Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 14
TPCEE 2026

A Hybrid Sizing and Supply Voltage Optimization Methodology for Energy-Efficient CMOS Digital Circuits

Zihao She [1, *]

[1] College of Electrical Engineering and Control Science, Nanjing Tech University, Nanjing, China

* Corresponding Author Email: 202421509048@njtech.edu.cn

Abstract. This paper presents a three-stage hybrid methodology for energy-efficient CMOS [Complementary Metal-Oxide-Semiconductor] digital circuit optimization under an explicit propagation-delay constraint. The method progressively explores transistor sizing and supply-voltage design variables rather than enumerating their full combinatorial space. First, Logical Effort is applied at the nominal supply voltage to obtain an initial sizing configuration and identify stages with high electrical effort. Second, timing-sensitivity-guided local sizing updates are performed to exploit available timing slack and reduce unnecessary capacitive loading. Third, the supply voltage is searched over a discrete range, while the transistor sizes are re-evaluated when voltage changes affect critical-path timing. A candidate configuration is accepted only when it satisfies the delay constraint and provides an energy improvement, and the search terminates when the improvement falls below the specified convergence criterion or the iteration limit is reached. The methodology is evaluated on a transistor-level 4-bit absolute-value detector using circuit-level SPICE simulation. With a timing allowance of 1.4 times the minimum-delay baseline, the proposed joint sizing and supply-voltage optimization reduces normalized energy from 120.83 to 50.83, corresponding to a 57.93% reduction, while meeting the target delay of 55.69. The results show that jointly optimizing device dimensions and supply voltage provides a more effective energy-delay trade-off than optimizing either variable independently.

Keywords: CMOS digital circuits; transistor sizing; dynamic energy; propagation delay; energy-efficient design.

1. Introduction

Continuous CMOS technology scaling has enabled increasingly complex digital systems, but the resulting growth in transistor density and switching activity has made energy efficiency a first-order design constraint, particularly for battery-powered and energy-constrained systems. Consequently, circuit-level techniques that reduce energy while preserving acceptable timing performance have become increasingly important.

Supply-voltage scaling is one of the most effective approaches for reducing CMOS dynamic energy because the dynamic component has an approximately quadratic dependence on the supply voltage. However, reducing V_{DD} also reduces transistor overdrive and drive current, which increases propagation delay. The design problem is therefore to identify an operating voltage that reduces energy without violating the timing requirement [1].

Transistor sizing provides a second important design variable. Borah et al. studied transistor sizing for minimum power under a specified delay constraint and showed that the optimum sizes depend strongly on circuit loading and timing requirements [2]. Logical Effort provides a systematic analytical framework for estimating delay and selecting appropriate gate sizes [3]. Pénzes et al. further formulated transistor sizing as an energy-delay optimization problem [4]. While Aizik and Kolodny demonstrated that nonuniform, circuit-aware sizing is important for obtaining energy-efficient solutions under delay constraints [5].

Energy-delay metrics can be used to formulate circuit optimization as an explicit performance constraint rather than as an informal trade-off [6]. Kabbani demonstrated a Logical-Effort-based framework for joint sizing and supply-voltage optimization [7]. These results, together with

Content from this work may be used under the terms of CC BY-NC 4.0 licence (https://creativecommons.org/licenses/by-nc/4.0/).
Published by Warwick Evans Publishing.

established low-power CMOS design principles, motivate treating sizing and voltage as coupled variables rather than as independent optimization stages [8].

This paper proposes a three-stage co-optimization flow. First, Logical Effort generates an initial sizing solution at the nominal supply voltage. Second, a timing-sensitivity-guided local sizing procedure modifies selected stages to exploit available timing slack while reducing capacitive energy. Third, a discrete V_{DD} sweep searches for the lowest-energy feasible operating point, with sizing updated when the voltage change alters the critical-path timing. The process is repeated until the energy improvement between successive iterations falls below the convergence criterion or the iteration limit is reached. The method is evaluated on a 4-bit absolute-value detector, with the CMOS implementation and circuit analysis following established transistor-level design principles [9].

