Charge Recycling in Power-Gated CMOS Circuits

Ehsan Pakbaznia, Student Member, Farzan Fallah*, Senior Member, Massoud Pedram, Fellow
University of Southern California, Los Angeles, CA 90089, U.S.A
*Fujitsu Laboratories of America, Sunnyvale, CA 94085, U.S.A

Abstract—Design of a suitable power gating (e.g., multi-threshold CMOS or super cutoff CMOS) structure is an important and challenging task in sub-90nm VLSI circuits where leakage currents are significant. In designs where the mode transitions are frequent, a significant amount of energy is consumed to turn on or off the power gating structure. It is thus desirable to develop a power gating solution that minimizes the energy consumed during mode transitions. This paper presents such a solution by recycling charge between the virtual power and ground rails immediately after entering the sleep mode and just before wakeup. The proposed method can save up to 43% of the dynamic energy wasted during mode transition while maintaining the wake up time of the original circuit. It also reduces the peak negative voltage value and the settling time of the ground bounce.

I. Introduction

As CMOS technology scales down, supply voltage is reduced to avoid device failure due to high electric fields in the gate oxide and the conducting channel under the gate. This supply voltage scaling reduces the dynamic component of circuit power dissipation, but unfortunately also decreases the switching speed of transistors. To compensate for this performance loss, the transistor threshold voltages are decreased, which in turn causes an exponential increase in the sub-threshold leakage current. Furthermore, to maintain the gate voltage control over the active region of the transistor, thickness of the dielectric between the gate and the channel region is reduced, which in turn results in an exponential increase in the gate leakage current. Please refer to [1] for a more detailed discussion.

Power gating, also known as Multi-threshold CMOS [2] or MTCMOS for short, is used to cut off the power to some functional blocks in a design. MTCMOS provides low leakage and high performance operation by utilizing high speed, low V_t (LVT) transistors for logic cell implementation and low leakage, high V_t (HVT) transistors for power gating switch implementation. The power gating switch itself is typically realized as a single (footer) NMOS or (header) PMOS transistor, which disconnects logic cells from ground or VDD rails to reduce the leakage when the circuit is in the sleep mode.

Some of the design challenges that must be considered when using the power gating technique are: (i) Placement and sizing of the sleep transistors; (ii) Automatic generation of sleep signal; (iii) Sleep signal scheduling for wakeup noise reduction; (iv) Mode transition energy minimization; (v) State retention; (vi) Support for multiple levels of sleep. In this paper, we focus on the problem of mode-transition energy minimization.

The remainder of this paper is organized as follows. In Section II, we review prior work in the area of power gating. Section III introduces the concept of charge recycling in power gated circuits. In Section IV we derive conditions for achieving the maximum energy saving with CR-MTCMOS. The impact of the proposed charge-recycling technique on leakage and ground bounce (GB) of the circuit is discussed in Section V. A few important variants of the charge-recycling technique in MTCMOS circuits and application of the technique to SCCMOS circuits are discussed in Section VI. Sections VII and VIII present our simulation results and conclusions, respectively.

II. Prior Work

MTCMOS is a leakage power saving solution that provides high active mode performance and low standby leakage power [3][4]. Sleep transistors slow down logic cells during the active mode operation of the circuit. This is due to the voltage drop across the functionally-redundant sleep transistors and the increase in the threshold voltage of logic cell transistors as a result of the body effect. The performance penalty of using a sleep transistor depends on its size and the amount of the current that flows through this transistor due to logic transitions in the active mode. A number of researchers have proposed methods for optimal sizing of sleep transistors in a given circuit to meet a performance constraint [5]-[9].

A large amount of current can flow during the sleep to active mode transition in an MTCMOS circuit. High peak rush current in the circuit can cause Electro-Migration (EM) problems in the power/ground rails. This rush current can also result in supply/ground bounces due to the *Ldi/dt* effect. In [10] the authors propose a wakeup strategy and a partitioning technique to limit the rush-through current. The authors of [11] tackle the problem of minimizing the wakeup time while limiting the current that flows to ground during the sleep to active mode transition. Their approach consists of first obtaining the discharge patterns of all logic cells and then grouping the circuit into a minimum number of clusters in such a way that the total discharge current of each cluster is below a given threshold. In [12] the authors introduce two power mode transition strategies to reduce the ground bounce while turning on the circuit. The first strategy uses single sleep transistor and gradually turns it on, while the second technique employs parallel-connected sleep transistors with increasing widths and turns them on one after the other starting from the transistor with the smallest width.

Due to the large amount of mode-transition energy overhead and large wakeup latency for the circuits, sometimes, for short standby periods, it is better to put the circuit in a drowsy mode instead of the sleep mode. The reason is that the wakeup latency of the drowsy circuit is much less than that of the circuit in sleep mode. The work in [13] presents multiple power modes for the circuit, but it needs multiple supply voltages (stable reference voltages to drive the gate terminal of the sleep transistor which will be operating in different points of the subthreshold conduction region during the sleep mode). In [14], the authors propose a power gating structure to support an intermediate (drowsy) power-saving mode and the traditional power cut-off mode. The idea is to add a PMOS transistor in parallel with each NMOS sleep transistor. By applying zero voltage to the gate of the PMOS transistor, the circuit can be put in an intermediate power saving mode whereby leakage reduction and data retention are both realized. Furthermore, by transitioning through this intermediate mode while changing between sleep and active modes, the magnitude of the voltage fluctuation of the power supply or ground during power-mode transitions is reduced. In the cut-off mode, the gate of the PMOS transistor is connected to V_{DD}.

None of these works attempt to minimize the power consumption during the sleep-to-active and active-to-sleep transitions or reduce wake up time and the noise generated by the power gating structure while maintaining the low standby leakage current. In this paper, we apply a charge-recycling technique to minimize the power consumption during the mode transition in a power gating structure while maintaining the wake up time. Through simulations, we show how the proposed technique also helps reduce the ground bounce in the sleep-to-active transition. Preliminary version of this work was published in [15].

III. Charge-Recycling Technique

Consider the coarse-grain MTCMOS configuration shown in Fig. 1. There are two different blocks in the circuit; one is power-gated by an NMOS sleep transistor which connects the virtual ground (VGND), i.e., node G in the figure, to the ground, whereas the other is power-gated by a PMOS sleep transistor which connects the virtual V_{DD} (VV_{DD}), i.e., node P in the figure, to the supply. In the active mode, sleep transistors S_N and S_P are in the linear region and the voltage values of the virtual ground and virtual V_{DD} are equal to 0 and V_{DD}, respectively. In the sleep mode, sleep transistors S_N and S_P are turned off; since they are high threshold voltage devices, very little subthreshold leakage current flows through them.

In practice (see below for precise conditions), all internal nodes of the gates in block C_1 and the virtual ground node, G, will be charged up to a voltage value very close to V_{DD}. This happens because G is floating and leakage current causes its voltage level to rise toward V_{DD}. Similarly, if the sleep period is long enough, all internal nodes of C_2 and the virtual supply node, P, will be discharged to a voltage very close to 0. We discuss this in more details in the following sub-section.

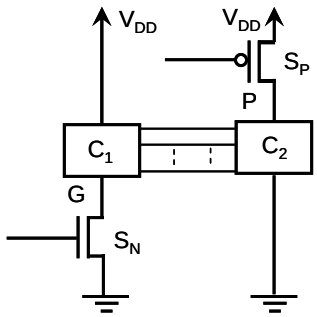

Fig. 1. The conventional power gating structure using an NMOS or a PMOS sleep transistor for each circuit block.

A. Virtual Node Voltage Values in the Sleep Mode

Consider sub-circuit C_1 in Fig. 1. We show that the assumption of the VGND node being charged to a value close to V_{DD} is invalid only when outputs of all logic cells in C_1 are forced to logic 1 (i.e., the pull-down sections of all cells are off) immediately before the active-to-sleep transition occurs. However, this case rarely happens in practice, because if there is at least one cell in C_1 with output value set to logic 0 (i.e., its pull-down section is on) before the active-to-sleep transition and if the sleep period is sufficiently long, then the steady-state value for the virtual ground voltage after entering the sleep mode will be close to V_{DD}. Considering that a sub-circuit will typically contain tens of logic cells, the probability of at least one of them having a logic 0 at its output (before entering the sleep mode) is almost 1, therefore, the voltage of the virtual ground of sub-circuit C_1 will rise and reach close to V_{DD} after sufficient time is spent in the sleep mode.

To empirically confirm the aforementioned claim, we show the voltage waveforms of the virtual ground node for four different cases in Fig. 2. In each case we have used an NMOS sleep transistor (the case with PMOS sleep transistor will be similar except that the corresponding output states are reversed). The first case is when there is only a single inverter cell in sub-circuit C_1 and the output of the inverter is logic 1 before entering the sleep mode. As the figure shows, after entering the sleep mode, the virtual ground voltage of the inverter cell rises to about 200mV, which is much less than V_{DD}=1.2V. The next case corresponds to the same sub-circuit C_1, this time with the output of the inverter forced to logic 0. Here, the virtual ground voltage rises to 0.95V, which is close to V_{DD}=1.2V and a suitable level for the charge-recycling purpose (cf. Section III.B). Next two cases correspond to C_1 comprising of 4 inverter cells each driven an input to C_1. In one case, three of the inverter outputs are 1 and only one inverter output is 0. In this case, the virtual ground voltage rises to even a higher level than case 2, resulting in a final steady sate voltage level of 1V, which is again suitable for the charge-recycling purpose (cf. Section III.B). In the last case, two inverter outputs are set to logic 1 while the others are set to logic 0. Clearly in this case, after entering the sleep mode, the virtual ground node is expected to rise and achieve a level even closer to V_{DD} than before. This is confirmed by the top waveform in the figure, which shows the virtual ground of sub-circuit C_1 reaches to a voltage close to 1.2V.

