The Fundamental Limit on Binary Switching Energy for Terascale Integration (TSI)

James D. Meindl and Jeffrey A. Davis

***Abstract*—The fundamental limit on signal energy transfer during a binary switching transition is $E_s(\min) = (ln2)kT$. Derivation of precisely this result based on two entirely distinct physical models confirms its validity. The first model is an ideal MOSFET operating in a CMOS inverter circuit at the limit of its capacity for binary signal discrimination. The second is an isolated interconnect treated as a noisy communication channel. Both models are relevant to modern terascale integration.**

***Index Terms*—Binary switching energy limits, fundamental limits, gigascale integration limits, limits, terascale integration limits.**

I. Introduction

THROUGHOUT the past four decades, semiconductor technology has advanced at exponential rates in both performance and productivity. In the real world, such exponential advances do not continue endlessly. Consequently, the intent of this paper is to establish conclusively that *the* fundamental limit on signal energy transfer during a binary switching transition is $E_s(\min) = (ln2)kT$, where k is Boltzmann's constant and T is absolute temperature. This limit is characterized as fundamental because its value is patently independent of the properties of any materials, devices, or circuits that may be used to implement the binary transition [1].

The first statement of this limit known to the authors is attributed to John von Neumann who "computed the thermodynamical minimum of energy per elementary act of information from the formula $kT\log_e N$" where $N = 2$ for a binary act [2]. Keyes observes that "the report of von Neumann's ideas fails to provide any justification of this assertion or explanation of the reasoning underlying it" [3].

Landauer analyzed a hypothetical binary device consisting of a particle in a bistable potential well [4]. Presence of the particle in the left-hand well defines the *zero* state and presence in the right-hand well, the *one* state. Calculation of the energy dissipation involved in restoring the particle to the *one* state "shows that if the lowering (of the potential barrier) of the right-hand well is carried out very slowly, the energy dissipation will, indeed, be $kT(\log_e 2)$" [3].

Based on the work of Swanson and Meindl [5], Section II of this paper rigorously derives the result that the minimum switching energy of an ideal MOSFET operating in an inverter circuit suitable for terascale integration (TSI) is $E_s(\min) = (\log_e 2)kT = (ln2)kT$. Precisely the same result is derived in Section III by treating an isolated interconnect, also suitable for TSI, as a communication channel described by Shannon's theorem for channel capacity [6]. The implication of the fundamental limit on binary switching energy for the minimum size of MOSFETs in TSI is discussed in Section IV. The conclusiveness of the limit is discussed in Section V to close the paper.

Manuscript received February 18, 2000; revised May 15, 2000. This work was supported by the Advanced Research Project Agency under Contract F33615-97-C1132 and the Semiconductor Research Corporation under Contract HJ-374.

The authors are with Georgia Institute of Technology, Atlanta, Georgia 30332-0269 USA (e-mail: james.meindl@ece.gatech.edu).

Publisher Item Identifier S 0018-9200(00)08708-4.

II. Minimum Switching Energy of a MOSFET

Lo [7] provides a convincing heuristic argument that establishes the following critical point. In order to fulfill the quintessential requirement of binary signal discrimination, the slope of the static transfer curve of a (CMOS) binary logic gate must be greater than unity in absolute value at the transition point where input and output voltage levels are equal. Imposing this requirement on the transfer curve of a CMOS inverter circuit with matched n- and p-channel MOSFET drain-current characteristics, the *minimum allowable* supply voltage of the circuit was derived by Swanson and Meindl [5] as

$$V_{dd}(\min) = 2(kT/q)[1 + C_{fs}/(C_{ox} + C_d)]\ln(2 + C_d/C_{ox}) \tag{1}$$

where C_{fs} is the fast surface state capacitance per unit area, C_{ox} is the gate-oxide capacitance per unit area, and C_d is the channel depletion region capacitance per unit area. In an ideal MOSFET (i.e., a device with a subthreshold slope of 60 mV/decade at 300° K) $C_{fs} \lll C_{ox}$ and $C_d \lll C_{ox}$ so that

$$V_{dd}(\min) \cong 2\,(\ln 2)kT/q = 1.38kT/q = 0.036 \text{ V at } 300^\circ\text{K}. \tag{2}$$

The signal energy stored on the gate capacitance C_g of a single MOSFET operating in a CMOS inverter circuit can be expressed as $E_s = (1/2)Q_gV_{dd}$ where Q_g is the gate charge. Assuming the minimum possible gate charge of a single electron $Q = q$ and the minimum allowable supply voltage given by (2) gives the minimum signal energy transfer during a binary switching transition of a MOSFET as

$$E_s(\min) = (ln2)kT = 0.693kT = 0.0179 \text{ eV at } 300^\circ\text{K}. \tag{3}$$

A CMOS inverter binary transition involves simultaneous switching transitions of two MOSFETs and consequently the minimum signal energy transfer of the circuit is triple the value given by (3). This result assumes that the p-channel width is double the n-channel width in order to compensate for electron and hole mobility differences and, therefore, obtain a symmetrical inverter transfer curve.

