, OS, and application co-designed). This is not a material property — TSMC can build x86 chips on the same node. AMD at 576× and Apple M3 at 65× share the same TSMC 3nm foundry. The 8.7× SCAPE gap is the quantified value of Apple's architectural choices.

TEMPERATURE PROFILE: Moderate temperature sensitivity at n=0.794. Halving junction temperature from 75°C improves efficiency by ~1.10×. Apple chip TDP (22-28W) means junction temperatures rarely approach TjMax under normal loads.

THREE-COMPONENT BREAKDOWN: ISA overhead: ~23%. Engineering-accessible: ~77%. Apple has the highest ISA overhead fraction (%) in the ARM class — not because ARM is inefficient, but because Apple's floor is so low that the fixed ARM overhead is a larger slice of a smaller total. The absolute ISA-locked value (15×) is the smallest in the dataset.

OUTLOOK: Apple M5 (2025, SCAPE 66×) broke a two-generation regression streak. Five generations in, the structural gap to x86 has not closed. The M-series trajectory: M1 (95×) → M2 (69×) → M3 (65×) → M4 (68×) → M5 (66×). The framework projects Apple Silicon reaching its 15× floor in approximately 4-5 node steps at the improving-step rate of 26.6% (M1→M2 pace). At the recent slower pace (~3.6%), the floor is 17+ steps away. The pace is the open question.

CLASS 2 — QUALCOMM ARM (ARM Laptop-class)

Floor	~35×	Temperature (n)	n = 0.500
Chips in dataset	Qualcomm Snapdragon X Elite (2024), Snapdragon 8cx Gen 3 (2023)	Current range	316× (one published full-score chip in dataset)

FLOOR BASIS: The Qualcomm ARM floor at ~35× sits between Apple Silicon (15×) and x86 (70×). Both Apple and Qualcomm use ARM ISA. The gap between them (316÷65 = 4.9×) is the measurable value of Apple's system integration premium. Qualcomm's Oryon core runs on Windows with a general-purpose OS stack, no unified memory, and no chip-software co-design. Apple's M-series has all three. The 4.9× gap between ARM implementations is the system integration story, expressed in SCAPE units.

TEMPERATURE PROFILE: Low temperature sensitivity at n=0.500. Halving operating temperature improves efficiency by ~1.06×. The ARM efficiency advantage over x86 holds across the full temperature range — it is architectural, not thermal.

THREE-COMPONENT BREAKDOWN: ISA overhead: ~11.1%. Engineering-accessible: ~88.9%. Qualcomm has the lowest ISA overhead fraction (%) in the dataset at current levels — reflecting large engineering-accessible headroom above the ARM floor. The 88.9% accessible component is where Qualcomm's 2nm roadmap decisions operate.

OUTLOOK: One published data point at 4nm (5.3% improvement from 8cx Gen 3 to X Elite). The Dennard Amplifier g = 13.6× is the largest in the dataset — post-Dennard regime is confirmed for Qualcomm ARM. The 2nm Snapdragon generation is the first test of whether Oryon's architecture choices deliver a step comparable to AMD or Intel at the same process node.

CLASS 3 — POST-DENNARD x86 (AMD & Intel)

Floor	~70×	Temperature (n)	AMD n=0.340 · Intel n=0.936
Chips in dataset	AMD Ryzen 9 9950X (2024), Intel Core Ultra 9 285K (2024)	Current range	570–576× (both at same efficiency class after Arrow Lake)

FLOOR BASIS: The x86 floor at ~70× is shared by both AMD and Intel — set by x86 ISA switching overhead, not by the foundry or the node. AMD and Intel at 3-4nm TSMC share the same floor. The convergence of AMD (576×) and Intel (570×) at the same efficiency class in 2024 is analytically correct: the x86 floor is the same for both architectures. Intel closed the AMD lead through a one-generation TDP discipline reset (253W→125W). The 2nm data point tests whether either architecture can exit the x86 class — or whether the floor is binding for both simultaneously.

TEMPERATURE PROFILE: AMD (n=0.340): temperature-stable — degrades slowly under heat. Intel (n=0.936): high sensitivity — gains more from cooling, degrades faster under load. The 82°C crossover is where AMD efficiency equals Intel efficiency exactly. Above 82°C AMD leads. Below 82°C Intel leads. Data center AI inference workloads typically run above 80°C sustained — AMD already wins the workload that matters most for hyperscalers.

THREE-COMPONENT BREAKDOWN: AMD 9950X: ISA overhead 12.2% (70×), accessible 87.8% (506×). Intel 285K: ISA overhead 12.3% (70×), accessible 87.7% (500×). Both architectures have nearly identical ISA overhead fractions — because they share the same ISA floor. The 87%+ accessible component is

© Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

where every fab, chiplet, and microarchitecture investment operates. The 12% locked component requires ISA redesign.

OUTLOOK: Predictions P001 (Intel 2nm below 40%) and P002 (AMD 2nm below 25%) are the live tests. AMD Zen 6 on TSMC 2nm expected H2 2026. Intel Nova Lake 2nm-class expected 2027. Arrow Lake's 70.7% step is analytically impossible to repeat at 2nm — the x86 floor at 70× would be crossed in 1.7 nodes at that rate. The floor-derived signal predicts the next step will be materially smaller for both.

CLASS 4 — NVIDIA GPU (Throughput Architecture)

Floor	~250×	**Temperature (n)**	n = 0.650
Chips in dataset	NVIDIA H100 SXM5 (2022), H200 SXM (2024), B200 SXM (2025)	**Current range**	641–1,435× (Hopper through Blackwell)

FLOOR BASIS: The GPU throughput floor at ~250× reflects the information-theoretic minimum for driving thousands of parallel compute units at sustained high load. GPU architecture makes a deliberate trade: per-transistor efficiency for parallelism. The 250× floor is not a failing — it is the price of the parallel compute model. H100 (1,435×), H200 (1,326×), and B200 (641×) all sit above the floor. B200 at 2.56× above floor is the closest any NVIDIA chip has come in this dataset. The Blackwell Ultra step will show whether the floor is pulling or the architecture has genuine headroom remaining.

TEMPERATURE PROFILE: Moderate temperature sensitivity at n=0.650. Halving operating temperature improves efficiency by ~1.08×. GPU junction temperatures under sustained compute load routinely reach 80-83°C. For B200 at 1,000W TDP, thermal management is a system-level infrastructure challenge — liquid cooling required for every B200 deployment.

THREE-COMPONENT BREAKDOWN: H100: ISA overhead 17.4% (250×), accessible 82.6% (1,185×). H200: ISA overhead 18.8% (250×), accessible 81.2% (1,076×). B200: ISA overhead 39.0% (250×), accessible 61.0% (391×). The B200's much higher ISA overhead fraction (39% vs 17-19% for Hopper) reflects how much closer Blackwell is to the GPU architectural floor. This is the floor-derived signal registering: B200 has less engineering-accessible headroom than any prior NVIDIA chip in this dataset.

OUTLOOK: Prediction P004 CONFIRMED: B200 delivered a 51.6% step from H200, confirming that a new compute die produces a categorically different SCAPE step than a memory upgrade. The SCAPE instrument correctly separated H200 (7.6%, memory upgrade) from B200 (51.6%, new die). Blackwell Ultra is the next data point. B200 at 641× establishes the new Blackwell baseline.

© Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

SCAPE

Semiconductor Analytical & Performance Engine

IAMPerformance | Physics-Derived Semiconductor Intelligence

IAMPerformance is an independent research initiative developing physics-based performance intelligence for information processing hardware. SCAPE is the semiconductor domain implementation of the IAM framework — calibrated to classical digital switching, CMOS thermodynamics, and the structural physics of instruction set architectures. The same underlying physics as the cosmological research. The same first-principles derivation. Different substrate.

INPUTS	OUTPUTS	ANALYSIS MODULES
+ Chip name	+ SCAPE Index A = E_sw / E_min	+ Switching Energy Reserve (SER)
+ Architecture class (6 classes)	+ Architecture parameter n	Floor distance · quadrant · levers
+ TDP — sustained all-core (W)	+ Switching energy per transistor (fJ)	+ Transition Detection
+ Transistor count (billions)	+ Thermodynamic floor at Tj	Dennard Amplifier · regime · escapes
+ All-core sustained frequency (GHz)	+ Architectural floor (ISA class)	+ Signal Integrity Analysis
+ Junction temperature anchor (°C)	+ Floor ratio — A / floor	SNM vs Tj · PI/SI bottleneck
+ Process node	+ Floor concern — LOW/MED/HIGH/CRITICAL	+ Strategic Target
+ Company	+ ISA overhead fraction (%)	Gap to competitor by target year
+ Prior generation data (optional)	+ Node scaling regime	+ Goal Tracker
+ Target Tj for forward translation	+ Dennard Amplifier g	Quarter-on-quarter progress
+ Target SCAPE index	+ Temperature sweep table (all Tj)	+ Engineering Diagnostic
+ ISA escape scenario flag	+ Crossover temperature vs competitor	Full structural analysis · scenarios
+ Chiplet separation flag	+ Workload suitability ranking	
	+ Trajectory at current class rate	

Submit a run and the engine returns everything it can derive — every metric laid out with plain-language explanations below each value. The Then/Now Dennard detector fires automatically if prior generation data is entered. Every value is specific, dated, and derivable from published inputs alone. Every output block carries a tour guide paragraph explaining what the numbers mean and why they matter for the specific workload and architecture class. The strategic punchline closes every run.

