

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2024.Doi Number

Datapath Width Design Space of 32-Bit RISC-V Processors for 2D Semiconductor Logic

Naoki Hayase¹

¹Center for EUV lithography, LASTI, University of Hyogo, NewSUBARU synchrotron radiation facility, Ako-gun Kamigori Kouto 1-1-2, Hyogo 678-1205, Japan

Corresponding author: N. Hayase (e-mail: hayase@lasti.u-hyogo.ac.jp).

This work was partially supported by Hyogo Earthquake Memorial 21st Century Research Institute, Hyogo Foundation for Science and Technology, Tanaka Kikinzoku Memorial Foundation, Takahashi Industrial and Economic Research Foundation, and Foundation of Kinoshita Memorial Enterprise.

ABSTRACT The first 32-bit RISC-V processor on a two-dimensional semiconductor, a 5,931-transistor MoS₂ chip, is bit-serial to fit a device-count-limited process. This work determines where the optimal datapath width departs from bit-serial execution. A digit-serial RV32I core of 1 to 32 bits is synthesized onto a model of that process, with every mapping and pipelining step formally proved. Serialization saves only about 2.5× in transistors, not 32×, because control and state do not scale with width. At the fabricated 1 kHz clock a 32-bit core thus delivers 13.5× the throughput at 5.2-5.5× lower energy per instruction. The energy ordering reverses when the throughput is fixed and the core cannot be power-gated. At the 2025 device parameters only yield selects bit-serial execution, above about 250 defects per million transistors if the 17-stage depth rule of the process is a delay budget, or 75 if it is a hard constraint. Mapper quality moves these values by up to 2.3×. Under the Poisson statistics every design yields below 1% at the published defect density, and a 10% die yield requires a 3.6-5.4× lower density, less with defect clustering. With an on-die register file, serialization saves only 11-13% of the transistors, and bit-serial execution is never the widest feasible design. Closed-form boundaries locate other technologies on this width map.

INDEX TERMS bit-serial processing, datapath width, defect density, design space exploration, monolithic three-dimensional integration, parallel-prefix adder, RISC-V, two-dimensional semiconductors.

I. INTRODUCTION

In 2025, a RISC-V microprocessor built from 5,931 MoS₂ transistors was reported, made feasible by serializing a 32-bit architecture [1]. A bit-serial datapath processes one bit per clock cycle and requires 32 serial steps to traverse a 32-bit word. The premise of that design is that serialization minimizes transistor count and static power, which are the binding constraints of a device-count-limited NMOS technology. The chip demonstrates the potential of two-dimensional (2D) integrated circuits.

2D semiconductors are maturing toward complementary field-effect transistor (CFET) logic and monolithic three-dimensional (M3D) integration [2-16], while the physical parameters that determine processor architecture, including device delay, interconnect load, integration density, leakage, and defect density, continue to develop. As a result, the optimal architecture for a 2D technology depends on its position in this parameter space. The key question is where the optimal datapath width shifts from bit-serial to parallel.

An evaluation of this problem requires four elements: (i) a family of cores that differ only in datapath width, making

width a controlled variable, (ii) a synthesis flow whose area, delay, and power models are tied to a fabricated 2D process rather than a silicon standard-cell library, (iii) verification sufficient to support architectural conclusions without new silicon, and (iv) a systematic sweep of physical parameters with boundaries that can be evaluated for other technologies [1,17]. This work provides all four through M2D, a Python-based RISC-V core derived from Minerva [18], and a design-space exploration that evaluates the datapath width within a defined set of candidate designs, an objective function, and a fixed implementation flow.

Once its inputs are known, the comparison follows from the static energy per instruction at a fixed clock and the Poisson die yield [19,20]. With static power dominant, the energy per instruction at a fixed clock is proportional to the transistor count T times the cycles per instruction (CPI), and a Poisson yield model converts differences in T into defect-density boundaries in closed form. The contribution of this work lies in the inputs, which only synthesis, verification, and depth legalization on the process model can supply. The first input already changes the picture: serialization saves 2.45-2.60× in

transistors, not the $32\times$ that a datapath-only estimate suggests, because the control, the program counter, the instruction register, and the sequential state do not scale with width (Section IV-A).

The primary scenario is the configuration that Ref. [1] fabricated, with the register file off the die (configuration A), at its 1 kHz clock, under fixed-clock or race-to-sleep operation and the Poisson defect statistics, with the expected throughput per fabricated die as the metric. Ref. [1] does not state whether its 17-stage limit is a delay budget or a hard cascade constraint, so every primary result is given as a pair for the two readings (Section V). Table 1 collects these results. The secondary cases are maximum-frequency operation and iso-throughput operation without power gating (Section IV-C), a bandwidth-limited memory (Section IV-D), the per-wafer-area metric (Section VI-B), defect clustering (Section VI-C and Appendix D), and the register file on the die (configuration B, Section VI-E). Under iso-throughput operation without power gating the energy ordering reverses in favor of bit-serial execution, and both orderings are reported.

(1) Depth compliance by construction (Section V). The fabricated chip obeys a limit of 17 stages between flip-flops that no synthesized netlist of this family meets. Sub-cycle pipelining built from the process cell library alone produces, at every design point, a netlist with no path longer than 17 stages, formally proved equivalent to the original over one architectural cycle and executed at gate level. It costs 2 to 8 sub-cycles per architectural cycle and 5,026 to 72,769 pipeline transistors per design. Under this hard constraint 16 bits with prefix carry becomes the energy optimum, and prefix carry becomes a prerequisite for a 32-bit datapath that beats bit-serial execution in energy.

(2) Closed-form selection boundaries (Section VI). Feasibility in device delay and wire load and Poisson yield in defect density are given in closed form with tabulated coefficients, so another process is located without rerunning the flow (Section VII-B). The crossover from bit-serial to parallel is a defect density of about 250 or 75 defects per million transistors, depending on the reading of the 17-stage rule, and mapper quality moves it by a factor of up to 2.3. Under the Poisson statistics every design yields below 1% at the published defect density, and a 10% die yield requires a $3.6\times$ (delay budget) or $5.4\times$ (cascade) lower density. Defect clustering lowers these targets (Section VI-C and Appendix D).

(3) The register file on the die (Section VI-E). With the register file integrated, proved, and executed at gate level, the serial-to-parallel transistor ratio collapses to 1.13-1.15, and bit-serial execution is never the widest feasible design. Under a 10% die-yield constraint with the Poisson statistics it survives only in windows narrower than 2 defects per million transistors, about $25\times$ (delay budget) or $36\times$ (cascade) below the published density.

TABLE 1. PRIMARY RESULTS AT THE 2025 PROCESS POINT FOR BOTH READINGS OF THE 17-STAGE RULE

quantity	delay budget	cascade	Sec.
transistors, $W = 32$ over $W = 1$	$2.45\times$ (r), $2.60\times$ (BK)	$3.61\times$ (BK)	IV-A, V
throughput, $W = 32$ over $W = 1$	$13.5\times$	$6.76\times$ (BK)	IV-C, V
energy per instruction, $W = 1$ over $W = 32$	$5.5\times$ (r), $5.2\times$ (BK)	$1.87\times$ (BK)	IV-C, V
energy optimum	$W = 32$ r	$W = 16$ BK	IV-C, V
bit-serial optimal above (ppm/T)	249	75	VI-B
first parallel width below it	$W = 8$ r	$W = 8$ BK	VI-B
range of that density over mapper quality (ppm/T)	154-441	33-133	IV-A, VI-C
density reduction for a 10% die yield	$3.6\times$	$5.4\times$	VI-C
same, register file on the die	$25\times$	$36\times$	VI-E

Configuration A, fixed-clock or race-to-sleep operation, zero-wait memory, the Poisson statistics, and the expected throughput per fabricated die, at 1 kHz (delay budget) or at the compliant clock of Section V (cascade). The range runs from the fabricated bit-serial chip in place of $W = 1$ (lower end) to a mapper $1.77\times$ better at every width (upper end). The last row uses configuration B. r: ripple carry, BK: Brent-Kung prefix carry.

II. RELATED WORK

TABLE 2. SCOPE OF THIS WORK AGAINST THE FABRICATED 2D PROCESSORS AND THE BENCHMARKING METHODOLOGY

work	technology, logic	devices	datapath, ISA	clock	establishes
[2] (2017)	MoS ₂ NMOS	115	1-bit custom	kHz	first 2D microprocessor
[1] (2025)	MoS ₂ E/D NMOS, 25 cells	5,931	bit-serial RV32I, register file off the die	1 kHz	first 32-bit 2D processor, process model of this work
[3] (2025)	MoS ₂ /WSe ₂ complementary	> 2,000	one-instruction-set computer	≤ 25 kHz	first complementary 2D processor
[17]	device-agnostic	-	fixed benchmark circuits	-	device-level benchmarking
[21]	FPGA soft core	-	bit-serial RV32I	-	microarchitecture of [1] and M2D
this work	model of [1]	10,474-212,955	W = 1-32, ripple and prefix, register file off and on die	1 kHz, $f_{\max}$	proved 17-stage-compliant netlists, closed-form width boundaries, register file on die

A. PROCESSORS ON 2D SEMICONDUCTORS

Table 2 shows this work against 2D processors and against the benchmarking methodology of Ref. [17]. The SERV core

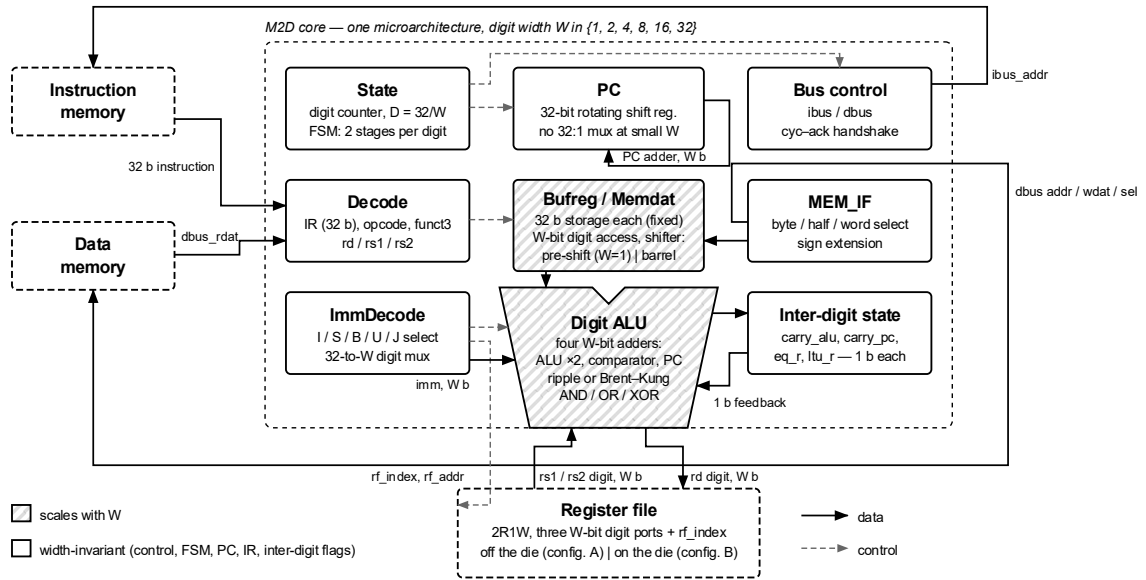


FIGURE 1. The M2D core. Only the hatched blocks scale with W : the digit access and shifter of the buffer and data registers, whose 32-bit storage is fixed, and the digit ALU with four W -bit adders (two ALU adders, the comparator, and the program-counter adder, which closes the PC update loop). The register file and both memories are external in configuration A, as in the fabricated chip, and configuration B (Section VI-E) moves the register file onto the die.