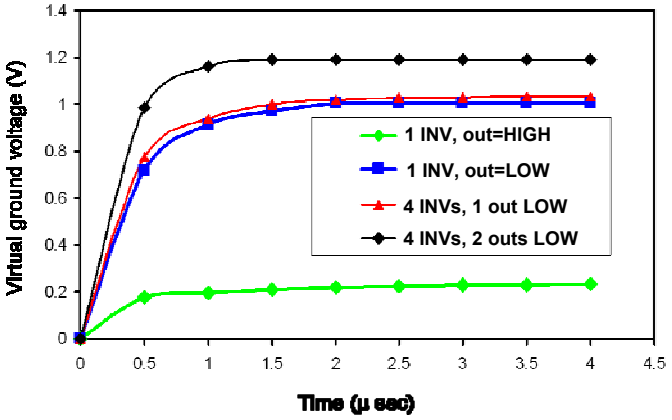

Fig. 2. The virtual ground voltage in the sleep mode, V_{DD}=1.2 V.

In summary, as long as there is a reasonably large number of logic cells in a sub-circuit that uses an NMOS sleep transistor, the probability that at least one of these cells will have a logic 0 output value before entering the sleep mode is close to one, so the virtual ground voltage of such a sub-circuit will gradually rise and stabilize to a voltage close to V_{DD}. This occurs in a relatively short period of sleep time (in the order of microseconds), which provides us with the opportunity for charge recycling between this sub-circuit and another one that uses a PMOS sleep transistor. The case that a PMOS sleep transistor is used instead of an NMOS transistor is similar and it can be shown that the VV_{DD} node is discharged to some value close to 0 during the sleep mode.

In practice, in a circuit block that uses an NMOS sleep transistor, the number and sizes of logic cells with 0 output values is sufficiently large so that the virtual ground voltage of this circuit after it enters the sleep mode rises to a value which is very close to V_{DD}. The same statement holds with respect to the VV_{DD} voltage of a circuit block that uses a PMOS sleep transistor. In this case, the VV_{DD} voltage drops to a value close to the ground voltage level after the circuit enters the sleep mode. In the analytical parts of this paper, we will assume that the virtual ground and V_{DD} voltages of circuits using NMOS and PMOS transistors will change to exactly V_{DD} and ground levels, respectively, after entering and staying in the sleep mode for a long enough time.

In the next sub-section we use this observation to propose a charge-recycling technique to achieve energy savings during mode transitions.

B. Charge recycling for Mode-Transition Energy Saving

When the sleep-to-active transition edge arrives at the gates of the sleep transistors in an MTCMOS circuit, the voltage of VGND node, G, starts to fall toward 0, whereas the voltage of VV_{DD} node, P, starts to rise toward V_{DD}. If we denote the total effective capacitance in the VGND and VV_{DD} nodes by C_G and C_P, respectively, we observe that during the active-to-sleep transition, C_G is charged up from 0 to V_{DD}, while C_P is discharged from V_{DD} to 0. The situation is reversed for the sleep-to-active transition, i.e., in this case C_G is discharged from V_{DD} to 0, while C_P is charged to V_{DD} from its initial value of 0. These charge and discharge events on the VGND and VV_{DD} nodes are wasteful from the energy dissipation point of view.

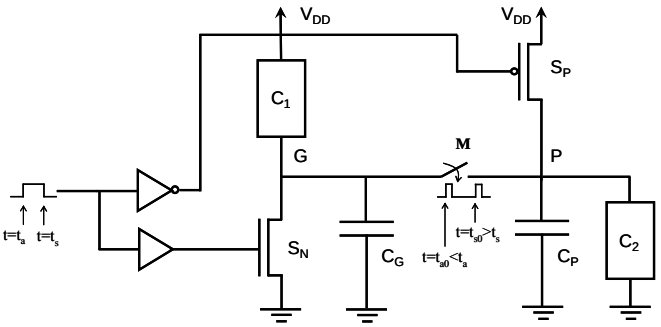

Fig. 3. The proposed charge-recycling configuration for power gating structures.

Our goal is to reduce the energy as we switch between active and sleep modes of the circuit. More precisely, we propose to use a charge-recycling technique to reduce the switching power consumption during the active-to-sleep and sleep-to-active transitions by adding a *charge sharing switch* between the VGND and VV_{DD} nodes as shown in Fig. 3. The proposed charge-recycling strategy works as follows. We turn on the charge sharing switch (i) immediately before turning on the sleep transistors while going from the sleep to the active mode, and (ii) just after turning off the sleep transistors while going from the active to the sleep mode. By turning on the switch at the end of the sleep mode as the circuit is about to go from sleep to active mode, we allow charge sharing between the completely charged up capacitance C_G and the completely discharged capacitance C_P. After the charge recycling is completed, the common voltage of the virtual ground and virtual supply is αV_{DD}, where α is a positive real number less than 1. The value of α depends on the relative sizes of C_G and C_P. As a result of this step, the mode-transition energy is reduced. The reason is that in this case, the voltage of VGND changes from αV_{DD} to 0 and the voltage of the VV_{DD} changes from αV_{DD} to V_{DD}, whereas in the conventional MTCMOS circuit, the transitions are from V_{DD} to 0 and from 0 to V_{DD} at VGND and VV_{DD} nodes, respectively. A similar analysis proves that the charge-recycling technique helps reduce the energy dissipated for transition from the active mode to the sleep mode as well.

In practice, we use a transmission gate (TG) to realize a switch (cf. Fig. 4). One may instead use other circuit realizations of a switch, such as pass transistors. Note that with a TG it is easier to achieve full charge sharing between the floating virtual ground and virtual V_{DD} nodes. We will use a TG in the rest of this paper.

The proposed charge-recycling technique is used for mode-transition energy saving in a coarse-grain MTCMOS design where each sleep transistor is used to disconnect ground/supply from *multiple* logic cells. In contrast, in fine-grain MTCMOS design, the standard cell library comprises of logic cells with integrated sleep transistors, i.e., each logic cell has its own built-in sleep transistor. Typically in such a library, virtual ground/supply nodes are considered as internal logic cell nodes. This means that the charge cycling technique cannot be applied. Of course if the logic cell library is designed such that we have access to the virtual nodes, the

charge-recycling technique can be used. In general, however, do not recommend applying the charge recycling technique at the individual cell level (fine-grain) since our basic requirement for energy saving due to charge recycling, i.e., the condition that virtual nodes change to the opposite rail values during sleep may be frequently violated under this scenario.

Another MTCMOS design configuration is the cluster-based style which we consider it as a mid-grain MTCMOS technique [6][7]. To implement the cluster-based charge recycling, we start by putting a group of, say, *n* logic cells that use NMOS sleep transistors together, connecting their virtual ground nodes to create a single virtual ground node. Similarly, a group of, say, *m* logic cells that use PMOS sleep transistors to make a single virtual supply node shared among the cells. Charge recycling can subsequently be performed between the virtual ground of one group and the virtual supply of the other group. Although cell clustering is an important optimization step, it falls outside the scope of the present paper.

In the next section we will analyze the energy saving achieved by applying the charge-recycling technique for a coarse-grain MTCMOS design.

IV. Energy Saving Analysis

In this section we first calculate the maximum achievable energy saving and discuss the conditions under which we can achieve this maximum saving. Then we quantitatively analyze the effect of threshold voltages and sizes of the transistors in the transmission gate realizing the charge sharing switch.

A. Energy Saving due to Charge Recycling

It is worth stating at the onset that for the purpose of analyzing energy consumption in CMOS circuits, energy is taken out of the V_{DD} rail only when a capacitive node is charged up through a direct connection to the V_{DD} rail. Energy that is dumped to the ground rail is the energy which was stored in that capacitive node and need not be accounted for again. The charge recycling between "floating" capacitive nodes (with possibly different initial voltage levels) does not extract any energy from the V_{DD} rail or dump any into the ground rail, instead some of the energy that was stored in the capacitors is consumed in the resistance of the switch that short circuits the two capacitive nodes while the remainder of the energy is appropriately distributed between the nodes.

To calculate energy saving of the charge-recycling technique, we consider two different transitions: wakeup transition, sleep-to-active, and sleep transition, active-to-sleep.

Case 1: *Wakeup Transition*

Let C_G and C_P represent the total capacitance in VGND and VV_{DD} nodes, respectively. We assume the sleep period is long enough so C_G is charged up to a voltage close to V_{DD}, while C_P is completely discharged to a voltage close to 0. This is a good assumption in most circuits. Otherwise, the voltages of C_G and C_P will be a function of the length of the sleep period. As stated earlier, to go from the sleep mode to the active mode, instead of simply turning on sleep transistors, we first allow charge recycling between C_G and C_P. This is done by closing switch M at time $t=t_{a0}$. Assuming ideal charge sharing between C_G and C_P, the common voltage value of nodes G and P after charge sharing is calculated by equating the total charge in both capacitances right before and after charge sharing:

$$V_f = \alpha V_{DD}$$
$$\alpha = \frac{C_G}{C_G + C_P} \tag{1}$$

The common voltage value, V_f, of VGND and VV_{DD} nodes at the end of the charge sharing is αV_{DD}. After the charge sharing is complete, i.e., at time $t=t_{a1}$, we open switch M and turn on the sleep transistors, S_N and S_P. As a result, there will be a path from the virtual ground to the (actual) ground going through S_N which would discharge C_G to 0. There will also be a path from the virtual V_{DD} to the (actual) V_{DD} going through S_P which would charge C_P to V_{DD}. For now we neglect the energy consumption in turning on and off the switch itself, so, the total energy drawn from the power supply is due to the process of charging capacitance C_P which can be obtained as follows:

$$E_{sleep\to active} = C_P V_{DD} \Delta V$$
$$= C_P V_{DD} \left(V_{DD} - V_f \right) \tag{2}$$

Substituting V_f from (1) into (2), we obtain the energy consumed during sleep-active transition:

$$E_{sleep\to active} = C_P V_{DD} \left(V_{DD} - \alpha V_{DD} \right)$$
$$= \left(1-\alpha\right) C_P V_{DD}^2 \tag{3}$$

Next we consider active-to-sleep transition.

