

Theory of circuit block switch-off

S. Henzler¹, J. Berthold², G. Georgakos², and D. Schmitt-Landsiedel¹

¹Lehrstuhl für Technische Elektronik, Technische Universität München, Theresienstrasse 90, D-80290 Munich, Germany

²Corporate Logic, Infineon Technologies AG, Balanstrasse 73, D-81541 Munich, Germany

Abstract. Switching-off unused circuit blocks is a promising approach to suppress static leakage currents in ultra deep sub-micron CMOS digital systems. Basic performance parameters of Circuit Block Switch-Off (CBSO) schemes are defined and their dependence on basic circuit parameters is estimated. Therefore the design trade-off between strong leakage suppression in idle mode and adequate dynamic performance in active mode can be supported by simple analytic investigations. Additionally, a guideline for the estimation of the minimum time for which a block deactivation is useful is derived.

1 Introduction

The static power dissipation of digital CMOS circuits in deep sub-micron technologies becomes one of the most challenging topics in digital system design. The exponentially increasing number of transistors per chip as well as the rising leakage currents per device make the standby power dissipation an appreciable portion of total power consumption. Various Circuit Block Switch-Off schemes (CBSO) have been proposed in order to reduce the static power dissipation of circuit blocks which are currently not used. Therefore, an additional transistor is used to separate the considered circuit block from either the v_{dd} or the v_{ss} potential. Thus the leakage current of this circuit block is reduced significantly during idle mode. The design of a Circuit Block Switch-Off scheme is driven by the trade-off between a strong leakage suppression, low area overhead due to the switch devices and a small performance degradation of the circuit in active mode. Additionally, the minimum time for which a block deactivation is useful has to be known in order to implement a reasonable power-down control logic. The fundamental approaches for answering these design questions are given in this paper. In the first part the static behaviour of a power

switched circuit block is considered. Following, the minimum power-down time is derived and finally a compact simulation approach to investigate the dynamic delay degradation in active mode is given.

2 Power-down state

After a circuit block has been cut-off, it reaches a certain steady state. This state has to be known in order to estimate the remaining leakage current. Usually the total transistor width in the logic block is much larger than the width of the switch device. Hence the potential of the virtual power-rail(s) has to change until the leakage current of the logic block is equal to the leakage of the switch. Figure 1 shows an equivalent circuit which models an n-block switched circuit. The cross coupled inverters are dimensioned very large in order to describe a large logic block with half of the nodes at logic high level and half of the nodes at logic low level. This simple equivalent circuit has a low transistor count, thus even large logic blocks can be modeled efficiently. The cross coupled inverter structure neglects interface effects between a power switched circuit block and the surrounding circuitry. Thus other models have to be used if these effects are to be investigated. In order to find the final state after switching-off a circuit block, the potential of the virtual power rail has been ramped up. The current through the switch device as well as the total current in the logic is given in Fig. 2. The steady state of the idle system is determined by the intersection of the two current voltage characteristics of the switch and the logic. Depending on the transistor widths and threshold voltages, the potential of the virtual power rail (v_{vss}) moves versus the non switched power rail (v_{dd}). In typical n-block switched circuits the v_{vss} potential as well as the potential of all internal signal nodes charges up to a value slightly below the v_{dd} potential. Thus all transistors operate in subthreshold region:

Correspondence to: S. Henzler
(stephan.henzler@ei.tum.de)

$$I_p = I_{0p} \exp\left(\frac{v_{dd} - v_1 + v_{tp}}{\eta v_T}\right) \left[1 - \exp\left(\frac{v_1 - v_{dd}}{v_T}\right)\right] \quad (1)$$

