Fundamental limits of energy dissipation in charge-based computing

Graham P. Boechler, Jean M. Whitney,[a] Craig S. Lent, Alexei O. Orlov, and Gregory L. Snider[b]
Department of Electrical Engineering, University of Notre Dame, Notre Dame, Indiana 46556, USA

(Received 4 June 2010; accepted 29 July 2010; published online 7 September 2010)

According to Landauer's principle, dissipation of energy is only necessary when information is erased, suggesting that vastly more efficient logical switches than transistors are possible. However, an influential analysis of binary switching suggests that representing information with electric charge is the root of the problem, that Landauer's principle is fundamentally flawed, and that any movement of charge, such as charging a capacitor, must dissipate at least $k_BT \ln(2)$. Here, using a RC circuit, an energy loss of much less than $k_BT \ln(2)$ is demonstrated while delivering energy of 100 $k_BT \ln(2)$ to the capacitor. This shows that there is no fundamental lower limit to energy dissipation in moving charge. © *2010 American Institute of Physics*. [doi:10.1063/1.3484959]

Is charge the problem? Modern computers use electric charge to physically represent information, and the manipulation and movement of this charge is used to perform computations. Digital transistors, however, dissipate energy as heat at a rate which currently limits scaling and packing density. Modern digital devices based on field-effect transistors (FETs) operate by moving electrons on and off a conducting gate, which modulates a barrier to current flowing through a channel. Scaling these structures to smaller sizes and faster switching speeds has fueled the exponential improvements in performance captured by "Moore's law." But we have now reached a point where heat dissipation limits further improvements in device speed and so alternatives such as multiple on-chip computational cores are becoming the preferred way to extract more performance. The semiconductor industry has therefore, perhaps belatedly, turned to a search for eventual replacements to the CMOS (complementary metal-oxide-semiconductor) FET. Candidates for the "next switch" include devices based on spin, molecular state, phase states, polarization, magnetization, collective effects, and sometimes electronic charge. Suspicion has fallen on charge as being intrinsic to the dissipation problem, and it is asserted that the key problem is moving charge—that modulating the potential on conductors necessarily entails energy dissipation of order k_BT per switching event.[1,2]

The oft-quoted analysis of Cavin *et al.*[1,2] has been very influential in the charge against charge. The Nanoelectronics Research Initiative, formed by the Semiconductor Industry Association and operated by the Semiconductor Research Corporation (SRC) based their conclusion that "to move any technology substantially beyond CMOS would require employing an alternative state variable other than electronic charge as the information token and/or new logic switching mechanisms that use significantly less power per computation,"[3] on references to the Zhirnov–Cavin analysis (see also Ref. 4).

Arguments about a fundamental connection between computation and heat dissipation have a long history. Landauer argued[5,6] that energy dissipation in computation is only required when information is erased, at a necessary cost of $k_BT \ln(2)$ per bit, an idea now referred to as Landauer's principle. Bennett then showed[7] that any logically irreversible computation could be in principle embedded in a logically reversible computation, and thus that there was no fundamental lower limit to the amount of energy that must be dissipated. In analyzing necessary dissipation for a switching event, Landauer considered a bistable two-state system with an energy barrier considerably larger than k_BT. He argued that one could switch a known bit by gradually changing the potential landscape to move from the original bistable state to a monostable state, and then back to the bistable state in the switched configuration, as shown schematically in Fig. 1 (after Ref. 5). Key to this Landauer switching process are (a) biasing the system toward the state it is already in, thus needing a copy of the bit outside the system to provide the biasing (this is crucially missing from the "Cavin's Demon" analysis of Ref. 1), and (b) being able to raise and lower potential barriers with no fundamental minimum energy cost. Direct calculation of such systems in a thermal environment bear out the Landauer/Bennett analysis—minimal dissipation is necessary unless information is erased.[8,9]

Cavin and Zhirnov have criticized Landauer's analysis because it relies on the ability to control the system at energy scales smaller than k_BT by gradually changing energy barriers. This can be done by moving charge on and off of electrodes, i.e., capacitors, again gradually. A controlled voltage source changes the potential on the electrodes—no current is meant to flow across a barrier. But unless the electrodes and wires are superconducting, there will always be some small residual resistance in the wires, and thus some residual dissipation. To maintain the signal in a thermal environment the barriers must at some point be raised to levels several times larger than k_BT. The key question is whether a charging process can create energy barriers larger than k_BT, yet keep the residual dissipation significantly less than k_BT. The SRC group[1,2] argues that this is impossible, thus indicting any charge-based computational switch. Here we present the first direct measurement at a sub-k_BT energy scale of the dissipation involved in charging a capacitor adiabatically. We show that charging energies of many times k_BT can be obtained with an energy dissipation much less than k_BT.

