MIT Reversible Computing Project
Memo #M3

Voltage Scaling and Limits to Energy Efficiency for CMOS-based SCRL

WORKING DRAFT
Revision: 1.27

Michael P. Frank
MIT AI Lab, Rm. 747
545 Technology Sq.
Cambridge, MA 02139
http://www.ai.mit.edu/~mpf

Started Thu., Dec. 19, 1996.
Revision Date: 1997/01/10 22:33:42 GMT.
Formatted April 25, 1997.

A current version is available at
http://www.ai.mit.edu/~mpf/rc/memos/M03_scrllimits.html

Abstract

This document explains in detail a simple analysis of the maximal energy efficiency of the SCRL [2, 1] adiabatic circuit technique when implemented using ordinary MOS devices, and of how its efficiency scales with varying threshold voltages and temperatures. The analysis is somewhat crude, and needs further development, but as a preliminary result, we find that the minimum energy per operation in SCRL circuits *decreases* as threshold and supply voltages increase, in contrast to standard CMOS where the opposite relation holds.

1 Brief Overview of SCRL

This document is not intended to introduce the reader to SCRL. For an introduction, the reader should refer to references [2, 1]. However, for illustrative purposes, figure 1 shows an example of an SCRL gate, in this case a NAND. It can be seen that the gate consists

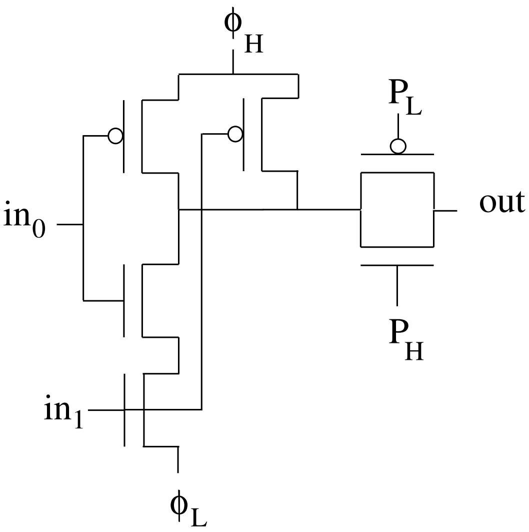

Figure 1: A typical SCRL gate: NAND.

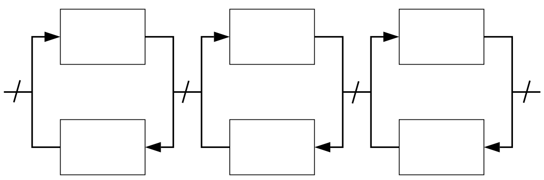

Figure 2: SCRL pipeline.

of a normal CMOS NAND structure, with the output fed through a transmission gate. In addition to its logical inputs, the gate requires four variable supply rails $\phi_{\mathrm{H}}, \phi_{\mathrm{L}}, \mathrm{P}_{\mathrm{H}}, \mathrm{P}_{\mathrm{L}}$.

The reader should keep in mind that the overall structure of a complete SCRL circuit consists of a number of such gates organized into a series of paired forward and reverse *stages*. Each stage may contain a number of SCRL gates in parallel, driving a bus of wires running between the stages. This large-scale structure is illustrated in figure 2, but the reader should consult [2, 1] for a detailed description.

Adjacent SCRL stages are driven by sup-

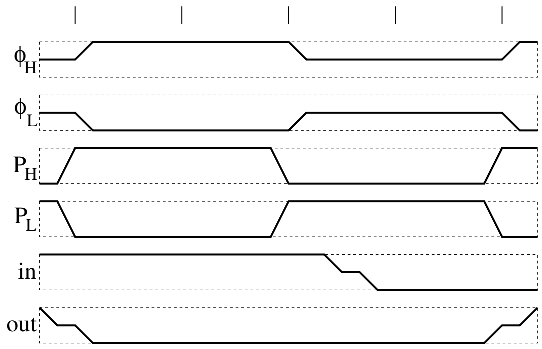

Figure 3: Timing diagram for an SCRL inverter.

ply rails whose timing is phase-shifted relative to each other, to cascade data through the pipeline while ensuring that two stages do not both try to drive the bus running between them at the same time. There exist a number of different timing disciplines for SCRL, having different numbers of distict phases. For reference, figure 3 shows an example of a timing diagram for an SCRL inverter in a 3-phase pipeline.

2 SCRL Switching Energy

2.1 Model

Switching energy is dissipated in an SCRL circuit whenever the voltages on some gate's power supply rails $\phi_{\mathrm{H}}, \phi_{\mathrm{L}}$ change. Energy is dissipated within the transistors of the gate's pullup/pulldown networks, and also in the transistors of the transmission gate attached to the gate's output. However, to simplify the analysis, we will lump together all the turned-on transistors within which dissipation occurs during a transition, and treat them as if they were a single transistor, as in figure 4.

We can consider a number of different cases for switching. A gate's output node voltage

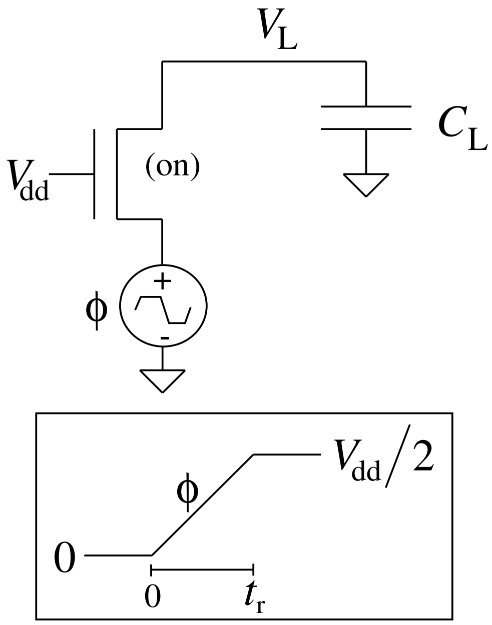

Figure 4: Circuit model for SCRL analysis.

may be switched either through the gate's pulldown network of NFETs or through its pullup network of PFETs. And the switching activity may either be to clear the output or to set the output. When an output node is cleared, its voltage goes from a valid level (0 or V_{dd}) to the neutral value $V_{dd}/2$; when it is set, its value goes from $V_{dd}/2$ to 0 or V_{dd}.

However, all these cases are symmetrically similar to each other with regards to how their energy dissipation scales with speed, threshold voltage, and temperature. Therefore, rather than analyzing them all separately, we will just consider one case: where the voltage V_L on the load capacitance C_L on the output node is charged up from 0V to $V_{dd}/2$, through a turned-on NFET which represents the gate's pull-down network and N pass transistor.

In our analysis, we will ignore any dissipation that occurs during switching in transistors along paths that do not actually connect all the way through to the gate's output. For example, we ignore energy dissipation that occurs when switching with the transmission gate turned off. As another example, referring back to the NAND gate in figure 1, we can see if in_0 is low and in_1 is high, then the transistor attached to in_1 will be turned on, and so there will be some small dissipation through it, even though it does not connect through to the output.

Later we will see that ignoring these dissipations is a simplification that is fairly well justified, because these dissipations involve driving relatively small capacitances. In adiabatic charging, there is a quadratic dependence of dissipation on the capacitance being driven. So the total dissipation we are ignoring should not be large, compared to the dissipation that we are including.

We assume that the PFETs and NFETs in the SCRL circuit have been sized so that their gain factors are equal, $k_n = k_p = k$ (matching the rise/fall delay times), and we assume that the PFET and NFET threshold voltages are also equal, $V_{t0n} = V_{t0p} = V_{t0}$, so that the analysis of the dissipation through the pulldown network comes out the same for the pullup network.

