Gain in Nanoelectronic Devices

William R. Frensley
Eric Jonsson School of Engineering and Computer Science
University of Texas at Dallas
Richardson, Texas 75083-0688

Abstract

The role of signal gain in successful computational technologies is explored, with an emphasis on applying the concept to emerging nanoelectronic technologies. Gain provides a means to protect the integrity of a computation against fluctuations in the properties of the components which comprise a digital system. The adverse consequences of inadequate gain are illustrated by a Monte Carlo simulation of simple models of reversible and irreversible logic gates. The factors affecting the gain of semi-classical electronic devices are examined, and it is pointed out that the process of scaling devices down into the quantum-scale regime will makes it difficult to maintain adequate gain. Finally, the definition of gain itself becomes problematic when a system contains too few electrons to permit the unique definition of voltages and currents.

1 Introduction

There is at present a great deal of interest in the development of nanoelectronic technologies which would directly exploit quantum-mechanical effects to perform digital computation. Much of this interest is based upon the simple notion that "quantization" is intrinsicly compatible with digital information processing. In this note I will examine some of the problems with this notion, and attempt to identify the key requirements for a successful nanoelectronic technology. This will be done by exploring at a fundamental physical level the question "Why does our present technology work?" The answer leads to the unsurprising assertion that gain is the key requirement of any electronic information-processing technology. However, when we scale devices below the limits of validity of the "circuit paradigm," we no longer know how to define gain.

2 Problems of Simple Quantum Systems

Let us first examine the idea of using simple quantum systems to perform digital processing. Such ideas have been studied in the context of reversible computation[1, 2, 3]. The fundamental idea of this work is that a complete set of Boolean operations can be implemented by unitary transformations, which in turn can be obtained from the time evolution of a Hamiltonian system. The problem with such systems is their sensitivity to errors in construction or initial state. To illustrate this, I will simulate the behavior of a long chain of inverters whose characteristics include random variations. The model reversible system will consist of a spin-$\frac{1}{2}$ particle which is acted upon by an "inverter" which reverses the spin direction, mathematically represented by the σ_x Pauli spin matrix. Now in a realistic system this reversal will not be exact, but some errors will be introduced by imperfections in the fabrication process. To model this, we take the imperfect inverter action to be $\sigma_x \cos\theta + \sigma_z \sin\theta$, which is still exactly unitary, as required by reversibility. The variable θ will be taken to be normally distributed with a mean of 0 and standard deviation 0.1 radian. We start the particle precisely in the "1" state and propagate it through a long chain of inverters. The result of a Monte Carlo calculation is shown in Fig. 1. The probability of finding the particle in the desired state executes a random walk, quickly losing all memory of the initial state. This illustrates the unsuitability of reversible quantum mechanical processes (or any process described by unitary transformations) for practical computation.

3 Logic Level Restoration

Of course such behavior is not tolerated in practical digital technologies. Such technologies achieve reliable processing of information by restoring the logic level

0-8186-3420-0/93 $03.00 © 1993 IEEE

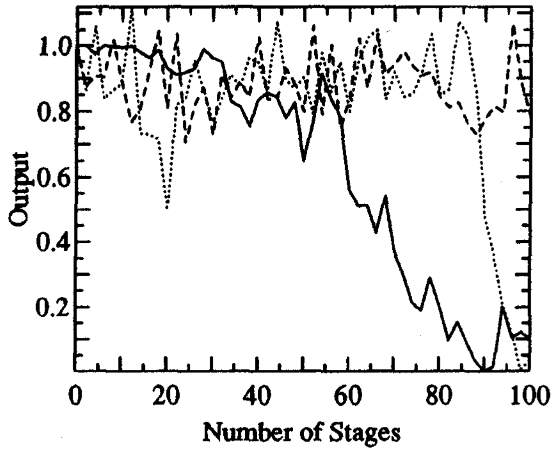

Figure 1: Simulation of a chain of inverters with random variation in the inverter characteristics. The outputs of only the even-numbered stages are plotted. The solid line shows the behavior of a model reversible inverter. It has no stability against random variations and all information about the initial state is quickly lost. The dashed line shows the behavior of an irreversible inverter with $A_v = 5$; reliable operation is achieved. The dotted line shows the behavior of an irreversible inverter with $A_v = 3$. Observe the occurrence of a bit error after about 90 stages.