We note that well-established previous work on adiabatic logic had as its goal lowering power dissipation in conven-

[a]Currently at the University of Wisconsin, Madison, USA.
[b]Electronic mail: snider.7@nd.edu.

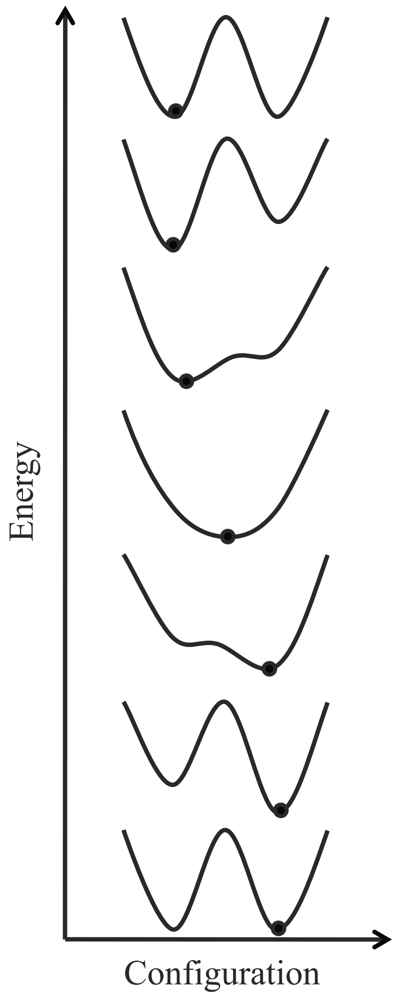

FIG. 1. Schematic energy profile at various times during Landauer switching of a known bit (after Ref. 5). The potential barrier between the states is much larger than k_BT for the initial and final states but is reduced to zero at the intermediate switching phase. Using this adiabatic switching approach minimal energy can be dissipated [less than $k_BT\ln(2)$], while maintaining a high barrier for robust bit discrimination in a thermal environment.

tional FET devices by using the same principles.[10] Present CMOS FET's are normally used in a mode that dissipates a relatively large amount of energy at each switching event. The energy stored when the gate, with capacitance C, is fully charged to a voltage V is $1/2CV^2$, and all of that energy is dissipated when the gate is discharged abruptly through a channel resistance. It has long been recognized that abruptly discharging the capacitor is wasteful and that by charging and discharging the capacitor gradually (adiabatically) the power dissipation can be significantly lowered, albeit at the cost of lowering the switching speed.[11] Because of the channel resistances and consequent charging times involved, this approach, termed adiabatic CMOS logic, has not proven to be a very attractive. The reduction in power dissipation in such circuits has been measured by thermoelectric techniques[12] but not at a level that could resolve energies of a few k_BT and thus address these fundamental questions. The question of taking adiabatic switching into the sub-k_BT domain arises in considering nontransistor-based alternatives for nanoelectronic information processing and the extreme demands of nanoscale packing densities.

The circuit to be considered here is a simple RC circuit, as shown in Fig. 2(a). The goal of the experiment is to see if an energy much greater than k_BT can be delivered to the

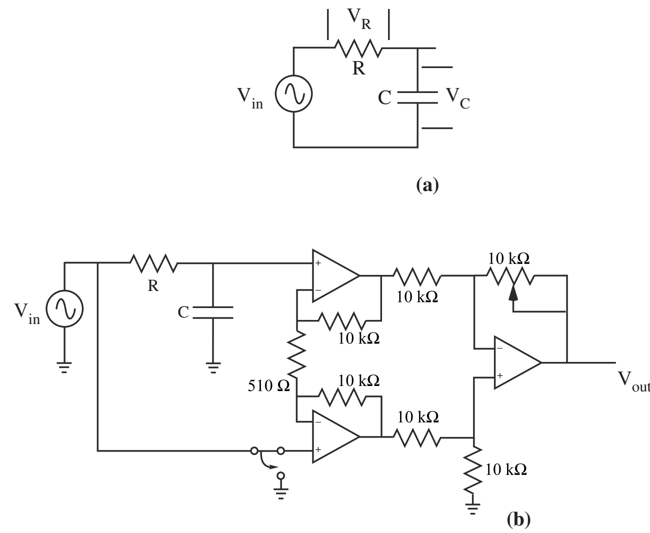

FIG. 2. (a) Schematic of RC circuit used in the experiment. (b) Schematic of the experimental setup including the RC circuit and instrumentation amplifier. The two first stage op-amps are from a ST TL072, and the second stage is a TI TLE 2141 op-amp.