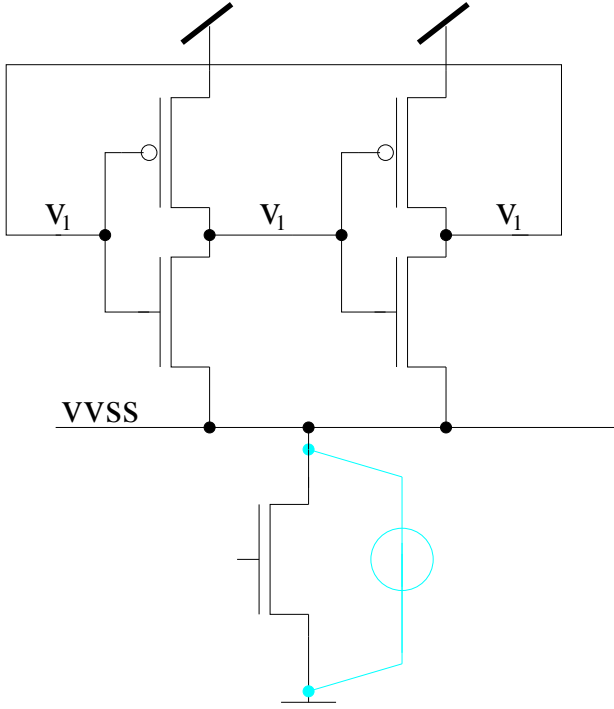


Fig. 1. Equivalent circuit to model a large power switched circuit block. Interface effects between the circuit block and the surrounding circuitry are neglected by this circuit. A voltage source between the virtual power rail and ground can be used to determine the current voltage characteristic of the circuit block and the switch device.

$$I_n = I_{0n} \exp\left(\frac{v_1 - vvss - v_{in}}{\eta v_T}\right) \left[1 - \exp\left(-\frac{v_1 - vvss}{v_T}\right)\right] \quad (2)$$

$$I_s = I_{0s} \exp\left(\frac{-v_{ts}}{\eta v_T}\right) \left[1 - \exp\left(-\frac{vvss}{v_T}\right)\right] \quad (3)$$

Hence the equivalent circuit of Fig. 3 can be used to derive an analytic expression for the steady state. The total logic width is assumed to be much larger than the width of the switch device. Thus Fig. 2 states that the potential of the virtual power rail is slightly below vdd. Therefore the $vvss$ dependency of the subthreshold current in the switch device can be neglected and the switch threshold voltage v_{ts} is approximated by the switch threshold voltage at maximum drain-to-source voltage $V_{ds}^{\text{switch}} = vdd$. As the deviation of the $vvss$ potential with respect to vdd is small, the potential V_1 of the signal nodes is also slightly below vdd . Thus a perturbation approach can be used to describe these voltages:

$$V_1 = vdd - \Delta V_1 \quad 0 < \Delta V_1 \ll vdd \quad (4)$$

$$vvss = V_1 - \Delta V_2 \quad 0 < \Delta V_2 \ll vdd \quad (5)$$

solving the linear equation system 1-3 results in

$$\Delta V_1^{1,2} = \frac{1}{2} \eta v_T \left(-1 \pm \sqrt{1 + \frac{4 I_{0s}}{\eta I_{0p}} \exp\left(-\frac{v_{ts} + v_{tp}}{\eta v_T}\right)} \right) \quad (6)$$

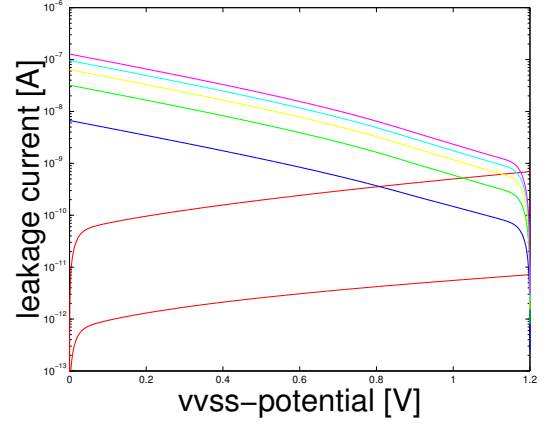


Fig. 2. Simulated leakage-voltage-characteristic of several logic circuits with different total transistor widths and subthreshold characteristic of two switch devices with low and high threshold voltage. The intersection of these curves determines the stationary state after block deactivation.

$$\Delta V_2^{1,2} = \frac{1}{2} \eta v_T \left(-1 \pm \sqrt{1 + \frac{4 I_{0s}}{\eta I_{0n}} \exp\left(\frac{v_{in} - v_{ts}}{\eta v_T}\right)} \right) \quad (7)$$