2.2 Analysis

To determine the energy dissipation of our model circuit (fig. 4), we would like to know the voltage on the load at each moment during the transision, $V_L(t)$, because this would tell us the instantaneous drain-to-source voltage $V_{DS}(t)$ across the transistor, which we could plug into the device's current-voltage relation to give us the instantaneous current $I(t)$, and thence the instantanous power, which we could integrate over time to find the total energy dis-

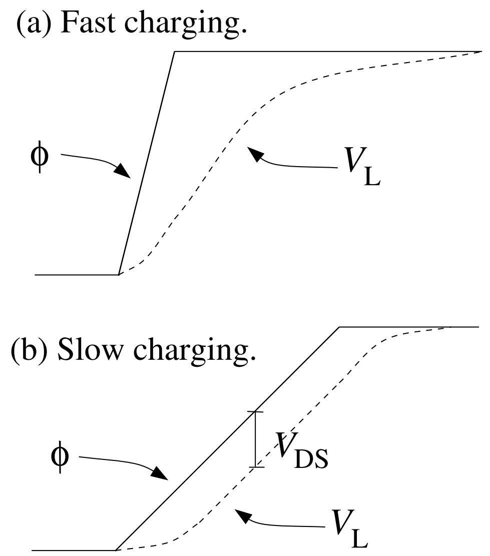

Figure 5: Power supply and output voltage curves for fast (a) and slow (b) charging.

sipation of the transition E_{tr}:

$$E_{\mathrm{tr}} = \int_{t=0}^{\infty} P(t)dt \tag{1}$$

$$= \int_{t=0}^{\infty} I(t)V_{\mathrm{DS}}(t)dt \tag{2}$$

Unfortunately, $V_{\mathrm{L}}(t)$ itself is determined by integrating the current $I(t)$ flowing into the load capacitance C_{L}, so that determining closed-form formulas for $I(t)$ and $V_{\mathrm{DS}}(t)$ requires solving a tricky differential equation, which we will not attempt here. Instead, we will approximate the energy dissipation by treating the limiting case where the supply rise time t_{r} is very large compared to the characteristic RC time constant of the circuit, where R is the effective resitance of the turned-on transistor. Cases where the rise time is about as small as RC will not be adequately addressed by the below analysis.

To understand this limiting case, refer to the diagrams in figure 5. Diagram (a) shows qualitatively what would happen if the supply rail were to rise very quickly compared to RC. Essentially the output voltage would rise at an exponentially-decaying rate and asymptotically approach the supply voltage, just as happens in a regular CMOS inverter whose input switches very quickly. The energy dissipation E_{fast} for this fast-switching case is well known to be

$$E_{\mathrm{fast}} = \frac{1}{2}C_{\mathrm{L}}(\Delta V)^2, \tag{3}$$

which in our case is (with $\Delta V = V_{\mathrm{dd}}/2$)

$$E_{\mathrm{fast}} = \frac{1}{8}C_{\mathrm{L}}V_{\mathrm{dd}}^2. \tag{4}$$

On the other hand, figure 5(b) shows what happens in the case which we will now analyze, where the supply rail rises very slowly. The output voltage V_{L} will initially rise slowly, but as the voltage drop V_{DS} across the transistor increases, the current $I(t)$ through the transistor will also rise, until an equilibrium is reached at which point V_{L} is rising at the same rate as the input voltage, but lagging behind it by a small amount $V_{\mathrm{DS}} = IR$. Then, when the input voltage stops rising, the output voltage will finish the approach to $V_{\mathrm{dd}}/2$ in asymptotic fashion, with an RC time constant.

We note that if the input rises slowly, V_{DS} is always small compared to $V_{\mathrm{dd}}/2$, and so $V_{\mathrm{L}}(t) \approx \phi(t)$. During the transition, $d\phi/dt$ is constant, and so the current $I = C_{\mathrm{L}}\frac{dV_{\mathrm{L}}}{dt}$ through the transistor will be approximately constant as well. I will be the quotient of the total charge $Q = C_{\mathrm{L}}V_{\mathrm{dd}}/2$ that is transfered to the load capacitance, divided by the supply rail rise time t_{r}, since that is the time during which almost all of this charge is transfered.

$$I(t) \approx I = Q/t_{\mathrm{r}} = \frac{C_{\mathrm{L}}V_{\mathrm{dd}}/2}{t_{\mathrm{r}}} \tag{5}$$

Now, armed with this constant current I, we can use the standard MOSFET triode-regime current-voltage formula to derive a closed form expression for V_{DS}. The reason we use the triode-regime rather than the saturation-regime formula is that turned-on transistors in SCRL are never in saturation.[1]

In the following, V_{GS} is the gate-to-source voltage, and V_{T} the threshold voltage. Everything except k (the transistor's gain factor) is here implicitly a function of t.

$$I = k\left((V_{\mathrm{GS}} - V_{\mathrm{T}})V_{\mathrm{DS}} - \frac{V_{\mathrm{DS}}^2}{2}\right) \tag{6}$$

Let's write $V_{\mathrm{GS}} - V_{\mathrm{T}}$ as V_{dr} (drive voltage) for conciseness.

$$I = k\left(V_{\mathrm{dr}}V_{\mathrm{DS}} - \frac{V_{\mathrm{DS}}^2}{2}\right) \tag{7}$$

We can easily solve this equation for V_{DS}, using the quadratic formula.

$$\frac{I}{k} = V_{\mathrm{dr}}V_{\mathrm{DS}} - \frac{V_{\mathrm{DS}}^2}{2} \tag{8}$$

$$\frac{1}{2}V_{\mathrm{DS}}^2 - V_{\mathrm{dr}}V_{\mathrm{DS}} + \frac{I}{k} = 0 \tag{9}$$

$$V_{\mathrm{DS}} = \frac{V_{\mathrm{dr}} \pm \sqrt{(-V_{\mathrm{dr}})^2 - 4\left(\frac{1}{2}\right)\left(\frac{I}{k}\right)}}{2\left(\frac{1}{2}\right)} \tag{10}$$

$$= V_{\mathrm{dr}} - \sqrt{V_{\mathrm{dr}}^2 - 2\frac{I}{k}} \tag{11}$$

Now, let us make a further simplification of eq. 11. We observe that our earlier approximation, that $I(t)$ was constant, assumed that t_{r} is large, and therefore that I is small (from eq. 5). With $I \ll kV_{\mathrm{dr}}^2$, this will allow us to approximate eq. 11 as follows. We observe that V_{DS} will be approximately linear in I for these small Is. V_{DS} will pass through 0 at $I = 0$, and the slope is given by dV_{DS}/dI:

$$\frac{dV_{\mathrm{DS}}}{dI} = \frac{d}{dI}\left(V_{\mathrm{dr}} - \sqrt{V_{\mathrm{dr}}^2 - 2\frac{I}{k}}\right) \tag{12}$$

$$= -\frac{1}{2}\left(V_{\mathrm{dr}}^2 - 2\frac{I}{k}\right)^{-\frac{1}{2}}\left(\frac{-2}{k}\right) \tag{13}$$

$$= \frac{1}{k\sqrt{V_{\mathrm{dr}}^2 - 2\frac{I}{k}}} \tag{14}$$

$$\approx \frac{1}{k\sqrt{V_{\mathrm{dr}}^2}} \quad \text{(for small } I\text{)} \tag{15}$$

$$= \frac{1}{kV_{\mathrm{dr}}}. \tag{16}$$

Given this slope, and the fact that $V_{\mathrm{DS}} = 0$ when $I = 0$, we can therefore simplify eq. 11 to the very concise form

$$V_{\mathrm{DS}} \approx I/kV_{\mathrm{dr}}. \tag{17}$$

Now, the drive voltage V_{dr} is itself actually time-dependent, because it is defined in terms of the gate-to-source voltage V_{GS}, and although the gate voltage is constant, the transistor source voltage changes linearly over time t_{r}, from 0 to $V_{\mathrm{dd}}/2$, following $\phi(t)$.

