Possible performance of capacitively coupled single-electron transistors in digital circuits

Alexander N. Korotkov,[a] Ruby H. Chen, and Konstantin K. Likharev
Department of Physics, State University of New York, Stony Brook, New York 11974-3800

(Received 23 February 1995; accepted for publication 19 April 1995)

We have carried out a theoretical analysis of the possible performance of single-electron transistors with capacitive coupling in simple logic and memory circuits. Both resistively loaded and complementary transistors have been analyzed, with a detailed account of parasitic factors including thermal fluctuations and background charge variations. The analysis shows that at optimal values of the parameters including the background charge, the maximum operation temperature is close to $0.025e^2/Ck_B$, where C is the capacitance of the smallest tunnel junction. At $T\sim 0.01e^2/Ck_B$ the parameter margins are relatively wide; for the structures with 2-nm minimum feature size, the latter temperature is close to 77 K. A typical margin for background charge fluctuations is on the order of $0.1e$; these fluctuations may be a major obstacle for practical ultradense single-electron circuits. © *1995 American Institute of Physics.*

I. INTRODUCTION

During recent years, considerable attention has been given to the effects of correlated single-electron tunneling in ultrasmall tunnel junctions and systems (for reviews, see, e.g., Refs. 1–3). These effects are based on the fact that tunneling of a single electron to/from a small but still macroscopic conductor may change its electrostatic potential considerably, in some experiments[4,5] by ~0.1 V. As a result of this change, the probability rate of tunneling of the succeeding electron to/from the conductor may be substantially changed, so that some correlation may be established between the single-electron tunneling events, the type and degree of the correlation depending upon the particular system.[1–3]

As far as applications are concerned, correlated tunneling makes it possible to control the motion of single electrons in solid state structures. Several "single-electron" devices based on these effects have been suggested and analyzed, and some of them have been tested experimentally (for a review, see Ref. 6). The prospects for wide application of such analog devices as the fundamental standards of dc current and ultrasensitive electrometers seem quite encouraging; in fact, the electrometers have already been used successfully in some unique physical experiments.

The situation with regard to the development of digital single-electron devices is quite different, despite the fact that much excitement has been generated by the prospect of single-electronic digital circuits of unparalleled integration scale.[1,2,6,7] Up to now, only a few experiments with the simplest devices have been reported (see, e.g., Refs. 8 and 9). Moreover, although several families of single-electron logic/memory circuits have been proposed,[6,7,10–14] we are not aware of any previous attempt at their detailed quantitative analysis. The objective of this paper is to present results of an analysis for one particular class of possible digital single-electron devices—voltage-stage logic using capacitively coupled single-electron transistors.

[a]Electronic mail: akorotkov@ccmail.sunysb.edu

The operation of devices of this class is based on the high charge sensitivity of current I flowing through a system of two small tunnel junctions separating a small conductor from two bulk conducting electrodes [Fig. 1(a)]. The current I as a function of the background charge Q_0 of the small conductor is a function oscillating with the period equal to the fundamental charge e. If tunnel conductances $G_{1,2}$ of the tunnel junctions and temperature T are low enough,

$$G_{1,2} \ll 1/R_Q, \tag{1}$$

$$k_B T \ll e^2/C_\Sigma, \tag{2}$$

where $R_Q=\pi\hbar/2e^2\simeq 6.5$ kΩ is the quantum unit of resistance, and $C_\Sigma=C_1+C_2$ is the total capacitance of the central conductor, then dc current is virtually blocked below a threshold piecewise linear line in the $[Q_0,V]$ plane, which oscillates with the period e [Fig. 1(b)]. This effect (now called "Coulomb blockade of tunneling") was first observed and interpreted by Giaver and Zeller.[15] Kulik and Shekhter developed[16] a quantitative semiclassical theory of this effect [strictly valid in the limit (1)]. In their theory, the background charge Q_0 was a constant fixed by workfunctions of the central conductor (W_0) and external electrodes ($W_{1,2}$):

$$Q_0=\frac{1}{e}[(W_1-W_0)C_1-(W_2-W_0)C_2]. \tag{3}$$

Averin and Likharev[17] (see also Ref. 7) recognized that Q_0 may be changed by a straightforward injection of external charge into the central conductor through any circuit component which allows continuous (on the scale of e) charge transfer. The simplest example of such a component is a capacitor C_0 [Fig. 1(c)]; in this case

$$Q_0 \to Q_0 + UC_0 \tag{4}$$

(the capacitor also causes the renormalization[7,18] $C_2\to C_0+C_2$, $C_\Sigma\to C_1+C_2+C_0$). This effect was first observed experimentally by Fulton and Dolan,[19] and then repeatedly studied for various implementations of the device (called the "single-electron transistor"[17]).

 0021-8979/95/78(4)/2520/11/$6.00

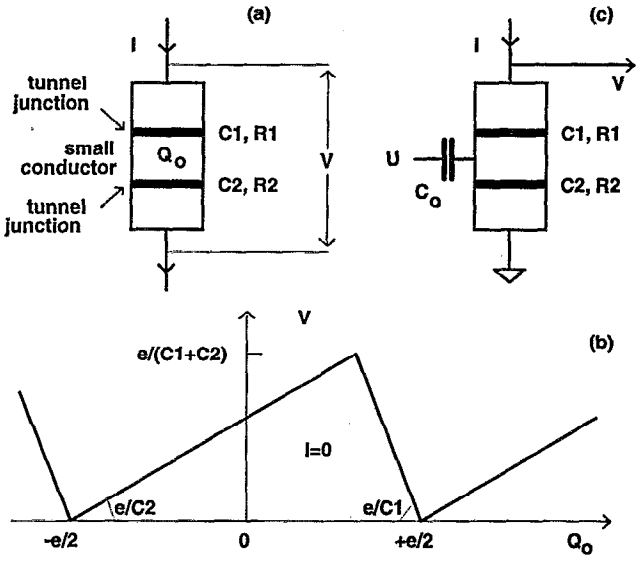

FIG. 1. Single-electron transistor: (a) the backbone double-junction system, (b) its threshold curve, and (c) capacitive coupling to the signal source.

Many publications have been devoted to the theoretical analysis of details of the dynamics of single-electron transistors (see, e.g., Refs. 20–26). There have been, however, surprisingly few papers discussing their possible performance in digital circuits. In the first work on this subject,[7] the maximum voltage gain

$$G=|\partial V_{\text{out}}/\partial V_{\text{in}}| \tag{5}$$

of the simplest inverter with resistive load [Fig. 2(a)], based on the capacitively coupled transistor (C-SET), was calculated (the result of this calculation was later confirmed by experiments[8,27]), and control characteristics of the inverter were plotted for several parameter values without an attempt to optimize the device or to calculate its maximum operation temperature. Using the e-periodicity of Coulomb blockade, single-electron transistors may be used to design complementary circuits without the need for new or additional fabrication steps (which are necessary in the CMOS technology). Several control curves for such a complementary

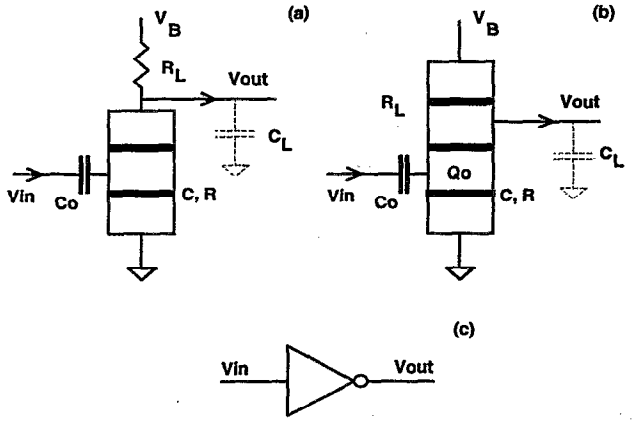

FIG. 2. Inverter/buffer stages using C-SET transistors: (a) with resistive load and (b) with tunnel-junction load. (c) Notation of the inverter.

inverter with resistive coupling to the signal source (R-SET) have been calculated in Ref. 7, again with no attempt at a quantitative optimization of the device. Later Tucker considered[12] a similar complementary inverter, but with capacitive coupling. His analysis was, however, restricted to the case of zero temperature and a certain set of parameters which did not correspond to the maximum operation temperature or parameter margins. Recently Lutwyche and Wada[28] undertook a semiquantitative analysis of digital circuits using single-electron transistors, stressing the necessity of taking their geometry into account. They showed that the finite height of the tunnel barrier may impose limits on the operation voltage and maximal temperature.