[21] revived bit-serial RISC-V and is the ancestor of RV32-WUJI [1] and of M2D. Width-scalable cores include QERV [22] (4 bits), FazyRV [23] (1 to 8 bits), FlexiBits [24] (1, 4, and 8 bits on flexible IGZO, with 2.65-3.50 $\times$ better energy efficiency per workload execution), and SAILOR [25] (1 to 32 bits), all in FPGA, silicon, or flexible metal-oxide technologies [26]. None addresses a ratioed 2D NMOS process with a 17-stage cascade limit, defect density as a selection axis, or the register file on the die.

B. PARALLEL-PREFIX ADDITION

On a 2D platform (4.06 μ s intrinsic stage delay, 2.5 pF per pin) logic depth is expensive and wire capacitance cheap. Brent-Kung (BK) [27] minimizes fan-out and wiring at the price of two extra levels, the right bias for a fan-out limit of four, whereas Kogge-Stone [28] and Sklansky [29] trade those resources for depth [30].

C. WIRES, RENT'S RULE, M3D, AND CFETS

Average wirelength follows Donath [31] with Rent parameters [32,33]. Monolithic 3D integration was demonstrated for emerging devices in N3XT [13,34], analyzed at system level in [35], and extended to 2D materials [14,15] and to complementary 2D CFET stacks [16]. This work uses these only as swept parameter axes (tier count k and static power p_s).

D. YIELD-DRIVEN ARCHITECTURE SELECTION

FlexiCores [36] measured die yield against width on flexible IGZO (81% at 4 bits, lower at 8 bits). This work makes

defect density an axis of the map, anchored by the yields of [1], which also constrain the negative-binomial clustering parameter [20].

E. LOGIC SYNTHESIS AND TECHNOLOGY MAPPING

No 2D process design kit accepted by commercial tools was available, so the mapper, analyzer, and rule checks were written for this study, with yosys [37] lowering the register-transfer level. The covering is DAG matching [38], the cross-check mapper is ABC [39], and the equivalence proofs use structural hashing and SAT [40]. The depth-compliant netlists of Section V relate to retiming [41] but keep the cycle counts. Switching activity, the standard vehicle for power estimation [42], is measured during program execution.

III. METHODOLOGY

A. WIDTH-PARAMETERIZED CORE

Fig. 1 shows M2D, a width-parameterized digit-serial [43] RV32I core that covers $W = 1, 2, 4, 8, 16$, and 32 and two carry schemes in one microarchitecture. Only the hatched blocks scale with W , and the register file and both memories are external, as in Ref. [1].

M2D derives from the Minerva soft processor [18] and implements RV32I [44] without FENCE, ECALL, and the CSR instructions, as in [1]. A finite-state machine of two stages per digit processes $D = 32/W$ digits per word operation, with every width-dependent state parameterized. The program counter is a rotating shift register, which removes the 32-to-1 read multiplexer and the 32-way write

demultiplexer at small W . The register file has three W -bit data ports plus index lines.

One-stage instructions (ALU-immediate, ALU-register, LUI, AUIPC, JAL) take $D + 2$ cycles, and two-stage instructions (branches, comparisons, shifts, and JALR, which needs an addition before the fetch is redirected) take $2D + 2$. Loads take $2D + 3$ and stores $D + 3$. At $W = 1$, right shifts add the shift amount because that point uses the transistor-minimal pre-shift scheme of [1] instead of a barrel shifter. The class costs are

$$\begin{aligned} c_1 &= D + 2, & c_2 &= 2D + 2, \\ c_{ld} &= 2D + 3, & c_{st} &= D + 3 \end{aligned} \quad (1)$$

The cycle count of a program follows from the retired class counts n_c and the summed right-shift amounts:

$$N_{cyc} = \sum_c n_c c_c(W) + [W = 1] \sum_r sh_r \quad (2)$$

where the bracket equals 1 at $W = 1$ and 0 otherwise. Appendix A validates both at zero tolerance at every design point. The Amaranth HDL [45] source is about 700 lines.

B. ADDERS

Both carry schemes are structural gate-level netlists, a ripple chain of W full adders (carry depth W) and a BK tree [27] with $2\lceil \log_2 W \rceil - 1$ carry levels, so that yosys cannot substitute its own carry-lookahead structure. Neither is fan-out-legal by construction, so both pass the same buffering pass, whose transistor cost is included in every number.

C. SYNTHESIS FLOW ONTO THE FABRICATED-PROCESS MODEL

The Python-based flow uses yosys [37] to lower the Amaranth RTLIL to primitive gates and maps them to the 2D_NMOS_lib of Ref. [1]: 25 enhancement/depletion NMOS cells at $V_{dd} = 4$ V, from INV through the NAND, NOR, AOI, and OAI families to MUX2, XOR2, and an 18-transistor flip-flop. Seven cells with a published noise margin below 5% of V_{dd} are excluded. A structural DAG-covering mapper [38] covers the gates, and inverter trees of even level count legalize fan-out to at most four. Static timing uses the linear stage-delay model

$$t = t_0 + s C_{load} \quad (3)$$

with $t_0 = 4.06$ μ s, $s = 0.6$ μ s/pF, and $C_{pin} = 2.5$ pF per pin. A stage driving F pins sees $C_{pin}F$, saturating at $F_{max} = 4$, which does not affect legalized netlists. The path delay includes the clock-to-output stage of the launching flip-flop:

$$f_{max} = 1/t_{path}, \quad t_{path} = t_{clkQ} + \sum_{i \in P} t_i \quad (4)$$

with $t_{clkQ} = t_0 + s F_{max} C_{pin} = 10.06$ μ s. Static power follows the transistor count T :

$$P_{stat} = p_s T \quad (5)$$

with $p_s = 72.67$ nW per transistor (0.431 mW over 5,931 transistors [1]), a uniform weighting discussed in Section VIII.

Ref. [1] publishes one delay number, which fixes only one combination of the two delay coefficients:

$$t_0 + F_{max} C_{pin} s = 10.06 \mu s \quad (6)$$

The split between t_0 and s is unidentified. Under any alternative split (t'_0, s') on the line of (6), a design of n stages (Table 6) with delay t has the delay

$$t' = nt'_0 + \frac{s'}{s} (t - nt_0) \quad (7)$$

Results at the fixed 1 kHz clock are invariant under the split, because a fixed clock removes the delay from throughput and energy. Over the admissible line the maximum-frequency yield ladder moves only from 41/92/198 to 39-43/86-95/185-204 defects per million transistors (ppm/T). Two flagged statements depend on the split: $W = 32$ with ripple carry loses 1 kHz feasibility for s' below 0.409 μ s/pF, and the maximum-frequency energy optimum can exchange between $W = 16$ and 32 with prefix carry (0.5% apart in Table 4).

Every mapping and buffering step is proved formally by structural hashing and SAT over all input and state values [40], and the proof refutes all 20 random single-cell mutants at every design point (Appendix A).

D. VERIFICATION STACK

Verification has four layers (Appendix A). (i) The per-class cycle counts of (1) match at RTL for all thirteen retirement classes at all nine design points (117 of 117 at zero tolerance). (ii) The checksum kernel executes on each mapped and legalized netlist in the RTL cycle count (9,393 cycles at $W = 1$ to 701 at $W = 32$) with the golden result of the instruction-set simulator (ISS) and supplies the measured switching and pad activity. (iii) ABC [39] maps the same RTLIL independently at three points (Table A1). (iv) Against the fabricated $W = 1$ chip this flow produces $1.77\times$ the transistors (10,474 against 5,931) and $2.0\times$ the logic depth (34 against 17 stages), which is why Section V meets the 17-stage limit by construction.

E. WORKLOADS AND THE CLOSED-FORM CPI MODEL

Two workload sets are used: five hand-written kernels (checksum, memcpy, softmul, sort, popcount) of 60 to 1,324 dynamic instructions, simulated at RTL at all nine points, and eighteen of the nineteen Embench-IoT benchmarks [46] compiled for RV32I at -Os (3.1×10^8 dynamic instructions). With the per-class costs exact, cycles per instruction (CPI) is a closed-form function of the class mix $f_c = n_c/N$ [47]:

$$CPI(W) = \sum_c f_c c_c(W) + [W = 1] \frac{1}{N} \sum_r sh_r \quad (8)$$

The headline numbers use the geometric mean of the five kernels, which matches all 45 cycle-accurate measurements within 0.005 CPI (compiled code: Appendix B).

F. WIRE MODEL

The per-pin load is modeled as

$$C_{eff} = C_G + \gamma_w c_w \frac{L(N)}{\sqrt{k}}, \quad L(N) = aN^{p-1/2} \quad (9)$$

where N is the cell count, $L(N)$ is the Donath average wirelength in gate pitches [31], the Rent exponent $p = 0.60$ and prefactor $a = 0.9$ [32,33], device input capacitance $C_G = 0.1$ pF, tier count k , and wire-technology factor γ_w . The wire capacitance per pitch, $c_w = 1.27$ pF, reproduces the measured 2.5 pF per pin for the single-layer $W = 1$ design. The effective wire factor $\lambda_{eff} = C_{eff}/C_{pin}$ rises to 1.106 at $W = 32$ with prefix carry, and no ordering changes for the Rent exponents from 0.5 to 0.75 (Appendix C).

G. EXTERNAL-INTERFACE COST MODEL

Five interface costs are charged (Section IV-D): the external pin count $N_{io}(W)$, pad energy $E_{io} = n_{io} C_{pad} V_{dd}^2$ from measured pad-bit toggles with $C_{pad} = 10$ pF (the maximum legal load), two pad crossings on every critical path, pad yield at ρ transistor-equivalents per pin, and memory bandwidth, measured by repeating all 45 CPI measurements with three wait states per bus transaction.

IV. OPERATING POINT

A. SYNTHESIS

Table 3 shows the nine legalized design points, with zero violations. Serialization saves only 2.45 to 2.60× in transistors (10,474 against 25,679 for ripple and 27,195 for prefix carry at $W = 32$), not 32×, because the control, program counter, instruction register, and sequential state are width-invariant. The carry scheme decides the logic depth: from $W = 1$ to 32 the ripple datapath deepens from 34 to 125 stages and the prefix datapath to 56. Every point meets 1 kHz, and none meets the 17-stage limit (Section V). Depth is given after buffering, with the pre-buffering value in parentheses. Delay is the critical path under Eqs. (3)-(4). Against legalized ABC this flow carries 1.44-1.58× the transistors (Table A1), a width-independent overhead that cancels from every serial-to-parallel ratio up to a ±5% spread. The yield boundaries depend on transistor-count differences and move with the mapper: ABC places the $W = 32$ -to-bit-serial boundary 1.4-1.7× higher and makes the $W = 32$ prefix netlist 3.9% smaller than the ripple one, which would give prefix carry to the $W = 32$ entries of the 1 kHz delay-budget ladders (Tables 7 and 10, Fig. 2-(a)). A mapper q times better at every width multiplies every boundary of (15) by q . The ripple-to-prefix delay ratio at $W = 32$ is 1.48 under ABC and 2.05 here, so the adder conclusion is mapper-independent.