$$V_{\mathrm{dr}}(t) \equiv V_{\mathrm{GS}}(t) - V_{\mathrm{T}}(t) \tag{18}$$

$$= (V_{\mathrm{G}} - V_{\mathrm{S}}(t)) - V_{\mathrm{T}}(t) \tag{19}$$

$$= \left(V_{\mathrm{dd}} - \frac{V_{\mathrm{dd}}}{2}\frac{t}{t_{\mathrm{r}}}\right) - V_{\mathrm{T}}(t) \tag{20}$$

$$= (V_{\mathrm{dd}} - V_{\mathrm{T}}(t)) - \frac{V_{\mathrm{dd}}}{2}\frac{t}{t_{\mathrm{r}}} \tag{21}$$

Moreover, $V_{\mathrm{T}}(t)$ as well will vary along with the supply voltage, due to the changing body

[1] This formula may not be appropriate for turned-on transistors if V_{dd} is about as small as the thermal voltage $k_{\mathrm{B}}T/q$, since then even turned-on transistors may only be in moderate or weak inversion, and the current may scale exponentially with V_{GS} rather than according to the triode formula. This is one area where the current analysis needs refinement.

effect as the source voltage changes. For example, when the supply voltage is at $V_{\rm dd}/2$, $V_{\rm T}$ might be perhaps (as a roughly estimated typical value) 50% above the minimum value $V_{\rm T0}$ that it has when $\phi = 0V$. Using the correct formulas for $V_{\rm GS}$ and $V_{\rm T}$, the energy integral in equation 2 would still a bit too complicated to conveniently evaluate, although if we really cared to do it, we could.

But instead, let's just make the rough approximation that $V_{\rm dr}(t)$ is constant, and is equal to

$$V_{\rm dr} = \frac{3}{4}V_{\rm dd} - b_{\rm avg}V_{\rm T0}, \qquad (22)$$

taking the average of the initial ($V_{\rm dd}$) and final ($V_{\rm dd}/2$) values of $V_{\rm GS}(t)$, with an average body-effect factor $b_{\rm avg} = V_{\rm T}/V_{\rm T0}$ for a typical body-effected $V_{\rm T}$. The reason for expressing the body-effected threshold $V_{\rm T}$ as a multiple of $V_{\rm T0}$ is that it will later allow us to derive a very simple expression for the switching energy.

Now, with our approximate constant expressions for I (eq. 5) and $V_{\rm dr}$ (eq. 22), we can consider $V_{\rm DS}$ as given by eq. 17 to be roughly constant, which allows us finally to approximate the transition energy integral (eq. 2) and derive a fairly simple expression for $E_{\rm tr}$ in the slow-transition limiting case. We set the upper bound on the integral to be time $t_{\rm r}$ rather than ∞, in observance of the fact that in the slow-transition limit, most of the energy dissipation occurs by time $t_{\rm r}$.

$$\begin{aligned}
E_{\rm tr} &= \int_{t=0}^{t_{\rm r}} I(t)V_{\rm DS}(t)dt && (23)\\
&\approx IV_{\rm DS}t_{\rm r} && (24)\\
&= I\left(\frac{I}{kV_{\rm dr}}\right)t_{\rm r} && (25)\\
&= \frac{I^2 t_{\rm r}}{kV_{\rm dr}} && (26)\\
&= \frac{\left(\frac{C_{\rm L}V_{\rm dd}/2}{t_{\rm r}}\right)^2 t_{\rm r}}{kV_{\rm dr}} && (27)\\
&= \frac{C_{\rm L}^2 V_{\rm dd}^2}{4t_{\rm r}kV_{\rm dr}} && (28)
\end{aligned}$$

Now, we would like to take another simplifying step, by assuming that our maximum power supply voltage $V_{\rm dd}$ is being scaled proportionately to $V_{\rm T0}$, and is equal to

$$V_{\rm dd} = n_{\rm dd}V_{\rm T0} \qquad (29)$$

where $n_{\rm dd}$ indicates the scaling factor used for determining $V_{\rm dd}/V_{\rm T0}$. SCRL will not work properly if $V_{\rm dd}$ is too close to the threshold voltage $V_{\rm T0}$. A reasonable value for $n_{\rm dd}$ for SCRL might be 4. Anyway, given eqs. 29 and 22, we can substitute $V_{\rm dd}$ and $V_{\rm dr}$ in eq. 28 to re-express it in terms of a single voltage parameter $V_{\rm T0}$, the zero-bias threshold voltage:

$$\begin{aligned}
E_{\rm tr} &= \frac{C_{\rm L}^2(n_{\rm dd}V_{\rm T0})^2}{4t_{\rm r}k\left(\frac{3}{4}n_{\rm dd}V_{\rm T0} - b_{\rm avg}V_{\rm T0}\right)} && (30)\\
&= \frac{C_{\rm L}^2 n_{\rm dd}^2 V_{\rm T0}^2}{4t_{\rm r}k\left(\frac{3}{4}n_{\rm dd} - b_{\rm avg}\right)V_{\rm T0}} && (31)\\
&= \left(\frac{n_{\rm dd}^2}{3n_{\rm dd} - 4b_{\rm avg}}\right)\frac{C_{\rm L}^2 V_{\rm T0}}{t_{\rm r}k}, && (32)
\end{aligned}$$

and let us finally just make this a bit more concise by renaming the factor containing $n_{\rm dd}$ as just

$$c_{\rm dd} \equiv n_{\rm dd}^2/(3n_{\rm dd} - 4b_{\rm avg}). \qquad (33)$$

To illustrate what a typical value of $c_{\rm dd}$ might be, if $n_{\rm dd} = 4$ and $b_{\rm avg} = 1.25$ (*i.e.*, average body-effected threshold 25% above $V_{\rm T0}$), then $c_{\rm dd} \approx 1.45$.

Anyway, we can now write the transition energy formula (32) as just

$$\boxed{E_{\rm tr} = c_{\rm dd}\frac{C_{\rm L}^2 V_{\rm T0}}{t_{\rm r}k}.} \qquad (34)$$

There are a couple of very interesting things to note about equation 34, when compared to equations like eq. 3 that govern the dissipation in fast SCRL transitions or ordinary CMOS transitions.

The first thing is that the transition energy in eq. 34 scales in proportion to the square of the load capacitance, in contrast to traditional CMOS where the CV^2 dissipation scales only linearly with capacitance. The reason is that higher capacitance means higher currents through our transistors, and thus a larger voltage drop across them, in addition to greater charge to move across that drop. So in designing SCRL circuits we must be even more careful to get load capacitances small than we are in regular CMOS. Unless most of the capacitance is in the interconnects, minimum-sized transistors are favored. If most capacitance is in transistor gates and PN junctions, then increasing transistor widths increases energy dissipation roughly linearly (not quadratically, because k is scaled too). The flip side of this coin is that SCRL benefits greatly from improved process technologies that allow smaller, less capacitive transistors.