All papers mentioned above lack the exact calculation of the maximal temperature; also, parameter margins were not calculated or estimated. Besides that, the problem of the background charge fluctuations has not been discussed quantitatively; in what follows we will see that this problem may in fact be a major obstacle for digital single electronics. This is why in the present paper we have focused on the calculation of the maximum possible operation temperature $T_{\max}$ for simple digital circuits using C-SET transistors and on the analysis of the background charge margins of these circuits.

The paper is organized as follows. In Sec. II we discuss the circuits to be analyzed, write down the basic equations of the "orthodox" theory of single-electronic systems, and discuss methods for their solution. Sections III and IV are devoted to results of analysis of devices based on the resistively loaded and complementary transistors, respectively. The results of our analysis are discussed and summarized in Sec. V.

II. MODEL, EQUATIONS, AND METHODS

Figure 2 shows the simplest voltage amplification/inversion stage based on the C-SET transistor. As we will see below, the load resistance R_L should be considerably higher than the tunnel junction resistances R, so that when input voltage V_{in} varies, the current through the transistor is nearly constant, and typically rather low ($I \ll e/RC$). This means that the bias points on the $[Q_0, V]$ plane [Fig. 1(b)] are typically not too far above the threshold line. Hence, if the effective background charge puts the transistor onto one of the negative-slope segments of the line,

$$n+\frac{1}{2}-\frac{C_1}{C_\Sigma}<\frac{Q_0+V_{\text{in}}C_0}{e}<n+\frac{1}{2}, \tag{6}$$

an increase of V_{in} will result in a decrease of V_{out} (see, e.g., Fig. 4). Hence, the circuit may play the role of inverter/amplifier, and its maximum voltage gain[7] (achieved at $I\to 0, T\to 0$) $G_{\max}=C_0/C_1$, may be larger than unity if $C_0>C_1$ (while the dc current gain is always infinite).

In order to increase the operation temperature T, capacitances $C_{1,2}$ of the tunnel junctions of the inverter should be as small as possible. For this reason in the following analysis we will accept the junctions to be similar:

$$C_1=C_2=C, \qquad R_1=R_2=R, \tag{7}$$

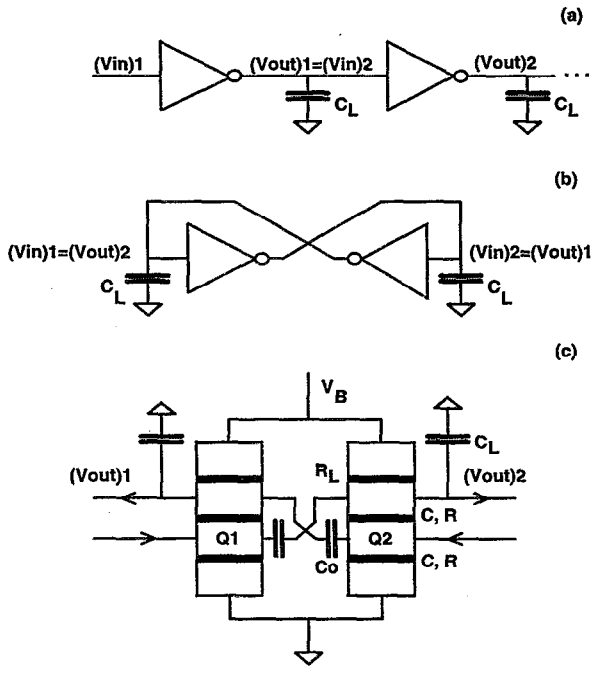

FIG. 3. Two simple digital circuits using inverters: (a) inverter string and (b) flip-flop. (c) Implementation of the flip-flop using the tunnel-junction-loaded C-SET transistors.

assuming that C is the smallest capacitance available in a given technology (we neglect the stray capacitance of the island). In all typical cases, capacitance C is much smaller than that (C_L) of the load (say, of the logic gate interconnects). This relation ($C_L \gg C$) allows us to neglect shot and thermal noises of the system and to consider the output voltage V_{out} as constant in time (for constant V_{in}), simplifying the analysis considerably. Note also that because the discreteness of the charge accumulated on C_L is not important in this limit, the replacement of the load resistor R_L by a tunnel junction with the same resistance [Fig. 2(b)] will not result in any change of the system dynamics. Such a replacement[29] may be preferable from the point of view of practical implementation.

Inverters with a voltage gain above unity are sufficient for forming two simple digital circuits: a long string of similar inverters, passing along the binary information [Fig. 3(a)] and a bistable flip-flop, i.e., a static memory cell [Fig. 3(b)]. It is well known that the dynamics of the two devices are very similar, and the minimum requirements for their operation coincide. Namely, the differential voltage gain (5) of the inverter at the point $V_{\text{in}}=V_{\text{out}}$ should exceed unity; equivalently, the control curve $V_{\text{out}}(V_{\text{in}})$ of the inverter should cross the transposed curve at at least three points, as shown in Fig. 4. In this case, the middle point $V_{\text{in}}=V_{\text{out}}$ is unstable (in the flip-flop the instability develops in time, while in the inverter string it develops in space), while two other crossing points are stable and correspond to two stationary voltage levels, which may be used to represent binary 0 and 1. In what follows, the existence of these two stable states, $V_{\text{high}}>V_{\text{low}}$, has been considered as a criterion for the operation of the circuit.

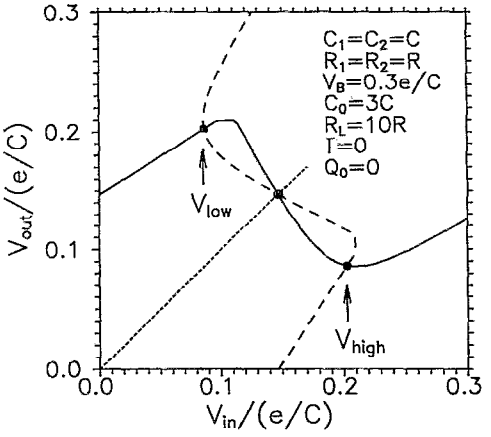

FIG. 4. A typical control curve of the resistively coupled inverter (solid line) and its transpose (dashed line). Filled (open) circles correspond to stable (unstable) states of the flip-flop or the inverter string.

In order to analyze the dependence $V_{\text{out}}(V_{\text{in}})$ quantitatively, we used evident equations of the circuits shown in Figs. 3(a) and 3(b), together with equations of the "orthodox" theory of correlated single-electron tunneling.[1–3] In this theory, electrons can tunnel through only one junction at a time, i.e., it does not take into account the cotunneling processes.[22] Cotunneling may play an important role in logic families in which the digital information is coded by single electrons[10,11,13] and hence a single undesirable cotunneling event may lead to an error. On the contrary, cotunneling is much less important for the logic which uses single-electron transistors, and the information is represented by the dc voltage levels, i.e., by many ($n \sim C_L/C$) electrons charging the interconnect capacitances $C_L \gg C$ (Figs. 2 and 3). As we will see later, in the typical case the current even through nominally "closed" transistor is on the order of $10^{-2}e/RC$, and the corresponding increase of the transistor resistance is only about 10 times. Comparison of this current with the estimates of the current due to cotunneling [see, e.g., Eq. (32) of Ref. 22] shows that the "orthodox" theory is sufficiently accurate for resistances $R \gtrsim 30R_Q \simeq 200$ kΩ. The only case when cotunneling cannot be neglected is estimates of power dissipation in the the complementary inverter in the "power saving" mode; this case is considered in Sec. IV.

Within the framework of the "orthodox" theory the probability rate Γ_{ij} of each tunneling event may be calculated independently as follows:

$$\Gamma_{ij}=\frac{1}{e}\,I(\Delta W_{ij}/e)\,\frac{1}{1-e^{-\Delta W_{ij}/T}}\,. \tag{8}$$

Here $\Delta W_{ij}=W_i-W_j$ is the change of electrostatic energy W of the system due to the tunneling event; this energy can be found without considering tunneling, from elementary electrostatics (see, e.g., Ref. 7). $I(V)$ is the current which would flow through the given junction if it were biased by a fixed voltage V; in what follows we accept the linear dependence $I(V)=V/R$, thus neglecting possible suppression of the junction barrier[28,30] by the applied voltage. T is the temperature in energy units.