III. Minimum Switching Energy of an Interconnect

Shannon's theorem [6] states that the maximum capacity of a communication channel contaminated by a white noise source is given by

$$C \leq B \log_2[1 + P_s/P_n] \quad \text{b/s} \tag{4}$$

where B is the channel bandwidth, P_s is the average signal power and P_n is the average noise power. Solving (4) for the energy per bit gives

$$E_{\text{bit}} = P_s/C \geq (P_n/C)[2^{C/B} - 1]. \tag{5}$$

Substituting P_n = kTB for the thermal noise power into (5) gives

$$E_{\text{bit}} = kT[2^{C/B} - 1][C/B]^{-1}. \tag{6}$$

Setting the derivative of (6) equal to zero or $dE_{\text{bit}}/d(C/B) = 0$ and employing L'Hospital's rule gives

$$E_{\text{bit}}(C/B \to 0) = E_{\text{bit}}(\min) = (ln2)kT. \tag{7}$$

Note that $(C/B) \to 0$ is tantamount to calculating the energy transfer of an infinitely long bit or a single binary transition.

The exact agreement between (3) and (7) conclusively establishes the value of the fundamental limit on signal energy transfer during a binary switching transition in terms that are most relevant to TSI. The remarkable agreement of this result with the prior efforts of von Neumann [2], Landauer [4], and Keyes [3] is indeed supportive.

IV. Implications for MOSFET Size

Consider the expressions for switching energy $E_s = (1/2)(Q_g)^2/C_g$ and gate capacitance $C_g = \varepsilon_{ox}(L_{\min})^2/T_{ox}$ where Q_g is gate charge, ε_{ox} is gate-oxide permittivity, T_{ox} is gate-oxide thickness, and $L_{\min}$ is the channel length of a minimum size MOSFET whose binary switching energy is $E_s(\min)$ given by (3) and whose gate charge is a single electron $Q_g = q$. Combining these expressions gives

$$\begin{aligned} L_{\min} &= \{2[E_s(\min)][T_{ox}/\varepsilon_{ox}][1/V_{dd}(\min)]^2\}^{1/2} \\ &= \{[T_{ox}/\varepsilon_{ox}]q^2/[2\,(ln2)kT]\}^{1/2}. \end{aligned} \tag{8}$$

Assuming the minimum value of gate-oxide thickness necessary for retention of the bulk properties of SiO_2, $T_{ox} = 1.5$ nm [8]. Substituting this value into (8) gives

$$L_{\min} = 13.9 \text{ nm} \tag{9}$$

as the benchmark channel length of a minimum switching energy single-electron MOSFET. Since $E_s(\min) = (ln2)kT \cong kT$, a descriptive term for this device is a "single electron, thermal energy" or SETE MOSFET. Further reductions in channel length through increasing ε_{ox} or V_{dd} are feasible.

V. Conclusion

The central thesis of this paper is that *the* fundamental limit on signal energy transfer during a binary switching transition is $E_s(\min) = (ln2)kT$. Derivation of precisely this result based on two entirely distinct physical models confirms its validity. The first model is an ideal MOSFET operating in a CMOS inverter circuit at the limit of its capacity for binary signal discrimination. The second is an isolated interconnect treated as a noisy communication channel. These models are relevant to modern TSI. The fundamental limit that they serve to establish is in full accord with prior results [2]–[4] predicated again on entirely different physical models. This further agreement serves well to bolster confidence in the fundamental limit on signal energy transfer during a binary switching transition.

References

[1] J. D. Meindl, "Low power microelectronics: Retrospect and prospect," *Proc. IEEE*, vol. 83, no. 4, pp. 619–635, Apr. 1995.

[2] J. von Neumann, *Theory of Self-Reproducing Automata*. Urbana, IL: Univ. Illinois Press, 1966, p. 66.

[3] R. W. Keyes, "Physical limits in digital electronics," *Proc. IEEE*, vol. 63, no. 5, pp. 740–767, May 1975.

[4] R. W. Landauer, "Irreversibility and heat generation in the computing process," *IBM J. Res. Develop.*, vol. 5, pp. 183–191, 1961.

[5] R. M. Swanson and J. D. Meindl, "Ion-implanted complementary MOS transistors in low-voltage circuits," *IEEE J. Solid-State Circuits*, vol. SC-7, pp. 1146–1152, Apr. 1972.

[6] C. Shannon, "The mathematical theory of communication," *Bell Syst Tech. J.*, vol. 27, pp. 379–423, Mar. 1948.

[7] A. W. Lo, "Physical realization of digital logic circuits," in *Micropower Electronics*, E. Keonjian, Ed. New York, NY: Macmillan, 1964, pp. 19–39.

[8] D. A. Muller *et al.*, "The electronic structure at the atomic scale of ultrathin gate oxides," *Nature*, vol. 399, pp. 758–761, June 24, 1999.