The other very interesting point is that given a constant n_{dd} ratio between supply and threshold voltages, and everything else but V_{T0} also constant, the switching energy of SCRL circuits *decreases only linearly with decreasing threshold voltage*, in contrast to the quadratic drop of traditional CMOS due to its CV^2 switching energy. Intuitively, the reason is because as voltages go down in SCRL, the effective on-resistance of our transistors increases, so the voltage drop across the transistors during transitions is increased, causing higher dissipation. In standard CMOS, the voltage drop across the transistors during switching is already as high as possible, and so making them more resistive doesn't affect the dissipation at all.

Equation 34 is interesting and useful on its own, because it allows us to predict the switching energy of SCRL circuits constructed in particular process technologies, and helps guide us in designing these circuits. But now, let's go a little further, and use eq. 34 as part of a more sophisticated analysis of SCRL energy dissipation that includes the effects of leakage.

3 Trading off Switching Energy and Leakage Energy

3.1 Adjusting Speed

One often-cited characteristic of the switching energy of adiabatic circuits, based on equations like eq. 34, is that it decreases linearly with increasing transition time t_{r}, leading to the conclusion that the energy per operation of SCRL circuits can be made arbitrarily small by just making the transition time larger. However, given current device technologies, this statement is somewhat misleading, because MOS transistors have a significant leakage power dissipation that is always present, and thus contributes a term to total energy per operation that increases linearly with increasing time per operation. This means that there is some speed at which the energy per operation of an SCRL circuit is minimized; at faster speeds, the switching energy dominates, and at lower speeds, the leakage energy dominates. In this section we derive a formula for the optimal rise time for minimizing total energy per operation.

Let us consider what happens to a signal wire in an SCRL circuit during a complete cycle, from the time it first holds one valid value to the time it first holds the next. During this time there will be two complete transitions on the wire: one from the old value to $V_{\mathrm{dd}}/2$, the other from $V_{\mathrm{dd}}/2$ to the new value. The total time for the complete cycle depends on the number of phases in the particular SCRL clocking discipline in question. A complete cycle of the 2-phase SCRL described by Younis [1] is the length of 18 transitions; 3-phase and 4-phase SCRL take 24 transitions, etc. These numbers are probably not minimal. Anyway, let n_t be the number of transitions per cycle; the total cycle time is then $T = n_\mathrm{t} t_\mathrm{r}$.

Now we can write down an expression for the total energy dissipation associated with this signal wire per complete cycle, including terms for both the transition energy and the leakage energy, where the leakage energy is expressed in terms of P_leak, the average leakage power associated with the signal wire:

$$\begin{aligned} E_\mathrm{tot} &= 2E_\mathrm{tr} + P_\mathrm{leak} T && (35) \\ &= 2c_\mathrm{dd}\frac{C_\mathrm{L}^2 V_\mathrm{T0}}{t_\mathrm{r} k} + P_\mathrm{leak} n_\mathrm{t} t_\mathrm{r}. && (36) \end{aligned}$$

where the multiplication by 2 comes from the above-mentioned fact that an SCRL wire undergoes two transitions per cycle.

We want to find the t_r that minimizes E_tot. First, let us collapse everything except t_r into coefficients a and b:

$$\begin{aligned} a &\equiv 2c_\mathrm{dd} C_\mathrm{L}^2 V_\mathrm{T0}/k && (37) \\ b &\equiv P_\mathrm{leak} n_\mathrm{t} && (38) \\ E_\mathrm{tot} &= \frac{a}{t_\mathrm{r}} + b t_\mathrm{r}. && (39) \end{aligned}$$

Figure 6 shows how the total energy in eq. 39 scales with t_r. We can see that at very high values of t_r, E_tot is high because of the high leakage energy, and at very low values of t_r, E_tot is high because of the high switching energy. In between, there is a point where the total energy is minimized.

We can find a formula for the t_r at this point; it's just where the derivative of eq. 39 equals zero, which turns out to be where the switching energy equals the leakage energy:

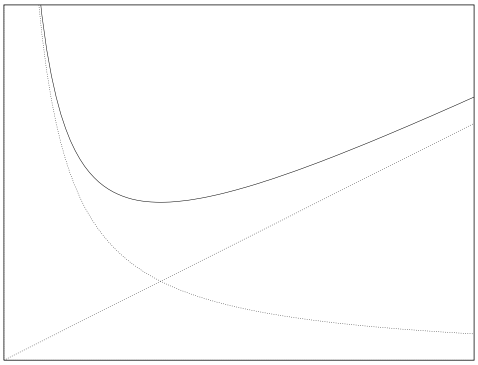

Figure 6: How total energy/operation scales with t_r in SCRL.

$$\frac{d}{dt_\mathrm{r}}\left(\frac{a}{t_\mathrm{r}} + b t_\mathrm{r}\right) = 0 \qquad (40)$$

$$-\frac{a}{t_\mathrm{r}^2} + b = 0 \qquad (41)$$

$$t_\mathrm{r} = \sqrt{\frac{a}{b}} = \sqrt{\frac{2c_\mathrm{dd} C_\mathrm{L}^2 V_\mathrm{T0}}{k P_\mathrm{leak} n_\mathrm{t}}} \qquad (42)$$

$$\boxed{t_\mathrm{r} = \left(\sqrt{\frac{2c_\mathrm{dd}}{n_\mathrm{t}}}\right) C_\mathrm{L} \sqrt{\frac{V_\mathrm{T0}}{k P_\mathrm{leak}}}} \qquad (43)$$

At this minimum-energy setting for t_r, the total energy dissipation is:

$$\begin{aligned} E_\mathrm{tot} &= \frac{a}{t_\mathrm{r}} + b t_\mathrm{r} && (44) \\ E_\mathrm{min} &= \frac{a}{\sqrt{a/b}} + b\sqrt{a/b} && (45) \\ &= \sqrt{\frac{a^2}{a/b}} + \sqrt{b^2 a/b} && (46) \\ &= \sqrt{ab} + \sqrt{ab} && (47) \\ &= 2\sqrt{ab} && (48) \\ &= 2\sqrt{\frac{2c_\mathrm{dd} C_\mathrm{L}^2 V_\mathrm{T0}}{k} P_\mathrm{leak} n_\mathrm{t}} && (49) \end{aligned}$$

$$\boxed{E_{\mathrm{min}} = \left(2\sqrt{2c_{\mathrm{dd}}n_{\mathrm{t}}}\right) C_{\mathrm{L}}\sqrt{\frac{V_{\mathrm{T0}}P_{\mathrm{leak}}}{k}}} \tag{50}$$

Looking at eq. 50, if we want the energy per operation of an SCRL circuit to be as low as possible, we will want to first minimize the wiring capacitance and other parasitic capacitances we need to drive. Then we'd want to maximize the gain factor k of our transistors. However, if we try to increase k by making the transistors wider, this also increases the capacitance, and the leakage power. So narrower transistors are favored.

Ideally we'd like to get a handle on minimum energy by adjusting the threshold voltage, so as to minimize the quantity $V_{\mathrm{T0}}P_{\mathrm{leak}}$ in eq. 50. But choosing the optimal V_{T0} is actually a bit tricky, since P_{leak} itself depends on V_{T0}, in a way which we will now analyze.