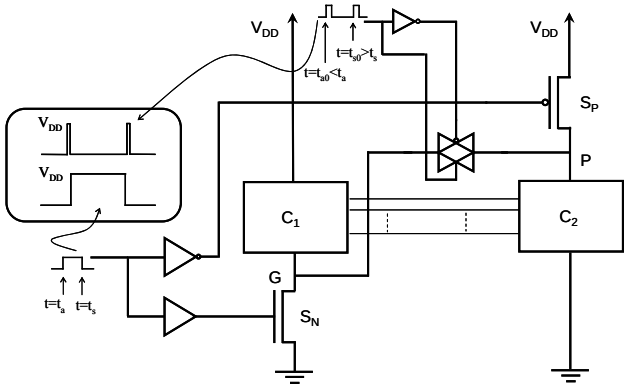

Fig. 4. The proposed charge-recycling configuration with a TG realization of the charge sharing switch.

Case 2: *Sleep Transition*

As mentioned earlier, to go from the active mode to the sleep mode, instead of simply turning off the sleep transistors, we do charge recycling between C_G and C_P as soon as the circuit enters the sleep mode. In other words, we close switch M at $t=t_{s0}$ which is the time when the sleep transistors are turned off. The voltage values of the VGND and VV_{DD} nodes at $t=t_{s0}$ are 0 and V_{DD}, respectively. Assuming ideal charge sharing between C_G and C_P, the common voltage value of nodes G and P after charge sharing is calculated by equating the total charge in both capacitances right before and after charge sharing:

$$V_f = \beta V_{DD}$$
$$\beta = \frac{C_P}{C_G + C_P} \quad (4)$$

Based on the above equation, the common voltage value, V_f, of VGND and VV_{DD} at the end of charge sharing is βV_{DD}. The charge recycling is complete at $t=t_{s1}$, so we open the switch. After opening the switch, there is a leakage path from the power supply to the virtual ground going through logic block C_1 which eventually causes C_G to be charged up to V_{DD}. There is also a leakage path from the virtual supply to the ground going through logic block C_2 which eventually causes C_P to be completely discharged to the ground. Again, if we neglect the energy consumption for turning on and off the switch, the total energy consumed is due to charging up the capacitance C_G; the energy consumption can be calculated as follows:

$$E_{active \to sleep} = C_G V_{DD} \Delta V = C_G V_{DD} \left(V_{DD} - V_f \right) \quad (5)$$

Substituting V_f from (4) into (5), we obtain:

$$E_{active \to sleep} = C_G V_{DD} \left(V_{DD} - \beta V_{DD} \right) = (1-\beta) C_G V_{DD}^2 \quad (6)$$

Since $\alpha+\beta=1$, the total energy consumption will be:

$$E_{CR-MTCMOS} = E_{active \to sleep} + E_{sleep \to active} = \alpha C_G V_{DD}^2 + \beta C_P V_{DD}^2 \quad (7)$$

where $E_{CR\text{-}MTCMOS}$ is the dynamic energy consumption during mode transition in the charge-recycling circuit.

We can calculate the total energy consumption of the corresponding conventional MTCMOS circuit, i.e., when no charge recycling is used using the following formula:

$$E_{MTCMOS} = C_G V_{DD}^2 + C_P V_{DD}^2 \quad (8)$$

From (7) and (8), and after substituting for α and β from (1) and (4), the *energy saving ratio* (ESR) would be:

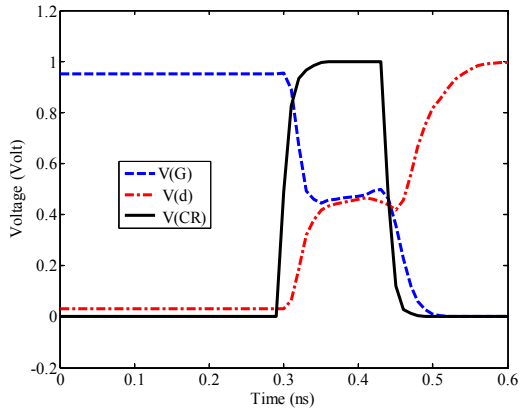

Fig. 5. The charge recycling waveforms for an inverter chain implemented in a 70 nm CMOS technology, V_{DD}=1V.

$$ESR(X) = \frac{E_{MTCMOS} - E_{CRMTCMOS}}{E_{MTCMOS}} = \frac{2X}{(1+X)^2} \quad (9)$$

where $X=C_G/C_P$ is the ratio of the VGND capacitance to the VV_{DD} capacitance. The optimum value of X which maximizes $ESR(X)$ is obtained by equating the derivative of $ESR(X)$ to zero which results in X=1, or $C_G=C_P$. In other words, in order to obtain the maximum energy saving, we need to have equal capacitances in VGND and VV_{DD}. The maximum energy saving is:

$$ESR_{\max} = ESR(X)\,|_{X=1} = 1/2 \quad (10)$$

This means a maximum energy saving of 50% can be achieved by using the charge-recycling method. However, considering the power needed to turn the TG on and off, the total saving ratio would be less than 50%.

Fig. 5 shows HSPICE waveforms when charge recycling is performed before transitioning from the sleep to the active mode for an inverter chain implemented in 70nm CMOS technology. Note that in the circuit, $C_G=C_P$. The figure shows the virtual ground voltage, V_G, the virtual V_{DD} voltage, V_P, and the charge-recycling signal, V_{CR}.

We denote the virtual ground and virtual supply capacitances as C_G and C_P, respectively. The total effective capacitance in the virtual ground (supply) comprises of the following components:

(a) **Diffusion capacitance (C_{diff}):** This component is calculated as the summation of the diffusion capacitances of transistors in logic gates connected to the virtual ground (supply).

(b) **Interconnect Capacitance (C_{wire}):** This component is the total rail capacitance in the virtual ground (supply) due to interconnect.

(c) **Internal node capacitance (C_{inte}):** This component is calculated as the total internal node capacitance of logic gates connected to virtual ground (supply) whose voltage values transition from V_{DD} to 0 or vice versa during mode transitions.

The total virtual node capacitances can thus be written as:

$$C_G = C_{diff_G} + C_{wire_G} + C_{inte_G}$$
$$C_P = C_{diff_P} + C_{wire_P} + C_{inte_P} \quad (11)$$

Now suppose each block C_1 and C_2 in Fig. 1 consists of a simple inverter. When charge recycling is performed, after the active-to-sleep transition, the value of C_G depends on the state of the inverter in C_1. To be more precise, we consider two cases as follows.

Case 1: When the input of the inverter is at logic zero, the NMOS transistor of the inverter is OFF, so the total capacitance, C_G, is the sum of the first two components in (11) (no internal node capacitance).

Case 2: When the input of the inverter is at logic one, the NMOS transistor of the inverter is ON, and the internal node capacitance contributes to C_G.

Similar discussion holds for the C_P capacitance and the state of the inverter in C_2 block. This makes C_G and C_P values input-pattern dependent for a general circuit, meaning that different input patterns applied to the circuit result in different logic values for the inputs of the circuit's gates which changes the contribution of the internal node capacitances to the total rail capacitance resulting in different C_G and C_P values. Fortunately, our simulations for circuit blocks containing a reasonable number of logic cells (e.g., more than 20 gates per block) show that the maximum change in the shared voltage value after charge recycling operation is less than 5% for different input patterns. In other words the impact

of the input vector on unbalancing the total virtual ground and virtual supply capacitance values is small and can be neglected.

Finally we point out that the energy saving ratio is only a weak function of the ratio between C_G and C_P. From (9), the maximum ESR is achieved when C_G=C_P. However, even when this condition is not satisfied, the energy saving ratio will not decrease dramatically, for example, for C_P=2×C_G which means X=1/2 in (9), ESR becomes 44%, and for C_P=3×C_G, X=1/3, ESR becomes 38%. Therefore, even in cases where C_G and C_P values are different by as much as 2 or 3 times, the energy saving ratio is still large.

Note all the equations we derived so far were based on the assumption of having an ideal charge recycling between C_G and C_P. Under this scenario, we assume that no energy is consumed to switch the TG on and off. We also assume that the TG is on while the charge recycling is in progress. However, because of the dynamic power consumption in the TG, and also the possibility of having incomplete charge sharing, this is not a perfect replacement in practice. In the following we study the effects of the TG threshold voltage and sizing on the energy saving ratio and the wakeup time of the charge-recycling configuration.

B. Effect of the Threshold Voltages of the TG

We first discuss the effect of threshold voltages of the NMOS and PMOS transistors of the TG on the energy saving and the delay of the circuit.