TABLE 3. SYNTHESIS RESULTS FOR THE NINE DESIGN POINTS (POST-BUFFERING, FAN-OUT ≤ 4, ZERO VIOLATIONS)

W	adder	cells	DFF	T (NMOS)	depth (pre-buf.)	delay (μs)
1	ripple	1,696	145	10,474	34 (29)	241.1
2	ripple	2,534	139	15,705	37 (30)	283.3
4	ripple	2,664	138	16,359	44 (35)	325.2
8	ripple	2,937	137	17,748	55 (46)	405.9
16	ripple	3,510	136	20,649	78 (69)	566.7
32	ripple	4,526	135	25,679	125 (114)	874.6
8	prefix	2,978	137	17,979	44 (35)	334.2
16	prefix	3,646	136	21,327	50 (39)	385.6
32	prefix	4,831	135	27,195	56 (41)	426.4

B. WORKLOAD BEHAVIOR

The geometric-mean CPI over the five kernels is 50.00, 26.00, 14.11, 8.16, 5.19, and 3.70 for $W = 1, 2, 4, 8, 16$, and 32, identical for both carry schemes. The serial-to-parallel ratio is 11.8 to 15.1 across the kernels (geometric mean 13.5) and 11.71 to 15.31 across the compiled benchmarks (13.87, Appendix B). The digit count $D = 32/W$ dominates it, and over all instruction mixes it is bounded exactly by 8.75 (stores only) and 24.25 (right shifts by 31 only).

C. THROUGHPUT AND ENERGY UNDER THREE OPERATING ASSUMPTIONS

The delay model is anchored at the 1 kHz clock of the fabricated process, so the primary results stand there, as model evaluations rather than measurements of parallel 2D circuits. Throughput is $\text{thr} = f/\text{CPI}$, and with static power dominant the energy per instruction is

$$E = p_s T \frac{\text{CPI}}{f} \quad (10)$$

Table 4 reports the fixed-clock point and, as a reference, each design at its own maximum frequency, extrapolated up to 5.1× beyond the calibration path of 171 μs. The $f_{\max}$ columns are extrapolated, and the 1 kHz columns are model evaluations at the fabricated clock.

TABLE 4. PERFORMANCE AND ENERGY

W	adder	$f_{\max}$ (Hz)	instr/s at $f_{\max}$	E/instr (μJ) at $f_{\max}$	instr/s at 1 kHz	E/instr (μJ) at 1 kHz
1	ripple	4,147.7	82.9	9.175	20.0	38.05
2	ripple	3,530.1	135.7	8.405	38.5	29.67
4	ripple	3,075.0	217.9	5.457	70.9	16.78
8	ripple	2,463.9	301.8	4.274	122.5	10.53
16	ripple	1,764.5	340.1	4.413	192.8	7.79
32	ripple	1,143.4	309.2	6.035	270.4	6.90
8	prefix	2,992.2	366.5	3.566	122.5	10.67
16	prefix	2,593.6	499.9	3.100	192.8	8.04
32	prefix	2,345.1	634.2	3.116	270.4	7.31

At the fixed 1 kHz clock the parallel core delivers 13.5× the throughput at 5.2-5.5× lower energy per instruction for 2.45-2.60× the transistors: with static power dominant, the CPI

ratio overwhelms the transistor ratio. At maximum frequency (extrapolated), $W = 32$ with prefix carry reaches 634 instructions per second, $7.7\times$ the bit-serial value, while ripple carry forfeits half of the achievable frequency.

The energy ordering depends on an operating assumption made explicit by Eq. (10). Under a fixed clock, or race-to-sleep with power gating, the parallel core wins by the CPI ratio divided by the transistor ratio. Under iso-throughput operation without power gating, the application fixes the instruction rate (here the bit-serial 20 instructions per second, met by the 32-bit ripple core at 74 Hz), $E \propto T$, and the bit-serial core wins by the transistor ratio (38.05 against 93.3 μJ), the reason Ref. [1] gave for serialization. Power gating is not modeled here, so which ordering applies depends on whether the technology can power-gate an idle core. For this reason the width map of Section VI uses throughput, which is assumption-free at a fixed clock.

D. EXTERNAL-INTERFACE COSTS AND MEMORY BANDWIDTH

The external pin count grows from 193 at $W = 1$ to 282 at $W = 32$, yet pad energy favors the parallel design, because the serial core drives few wires many times (318 against 73 pad-bit toggles per instruction), and at 10 pF both pad energies (0.051 and 0.012 μJ per instruction) are small beside the static budget. The pad crossings leave all nine points feasible at 1 kHz. Charging $\rho = 300$ or 1,000 transistor-equivalents per pin keeps the envelope order of Section VI-B, lowers every boundary, and opens a narrow $W = 4$ window between $W = 8$ and bit-serial.

With three wait states per bus transaction (an 8-bit bus serving 32-bit words), the serial-to-parallel CPI ratio falls from 13.5 to 7.3 and the fixed-clock energy optimum moves inward from $W = 32$ to $W = 16$ with ripple carry (13.24 against 13.66 μJ), while the bit-serial core still costs $3.1\times$ the energy (40.85 μJ).

E. THE ADDER QUESTION

At $W = 32$ the standalone ripple adder needs 96 mapped stages and 636.8 μs , and BK needs 20 stages and 149.8 μs for $1.38\times$ the transistors. At core level prefix carry costs 5.9% more transistors and doubles the maximum frequency from 1,143 to 2,345 Hz. With every adder edge annotated by its span in bit pitches, the critical wire paths of both adders traverse the same W pitches, so no linear wire quality reverses the ordering, and long unbuffered distributed-RC wires reverse it only above 2.62 pF per pitch, more than the total measured per-pin load. A minimum-depth network such as Kogge-Stone lengthens the critical wire path and could admit a crossover.

V. THE TWO READINGS OF THE 17-STAGE RULE

The fabricated chip has at most 17 logic stages between flip-flops, and Ref. [1] does not state whether this is a delay budget or a noise-margin and level-restoration constraint of

enhancement/depletion NMOS cascades [48]. This work builds the netlists for both readings.

Under the delay-budget reading, a 17-stage path at 171 μs uses 17% of the 1 kHz period, the deepest point of Table 3 uses 87%, and all nine netlists meet timing as they stand.

Under the cascade reading, no netlist of Table 3 complies, so every point is implemented again to satisfy the three netlist-level rules of Ref. [1] (at most 17 stages between flip-flops, fan-out at most four, no cells below the noise-margin threshold), which establishes rule compliance within the model, not manufacturability. The construction is a sub-cycle pipelining with 2D_NMOS_lib cells alone that keeps the architecture and its cycle counts. A one-hot ring counter divides each architectural cycle into S physical periods, and every architectural flip-flop latches its next state in the last period through a hold multiplexer, omitted where the sequential check proves it unnecessary. $S - 1$ cuts, placed by a dynamic program that minimizes the register count, divide the logic into bands of at most 17 levels, with a flip-flop on every logic-driven net that crosses a cut. Flip-flop outputs and input pads are full-swing sources, and primary outputs hold their settled value, so neither needs a register. All designs run at the compliant period given by Eq. (11)

$$t_c = D_{\max} (t_0 + sF_{\max}C_{\text{pin}}) = 171.02 \mu\text{s},$$

$$f_c = 5,847 \text{ Hz} \quad (11)$$

and each architectural cycle takes $S = \lceil L_{\text{end}}/17 \rceil$ physical periods, so $\text{CPI}_c = S \times \text{CPI}$. The compliant transistor count $T_c = T + \Delta T_{\text{pipe}}$ is counted on the netlist (pipeline registers, hold multiplexers, ring counter, and legalization inverters).

Every compliant netlist passes the analyzer of Section III-C at 17 stages with zero fan-out violations and passes the noise-margin check (critical path 136.1 to 149.6 μs per period). Each is proved equivalent to its mapped netlist over one architectural cycle for all states, pipeline contents, and inputs, with arbitrary values on every unsettled input. Every cycle after reset starts in ring phase 0, so the proof covers every cycle, and a ring of $S - 1$ phases is refuted at every point. All nine netlists also execute the checksum program at gate level from a random initial state in the RTL cycle count with the golden result.

CPI_c is S times the geometric-mean CPI of Section IV-B. Throughput and energy are at the compliant clock of 5,847 Hz with the static power of Eq. (5). Three things change under the cascade reading. First, the throughput optimum stays at $W = 32$ with prefix carry ($6.76\times$ the compliant bit-serial core for $3.61\times$ the transistors), but $W = 16$ with prefix carry comes within 5% of it with 25% fewer transistors and is the energy optimum (8.09 μJ , $2.38\times$ better than bit-serial, against $1.87\times$ for $W = 32$ prefix). Second, ripple carry at $W = 32$ needs 8 sub-cycles against 4 and $1.76\times$ the transistors of prefix carry, and costs $1.88\times$ the energy of the compliant bit-serial core, so prefix carry becomes a prerequisite for a 32-bit datapath that beats bit-serial execution in energy. Third, the transistor gap widens from $2.6\times$ to $3.6\times$, which moves the

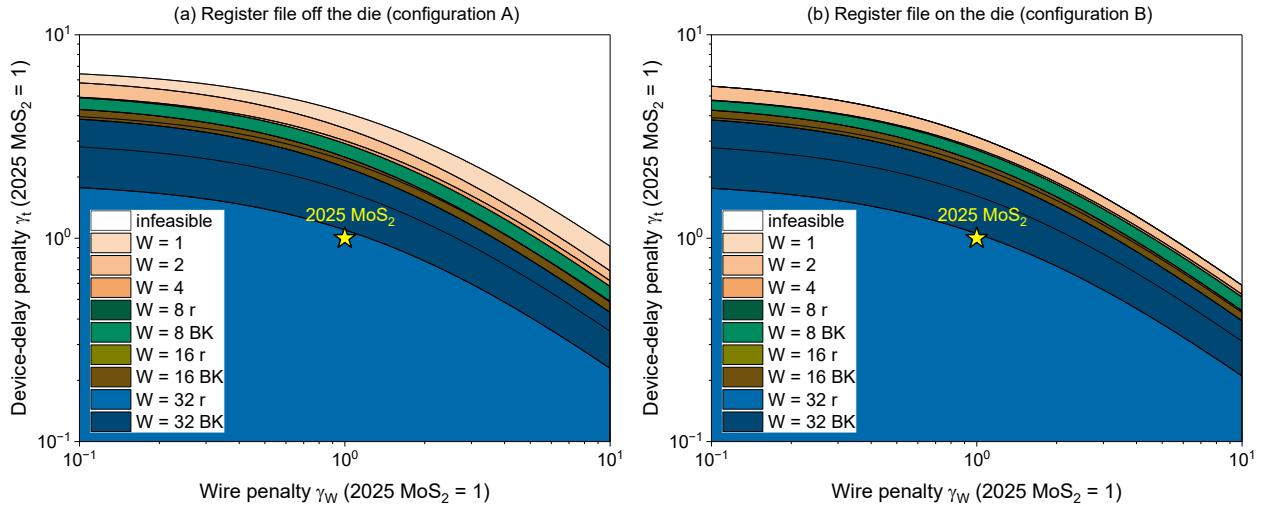


FIGURE 2. Optimal datapath width at 1 kHz over the wire penalty γ_w and the device-delay penalty γ_t for a single tier, (a) with the register file off the die and (b) on the die (Table 6). Colors give the widest feasible design (legend), and black curves the frontiers of (12). The star marks the 2025 MoS₂ process, which in (a) lies near the sub-boundary between ripple and prefix carry at $W = 32$, where the color depends on the split of (6).

yield boundaries. A 1 kHz physical clock would scale every throughput and energy of Table 5 by one factor and leave every ratio unchanged.