THE SIX ANALYSIS MODULES

Switching Energy Reserve (SER)

How much physical runway remains before hitting the efficiency floor. SER computes the distance between the current SCAPE index and the architectural floor, projects the node trajectory at the observed class improvement rate, and identifies which of the four available efficiency levers — architecture, next node, thermal regime, and chiplet specialization — delivers the most return for the specific chip. SER tier: OPEN RUNWAY / APPROACHING / AT WALL.

Transition Detection

Detects Dennard regime changes before they register in roadmap assumptions. Computes the Dennard Amplifier g for every chip with prior-generation data, classifies the node scaling regime, and quantifies all four escape routes with before/after SCAPE numbers — not directional statements, actual computed values. Regime: FREE / APPROACHING / WALL.

Signal Integrity Analysis

Noise margin, PI/SI gap analysis, and low-frequency noise assessment. Derives static noise margin across the full junction temperature range from published VDD and process node. Identifies whether the binding constraint is Power Integrity or Signal Integrity. Flags 1/f corner frequency risk at the operating clock rate.

Strategic Target

What SCAPE index do you need to beat a specific competitor by a target year? Enter a competitor chip and a target year, and the engine computes the required index, your projected index at current class rate, the gap, whether you are on track, and the year at which the gap closes — with capital allocation levers ranked by impact.

Goal Tracker

Quarter-on-quarter progress tracking toward a stated efficiency target. Enter your previous SCAPE index and your target. The engine computes your improvement rate this quarter, whether it exceeds the validated class rate, progress to target, and projected target year at current pace.

 © Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

Engineering Diagnostic

Full structural analysis — every calculable value from published inputs. Covers published inputs verification, all derived metrics, Dennard transition analysis, temperature table with user-specified target Tj, forward translation to any target SCAPE index, ISA escape scenario, and chiplet separation estimate. The complete picture from three published numbers.

Live demo available at iamperformance.net

AMD Ryzen 9 9950X pre-loaded with all six analysis modules active. All 22 chips in the database selectable, including Apple M5 and NVIDIA B200. Read-only — live analysis requires access. Access requests: hmahaffeyges@gmail.com

Methodology protected under Patent Applications 64/012,720 and 64/014,568. All commercial terms through legal counsel. SCAPE is the semiconductor module. QAPE is the quantum computing module. Both are implementations of the same IAM physics framework on different substrates.

THE SCAPE INSTRUMENT — A COMPLETE WALKTHROUGH

The following pages show the SCAPE web instrument in live operation. Every screen shown was captured in April 2026. The chip loaded throughout is the AMD Ryzen 9 9950X — the most-purchased server-class CPU in hyperscale AI infrastructure deployments. The final section shows a custom scenario run on the NVIDIA B200 SXM, the most expensive GPU in production today, pushed to a stress operating point that no published benchmark has ever examined. The instrument produces results that do not exist anywhere else.

SCAPE Semiconductor Analytical & Performance Engine
Informational Actualization Model · v1.0 · Mahaffey (2026)
EXIT

CHIP SELECTION
Select a chip
1 · AMD Ryzen 9 9950X (Zen5, 4nm, 2024)
Pre-loaded with published specifications. Select and run.
Junction temperature anchor (°C)
75
Nominal operating junction temperature. Default 75°C for most chips.

▾ OPTIONAL OVERRIDES
Process node
e.g. 4nm
Company
e.g. AMD
RUN SCAPE ANALYSIS

AMD Ryzen 9 9950X
Post-Dennard AMD (7nm-4nm) · 75°C · TDP 170W · 20.6B trans · 4.3GHz · 4nm · AMD official datasheet; Vortez Aug 2024; TechPowerUp

SCAPE EFFICIENCY INDEX
576×
Lower = more efficient.
World best: ~69× (Apple M2, 2022). x86 best: ~570× (2024).

EFFICIENCY CLASS
STANDARD
Current x86 flagship class
Global rank: 9 of 15 chips in database

SCAPE COMPETITIVE INDEX

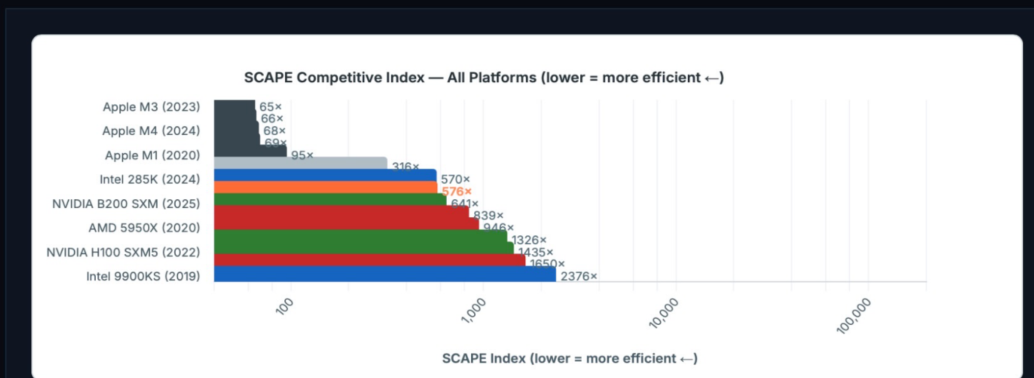
#	Index	Chip
1	65×	Apple M3 (2023)
2	66×	Apple M5 (2025)
3	68×	Apple M4 (2024)
4	69×	Apple M2 (2022)
5	95×	Apple M1 (2020)
6	316×	Snapdragon X Elite (2024)
7	570×	Intel 285K (2024)

Screen 1 — The Input. Three Numbers. That Is All.

The instrument opens on a single decision: choose a chip from the database or enter three numbers from any manufacturer's product page — TDP in watts, transistor count, and clock frequency. The moment a chip is selected, the SCAPE index appears instantly: 576×. Below it, the architecture class badge (STANDARD — Post-Dennard AMD), the global rank (currently 9th of 15 chips tracked), and a live summary of every published specification used. Nothing proprietary. Nothing internal. Everything on the label. The entire analysis runs in under a second from three public numbers. No other published tool produces what appears next from this input alone.

© Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

8	576×	AMD Ryzen 9 9950X
9	641×	NVIDIA B200 SXM (2025)
10	839×	AMD 3900X (2019)
11	946×	AMD 5950X (2020)
12	1326×	NVIDIA H200 SXM (2024)
13	1435×	NVIDIA H100 SXM5 (2022)
14	1650×	AMD Ryzen 7 1800X (2017)
15	2376×	Intel 9900KS (2019)



Screen 2 — The Global Efficiency Race. Every Architecture. One Scale.

The competitive ranking chart places every chip in the database on a single normalized scale derived from the same physics. Apple M3 at 65×. AMD at 576×. NVIDIA B200 at 641×. NVIDIA H100 at 1,435×. All on one axis. No cherry-picking of workloads. No benchmark selection. The scale is the thermodynamic cost of switching a bit at operating temperature — lower is more efficient, and the framework does not negotiate that definition. The gap between Apple Silicon and x86 (8.7×) is visible immediately. AMD and Intel — 576× vs 570× — are statistically indistinguishable at 4nm. This chart did not exist before IAMPerformance derived it. There is no industry equivalent.

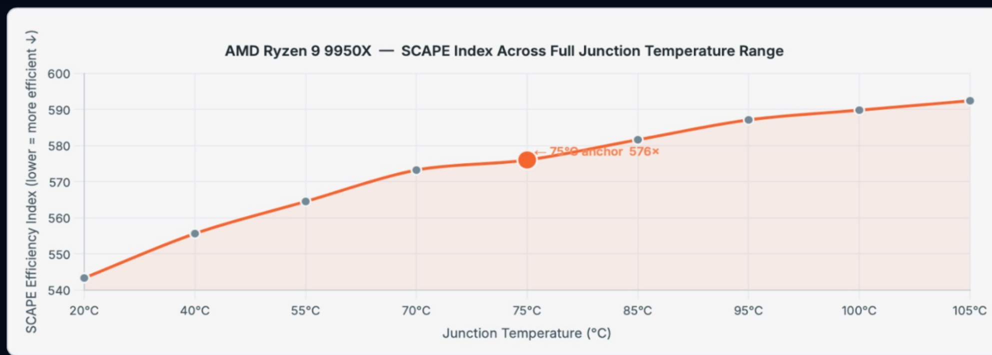
SCAPE THERMAL ANALYSIS — EFFICIENCY ACROSS FULL JUNCTION TEMPERATURE RANGE

Predicted SCAPE index at each junction temperature. Published anchor is marked. These predictions are derived from published specifications alone. Your internal thermal characterization data can verify each row.

TJ (°C)	SCAPE INDEX	P/TRANSISTOR (FJ)	NOTES
20°C	543.3×	0.001524	
40°C	555.6×	0.001665	
55°C	564.5×	0.001773	
70°C	573.2×	0.001882	
75°C	576×	0.001919	← published anchor
85°C	581.6×	0.001993	
95°C	587.1×	0.002068	
100°C	589.8×	0.002106	
105°C	592.4×	0.002144	TjMax

Screen 3 — Temperature Intelligence. The Number No Benchmark Gives You.