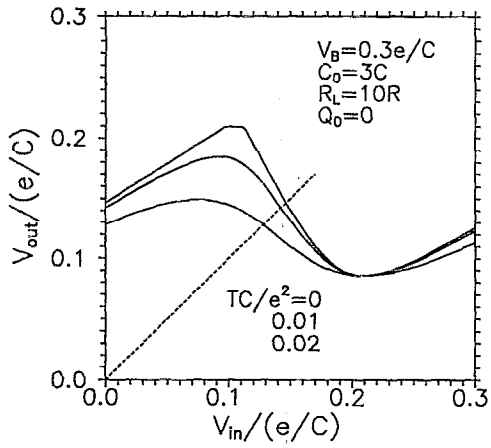

FIG. 5. The control curves of the inverter for different temperatures. The lowest curve corresponds to the case when digital applications are impossible because the voltage gain G is less than unity.

There are two alternative ways to analyze the behavior of the circuits using Eq. (8). The first is the Monte Carlo approach, using Eq. (8) to simulate real motion of electrons in the system.[31] This approach is very convenient for crude fast analyses of novel single-electron circuits[29,31–33] and we used one such program[34] as a reference. The simulation method is, however, too slow for the optimization of single-electron circuits in the multi-dimensional space of parameters. That is why our final results were obtained using a different ("Fokker-Planck-type") approach, in which Eq. (8) is used in the "master equation"[1–3]

$$\dot{P}_i = \sum_j \Gamma_{ji} P_j - \Gamma_{ij} P_i \tag{9}$$

for the probability of a particular (ith) charge configuration. A general drawback of this approach is the potentially large number of charge states of complex single-electronic systems (see, e.g., Ref. 35) and hence the large size of the matrix Γ_{ij}. In our case, however, there are two time scales for the system dynamics: one determined by $\tau = RC$ constant of the transistor itself, and another $\tau_L = RC_L$ determined by the much larger load capacitances. Hence processes in the transistor may first be averaged over a time interval $\tau \ll \Delta t \ll \tau_L$; at this stage all input and output voltages may be considered as constants. This means that a charge state is defined by only one integer i (which determines the island charge $Q_i = Q_0 + C_0 V_{in} + ie$), and stationary probabilities P_i satisfying Eq. (9) [and hence average current $I = I(V,U)$] may be calculated very fast (typically, ~1 ms per point on an Intel-486-based PC). We used this algorithm as a fast subroutine for solving the set of ordinary differential equations of the circuit as a whole, which allowed the study of the stationary solutions and the dynamics of the system on the time scale τ_L.

III. RESISTIVELY LOADED C-SET TRANSISTORS

Calculations show that thermal fluctuations lead to fast smearing of the control curve of the inverter based on the resistively loaded transistor (Fig. 2). For example, Fig. 5 shows this smearing for "typically good" parameters, including relatively high coupling capacitance $C_0 = 3C$ and load resistance $R_L = 10R$. The figure shows that the voltage gain G [Eq. (5)] at $V_{in} = V_{out}$ becomes less than 1 at temperature as low as $\sim 0.015e^2/C$. (In contrast to this rapid loss of its digital capabilities, the reasonable analog performance by the transistor persists up to much higher temperatures. For example, output voltage modulation depth decreases below 5% at $T \simeq 0.15e^2/C$ for $C_0 \ll C$.)

Let us see how the circuit parameters affect the temperature range and how the parameter window shrinks with increasing temperature. Figure 6 shows the operation range (corresponding to bistability of the flip-flop or inverter string) at different temperatures in the plane of dc bias voltage V_B and coupling capacitance C_0, for several values of load resistance R_L (the background charges are assumed to be zero). It can be seen that an increase of R_L widens the parameter window for a given temperature, and also increases T_{max}. An important conclusion is that the device performance degrades rapidly when ratio R_L/R becomes less than ~10, and even above this level the effect of ratio R_L/R on the performance is still quite considerable. The best choice of ratio C_0/C (the approximate center of the parameter window) is roughly between 2 and 4 and depends both on ratio R_L/R and temperature. An increase in temperature or R_L/R decreases the optimal C_0/C.

Another conclusion which may be drawn from examination of Fig. 6 is that if T is not too close to T_{max} (but is, say, a factor of 2 lower), the margins for C_0 and V_B are relatively wide (of the order of ±50%). A similar conclusion could be made for the background charge Q_0, if it were assumed that Q_0 is similar for all the transistors of the circuit [Fig. 7(a)]. To study the influence of independent variations of background charge we consider a flip-flop with background charges Q_1 and Q_2 on its transistors. Figure 7(b) shows that the asymmetric deviation of Q_1 and Q_2 is much more dangerous than the symmetric mode. The reason is simple: additional input signals [Fig. 3(c)] of opposite signs destroy the symmetry of the system, both pulling the unstable stationary state in the same direction (Fig. 4); when this state reaches a stable state (V_{high} or V_{low}) bistability becomes impossible.

Figure 8 shows the parameter window in the plane of the two background charges of the flip-flop transistors, for a fixed dc bias voltage. This window shrinks with an increase of the temperature as well as with a decrease of the ratio R_L/R (in the limit $T \to 0$, $R_L/R \to \infty$, $C_0/C \to \infty$ the window includes all possible background charge values). An important observation is that the point corresponding to zero background charge is typically close to the optimal one (approximate center of the parameter window). However, the optimal background charge is not exactly zero and depends on other circuit parameters and temperature.

The absolute maximum temperature at which the flip-flop and the inverter string still have two stable states corresponds to the following parameter set:[36]

$$T_{max} \simeq 0.026 \frac{e^2}{C} \quad \text{at } C_0/C \simeq 2,\ Q_0 \simeq 0.1e,$$
$$V_B/R_L \simeq 0.02e/RC,\ R_L \gg R. \tag{10}$$

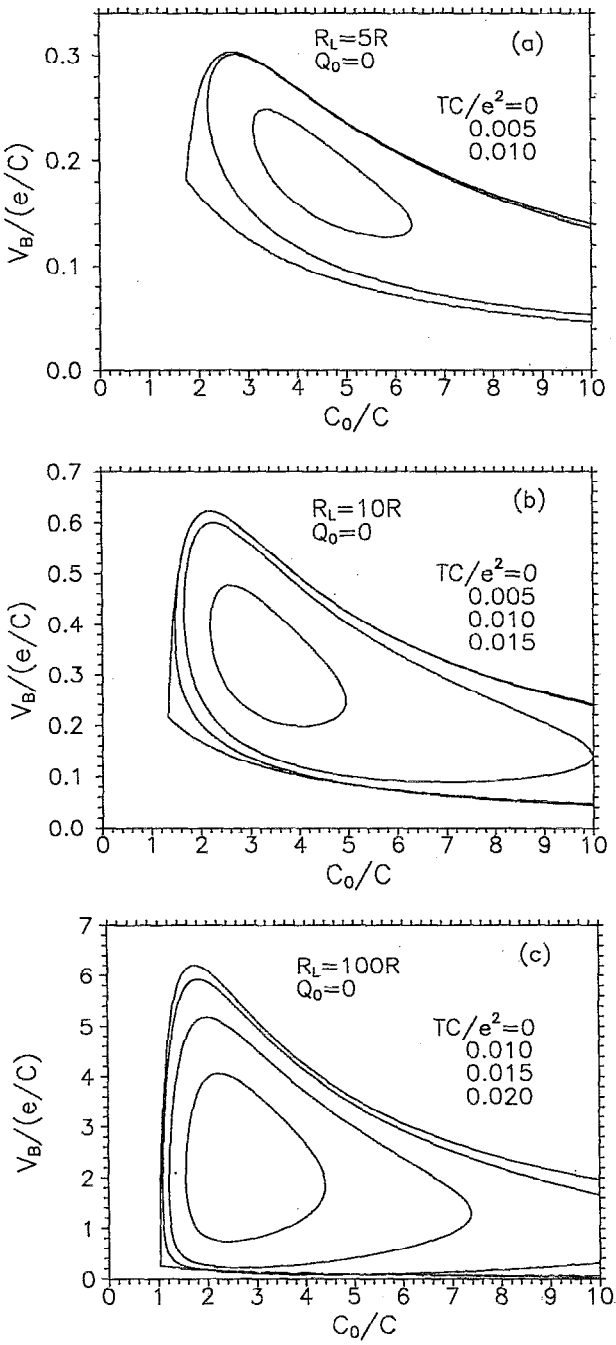

FIG. 6. Parameter window in the $[C_0, V_B]$ plane, corresponding to the existence of two stable states of the flip-flop or inverter string. Increase of temperature and decrease of load resistance decreases the window. Background charges are taken to be zero.