TABLE 5. STRICT DEPTH COMPLIANCE AT 17 STAGES

W	adder	S	cut regs	hold MUX	T_c	CPI _c	instr/s	E/instr (μ J)
1	ripple	2	198	134	15,500	100.0	58.5	19.26
2	ripple	3	549	139	27,127	78.0	75.0	26.30
4	ripple	3	767	137	31,703	42.3	138.1	16.68
8	ripple	4	1,068	137	38,565	32.6	179.1	15.64
16	ripple	5	1,953	135	57,447	26.0	225.3	18.53
32	ripple	8	3,953	135	98,448	29.6	197.5	36.22
8	prefix	3	822	137	34,341	24.5	238.9	10.45
16	prefix	3	1,050	136	41,830	15.6	375.5	8.09
32	prefix	4	1,514	135	55,960	14.8	395.1	10.29

VI. THE WIDTH MAP

The map gives every boundary between widths in a form that a reader with another process can evaluate, with the register file off and on the die (Section VI-E).

A. FEASIBILITY BOUNDARIES IN CLOSED FORM

Under scaled parameters the critical path of every design is linear in two scaling factors at zero residual, as shown in Eq. (12).

$$\text{delay} = \gamma_t(D_0 + D_1(k)\gamma_w) = 1/f_{op} \quad (12)$$

where γ_t scales the device delay through both t_0 and s , γ_w scales the wire capacitance, and the two coefficients follow from the measured delay t of a design with n stages on its critical path, as shown in Eq. (13).

$$D_0 = nt_0 + (t - nt_0) \frac{C_G}{C_{pin}},$$

$$D_1(k) = (t - nt_0) \frac{c_w L(N)}{\sqrt{k} C_{pin}} \quad (13)$$

Figs. 2-(a) and (b) show the feasibility boundaries, which are the closed-form curves of Eq. (12) and are reproducible from Table 6 without rerunning the flow. A change of operating frequency rescales only the γ_t axis. At $\gamma_w = 1$, $D_0 + D_1$ equals the measured delay at $W = 1$ and exceeds it elsewhere by 1.8 to 4.9% through the architecture-dependent wire factor.

TABLE 6. FEASIBILITY-FRONTIER COEFFICIENTS OF Eqs. (12)-(13) IN MICROSECONDS, WITH THE REGISTER FILE OFF (A) AND ON (B) THE DIE

W	adder	A: register file off the die			B: register file on the die		
		stages	D_0	D_1	stages	D_0	D_1
1	ripple	35	146.1	95.0	39	163.6	154.2
2	ripple	38	159.4	128.9	39	163.6	154.2
4	ripple	45	188.4	143.1	46	192.6	169.6
8	ripple	56	234.5	181.0	56	234.5	208.7
16	ripple	79	330.6	254.0	79	330.6	288.6
32	ripple	126	526.1	384.4	126	525.9	423.6
8	prefix	45	188.8	153.9	46	192.8	175.5
16	prefix	51	214.2	185.0	51	214.1	206.1
32	prefix	57	239.2	207.9	57	239.2	230.9

The stage count includes the clock-to-output stage of the launching flip-flop. D_1 is given for a single tier and scales as $1/\sqrt{k}$ with the tier count k .

At 1 kHz the widest feasible design is both the throughput and the fixed-clock energy optimum, because CPI and $T \times$ CPI both decrease at every step in W (a property of this family that fails under the bandwidth-limited memory of

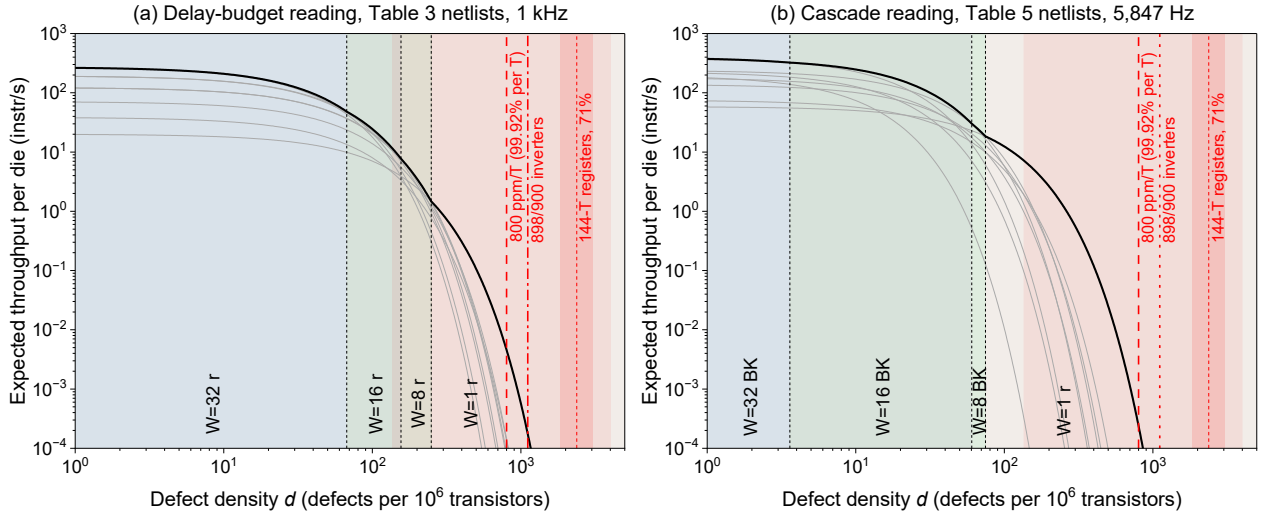


FIGURE 3. The yield ladder. Expected throughput per fabricated die, (14), versus defect density for the nine designs (grey) and their envelope (black), (a) under the delay-budget reading at 1 kHz and (b) under the cascade reading at the compliant clock. Shading marks the optimal width, dotted lines the boundaries of Table 7, and red lines the yields of [1] (dashed: 800 ppm/T, dash-dot: inverters, dotted: 144-transistor registers, not used in the ladders) with 95% bands.

Section IV-D). At unit wire penalty, $W = 32$ survives to $\gamma_t = 2.24$ with prefix carry (1.10 with ripple carry), $W = 16$ to 2.51, $W = 8$ to 2.92, $W = 4$ to 3.02, and $W = 2$ to 3.47, so bit-serial is optimal only for devices 3.47 to $4.15\times$ slower than the 2025 devices against the clock target. Along the wire axis the staircase runs from 1.23 ($W = 32$ ripple) and 3.66 ($W = 32$ prefix) to 8.99 ($W = 1$). Prefix carry extends the parallel region $3.0\times$ along the wire axis and $2.0\times$ along the device axis.

Tiers scale D_1 by $1/\sqrt{k}$, so M3D does not change which width is optimal at a technology point but moves every point $\sqrt{k}$ deeper into the parallel region, within the validity of that assumption (Appendix C).

B. THE YIELD LADDER UNDER BOTH READINGS AND BOTH METRICS

The yield axis uses a Poisson defect model per transistor with no redundancy. The *per-die* metric is the expected throughput per fabricated die, given by Eq. (14).

$$E[thr] = e^{-dT} thr \quad (14)$$

when the pad ring sets the die area, as for the 6 mm $\times$ 6 mm die of [1]. The per-wafer-area metric divides Eq. (14) by T_i , appropriate when the core sets the die area. The boundary between two designs sits where their metrics are equal, as shown in Eq. (15).

$$d_{die}^* = \frac{\ln(thr_i/thr_j)}{T_i - T_j},$$

$$d_{area}^* = \frac{\ln(thr_i/thr_j) - \ln(T_i/T_j)}{T_i - T_j} \quad (15)$$

Figs. 3-(a) and (b) show the per-die ladders under the delay-budget reading at 1 kHz and under the cascade reading at the compliant clock. Under the delay-budget reading, thr and T are those of Tables 3 and 4, and under the cascade reading those of Table 5. Table 7 lists the boundaries for both readings and both metrics, with the following features.

TABLE 7. THE YIELD LADDER WITH THE REGISTER FILE OFF THE DIE (CONFIGURATION A)

reading	clock	metric	envelope	d* (ppm/T)	margin
delay budget	1 kHz	die	32r→16r→8r→1	67 / 156 / 249	3.2×
delay budget	1 kHz	area	32r→16r→8r→1	24 / 104 / 177	4.5×
delay budget	f _{max}	die	32BK→16BK→8BK→1	41 / 92 / 198	4.0×
delay budget	f _{max}	area	16BK→8BK→1	- / 42 / 126	6.3×
cascade	f _c , 1 kHz	die	32BK→16BK→8BK→1	3.6 / 60 / 75	10.7×
cascade	f _c , 1 kHz	area	16BK→1	- / - / 33	24×

- (1) The boundaries are insensitive to the wire penalty.** They move by at most 13% across γ_w from 0.3 to 10, and by at most 3% under the compiled-benchmark mixes.
- (2) The first parallel width beyond the crossover.** $W = 2$ and 4 never reach the envelope at zero pad cost (a pad charge of $\rho \geq 300$ opens a narrow $W = 4$ window, Section IV-D).
- (3) The cascade reading lowers every boundary of the 1 kHz delay-budget ladders by 2.6 to 19×**. The $W = 32$ window shrinks to 3.6 ppm/T per die and vanishes per area, where $W = 16$ with prefix carry is the widest design on the envelope.

C. THE 2025 ANCHOR AND THE DIE-YIELD CONSTRAINT

Fig. 4 shows the yield data of Ref. [1]. Under the Poisson model $Y = \exp(-dT)$, the transistor yield of 99.92% gives $d = 800$ ppm/T, and 898 functional inverters of a 30×30 array (Supplementary Fig. 4 of Ref. [1]) give 1,112 ppm/T (exact 95% interval 135-4,018). One wafer of 28 matrices of eight 8-bit shift registers of 144 transistors (Supplementary Fig. 7 of Ref. [1]) yields 71% of the 224 8-bit registers and 2 of 28 64-bit registers of 1,152 transistors, that is, 2,380 and 2,291 ppm/T, three times the per-transistor value. Spatial clustering raises chip yield at fixed device yield, and more for the larger design under the negative-binomial model [19], [49], as shown in Eq. (16).

$$Y(T) = \left(1 + \frac{dT}{\alpha}\right)^{-\alpha} \quad (16)$$

Every boundary moves to a higher defect density, while the anchor stays within 0.05% of $-\ln Y_T$ for $\alpha \geq 1$, so the Poisson limit ($\alpha \rightarrow \infty$) is conservative in the direction that favors serialization. A binomial maximum-likelihood fit to the two register sizes gives $d = 2,394$ ppm/T (95% interval 1,896-3,260) and $\alpha = 30$ with a 95% lower bound of 1.3, and does not reject the Poisson limit ($p = 0.89$). Each 64-bit register consists of the eight 8-bit registers of one matrix, so the two counts are not independent and these intervals are approximate. The nested data give a direct check: independent 8-bit failures at the observed 71% predict 1.8 fully functional matrices of 28, against 2 observed.