The upper solution is physical meaningful and for $I_{0s} \ll I_{0p}$, I_{0n} Eq. (6) and Eq. (7) can be approximated by

$$\Delta V_1 = v_T \frac{I_{0s}}{I_{0p}} \exp\left(-\frac{v_{ts} + v_{tp}}{\eta V_T}\right) \quad (8)$$

$$\Delta V_2 = v_T \frac{I_{0s}}{I_{0n}} \exp\left(\frac{v_{in} - v_{ts}}{\eta V_T}\right) \quad (9)$$

These equations state that the voltage difference of the signal- and $vvss$ -nodes decreases exponentially with increasing switch threshold voltage and linear with decreasing width ratio $\frac{W_s}{W_{\log}}$.

3 Leakage reduction ratio

The target of any CBSO scheme is the reduction of the leakage current of unused circuit blocks. The quality of this leakage reduction can be quantified by the leakage reduction ratio (LRR):

$$LRR := \frac{I_L^{\text{no switch}}}{I_L^{\text{switch off}}} \quad (10)$$

$I_L^{\text{no switch}}$ is the total leakage current of the considered circuit block if no cut-off switch is used and $I_L^{\text{switch off}}$ is the remaining leakage current of this circuit block if a cut-off switch is added and turned off. Without cut-off switch, the leakage of an arbitrary static CMOS circuit can be calculated according to

$$I_L^{\text{no switch}} = \sum_{\{LV\}} \left(W_p^{eff} I_p' e^{\frac{v_{ip}}{\eta v_T}} + W_n^{eff} J_n^n \right) + \sum_{\{HV\}} \left(W_n^{eff} I_n' e^{\frac{v_{in}}{\eta v_T}} + W_p^{eff} J_p^p \right) \quad (11)$$

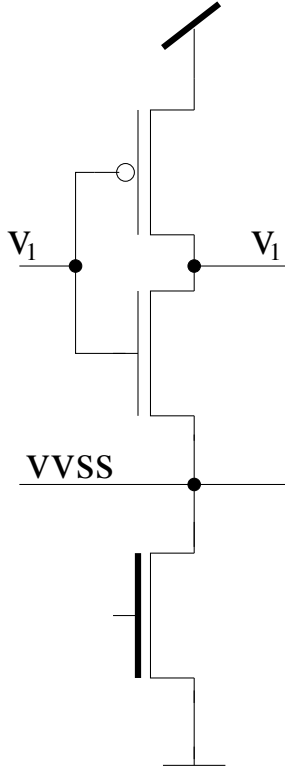


Fig. 3. Reduced equivalent circuit to determine the stationary cut-off state analytically.

$\{L_v\}$ describes the set of gates with a low output level and $\{H_v\}$ the disjoint set of gates with a high output level. I'_n and I'_p respectively describe the subthreshold leakage per transistor width and j_T^n and j_T^p refer to the gate tunneling currents per transistor width. (Assuming all transistors have the same channel length.) Averaged over all system states, the leakage current can be approximated by a common effective logic width W_L^{eff} :

$$I_L^{no\ switch} \approx \frac{1}{2} W_L^{eff} \left(I'_p e^{\frac{v_{TP}}{n v_T}} + I'_n e^{-\frac{v_{TN}}{n v_T}} + j_T^n + j_T^p \right) \quad (12)$$

In an n-block switching scheme, the virtual ground potential charges up to a voltage slightly below vdd . Thus the residual leakage current can be estimated by

$$I_L^{switch\ off} = W_s \left(I'_s e^{-\frac{v_{TS}}{n v_T}} + j_T^s \right) \quad (13)$$

Insertion of Eqs. (12) and (13) into (10) results in the LRR of the n-block switching scheme.