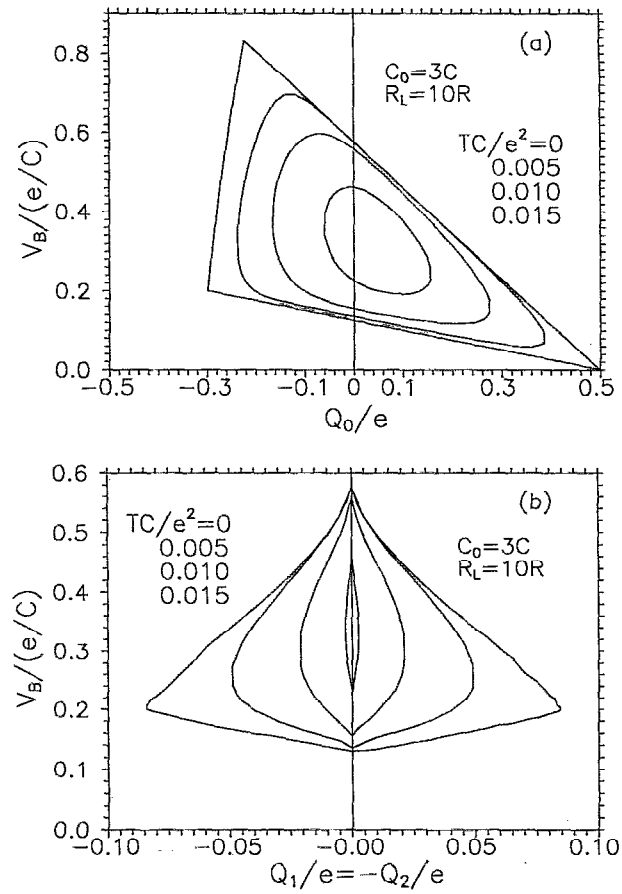

FIG. 7. (a) Parameter windows (for the flip-flop or inverter string) in the $[Q_0, V_B]$ plane for similar background charges on all transistors. (b) Parameter windows for the flip-flop with asymmetric background charges, $Q_1 = -Q_2$.

Figure 9 shows the parameter window which opens at lower temperatures for the above value of the background charge.

Figure 10 shows the margin for the magnitude of the charge deviation vector $\Delta\mathbf{Q}=\{\Delta Q_1, \Delta Q_2\}$ for the flip-flop, provided that the ratio $\Delta Q_1/\Delta Q_2$ is arbitrary, calculated for the parameter set (10) and two other sets. One may see that at $T \simeq 0.01e^2/C$, which is close to $T_{max}/2$, the margins for V_B and C_0 are very wide, while the background charge margin even for large ratio R_L/R is only about $0.1e$. For a more moderate load resistance, $R_L \simeq 10R$, the margin for ΔQ (at $T \simeq 0.01e^2/C$) reduces to about $0.03e$.

The question arises whether the condition $R_L \gg R$ leads to unacceptably slow switching of the single-electron transis-

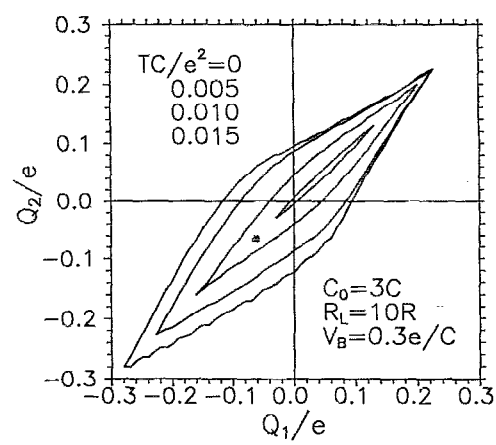

FIG. 8. Parameter windows (for the flip-flop) in the plane of background charges Q_1 and Q_2 at different temperatures for a particular parameter set.

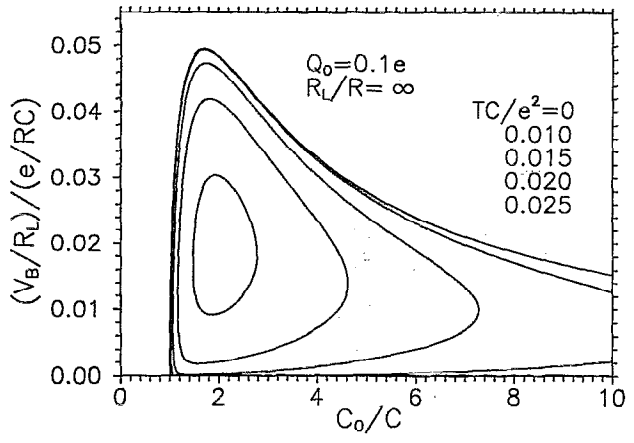

FIG. 9. Parameter windows (for the flip-flop or inverter string) in the plane $[C_0, V_B]$ for $Q_0=0.1e$, $R_L/R=\infty$ (this choice corresponds to the maximum temperature).

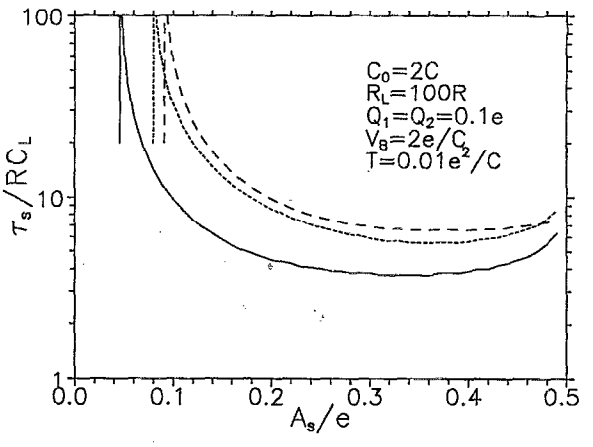

FIG. 11. The minimum duration τ_s of the input charge pulse sufficient to switch the flip-flop [with parameters satisfying Eq. (10) and for $T=0.01e^2/C$] from one stable state to the other, as a function of the pulse amplitude A_s. The solid line corresponds to simultaneous pulses of different polarity fed into both inputs [Fig. 3(c)]; while the dashed (dotted) line shows the effect of only one pulse at the input of the "open" ("closed") transistor. The vertical lines are the asymptotes corresponding to $\tau_s\to\infty$, i.e., to thresholds of switching by infinitely long signals.

tor. Figure 11 demonstrates that the answer is no. This figure shows how long a rectangular pulse must be applied to inputs of the flip-flop to switch it into the opposite state.[37] While the exact value of the necessary pulse duration τ_s depends on its charge amplitude A_s, the minimum value of τ_s is of the order of $\tau=RC_L$ rather than $\tau_L=R_LC_L$. The reason for this relatively fast switching is that in a stationary state both transistors are more or less open (see the very typical control curve in Fig. 4). This means that during switching, both load capacitances are recharged through the relatively low-resistance transistors rather than through the high-resistive loads.