The anchor tests the model only weakly. It lies 3.2 to $24\times$ inside the bit-serial region of every ladder of Table 7, and the register blocks 9.2 to $72\times$, consistent with the fabricated choice, but a model that penalizes transistor count reproduces a choice made on transistor count. The check holds for clustering parameters down to $\alpha = 5.03$ (1 kHz) and 3.68 ($f_{\max}$) per die and 3.14 and 2.09 per area under the delay-budget reading, and 2.07 per die and 0.91 per area under the cascade reading (Table D1), and for uniform mapper factors q from 1 to 1.77. All six break-even values lie within or at the edges of the range of about 1 to 5 that silicon fits of α give [20], and the register data, with a 95% lower bound of 1.3 on α , guarantee only the last check (0.91). The anchor is consistent with the fabricated choice under every reading and metric only for $\alpha \geq 5.03$, that is, for clustering at or beyond the weak end of the silicon range (Appendix D).

The hand-optimized chip is not uniformly better than this flow, because its savings are specific to serial execution (Appendix A, V3). Replacing the $W = 1$ entry alone by the fabricated chip (5,931 transistors at the M2D cycle counts, and $S = 1$ under the cascade reading because it meets the 17-stage rule) bounds this case. The per-die crossover then falls from 249 to 154 ppm/T (delay budget) and from 75 to 33 ppm/T (cascade), with 16 bits as the first parallel width, and the cascade per-area metric favors bit-serial execution at

every density. The anchor stays inside the bit-serial region in every case.

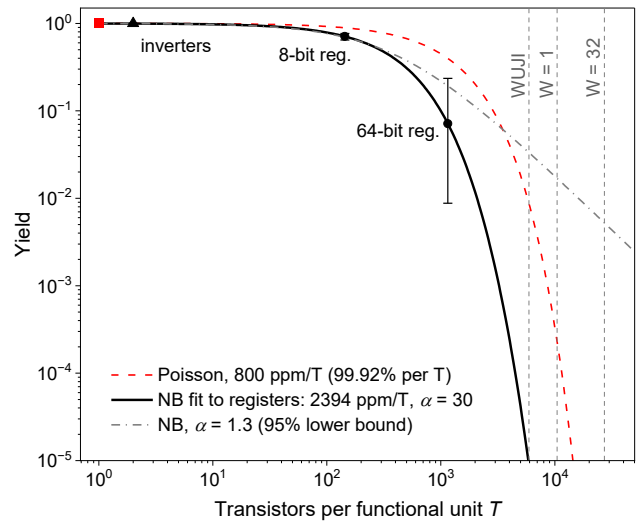


FIGURE 4. Yield data of [1] against the transistor count of the functional unit: per-transistor yield (square), inverters (triangle), and 8-bit and 64-bit registers (circles, exact 95% intervals). Red: Poisson at 800 ppm/T. Black: negative-binomial fit to the registers, $d = 2,394$ ppm/T and $\alpha = 30$. Grey: the fit at the 95% lower bound of α , 1.3. Dotted lines: the fabricated die (5,931 T) and the $W = 1$ and $W = 32$ prefix cores of Table 3.

(1) Die-yield constraint. In the Poisson limit the anchor also lies where every design has a die yield below 1% (0.023% for the bit-serial core of this flow and 0.87% for the fabricated 5,931-transistor chip at 800 ppm/T), so the practical selection rule is a die-yield constraint. With a 10% minimum, the width is selected among designs with $d \leq \ln 10/T$. The bit-serial core meets it below 220 ppm/T ($T = 10,474$) or 149 ppm/T ($T_c = 15,500$), and every wider design only at a lower density. Below these limits the constraint removes the wide designs one by one, and bit-serial execution is selected per die at 141-220 ppm/T (delay budget) and 67-149 ppm/T (cascade). The lower ends are the yield limits of $W = 4$ ripple and $W = 8$ prefix, and under the delay-budget reading $W = 4$ is selected at 130-141 ppm/T. With the fabricated bit-serial chip of the bound above, the windows widen to 130-388 and 33-388 ppm/T.

(2) Process targets. Read in the other direction, these limits are process targets: the defect density must fall $3.6\times$ (delay budget) or $5.4\times$ (cascade) below the 800 ppm/T anchor before any design of this family reaches a 10% die yield. These die yields and targets hold in the Poisson limit ($\alpha \rightarrow \infty$ in Table D1). Clustering raises die yield at a fixed anchor, to 2.1-17% for the fabricated chip over $\alpha = 10$ to 1 (Table D1), and for α from 5 down to the register-data bound of 1.3 the required reduction falls to 2.9-1.3 $\times$ (delay budget) and 4.2-2.0 $\times$ (cascade) (Appendix D).

(3) Choice of anchor. The anchor is a second source of spread. The functional register blocks of Ref. [1] give 2,394 ppm/T, three times the per-transistor value, and raise the

Poisson targets to $11\times$ and $16\times$. The per-transistor value remains the anchor because it is the figure that Ref. [1] reports for the process, and the register value shows how far the targets grow when circuit-level failures are counted.

D. STATIC POWER: WHERE DYNAMIC ENERGY TAKES OVER

Complementary 2D CFETs remove the static current. With textbook CMOS transistor counts and the switching activity α_{sw} measured per design at gate level (fan-out-weighted toggles at cycle boundaries, glitches excluded), the dynamic energy per instruction is shown in Eq. (17).

$$E_{dyn} = \alpha_{sw} N_{pins} C_{eff} V_{dd}^2 CPI \quad (17)$$

The measured activity rises from 0.045 at $W = 1$ to 0.12 at $W = 16$ and eases at $W = 32$, because a serial datapath holds most of the netlist stable while one digit churns. It cuts the parallel advantage in dynamic energy from $4.6\times$ under a uniform activity of 0.10 [42] to $2.0\times$.

Fig. 5 shows the energy-optimal width at maximum frequency as the static power is swept from 0.01 to 100 nW per transistor: $W = 16$ with prefix carry above 14-82 nW, a range that contains the measured 72.67 nW for γ_w below 3.3, $W = 32$ with prefix carry across the broad middle, and $W = 32$ with ripple carry below 0.02-0.28 nW, the CFET regime in which dynamic energy favors the design with fewer gate input pins (N_{pins} in Eq. (17)). Under fixed-clock and race-to-sleep operation no bit-serial region exists across five decades of static power.

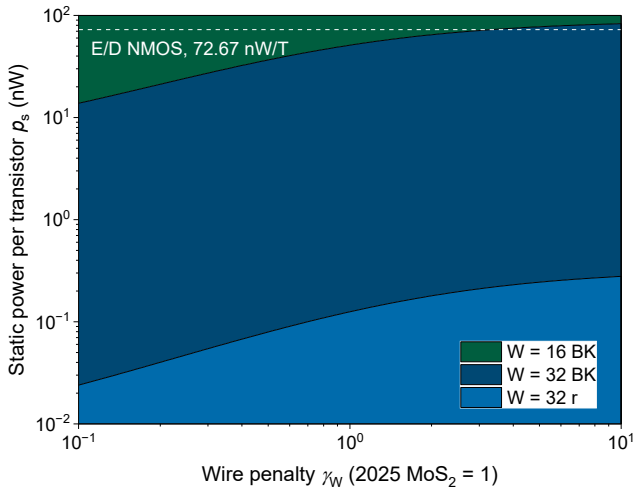


FIGURE 5. Energy-optimal width at maximum frequency over the wire penalty and the static power per transistor, with the measured per-design switching activity, for a single tier. The dashed line marks the measured value of enhancement/depletion NMOS, 72.67 nW per transistor.

E. THE REGISTER FILE ON THE DIE

Configuration A places the register file off the die, as in RV32-WUJI [1], which accesses it through a memory port.

Since the register file is among the largest and slowest blocks of any datapath, configuration B puts it on the die.

The digit-serial register file of Fig. 1 has 31×32 flip-flops ($x0$ is hardwired), two combinational W -bit read ports that select a register and a digit, and one W -bit digit write port with decoded enables. The library flip-flop has no enable, so every bit carries a hold multiplexer. At all nine points the register file is integrated with the core into one netlist (Table 8), pipelined to 17 stages as in Section V, and proved as in Sections III-C and V. The integrated RTL executes all five kernels with ISS results and core-only cycle counts, and the checksum kernel executes at gate level on each integrated netlist. T_{RF} is the register file alone (its depth is 20 to 16 stages from $W = 1$ to 32). T_{sys} , depth, and delay are the integrated netlist after fan-out legalization. E is the fixed-clock energy of Eq. (10) with T_{sys} . S and $T_{sys,c}$ are the depth-compliant integrated netlist.

TABLE 8. THE REGISTER FILE AND THE INTEGRATED CORE-PLUS-REGISTER-FILE NETLISTS (CONFIGURATION B)

W	adder	T_{RF}	T_{sys}	depth	Delay (μs)	E (μJ)	S	$T_{sys,c}$
1	ripple	62,420	72,869	38	290.3	264.8	3	104,581
2	ripple	59,148	74,861	38	290.3	141.4	3	114,645
4	ripple	57,562	73,939	45	332.3	75.8	3	114,079
8	ripple	57,364	75,144	56	405.9	44.6	4	128,085
16	ripple	56,770	77,468	78	566.7	29.2	5	155,112
32	ripple	56,290	82,020	125	870.1	22.1	8	212,955
8	prefix	57,364	75,379	45	336.8	44.7	3	116,402
16	prefix	56,770	78,146	50	382.6	29.5	3	123,853
32	prefix	56,290	83,540	56	426.4	22.5	4	139,817

(1) Transistors. The register file costs 56,290 to 62,420 transistors, 5.4 to $6.0\times$ the bit-serial core, and is largest at $W = 1$ because the digit multiplexer disappears as W grows. With it on the die the serial-to-parallel transistor ratio collapses from 2.45 - 2.60 to 1.13 - 1.15 : serialization saves 11 to 13% of the die, not a factor of 2.5.

(2) Delay. The write port drives the result digit through an inverter fan-out tree into the hold multiplexer of every register bit. This path is critical at $W \leq 4$ and at $W = 8$ with prefix carry (38 stages and 290 μs at $W = 1$ against 34 stages and 241 μs for the core alone), and elsewhere the delay changes by at most 0.8%. The register file slows only the serial end of the family, and all nine integrated netlists meet 1 kHz.