3.2 Adjusting Threshold

In a single transistor across which there is a voltage drop of $V_{\mathrm{DS}} = V_{\mathrm{dd}}$, which we will later see suffices to model the leakage through all the transistors attached to a given SCRL signal wire, the leakage power P_{leak} is given by

$$\begin{aligned} P_{\mathrm{leak}} &= I_{\mathrm{leak}}V_{\mathrm{dd}} && (51)\\ &= I_{\mathrm{leak}}n_{\mathrm{dd}}V_{\mathrm{T0}} && (52)\end{aligned}$$

and I_{leak} for transistors that are supposed to be "off" ($V_{\mathrm{GS}} \leq V_{\mathrm{T}}$) is given by a standard formula

$$I_{\mathrm{leak}} = I_0 e^{(V_{\mathrm{GS}}-V_{\mathrm{T}})/((1+\alpha)k_{\mathrm{B}}T/q)} \tag{53}$$

where I_0 denotes the leakage current when the transistor is just barely on the edge of being off (*i.e.*, when $V_{\mathrm{GS}} = V_{\mathrm{T}}$). k_{B} is Boltzmann's constant, T is the absolute temperature, q is the magnitude of the electron charge, and α is a technology-dependent constant fudge factor, which is ideally 0 but in practice is perhaps closer to 1. α is needed because real devices are found empirically to have a greater dependence of leakage on temperature than is predicted by the theoretical ideal.

Now, the leakage in SCRL circuits is not really continuous, but fluctuates during the SCRL cycle as different rails split and merge. In static versions of SCRL such as Younis's 3-phase clocking scheme, we can identify two types of leakage: (1) leakage through the middle of a logic gate across a voltage drop of V_{dd} when the gate's supply rails are split, and (2) leakage through a turned-off pass transistor across a voltage drop of $V_{\mathrm{dd}}/2$. All these leakages occur through off devices that have a V_{GS} of zero; other off devices with $V_{\mathrm{GS}} < 0$ have exponentially less leakage, and so we ignore them. During some transitions, there are also leakages across voltage drops smaller than $V_{\mathrm{dd}}/2$. Some of these happen when $V_{\mathrm{GS}} < 0$, and the others contribute small amounts to the total leakage power.

One may carry out a careful analysis of leakage based on the timing diagram of Younis's 3-phase clocking cycle. We will not relate the analysis in detail here. However, one finds that for each signal wire, there is leakage inside one of the logic gates that drive that wire during $\frac{22}{24}$ of each cycle, and leakage through a pass transistor for about $\frac{19}{24}$ of the cycle (this latter figure is adjusted to take into account the smaller voltage drops that occur during transitions).

Further, the I_0 for the leakage inside logic gates may be different than the I_0 for the leakage through the pass transistors, depending on how the devices are sized relative to each other, and also remembering that if a logic gate is not a simple inverter but rather contains several parallel paths, there may be leakage through all of the paths.

However, all of these factors can incorporated into our definition of the effective I_0 for the SCRL signal wire, as follows. Let $I_{0\mathrm{G}}$ be the effective I_0 in the pullup/pulldown networks of our logic gates (taking into account the widths of devices and number of parallel paths). Let I_{0P} be the I_0 through our pass transistors (taking into account their widths). Then we just define the effective I_0 for the single-transistor equivalent model of the SCRL signal wire's average leakage as

$$I_0 = I_{0G}\frac{22}{24} + I_{0P}\frac{1}{2}\cdot\frac{19}{24} \tag{54}$$

where the $\frac{1}{2}$ compensates for the fact that the leakage through the pass transistors involves a voltage drop of $V_{dd}/2$ rather than V_{dd}. This substitution is valid because the other factor in eq. 53 (the exponential) doesn't depend on the magnitude of the V_{DS} voltage drop or on which kind of leakage we are looking at, since $V_{GS} = 0$ for all the significant leakage.

We further note that almost all of the leakage takes place when $V_{GS} = 0$ and $V_{SB} = 0$, so that at these times $V_T = V_{T0}$, and we can substitute V_{T0} for V_T in eq. 53. Further, for conciseness let's define convenient notations for the thermal voltage $k_B T/q$ with and without the $(1+\alpha)$ fudge factor.

$$\phi_t \equiv k_B T/q \tag{55}$$

$$\phi_t' \equiv (1+\alpha)\phi_t \tag{56}$$

Now we can re-express the leakage current as just

$$I_{leak} \approx I_0 e^{-V_{T0}/\phi_t'}. \tag{57}$$

Although the above method for estimating I_{leak} was developed for the particular case of static 3-phase SCRL, it is fairly clear that the same approach could be carried out similarly for other SCRL clocking schemes as well, with appropriate modifications to eq. 54. Remember, however, that in 2-phase SCRL, nodes are not always being actively driven, and so high leakages can harm functionality as well as dissipating power; therefore the analysis later in this section will probably not be appropriate for dynamic 2-phase clocking.

Now that we've gotten I_{leak} expressed in terms of V_{T0}, let's merge eqs. 52 & 57 back into our expression for E_{min} (eq. 50):

$$\begin{aligned} E_{min} &= \left(2\sqrt{2c_{dd}n_t}\right) C_L \sqrt{\frac{V_{T0}P_{leak}}{k}} && (58)\\ &= \left(2\sqrt{2c_{dd}n_t}\right) C_L \cdot \\ &\quad \sqrt{\frac{V_{T0}(n_{dd}V_{T0})I_0 e^{-V_{T0}/\phi_t'}}{k}} && (59)\\ &= \left(2\sqrt{2c_{dd}n_t n_{dd}}\right)\cdot \\ &\quad C_L V_{T0}\left(\sqrt{\frac{I_0}{k}}\right) e^{-\frac{1}{2}V_{T0}/\phi_t'} && (60)\end{aligned}$$

To make this formula easier to work with, we'll express the factor involving the SCRL power and timing parameters n_{dd} and n_t as just s. Also, we note that since I_0 and k both scale roughly proportionally to transistor width, the voltage factor $\sqrt{I_0/k}$ is basically independent of transistor width. It scales up with increasing length however (because k scales down proportionally, but I_0 does not scale down as much), indicating that SCRL favors designing with minimum-length devices and small gate fan-ins. (Larger fan-ins yield a larger effective length.) In such designs, $\sqrt{I_0/k}$ can be thought of as a width-independent voltage v_c that is characteristic of the particular device technology being used. It can be interpreted as the drive voltage required to turn on a standard-length transistor strongly enough to conduct current at some fixed multiple of the transistor's zero-drive leakage current I_0.[2]

Given the above definitions, we can re-express the minimum energy as

$$s \equiv 2\sqrt{2c_{dd}n_t n_{dd}} \tag{61}$$

$$= 2\sqrt{\frac{2n_t n_{dd}^3}{3n_{dd} - 4b_{avg}}} \tag{62}$$

[2] Perhaps v_c is related to the drive voltage needed for strong inversion. I need to look into this sometime.

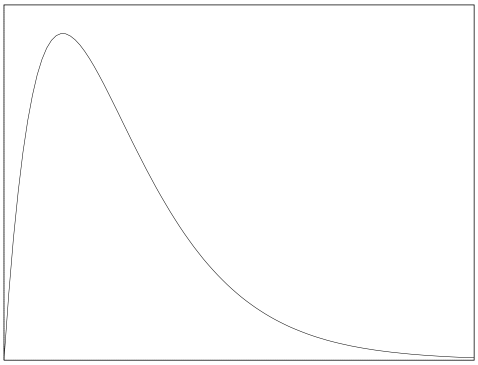

Figure 7: How minimum energy/operation scales with V_{T0} in SCRL.

$$v_{\mathrm{c}} \equiv \sqrt{I_0/k} \tag{63}$$

$$E_{\min} = sC_{\mathrm{L}} v_{\mathrm{c}} V_{\mathrm{T0}} e^{-\frac{1}{2} V_{\mathrm{T0}}/\phi'_{\mathrm{t}}} \tag{64}$$

Figure 7 shows qualitatively how $E_{\min}$ scales as V_{T0} is changed. Perhaps surprisingly, above a certain point, the minimum energy/op of SCRL actually decreases exponentially as the threshold voltage is increased! This contrasts with the situation in standard CMOS, where higher thresholds mean quadratically larger switching energy, determined by equations like eq. 3. The difference in SCRL is that higher thresholds mean exponentially smaller leakage power, which allows us to run at exponentially slower speeds and still not have leakage dominate the total energy, which thus allows exponentially less energy to be dissipated during our quasistatic charging at high thresholds.