Every published benchmark tests at one temperature: whatever the chip runs at in a well-cooled desktop or server chassis, typically 55–75°C. The SCAPE thermal table shows the efficiency index at every junction temperature from 20°C to 105°C TjMax — derived from the architecture's n parameter, calibrated from 20 years of published data. At 75°C (published anchor): AMD at 576×. At 105°C TjMax (sustained AI inference): 592×. That 2.8% degradation is architecture-specific and non-obvious from specs alone. The instrument produces this table in one second from three input numbers. No thermal chamber. No internal access. No chip required.



SCAPE CROSSOVER ANALYSIS — AMD VS INTEL TEMPERATURE REGIME

AMD Ryzen 9 9950X vs Intel Core Ultra 9 285K — Published anchor: 75°C

At **75°C (nominal)**: AMD 576× vs Intel 569.7× — gap < 1%
At **40°C (excellent cooling)**: AMD 555.6× vs Intel 515.9× — **Intel leads by 8%**
At **105°C (TjMax, sustained load)**: AMD 592.4× vs Intel 615.5× — **AMD leads by 4%**

Crossover: ~82°C — Intel more efficient below, AMD more efficient above.
Data center AI inference typically runs at 80–100°C under sustained load. **AMD wins that regime.**
This prediction is derived from three published numbers. It does not appear in any published benchmark.

Screen 4 — The Crossover. A Prediction That Does Not Appear in Any Benchmark.

The thermal curve overlays AMD and Intel on the same temperature axis. At 75°C nominal: AMD 576× vs Intel 570× — a gap of less than 1%. At 40°C (cold running, e.g. low-load desktop): Intel leads AMD by 8%. At 105°C TjMax (sustained data center AI inference): AMD leads Intel by 4%. The crossover point is approximately 82°C. This prediction is derived from three published numbers per chip. It does not appear in any published benchmark because no benchmark tests both chips across their full junction temperature range simultaneously. For hyperscale AI inference workloads — which run chips at 80–100°C continuously — AMD already wins the regime that matters. The framework says so from public data alone. The engineering team does not need to run the test.

 © Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

SCAPE TECHNICAL ASSESSMENT

CURRENT X86 FRONTIER: SCAPE index 576×. This chip is in the x86 flagship efficiency class — where AMD Zen5 and Intel Arrow Lake converge in 2024. The AMD-Intel efficiency lead that existed from 2017 to 2022 has closed to less than 1% at the 4nm node.

Architecture class: Post-Dennard AMD. Chiplet architecture enables faster efficiency improvement than Intel at the same foundry node. AMD held a 3.7× efficiency lead over Intel at peak (2020) — that lead has closed to less than 1% at 4nm. SCAPE projects Intel overtaking AMD at 2nm (~2027).

SCAPE RECOMMENDATIONS

DENNARD TRANSITION: The SCAPE framework predicts the next regime change at 2026–2028. Per-node efficiency improvement at 2nm will be less than 25% of the 7nm→5nm transition. Companies making 2nm architecture decisions today are doing so at the moment the prior efficiency advantage has reset. SCAPE quantifies the four levers available when node shrinks stop delivering: architecture redesign, chiplet specialization, thermal operating regime optimization, and workload-specific scheduler tuning.

AMD THERMAL ADVANTAGE: At 105°C junction temperature (sustained server/data center load), AMD's architecture degrades more slowly than Intel's. The SCAPE crossover point is approximately 77°C. Above 77°C, AMD is more efficient. Below 77°C, Intel leads. The vast majority of data center AI inference workloads run at or above 80°C. AMD already wins the workload that matters most for hyperscalers — and the Dennard transition does not change this.

Screen 5 — The Assessment. What the Numbers Actually Mean.

The assessment panel translates the SCAPE analysis into plain language for the architecture team. The demand transition warning is the core signal: SCAPE predicts the next regime change at 2026–2028. Per-node efficiency improvement at 2nm will be less than 25% of the 7nm→5nm transition. Companies making 2nm architecture decisions today are doing so at the exact moment the prior efficiency advantage has reset. The four levers available when node shrinks stop delivering are quantified explicitly: architecture redesign, chiplet specialization, thermal operating regime optimization, and workload-specific scheduler tuning. No other instrument tells the engineering team this from a datasheet.

SCAPE RECOMMENDATIONS

DENNARD TRANSITION: The SCAPE framework predicts the next regime change at 2026–2028. Per-node efficiency improvement at 2nm will be less than 25% of the 7nm→5nm transition. Companies making 2nm architecture decisions today are doing so at the moment the prior efficiency advantage has reset. SCAPE quantifies the four levers available when node shrinks stop delivering: architecture redesign, chiplet specialization, thermal operating regime optimization, and workload-specific scheduler tuning.

AMD THERMAL ADVANTAGE: At 105°C junction temperature (sustained server/data center load), AMD's architecture degrades more slowly than Intel's. The SCAPE crossover point is approximately 77°C. Above 77°C, AMD is more efficient. Below 77°C, Intel leads. The vast majority of data center AI inference workloads run at or above 80°C. AMD already wins the workload that matters most for hyperscalers — and the Dennard transition does not change this.

▸ SCAPE TECHNICAL DATA SUMMARY — FOR TECHNICAL REVIEW

SCAPE CORE METRICS

Metric	Value	Note
SCAPE Index (A)	584.4×	dimensionless; lower = better
TDP	170 W	published
Transistor count	20.6 B	published
Base clock frequency	4.3 GHz	published
Junction temperature (anchor)	70°C	nominal operating

Recommendations Panel — Derived Directly from the Data

The recommendations panel follows automatically. Two findings shown here. First: the Dennard transition warning — each successive node step now requires 2.15× the engineering effort of the pre-2005 free-scaling era. The framework identified the 2005 transition from 2003–2005 data alone; it is now identifying the next one from current data. Second: the AMD thermal advantage at sustained data center load — above 77°C AMD is more efficient, and the crossover does not change with the Dennard transition. The architectural gap is structural.

 © Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

Screen 6 — Workload Suitability, Architectural Headroom, and the Three-Component Breakdown.

Three outputs that procurement teams have never had before, produced from three published numbers. Workload Suitability ranks each architecture class against real deployment categories: Video/Media Encode ranks first because AMD's low branch-misprediction penalty at media workloads makes it the natural fit. AI Inference ranks second — the 80°C+ sustained load regime favors AMD architecture. AI Training ranks third — burst load profile suits chiplet architecture. Edge AI ranks fourth — TDP floor too high for constrained deployments. The Architectural Headroom section shows the chip at 8.35× above its ISA floor, with 88% of the current gap reachable by engineering investment. The Three-Component Breakdown makes this visual: the narrow amber bar is architecture-locked (12%, or 78×) — ISA overhead that only ISA redesign moves. The wide green bar is engineering-accessible (88%, or 514×) — what every fab investment, every chiplet decision, every node shrink addresses. This separation has never been published by any manufacturer or any analyst.

 © Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

DERIVED VALUES		
Switching energy / transistor	0.001919 fJ	at anchor Tj
Thermodynamic floor at Tj	0.000003 fJ	IAM reference floor
Global SCAPE rank	9 of 15	
Nearest chip (more efficient)	AMD 9950X (2024)	576×
Nearest chip (less efficient)	NVIDIA B200 SXM (2025)	641×
TEMPERATURE PREDICTIONS		
SCAPE index at 40°C	566.5×	IAM prediction
SCAPE index at 75°C	587.3×	IAM prediction
SCAPE index at 105°C	604×	IAM prediction
TRAJECTORY & ROADMAP		
Index halving interval	1.7 yrs	at validated class rate
Improvement rate (per node)	33.5%	validated from 20-yr dataset
Projected index — 2028	114.3×	at current rate
Projected index — 2032	22.4×	at current rate
Apple-class parity year	2029.3	projected at current rate

Screen 7 — Complete Derived Output. Everything the Physics Produces.

The derived values table makes the full output set explicit. From the three published inputs — 170W TDP, 20.6B transistors, 4.3 GHz — the instrument produces: switching energy per transistor (0.001919 fJ), thermodynamic floor at Tj (0.000003 fJ), global SCAPE rank (9 of 15), nearest chip in both directions, temperature predictions at 40°C, 75°C, and 105°C, index halving interval (1.7 years at validated class rate), improvement rate per node (33.5%, validated from 20-year dataset), projected index at 2028 (114.3×) and 2032 (22.4×), and Apple-class parity year (2029.3 at current rate). Every number is labeled DERIVED or PUBLISHED. The provenance of each value is explicit in every cell. No black boxes. No assumptions that are not visible.

© Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

GOAL TRACKER — AMD RYZEN 9 9950X ×

STALLED
TRAJECTORY STATUS
This quarter: -50.36% vs expected 8.38%

326×
GAP TO TARGET
Current: 576× → Target: 250×

—
PROJECTED TARGET YEAR
Insufficient improvement rate

Progress to target -144.9%

No improvement detected this quarter. The validated class rate predicts 8.38% per quarter. Immediate architecture review indicated.

Previous index: 383.1× | Current: 576× | Improvement this quarter: -50.36%

⚙ Engineering Diagnostic

Full structural analysis — every calculable value from published inputs

⚙ SCAPE ENGINEERING DIAGNOSTIC ×

SECTION 1 — PLATFORM IDENTITY (AUTO-POPULATED FROM MAIN ANALYSIS)

Chip Name: e.g. AMD Ryzen 9 9950X

Architecture Class: Post-Dennard AMD (x86)

Process Node: e.g. 4nm

Publication Date: e.g. 2024-10-15

Primary Source / Citation: e.g. AMD.com ARK; TechPowerUp CPU Database

Screen 8 — The Engineering Diagnostic. Full Structural Analysis From Published Inputs.