4 Power-down process

When a circuit block is not used it can be turned off by a cut-off switch. As the circuit block includes inner capacities, the

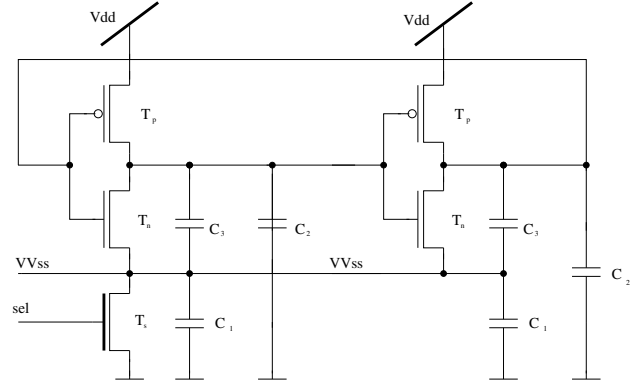


Fig. 4. Equivalent Circuit of a large logic block with internal capacities

leakage current does not collapse instantaneously. Instead, as shown in Fig. 5 there is a smooth transition between the active leakage and the residual leakage in idle mode. During this transition the v_{SS} rail as well as the internal signal nodes charge up to a potential usually slightly below vdd . Figure 4 shows the equivalent circuit introduced in Sect. 2 and its internal capacities. The capacities labeled by C_2 represent the capacity of each signal node versus the vdd -rail and the capacities labeled by C_3 represent the capacitance with respect to the virtual ground rail. Hence the collapsing supply current charges the C_1 and C_2 capacitances. The capacities C_3 however are discharged. Thus the intrinsic energy dissipation W_{intr} due to charging the inner capacities can be estimated by

$$\begin{aligned} W_{intr} &= vdd^2 \left(\sum C_1 + \sum_{\{L_v\}} C_2 - \sum_{\{H_v\}} C_3 \right) \\ &\approx vdd^2 \left(\sum C_1 + \frac{1}{2} \sum C_2 - \frac{1}{2} \sum C_3 \right) \\ &= \frac{1}{2} vdd^2 \sum (2C_1 + C_2 - C_3) \end{aligned} \quad (14)$$

When the circuit is turned on again, about half of the signal nodes keep a logic high level and hence the energy $W_{boot, cap} = \frac{1}{2} vdd^2 C_3$ is dissipated in order to charge up the C_3 capacitances associated to these nodes.

In order to estimate the minimum time T_{min} for which it is useful to turn off the circuit block, the collapse of the supply current after the switch is turned off has to be modeled properly. The model of this current has to be accurate but also simple enough to be used in analytic expressions. The leakage current after turning off the system can be decomposed according to

$$i_L(t > t_{off}) = i_L^{off} + i_{inertia}(t) \quad \lim_{t-t_{off} \rightarrow \infty} i_{inertia} \rightarrow 0 \quad (15)$$

into the stationarily remaining leakage current I_L^{off} and the so called inertial current $i_{inertia}$. The power dissipation due to the non instantaneously collapsing leakage current ist given

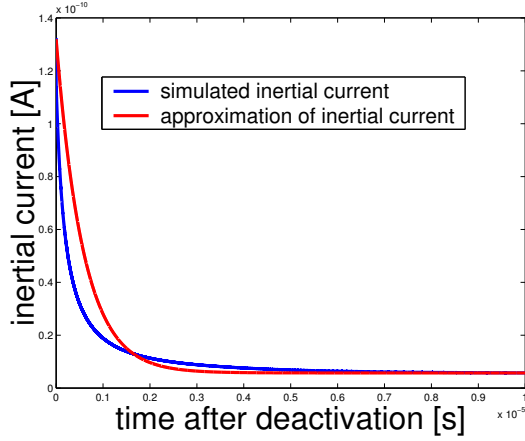


Fig. 5. Time dependence of inertial current and its approximation by an exponential function

by $\int_{t_{\text{off}}}^{\infty} i_{\text{inertia}}(t) dt$. If the leakage current was collapsing instantaneously the time which one has to wait in order to save the power penalty due to W_{intr} is given by

$$T_{\text{inertia}}^{\text{eff}} = \frac{\int_{t_{\text{off}}}^{\infty} i_{\text{inertia}}(t) dt}{(I_L^{\text{on}} - I_L^{\text{off}})} = \frac{W_{\text{intr}}}{vdd (I_L^{\text{on}} - I_L^{\text{off}})} \quad (16)$$

This time should be called effective inertial time. A simple approximation for the inertial current is the exponential function:

$$i_{\text{inertia}}^{\text{approx}} = (I_L^{\text{on}} - I_L^{\text{off}}) \exp\left(-\frac{t - t_{\text{off}}}{\tau}\right) \quad (17)$$

It is suggestive to determine the time constant τ such that the total energy dissipation due to the exponential function is equal to the real energy consumption. Equating this postulation results in $\tau = T_{\text{inertia}}^{\text{eff}}$.