2. Methodology and Problem

2.1. CMOS Dynamic Energy

Dynamic energy is mainly associated with charging and discharging capacitive nodes during logic transitions. For a CMOS circuit, the average dynamic power can be approximated by

$$P_{dyn} = \alpha\, C_L\, V_{DD}{}^2 f \tag{1}$$

where α is the switching activity factor, CL is the effective load capacitance, VDD is the supply voltage, and f is the operating frequency. This conventional dynamic-power relationship is widely used in low-power CMOS analysis [8], [9]. Since the dynamic power in (1) is drawn over a clock period T = 1/f, the energy associated with one switching event can be written as

$$E_{dyn} = P_{dyn} T = \alpha\, C_L V_{DD}{}^2 \tag{2}$$

Thus, (2) represents energy per switching event and is independent of frequency, as expected for a per-transition energy metric. Reducing V_{DD} therefore provides a strong mechanism for lowering dynamic energy. At the same time, increasing transistor width increases gate and diffusion capacitances, which can increase switching energy even though the larger device may reduce the effective resistance.

For circuit-level evaluation, the energy consumed during a specified switching event is obtained from the instantaneous supply power:

$$E = \int [t1, t2]\, P(t)\, dt \tag{3}$$

The integration interval [t1, t2] is chosen to cover the entire switching event of interest, ensuring that the supply current returns to its quiescent value before t2. The same interval, input transition, and loading conditions are used for all candidate configurations to ensure a fair comparison.

2.2. CMOS Propagation Delay

For a simplified long-channel CMOS gate, the propagation delay can be approximated by the RC model

$$t_p \approx 0.69 R_{eq} C_L \tag{4}$$

where R_{eq} is the effective resistance of the conducting transistor network and C_L is the effective load capacitance. Equation (4) is used only for analytical insight and initial sizing guidance. Because the target circuit is evaluated at a deep-submicron technology node, short-channel effects and velocity saturation can make the simple RC relationship inaccurate[10]. Therefore, the final delay values of

candidate designs are obtained from transistor-level SPICE simulations using the BSIM device model for the target technology.

Increasing transistor width generally decreases R_{eq} and can improve speed, but it simultaneously increases parasitic capacitance. Consequently, excessive upsizing produces diminishing delay benefits and can increase energy. Similarly, reducing V_{DD} lowers the available overdrive and drive current, increasing propagation delay. These opposite trends make delay a natural constraint in the proposed energy-minimization problem.

2.3. Logical Effort and Transistor Sizing

Logical Effort is used to obtain an analytical starting point for transistor sizing [3]. The normalized delay of a logic stage is expressed as

$$d = g h + p \tag{5}$$

where g is the logical effort, h is the electrical effort, and p is the parasitic delay. The electrical effort is

$$h = C_L / C_{in} \tag{6}$$

where C_{in} is the input capacitance of the gate and C_L is the total capacitance driven by the stage, including the fanout and other effective loading terms. Logical Effort therefore provides a compact way to estimate stage delay and distribute driving capability along a path [3].

In this work, Logical Effort is not treated as the final circuit model. Instead, it supplies the initial sizing direction, after which candidate configurations are evaluated using transistor-level simulation. This combination avoids blind enumeration of the full sizing space while retaining circuit-level accuracy.

2.4. Energy–Delay Trade-off and Optimization Problem

Energy and propagation delay exhibit conflicting trends. Lowering V_{DD} reduces dynamic energy but increases delay, whereas increasing transistor size can reduce resistance but increases capacitance. The optimization therefore treats delay as an explicit constraint and minimizes energy inside the feasible timing region:

$$in_{W_i, V_{DD}} E(W_i, V_{DD}) \tag{7}$$

$$subjecttot_p(W_i, V_{DD}) \leq t_{max} \tag{8}$$

where W_i denotes the size of the i-th transistor or sizing variable, V_{DD} is the supply voltage, and t_{max} is the maximum allowable propagation delay. The experimental timing target is defined as 1.4 times the measured minimum-delay baseline. With a baseline delay of 39.78, the target is approximately 55.69 normalized time units.

The optimization uses a progressive search rather than an unspecified generic iteration. The initial sizing is obtained analytically, timing-sensitive stages are then adjusted locally, and V_{DD} is swept over a discrete range. A sizing change is retained only when it decreases energy while satisfying the delay constraint. The process terminates when the energy improvement between successive accepted configurations is less than 0.1% or when 20 iterations have been completed.

3. Case Study: A 4-bit Absolute Value Detector

A 4-bit absolute-value detector is selected to evaluate the proposed methodology. The circuit processes a 4-bit two's-complement input A = A3A2A1A0 and a 3-bit threshold T = T2T1T0. A3 is

the sign bit. The logic network contains multiple stages implemented using AND, OR, XOR, and buffer functions, providing different electrical efforts, path lengths, and capacitive loads.