capacitor while dissipating much less than k_BT in the resistor. Instead of abruptly charging and discharging the capacitor we consider a sinusoidal input voltage $V_{in}=V_a\sin(\omega t)$, where V_a is the amplitude and ω is the applied angular frequency and $\omega_0\equiv 1/(RC)$ is the characteristic angular frequency of the circuit. The energy delivered to the capacitor in one period is

$$E_C(\omega) = CV_a^2\left[\frac{1}{1+(\omega/\omega_0)^2}\right], \tag{1}$$

$$E_C(\omega \ll \omega_0) = CV_a^2. \tag{2}$$

During each half of the period $1/2E_C$ is transferred on and off the capacitor. The energy dissipated in the resistor in one period is

$$E_R(\omega) = \frac{1}{2}\frac{V_a^2}{R}\frac{2\pi}{\omega}\left[\frac{(\omega/\omega_0)^2}{1+(\omega/\omega_0)^2}\right], \tag{3}$$

$$E_R(\omega \ll \omega_0) = \pi\frac{\omega}{\omega_0}[CV_a^2]. \tag{4}$$

Equation (4) shows that at low frequencies the energy dissipated in the resistor can be much lesser than the energy stored on the capacitor. Zhirnov and Cavin argue that this elementary analysis fails when the energies approach those of thermal fluctuations with the consequence that the dissipation must be at least $k_BT\ln(2)$. We here put this to the test of direct experimental measurement.

Circuit parameters were chosen to accommodate the need for measuring small energy dissipation. The resistance in an adiabatic control electrode is deliberately very low—just the resistance of the conductors themselves.[13] In our experiment, a much larger resistor is used to make the dissipation measurable at reasonable frequencies. Other considerations in the choice of the resistor and capacitor are the 100 kHz maximum measurement frequency of the lock-in amplifier, and the requirement that the frequencies where E_R is less than $k_BT\ln(2)$ be high enough that 1/f noise from the instrumentation amplifier used in measurement does not degrade

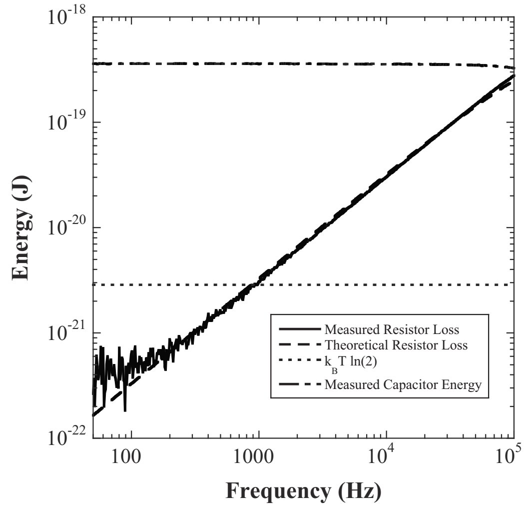

FIG. 3. Plot of the measured energy delivered to the capacitor over one period, along with the measured energy loss in the resistor and the theoretical loss in the resistor, as a function of frequency. A horizontal line shows $k_B T \ln(2)$ at room temperature. The energy lost in the resistor is less than $k_B T \ln(2)$ for frequencies less than 900 Hz.

the signal. In addition the capacitor value should be chosen small enough that the energy delivered is approximately 100 $k_B T \ln(2)$ but large enough that parasitic capacitances will not dominate. The measured values of the components used including parasitics are $R=10.01\ k\Omega$ with $C=43.7\ pF$ so that $\omega_0/2\pi=364\ kHz$.

The challenge in this measurement is to measure both the voltage across the capacitor and the voltage across the resistor. An instrumentation amplifier was built and connected to the RC circuit as shown in Fig. 2(b). This amplifier topology was chosen for its high common-mode rejection, and high input resistance, $\sim 10^{12}\ \Omega$ that causes negligible loading in the measurement of either the resistor or capacitor voltage. The input capacitance of the amplifier, approximately 4.7 pF, is measured independently and included in the calculation. The amplifier can be switched, as shown in the schematic, so that either the voltage across the capacitor or the voltage across the resistor can be measured. The voltage gain of the instrumentation amplifier is 40.0, and standard lock-in techniques are used with a time constant of 10 s to reduce the noise in the measurement.