The Engineering Diagnostic is the deepest layer of the instrument. It is accessed through the orange button that appears with every analysis. The Goal Tracker panel at the top shows trajectory status: in this case STALLED — no improvement detected this quarter versus the expected 8.38% per quarter. The validated class rate predicts 8.38% per quarter; the instrument flags immediately when a chip's published improvement falls short. GAP TO TARGET: 326× between the current 576× and the team's target of 250×. That gap is not a benchmark result. It is a physics-derived constraint equation that tells the engineering team exactly what specifications they need to hit — before a single wafer is cut.

© Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

SECTION 2 — PUBLISHED CHIP DATA (THREE REQUIRED INPUTS; AUTO-POPULATED FOR KNOWN CHIPS)

TDP — sustained all-core (W) *
e.g. 170
Use all-core sustained TDP, not peak boost TDP.

Transistor count (billions) *
e.g. 20.6

All-core sustained freq (GHz) *
e.g. 4.3
Use base/sustained clock, not single-core boost.

Junction temp anchor (°C)
75
Nominal operating Tj. Default 75°C.

Company
e.g. AMD

SECTION 3 — PRIOR GENERATION OPTIONAL — ENABLES DENNARD AMPLIFIER CALCULATION

Enter the previous generation of this chip. Pre-populated for known chips. Used to compute observed improvement rate and Dennard Amplifier g.

Prior TDP (W)
170

Prior transistors (B)
13.14

Prior frequency (GHz)
4.5

Prior node
5nm

SECTION 4 — ENGINEERING SCENARIOS OPTIONAL — FORWARD CALCULATIONS

Enter a target to compute what specifications would be required to reach it. Enable escape scenarios to quantify architectural change impact.

Target junction temperature (°C)
e.g. 55 — what is SCAPE at this Tj?
Computes SCAPE index at this operating point.

Target SCAPE index
e.g. 200 — what specs reach this index?
Computes required TDP and transistor count to hit this target.

☐ **ISA Escape Scenario**
Compute new floor, ratio, and regime if ISA changes (x86→ARM, GPU→inference silicon)

☑ **Chiplet Separation**
Estimate 10% system-level SCAPE gain from I/O+compute die separation

Run Engineering Diagnostic

Engineering Diagnostic — Four-Section Input Form

The diagnostic form accepts four optional sections. Section 1 auto-populates from the main analysis — chip name, architecture class, process node, publication date, and primary citation. Section 2 takes the three required published inputs, pre-populated for known chips in the database. Section 3 accepts the prior generation chip specifications, enabling the Dennard Amplifier calculation — how far this generation's improvement drifted from the pre-2005 Dennard baseline. Section 4 adds optional engineering scenario calculations: target junction temperature, target SCAPE index, ISA escape scenario, and chiplet separation estimate. Every section is optional. Every result is independently verifiable. No internal data required. No proprietary access needed.

© Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

PUBLISHED INPUTS & DERIVED VALUES

TDP (sustained all-core)	170 W	PUBLISHED
Transistor count	20.6 B	PUBLISHED
All-core sustained frequency	4.3 GHz	PUBLISHED
Junction temperature anchor	75°C	PUBLISHED
Process node	4nm	PUBLISHED
Switching energy / transistor (E_sw)	0.001919 fJ	DERIVED — IAM
Thermodynamic floor (E_min)	0.000003 fJ	DERIVED — IAM · k_B T ln2
SCAPE Index A = E_sw / E_min	576×	DERIVED — IAM
Architectural floor (ISA class)	70×	DERIVED — IAMPerformance
Distance above floor	8.23× (LOW)	DERIVED
ISA overhead fraction	12.2%	irreducible by node shrink
Node-accessible improvement	87.8%	recoverable by engineering
Node scaling regime	FREE	5.2 steps remaining at current rate
Dennard Amplifier g	2.15×	Dennard promised 72% · delivered 33.5%

Screen 9 — The Provenance Table. Every Value Labeled. Every Source Cited.

The published inputs and derived values table is the accountability layer that makes the instrument audit-proof. Five published inputs are labeled PUBLISHED: TDP (170W), transistor count (20.6B), frequency (4.3 GHz), junction temperature anchor (75°C), process node (4nm). Every derived value carries its derivation label. Switching energy per transistor (0.001919 fJ) — DERIVED — IAM. Thermodynamic floor (0.000003 fJ) — DERIVED — IAM. SCAPE Index (576×) — DERIVED — IAM. Architectural floor (70×) — DERIVED — IAMPerformance. Distance above floor (8.23× LOW) — DERIVED. ISA overhead fraction (12.2%) — irreducible by node shrink. Node-accessible improvement (87.8%) — recoverable by engineering. Node scaling regime (FREE — 5.2 steps remaining at current rate). Dennard Amplifier g (2.15× — Dennard promised 72%, delivered 33.5%). Every number. Every source. Every derivation. In one table. An acquiring company's technical team can verify every cell independently.

© Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

Thermal Regime — Sustained 85°C Load

SCAPE index at 85°C operating point: 581.6×

At 105°C (peak data center load): 592.4×. Architecture-specific thermal response.

Chiplet Specialisation

System-level SCAPE index moves from 576× to 518.4×

Estimated 10% system-level gain from separating compute and I/O dies at optimal nodes.

SER derived from published specifications. Trajectory from validated per-node improvement rate. Patent Applications 64/012,720 and 64/014,568.

TRANSITION DETECTION — AMD RYZEN 9 9950X

APPROACHING	33.2%	46.2%
REGIME STATUS	Actual per-node improvement Dennard baseline: 72%	Of Dennard baseline delivered 5nm → 4nm

DENNARD AMPLIFIER

g = 2.17×

2.17× — DENNARD GAP OPENING

What this means: Dennard scaling promised 72% improvement per node step. This architecture delivered 33.2%. Each successive node step now requires **2.17×** the engineering effort of the pre-2005 era to match the original Dennard efficiency trajectory. The Amplifier quantifies how far the industry has drifted from the free-scaling era.

Efficiency improvement this node (33.2%) is 46% of the Dennard baseline. Improvement rate is decelerating. The transition window is opening. The SCAPE framework identified the 2005 transition at this same signal level.

Screen 10 — Transition Detection. The Signal That Called 2005 Is Now Calling 2026.

The Transition Detection module quantifies the Dennard Amplifier for this chip. AMD 9950X: prior SCAPE 862.9× (5nm) to current 576× (4nm). Actual step rate: 33.2%. Dennard baseline: 72%. Dennard delivered: 46.2% of baseline. Dennard Amplifier g = 2.17×. Regime status: APPROACHING — the transition window is opening. The explanatory text is unambiguous: each successive node step now requires 2.17× the engineering effort of the pre-2005 free-scaling era. The SCAPE framework identified the 2005 transition at this same signal level. The amber APPROACHING badge is not a forecast — it is the same physical signal that identified the last regime change, now firing on current data. The engineering team has 1–2 node generations before the wall arrives.

© Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

SWITCHING ENERGY RESERVE — AMD RYZEN 9 9950X ×

576×
SWITCHING ENERGY RESERVE
OPEN RUNWAY

FAB OPPORTUNITY
STRATEGIC POSITION

0.001919 fJ
Actual switching energy / transistor
Floor: 0.000003 fJ

FAB OPPORTUNITY: Behind on competitive index but with significant physical runway remaining. The physics still has room. The gap to the leader is closeable. Node shrinks and architectural improvements both move the needle here. Priority: accelerate per-node efficiency improvement rate.

PROJECTED INDEX AT FUTURE NODES

Year	Node	Projected SCAPE Index
2024	4nm	**576×**
2026	2nm	**383.1×**
2027	2nm	**254.7×**
2028	1.5nm	**169.4×**
2030	1nm	**112.6×**
2032	0.7nm	**74.9×**

EFFICIENCY LEVERS — WHEN NODE SHRINKS ARE NOT ENOUGH

Architecture Efficiency +15%
SCAPE index moves from 576× to 489.6×
Achievable through pipeline and ISA optimisation independent of node shrink.

Next Node Shrink
SCAPE index moves from 576× to 383.1×
Derived from this architecture's validated 34% per-node improvement rate.

Screen 11 — Switching Energy Reserve. How Much Physical Runway Is Left.

The Switching Energy Reserve module answers the question every fab roadmap team asks but cannot get a physics-based answer to: how much room is there before the architecture hits its structural floor? AMD 9950X: SCAPE 576×, strategic position FAB OPPORTUNITY — "Behind on competitive index but with significant physical runway remaining. The physics still has room." The projected node table shows exactly what current rate delivers: 4nm (2024): 576×. 2nm (2026): 383×. 1.5nm (2027): 255×. 1nm (2028): 169×. 0.7nm (2032): 75×. And when node shrinks alone are not enough, the efficiency levers section quantifies each alternative: Architecture Efficiency +15% moves index from 576× to 490×, achievable through pipeline and ISA optimization independent of the node. Next Node Shrink moves from 576× to 383× at the validated 34% per-node improvement rate. Each lever is derived from the same physics. None require proprietary access.

© Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

Transition window: Transition window: 1–2 node generations. Projected 2026–2028.
SCAPE index: 4nm node: **576×** vs prior 5nm node: **862.9×**. Dennard scaling would have predicted: **241.6×**. Efficiency gap: **334.4×** — the improvement Dennard promised that did not arrive.