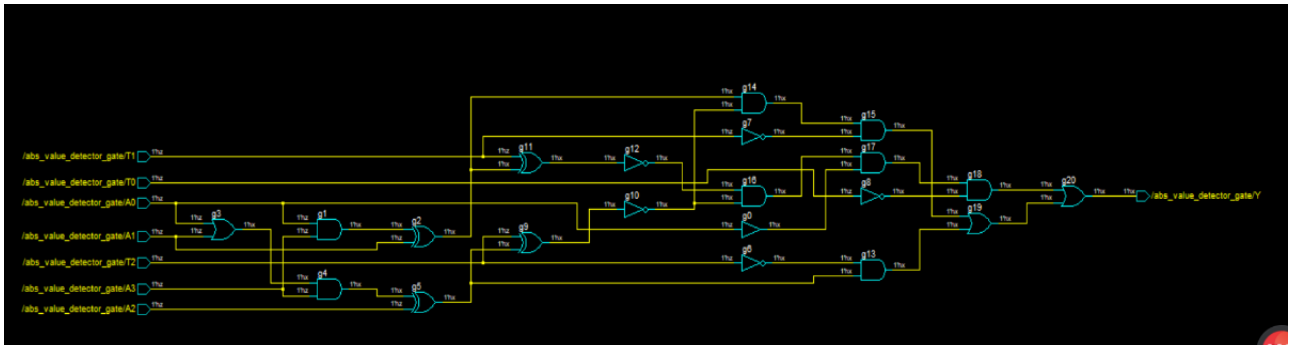

Figure 1. Gate-level schematic of the 4-bit absolute-value detector, showing inputs A [3:0], T [2:0], internal logic stages, and output Y. (Picture credit: Original)

Figure 1 shows the gate-level architecture used as the case-study circuit, including the seven primary inputs and the output Y.

3.1. Circuit Architecture

The seven primary inputs are A3, A2, A1, A0, T2, T1, and T0, and the circuit produces one output Y. The architecture is implemented as a combinational CMOS network. Because different input transitions activate different logic paths, the worst-case delay is determined by exhaustive or sweep-based evaluation of the output-changing transitions rather than by a single nominal input pattern.

3.2. Transistor-Level CMOS Implementation

The reference CMOS inverter used for the normalization basis is shown in figure 2.The logic functions are implemented using complementary CMOS transistor networks. PMOS devices form the pull-up networks and NMOS devices form the pull-down networks. The reference unit inverter uses an NMOS width of 200 nm, a PMOS width of 300 nm, and a channel length of 100 nm. The transmission-gate XOR uses Wp = Wn = 200 nm as the unit transmission-gate width. These values define the normalization basis for the sizing variables.

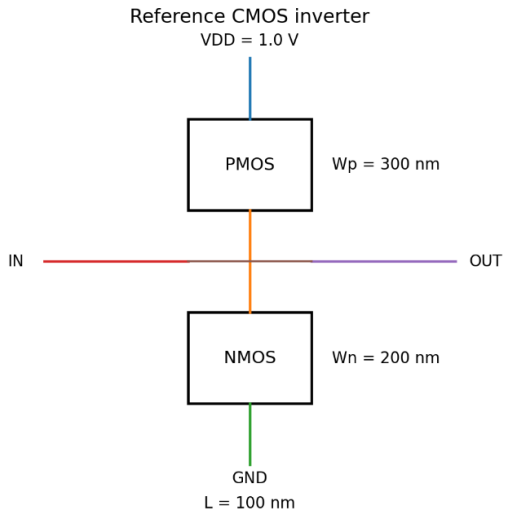

Figure 2. Reference CMOS inverter used for sizing normalization, with Wp = 300 nm, Wn = 200 nm, and L = 100 nm. (Picture credit: Original)

The design does not scale all devices uniformly. Timing-sensitive stages are assigned larger sizes when necessary, whereas non-critical stages remain near the unit size to avoid unnecessary capacitance. The final sizing variables used in the hybrid optimization are reported as relative size factors with respect to the reference unit.

3.3. Critical Path and Delay Evaluation

The worst-case propagation delay is obtained by applying the specified input transitions and measuring the corresponding output transition. Both rising and falling output transitions are considered. The largest measured value is used as the representative circuit delay. This procedure prevents the optimization from focusing on a non-critical path.

For the baseline configuration, the minimum measured delay is 39.78 normalized time units. The allowable delay for energy optimization is therefore 55.69, corresponding to a 40% relaxation relative to the minimum-delay design.

3.4. Supply Voltage Sweep

The supply voltage is swept over the feasible range from 0.80 V to 1.00 V with a 0.01 V step. At each voltage, the same input transitions, output loading, and measurement interval are retained. Configurations whose worst-case delay exceeds 55.69 are rejected. Among the remaining points, the minimum-energy operating point is selected.