(3) Feasibility. Table 6 (configuration B) shows the coefficients of (12) with c_w kept at the core calibration, and the larger cell count raises the wire factor to 1.21 - 1.22 . In Fig. 2-(b) the serial end moves down (the $W = 1$ limits fall from 4.15 to 3.15 in device delay and from 8.99 to 5.42 in wire load) while the parallel end moves little ($W = 32$ prefix: 2.24 to 2.13 and 3.66 to 3.29). Because the write-port path gives $W = 1$ and 2 the same critical path and $W = 2$ has half the CPI, bit-serial execution is never the widest feasible design.

(4) Energy and the cascade reading. At 1 kHz the 32-bit cores use $12.0\times$ (ripple) and $11.8\times$ (prefix) less energy per instruction than the bit-serial core ($5.2\text{--}5.5\times$ in configuration A), and at iso-throughput without power gating the bit-serial advantage shrinks to $1.13\text{--}1.15\times$. Under the cascade reading the bit-serial system needs $S = 3$, because its 38-stage path into the write port exceeds two bands of 17, so the compliant 32-bit prefix system delivers $10.1\times$ its throughput at $7.6\times$ less energy ($6.76\times$ and $1.87\times$ in configuration A).

(5) Yield. Table 9 gives the ladders of (15) with T_{sys} and $T_{\text{sys,c}}$. Every boundary moves to a higher defect density, because the transistor differences shrink while the throughput ratios do not, and under the delay-budget reading a $W = 4$ window opens and bit-serial wins only above 1,182 ppm/T. The die yield of an integrated die is decisive: 5×10^{-26} for the 72,869-transistor bit-serial system at 800 ppm/T, and 10% only below 32 ppm/T (22 ppm/T under the cascade reading), a $25\times$ or $36\times$ reduction from the anchor. Under a 10% die-yield constraint the selection steps from 32 through 16, 8, and 4 bits to bit-serial, which is selected only at 31.1–31.6 ppm/T (delay budget) or 20.2–22.0 ppm/T (cascade), just below the density above which no integrated design qualifies. With clustering ($\alpha = 1$ to 10) the width selected at the anchor is 8 bits or wider, so the consistency check of Section VI-C holds only for configuration A, the one [1] built. The last column of Table 9 is the width selected at the anchor, or the anchor's margin inside the bit-serial region.

TABLE 9. THE YIELD LADDER WITH THE REGISTER FILE ON THE DIE (CONFIGURATION B)

reading	clock	metric	envelope	d^* (ppm/T)	anchor
delay budget	1 kHz	die	$32r \rightarrow 16r \rightarrow 8r \rightarrow 4r \rightarrow 1$	74 / 195 / 454 / 1,182	$W = 4$
delay budget	1 kHz	area	$32r \rightarrow 16r \rightarrow 8r \rightarrow 4r \rightarrow 1$	62 / 182 / 441 / 1,169	$W = 4$
delay budget	f_{max}	die	$32BK \rightarrow 16BK \rightarrow 8BK \rightarrow 4r \rightarrow 1$	43 / 117 / 371 / 1,056	$W = 4$
delay budget	f_{max}	area	$32BK \rightarrow 16BK \rightarrow 8BK \rightarrow 4r \rightarrow 1$	30 / 104 / 358 / 1,042	$W = 4$
cascade	f_c , 1 kHz	die	$32BK \rightarrow 16BK \rightarrow 8BK \rightarrow 1$	3 / 61 / 153	$5.2\times$ inside
cascade	f_c , 1 kHz	area	$16BK \rightarrow 8BK \rightarrow 1$	- / 52 / 144	$5.5\times$ inside

VII. WHERE THE CROSSOVER IS, AND HOW TO USE THE MAP

A. THE CROSSOVER

Table 10 shows the width selected at the 2025 process point under fixed-clock or race-to-sleep operation. With the register file off the die, $W = 32$ is the throughput and fixed-clock energy optimum, and under the cascade reading $W = 16$ with prefix carry becomes the energy optimum. At the device and wire parameters of this point, yield is the only axis that selects bit-serial execution: above 249 ppm/T under the delay-budget reading and above 75 under the cascade

reading, with 8 bits as the first parallel width. These values carry the mapper uncertainty of Section IV-A. A uniformly better mapper raises them by its factor q , up to 1.77 for the gap to the fabricated chip, whereas an improvement confined to the serial design lowers them, to 154 and 33 ppm/T for the fabricated chip (Section VI-C). Mapper quality moves the crossover by up to $2.3\times$ from its nominal value. In the Poisson limit the 2025 process lies inside the bit-serial region, consistent with the fabricated design, but where every design yields below 1%, and a 10% die-yield constraint confines bit-serial execution to 141–220 or 67–149 ppm/T. Read as a process target, a 10% die yield requires a $3.6\times$ (delay budget) or $5.4\times$ (cascade) lower defect density than the 2025 anchor, or $1.3\text{--}2.9\times$ and $2.0\text{--}4.2\times$ for clustering parameters from 1.3 to 5 (Section VI-C).

With the register file on the die, the transistor ratio falls to 1.13–1.15, bit-serial execution is never the widest feasible design, and under a 10% die-yield constraint with the Poisson statistics it is selected only in windows of 0.5 and 1.8 ppm/T at the edge of feasibility. That edge lies $25\times$ (delay budget) or $36\times$ (cascade) below the 2025 anchor.

TABLE 10. WIDTH SELECTED AT THE 2025 PROCESS POINT AND EXTENT OF THE BIT-SERIAL REGION, BY REGISTER-FILE PLACEMENT AND READING OF THE 17-STAGE RULE

RF	17-stage reading	optimum at the 2025 point	bit-serial above (ppm/T)	at 800 ppm/T	bit-serial under $Y \geq 10\%$ (ppm/T)
off (A)	delay budget	$W = 32$ (r or BK)	249	$W = 1$	141–220
off (A)	cascade	32BK (thr), 16BK (E)	75	$W = 1$	67–149
on (B)	delay budget	$W = 32$ (r or BK)	1,182	$W = 4$	31.1–31.6
on (B)	cascade	32BK (thr), 16BK (E)	153	$W = 1$	20.2–22.0

Fixed-clock or race-to-sleep operation, zero-wait memory, and Poisson per-die metric at 1 kHz (delay budget) or at the compliant clock (cascade). The last column is the range in which $W = 1$ is optimal among designs with a die yield of at least 10%, with $\ln 10$ over the bit-serial transistor count as upper end. At iso-throughput without power gating, bit-serial execution is energy-optimal in every row. thr: throughput, E: energy per instruction, r: ripple carry.

B. PROCEDURE FOR LOCATING ANOTHER TECHNOLOGY

A technology is located by four measurements: (i) the per-transistor yield, which gives d , (ii) the stage-delay and wire parameters relative to the 2025 MoS₂ values (γ_t , γ_w) and the tier count k , (iii) the static power per transistor p_s , and (iv) the cascade limit, read as a delay budget or a hard constraint, together with a target clock and an operating assumption.

(1) Feasibility. Evaluate Eq. (12) with Table 6. The widest feasible design is the throughput optimum and, under fixed-clock or race-to-sleep operation with zero-wait memory, the

energy optimum. Under the cascade reading Table 5 gives the optimum ($W = 16$ with prefix carry in energy).

(2) Yield. Read Table 7 or 9 for the applicable reading and metric, multiply every boundary by q for a mapper q times better (Section VI-C bounds an improvement confined to the serial design), and budget the pad exposure p (Section IV-D). Under a minimum die yield, drop the designs that miss it before selecting (Section VI-C).

(3) Adder. Use BK prefix carry whenever speed, the extent of the parallel region, or a cascade rule matters. It is mandatory at $W = 32$ under a hard 17-stage rule. Ripple carry fits only fixed low-frequency operation with ample margin and the deep-CFET regime of Fig. 5.

(4) Operating assumption. A bandwidth-limited memory moves the interior fixed-clock energy optimum to $W = 16$ but never favors bit-serial. If the application fixes the throughput and the technology cannot power-gate, invert the energy conclusion: the narrowest feasible datapath is energy-optimal.

(5) Reuse conditions. Tables 6, 7, and 9 belong to netlists in 2D_NMOS_lib (25 cells, fan-out four, an 18-transistor flip-flop, linear stage delay). They transfer without rerunning the flow to a process that keeps these cells and rules and changes only t_0 and s (through γ_t and Eq. (7)), C_{pin} and c_w (through γ_w), k , p_s , and d , and otherwise require remapping with the released flow. A per-stage delay should be measured in either case, because the split of (6) is otherwise unidentified.

(6) Worked example (the 2025 process). $d = 800$ ppm/T, $\gamma_t = \gamma_w = 1$, $k = 1$, $p_s = 72.67$ nW/T, 1 kHz. Step 1 selects $W = 32$, and step 2 overrides it to $W = 1$ under both readings, at a density where every design yields below 1% in the Poisson limit. The map returns the fabricated design.

VIII. LIMITATIONS

(1) Relative quality of results. The flow carries about $1.5\times$ the transistors of legalized ABC and $1.77\times$ those of the hand-optimized chip. It predicts relative trends across widths within one flow, not the quality of hand design at any single width (bounded for the serial design in Section VI-C), and the gap to a commercial mapper with a real process design kit is unmeasured.

(2) Register file. Configuration B synthesizes one register file, a flip-flop array with digit-addressed ports (not shift-register, latch-based, or macro storage), with the wire model calibrated at the core alone.

(3) Uniform microarchitecture. The fixed finite-state machine forgoes width-specific optimizations on both sides. The four unshared adders inflate the parallel transistor counts, which is conservative against the parallel conclusion.

(4) Stated assumptions in the projections. The $1/\sqrt{k}$ tier rule (inconsistent with the Donath scaling of the architecture axis, Appendix C), the CFET cell counts, the device input capacitance, and the Rent exponent are assumptions, and no model covers inter-tier vias or the thermal limits of M3D.

(5) Yield model. Defects are counted per transistor with no redundancy, because the dominant failure mechanisms of

current 2D processes (channel-film defects, contact and threshold non-uniformity) attach to devices, and device-level yield is what Ref. [1] and its precursor [4] optimize and report. One reported die yield would decide the clustering test of Section VI-C. The die-yield statements of Sections VI-C and VI-E are Poisson values, and clustering lowers the density reductions they require (Appendix D).

(6) Frequency extrapolation. The maximum-frequency columns extrapolate the delay model up to $5.1\times$ beyond the calibration path, but the primary conclusions stand at 1 kHz, and the split of (6) is swept.

(7) Activity and clock energy. The measured activity excludes glitches, and the clock network is not modeled. Both omissions are conservative for Section VI-D, and only the low-static-power corner of Fig. 5 could move under a glitch-aware simulator.

(8) Compiled workloads. The compiled CPIs come from the validated closed-form model (Appendix B), no compiled benchmark has run to completion at RTL or gate level, and gate-level execution covers one program per netlist.

(9) Scope. Depth compliance is rule compliance within the model, without placement, routing, or electrical cascade simulation. The map decides how wide the core should be, not how a complete system should be partitioned.

(10) Static power per transistor. Static power is charged per transistor at the chip average of Ref. [1], whereas in ratioed enhancement/depletion NMOS it depends on the logic state of each gate. Every energy inherits this weighting, while the throughput, feasibility, and yield results do not.