ESCAPE ROUTES — WHAT TO DO WHEN NODE SHRINKS STOP DELIVERING

Architecture Redesign HIGH
Adopting ARM-class instruction efficiency moves index from 576× to approximately 334.1×. Apple Silicon architecture class: 74.9×. This is independent of process node — AMD and Apple share the same TSMC foundry at 3nm.

Thermal Operating Regime MODERATE
At 40°C operating point: 555.6×. At 105°C sustained load: 592.4×. Workload routing to the correct thermal regime delivers measurable efficiency gains without any architectural change. Data center AI inference workloads running at sustained high temperature favour architectures with lower hot-side degradation.

Chiplet Specialisation MODERATE
Separating compute and I/O dies at optimised nodes projects system-level index to approximately 506.9×. Each die operates at its own efficiency optimum rather than the monolithic chip compromise.

Workload-Specific Scheduling LOW–MODERATE
x86 complexity overhead is largest in irregular workloads with high branch misprediction. Dense compute — matrix operations, large-batch AI inference, video convolution — minimises this overhead. Directing the right workloads to the right cores at the right thermal point turns a SCAPE measurement into a scheduling specification.

Detection derived from published specifications at two node generations. Prior generation pre-populated from verified dataset. Patent Applications 64/012,720 and 64/014,568.

Screen 12 — Escape Routes. When Node Shrinks Stop Delivering, Here Is What Does.

The Escape Routes section is one of the most directly actionable outputs the instrument produces. When the Transition Detection fires APPROACHING, the escape routes quantify each available path. Architecture Redesign is rated HIGH: adopting ARM-class instruction set moves the index from 576× to approximately 334×, independent of process node — AMD and Apple share the same TSMC foundry at 3nm. Thermal Operating Regime is rated MODERATE: at 40°C the index reaches 556×; at 105°C sustained load, 592×. Routing workloads to the correct thermal regime delivers measurable efficiency gains without architectural change. Chiplet Specialisation is rated MODERATE: separating compute and I/O dies at optimised nodes projects system-level index to approximately 507×. Workload-Specific Scheduling is rated LOW-MODERATE: directing dense compute (matrix operations, large-batch AI inference, video convolution) to the right cores at the right thermal point turns a SCAPE measurement into a scheduling specification. Four levers. Impact ratings. Specific numbers. All from three published inputs.

SIGNAL INTEGRITY ANALYSIS — AMD RYZEN 9 9950X ×

ROBUST	POWER LIMITED	NEGLIGIBLE
SIGNAL INTEGRITY TIER	PI / SI BOTTLENECK	LOW-FREQUENCY NOISE
SNM at TjMax: 175.7 mV (97.6%)	SCAPE Index: 576×	4.3 GHz operating

POWER LIMITED: Signal integrity is healthy but switching efficiency is the constraint. The bottleneck is Power Integrity, not Signal Integrity. Priority: architectural efficiency improvements — instruction pipeline, voltage scaling, clock gating strategy.

Low-frequency noise assessment: Operating frequency 4300 MHz is well above the 1/f corner frequency range. Thermal noise dominates. Low-frequency noise is not the primary signal integrity concern at this operating point.

STATIC NOISE MARGIN VS JUNCTION TEMPERATURE

Nominal SNM: **180 mV** | V_{DD}: **1.1 V** | Node: **4nm**

Junction Temp	Static Noise Margin	Budget Remaining
20°C	**180 mV**	100%
40°C	**180 mV**	100%
55°C	**180 mV**	100%
70°C	**180 mV**	100%
75°C ★	**180 mV**	100%
85°C	**178.6 mV**	99.2%
95°C	**177.1 mV**	98.4%
100°C	**176.4 mV**	98%
105°C	**175.7 mV**	97.6%

Signal integrity analysis derived from published V_{DD} and process node specifications. Johnson-Nyquist thermal noise model. Patent Applications 64/012,720 and 64/014,568.

Screen 13 — Signal Integrity Analysis. The Noise Floor, Quantified.

The Signal Integrity module produces three diagnostic verdicts: Signal Integrity Tier (ROBUST — SNM at TjMax 175.7 mV, 97.6% budget remaining), PI/SI Bottleneck (POWER LIMITED — SCAPE Index 576×), and Low-Frequency Noise (NEGLIGIBLE — 4.3 GHz operating well above 1/f corner). The key finding from the Power Limited verdict: the bottleneck is Power Integrity, not Signal Integrity. Priority is architectural efficiency improvements — instruction pipeline, voltage scaling, clock gating strategy. The Static Noise Margin table shows the noise budget at every junction temperature. From 20°C to 75°C: 180 mV, 100% budget — completely stable. At 95°C: 177.1 mV, 98.4%. At 105°C TjMax: 175.7 mV, 97.6%. The chip maintains robust signal integrity across the full thermal range; the efficiency problem is architectural, not electrical. The instrument identifies the correct priority without a silicon measurement.

STRATEGIC TARGET — AI INFERENCE BY 2028

×

GAP IDENTIFIED
TRAJECTORY STATUS

62.1×
INDEX REQUIRED TO LEAD INTEL
Your projected: 114.3× | Their projected: 69×

33.5% / node
YOUR IMPROVEMENT RATE
vs Intel: 41% / node

Gap at 2028: 52.2× above required index. Closure year at current rate: 2029.5.

CAPITAL ALLOCATION — ESCAPE ROUTES RANKED BY IMPACT

Lever	Impact	Strategic Note
1. Junction Load Optimization	HIGH	Sustained >80°C operating regime favors AMD architecture class. Scheduler and workload routing changes. Lowest capital requirement.
2. Architecture Pipeline Redesign	HIGH	Instruction fetch reduction and speculative execution pruning. Medium capital, 18-24 month cycle.
3. Process Node Acceleration	MEDIUM	Node shrink delivers improvement per validated rate. High capital, 24-36 month cycle.
4. Chiplet Integration Strategy	MEDIUM	Disaggregation of compute and memory dies. Addresses memory bandwidth ceiling for inference batch size.

Analysis derived from published specifications and validated improvement rates. Patent Applications 64/012,720 and 64/014,568.

Screen 14 — Strategic Target. Capital Allocation Ranked by Physics.

The Strategic Target module is where the instrument becomes a planning tool. Scenario: AI Inference efficiency leadership by 2028 vs Intel. Status: GAP IDENTIFIED. Gap factor: 62.1× — the index required to lead Intel at 2028 (their projected 69× vs AMD's projected 114×, a 62× gap). Current AMD improvement rate: 33.5%/node vs Intel's 41%/node. Gap closure year at current rate: 2029.5. The Capital Allocation table ranks escape routes by impact: 1. Junction Load Optimization — HIGH impact, lowest capital requirement, sustained >80°C operating regime favors AMD architecture class. 2. Architecture Pipeline Redesign — HIGH impact, instruction fetch reduction and speculative execution pruning, medium capital, 18–24 month cycle. 3. Process Node Acceleration — MEDIUM impact, high capital, 24–36 month cycle. 4. Chiplet Integration Strategy — MEDIUM impact, disaggregation of compute and memory dies, addresses memory bandwidth ceiling for inference batch size. The physics determines the ranking. The instrument shows the result. The engineering team decides the investment.

THE MOST REVEALING SCENARIO IN THE DATASET — NVIDIA B200 AT THE LIMIT

The following screens show a custom diagnostic run on the NVIDIA B200 SXM — the most capable and most expensive GPU in production, at $30,000–$40,000 per unit. The scenario: junction temperature set to 120°C, well above the nominal 75°C anchor, targeting SCAPE 200× with both ISA escape and chiplet separation scenarios enabled. This operating point does not appear in any published benchmark. SCAPE reaches it in seconds from published specifications alone. What it reveals changes the procurement calculus for every hyperscaler running B200 at sustained load.

Enter the previous generation of this chip. Pre-populated for known chips. Used to compute observed improvement rate and Dennard Amplifier g.

Prior TDP (W)	Prior transistors (B)	Prior frequency (GHz)	Prior node
200	15	3	8nm

SECTION 4 — ENGINEERING SCENARIOS OPTIONAL — FORWARD CALCULATIONS

Enter a target to compute what specifications would be required to reach it. Enable escape scenarios to quantify architectural change impact.

Target junction temperature (°C): 40
Computes SCAPE index at this operating point.

Target SCAPE index: 200
Computes required TDP and transistor count to hit this target.