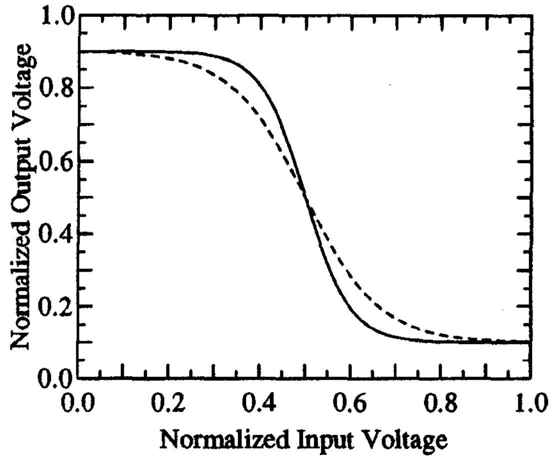

Figure 2: Idealized dissipative inverter characteristics showing the relationship between noise margins and voltage gain. The solid line shows an inverter with $A_v = 5$, leading to noise margins of about 30% of the voltage swing. The dashed line shows an inverter with $A_v = 3$, whose noise margins are much smaller.

of the signal at each gate, by throwing away the small deviations due to device variations. It is well-known that this process of level restoration implies, via the connection between information theory and thermodynamics, that the circuits producing the level restoration must be irreversible, and thus energy-dissipating, systems [4, 5]. One can think of this as a merging of nearby trajectories, which of course requires a nonconservative force.

What is required to achieve level restoration? An obvious answer is a nonlinear, saturating transfer characteristic. This, however, is not sufficient, because one wants the output of any gate to saturate at a level sufficient to drive the input of a similar gate. (One of the requirements of universal computation is that it must be possible to carry out an unlimited number of operations. This is not possible if the signal is attenuated by each gate.) The combination of these two requirements imply that an inverter must have significant voltage gain A_v (presuming that voltage is the quantity used to define the logic state). The situation is illustrated in Fig. 2, which shows the transfer characteristics of ideal inverters. If one is to have the relatively flat regions of the transfer curves which provide the noise margins, these regions must be connected by curve whose maximum slope is considerably greater than unity. This slope is the voltage gain. The magnitude of the voltage gain determines the available noise margins, as illustrated by the comparison between inverters with $A_v = 5$ and $A_v = 3$.

To illustrate the relationship between the gain and the reliability of logic operations, Monte Carlo simulations were performed of a chain of idealized inverters with varying characteristics. The transfer function of each inverter was taken to be

$$V_{\rm out} = \mathbf{V}_0 + \frac{(\mathbf{V}_1 - \mathbf{V}_0)}{1 + \exp[4A_v(V_{\rm in} - \mathbf{V}_t)]}, \qquad (1)$$

where the boldface quantities were normally distributed random variables with standard deviation equal to 0.1. The dashed curve in Fig. 1 shows the output from the even-numbered stages of a chain of inverters with $A_v = 5$. Though the level varies, it is consistently within the limits that define a "1." If the voltage gain is reduced to $A_v = 3$, however, the noise margins are not adequate and bit error eventually occurs, as the signal is interpreted as a "0" and this interpretation is reinforced by succeeding stages. (The gain of CMOS inverters used in present integrated circuits typically exceeds 50.)

I assert that gain is the fundamental requirement of any successful computational technology, electronic or otherwise. What is the physics of gain? If the signal power coming out of the system is greater than the signal power going into the system, there must obviously be another source of energy available. This invariably takes the form of at least two reservoirs with different thermodynamic properties (such as the chemical potential), and a system operating between such reservoirs cannot be a Hamiltonian system [6]. Thus it nearly certain that systems displaying gain are necessarily dissipative systems.

If a system can be described adequately by electrical network theory, the definitions of various aspects of gain are well-established. However, to be so described, the system must fit within the *circuit paradigm*, which means that it must satisfy the following criteria:

- The system must be naturally partitionable into a set of "nodes" which form the vertices of a (mathematical) graph, interconnected with "components" which form the links of the graph.
- Each node must contain enough electrons (or other charge carriers) to be viewed as a particle reservoir with a well-defined chemical potential. (This is the node voltage.)
- Electrons traversing a node must lose all phase coherence. More generally, the information contained in all moments of the electron distribution function above the first must be lost, so that the voltage and current are the only variables coupled from one component to another through a node.

Notice that it is really the properties of the nodes (or wires) of a circuit which determines the applicability of circuit paradigm, not the properties of the components.

The currents through the components are driven by chemical potential differences. However, the current flow in any two-port active device is ultimately controlled by a varying electrostatic potential. The coupling between the voltage (chemical potential) on the control electrode and this electrostatic potential depends upon the existence of a region of sufficient size and carrier density that charge neutrality is assured. The electrostatic potential in such a region is then tied to the chemical potential by a fixed work function.

4 Gain of Semi-Classical Devices

To begin to understand the factors which determine the gain of a device, let us consider the analysis of the small-signal gain of a linear two-port network, with the idea of eventually applying the ideas to quantum devices. We usually analyze quantum-mechanical systems by determining a potential from which the wavefunctions and thus the current density are calculated. Thus it is convenient to model a device using the y-parameters, which express the current as function of the applied voltages. Let port 1 represent the input electrode and port 2 the output. Then the small-signal model is

$$\begin{bmatrix} i_1 \\ i_2 \end{bmatrix} = \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}, \tag{2}$$

where the elements y represent the various admittances. Consider such a device with its input connected to an ideal voltage source and its output loaded with resistance R_L. Then the voltage gain of the resulting amplifier is

$$A_v = \frac{v_2}{v_1} = -\frac{y_{21} R_L}{1 + y_{22} R_L}. \tag{3}$$

The first thing to notice about this expression is that there is no gain if there is no load resistance, another reminder of the necessity of dissipation. Secondly, to get adequate gain, one needs an adequate value of y_{21}, which is just the transconductance (often written as g_m) of the device. Notice, however, that the effect of y_{22} is to reduce the voltage gain. y_{22} is the output conductance; it measures the sensitivity of the device to voltage swings on its output terminal.