Consider the charge sharing configuration shown in Fig. 6 where V_1 and V_2 are set to V_{DD} and 0 levels initially. After the TG is closed, the common node voltage is referred to as V_f. To have a complete charge sharing, the TG has to stay on for the whole duration of the charge sharing process. In order to have this property, the absolute values of the threshold voltages of the N and P transistors of the TG have to be small enough. To guarantee this, the common final voltage value of virtual ground and virtual supply, V_f, has to satisfy at least one of the following two inequalities:

$$\begin{cases} V_{t,n} \le V_{DD} - V_f & \text{or} \\ \left|V_{t,p}\right| \le V_f \end{cases} \qquad (12)$$

where $V_{t,n}$ and $V_{t,p}$ denote threshold voltages of the NMOS and PMOS transistors in the TG accounting for the body effect. Notice that V_f can be obtained from (1) for the active to sleep case and from (4) for the sleep to active case. The inequalities in (12) guarantee that at least one of the transistors in the TG remains on for the complete duration of charge sharing.

In the case of equal virtual node capacitances, C_G=C_P, a complete charge sharing in both active-to-sleep and sleep-to-active cases results in a common final voltage value of V_f=V_{DD}/2, and (12) translates into $Min\{V_{t,n}, |V_{t,p}|\} \le V_{DD}/2$.[1] Now, if V_{tn}=$|V_{tp}|$≤V_{DD}/2, a TG may be replaced with a pass transistor while still achieving full charge sharing. Note in current CMOS technologies this condition is easily satisfied for both LVT and HVT devices so as to have acceptable static DC noise margins. In the future CMOS technologies that use sub-1V power supply level, as it will be discussed in Section VI, turning on the HVT devices will be difficult, and that is why Super Cut-off CMOS (which uses voltage over or under drive) was developed in [16]. Therefore, for sub-1V technologies, we recommend using CR-SCCMOS instead of CR-MTCMOS (cf. Section VI). In this case, transistors of the transmission gate will be LVT and V_{tn}, $|V_{tp}| \le V_{DD}/2$ will be automatically satisfied (otherwise, even the CMOS logic cells inside the logic blocks, which all use LVT transistors, would fail).

[1] If $Min\{V_{t,n}, |V_{t,p}|\} > V_{DD}/2$, the charge-recycling will not be complete, and the ESR value will be less than what we have predicted.

C. Effect of the Transistor Sizes of the TG

Sizing of the TG is another factor that affects the ESR as well as the wake up time of the circuit. In case of the original configuration when there is not any charge recycling, the wakeup time is typically defined as the time that it takes for the voltage of the virtual ground or virtual V_{DD} to reach within 10% of their final values after we turn on the sleep transistor. In the proposed charge recycling solution, we first turn on the TG in order to perform charge sharing between C_G and C_P, and next we switch on the sleep transistors to complete the mode transition. Therefore, in the charge-recycling circuit, the wakeup time is defined as the time that it takes for the voltage of the virtual ground or virtual V_{DD} to reach within 10% of their final values after we turn on the TG. In the following discussion, we consider the effect of the dynamic power consumption of the TG on the ideal energy saving ratio, ESR, which we previously calculated.

Consider TG with its control signal (the complement of the control signal is produced by a CMOS inverter). Assume a total input capacitance of C_{tg} for the NMOS and PMOS transistors of the TG. In each active-sleep-active cycle, we need to turn on the TG twice, once before turning the sleep transistors on and once after turning them off. Every time we turn the TG on and off, we charge and discharge C_{tg}. We have to turn off the TG after the charge sharing is complete.

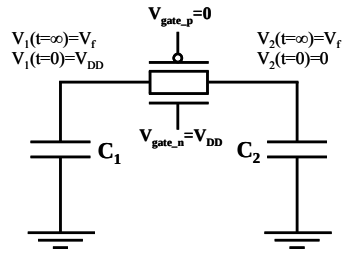

Fig. 6. Charge sharing between C_1 and C_2 when using a TG to realize the charge sharing switch.

Therefore, we can calculate the dynamic energy consumption of the TG for one complete active-sleep-active cycle as follows:

$$E_{TG} = 2C_{tg}V_{DD}^2 \qquad (13)$$

Therefore, the actual energy saving ratio (ESR) can be calculated by subtracting the correction ratio E_{TG}/E_{MTCMOS} from the ideal ESR in (9). The correction ratio can be calculated as:

$$\frac{E_{TG}}{E_{MTCMOS}} = \frac{2C_{tg}V_{DD}^2}{(C_G + C_P)V_{DD}^2} = \frac{2C_{tg}}{C_G + C_P} \qquad (14)$$

This correction ratio is proportional to the sizes of the TG's transistors since C_{tg} is proportional to the size of the TG. Because many gates are usually connected to the virtual ground and the virtual V_{DD}, C_G+C_P is usually much larger than C_{tg}. Thus, the correction ratio is usually few percents making the actual ESR to be less than the ideal ESR, i.e., 50%, by only a few percentage points.

Fig. 7 shows the ESR versus total transistor width used in the TG. As seen the ESR is reduced as we increase the TG size.

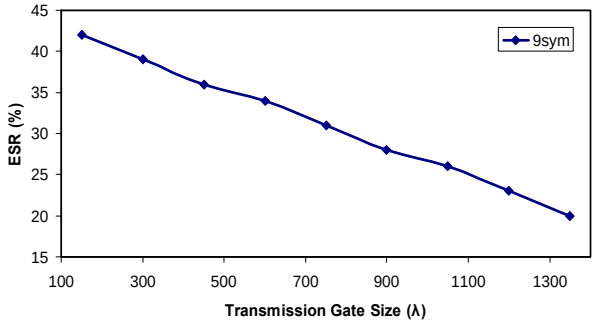

Fig. 7. Percentage of Energy Saving Ratio (ESR) versus size of the transmission gate used for 9sym benchmark circuit.

By changing the TG size, we can change the speed of charge sharing operation and as a result, minimize the wakeup time; however, charge-sharing operation only changes the virtual node voltages from their initial values to $V_{DD}/2$. The rest of the wakeup operation is performed by the sleep transistors and its duration depends on the sizes of the sleep transistors. Clearly, increasing the TG size does not affect the speed by which the sleep transistors can change the virtual node voltages from $V_{DD}/2$ to V_{DD} or 0 as the case may require. Therefore, the total wakeup time of the circuit is expected to decrease when we increase the TG size, but then it saturates at some point.

Fig. 8 shows the circuit wakeup time versus the total transistor width used in TG. Finally note that although increasing the TG size reduces the wakeup time, it also increases the correction ratio given in (14), thereby, changing the energy saving ratio of the circuit. In other words, there is a tradeoff between the wakeup time and the energy saving ratio.

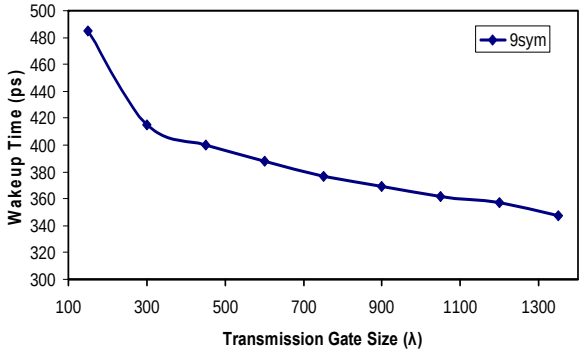

Fig. 8. Wakeup time versus size of the transmission gate used for 9sym benchmark circuit.

V. Leakage Current and Ground Bounce Analysis

We analyze two important issues for the proposed charge-recycling MTCMOS configuration, namely the leakage current and the ground bounce (GB).

A. Leakage Current

In the sequel, we derive the subthreshold leakage current equations for both MTCMOS and CR-MTCMOS circuits. The leakage current of a MOS transistor can be written as follows [16]:

$$I_{leakage} = \mu_0 \frac{\varepsilon_{ox}}{T_{ox}} \frac{W}{L} \nu_T^2 e^{1.8} \mathrm{e}^{\frac{|V_{gs}-V_{th}|}{S\nu_T}} \left(1 - \mathrm{e}^{-\frac{|V_{ds}|}{\nu_T}}\right) \qquad (15)$$

where V_{gs} and V_{ds} are the gate-source and drain-source voltages of the transistor and W/L is the width to the length ratio of the transistor. In the sleep mode, all sleep and charge-recycling transistors are off, i.e., they all have $V_{gs}=0$. Here, V_{ds} for each charge-recycling transistor is the absolute voltage difference between VGND and VV_{DD} nodes in the sleep mode, which is approximately equal to V_{DD} based on the discussion in Section III. From (15), we can ignore the dependence of the transistor's subthreshold leakage current on V_{ds} since $V_{ds} \geq 75mv$. There are two leakage current components corresponding to the two leakage paths in the conventional MTCMOS circuit: the NMOS sleep transistor leakage current (I_{Ln}) and the PMOS sleep transistor leakage current (I_{Lp}). Assuming the widths of NMOS and PMOS sleep transistors are W_n and W_p, respectively, I_{Ln} and I_{Lp} can be written as:

$$\begin{aligned} I_{Ln} &= \mu_n \frac{\varepsilon_{ox}}{T_{ox}} \frac{W_n}{L} \nu_T^2 e^{1.8} \mathrm{e}^{\frac{-V_{tH}}{S\nu_T}} \\ I_{Lp} &= \mu_p \frac{\varepsilon_{ox}}{T_{ox}} \frac{W_p}{L} \nu_T^2 e^{1.8} \mathrm{e}^{\frac{-V_{tH}}{S\nu_T}} \end{aligned} \qquad (16)$$

where V_{tH} is the threshold voltage of the sleep transistors. The total leakage current of the MTCMOS circuit is the sum of I_{Ln} and I_{Lp}:

$$I_{leakage}^{MTCMOS} = \left(\mu_n W_n + \mu_p W_p\right) \frac{\varepsilon_{ox}}{L T_{ox}} \nu_T^2 e^{1.8} \mathrm{e}^{\frac{-V_{tH}}{S\nu_T}} \qquad (17)$$