5 Minimum powerdown time

Turning a circuit block off and on causes not only intrinsic energy dissipation W_{intr} but also extrinsic losses W_{extr} . These losses consist of losses in the switch device, in the switch driver and in the power-down-logic. Additionally there are losses during the power-up process (power-up glitches, $W_{\text{boot, cap}}$). Summing up all these energy dissipations results in the total energy penalty due to the power switching: $W_{\text{tot}} = W_{\text{intr}} + W_{\text{extr}}$.

The power saving due to the power switching at time $t > t_{\text{off}}$ can be calculated by

$$P_{\text{save}}(t) = (\Delta I_L - i_{\text{inertia}}(t)) vdd \quad \Delta I_L := I_L^{\text{on}} - I_L^{\text{off}} \quad (18)$$

If the circuit block is switched on again at time $t_{\text{off}} + T$ the energy penalty and the saved energy are given by

$$W_{\text{loss}} = vdd \int_{t_{\text{off}}}^{T+t_{\text{off}}} i_{\text{inertia}}(t) dt + W_{\text{extr}} \quad (19)$$

$$W_{\text{saved}} = vdd \int_{t_{\text{off}}}^{T+t_{\text{off}}} \Delta I_L dt \quad (20)$$

Equating these two expressions results in the so called energy equivalent powerdown time T_{eeq} which describes the minimum time for which it is useful to switch the circuit block off under power considerations:

$$\frac{W_{\text{extr}}}{vdd} = \Delta I_L \left[T_{\text{eeq}} + T_{\text{inertia}} \left(\exp\left(-\frac{T_{\text{eeq}}}{T_{\text{inertia}}}\right) - 1 \right) \right] \quad (21)$$

If the extrinsic losses are large, $T_{\text{eeq}} \gg T_{\text{inertia}}$ and therefore

$$T_{\text{eeq}} \approx \frac{W_{\text{extr}}}{vdd \Delta I_L} + T_{\text{inertia}} \quad (22)$$

Although this time is the main contributor to the minimum powerdown time T_{min} there are additional terms due to the signal propagation delay in the power-down-logic and the switch driver. Additionally, there is a certain settling time $T_{\text{power on}}^{CB}$ before the circuit block can be used after block activation. The minimum power-down time is the sum of all these contributors:

$$T_{\text{min}} = 2T_{\text{tot}}^{\text{switch}} + T_{\text{eeq}} + T_{\text{power on}}^{CB} \quad (23)$$

6 Dynamic behaviour of CBSO-systems

As the cut-off device in a CBSO-system has a finite on-resistance, the supply current through the logic block causes a voltage drop across the switch. Hence the effective supply voltage is reduced and the signal propagation delay in the circuit increases. The voltage-delay dependence of static CMOS gates can be derived by analytic MOSFET models like the alpha-power-law:

$$d \propto \frac{vdd}{(vdd - v_{th})^\alpha} \quad (24)$$

In this equation vdd is the nominal supply voltage, v_{th} is the threshold voltage and α is a technology dependent coefficient with values between 1 and 2. As the whole supply current of a power switched circuit flows across the cut-off switch, the delay degradation depends on the switching activity of all gates assigned to the switch. Therefore not only the time critical paths but the ensemble of all gates which switch their outputs while the signal propagates through the critical paths determine the overall delay of the circuit. Furthermore, the voltage-delay characteristic Eq. (24) is nonlinear and hence it is not a trivial task to determine the optimum dimensioning of the switch device. A large switch transistor reduces the delay degradation but suffers from large area consumption and poor leakage suppression. In order to estimate the delay degradation for a given circuit if a certain cut-off switch is added, the supply current profile of the circuit without cut-off switch is assumed to be known. The addition of a cut-off switch causes delay degradation but does not affect the logic function of the circuit. Hence the switching events are the same whether a cut-off switch is inserted or not and the current profiles of the two cases are similar.