This fact, although very favorable for device speed, is unfavorable in terms of power dissipation. For example, in the "typically good" stationary points shown in Fig. 4, the average dc power consumption per inverter is as high as $\sim 5\times 10^{-3}e^2/C^2R$. Another clear disadvantage of the resistively loaded logic stages is the necessity to provide either Ohmic resistors of small geometric size and high nominal resistance ($R_L \gtrsim 10R \gtrsim 10^2 R_Q \sim 1$ MΩ), or tunnel junctions

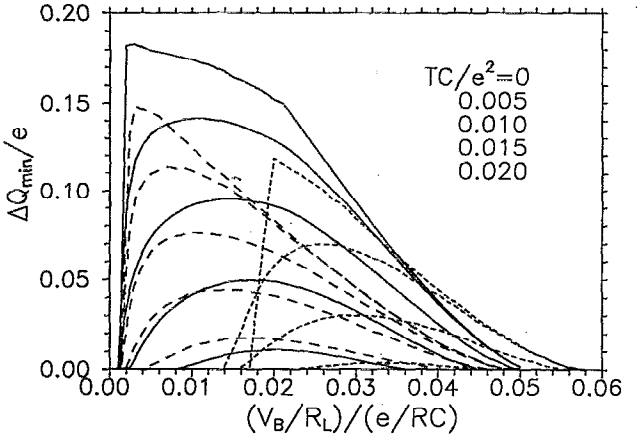

FIG. 10. The minimum fluctuation $\Delta Q=(\Delta Q_1^2+\Delta Q_2^2)^{1/2}$ of background charges which destroys the bistability of the flip-flop. Solid line: $C_0=3C$, $Q_1=Q_2=0$, $R_L=100R$, dashed line: $C_0=2C$, $Q_1=Q_2=0.1e$, $R_L=100R$, dotted line: $C_0=3C$, $Q_1=Q_2=0$, $R_L=10R$.

with resistances very different from those of the transistor junctions. Both facts push us towards using nonlinear loads, notably the complementary transistors.[7,12]

IV. COMPLEMENTARY C-SET TRANSISTORS

Figure 12 shows the equivalent circuit of the "complementary" inverter consisting of two similar C-SET transistors,[12] and the flip-flop using this inverter. The device is only slightly more complex than its resistively loaded counterpart [Fig. 3(c)], but has the important advantage that all its tunnel junctions may be similar, and there is no need for Ohmic resistors. In contrast to Ref. 12, we are not using additional relatively large capacitors between islands and bias leads. Instead, it is assumed that background charges Q_u and Q_d (Fig. 12) are controllable, for example, with the help of small capacitors leading to relatively large external potentials. (This substitution allows C_Σ to be reduced and T_{max} increased considerably.) On the other hand, as we will see below, the complementary inverter may work well even when $Q_u=Q_d=0$.

An important new feature of the complementary inverter in comparison with the resistively loaded one is that at $T\to 0$ both of the serially connected C-SET transistors may be "closed," i.e., be in the Coulomb blockade state ($I=0$). This feature appears clearly in the control curves of the inverter (Fig. 13). For example, curve 1 in this figure (corresponding to $Q_u=-Q_d=0.15e$) widens into an "uncertainty region" at $V_{in}\sim 0.15e/C$. Inside this region both transistors are closed and the current vanishes, so that the output voltage V_{out} is arbitrary within limits.

The uncertainty region widens with a decrease in the dc bias voltage V_B. Its dependence on the background charges of the transistors (Q_u and Q_d) is more complex; in particular the uncertainty regions may appear either near the borders of the control curve (i.e., at $V_{in}\simeq 0, V_B$) or near its center

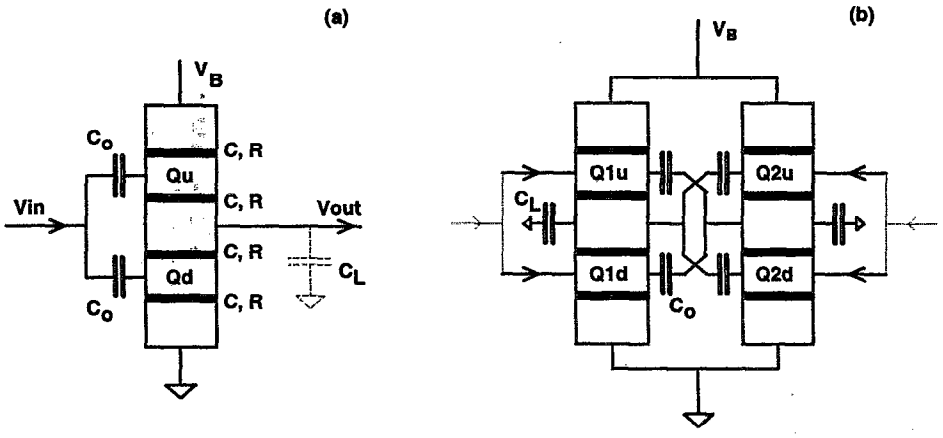

FIG. 12. (a) The complimentary inverter stage consisting of two C-SET transistors, and (b) the flip-flop using two such inverters.

($V_{in} \simeq V_{out}$); these two cases are represented by curves 1 and 4 in Fig. 13. Note that we have plotted results for asymmetric background charges only, $Q_u = -Q_d$, because the symmetric charge variation simply shifts control curves along the horizontal axis, i.e., destroys the symmetry of the complementary inverter.

For finite but low temperatures ($T \ll e^2/C_\Sigma$) the boundaries of the Coulomb blockade are no more exact. Formally, it is possible to calculate the single-valued dependence of the output voltage corresponding to the exponentially small current I through transistors. This single-valued dependence can be calculated even at $T=0$ if we take the cotunneling[22] into account. However, such a calculation does not have much sense for the analysis of the parameter margins of the device. This is because if the currents are too small, the time necessary for charging the load capacitance to reach the stationary state may be very long. Hence, if we need digital circuits operating at reasonable frequencies, we should consider the output voltage of the inverter as uncertain within some limits. That is why even for finite temperature we continue to consider uncertainty regions on the control curves, establishing their boundaries as corresponding to a small fixed current I_m through a transistor (we can choose, for example, $I_m = 10^{-3} e/RC$). The value of this current crudely determines the upper bound for the time $\tau \sim eC_L/CI_m$ of transient inside the uncertainty region. The introduction of I_m is mostly for the sake of mathematical correctness, because results do not depend on I_m if the bias voltage is not too close to its lower limit or the temperature is not too low ($T \gtrsim 0.01 e^2/C$).

In the presence of the uncertainty regions the former criterion of the operation of the flip-flop or inverter string (three intersections of the control curve and its transpose) becomes ambiguous, because the intersections may occur not at points but in regions. In this case we have generalized the operation criterion in the following way: the inverter was assumed to be operational if there were three separate regions (or points) of intersection. This means that despite the fact that the exact value of the stationary output signal may not be well defined, we can easily distinguish states corresponding to logical 1 or 0.

Figure 14 shows the operation windows of the flip-flop (or inverter string) for complementary inverters with zero background charge. At sufficiently low temperatures the low-bias margin widens with the temperature (compare the

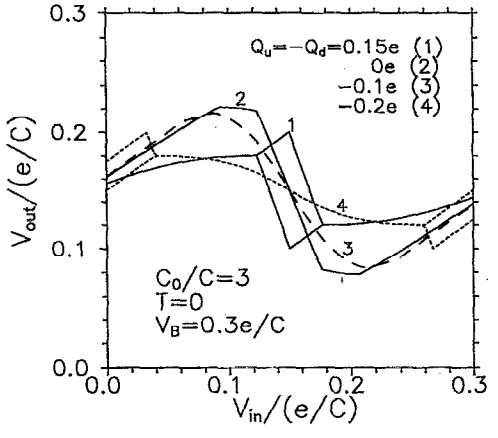

FIG. 13. Control curves of the complementary inverter for different background charges at $T=0$. The lines widen into "uncertainty regions" which correspond to the Coulomb blockade of tunneling in both transistors.

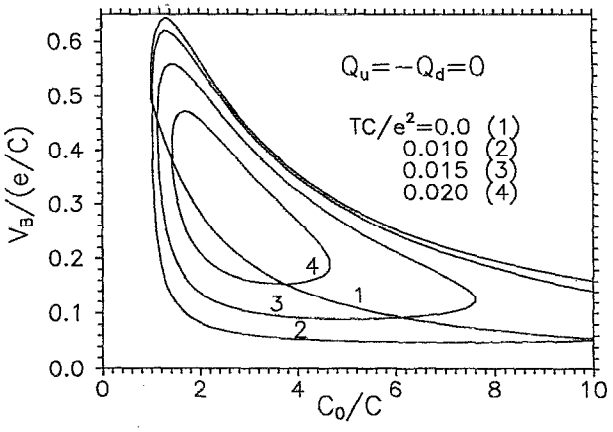

FIG. 14. Parameter windows in the $[C_0, V_B]$ plane for the complementary flip-flop or inverter string.