IX. CONCLUSION

On a model calibrated to the 2025 MoS₂ process, bit-serial RISC-V execution on 2D semiconductors is a strategy for yield, for feasibility with very slow devices, and, under an iso-throughput requirement without power gating, for static energy. It is not a strategy for throughput or fixed-clock energy. Serialization saves only $2.45\text{--}2.60\times$ in transistors while it raises the cycles per instruction $13.5\times$, so at the fabricated 1 kHz clock a 32-bit core delivers $13.5\times$ the throughput at $5.2\text{--}5.5\times$ lower energy per instruction.

Sub-cycle pipelining with the process cell library meets the 17-stage limit at every design point, with each netlist proved equivalent to the original and executed at gate level. Under this hard constraint the 32-bit prefix core keeps $6.76\times$ the throughput of the bit-serial core at $1.87\times$ lower energy, 16 bits with prefix carry becomes the energy optimum, and prefix carry becomes a prerequisite for a 32-bit datapath that beats bit-serial execution in energy.

The crossover from bit-serial to parallel is a defect density in closed form, about 250 defects per million transistors when the 17-stage rule is a delay budget and 75 when it is a hard constraint, with 8 bits as the first parallel width. Mapper quality moves these values by a factor of up to 2.3. Under the Poisson statistics the yield data of Ref. [1] lie inside the bit-serial region, consistent with the fabricated choice, and a

10% die yield for any design requires the defect density to fall $3.6\times$ or $5.4\times$ below the published value. Defect clustering lowers this target and, if it is as strong as in silicon, can move the published point out of the bit-serial region.

With the register file on the die, the serial-to-parallel transistor ratio collapses to 1.13-1.15, bit-serial execution is never the widest feasible design, and under the Poisson statistics a 10% die yield requires a $25\times$ or $36\times$ lower density than the published value. Tables 6, 7, and 9 and the procedure of Section VII-B locate other device-count-limited technologies on this map.

The data that support the findings of this study are available from the corresponding author on request and <https://github.com/nhayase/M2D>.

APPENDIX A. VALIDATION DETAILS

(1) Calibration. The delay pair (t_0, s) is calibrated jointly so that a 17-stage path at full fan-out reproduces the published $171\ \mu\text{s}$: $17 \times (4.06 + 0.6 \times 10) = 171.0\ \mu\text{s}$. Ref. [1] publishes the total and no per-stage data, so the publication alone supports no independent fit of the two coefficients. The same holds for the static power per transistor and the 18-transistor flip-flop.

(2a) Cycle-exact class validation. Two programs with n_1 and n_2 copies of one instruction give its class cost $c = [N_{\text{cyc}}(n_2) - N_{\text{cyc}}(n_1)]/(n_2 - n_1)$, free of the setup and termination costs. All thirteen retirement classes, including LUI, AUIPC, JAL, and JALR, match at all nine design points (117 of 117 at zero tolerance), and JALR costs $2D + 2$ rather than $D + 2$ cycles (Appendix B).

(2b) Gate-level execution. The checksum kernel runs on each of the nine mapped and legalized netlists against the same external memory and register-file environment as the RTL benches, finishing in a cycle count asserted equal to the RTL and producing the golden checksum 0x353761ae.

(2c) Independent mapper. The ABC cross-check uses the built-in gate classes with a one-to-one counterpart in 2D_NMOS_lib. Each netlist is translated cell for cell, analyzed and legalized like this flow, and re-verified against the emitted netlist by randomized equivalence over 8,192 patterns. Legalization costs ABC 8.7% to 13.7% of its transistors against 9.2% to 10.1% for this flow. Between legal netlists this flow is $1.48\times$ ($W = 1$) and $1.69\times$ ($W = 32$) deeper for ripple carry and $1.12\times$ for prefix carry. The ABC gate set lacks the three-input NAND and NOR and the larger AOI/OAI cells, so $q = 1.48$ is a lower bound for a production mapper.

(2d) Coverage. Every validation stage ran at all nine design points except the ABC cross-check. The co-simulation suite of 122 tests passes, configuration B adds five kernels at RTL and the checksum kernel at gate level per point, and the formal proofs of Sections III-C, V, and VI-E hold at all nine points of both configurations (11 to 108 outputs per core netlist and 1,000 to 1,067 per integrated netlist reach SAT).

(3) Chip-level gap. The 10,474 transistors of this flow at $W = 1$ comprise 2,610 T of flip-flops, 884 T of legalization inverters, 3,249 T of multiplexer trees, and 3,731 T of other combinational logic. Serial immediate decoding, register-port sharing, and global restructuring in Ref. [1] account for the $1.77\times$ gap. ABC emits a 15-stage netlist for the same RTL with fan-out violations on 61 nets, and legalizing it costs 283 inverters and raises the depth to 23 stages, so the depth gap lies in hand optimization and not in the mapper. ABC enforces no fan-out rule, so Table A1 lists its netlists as emitted and after the same legalization, and the legalized rows are the like-for-like comparison. The as-emitted delays are optimistic bounds, because the model saturates the load at four pins while those netlists reach fan-outs of 40 to 84.

TABLE A1. INDEPENDENT-MAPPER CROSS-CHECK WITH BOTH MAPPERS FAN-OUT-LEGALIZED

design	mapper	cells	T	depth	delay (μs)	max FO	FO > 4
W = 1 ripple	this flow (legalized)	1,696	10,474	34	241.1	4	0
W = 1 ripple	ABC (as emitted)	851	6,490	15	120.5	40	61
W = 1 ripple	ABC (legalized)	1,134	7,056	23	175.4	4	0
W = 32 ripple	this flow (legalized)	4,526	25,679	125	874.6	4	0
W = 32 ripple	ABC (as emitted)	2,821	15,718	60	474.2	84	214
W = 32 ripple	ABC (legalized)	3,894	17,864	74	570.0	4	0
W = 32 prefix	this flow (legalized)	4,831	27,195	56	426.4	4	0
W = 32 prefix	ABC (as emitted)	2,616	15,326	40	309.0	71	170
W = 32 prefix	ABC (legalized)	3,536	17,166	50	384.5	4	0

APPENDIX B. COMPILED WORKLOADS AND THE RANGE OF INSTRUCTION MIXES

Eighteen of the nineteen Embench-IoT benchmarks [46] were built with riscv-none-elf-gcc 15.2.0 at $-\text{Os}$ ($-\text{march}=rv32i$ $-\text{mabi}=ilp32$ $-\text{mstrict-align}$, without wikisort, which needs facilities outside the implemented subset) and run to completion on the ISS. Under Eq. (8) their serial-to-parallel CPI ratio ranges from 11.71 (statemate) to 15.31 (aha-mont64), with a geometric mean of 13.87, and absolute CPI exceeds that of the kernels by 1.0% at $W = 32$ and 3.6% at $W = 1$ through procedure calls and returns. At 1 kHz the compiled geometric mean gives $39.45\ \mu\text{J}$ per instruction at $W = 1$ against 6.97 (ripple) and 7.38 (prefix) at $W = 32$.

The first 6,000 retired instructions of four binaries (slre, ud, sglib-combined, qrduino) ran on the RTL core at all nine design points, and in all 36 runs the cycle count equals the closed-form prediction and the data memory matches the ISS. Two prefixes retire JALR (137 and 111 times), on which a

model treating JALR as a one-stage jump would be short by 0.73 and 0.59 CPI at $W = 1$, so the model is exact on compiled code, not fitted to the kernels.

Since the per-class costs of (1) are fixed integers and CPI is linear in the class-frequency vector, $CPI(1)/CPI(32)$ is a ratio of linear forms on the probability simplex and attains its extrema at vertices (single-class programs), as shown in Eq. (18).

$$\min_c \frac{c_c(1)}{c_c(32)} \leq \frac{CPI(1)}{CPI(32)} \leq \max_c \frac{c_c(1)}{c_c(32)} \quad (18)$$

The envelope runs from 8.75 (stores alone, 35 against 4 cycles) to 24.25 (right shifts by 31 alone, 97 against 4). With the transistor ratio of 2.45 to 2.60, the parallel design wins fixed-clock energy per instruction by 3.4× (prefix carry, store vertex) to 9.9× (ripple carry, right-shift vertex) and throughput by 8.8 to 24.3× for every conceivable workload.

APPENDIX C. WIRE MODEL CAVEATS AND RENT SENSITIVITY

Three caveats attach to (9). First, the cell count spans 2.85× across the family, so the wire factor carries a small architecture-dependent penalty (+10.6% on the widest design) on which no conclusion rests. Second, any constant tail factor between the average and the critical-path wirelength multiplies c_w and is absorbed by the calibration and the γ_w axis. Third, the tier axis uses the geometric $1/\sqrt{k}$ of the 3D-integration convention [35], whereas Donath scaling would give $k^{p-1/2} = 0.87$ at $k = 4$ instead of 0.50, and Section VI-A gives the effect as a factor $\sqrt{k}$ that a reader may replace. Rederiving the wire model for Rent exponents 0.5, 0.6, and 0.75 and prefactors 0.5, 0.9, and 2.0, recalibrated to 2.5 pF per pin at each pair, removes the prefactor from every result. The exponent moves the wire factor of the widest design from 1.000 to 1.287, the per-die boundary at which $W = 32$ leaves the maximum-frequency envelope, evaluated with Eq. (12) at $\gamma_w = 1$, from 40.5 ppm/T ($p = 0.5$, where every load equals the uniform 2.5 pF of Table 7) to 34.7 ppm/T ($p = 0.75$), +5% to -10% about 38.4 at $p = 0.6$, and the $W = 32$ ripple wire limit from 1.36 to 1.06, +11% to -14% about 1.23. No ordering changes anywhere in the swept box.

APPENDIX D. CLUSTERING: WIDTH SELECTED AT THE ANCHOR

Table D1 shows the width selected by the per-die envelope at the 800 ppm/T anchor (configuration A) under the negative-binomial model Eq. (16) as the clustering parameter is swept, with the die yield that the anchor implies for the fabricated 5,931-transistor chip. The last row gives the break-even value of each column, the value of α above which the anchor selects $W = 1$, and the table note gives the per-area values. The anchor itself moves by less than 0.05%, and a reported die yield would select a row. The die-yield column shows that the die-yield statements of Section VI-C hold only in the Poisson limit. At $\alpha = 1$ the fabricated chip already exceeds a

10% die yield at the anchor. With clustering the bit-serial core of this flow reaches a 10% die yield below $d = \alpha(10^{1/\alpha} - 1)/T$, which for α from 5 to 1.3 lowers the required reduction from 3.6× to 2.9-1.3× (delay budget, $T = 10,474$) and from 5.4× to 4.2-2.0× (cascade, $T = 15,500$).