For the charge recycling MTCMOS (CR-MTCMOS), however, there is an additional leakage component due to the charge-recycling switch (I_{Lcr}). For the purpose of this section, assume instead of a TG, a single NMOS transistor with the width W_{cr} is used for charge recycling. Using (15) I_{Lcr} can be written as:

$$I_{Lcr} = \mu_n \frac{\varepsilon_{ox}}{T_{ox}} \frac{W_{cr}}{L} \nu_T^2 e^{1.8} \mathrm{e}^{\frac{-V_{tH}}{S\nu_T}} \qquad (18)$$

Using (17) and (18), the ratio of the leakage current for MTCMOS and CR-MTCMOS can be written as:

$$\begin{aligned} \frac{I_{leakage}^{CR-MTCMOS}}{I_{leakage}^{MTCMOS}} &= \frac{\mu_n W_n + \mu_n W_{cr} + \mu_p W_p}{\mu_n W_n + \mu_p W_p} \\ &= 1 + \frac{W_{cr}}{W_n + \left(\mu_p/\mu_n\right) W_p} \end{aligned} \qquad (19)$$

Assuming $\mu_n=2\mu_p$ and $W_n=0.5W_p$:

$$\frac{I_{leakage}^{CR-MTCMOS}}{I_{leakage}^{MTCMOS}} = 1 + \frac{W_{cr}}{2W_n} \quad (20)$$

Since the charger-recycling transistor is usually much smaller than the sleep transistors, the leakage-increase ratio given in (20) is usually too small when compared to the power saving achieved by using the charge-recycling technique.

B. Ground Bounce

Ground and power line bounces are one of the most important design concerns when power gating is used [12]. Ground Bounce (GB) or power bounce may occur in power gating structures at the sleep to active transition edge. In this section we discuss about how charge-recycling technique affects the ground bounce. Consider the circuit in Fig. 9. Large current flows into the ground after the sleep transistor is turned on at the end of the sleep period. We adopt a simple RL model for the purpose of GB analysis. Because of the large *di/dt* at the turn-on time, a large voltage, i.e., *Ldi/dt*, appears across the inductance. We next study the effect of the proposed charge-recycling technique on the GB of the circuit.

Fig. 9 shows the virtual ground capacitance, C_G, connected to the RL circuit (modeling the pin-package parasitics of the IC), via the sleep transistor, S_N. The sleep transistor is turned on at t=0 when the initial voltage of C_G is V_0, i.e., $V_G(t=0)=V_0$. Based on the results of [20], the positive peak of the GB occurs during the time when S_N operates in the saturation region. If we neglect the channel length modulation effect, the saturation current of S_N does not depend on V_0. Therefore, we expect the proposed charge-recycling technique, which changes V_0 from V_{DD} to $V_{DD}/2$, would not change the GB's positive peak. However, due to the channel length modulation effect, the saturation current of the sleep transistor, S_N, is somewhat smaller for the CR-MTCMOS compared to the MTCMOS circuit. This results in smaller GB for the CR-MTCMOS circuit. In addition, the negative peak and the settling time of GB are functions of V_0 i.e., they both decrease when V_0 decreases [20]. Therefore, both the negative peak value and the settling time of the GB voltage are expected to decrease for the CR-MTCMOS circuit.

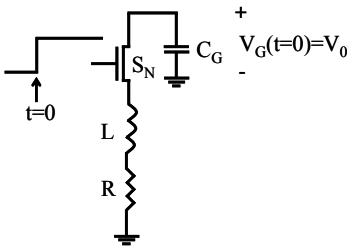

Fig. 9. The RL equivalent model of the ground used to analyze the GB effect in MTCMOS circuits.

The amounts of improvement in the negative peak and settling time depend on the relative values of L, C_G, R, V_{DD}, and the sleep transistor parameters. Fig. 10 compares GB waveforms for the conventional and the charge-recycling power gating structures used for an inverter chain implemented in 70nm CMOS technology. As expected, the positive peak value is almost the same in both cases; however, the negative peak value and the settling time are smaller for the charge-recycling MTCMOS structure.

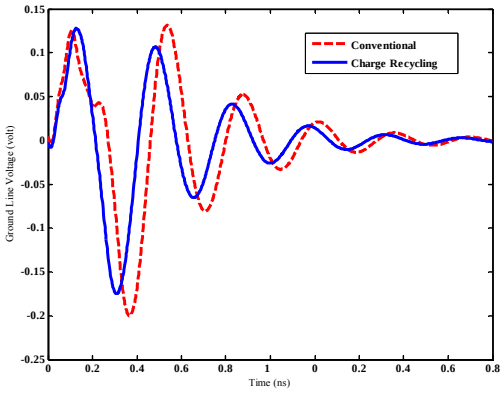

Fig. 10. The GB waveforms in the conventional and the CR structures for an inverter chain implemented in a 70nm CMOS technology, V_{DD}=1V.

VI. VARIANTS OF THE CHARGE-RECYCLING TECHNIQUE

Previously, we presented a certain type of charge-recycling technique that uses both NMOS and PMOS sleep transistors. Charge recycling was then applied between VGND and VV_{DD} nodes. In this section we discuss three variations of the proposed charge-recycling technique for the MTCMOS circuits.

A. Charge Recycling Between the Same Type of Virtual Rails

Consider **Error! Reference source not found.**.a where two circuit blocks C_1 and C_2 are using the same type of sleep transistors, e.g., NMOS transistors. Suppose C_1 and C_2 work in "orthogonal" modes, i.e., when C_1 is in active mode, C_2 is in sleep mode and vice versa. For example, C_1 and C_2 can be integer and floating-point arithmetic blocks of a processor. When the integer arithmetic block is used, the floating-point block will be idle and conversely. We show charge recycling can be performed between VGND nodes of blocks C_1 and C_2, denoted by $VGND_1$ and $VGND_2$, respectively.

First assume C_1 is in the active mode and C_2 is in the sleep mode. Voltages of $VGND_1$ and $VGND_2$ are 0 and V_{DD}, respectively. When C_1 is switched to the sleep mode, C_2 is switched to the active mode and the voltages of $VGND_1$ and $VGND_2$ change to V_{DD} and 0, respectively. Therefore, the charge recycling can be done between $VGND_1$ and $VGND_2$ nodes to save the mode transition energy.

The energy consumptions for the MTCMOS and CR-MTCMOS circuits in a full active-sleep-active cycle are:

$$E_{MTCMOS} = \left(C_{G_1} + C_{G_2}\right)V_{DD}^2$$
$$E_{CR-MTCMOS} = C_{G_1}V_{DD}\Delta V_1 + C_{G_2}V_{DD}\Delta V_2 \quad (21)$$

where ΔV_1 and ΔV_2 are the voltage differences between the final charge-recycling voltage value and the supply voltage values of the two blocks and are calculated as follows:

$$\Delta V_1 = V_{DD} - \frac{C_{G_2}}{C_{G_1} + C_{G_2}} V_{DD}$$
$$\Delta V_2 = V_{DD} - \frac{C_{G_1}}{C_{G_1} + C_{G_2}} V_{DD} \quad (22)$$

Substituting ΔV_1 and ΔV_2 from (22) into (21), we can calculate the energy saving ratio as:

$$\frac{E_{CR-MTCMOS}}{E_{MTCMOS}} = \frac{C_{G_1}^2 + C_{G_2}^2}{\left(C_{G_1} + C_{G_2}\right)^2} \quad (23)$$

which is similar to the regular charge-recycling case. The maximum energy saving of 50% is achieved when C_{G1}=C_{G2}. Similarly, the charge-recycling technique may be applied between the VV_{DD} nodes of two blocks that use PMOS sleep transistors.

B. Charge Recycling for Blocks with Different Power Supply Levels

Consider **Error! Reference source not found.**.b where two circuit blocks C_1 and C_2 use two different power supply levels, V_{DD1} and V_{DD2}, respectively.

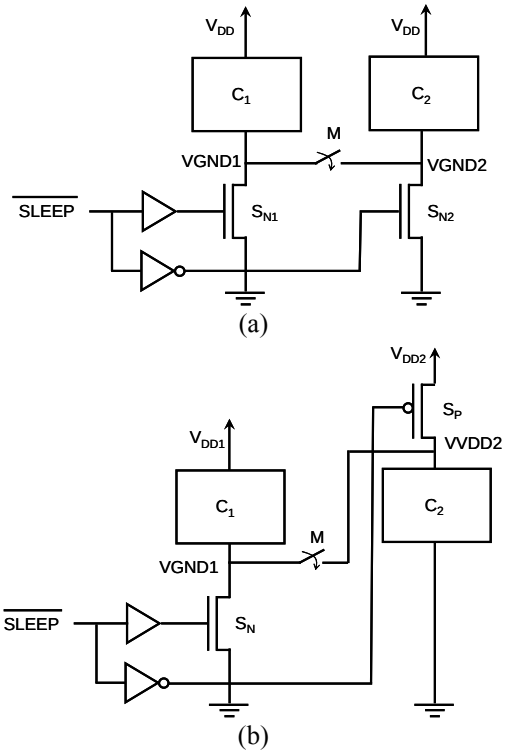

Fig. 11. (a) Charge recycling between two virtual grounds: VGND1 and VGND2 (b) Charge recycling between the virtual rails of blocks with different supply levels.

If C_1 and C_2 use different types of sleep transistors, for example, C_1 uses an NMOS while C_2 uses a PMOS sleep transistor and if C_1 and C_2 are always in the same mode of operation (i.e., they are both in the sleep mode or they are both in the active mode), then the charge-recycling technique may be applied between the virtual ground of C_1, VGND1, and the virtual supply of C_2, VV_{DD2}.