In all three-terminal devices one can identify a critical electrostatic potential ϕ_c which determines the current through the device at a given operating condition. (The location at which ϕ_c may be defined is not necessarily unique, but for small signals, the potentials at all such points are linearly dependent.) The output y-parameters are determined by the sensitivity of ϕ_c to the terminal voltages:

$$\begin{aligned} y_{21} &= \alpha \frac{\partial \phi_c}{\partial v_1}, \\ y_{22} &= \alpha \frac{\partial \phi_c}{\partial v_2}, \end{aligned}$$

where α is some constant. The successful classical devices achieve an adequate y_{21}/y_{22} ratio by designs which provide a mechanism for screening the electric fields originating in the output electrode. This usually involves arranging the electrodes so that the control electrode is much closer to site at which the current flow is determined than is the output electrode. But still, screening requires a significant population of electrons (or other charge carriers).

A large number of three-terminal devices have been demonstrated or proposed which employ quantum-mechanical effects such as tunneling or split-path interference to control the current transport [7, 8]. Of those that have been demonstrated, many suffer from excessive y_{22} (as evidenced by steeply sloping $I(V)$ characteristics). One can also expect large y_{22} from many of the devices which have not yet been demonstrated, based upon inadequate electrostatic isolation of the output terminal.

5 Beyond the Circuit Paradigm

Quantum-mechanical models of electronic circuits are well established [9]. But as we scale devices below the limits at which they may be described within the circuit paradigm, we encounter both conceptual and practical problems. The conceptual problems center upon what will replace the circuit-level notions of "voltage" and "current" when there are too few electrons in and between the devices to constitute a reservoir. How does one define gain with such small devices? Let me point out that any notion of gain will involve the transfer of power out of the system, and therefore gain is meaningful only for open systems. The theory of open quantum systems is not at all well established.

The practical problems (aside from the obvious ones concerning fabrication) center upon the problems of few-electron devices. At some point one will lack an adequate number of mobile electrons to provide the screening of applied electric fields which both permits current control and provides isolation in classical devices. There is also the more obvious problem of small-number statistics in the definition of logic states.

In summary, it appears that gain is a necessary attribute of any successful digital technology. One can find any number of quantum and classical phenomena which can be made to implement Boolean operations. But if each gate requires a sense amplifier to detect and condition its output, the technology will not be competitive, for one can throw away the gates and use the amplifiers to do the logic. How small can an amplifier be made? There is at present no answer to this question because there is no adequate definition of gain in quantum systems.

References

[1] Feynman, R. P., "Quantum Mechanical Computers," Optics News, vol. 11, no. 2, pp. 11–20 [Feb. 1985].

[2] Benioff, P. A., "Quantum Mechanical Hamiltonian Models of Computers," in "New Ideas and Techniques on Quantum Measurement Theory," Ann. New York Acad. Sci., vol. 480, pp. 475–86 [1986].

[3] Margolus, N., "Parallel Quantum Computation," in "Complexity, Entropy, and the Physics of Information, SFI Studies in the Sciences of Complexity," W. H. Zurek, ed., vol. 8, pp. 273–87 [1990].

[4] Landauer, R., "Irreversibility and Heat Generation in the Computing Process," IBM J. Res. Develop., vol. 3, pp. 183–191 [1961].

[5] Bennett, C. H., "The Thermodynamics of Computation — a Review," Int. J. Theor. Phys. vol. 21, pp. 905–40 [1982].

[6] Frensley, W. R., "Boundary Conditions for Open Quantum Systems Driven Far from Equilibrium," Rev. Mod. Phys. vol. 62, pp. 745–91 [1990].

[7] M. A. Reed, W. R. Frensley, R. J. Matyi, J. N. Randall, A. C. Seabaugh, and H. L. Tsai, "Realization of a Three-Terminal Resonant Tunneling Device: The Bipolar Quantum Resonant Tunneling Transistor (BiQuaRTT)," Appl. Phys. Lett. vol. 54, 1034 (1989).

[8] See the several papers in chapter 4 of M. A. Reed and W. P. Kirk, eds., Nanostructure Physics and Fabrication, Proc. Int. Symp., College Station, TX, March 1989. (Academic Press, 1989).

[9] Yurke, B., and J. S. Denker, "Quantum Network Theory," Phys. Rev. A, vol. 29, pp. 1419–1437 [1984].