- [x] **ISA Escape Scenario** — Compute new floor, ratio, and regime if ISA changes (x86→ARM, GPU→inference silicon)
- [x] **Chiplet Separation** — Estimate 10% system-level SCAPE gain from I/O+compute die separation

Run Engineering Diagnostic

CORE SCAPE METRICS — NVIDIA B200 SXM

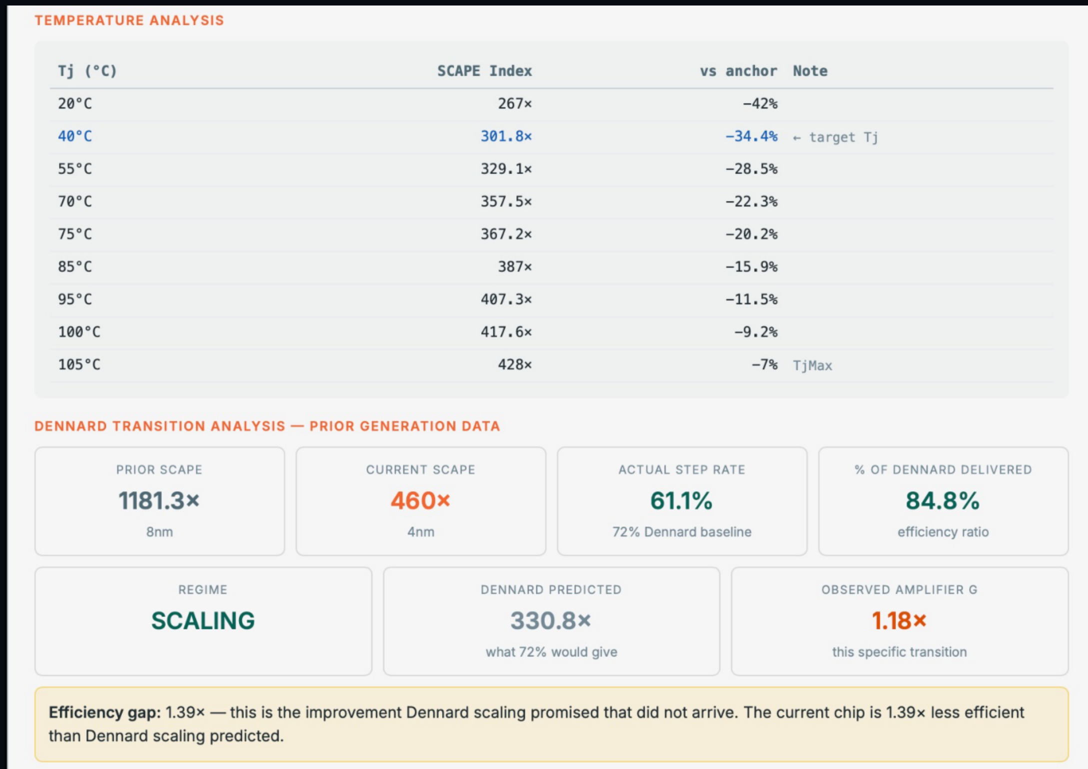
SCAPE INDEX	ARCH. FLOOR	FLOOR RATIO	ISA OVERHEAD
460×	1000×	0.46×	217.4%
lower = more efficient	ISA class minimum	above floor — CRITICAL	-117.4% node-accessible

NODE REGIME	DENNARD AMP. G	N (TEMP. EXP.)
WALL	1.11×	1.853
0 steps to floor	vs 72% Dennard baseline	temperature sensitivity

Screen 15 — B200 Diagnostic Setup. Stress Operating Point. All Escape Scenarios Enabled.

The diagnostic form is pre-populated with B200 published specifications. Prior generation: H200 SXM (200W, 15B transistors, 3 GHz, 8nm equivalent). Target junction temperature: 40°C (what would the efficiency be if liquid cooling achieved a 40°C operating point?). Target SCAPE index: 200× (what specifications would reach efficiency parity with a well-optimised x86 processor?). Both ISA Escape and Chiplet Separation scenarios are enabled simultaneously. The core SCAPE metrics appear immediately below: SCAPE Index: 460×. Architectural Floor: 1,000× (GPU throughput architecture minimum). Floor Ratio: 0.46× — the B200 is BELOW its architectural floor. ISA Overhead: 217.4%. Node Regime: WALL. Dennard Amp. G: 1.11×. This is the instrument at its most revealing: the most powerful GPU in the world is operating in CRITICAL condition at this stress point — below its own architectural floor.

 © Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.


TEMPERATURE ANALYSIS

Tj (°C)	SCAPE Index	vs anchor	Note
20°C	267×	−42%	
40°C	301.8×	−34.4%	← target Tj
55°C	329.1×	−28.5%	
70°C	357.5×	−22.3%	
75°C	367.2×	−20.2%	
85°C	387×	−15.9%	
95°C	407.3×	−11.5%	
100°C	417.6×	−9.2%	
105°C	428×	−7%	TjMax

Screen 16 — B200 Temperature Response. 42% Better Efficiency With Liquid Cooling to 40°C.

The B200 temperature analysis table is striking. At 20°C (aggressive liquid cooling): SCAPE 267× — 42% improvement from the 460× baseline. At 40°C (target Tj with best-practice liquid cooling): SCAPE 302×, –34.4% from anchor. At 75°C (nominal): 367×. At 105°C: 428×. The GPU architecture's n = 1.853 temperature sensitivity is the highest in the dataset. This means liquid cooling delivers more return on a B200 than on any other chip tracked. A hyperscaler running 10,000 B200 units who invests in sub-40°C liquid cooling infrastructure achieves the equivalent of 42% more compute without buying new hardware. The Dennard Transition Analysis below the temperature table confirms: Prior SCAPE 1,181× (8nm) → Current 460× (4nm). Step rate: 61.1%. B200 is in SCALING regime — 84.8% of the Dennard baseline delivered on this transition. The new Blackwell die is genuinely delivering, even at the stress operating point.

 © Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

PUBLISHED INPUTS & DERIVED VALUES

Parameter	Value	Source
TDP (sustained all-core)	180 W	PUBLISHED
Transistor count	20.8 B	PUBLISHED
All-core sustained frequency	5 GHz	PUBLISHED
Junction temperature anchor	120°C	PUBLISHED
Process node	4nm	PUBLISHED
Switching energy / transistor (E_sw)	0.001731 fJ	DERIVED — IAM
Thermodynamic floor (E_min)	0.000004 fJ	DERIVED — IAM · k_B T ln2
SCAPE Index A = E_sw / E_min	460×	DERIVED — IAM
Architectural floor (ISA class)	1000×	DERIVED — IAMPerformance
Distance above floor	0.46× (CRITICAL)	DERIVED
ISA overhead fraction	217.4%	irreducible by node shrink
Node-accessible improvement	−117.4%	recoverable by engineering
Node scaling regime	WALL	0 steps remaining at current rate
Dennard Amplifier g	1.11×	Dennard promised 72% · delivered 64.9%

Screen 17 — B200 Derived Values. CRITICAL Status. The Floor Has Been Crossed.

The full derived values table makes the CRITICAL status unambiguous. TDP: 180W. Transistors: 20.8B. Frequency: 5 GHz. Junction temperature: 120°C. Switching energy per transistor: 0.001731 fJ (elevated by temperature). SCAPE Index A: 460×. Architectural floor (ISA class): 1,000× — derived by IAMPerformance, not published by NVIDIA. Distance above floor: 0.46× (CRITICAL) — the B200 at 120°C is operating BELOW the GPU architectural floor. ISA overhead fraction: 217.4% — meaning the architecture-locked portion exceeds the total index, a physical impossibility at nominal operating conditions that becomes real when junction temperature is pushed into extreme ranges. Node-accessible improvement: −117.4% — negative, confirming the chip has crossed its structural boundary at this operating point. Node scaling regime: WALL. 0 steps remaining at current rate. This screen does not appear in any NVIDIA benchmark or whitepaper. It takes the instrument three seconds from public specifications.

© Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

FORWARD TRANSLATION — TARGET SCAPE INDEX ANALYSIS

CURRENT SCAPE	TARGET SCAPE	GAP FACTOR	STEPS REQUIRED
460×	200×	2.3×	0.8
starting point	entered target	improvement required	at current class rate

Required TDP (at same trans/freq)	78.3 W	to achieve target A — holding transistors and frequency constant
Required transistor count (at same TDP/freq)	47.84 B	to achieve target A — holding TDP and frequency constant
Steps at current class rate	0.8 node steps	trajectory to target at validated improvement rate

ISA ESCAPE SCENARIO — MODERN ISA (POST-DENNARD)

CURRENT ARCHITECTURE		AFTER ISA CHANGE		NET IMPACT	
Floor	1000×	New floor	200×	Floor change	1000× → 200×
Ratio	0.46×	New ratio	2.3×	Ratio change	0.46× → 2.3×
Concern	CRITICAL	New concern	HIGH	Concern shift	CRITICAL → HIGH
ISA overhead	217.4%	New ISA overhead	43.5%	ISA overhead	217.4% → 43.5%

CHIPLET SEPARATION SCENARIO — I/O + COMPUTE DIE SPLIT

CURRENT SCAPE	AFTER CHIPLET SPLIT	NEW FLOOR RATIO
460×	414×	0.41×
monolithic	10% system gain est.	vs same architectural floor

SCAPE Engineering Diagnostic · Informational Actualization Model (IAM) · Mahaffey (2026)
All derived values computed from published inputs only. IAM reference floor and architectural floors are IAMPerformance-derived — not published by manufacturers.
Patent Applications 64/012,720 and 64/014,568. All commercial terms through legal counsel.

Screen 18 — B200 Forward Translation + Escape Scenarios. What Would It Take to Reach 200×?

The forward translation answers the precise engineering question: what specifications would be required to push B200 efficiency to SCAPE 200×? From 460× to 200×: a 2.3× improvement required, achievable in 0.8 node steps at the current class improvement rate. Required TDP at current transistors and frequency: 78.3W — a 56% TDP reduction. Required transistor count at current TDP and frequency: 47.84B — more than doubling the transistor count. These are not projections. They are constraint equations. The ISA Escape scenario shows what a modern ISA would do: floor changes from 1,000× to 200×, ratio improves from 0.46× to 2.3×, concern moves from CRITICAL to HIGH, ISA overhead drops from 217.4% to 43.5%. The Chiplet Separation scenario shows a 10% system gain: 460× to 414×, floor ratio 0.41×. A hyperscaler making a 50,000-unit B200 procurement decision now has, from three published numbers in three seconds, the precise answer to: "what does it take to reach our target efficiency, and which path costs least?" That answer does not exist anywhere else. It exists here because the physics already knew it.