In this case, the energy consumptions for the MTCMOS and CR-MTCMOS circuits can be written as follows:

$$E_{MTCMOS} = C_{G_1} V_{DD_1}^2 + C_{P_2} V_{DD_2}^2$$
$$E_{CR\,MTCMOS} = C_{G_1} V_{DD_1} \Delta V_1 + C_{P_2} V_{DD_2} \Delta V_2 \quad (24)$$

where ΔV_1 and ΔV_2 are the voltage differences between the final charge-recycling voltage value and the supply voltage values of the two blocks and are calculated as follows:

$$\Delta V_1 = V_{DD_1} - \frac{C_{P_2}}{C_{G_1} + C_{P_2}} V_{DD_2}$$
$$\Delta V_2 = V_{DD_2} - \frac{C_{G_1}}{C_{G_1} + C_{P_2}} V_{DD_1} \quad (25)$$

Substituting ΔV_1 and ΔV_2 from (25) into (24), we can calculate the energy saving ratio as:

$$ESR = \frac{E_{MTCMOS} - E_{CR-MTCMOS}}{E_{MTCMOS}} = \frac{2C_{G1}C_{P2}V_{DD1}V_{DD2}}{\left(C_{G1} + C_{P2}\right)\left(C_{G1}V_{DD1}^2 + C_{P2}V_{DD2}^2\right)} \quad (26)$$

One can see from (26), the energy saving ratio in this case depends not only on the capacitance values in the virtual rails, but on both supply voltage values. Notice that if $V_{DD1} = V_{DD2}$ then (26) is reduced to (9).

TABLE I shows the energy saving results for the two variants of the charge-recycling technique discussed in parts A and B of this Section. This table includes three example cases of charge recycling for the same type of virtual rail. In each case, we have used two blocks of the same circuit when they both employ NMOS sleep transistors. TABLE I also includes a charge-recycling case for blocks with different supply levels. In this case we put together two circuit blocks 9sym and C880 where 9sym employed a PMOS sleep transistor and a supply voltage of V_{DD1}=1.3V whereas C880 used an NMOS sleep transistor and a supply voltage of V_{DD2}=1.0V. The results show that the energy consumption during mode transition for CR-MTCMOS is less than that for MTCMOS by an average of 36%.

C. Charge Recycling for Super Cut-off CMOS

Turning on HVT devices is difficult in sub 1-V CMOS technologies [16][18]. In 45nm technology, the best corner V_{DD} is 0.9V while the standard threshold voltage, SVT, is about 0.5V. For acceptable leakage saving, the high threshold voltage must be at least 0.65V. This leaves only a 0.25V margin for the gate-source voltage ($0.65 < V_{GS} < 0.9$V) of a turned on NMOS sleep transistor when MTCMOS is used. Therefore, high threshold voltage (HVT) sleep transistors are too slow and hard to turn on in sub 1-V technologies. Super Cut-off CMOS (SCCMOS) circuits solve this problem by using a low threshold voltage (LVT) device for cutting off ground or V_{DD} [16]. Instead of using HVT devices for leakage reduction, SCCMOS circuits overdrive the LVT PMOS sleep transistors by applying a positive overdrive voltage of ΔV_{DD} in excess of V_{DD} to their gate terminals. Similarly, they under drive the LVT NMOS sleep transistors by applying a negative voltage of $-\Delta V_{DD}$ to their gate terminals. It has been shown the SCCMOS circuits achieve the same leakage reduction as the corresponding MTCMOS circuits with shorter wakeup times due to the use of LVT transistors.

TABLE I
COMPARISON OF THE DYNAMIC ENERGY CONSUMPTION OF MTCMOS AND VARIANTS OF CR-MTCMOS CIRCUITS.

Circuit Blocks	Type	Avg. # of Cells per block	Avg. SLP TX width per block (λ)	Total CR TX width (λ)	Dynamic Energy Dissipation for Mode Trans. (Femto Jules)		ESR (%)
					MTCMOS	CR-MTCMOS	
9sym/9sym	2VGND	276	1200	600	866	525	39.4
C432/C432	2VGND	204	1000	450	673	375	44.3
C880/C880	2VGND	432	1600	1050	1356	851	37.2
C880/9sym	2VDD	354	2200	750	2373	1800	24.1
Avg.	-	-	-	-	1317	887.8	36.3

Similar to MTCMOS, conventional SCCMOS circuits suffer from wasteful mode transition energy consumption. Both NMOS and PMOS sleep transistors may be used to cut off power or ground from the gates inside a circuit. During the standby mode, due to leakage, the VGND node will be charged to a value close to V_{DD} while the VV_{DD} node will be discharged to a voltage close to zero [18]. The opposite situation occurs in the active mode. Consequently, charge recycling may be applied to SCCMOS circuits to save the mode transition energy in the same fashion as it is applied to MTCMOS circuits. Fig. 12 shows the configuration of the circuit used for charge-recycling SCCMOS (CR-SCCMOS).

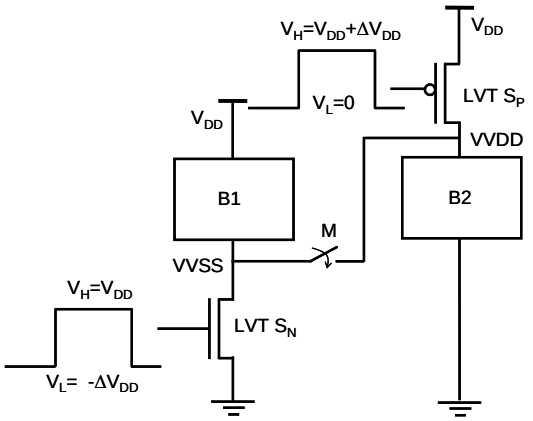

Fig. 12. Charge recycling for SCCMOS circuits.

TABLE III reports the results of applying the charge-recycling technique to SCCMOS circuits. In order to have a fair comparison between each MTCMOS and its SCCMOS counterpart, the value of the overdrive voltage for a PMOS sleep transistor in the SCCMOS circuit, i.e., ΔV_{DD}, is set to the threshold voltage difference between the HVT and LVT PMOS devices in the MTCMOS circuit. Similarly, the value of the underdrive voltage for an NMOS sleep transistor in the SCCMOS circuit, $-\Delta V_{DD}$, is set to the threshold voltage difference between the LVT and HVT NMOS devices in the MTCMOS circuit.

VII. SIMULATION RESULTS

We used the ISCAS-85 circuit benchmark suite to generate our experimental results. All benchmark circuits are first optimized using "script.rugged" in SIS. We used a 90nm cell library to perform timing-driven technology mapping. The LVT value is 0.25V, whereas the HVT value is 0.55V for NMOS transistors. Similarly, for PMOS transistors LVT value is -0.22V, whereas the HVT value is -0.52V. The supply voltage's value is V_{DD}=1.2V.

Starting with an optimized and technology mapped ISCAS-85 circuit, we first generate the MTCMOS version of the same circuit as follows. We use a single NMOS sleep transistor to cut off the ground from the virtual ground node during the sleep time. The size of this sleep transistor is set to ensure a voltage drop of no more than 5% of V_{DD} across its R_{DS}(ON) when the circuit is active. This limits the performance penalty of the power gating structure. The exact solution to this problem requires an optimization that falls outside the scope of this paper. Interested readers may refer to [6][7][21] for different ways in which the problem can be formulated and solved. Let N denote the number of logic gates in the circuit. In our experiments, we assumed at most 10% of logic gates in the circuit exhibit a simultaneous high-to-low output transition in any given cycle, each transition contributing an average of ΔI_{avg} current to the total current flowing through the on sleep transistor, and therefore,

$$R_{ds,n}(ON) = \frac{\Delta V}{I} = \frac{0.05V_{DD}}{\frac{N}{10}\Delta I_{avg}} = \frac{V_{DD}}{2N\Delta I_{avg}}$$
$$\left(\frac{W}{L}\right)_n = \frac{1}{R_{ds,n}(ON)\mu_n C_{ox}(V_{DD} - V_{tH,n})} \quad (27)$$

This simple derivation produces reasonably good results for the size of the MTCMOS sleep transistor in our benchmark suite. However, in general, a more sophisticated sizing technique is needed to guarantee that the worst-case path delay increase is below some pre-specified target level. In the table of results, we use the notation ST-MTCMOS to refer to standard MTCMOS version of circuits.

Next, we generate a version of the circuit benchmarks that uses both NMOS and PMOS sleep transistors. In particular we partition circuit C into two blocks C_1 and C_2, where C_1 uses an NMOS sleep transistor, while C_2 uses a PMOS one. Furthermore, the partitioning is done such that the total capacitance of the virtual ground node of C_1 is equal to the total capacitance of the virtual voltage node of C_2. The sizing of the NMOS and PMOS sleep transistors for each circuit block is done similar to the ST-MTCMOS case (accounting for the difference between hole and electron mobility). We

refer to this version as the NP-MTCMOS because it uses both types of sleep transistors, yet it does not perform any charge recycling.

We incorporate the charge recycling technique into NP-MTCMOS by using an appropriately sized TG as the switch between the VGND of C_1 and VV_{DD} of C_2. The size of this TG is selected such that the wakeup times of the NP-MTCMOS and the CR-MTCMOS are approximately equal. The optimization is performed by measuring the wakeup time of the NP-MTCMOS and sweeping the TG size (using SPICE) while monitoring the wakeup time of the CR-MTCMOS circuit.