The curve in fig. 7 also suggests that at very low thresholds, the energy/op can be made arbitrarily small as well. However, this part of the curve is probably not accurate. Appendix A shows that the maximum point on the curve occurs when $V_{\mathrm{T0}} = 2\phi'_{\mathrm{t}}$, twice the adjusted thermal voltage. At thresholds near or below the thermal voltage, a V_{dd} that is only a small fixed multiple of the threshold voltage will probably not be high enough to produce strong inversion, and the square-law equation (6) will probably not accurately represent the source-drain current of our transistors, upon which the above analysis was based. Moreover, at low thresholds, the high leakage power will call for a very short rise time from eq. 43; if the rise time is too short, it will not be large compared to the effective RC of our transistors, which will invalidate the assumptions upon which the analysis of section 2.2 was based.

Therefore, let's now make an effort to determine some expressions for the range of validity of the above analysis.

3.3 Range of Validity

...

In any case, assuming we are operating within a regime where the above analysis can be validly applied, the burning question now is, "Is SCRL's minimum energy/operation (as given by eq. 64) actually better than that achievable via voltage scaling in standard CMOS?" That is the question we will address in the next section.

4 Comparison vs. CMOS

Having successfully determined the minimum energy dissipation of SCRL circuits, we would now like to perform a similar analysis for low-threshold static-CMOS circuits, so that we can compare the results, and determine which circuit technique is actually better for achieving minimal energy dissipation, and how this relationship varies with temperature.

For reference, figure 8 shows an ordinary static CMOS inverter, which, for our purposes of determining minimum energy, will serve as our model for general CMOS logic gates. As with our SCRL analysis, we will assume

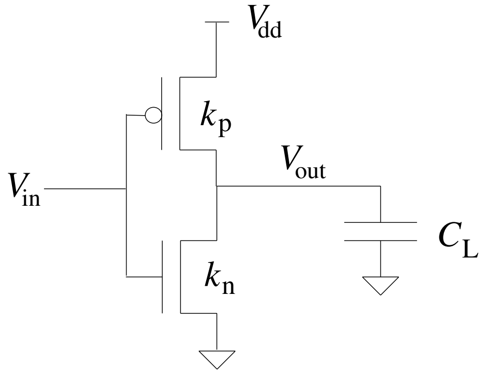

Figure 8: Ordinary CMOS inverter.

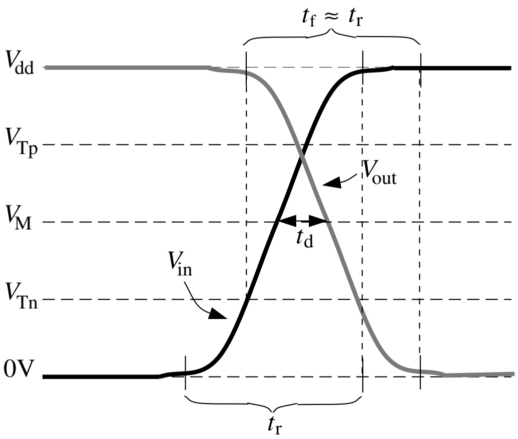

Figure 9: CMOS inverter dynamic behavior.

that the pullup and pulldown networks behave equivalently to single PMOS and NMOS transistors, with gain factors $k_{\mathrm{n}} = k_{\mathrm{p}} = k$ which are assumed to be made equal via appropriate sizing (matching the rise and fall times).

Figure 9 is a familiar illustration of the dynamic behavior of the CMOS inverter. The input voltage V_{in} rises (the falling case is similar) from 0V to V_{dd} in a time t_{r}. When V_{in} exceeds the threshold voltage V_{Tn} of the NFET, the NFET starts conducting significantly and pulling the output voltage V_{out} low; meanwhile the conductance of the PFET is decreasing,

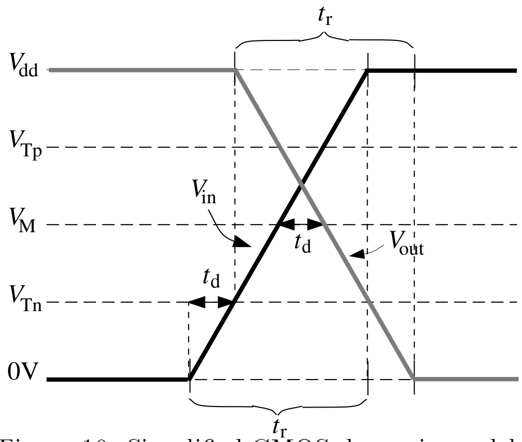

Figure 10: Simplified CMOS dynamic model.

and turns almost completely off when the input voltage rises above the PFET threshold. As a result, the output voltage V_{out} falls almost all the way to 0V, in a time t_{f}. The falling edge is delayed from the rising edge by an amount t_{d}. Since we presume that the inverter is driven by a similar gate, and that the gain factors of the N and P devices are the same, we suppose that $t_{\mathrm{f}} = t_{\mathrm{r}}$, and that the falling edge has the same shape as the rising edge. Therefore, t_{d} should be about equal to the time for the rising edge to exceed the V_{Tn} threshold, since that is the delay between the rising edge starting to rise and the falling edge starting to fall.

To simplify the analysis further, we approximate the rise/fall curves with straight lines, as shown in figure 10. These allow us to begin writing down some simple dynamic equations for the system. Taking $t = 0$ to be the time at the start of the rising edge, the input voltage during the rising edge is given by

$$V_{\mathrm{in}}(t) = \frac{t}{t_{\mathrm{r}}} V_{\mathrm{dd}}. \qquad (65)$$

Given this, we can derive an expression for the delay t_{d}, because as stated earlier, we can approximate the delay as the time for the input voltage to exceed the NFET threshold V_{Tn}. Since the source-to-bulk voltage for these devices is zero,[3] there is no body effect, and $V_{\mathrm{Tn}} = V_{\mathrm{T0}}$, our normal threshold voltage. We will find that $V_{\mathrm{in}}(t) = V_{\mathrm{T0}}$ at time

$$\begin{aligned} t_{\mathrm{d}} &= \frac{V_{\mathrm{T0}}}{V_{\mathrm{dd}}} t_{\mathrm{r}} && (66) \\ &= t_{\mathrm{r}}/n_{\mathrm{dd}} && (67) \end{aligned}$$

where n_{dd} is some standard ratio of $V_{\mathrm{dd}}/V_{\mathrm{T0}}$ such as 4, just like we had in section 2 (eq. 29).

The reason we are interested in t_{d} for our energy analysis is because it will determine how long we must wait before cycling new inputs into a pipeline of CMOS gates, which will determine the amount of energy that is dissipated each cycle by leakage.

Now, the transition time t_{r} of the output is just the time required for the (assumed constant) net output current I to charge up the load capacitance C_{L} from 0 to V_{dd}, *i.e.*, just

$$\begin{aligned} t_{\mathrm{r}} &= \frac{C_{\mathrm{L}} V_{\mathrm{dd}}}{I} && (68) \\ &= \frac{C_{\mathrm{L}} n_{\mathrm{dd}} V_{\mathrm{T0}}}{I} && (69) \end{aligned}$$

Estimating the net output current is tricky. It is not constant or linear during the transition since the NFET will be turning on, and the PFET off, according to a square law current equation such as eq. 6. And V_{DS} itself is not constant either, unlike in our SCRL analysis.