An input voltage of $V_a=62.9\ \mu V$, with a slowly varying frequency, $f=\omega/2\pi$, from 50 Hz to 100 kHz, is applied to the RC circuit, and the measurement is carried out at room temperature. The voltage across the capacitor and resistor are measured as a function of frequency, and the energy lost in the resistor and delivered to the capacitor over one period is calculated using the following equations.

$$E_R = \frac{\pi V_R^2}{\omega R}, \quad (5)$$

$$E_C = C V_C^2, \quad (6)$$

where V_R and V_C are the amplitude of the measured voltages across the resistor and capacitor. The measured capacitor and resistor energies are plotted in Fig. 3, along with the theoretical energy for E_R, and a horizontal line at $k_B T \ln(2)$, ~ 3 zJ at room temperature. As expected, the energy delivered to the capacitor over one period is constant at low frequencies with a value of approximately 100 $k_B T \ln(2)$. The energy lost in the resistor over one period increases linearly with frequency, and is below $k_B T \ln(2)$ for frequencies below 900 Hz. At frequencies below 100 Hz the measured E_R saturates at approximately 500 yJ due to noise in the amplifier. These results clearly show that an energy much greater than $k_B T$ can be delivered to a capacitor while dissipating less than $k_B T \ln(2)$ in the resistor.

There is considerable controversy in the electronics industry over the suitability of charge as a state variable for information. The work of Cavin *et al.*[1,2] has suggested that other state variables, such as spin, are necessary because the manipulation of information in the form of charge cannot be done without excessive dissipation. Spin may or may not prove advantageous but the fundamentals of energy dissipation in a bistable system are the same regardless of the state variable. Moreover most noncharge state variables would be manipulated electronically, through either changing potentials on control electrodes or current through conductors to change applied magnetic fields.

Charge is not the problem. The results of our experiments directly demonstrate a charge transfer with energy much greater than $k_B T$, sufficient to robustly represent a bit, with an associated energy dissipation that is far less than $k_B T \ln(2)$. This proves, in agreement with Landauer's principle, that there is no fundamental lower limit for the energy dissipation in information processing that is based on the position of charge. The transistor may have hit its limit in terms of power dissipation[14] but there is no reason to think that the solutions will not involve encoding information with electric charge.

This work was partially supported by National Science Foundation under Grant Nos. CCR-0210153, CCF04-03760, and ECS-0725794. The authors also thank Aaron Prager, Hubert George, and Robin Joyce for their assistance in the measurements.

[1]R. K. Cavin, V. V. Zhirnov, J. A. Hutchby, and G. I. Bourianoff, Fluct. Noise Lett. **5**, C29 (2005).
[2]V. V. Zhirnov, R. K. Cavin, J. A. Hutchby, and G. I. Bourianoff, Proc. IEEE **91**, 1934 (2003).
[3]J. J. Welser, G. I. Bourianoff, V. V. Zhirnov, and R. K. Cavin, J. Nanopart. Res. **10**, 1 (2008).
[4]K. Galatsis, A. Khitun, R. Ostroumov, K. L. Wang, W. R. Dichtel, E. Plummer, J. F. Stoddart, J. I. Zink, J. Y. Lee, Y.-H. Xie, and K. W. Kim, IEEE Trans. Nanotechnol. **8**, 66 (2009).
[5]R. W. Keyes and R. Landauer, IBM J. Res. Dev. **14**, 152 (1970).
[6]R. Landauer, IBM J. Res. Dev. **5**, 183 (1961).
[7]C. H. Bennett, IBM J. Res. Dev. **17**, 525 (1973).
[8]C. S. Lent, Nanotechnology **18**, 298002 (2007).
[9]C. S. Lent, M. Liu, and Y. H. Lu, Nanotechnology **17**, 4240 (2006).
[10]S. G. Younis and T. F. Knight, *Research on Integrated Systems* (MIT Press, Seattle, WA, 1993), pp. 234–250.
[11]A. G. Dickinson and J. S. Denker, IEEE J. Solid-State Circuits **30**, 311 (1995).
[12]P. M. Solomon and D. J. Frank, *Proceedings of the International Symposium on Low Power Design*, San Jose, CA, 1995 (IEEE, New York, NY, 1995), pp. 18–19.
[13]E. P. Blair, E. Yost, and C. S. Lent, J. Comput. Electron. **9**, 49 (2010).
[14]J. D. Meindl and J. A. Davis, IEEE J. Solid-State Circuits **35**, 1515 (2000).