NMOS transistors have higher drive strength compared to PMOS transistors; thus, from a layout area point of view it is better to use NMOS sleep transistors. However, the sleep transistor size is not the only factor determining whether NMOS or PMOS sleep transistors must be used. Other factors such as leakage and noise on power/ground rails are also important. For example, PMOS transistors have lower leakage. In any case since the total area overhead of the sleep transistors is relatively small (it is typically less than 5% of the total logic cell area), using NMOS vs. PMOS sleep transistors does not make a big difference in terms of the total area. An important issue is the cost of implementing PMOS or NMOS sleep transistors in the given process technology as follows. If NMOS sleep transistors are used, body connections of the NMOS transistors of logic cells have to be tied to the VGND node in order to minimize the body effect. On the other hand, the body connection of the NMOS sleep transistor has to be tied to the actual ground. Thus, a three-well CMOS process is required, which is more expensive than a typical two-well CMOS process. In contrast if PMOS sleep transistors are used, the p-substrate easily separates the n-well of these transistors from other n-wells which contain PMOS transistors used in the normal cells.

TABLE II
COMPARISON OF THE DYNAMIC ENERGY CONSUMPTION OF ST-MTCMOS, NP-MTCMOS AND CR-MTCMOS CIRCUITS.

Circuit	# Cells connected to VGND	# Cells connected to VVDD	Total SLP TX width (λ)	Total CR TX width (λ)	Dynamic Energy Dissipation for Mode Transitions (Femto Jules)			ST-ESR (%)	NP-ESR (%)
					ST-MTCMOS	NP-MTCMOS	CR-MTCMOS		
9sym	145	131	1,620	300	1240	1600	930	25	39.0
C432	128	76	1,120	240	890	1060	660	25.8	37.3
C880	232	200	2, 528	480	1880	2400	1470	21.8	38.7
C1355	296	230	3,024	480	2230	2820	1700	23.8	39.5
C3540	745	550	7,580	900	5670	7340	4290	24.4	41.6
C5315	1,017	710	9,748	900	7210	9230	5270	26.8	42.8
Avg.	-	-	-	-	3187	4075	2387	24.6	39.8

We generate NP-SCCMOS circuits by taking the NP-MTCMOS and scaling both the NMOS and PMOS sleep transistors by the following factor:

$$\frac{(V_{DD} - V_{tH,*})}{(V_{DD} - V_{tL,*})}$$

where $V_{tH,*}$ and $V_{tL,*}$ denote the HVT and LVT values of NMOS or PMOS devices.

Finally, we generate CR-SCCMOS by enabling charge sharing with an appropriately sized TG. Similar to CR-MTCMOS case, the size of this TG is determined through SPICE simulation with the goal of equating the wakeup times of NP-SCCMOS and CR-SCCMOS.

The control signal for the transmission gate needs to be synchronized with the sleep signal generated by the power management unit. The pulse duration has to be long enough to enable charge sharing but not overly long since it adds up to the wakeup time. Typically 20%-30% of the total cycle time is sufficient for the charge-recycling operation to finish. For example, in a 90nm CMOS technology with a clock frequency of 2.5GHz, the cycle time is 400ps. Thus, a 100ps pulse-width is a good choice for the charge-recycling operation. The task of synchronizing this pulse with the clock and power management control signal is similar to meeting other timing constraints in nanoscale CMOS designs.

TABLE II shows the energy saving results for various ST-MTCMOS circuits and their corresponding NP-MTCMOS and CR-MTCMOS ones. As one can see the energy consumption during mode transition for CR-MTCMOS is less than ST-MTCMOS and NP-MTCMOS by an average of about 25% and 40%, respectively. Note, in all reported cases, the wakeup times are equal. We have observed that the total sleep transistor area overhead in the NP/CR-MTCMOS is 50% more than that in the ST-MTCMOS. Since this area overhead is only a small percentage of the total chip area (less than 5%), the actual sleep transistor area overhead due to using CR-MTCMOS compared to ST-MTCMOS is small.

TABLE III shows the energy saving results for various NP-SCCMOS and corresponding CR-SCCMOS circuits. As it was explained in Section VI, in order to have a fair comparison between MTCMOS and SCCMOS circuits, the value of the overdrive voltage for a PMOS super cut-off switch in the SCCMOS circuit is set to the threshold voltage difference between the HVT and LVT PMOS devices in the MTCMOS circuit. Similarly, the value of the underdrive voltage for an NMOS switch in the SCCMOS circuit is set to the threshold voltage difference between the HVT and LVT NMOS devices in the MTCMOS circuit. As one can see the energy saving of CR-SCCMOS over NP-SCCMOS is about 36% on an average for the same wakeup time.

Reducing ground and power rail bounces is among the important issues in designing MTCMOS circuits. As it was

discussed in Section V, the proposed charge-recycling technique reduces the ground (power) bounce of the MTCMOS circuits. TABLE IV validates this expectation by reporting the positive and negative peaks of the ground bounce for various NP-MTCMOS circuits and the corresponding CR-MTCMOS circuits. As one can see the negative peak ground bounce value of the CR-MTCMOS has decreased by an average of 33% compared to NP-MTCMOS.

Next, we compare ST-MTCMOS and CR-MTCMOS circuits in terms of their total energy consumptions. The total energy consumptions in the ST-MTCMOS and CR-MTCMOS circuits may be written as the summation of their corresponding active and sleep mode energy consumptions plus the energy consumption due to the mode transition in these circuits:

$$E_{total}^{ST-MTCMOS} = E_{active}^{ST-MTCMOS} + E_{sleep}^{ST-MTCMOS} + E_{mt}^{ST-MTCMOS}$$
$$E_{total}^{CR-MTCMOS} = E_{active}^{CR-MTCMOS} + E_{sleep}^{CR-MTCMOS} + E_{mt}^{CR-MTCMOS} \quad (28)$$

The active-mode energy consumption for both cases consists of two parts: dynamic component and static (leakage) component.

TABLE III
MODE TRANSITION ENERGY CONSUMPTION OF NP-SCCMOS AND CR-SCCMOS CIRCUITS, V_{DD}=1.2V.

Circuit	# Cells Connected to VGND	# Cells Connected to VVDD	Total SLP TX width (λ)	Total CR TX width (λ)	Dynamic Energy in Mode Transition (Femto Jules)		ESR (%)
					NP-SCCMOS	CR-SCCMOS	
9sym	145	131	972	450	860	590	31.4
C432	128	76	672	330	590	410	30.5
C880	232	200	1,517	600	1270	840	33.8
C1355	296	230	1,815	480	1480	920	37.8
C3540	745	550	4,548	900	3930	2330	40.7
C5315	1,017	710	5,849	900	5020	2910	42
Avg.	-	-	-	-	2192	1333	36

TABLE IV
GROUND BOUNCE COMPARISON BETWEEN MTCMOS AND CR-MTCMOS CIRCUITS, V_{DD}=1.2, L=5NH, R=5Ω.

Circuit	Positive Peak GB (mV)			Negative Peak GB (mV)		
	MTCMOS	CR-MTCMOS	GB Reduction (%)	MTCMOS	CRMTCMOS	GB Reduction (%)
9sym	514	479	6.8	406	270	33.5
C432	476	437	8.2	375	225	40.0
C880	455	417	8.4	324	181	44.1
C1355	431	398	7.7	311	151	51.4
C3540	315	293	7.0	202	155	23.2
C5315	228	202	11.4	206	193	6.3
Avg.	397	364	8.25	291	177	33.1

The active-mode energy components in ST-MTCMOS and CR-MTCMOS circuits can be written as:

$$E_{active}^{ST-MTCMOS} = E_{active}^{CR-MTCMOS} = \left(c_{sw}V_{DD}{}^2 f_{clk} + I_{la}V_{DD}\right)t_{active} \quad (29)$$

where c_{sw} denotes the average *switched capacitance* for the circuit in each clock cycle, f_{clk} is the clock frequency, I_{la} denotes the average active leakage current in the circuit, and t_{active} is the total time the circuit is active. Let N_{clk} denote the number of the clock cycles over which energy calculations are performed.

We can write:

$$t_{active} = \alpha N_{clk} T_{clk}$$
$$t_{sleep} = (1-\alpha) N_{clk} T_{clk} \quad (30)$$

where T_{clk}=1/f_{clk} is the clock period, and α denotes the (*active*) *duty factor* which is defined as the percentage of the total time during which the circuit is in the active mode.

The sleep-mode energy consumptions for the two circuits can be written as:

$$E_{sleep}^{ST-MTCMOS} = I_{ls_n}^{ST} V_{DD} t_{sleep}$$
$$E_{sleep}^{CR-MTCMOS} = \left(I_{ls_n}^{CR} + I_{ls_p}^{CR} + I_{lcr}^{CR}\right)V_{DD} t_{sleep} \quad (31)$$

where $I_{ls_n}^{ST}$ is the subthreshold leakage current through the sleep transistor in the ST-MTCMOS circuit during the sleep mode. $I_{ls_n}^{CR}$, $I_{ls_p}^{CR}$ and $I_{ls_{cr}}^{CR}$ denote the subthreshold leakage currents through the NMOS and PMOS sleep transistors and the charge-recycling transistors in the CR-MTCMOS circuit during the sleep mode, respectively. Typically, the

leakage current through sleep transistors in both cases are in the same order, however, since the TG is much smaller than the sleep transistors, usually smaller than 1/10th, $I_{ls_{cr}}^{CR}$ in (31) is much smaller, smaller than 1/10th, than $I_{ls_n}^{CR} + I_{ls_p}^{CR}$.