At this point we might throw up our hands and give up, but let's keep in mind that really all we care about in this document is to get a picture of the overall scaling laws, with respect to threshold/supply voltage, and so we do not care so much small constant factor errors in our formulas. So let's just come up with a simple formula for an average I that has the right overall order of magnitide and the right scaling properties.

[3] Actually, this is only true for the inverter; in more complex gates, V_{SB} of some transistors might not be 0 until some interior nodes have been charged to the right level; however if we imagine that interior node capacitance is small we might say this charging happens quickly enough to be ignored.

Without further justification, we announce that we will approximate the average net current as just half of the maximum saturation current through the NFET. This is very crude but suitable for our purposes. Some rough hand-estimates show this formula to be approximately right for the linear model in 10 where $n_{\mathrm{dd}} = 4$, but we are not too confident in the accuracy of this formula if n_{dd} is very different from 4.

$$I \approx I_{\mathrm{sat}}/2. \qquad (70)$$

The standard formula for the saturation current (ignoring short-channel effects) is

$$I_{\mathrm{sat}} = \frac{k}{2}(V_{\mathrm{GS}} - V_{\mathrm{T}})^2. \qquad (71)$$

As noted earlier, $V_{\mathrm{T}} = V_{\mathrm{T0}}$. The maximum value of V_{GS} is $V_{\mathrm{dd}} = n_{\mathrm{dd}} V_{\mathrm{T0}}$. So the maximum saturation current is

$$\begin{aligned} I_{\mathrm{sat}} &= \frac{k}{2}[(n_{\mathrm{dd}} - 1)V_{\mathrm{T0}}]^2 && (72) \\ &= \frac{k}{2}(n_{\mathrm{dd}} - 1)^2 V_{\mathrm{T0}}^2. && (73) \end{aligned}$$

So now we have an equation for the average I that shows how it scales with k and V_{T0}.

$$I = \left(\frac{(n_{\mathrm{dd}} - 1)^2}{4}\right) k V_{\mathrm{T0}}^2. \qquad (74)$$

It is left as an exercise for the reader to come up with a more accurate formula for I, but it is expected that the overall scaling behavior implied here (except perhaps with regards to n_{dd}) will be found to be essentially correct.

Now, let's plug eq. 74 back into eq. 69.

$$
\begin{aligned}
t_{\mathrm{r}} &= \frac{C_{\mathrm{L}} n_{\mathrm{dd}} V_{\mathrm{T0}}}{\frac{(n_{\mathrm{dd}}-1)^2}{4} k V_{\mathrm{T0}}^2} && (75)\\
&= \left(\frac{4 n_{\mathrm{dd}}}{(n_{\mathrm{dd}}-1)^2}\right) \frac{C_{\mathrm{L}}}{k V_{\mathrm{T0}}}. && (76)
\end{aligned}
$$

This says, fairly intuitively, that the rise time scales up proportionally to load capacitance, and scales down with the transistor gain factor. Interestingly, with fixed n_{dd}, the rise time scales up, making the circuit slower, as V_{T0} goes down. Contrast this with the situation in SCRL, where the switching rise time for minimum energy actually goes down as the threshold decreases. One might at first think that this implies that SCRL will be faster than CMOS at sufficiently low thresholds. But actually, we will see later that at the lowest feasible thresholds, minimum-energy SCRL is still slower than CMOS, and at high thresholds, minimum-energy SCRL is very much slower than CMOS.

Anyway, let's go ahead and plug eq. 76 back into eq. 67 to get our new formula for the delay.

$$
t_{\mathrm{d}} = \left(\frac{4}{(n_{\mathrm{dd}}-1)^2}\right) \frac{C_{\mathrm{L}}}{k V_{\mathrm{T0}}}. \qquad (77)
$$

Now it's time to begin analyzing the total energy dissipation for a CMOS circuit, as a sum of the switching energy, short-circuit energy, and leakage energy.

$$
E_{\mathrm{tot}} = E_{\mathrm{sw}} + E_{\mathrm{ss}} + E_{\mathrm{leak}}. \qquad (78)
$$

We saw the general equation for switching energy back in section 2.2, eq. 3. In normal CMOS, the voltage change during switching is $\Delta V = V_{\mathrm{dd}}$. We multiply this by an activity factor α_{sw} giving the probability of switching during a given operation to yield the expected switching energy per operation E_{sw}.

$$
\begin{aligned}
E_{\mathrm{sw}} &= \alpha_{\mathrm{sw}} \cdot \frac{1}{2} C_{\mathrm{L}} V_{\mathrm{dd}}^2 && (79)\\
&= \left(\frac{\alpha_{\mathrm{sw}} n_{\mathrm{dd}}^2}{2}\right) C_{\mathrm{L}} V_{\mathrm{T0}}^2. && (80)
\end{aligned}
$$

Now, let's look at the short circuit energy E_{ss}. Short-circuit energy is dissipated by the current that flows through the PFET and the NFET during the period of switching when both devices are turned on, which will happen if we assume that $n_{\mathrm{dd}} < 2$. Given our linear model in fig. 10, the length of this period is $(n_{\mathrm{dd}} - 2)/n_{\mathrm{dd}}$ of the total transition time t_{r}. The current during this transition we will crudely estimate as being the same I (half the saturation current) from eq. 74 that we used in expressing the net output current. (Hey, it's probably within a factor of two; we'll worry about making better approximations in a later version of this document.) The voltage drop across which this current falls is $V_{\mathrm{dd}} = n_{\mathrm{dd}} V_{\mathrm{T0}}$. Also, short-circuit dissipation only occurs if the input actually changes, so we will multiply all this by the activity factor α_{sw} to get the average short-circuit energy:

$$
E_{\mathrm{ss}} = \left(\alpha_{\mathrm{sw}} \frac{n_{\mathrm{dd}} - 2}{n_{\mathrm{dd}}}\right) t_{\mathrm{r}} I n_{\mathrm{dd}} V_{\mathrm{T0}}. \qquad (81)
$$

We can substitute our expressions for t_{r} (eq. 76) and I (eq. 74) to get

$$
\begin{aligned}
E_{\mathrm{ss}} = &\left(\alpha_{\mathrm{sw}} \frac{n_{\mathrm{dd}} - 2}{n_{\mathrm{dd}}}\right) \left(\frac{4 n_{\mathrm{dd}}}{(n_{\mathrm{dd}}-1)^2}\right) \frac{C_{\mathrm{L}}}{k V_{\mathrm{T0}}} \cdot \\
&\left[\frac{(n_{\mathrm{dd}}-1)^2}{4} k V_{\mathrm{T0}}^2\right] n_{\mathrm{dd}} V_{\mathrm{T0}}. \qquad (82)
\end{aligned}
$$

which simplifies to

$$
E_{\mathrm{ss}} = \left(\alpha_{\mathrm{sw}} \frac{n_{\mathrm{dd}} - 2}{n_{\mathrm{dd}}}\right) C_{\mathrm{L}} V_{\mathrm{T0}}^2, \qquad (83)
$$

showing that the short-circuit energy scales with CV^2 just like the switching energy from eq. 80. Thus the sum of switching and short-circuit energy can be conveniently expressed as a constant times CV^2.

$$E_{\rm sw}+E_{\rm ss} = \left[\alpha_{\rm sw}\left(\frac{n_{\rm dd}^2}{2}+\frac{n_{\rm dd}-2}{n_{\rm dd}}\right)\right]\cdot C_{\rm L}V_{\rm T0}^2. \quad (84)$$

Now let's analyze the leakage energy $E_{\rm leak}$, which is the third and final component of the total CMOS energy $E_{\rm tot}$ (eq. 78). Now, the leakage energy is going to depend on the rate at which we will be clocking the circuit, which we have not yet specified. As with SCRL, the leakage energy will be greater the longer the cycle is. In SCRL we saw that the trade-off between leakage and switching energy led to an expression for the optimal cycle length (eq. 43).