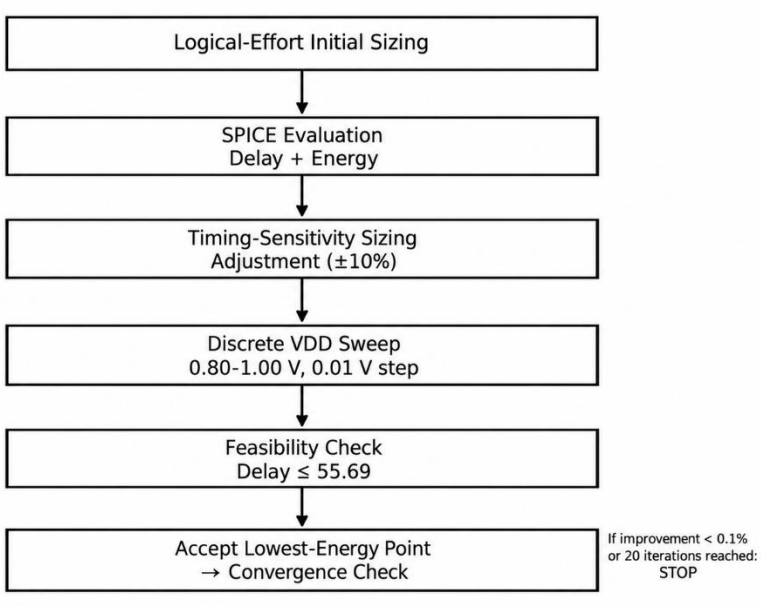

Figure 3. Flowchart of the proposed hybrid transistor-sizing and supply-voltage optimization procedure. (Picture credit: Original)

This discrete sweep is subsequently repeated after accepted sizing adjustments because a change in transistor size modifies the minimum feasible supply voltage. The complete iterative procedure is summarized in figure 3.

3.5. Hybrid Sizing and Voltage Optimization

The co-optimization flow consists of explicit, reproducible steps. First, the unit-inverter and Logical Effort model generate the initial transistor-size factors. Second, the circuit is simulated to establish its delay and energy. Third, the stage with the largest timing sensitivity is adjusted: if the delay violates the target, its size is increased by 10%; if sufficient timing slack exists and energy can be reduced, a non-critical stage is reduced by 10%. Each candidate is simulated before it is accepted. Fourth, V_{DD} is swept from 0.80 V to 1.00 V in 0.01 V increments, and the lowest-energy feasible point is retained. The sizing and voltage steps are repeated until the relative energy improvement is below 0.1% or 20 iterations are reached.

Initial Logical-Effort Sizing → SPICE Evaluation → Timing-Sensitivity Sizing Adjustment → V_{DD} Sweep → Feasibility Check → Accept Lowest-Energy Point → Convergence Check

This procedure explicitly specifies the direction and magnitude of sizing adjustment, the voltage-search step, and the convergence criteria. The use of local updates limits unnecessary exploration while preserving the coupling between sizing and supply voltage.

3.6. Energy Measurement and Comparison

For each candidate configuration, energy is calculated from the supply voltage and supply current over the same switching interval:

$$Emin_{W_i,V_{DD}} E(W_i, V_{DD}) \tag{9}$$

4. Experimental Results and Discussion

4.1. Baseline Performance and Timing Constraint

The baseline design operates at V_{DD} = 1.00 V and has a measured delay of 39.78 normalized time units and energy of 120.83 normalized energy units. The prescribed 1.4× timing target is therefore 55.69. This timing relaxation provides the design freedom required to trade speed for lower energy.

4.2. Comparison of Optimization Strategies

Table 1 compares the baseline, sizing-only, voltage-only, and joint sizing-plus-voltage configurations under the same timing constraint.

Table 1. Comparison of baseline and optimized configurations.

Design	Delay	Energy	V_{DD}	ΔDelay	ΔEnergy
Baseline	39.78	120.83	1.00 V	0%	0%
W only	55.69	61.55	1.00 V	+40%	−49.06%
VDD only	55.69	78.82	0.81 V	+40%	−34.77%
W + VDD	55.69	50.83	0.856 V	+40%	−57.93%

The sizing-only configuration reduces energy from 120.83 to 61.55, corresponding to a 49.06% reduction, while the delay increases to the 55.69 target. The voltage-only configuration reduces V_{DD} to 0.81 V and decreases energy to 78.82, corresponding to a 34.77% reduction. The joint W + V_{DD} configuration achieves the lowest energy of 50.83 at an optimized V_{DD} of 0.856 V, corresponding to a 57.93% reduction from the baseline while satisfying the same timing constraint.