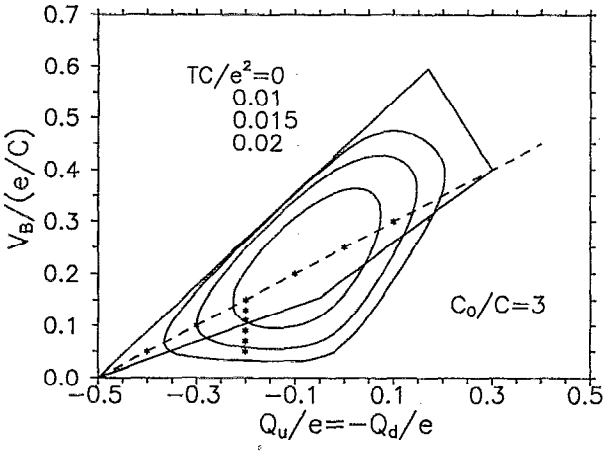

FIG. 15. Bias voltage window as a function of background charges $Q_u=-Q_d$ for the complementary flip-flop or inverter string. The dashed line corresponds to Eq. (12), and stars represent the parameter sets used in Fig. 18.

curves for $T=0$ and $T=0.01e^2/C$). This is a consequence of our new operation criterion: for a sufficiently low bias voltage at $T=0$ the uncertainty region is so large that logical 0 and 1 cannot be distinguished. The temperature increase makes the intermediate states unstable and hence decreases the lower boundary of the operation window. The position of the lower boundary in this case (which takes place for $T\lesssim 0.01e^2/C$) depends on the value of I_m. For higher temperatures ($T\gtrsim 0.01e^2/C$) the Coulomb blockade is sufficiently smeared out; hence, the new criterion coincides with the old one, the operation window does not depend on I_m and shrinks with temperature.

A comparison of Fig. 14 with Fig. 6 shows that the parameter margins of the complementary inverter are close to those of its resistively loaded counterpart with $R_L/R\to\infty$, and are considerably wider than those for $R_L/R\lesssim 10$, thus confirming the arguments in favor of the complementary inverter.

Figure 15 shows the parameter window in the $[Q_0, V_B]$ plane for the case of the asymmetric shift of background charges, $Q_u=-Q_d$, identical in all inverters of the circuit. It can be seen that the lower boundary of the window for $T=0$ is higher than that for $T=0.01e^2/C$ for a the considerable range of background charges. This is again a consequence of new operation criterion. Note that the lower boundary of the window in Fig. 15 shows noticeable dependence on the choice of I_m for temperatures below $\sim 0.01e^2/C$ (cf. Fig. 14).

The lower left part of the operation window shown in Fig. 15 (with $Q_u=-Q_d$ approximately between $-0.2e$ and $-0.4e$) corresponds to the case of low power consumption (see below), however, background charges closer to zero provide considerably higher operation temperatures. The maximum temperature corresponds to the following parameter set:[38]

$$T_{\max}\simeq 0.026e^2/C \quad \text{at } C_0/C\simeq 2,$$
$$Q_u=-Q_d\simeq -0.1e, \quad V_B\simeq 0.27e/C, \tag{11}$$

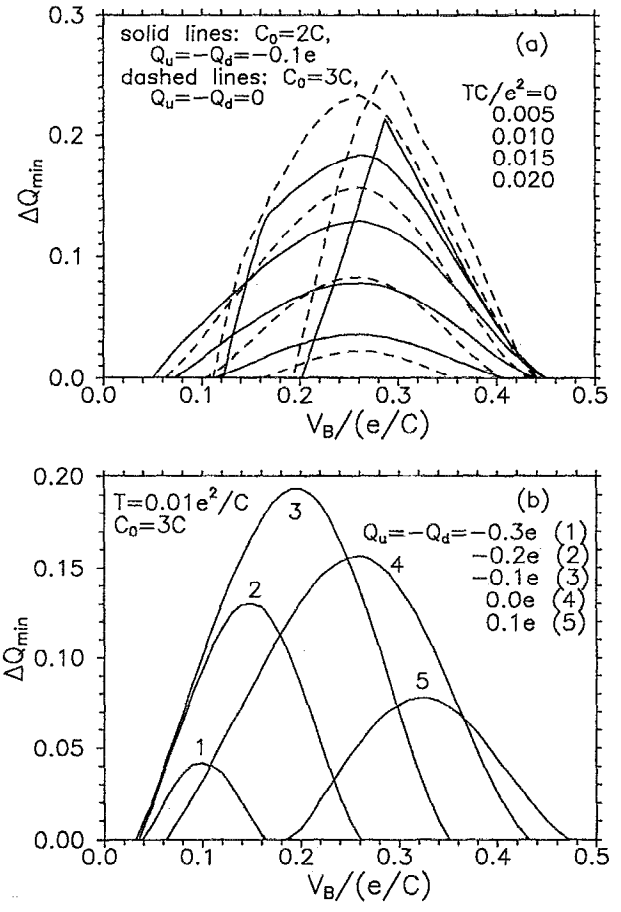

FIG. 16. Critical margin for the charge deviation vector magnitude $\Delta Q=(\Delta Q_{1u}^2+\Delta Q_{1d}^2+\Delta Q_{2u}^2+\Delta Q_{2d}^2)^{1/2}$ for the complementary flip-flop as a function of the bias voltage at (a) several temperatures, and (b) several values of the background charge at $T=0.01e^2/C$.

and *exactly* coincides with the maximum temperature (10) of a resistively loaded inverter with $R_L/R\to\infty$. This fact may be proven from the following symmetry arguments. In the limiting case of maximum temperature the stable states almost coincide with each other and correspond to the center of the control curve ($V_{\text{in}}\approx V_{\text{out}}\approx V_B/2$) of the symmetrical complementary inverter. In this case a small input signal opens one transistor as much as it closes the other one, so that the current does not change. This means that each transistor is biased by a fixed current, exactly as the resistively loaded transistor at $R_L/R\to\infty$.

Note that for the parameter set of the complimentary inverter accepted by Tucker[12] the maximum operation temperature is as low as $0.0056e^2/C$, where C is the minimum capacitance of tunnel junctions (which were different in Ref. 12). This large reduction in $T_{\max}$ in comparison with Eq. (11) is due to the use of large additional capacitances and a power-saving operating point (see below).

Similar to the resistively loaded case, if fluctuations of the background charge in different inverters (and in transistors of the same inverter) are independent, they should be relatively small to ensure logic operation. Figure 16(a) shows the critical margin for $\Delta Q=(\Delta Q_{1u}^2+\Delta Q_{1d}^2+\Delta Q_{2u}^2+\Delta Q_{2d}^2)^{1/2}$ describing the background charge fluctuations in

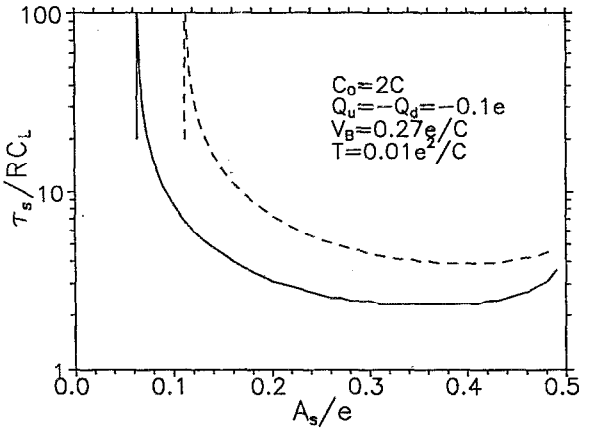

FIG. 17. The minimum duration τ_s of the switching pulse for the complementary flip-flop as a function of the charge pulse amplitude A_s. The solid line corresponds to simultaneous signals at both inputs, while the dashed line corresponds to a signal applied to only one inverter.

four transistors of the complementary flip-flop [Fig. 12(b)] for two parameter sets at different temperatures. At $T\simeq 0.01e^2/C$ the typical margin is $\Delta Q_{\min}\simeq 0.15e$ which is close to that in the resistively loaded case (Fig. 10) for $R_L/R\to\infty$ (for a just comparison of $\Delta Q_{\min}$ for two- and four-dimensional spaces of charges, in the latter case this parameter should be divided by $\sqrt{2}$). Figure 16(b) shows the critical margin $\Delta Q_{\min}$ for various initial background charges at $T=0.01e^2/C$. The largest margins are achieved when the initial charges are relatively close to zero (similar to the temperature optimization — cf. Fig. 15).