The mode-transition energy consumption for two circuits can be written as:

$$E_{mt}^{ST-MTCMOS} = \left(c_{slp_{st}} + c_{G_{st}}\right) V_{DD}^2 \beta N_{clk}$$
$$E_{mt}^{CR-MTCMOS} = \left(c_{slp_{cr}} + \frac{1}{2}\left(c_{G_{cr}} + c_{P_{cr}}\right)\right) V_{DD}^2 \beta N_{clk} \quad (32)$$

where $c_{slp_{st}}$ and $c_{slp_{cr}}$ denote the total sleep transistor input capacitance, and $c_{G_{st}}$ denotes the total virtual ground capacitance in the ST-MTCMOS circuit while $c_{G_{cr}}$ and $c_{P_{cr}}$ denote the total virtual ground and virtual V_{DD} capacitances in the CR-MTCMOS circuit, respectively. Finally, β is the *mode transition frequency*, that is, the average number of mode transitions per clock cycle. We also define the *mode transition factor* in some time window T as the β value times the number of clock cycles in T.

From (29), the active mode energy consumption is the same for both circuits which means that charge-recycling technique does not have any influence on the active mode energy consumption; therefore, we do not consider the active mode energy consumption component of (28) for the remainder of the discussion. Therefore, (28) can be rewritten as:

$$E_{slp,mt}^{ST-MTCMOS} = E_{sleep}^{ST-MTCMOS} + E_{mt}^{ST-MTCMOS}$$
$$E_{slp,mt}^{CR-MTCMOS} = E_{sleep}^{CR-MTCMOS} + E_{mt}^{CR-MTCMOS} \quad (33)$$

Substituting (30), (31) and (32) into (33), and ignoring the terms related to the sleep transistors, we obtain:

$$E_{slp,mt}^{ST-MTCMOS} = \left(I_{ls_n}^{ST} V_{DD} (1-\alpha) T_{clk} + c_{G_{st}} V_{DD}^2 \beta\right) N_{clk}$$
$$E_{slp,mt}^{CR-MTCMOS} = \left(\left(I_{ls_n}^{CR} + I_{ls_p}^{CR} + I_{lcr}^{CR}\right) V_{DD} (1-\alpha) T_{clk} + \frac{1}{2}\left(c_{G_{cr}} + c_{P_{cr}}\right) V_{DD}^2 \beta\right) N_{clk} \quad (34)$$

TABLE V shows active mode and sleep mode leakage current and mode transition energy consumption values for a LVT inverter in the library. The table contains two sets of data corresponding to differently sized HVT NMOS sleep transistors, targeting in 10% and 15% active mode delay overheads for the inverter, respectively.

Fig. 13 shows the percentage of the total energy saving of CR-MTCMOS over ST-MTCMOS as a function of the mode-transition frequency for three different duty factor values for one of the ISCAS-85 benchmark circuits, 9sym. As we increase the mode-transition factor, the percentage of energy saving increases for each case. This is because the charge-recycling technique can save energy during mode transition only. As we increase the duty factor α, the total sleep time will decrease and the total saving will consequently increase. This can be seen in Fig. 13 by looking at energy saving plots for different activity factors. For large values of α (e.g., 0.9), and β, the sleep plus mode-transition energy saving ratio will be approximately equal to the mode-transition energy saving ratio (as was reported in TABLE II).

TABLE V
LEAKAGE CURRENT AND MODE TRANSITION COST OF AN INVERTER CELL IN THE LIBRARY FOR TWO DIFFERENT ACTIVE MODE DELAY PENALTIES.

Delay increase in Active Mode	CMOS (no power gating) (nA)	MTCMOS (footer power gating (nA)	Mode Transition Energy (FJ)
10%	260	0.291	3.52
15%	260	0.388	3.88

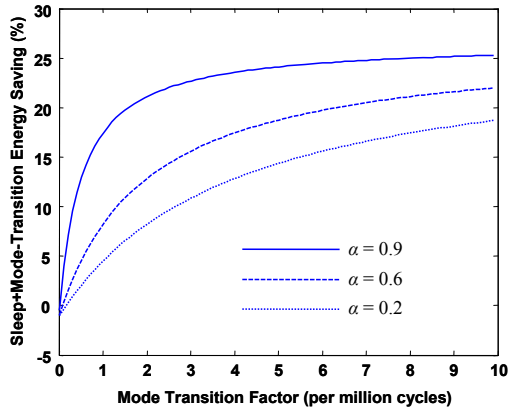

Fig. 13. Percentage of energy saving versus mode-transition factor for different duty factors for 9sym circuit, f_{clk}=4GHz.

VIII. CONCLUSION

In this paper we introduced the concept of charge-recycling (CR) in MTCMOS and SCCMOS circuits. We showed by applying charge recycling to MTCMOS or SCCMOS circuits, we can save up to 43% of the energy wasted during mode transition while maintaining the wake up time of the original MTCMOS or SCCMOS circuit. We also showed that by using the proposed technique, we can reduce the peak voltage and the settling time of the ground bounce occurred while waking up the circuit. Since the charge-recycling transistors are much smaller than the sleep transistors, the leakage increase due to the additional sneak path in the proposed technique is usually quite small.

REFERENCES

[1] Y. Taur, "CMOS design near the limit of scaling," *IBM J. Res. & Dev.,* Vol. 46 No. 2/3, pp. 213-222, Mar/May 2002.

[2] S. Mutoh, T. Douseki, Y. Matsuya, T. Aoki, S. Shigematsu, J. Yamada, "1-V Power Supply High-speed Digital Circuit Technology with Multi-threshold-Voltage CMOS," JSSC, vol. 30, no. 8, pp. 847-854, August 1995.

[3] S. Mutoh, S. Shigematsu, Y. Matsuya, H. Fukada, J. Yamada, "1V Multi-Threshold CMOS DSP with an Efficient Power Management Technique for Mobile Phone Application", *Proc. ISSCC*, pp. 168-169, 1996.

[4] J. Kao, S. Narendra, and A. Chandrakasan, "Subthreshold leakage modeling and reduction techniques," *Proc. Int'l Conference on Computer Aided Design*, pp. 141–148, Nov. 2002.

[5] J. Kao, A. Chandrakasan, and D. Antoniadis, "Transistor Sizing Issues and Tool for Multi Threshold CMOS Technology," *Proc. Design Automation Conference,* pp. 409-414, 1997.

[6] J. Kao, S. Narenda, and A. Chandrakasan, "MTCMOS hierarchical sizing based on mutual exclusive discharge patterns," *Proc. Design Automation Conference*, pp. 495-500, 1998.

[7] M. Anis, S. Areibi, M. Mahmoud, and M. Elmasry, "Dynamic and leakage power reduction in MTCMOS circuits using an automated efficient gate clustering technique," *Proc. Design Automation Conference*, pp. 480-485, 2002.

[8] E. Pakbaznia and M. Pedram, "Coarse-Grain MTCMOS Sleep Transistor Sizing Using Delay Budgeting," *Proc. Design Automation and Test in Europe,* pp. 385-390, 2008.

[9] C. Long and L. He, "Distributed sleep transistor network for power reduction," IEEE Trans. on VLSI Systems, Vol. 12, No. 9, pp. 937- 946, September 2004.

[10] A. Davoodi and A. Srivastava, "Wake-up protocols for controlling current surges in MTCMOS-based technology," *Proc. of the Asia South Pacific Design Automation Conference*, pp. 868-871, 2005.

[11] A. Abdollahi, F. Fallah, and M. Pedram, "An effective power mode transition technique in MTCMOS," *Proc. Design Automation Conference*, pp. 37-42, 2005.

[12] S. Kim, S.V. Kosonocky, Stephen, and D.R. Knebel, "Understanding and minimizing ground bounce during mode transition of power gating structures", *Proc. Int'l Symp. on Low Power Electronics and Design*, pp. 22-25, 2003.

[13] K. Agarwal, H. Deogun, D. Sylvester, K. Nowka, "Power Gating with Multiple Sleep Modes," *Proc. Int'l Symposium on Quality Electronic Design*, pp. 633 – 637, 2006.

[14] S. Kim, S.V. Kosonocky, D. R. Knebel, and K. Stawiasz, "Experimental measurement of a novel power gating structure with intermediate power saving mode," *Proc. Int'l Symp. on Low Power Electronics and Design*, pp. 20-25, 2004.

[15] E. Pakbaznia, F. Fallah and M. Pedram "Charge recycling in MTCMOS circuits: concept and analysis," *Proc. Design Automation Conference*, pp. 97-102, 2006.

[16] H. Kawaguchi, K. Nose and T. Sakurai, "A Super Cut-Off CMOS (SCCMOS) Scheme for 0.5-V Supply Voltage with Picoampere Stand-By Current", *IEEE Journal of Solid-State Circuits,* vol. 35, No. 10, pp. 1498-1501, October 2000.

[17] S. Mukhopadhyay and K. Roy, "Modeling and Estimation of Total Leakage Current in Nano-scaled CMOS Devices Considering the Effect of Parameter Variation", *Proc. Int'l Symposium on Low Power Electronics and Design*, pp. 172-175, 2003.

[18] K. S. Min and T. Sakurai, " Zigzag Super Cut-off CMOS (ZSCCMOS) Scheme with Self-Saturated Virtual Power Lines for Subthreshold-Leakage-Suppressed Sub-1-V-V_{DD} LSI's", *Proc. European Solid-State Circuits Conference,* pp. 679-682, 2002**.**

[19] A. Abdollahi, F. Fallah, and M. Pedram, "A robust power gating structure and power mode transition strategy for MTCMOS design," *IEEE Trans. on VLSI Systems,* Vol. 15, No., 1, Jan. 2007, pp. 80-89.

[20] P. Heydari and M. Pedram, "Ground bounce in digital VLSI circuits," *IEEE Trans. on VLSI systems*, pp. 180-193, Apr. 2003.

[21] A. Ramalingam, B. Zhang, A. Devgan and D. Pan, "Sleep transistor sizing using timing criticality and temporal currents," *Proc. Asia South Pacific Design Automation Conference,* pp. 1094-1097, 2005.