TABLE D1. WIDTH SELECTED AT THE ANCHOR VERSUS CLUSTERING (PER-DIE METRIC)

α	regime	delay budget (Tables 3, 4), $f_{\max}$	delay budget (Tables 3, 4), 1 kHz	cascade (Table 5)	Y of a 5,931- T die
∞	Poisson	$W = 1$	$W = 1$	$W = 1$	0.87 %
10	weak	$W = 1$	$W = 1$	$W = 1$	2.06 %
5	weak	$W = 1$	$W = 8$ ripple	$W = 1$	3.55 %
3	moderate	$W = 8$ prefix	$W = 16$ ripple	$W = 1$	5.81 %
2	moderate	$W = 16$ prefix	$W = 16$ ripple	$W = 16$ prefix	8.79 %
1	strong	$W = 32$ prefix	$W = 32$ ripple	$W = 16$ prefix	17.40 %
	break-even	3.68	5.03	2.07	—

The break-even row gives the value of α above which $W = 1$ is selected at the anchor. Under the per-wafer-area metric the break-even values are 2.09 ($f_{\max}$), 3.14 (1 kHz), and 0.91 (cascade). The bit-serial core of this flow (10,474 transistors) has a die yield of 0.023%, 0.23%, 0.73%, 1.8%, 3.7%, and 10.7% at the anchor for the six rows.

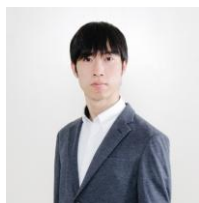
ACKNOWLEDGMENT

Claude.ai was used to draw Fig. 1 and correct the M2D code errors.

REFERENCES

- [1] M. Ao, X. Zhou, X. Kong, S. Gou, S. Chen, X. Dong, Y. Zhu, Q. Sun, Z. Zhang, J. Zhang, Q. Zhang, Y. Hu, C. Sheng, K. Wang, S. Wang, J. Wan, J. Han, W. Bao, and P. Zhou, "A RISC-V 32-bit microprocessor based on two-dimensional semiconductors," *Nature*, vol. 640, pp. 654–661, 2025.
- [2] S. Wachter, D. K. Polyushkin, O. Bethge, and T. Mueller, "A microprocessor based on a two-dimensional semiconductor," *Nat. Commun.*, vol. 8, art. no. 14948, 2017.
- [3] S. Ghosh, Y. Zheng, M. Rafiq, H. Ravichandran, Y. Sun, C. Chen, M. Goswami, N. U. Sakib, M. U. K. Sadaf, A. Pannone, S. Ray, J. M. Redwing, Y. Yang, S. Sahay, and S. Das, "A complementary two-dimensional material-based one instruction set computer," *Nature*, vol. 642, pp. 327–335, 2025.
- [4] X. Chen et al., "Wafer-scale functional circuits based on two dimensional semiconductors with fabrication optimized by machine learning," *Nat. Commun.*, vol. 12, art. no. 5953, 2021.
- [5] D. Fan et al., "Two-dimensional semiconductor integrated circuits operating at gigahertz frequencies," *Nat. Electron.*, vol. 6, pp. 879–887, 2023.
- [6] S. Das et al., "Transistors based on two-dimensional materials for future integrated circuits," *Nat. Electron.*, vol. 4, pp. 786–799, 2021.
- [7] Y. Liu et al., "Promises and prospects of two-dimensional transistors," *Nature*, vol. 591, pp. 43–53, 2021.
- [8] H. Wang et al., "Integrated circuits based on bilayer MoS₂ transistors," *Nano Lett.*, vol. 12, pp. 4674–4680, 2012.
- [9] L. Yu et al., "Design, modeling, and fabrication of chemical vapor deposition grown MoS₂ circuits with E-mode FETs for large-area electronics," *Nano Lett.*, vol. 16, pp. 6349–6356, 2016.
- [10] W. Li et al., "Approaching the quantum limit in two-dimensional semiconductor contacts," *Nature*, vol. 613, no. 7943, pp. 274–279, 2023.

- [11] G. Migliato Marega et al., "Logic-in-memory based on an atomically thin semiconductor," *Nature*, vol. 587, pp. 72–77, 2020.
- [12] P.-C. Shen et al., "Ultralow contact resistance between semimetal and monolayer semiconductors," *Nature*, vol. 593, no. 7858, pp. 211–217, 2021.
- [13] M. M. Shulaker et al., "Three-dimensional integration of nanotechnologies for computing and data storage on a single chip," *Nature*, vol. 547, pp. 74–78, 2017.
- [14] J.-H. Kang et al., "Monolithic 3D integration of 2D materials-based electronics towards ultimate edge computing solutions," *Nat. Mater.*, vol. 22, pp. 1470–1477, 2023.
- [15] S. Ghosh et al., "Monolithic and heterogeneous three-dimensional integration of two-dimensional materials with high-density vias," *Nat. Electron.*, vol. 7, no. 10, pp. 892–903, 2024.
- [16] A. Pal et al., "Three-dimensional transistors with two-dimensional semiconductors for future CMOS scaling," *Nat. Electron.*, vol. 7, pp. 1147–1157, 2024.
- [17] D. E. Nikonov and I. A. Young, "Overview of beyond-CMOS devices and a uniform methodology for their benchmarking," *Proc. IEEE*, vol. 101, no. 12, pp. 2498–2533, Dec. 2013.
- [18] M. Nguyen et al., *Minerva: A 32-bit RISC-V soft processor*. <https://github.com/minerva-cpu/minerva> Accessed: Oct. 8, 2026.
- [19] C. H. Stapper, F. M. Armstrong, and K. Saji, "Integrated circuit yield statistics," *Proc. IEEE*, vol. 71, no. 4, pp. 453–470, 1983.
- [20] J. A. Cunningham, "The use and evaluation of yield models in integrated circuit manufacturing," *IEEE Trans. Semicond. Manuf.*, vol. 3, no. 2, pp. 60–71, May 1990.
- [21] O. Kindgren et al., *SERV: The SERIAL RISC-V CPU*. <https://github.com/olofk/serv> Accessed: Oct. 8, 2026.
- [22] O. Kindgren, *QERV*. <https://github.com/olofk/qerv> Accessed: Oct. 8, 2026.
- [23] M. Kissich and M. Baunach, "FazyRV: Closing the gap between 32-bit and bit-serial RISC-V cores with a scalable implementation," *Proc. 21st ACM Int. Conf. Computing Frontiers (CF)*, 2024, pp. 240–248.
- [24] S. Prakash et al., "Lifetime-aware design for item-level intelligence at the extreme edge," *Proc. 31st ACM Int. Conf. Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, vol. 2, 2026, pp. 1089–1112.
- [25] C. Ewert et al., "SAILOR: A scalable and energy-efficient ultra-lightweight RISC-V for IoT security," arXiv:2602.24166, 2026.
- [26] A. Raisiardi, K. Iordanou, J. Kufel, K. Gudimetla, K. Myny, and E. Ozer, "Flexing RISC-V instruction subset processors to extreme edge," in *Proc. 58th IEEE/ACM Int. Symp. Microarchitecture (MICRO)*, 2025, pp. 1147–1159.
- [27] R. P. Brent and H. T. Kung, "A regular layout for parallel adders," *IEEE Trans. Comput.*, vol. C-31, no. 3, pp. 260–264, 1982.
- [28] P. M. Kogge and H. S. Stone, "A parallel algorithm for the efficient solution of a general class of recurrence equations," *IEEE Trans. Comput.*, vol. C-22, no. 8, pp. 786–793, 1973.
- [29] J. Sklansky, "Conditional-sum addition logic," *IRE Trans. Electron. Comput.*, vol. EC-9, pp. 226–231, 1960.
- [30] R. E. Ladner and M. J. Fischer, "Parallel prefix computation," *J. ACM*, vol. 27, no. 4, pp. 831–838, 1980.
- [31] W. E. Donath, "Placement and average interconnection lengths of computer logic," *IEEE Trans. Circuits Syst.*, vol. 26, no. 4, pp. 272–277, 1979.
- [32] B. S. Landman and R. L. Russo, "On a pin versus block relationship for partitions of logic graphs," *IEEE Trans. Comput.*, vol. C-20, no. 12, pp. 1469–1479, 1971.
- [33] P. Christie and D. Stroobandt, "The interpretation and application of Rent's rule," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 8, no. 6, pp. 639–648, 2000.
- [34] M. M. Sabry Aly et al., "Energy-efficient abundant-data computing: The N3XT 1,000x," *Computer*, vol. 48, no. 12, pp. 24–33, 2015.
- [35] A. Rahman and R. Reif, "System-level performance evaluation of three-dimensional integrated circuits," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 8, no. 6, pp. 671–678, Dec. 2000.
- [36] N. Bleier, C. Lee, F. Rodriguez, A. Sou, S. White, and R. Kumar, "FlexiCores: Low footprint, high yield, field reprogrammable flexible microprocessors," *Proc. 49th Annu. Int. Symp. Computer Architecture (ISCA)*, 2022, pp. 831–846.
- [37] C. Wolf, *Yosys Open Synthesis Suite*. <https://yosyshq.net/yosys/> Accessed: Oct. 8, 2026.
- [38] K. Keutzer, "DAGON: Technology binding and local optimization by DAG matching," in *Proc. 24th ACM/IEEE Design Automation Conf. (DAC)*, 1987, pp. 341–347.
- [39] A. Mishchenko, S. Chatterjee, and R. K. Brayton, "DAG-aware AIG rewriting: A fresh look at combinational logic synthesis," *Proc. 43rd Design Automation Conf. (DAC)*, San Francisco, CA, USA, 2006, pp. 532–535.
- [40] A. Mishchenko, S. Chatterjee, R. Brayton, and N. Eén, "Improvements to combinational equivalence checking," *Proc. IEEE/ACM Int. Conf. Computer-Aided Design (ICCAD)*, 2006, pp. 836–843.
- [41] C. E. Leiserson and J. B. Saxe, "Retiming synchronous circuitry," *Algorithmica*, vol. 6, no. 1, pp. 5–35, 1991.
- [42] F. N. Najm, "A survey of power estimation techniques in VLSI circuits," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 2, no. 4, pp. 446–455, Dec. 1994.
- [43] R. I. Hartley and K. K. Parhi, *Digit-Serial Computation*. Norwell, MA, USA: Kluwer Academic, 1995.
- [44] A. Waterman and K. Asanović, Eds., *The RISC-V Instruction Set Manual, Volume I: Unprivileged ISA*, RISC-V Foundation, 2019.
- [45] *Amaranth HDL: A modern hardware definition language and toolchain based on Python*. <https://github.com/amaranth-lang/amaranth> Accessed: Oct. 8, 2026.
- [46] Embench, *Embench IoT Benchmark Suite*. <https://embench.org> Accessed: Oct. 8, 2026.
- [47] J. L. Hennessy and D. A. Patterson, *Computer Architecture: A Quantitative Approach*, 6th ed. Cambridge, MA, USA: Morgan Kaufmann, 2019.
- [48] C. Mead and L. Conway, *Introduction to VLSI Systems*. Reading, MA, USA: Addison-Wesley, 1980.
- [49] I. Koren and Z. Koren, "Defect tolerance in VLSI circuits: Techniques and yield analysis," *Proc. IEEE*, vol. 86, no. 9, pp. 1819–1838, 1998.



Naoki Hayase received the B.E. degree in electrical engineering from Tokyo University of Science, Japan in 2018, and the M.E. degree in electrical engineering and information systems from the University of Tokyo, Japan in 2020. He has been an Assistant Professor at University of Hyogo since 2025. His research interests include thin solid films, interfacial engineering, X-ray optics, and 2D electronics.