Figure 17 shows the switching speed of the complementary flip-flop. Similarly to Fig. 11 (for the resistively loaded case) we have used the parameter set which maximizes the operation temperature. A comparison between these two figures shows that the complementary inverter is slightly faster than the resistively loaded one. This is to be expected since in the complementary case, the load capacitance is charged through two transistors in parallel, in contrast to one transistor in the resistively-loaded device with $R_L\gg R$. Note also that in the complementary case the switching by a signal applied to either inverter of the flip-flop is equally fast (Fig. 17), while in the resistively loaded case the signal fed into the "closed" transistor switches the flip-flop slightly faster than the signal applied to the "open" transistor (Fig. 11).

To summarize, the performance of the complementary inverter is comparable to that of its resistively loaded counterpart for parameters which allow relatively high operation temperatures. Unfortunately, for these parameters the complimentary inverter does not provide a considerable advantage in power consumption P. For example, $P\simeq 2\times 10^{-3}e^2/C^2R$ per inverter for the parameter set (11) at zero temperature. On the other hand, in semiconductor electronics the complementary logic makes a sizeable reduction in power consumption possible. Such a reduction is possible for the complementary single-electron inverter as well, but with the price of a considerable reduction in operation temperature and parameter margins. The power-saving mode of operation may be achieved in the lower left part of the operation window shown in Fig. 15. Figure 18 shows power consumption per inverter as a function of temperature for several parameter sets. The sets used in Fig. 18(a) satisfy the equation

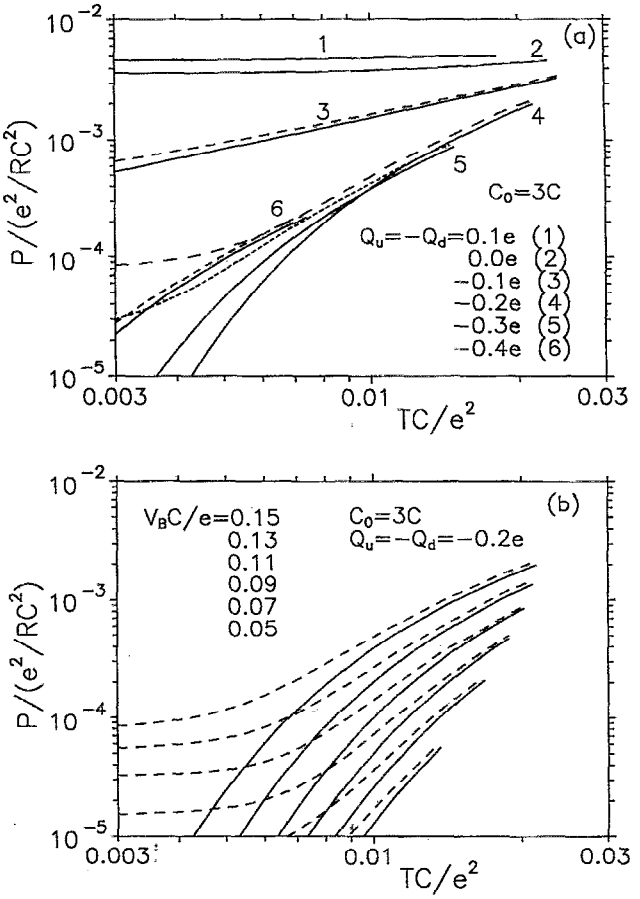

FIG. 18. Power consumption per complementary inverter as a function of the temperature for different parameter sets: (a) the bias voltage changes together with the background charge in accordance with Eq. (12); (b) the charge is fixed while the voltage changes. Each curve is cut off at the maximum operation temperature for the given parameter set. Solid lines represent the results of the "orthodox" theory, while dashed lines are calculated taking cotunneling into account for $R=30R_Q\simeq 200$ kΩ [in (a) the dashed lines are shown only for curves 3–6].

$$V_B=\frac{e}{C_0+C}(Q_u-Q_d+1),\quad Q_u=-Q_d \tag{12}$$

which corresponds to the dashed line in Fig. 15 going roughly along the middle of the operation region. [Equation (12) gives the threshold voltage for Coulomb blockade in the central part of the control curve of inverter at $T=0$.] The parameter sets of Fig. 18(b) differ in only V_B and correspond to the points along the vertical line in Fig. 15. It can be seen that when the background charges are relatively far from zero ($Q_u=-Q_d\sim -0.2e$), static power consumption at $T\lesssim 0.01e^2/C$ can be made very small, less than $10^{-4}e^2/RC^2$ per inverter. In this case both transistors are well below the Coulomb blockade threshold, and cotunneling processes[22] should be taken into account. We have used the approximation of Ref. 39 to calculate the current due to

cotunneling and simply added it to the "orthodox" thermoactivated current (the two contributions can be merely added, because well below the blockade threshold both of them are small and do not influence each other). The results of such calculation for $R=30R_Q\simeq 200k\Omega$ are represented in Fig. 18 by dashed lines while the solid lines show the results of "orthodox" theory. One can see that contribution of cotunneling becomes very important when the static current through inverter is less than $\sim 10^{-3}e/RC$. The cotunneling contribution decreases with the junction resistance increase.

V. DISCUSSION

We have analyzed the simplest digital circuits using capacitively coupled single-electron transistors (both resistively loaded and complementary) and have found their maximum operation temperature and typical parameter margins.[40] The results are strongly dependent on the randomness of the background charge Q_0. If the background charge of each transistor in the circuit may be tuned to its optimum value, the maximum operation temperature is close to $0.025e^2/Ck_B$, and may be relatively high for devices which could be implemented using even present-day nanofabrication techniques. Even if we leave aside scanning-probe methods of single-atom manipulation,[41,42] which are forbiddingly slow at the present stage of their development, direct e-beam writing has been demonstrated to provide features as small as $\sim$ 2 nm (see, e.g., Refs. 43 and 44). The capacitance of a thin conducting island with such a diameter on a silicon substrate is close to 5×10^{-19} F. We may expect a typical capacitance $\sim 3\times 10^{-19}$ F for a tunnel junction formed between the island and "wires" of a similar width placed at distance ~ 2 nm from each other (tunneling may be achieved by appropriate doping of the semiconductor substrate). For digital circuits using such junctions, the maximum temperature of operation should be close to ~ 150 K. In this case relatively wide parameter margins would exist at the liquid nitrogen temperatures ($T\simeq 77$ K).

It is easy to imagine that utilization of substrates with lower dielectric constants or further reduction of the minimum feature size to ~ 1 nm could make the operation feasible even at room temperature. However, it is important to remember that at the scale of 1 nm the discreteness of the electron spectrum in the dot should become noticeable,[45] so that the system behavior will be quantitatively different from that given by the orthodox theory used in this work. The finite height of the tunnel barrier[28,30] can also give quantitative corrections to properties of devices of this scale. Moreover, the stray capacitance C_{str} of conducting island may become comparable to the capacitance of the tunnel junctions. This factor would lead to a decrease of the maximal temperature.

The optimistic estimate of the possible operation temperature presented above is only valid if the background charge of each transistor is close to its optimum value. For circuits of low integration scale the problem may be solved in a straightforward manner by adjusting each transistor individually. At high integration scale the only hope is that the background charge would have a *naturally* narrow statistical distribution.[46] This might be expected to happen in systems reproducible on the atomic level. In real structures there is always some amount of charged impurities, lattice defects, etc., leading to fluctuations of the background charge (see, e.g., Refs. 5, 8, and 47). Nevertheless, there is hope that in small structures the impurities would be eliminated due to internal electrostatic forces (for example, the attraction of the charge to the conducting surface by the image charge). There is some experimental evidence[48] of such a "purging" in granular systems with the grain of a relatively large size (~ 10 nm). In small structures the effect should be stronger, but additional experimental studies of the "purging" are clearly necessary.

ACKNOWLEDGMENTS

Useful discussions with D. V. Averin, L. Fonseca, and P. Solomon are gratefully acknowledged. The work was supported in part by AFOSR Grant No. 91-0445 and ONR/ARPA Grant No. N00014-93-1-0880.