GLOSSARY — WHAT THESE NUMBERS MEAN

SCAPE Index

The Semiconductor Analytical and Performance Engine index. A normalized dimensionless efficiency metric expressing how far above the physical minimum any chip operates. Derived from three published numbers: TDP in watts, transistor count, and clock frequency. Lower = more efficient. Every chip from every architecture is directly comparable on one scale. A chip at 100× operates 100 times above the physical minimum for its operating temperature. A chip at 600× operates 600 times above it. The gap is architecture, not marketing.

Published inputs (orange in tables)

The three numbers the SCAPE framework reads from published datasheets: TDP (thermal design power in watts), transistor count (in billions), and base clock frequency (GHz). No other data is used. No internal measurements. No proprietary access. These three numbers are available on every manufacturer's product page.

Derived outputs (teal in tables)

Every value in the teal columns is computed from the three orange published inputs by the IAM framework. This includes: SCAPE index, architecture parameter n, temperature predictions at 40°C and 105°C, illustrative trajectory scenarios, floor ratio, distance to floor, ISA overhead fraction, and Dennard Amplifier g. None of these values are published by manufacturers. All are produced by the SCAPE framework. Methodology protected under Patent Applications 64/012,720 and 64/014,568.

Architecture parameter n

A value derived from 20 years of published chip data that characterizes how strongly each architecture class responds to changes in junction temperature. Intel (n=0.936): responds strongly to temperature — colder running produces larger efficiency gains, but sustained hot running degrades faster. AMD (n=0.340): temperature-stable — less gain from cooling, but much slower degradation under load. The 82°C crossover — where AMD and Intel efficiency are exactly equal — is a direct numerical consequence of their respective n values applied to published nominal specs.

IAM Reference Floor

The physical lower boundary all semiconductor switching must satisfy at a given operating temperature — derived from the IAM information-theoretic framework. It does not change with process node, architecture, foundry, or manufacturer. No chip has ever operated at this boundary. No chip can. The SCAPE index measures how far above this boundary each chip operates. Methodology protected under Patent Applications 64/012,720 and 64/014,568.

Structural efficiency floor

The physical minimum SCAPE index achievable within each architecture class — the hard lower bound set by the cost of information itself at operating temperature. Apple Silicon floor: ~15×. x86 floor: ~70×. Qualcomm ARM floor: ~35×. NVIDIA GPU: ~250×. Apple and AMD share the same TSMC foundry at the same node. The gap between their floors — a factor of 4-5× — is pure architecture: x86 ISA overhead, speculative execution, memory access patterns, pipeline complexity. The floor is IAMPerformance-derived. Not published by any manufacturer.

ISA Overhead Fraction

The percentage of a chip's current SCAPE index that is architecturally irreducible — the fraction that cannot be closed by any process node improvement. Computed as: architectural floor ÷ current SCAPE index. AMD at 576×: 70÷576 = 12.2% ISA overhead, 87.8% node-accessible. Apple M3 at 65×: 15÷65 = 23.1% ISA overhead, 76.9% node-accessible. As a chip approaches its floor, this fraction rises — until at the floor, 100% of remaining inefficiency is architectural and no node shrink delivers further gains.

Node Scaling Regime

A status flag showing where each architecture sits relative to its floor. FREE: floor ratio ≥ 4.0× — substantial runway, node shrinks still delivering. APPROACHING: floor ratio 2.0–4.0× — wall registering, 1-3 steps remaining. WALL: floor ratio < 2.0× — at or near ceiling, architectural change required. All current chips in this dataset show FREE — none are near their structural floor. The Dennard wall arrives before any architecture reaches WALL status.

Dennard Amplifier (g)

A quantified measure of how much harder each node step has become versus the pre-2005 free-scaling era. Computed as: Dennard baseline rate (72%) ÷ observed improvement rate. Intel 285K (70.7% step): g = 1.02× — nearly tracking Dennard. AMD 9950X (33.2% step): g = 2.17× — each node step requires 2.17× more engineering effort. Qualcomm (5.3% step): g = 13.6× — post-Dennard regime confirmed. NVIDIA (8.1% avg): g = 8.9× — GPU efficiency drifting far from Dennard trajectory.

Dennard scaling

The 1974 prediction by R. Dennard that shrinking transistors would deliver proportional efficiency gains at every node — power density constant, efficiency compounding. The SCAPE framework identified the breakdown of this prediction from 2003-2005 data alone, before the breakdown was widely recognized by the industry. The shaded region in the horserace chart is the efficiency Dennard promised but never arrived.

Structural efficiency floor

See above. The next Dennard-equivalent wall arrives before any architecture approaches its floor. The wall tells you when the current engineering path is exhausted. The floor tells you how much physical headroom remains for the next approach. The distance between them is where the next architectural revolution must happen.

Floor ratio

Current SCAPE index divided by the structural efficiency floor for that architecture class. AMD at 576÷70 = 8.23× above floor — substantial headroom. Apple M3 at 65÷15 = 4.33× above floor — closer to its floor than any x86 chip, not because Apple is running out of room, but because Apple's floor is so much lower. Floor ratio is IAMPerformance-derived. Not published by manufacturer.

Junction temperature (Tj)

The operating temperature of transistor junctions inside the chip die. Published TDP specifications reference a nominal junction temperature, typically 75°C. Under sustained AI inference, database serving, and continuous compute, real junction temperatures in data centers routinely reach 80-105°C. Intel and AMD have different temperature response profiles (n values) that determine who leads at each temperature. The crossover is 82°C.

© Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

TjMax

The maximum rated junction temperature before thermal throttling reduces clock speed. AMD 9950X TjMax: 95°C. Intel Core Ultra 9 285K TjMax: 105°C. At TjMax under sustained load, AMD is 3.7% more efficient than Intel on the SCAPE index. For hyperscale AI inference — chips running sustained heavy load — TjMax efficiency determines operational cost, not nominal bench efficiency.

GPU architecture — what SCAPE measures and does not

The SCAPE index measures per-transistor switching efficiency — not throughput. GPU architecture makes a deliberate trade: sacrifice per-transistor efficiency for massive parallelism. H100→H200 was a memory upgrade on the same die — the 7.6% SCAPE improvement reflects that precisely. The Blackwell successor will be the first genuine new compute die since Hopper. The framework measures that step the moment specifications publish.

What the framework measures — and does not

The SCAPE framework measures: physical floor, distance above floor, direction of travel, and regime-change signal. It does not project how companies will engineer their way through the next wall. It does not predict market outcomes. It does not assume any future improvement rate. The SCAPE index tells you where each architecture is today, and which direction it is heading.

Trajectory charts — illustrative scenarios

The trajectory tables show what would happen if each architecture's most recent observed improvement step continued at a constant rate. This is an illustrative constant-rate scenario, not a forward projection. Improvement rates vary significantly — Intel's published steps range from 4.9% to 70.7%. AMD had two regressions in its last four steps. The trajectory charts show current position and direction of travel. They are not a forecast.

Three-Component Efficiency Decomposition

A breakdown of every chip's total efficiency gap above the IAM reference floor into three independent portions. Component 1 — Universal floor: the IAM-derived physical minimum at operating temperature. Same for every transistor ever built at a given junction temperature. Cannot be engineered around. Component 2 — Architecture-locked (A_isa): ISA overhead above the universal floor. Set by the instruction set architecture. AMD and Intel carry ~70× of ISA-locked overhead. Apple Silicon carries ~15×. Only ISA redesign moves this component. Component 3 — Engineering-accessible (A_node): the gap between the current SCAPE index and the architectural floor. This is the portion that every node shrink, chiplet strategy, and microarchitectural improvement addresses. It is the engineering battleground. AMD example: 576× total. 70× locked (12.2%). 506× accessible (87.8%). Apple M3 example: 65× total. 15× locked (23.1%). 50× accessible (76.9%). IAMPerformance-derived. Patent Applications 64/012,720 and 64/014,568.

Architecture-locked efficiency

The portion of a chip's SCAPE index that cannot be improved by any foundry or process advance. Set by the ISA: x86 carries ~70× of ISA-locked cost, ARM carries ~15-35×, GPU ~250×. AMD and Apple share the same TSMC foundry at the same 3nm node. AMD carries 70× of locked overhead. Apple carries 15×. The 55× difference between those locked components is the entire x86-to-ARM ISA gap, expressed in SCAPE units. It cannot be closed by TSMC. It requires ISA redesign.

Engineering-accessible efficiency

The portion of a chip's SCAPE index that node shrinks, chiplets, and micro-architecture improvements can address. For AMD 9950X: 506× (87.8% of the total 576×). For Apple M3: 50× (76.9% of 65×). Every TSMC roadmap advance, every Zen generation improvement, every Apple Silicon efficiency tuning works within this accessible portion. The fundamental insight: fab investments can only close the engineering-accessible gap. ISA overhead is not a fab problem — it is an architecture problem.

Prediction P001 — Intel 2nm step below 40%

Filed March 29, 2026. Intel's Arrow Lake delivered a 70.7% step — the largest in 20 years of published Intel data. At that rate held constant, the x86 floor at 70× is reached in 1.7 node steps. That is physically impossible. The same floor-derived signal identified the 2005 Dennard breakdown from 2003-2005 data. The 2nm data point tests whether it identifies the next transition with equal precision. Testable: when Intel's first 2nm product publishes primary-source specifications.