The results demonstrate that the two optimization variables provide complementary benefits. Sizing primarily controls the resistance and capacitance of individual stages, while voltage scaling reduces the voltage-dependent switching energy. Joint optimization therefore explores operating points that cannot be obtained efficiently by optimizing either variable alone.

4.3. Optimized Sizing

The hybrid optimization produces nonuniform relative sizing factors. The principal optimized variables reported by the optimization procedure are approximately 2.7379, 7.8343, 25.1954, and 109.7851. The nonuniform distribution is consistent with the different electrical efforts and timing sensitivities of the corresponding stages. Large sizes are reserved for stages where additional drive capability provides useful timing benefit, while smaller stages avoid unnecessary capacitive energy.

5. Conclusion

This paper presented a hybrid methodology for energy-efficient CMOS digital circuit design through the joint optimization of transistor sizing and supply voltage. The proposed flow combines Logical-Effort-based initial sizing, timing-sensitivity-guided local sizing updates, and discrete supply-voltage

search. Unlike independent sizing and voltage optimization, the proposed procedure explicitly updates both variables to account for their coupling through resistance, capacitance, and delay.

A 4-bit absolute-value detector was used as the case study. The baseline design at 1.00 V exhibits a delay of 39.78 and energy of 120.83 in the adopted normalized metrics. With a timing allowance of 1.4 times the minimum delay, the target delay becomes 55.69. Sizing-only optimization reduces energy by 49.06%, while voltage-only optimization reduces energy by 34.77%. The proposed joint $W + V_{DD}$ optimization achieves the best result, reducing energy to 50.83, corresponding to a 57.93% reduction, at an optimized supply voltage of 0.856 V and the specified 55.69 delay target.

These results confirm that significant energy savings can be obtained by exploiting timing slack and jointly adjusting device dimensions and supply voltage. Future work will extend the methodology to larger combinational networks and will consider leakage energy, process variation, temperature, and more detailed technology-dependent device models.

References

[1] S. Hanson, D. Sylvester, and D. Blaauw, “A new technique for jointly optimizing gate sizing and supply voltage in ultra-low energy circuits,” in Proc. Int. Symp. Low Power Electron. Design (ISLPED), 2006, pp. 338–341, doi: 10.1145/1165573.1165653.

[2] M. Borah, R. M. Owens, and M. J. Irwin, “Transistor sizing for minimizing power consumption of CMOS circuits under delay constraint,” in Proc. Int. Symp. Low Power Design, 1995, pp. 167–172, doi: 10.1145/224081.224111.

[3] I. Sutherland, R. F. Sproull, and D. Harris, Logical Effort: Designing Fast CMOS Circuits. San Francisco, CA, USA: Morgan Kaufmann, 1999.

[4] P. I. Pénzes, M. Nyström, and A. J. Martin, “Transistor sizing of energy-delay-efficient circuits,” in Proc. Int. Symp. Low Power Electron. Design, 2002, pp. 333–338, doi: 10.1145/589411.589439.

[5] Y. Aizik and A. Kolodny, “Finding the energy efficient curve: Gate sizing for minimum power under delay constraints,” VLSI Design, vol. 2011, Art. no. 845957, 2011, doi: 10.1155/2011/845957.

[6] M. Alioto, E. Consoli, and G. Palumbo, “From energy-delay metrics to constraints on the design of digital circuits,” Int. J. Circuit Theory Appl., vol. 40, no. 8, pp. 815–834, 2012, doi: 10.1002/cta.757.

[7] A. Kabbani, “Logical effort based dynamic power estimation and optimization of static CMOS circuits,” Integration, the VLSI Journal, vol. 43, no. 3, pp. 279–288, 2010, doi: 10.1016/j.vlsi.2010.02.002.

[8] A. P. Chandrakasan, S. Sheng, and R. W. Brodersen, “Low-power CMOS digital design,” IEICE Trans. Electron., vol. E75-C, no. 4, pp. 371–382, 1992.

[9] N. H. E. Weste and D. M. Harris, CMOS VLSI Design: A Circuits and Systems Perspective, 4th ed. Boston, MA, USA: Pearson, 2011.

[10] J. M. Rabaey, A. P. Chandrakasan, and B. Nikolić, Digital Integrated Circuits: A Design Perspective, 2nd ed. Upper Saddle River, NJ, USA: Pearson Education, 2003.