[1] K. K. Likharev, IBM J. Res. Dev. **32**, 144 (1987).
[2] D. V. Averin and K. K. Likharev, in *Mesoscopic Phenomena in Solids*, edited by B. Altshuler, P. Lee, and R. Webb (Elsevier, Amsterdam, 1991), p. 173.
[3] *Single Charge Tunneling*, edited by H. Grabert and M. Devoret (Plenum, New York, 1992).
[4] R. Wilkins, E. Ben-Jacob, and R. C. Jaklevic, Phys. Rev. Lett. **63**, 801 (1989).
[5] C. Schönenberger, K. van Houten, and H. C. Donkersloot, Europhys. Lett. **20**, 249 (1992).
[6] D. V. Averin and K. K. Likharev, in Ref. 3, p. 311.
[7] K. K. Likharev, IEEE Trans. Magn. **23**, 1142 (1987).
[8] G. Zimmerli, R. L. Kautz, and J. M. Martinis, Appl. Phys. Lett. **61**, 2616 (1992).
[9] P. Dresselhaus, L. Ji, S. Han, J. Lukens, and K. Likharev, Phys. Rev. Lett. **72**, 3226 (1994).
[10] K. K. Likharev and V. K. Semenov, in Extended Abstracts of Int. Supercond. Electron. Conference, Tokyo, 1987, p. 182.
[11] K. K. Likharev, S. F. Polonsky, and S. V. Vyshenskii (unpublished, 1990).
[12] J. R. Tucker, J. Appl. Phys. **72**, 43339 (1992).
[13] K. Nakazato and J. D. White, in *Proceedings of IEDM'92* (IEEE, New York, 1992), p. 487; K. Nakazato, B. Blaikie, and H. Ahmed, J. Appl. Phys. **75**, 5123 (1994).
[14] P. D. Tougaw and C. S. Lent, J. Appl. Phys. **75**, 1818 (1994).
[15] I. Giaver and H. R. Zeller, Phys. Rev. Lett. **20**, 1504 (1968).
[16] I. O. Kulik and R. I. Shekhter, Zh. Exp. Teor. Phys. **68**, 623 (1975) [Sov. Phys. JETP **41**, 308 (1975)].
[17] D. V. Averin and K. K. Likharev, J. Low Temp. Phys. **62**, 345 (1986).
[18] In fact, renormalizations containing arbitrary parameter α are possible: $Q_0\rightarrow Q_0+UC_0-\alpha C_0V$, $C_1\rightarrow C_1+\alpha C_0$, $C_2\rightarrow C_2+(1-\alpha)C_0$; Eq. (4) corresponds to the particular case $\alpha=0$.
[19] T. A. Fulton and G. J. Dolan, Phys. Rev. Lett. **59**, 109 (1987).
[20] M. Amman, K. Mullen, and E. Ben-Jacob, J. Appl. Phys. **65**, 339 (1988).
[21] L. I. Glazman and R. I. Shekhter, J. Phys. Condensed Matter **1**, 5811 (1989).
[22] D. V. Averin and A. A. Odintsov, Phys. Lett. A **140**, 251 (1989).
[23] J. C. Wan, K. A. McGreer, L. I. Glazman, A. M. Goldman, and R. I. Shekhter, Phys. Rev. B **43**, 9381 (1991).
[24] M. Amman, R. Wilkins, E. Ben-Jacob, P. D. Maker, and R. C. Jaklevic, Phys. Rev. B **43**, 1146 (1991).
[25] A. N. Korotkov, D. V. Averin, K. K. Likharev, and S. A. Vasenko, in *Single-Electron Tunneling and Mesoscopic Devices*, edited by H. Koch and H. Lubbig (Springer, Berlin, 1992), p. 45.
[26] D. V. Averin, Physica B **194-196**, 979 (1994).
[27] E. H. Visscher, S. M. Verbrugh, J. Lindeman, P. Hadley, and J. E. Mooij, Appl. Phys. Lett. **66**, 305 (1995).
[28] M. T. Lutwyche and Y. Wada, J. Appl. Phys. **75**, 3654 (1994).
[29] H. Fukui, M. Fujishima, and K. Hoh, in Extended Abstracts of the 1994

International Conference on Solid State Devices and Materials, Yokohama, 1994, p. 331.
[30] A. N. Korotkov and Yu. V. Nazarov, Physica B **173**, 217 (1991).
[31] K. K. Likharev, N. S. Bakhvalov, G. S Kazacha, and S. I. Serdyukova, IEEE Trans. Magn. **25**, 1436 (1989).
[32] M. Kirihara, N. Kuwamura, K. Taniguchi, and C. Hamaguchi, in Extended Abstracts of the 1994 International Conference on Solid State Devices and Materials, Yokohama, 1994, p. 328.
[33] N. Yoshikawa, H. Ishibashi, and M Sugahara, Ref. 32, p. 334.
[34] MOSES (Monte Carlo Single-Electronics Simulator), written by Ruby Chen. This program allows the simulation of single-electron circuits presenting arbitrary connections of tunnel junctions, capacitors, Ohmic resistances, and external wires (dc or ac biased). MOSES can calculate average currents and average charges, spectral densities, etc. The program will be described in the paper by R. Chen, K. Matsuoka, and K. K. Likharev (in preparation), and is available for free noncommercial use (send an e-mail to rchen@max.physics.sunysb.edu).
[35] L. R. C. Fonseca, A. N. Korotkov, K. K. Likharev, and A. A. Odintsov (unpublished).
[36] The constraint of zero background charge, $Q_0=0$, only slightly reduces the maximum temperature, $T_{max}\simeq 0.024\ e^2/C$ at $C_0/C\simeq 2.4$, $V_B/R_L\simeq 0.02e/RC$, $R_L\gg R$.
[37] In the standard RAM the switching of the memory cell occurs when two different signals (from "word" and "bit" lines) are applied simultaneously, while each signal alone is not sufficient. This mode of operation corresponds to the region between the solid and dotted lines in Fig. 11. The margin is about 25% for the signal duration and much wider for the signal amplitude. The margins are even wider in complimentary case—see Fig. 17.
[38] For $Q_u=Q_d=0$ the maximum temperature $T_{max}\simeq 0.024e^2/C$ is achieved at $C_0/C\simeq 2.4$, $V_B\simeq 0.3e/C$.
[39] H. D. Jensen and J. M. Martinis, Phys. Rev. B **46**, 13407 (1992).
[40] A similar analysis of more complex logic gates (AND, OR, etc.), as well as write-in and read-out circuits for static RAMs would be of considerable interest. At this stage, we have confirmed the operation of SET-based complementary NOR and NAND gates for the parameter set $C_0=3C$, $V_B=0.25e/C$, $Q_0=0$, $T=0.01e^2/C$ (which is a typical operation point of complementary inverter—see Figs. 14 and 15). These gates were structured in complete analogy to corresponding gates established in the CMOS technology.
[41] D. M. Eigler and E. K. Schweizer, Nature **344**, 2001 (1990).
[42] J. W. Lyding, T.-C. Shen, J. S. Hubacek, J. R. Tucker, and G. C. Abeln, Appl. Phys. Lett. **64**, 2010 (1994).
[43] G. S. Chen, C. B. Boothroyd, and C. J. Humphreys, Appl. Phys. Lett. **62**, 1949 (1993).
[44] H. I. Liu, D. K. Biegelsen, F. A. Ponce, N. M. Lohnson, and R. F. W. Pease, Appl. Phys. Lett. **64**, 1383 (1994).
[45] D. V. Averin and A. N. Korotkov, Zh. Éksp. Teor. Phys. **97**, 1661 (1990) [Sov. Phys. JETP **70**, 937 (1990)].
[46] It might be expected that a solution to the problem of background charge would be to use resistively coupled (Ref. 7) single-electron transistors instead of capacitively coupled ones, because continuous charge transfer through a resistor automatically compensates any background charge (Refs. 1–3). However, our preliminary analysis of these transistors show that the Nyquist noise of the coupling resistance R_0 considerably affects the performance of the inverter, reducing the maximum temperature of the operation. This temperature increases with R_0; however even for $R_0\sim 20R$ the maximum temperature is still of the order of $10^{-3}e^2/C$.
[47] J. G. A. Dubois, E. N. G. Verheijen, J. W. Gerritsen, and H. van Kempen, Physica B **194-196**, 997 (1994).
[48] J. Lambe and R. C. Jacklevic, Phys. Rev. Lett. **22**, 1371 (1969).