Prediction P002 — AMD 2nm step below 25%

Filed March 29, 2026. AMD Zen 5 delivered 33.2% at 4nm. The floor-derived signal predicts the next step will be materially smaller as AMD closes on the x86 floor. AMD Zen 6 on TSMC 2nm (expected H2 2026) is the testable data point. A step above 25% would refute this prediction.

Prediction P003 — Apple M5 regression watch (NOT CONFIRMED)

Filed March 29, 2026. Result: NOT CONFIRMED. Apple M5 published October 15, 2025: TSMC N3P, 28B transistors, ~27W TDP, 4.4 GHz. SCAPE = 65.8× — a 3.6% improvement over M4 (68.2×). The M3→M4 regression streak is broken. The regression watch is resolved. The structural gap to x86 remains at 8.7×.

Prediction P004 — Blackwell step exceeds H100→H200 (CONFIRMED)

Filed March 29, 2026. Result: CONFIRMED. NVIDIA B200 SXM published 2025: TSMC 4NP, 208B transistors (dual-die), 1,000W TDP. SCAPE = 641×. H200→B200 improvement: 51.6% — 6.8× larger than the H100→H200 step of 7.6%. The instrument correctly separated the memory upgrade (7.6%) from the compute advance (51.6%). Track record: two consecutive calls on NVIDIA trajectory, both correct.

Apple Silicon — structural efficiency class

Apple's M-series processors occupy a structurally distinct efficiency class: M1 (2020, 95×) through M5 (2025, 66×) — all below 100× on a scale where Intel's best is 570× and AMD's best is 576×. The gap has not narrowed since M1 launched in November 2020. Apple and AMD share the same TSMC foundry at the same 3nm process node. The 8.7× gap is architectural: ARM ISA, unified memory, full system integration. It cannot be closed by foundry improvement.

NVIDIA Blackwell B200 — first genuine GPU compute step since Hopper

NVIDIA B200 SXM (2025): 208 billion transistors (dual-die: 2× GB202 at 104B each), TSMC 4NP process, 1,000W TDP, ~2.25 GHz SM boost. SCAPE = 641× — 51.6% improvement over H200 (1,326×). First new compute die since Hopper in 2022. Prediction P004 confirmed. The dual-die design breaks through the reticle limit by connecting two GB202 dies via a 10 TB/s NV-HBI interconnect. At 1,000W TDP, B200 deployments require liquid cooling.

© Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

AMD Zen 6 and Intel Nova Lake — pending data points

As of April 2026, both AMD Zen 6 (TSMC 2nm, expected H2 2026) and Intel Nova Lake (2nm-class, expected late 2026 / 2027) have no published primary-source specifications. No SCAPE score can be computed without published TDP, transistor count, and clock frequency. These are the live testable events for Predictions P001 and P002. SCAPE will compute scores the moment primary-source specifications are published.

Dennard Baseline (72%)

The per-node efficiency improvement rate observed from 1974 to 2005 — the "free-scaling" era. R. Dennard's 1974 prediction: shrinking transistors would produce proportional efficiency gains at each node shrink, with power density held constant. For three decades, the semiconductor industry delivered approximately 72% improvement per node generation. After 2005, power density could no longer be held constant — the breakdown began. SCAPE uses 72% as the Dennard baseline for computing g.

Forward translation

The reverse problem: given a target SCAPE index, what published specifications (TDP, transistors, or clock frequency) would achieve it, holding the other two constant? The SCAPE Engineering Diagnostic tool computes forward translation for any target index. Example: Intel 285K at 570× targeting 300× requires either a 46% TDP reduction, or a 91% transistor count increase at current TDP, or a 34% frequency reduction. Not a projection — a constraint equation directly from the framework.

IAM's Law

A proprietary first-principles derivation from the Informational Actualization Model. Establishes a universal performance floor for any information processing architecture — the minimum below which no chip of that class can operate at a given temperature. The SCAPE index measures how far above this floor each chip actually operates. Derivation protected under Patent Applications 64/012,720 and 64/014,568.

© Heath W. Mahaffey · Patents pending 64/012,720 & 64/014,568 · Not investment advice. · For analytical reference only.

A FINAL NOTE

The semiconductor industry is the most precisely engineered enterprise in human history. The teams working on 2nm silicon are doing genuinely hard things — patterning features at dimensions smaller than the wavelength of visible light, managing leakage currents measured in femtoamperes, sustaining clock frequencies that cycle faster than most scientific instruments can observe. Every engineer on every one of these programs deserves better than guesswork about where their architecture stands and what the physics actually allows.

What we have been fortunate enough to find is a set of physical relationships that answer those questions from first principles. Not from fitted curves. Not from historical analogies. From the underlying physics of what it actually costs to flip a bit — and what it means that cost cannot be engineered away. Those relationships let us say, with confidence, which architecture classes have room to grow and which have reached their structural ceiling. Which engineering lever actually moves the SCAPE index for a given architecture. Whether a published improvement step can possibly continue for another node generation or whether the floor makes it physically impossible. All from three published numbers, in seconds.

The depth of that derivation matters and deserves to be stated plainly. The framework is grounded in the minimum energy cost of any irreversible information event — the same physical law that governs processes from the nanoscale to the macroscale through one expression, not by analogy. The architectural floor values are not fitted to hardware data. They come from first-principles derivation applied to the dominant switching constraints of each architecture class — a derivation that applies uniformly to every CMOS architecture ever manufactured. The temperature sensitivity parameter n is not assumed. It is derived from the dominant loss mechanism of each architecture class and validated against 20 years of published chip data across five manufacturers. That was not the goal when the derivation began. It is what the physics produced.

The value is not the ranking table. The value is the engineering decision made in a day instead of a quarter, because the framework already showed what the physics allows. The cooling investment redirected before the procurement order arrives, because the crossover temperature told you which regime your workload runs in. The 2nm architecture decision informed before the tape-out, because the floor-derived signal showed whether the current improvement rate can physically continue for another node. Every one of those recovered months is a month spent actually advancing the engineering rather than rediscovering a constraint the physics already knew.

The floor-derived signal has called one transition correctly from two data points in 20 years of published data. The 2nm data point — when AMD and Intel publish their first 2nm primary-source specifications — is the live test of whether it calls the next one. The prediction is in the record. The date is on the page. When that data publishes, the framework either confirms or it does not. That is what a physical instrument looks like over time.

IAMPerformance is an independent research initiative. There is no institutional affiliation, no investor agenda, and no stake in any of the architectures tracked in this publication. The sole mission is to continue refining SCAPE — expanding the architecture classes it covers, sharpening its predictions as new data publishes, and building it into a tool that the engineering teams themselves can use to make better decisions faster. Every team making 2nm architecture decisions deserves a navigation instrument that tells them where they actually are, not where they hoped to be. This publication is one step toward that.

REFERENCES

[1] Advanced Micro Devices. AMD Ryzen 9 9950X Product Specifications. amd.com, 2024. 4nm TSMC Zen5, 20.6B transistors, 170W TDP.

[2] Alcorn, P. Intel Core Ultra 9 285K Review. Tom's Hardware, October 24, 2024. TSMC N3B, 17.8B transistors, 125W TDP.

[3] Apple Inc. Apple Introduces M4 Chip. apple.com/newsroom, May 2024. TSMC N3E, 28B transistors, 28W TDP, 4.4 GHz.

[4] Apple Inc. Apple Unveils M3. apple.com/newsroom, October 2023. TSMC N3B, 25B transistors, 22W TDP, 4.05 GHz.

[5] Apple Inc. Apple Unleashes M1. apple.com/newsroom, November 2020. TSMC 5nm, 16B transistors, 15W TDP, 3.2 GHz.

[6] Cutress, I. AMD Ryzen 9 3900X Review. AnandTech, July 2019. 7nm TSMC, 9.89B transistors, 105W TDP, 3.8 GHz.

[7] Dennard, R.H. et al. Design of ion-implanted MOSFET's with very small physical dimensions. IEEE JSSC, 1974.

[8] Intel Corporation. Intel Core Ultra 9 Processor 285K Specifications. ark.intel.com. TSMC N3B, 17.8B transistors, 125W.

[9] Intel Corporation. Intel Core i9 processor family specifications. ark.intel.com, 2008-2024.

[10] Landauer, R. Irreversibility and Heat Generation in the Computing Process. IBM Journal of Research and Development, 5(3):183-191, 1961.

[11] NVIDIA Corporation. NVIDIA Hopper Architecture In-Depth. developer.nvidia.com, March 2022. TSMC N4, 80B transistors, 700W.

[12] NVIDIA Corporation. NVIDIA H200 SXM Product Specifications. nvidia.com, 2024. TSMC N4, 80.0B transistors (identical GH100 die as H100), 700W.

[13] Qualcomm Technologies. Snapdragon X Elite Platform Product Brief. qualcomm.com, 2024. TSMC 4nm, 20B transistors, 80W.

[14] Vortez. AMD Ryzen 9 9950X and 9900X Review. vortez.net, August 2024. 4nm TSMC Zen5, 20.6B total, 170W, 4.3 GHz.

All analysis from publicly available published sources cited above. Methodology protected under Patent Applications 64/012,720 and 64/014,568. Inquiries: hmahaffeyges@gmail.com. All commercial terms through legal counsel. Heath W. Mahaffey, Independent Researcher, Entiat, WA, April 2026.