The supply current of a static CMOS gate is given by the sum of the static leakage current, the short circuit current

during the output transition and the dynamic current which charges the load capacitance. As the effective supply voltage is reduced due to the voltage drop across the switch device, the charge which is necessary to charge the load capacity is reduced. The voltage swing of the virtual ground line in a n-block-switching scheme is small in order to keep delay degradation low. Hence the charge on the nonlinear load capacitance of an arbitrary gate can be estimated by linear Taylor approximation: $Q(v)=c(v)v$ describes the dependence of the charge on this capacity on the voltage. In the vicinity of v_{dd} , $c(v)$ is assumed to be a weak function of v . If the effective supply voltage is reduced by v_{vss} , the charge necessary to charge up the load capacity can be expressed by

$$\begin{aligned} Q(v_{dd} - v_{vss}) &= Q(v_{dd}) - \frac{dQ}{dv}(v_{dd})v_{vss} \\ &= c(v_{dd})v_{dd} - c(v_{dd})v_{vss} - \frac{dc}{dv}(v_{dd})v_{dd}v_{vss} \\ &\approx c(v_{dd})[v_{dd} - v_{vss}]. \end{aligned} \quad (25)$$

A similar approach can be done for the charge that is injected into the virtual power rail due to short circuit and leakage currents. Hence the supply current of a logic circuit can be decomposed according to

$$i(t) = \frac{dq}{dt} = \zeta(t)v_{dd}(t) \quad (26)$$

into an effective conductivity $\zeta(t)$ and the momentary supply voltage. If there is no cut-off switch the supply voltage is constant and ζ is given by $\zeta(t) = \frac{i(t)}{v_{dd}}$. Using this expression the charge which is injected into the ground node during the time interval $[t; t+dt]$ can be expressed by $dq = \zeta(t)dt v_{dd}$. Assuming that the supply current profile of a large circuit block is given and assuming further that this current profile results from many discrete switching events, this current profile can be partitioned into small current pulses of width $d\tau$. The distinct current pulses are shaped in the following way in order to estimate the current profile of the circuit with cut-off switch: The charge which is injected into the virtual ground node is reduced due to the reduced effective supply voltage. Hence ζ has to be multiplied by this effective supply voltage $v_{dd} - v_{vss}(t)$. Moreover the time interval during that this charge is injected is stretched with respect to the time interval if there was no switch device. Assume $i(\tau)$ is the current profile of a given circuit block without cut-off switch for a certain input transition. The dynamic of this circuit is described in the time variable τ . The current profile $i(t)$ of the circuit block with cut-off switch has to be determined by shaping the given profile. The time variable t describes the dynamic of the power switched circuit block. The charge which is injected into the virtual ground node during the time interval $[t; t + dt]$ is given by

$$dq(t) = \zeta(\tau(t)) \cdot [v_{dd} - v_{vss}(t)] d\tau(t) \quad (27)$$

The relation between the two infinitesimal time elements dt and $d\tau$ is given by the normalized supply voltage dependent delay degradation $\delta(v)$ of a CMOS gate:

$$dt = \delta(v_{dd} - v_{vss}(t)) d\tau \quad (28)$$

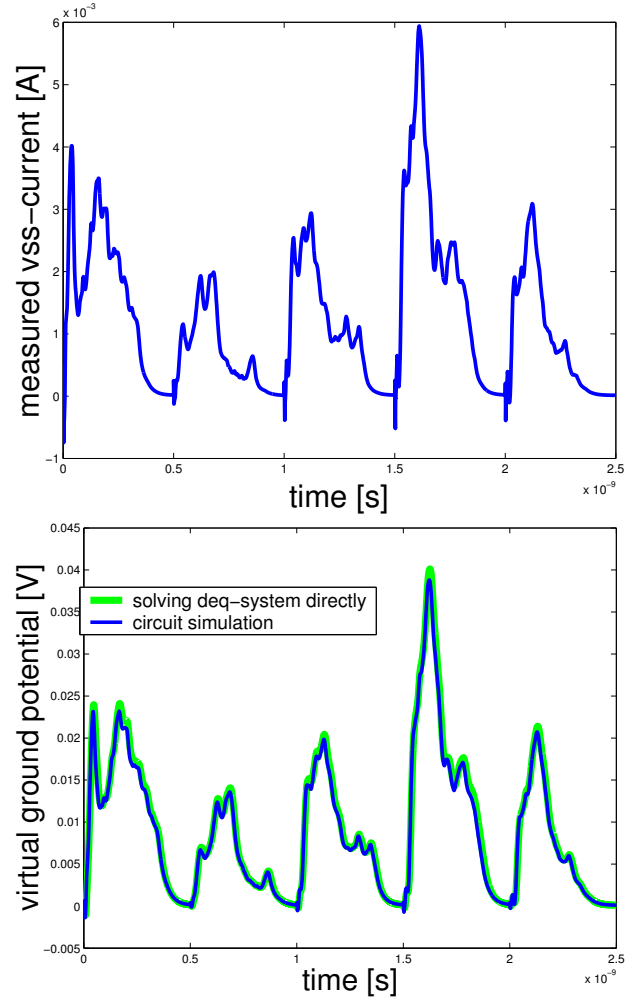


Fig. 6. Example for a current supply-current-profile and the resulting potential of the virtual ground line. The congruence of the curve determined by analog circuit simulation and the direct solving of the derived deq-system is very good.

Combining Eqs. (27) and (28) results in an expression for the current profile in the t domain which can be inserted into the Kirchhoff current equation of the virtual ground node:

$$\begin{aligned} \frac{dQ}{dv_{vss}}v_{vss}(t) + I(v_{vss}(t)) \\ = \zeta(\tau(t))\delta^{-1}(v_{dd} - v_{vss}(t))[v_{dd} - v_{vss}(t)] \end{aligned} \quad (29)$$

$$dt = \delta(v_{vss})d\tau \quad (30)$$

Solving this system of ordinary differential equations results in an estimation of the current profile of the circuit with current switch. The nonlinear functions $q_c = Q(v_{vss})$ and $i_R = I(v_{vss})$ describe the charge on the v_{vss} -capacitance and the current through the switch device respectively. These two functions have to be determined once by analog simulation and then many different current profiles can be evaluated by the fast current-profile-shaping algorithm. Figure 6 shows a given current profile of a 32-bit Kogge-Stone adder for five input transitions. The resulting v_{vss} -potential after the

insertion of a certain cut-off switch is given in the lower plot. The congruence of the voltage profile acquired by analog circuit simulation and by solving the deq-system respectively is very good.

Thus many different input transistions, switch sizes or switch device types can be examined by this fast compact simulation methodology.

7 Conclusion

In this paper we have described analytically the behaviour of a power switched circuit block during and after the block deactivation. The minimum power down time has been defined and the main contributors have been derived. Finally a fast analytic methodology to estimate the influence of a cut-off switch on the dynamic performance of a circuit has been proposed.

References

- Henzler, S., Koban, M., Berthold, J., Georgakos, G., and Schmitt-Landsiedel, D.: Design Aspects and Technological Scaling Limits of ZigZag Circuit Block Switch-Off Schemes, IFIP International Conference on Very Large Scale Integration of System-on-Chip, 2003.
- Mutoh, S., Douseki, T., Matsuya, Y., Aoki, T., Shigematsu, S., and Yamada, J.: 1-V Power Supply High-Speed Digital Circuit Technology with Multithreshold-Voltage CMOS, IEEE Journal of Solid-State Circuits, 30, 8, 847–854, 1995.
- Kawaguchi, H., Nose, K., and Sakurai, T.: A Super Cut-Off CMOS (SCCMOS) Scheme for 0.5-V Supply Voltage with Picoampere Stand-By Current, IEEE Journal of Solid-State Circuits, 35, 10, 1498–1501, 2000.
- Min, K., Kawaguchi, H., Sakurai, T.: Zigzag Super Cut-off CMOS (ZSCCMOS) Block Activation with Self-Adaptive Voltage Level Controller: An Alternative to Clock-Gating Scheme in Leakage Dominant Era, IEEE Solid State Circuits Conference, 2003.
- Meer, P. R. and Staveren, A.: New standby-current reduction technique for deep sub-micron VLSI CMOS circuits: Smart Series Switch, European Solid State Circuits Conference, 2002.