However, in CMOS, we just saw $E_{\rm sw}+E_{\rm ss}$ does not depend on the cycle time, so the total energy will be minimized when the leakage energy is minimized, or in other words when the cycle time is made as short as possible. Cycle time is partly an architectural issue; it can be decreased by using shorter pipeline stages or by carefully matching delays of parallel circuit paths.

However, to ensure correct functionality, the cycle time, being the time from the start of one input transistion to the next, will probably not be able to be much smaller than, say, the transition time plus the delay. If we take this as our cycle time, we are saying that our gates will wait until the output reaches a valid level before beginning the next input transition. It is probably being kind to CMOS to make the cycle time this short, so this will make our evaluation of SCRL's benefit a little conservative.

Anyway, let's write the cycle time as described and expand the expression using eqs. 76 and 77:

$$t_{\rm cyc} \approx t_{\rm r}+t_{\rm d} \quad (85)$$

$$= \left(\frac{4(n_{\rm dd}+1)}{(n_{\rm dd}-1)^2}\right)\frac{C_{\rm L}}{kV_{\rm T0}}. \quad (86)$$

Now, we can write the leakage energy as the leakage current times the supply voltage times the cycle time. I am not going to go through the detailed justification of the leakage current factor below, because it is about the same as we used for SCRL earlier. However, the reader should be aware that the I_0 used here will not in general be the same as the effective I_0 for the equivalent SCRL circuit which we saw in sec. 3.2, eq. 54.

$$\phi_{\rm t}' \equiv (1+\alpha)\phi_{\rm t} \quad (87)$$

$$E_{\rm leak} = \left(I_0 e^{-V_{\rm T0}/\phi_{\rm t}'}\right)(n_{\rm dd}V_{\rm T0})\cdot \left(\frac{4(n_{\rm dd}+1)}{(n_{\rm dd}-1)^2}\right)\frac{C_{\rm L}}{kV_{\rm T0}} \quad (88)$$

$$= \left(\frac{4n_{\rm dd}(n_{\rm dd}+1)}{(n_{\rm dd}-1)^2}\right)\frac{C_{\rm L}I_0}{k}e^{-V_{\rm T0}/\phi_{\rm t}'} \quad (89)$$

Now, we can finally write a complete formula for $E_{\rm tot}$.

$$c_0 \equiv \alpha_{\rm sw}\left(\frac{n_{\rm dd}^2}{2}+\frac{n_{\rm dd}-2}{n_{\rm dd}}\right) \quad (90)$$

$$c_1 \equiv \left(\frac{4n_{\rm dd}(n_{\rm dd}+1)}{(n_{\rm dd}-1)^2}\right)\frac{I_0}{k} \quad (91)$$

$$E_{\rm tot} = c_0C_{\rm L}V_{\rm T0}^2 + c_1C_{\rm L}e^{-V_{\rm T0}/\phi_{\rm t}'} \quad (92)$$

$$= C_{\rm L}\left(c_0V_{\rm T0}^2+c_1e^{-V_{\rm T0}/\phi_{\rm t}'}\right) \quad (93)$$

Now, let's set the derivative of this formula with respect to $E_{\rm T0}$ equal to zero, to find the point of minimal $E_{\rm T0}$. (We can see that this point is a minimum rather than a maximum by inspecting the graph of the formula.)

$$\frac{dE_{\rm tot}}{dV_{\rm T0}} = 0 \quad (94)$$

$$\frac{d\left(c_0V_{\rm T0}^2+c_1e^{-V_{\rm T0}/\phi_{\rm t}'}\right)}{dV_{\rm T0}} = 0 \quad (95)$$

$$2c_0V_{\rm T0}-\frac{c_1}{\phi_{\rm t}'}e^{-V_{\rm T0}/\phi_{\rm t}'} = 0 \quad (96)$$

$$
\begin{aligned}
2c_0 V_{\mathrm{T0}} &= \frac{c_1}{\phi'_{\mathrm{t}}} e^{-V_{\mathrm{T0}}/\phi'_{\mathrm{t}}} && (97)\\
r &\equiv V_{\mathrm{T0}}/\phi'_{\mathrm{t}} && (98)\\
2c_0 r \phi'_{\mathrm{t}} &= \frac{c_1}{\phi'_{\mathrm{t}}} e^{-r} && (99)\\
re^r &= \frac{c_1}{2c_0 \phi'^2_{\mathrm{t}}} && (100)\\
c_2 &\equiv \frac{c_1}{2c_0 \phi'^2_{\mathrm{t}}} && (101)\\
re^r &= c_2 && (102)
\end{aligned}
$$

We'd like to solve this for r, to get a formula for how the optimal ratio of threshold voltage to thermal voltage scales with temperature, but unfortunately, I know of no way to solve eq. 102 analytically. Instead, we must resort to numerical methods to solve r for particular values of c_2. To find typical values, let's plug in some numbers. With $n_{\mathrm{dd}} = 4$,

$$
\begin{aligned}
c_2 &= \frac{c_1}{2c_0 \phi'^2_{\mathrm{t}}} && (103)\\
&= \frac{\left(\frac{4n_{\mathrm{dd}}(n_{\mathrm{dd}}+1)}{(n_{\mathrm{dd}}-1)^2}\right)\frac{I_0}{k}}{2\alpha_{\mathrm{sw}}\left(\frac{n^2_{\mathrm{dd}}}{2}+\frac{n_{\mathrm{dd}}-2}{n_{\mathrm{dd}}}\right)\phi'^2_{\mathrm{t}}} && (104)\\
&= \frac{80}{153}\frac{I_0}{k}\frac{1}{\alpha_{\mathrm{sw}}\phi'^2_{\mathrm{t}}} && (105)
\end{aligned}
$$

Let's assume random input bits, so that the probability of switching α_{sw} is 0.5. Let's assume $\sqrt{I_0/k} = 70\mathrm{mV}$ since that was the value I calculated earlier for an inverter in the HP14 process. Let's also plug in a typical value of 1 for α, the fudge factor in the leakage current. So I get

$$c_2 \approx (35.8mV/\phi_{\mathrm{t}})^2. \qquad (106)$$

Interestingly, with the above choices of parameters, at room temperature c_2 is about 1.9, and numerical solution of eq. 102 gives $r = 0.83$, or $V_{\mathrm{T0}} \approx 21.5\mathrm{mV}$. Thus, at room temperature, the optimal threshold voltage for standard CMOS, subject to all the above crude approximations, assumptions, and parameter choices, is *slightly below the thermal voltage.*

T (°K)	$r = V_{\mathrm{T0}}/\phi'_{\mathrm{t}}$	V_{T0} (mV)
450	0.51	19.8
400	0.59	20.3
350	0.70	21.1
300	0.83	21.5
250	1.01	21.7
200	1.24	21.4
150	1.58	20.4
100	2.10	18.1
50	3.01	13.0
10	5.71	4.9
1	9.77	0.8

Table 1: How the optimal threshold voltage in CMOS scales with temperature.

Using the parameter choices above, table 1 shows how r scales among a range of temperatures. These numbers were derived through numerical solution of eq. 102. We can see that r increases somewhat as the temperature goes down, especially as absolute zero is approached, but for most of the reasonable temperature range, it stays fairly close to 1. So it is fair to say that the optimal threshold voltage for CMOS is close to the thermal voltage for a wide range of temperatures. Additionally, until the temperature goes below about 100°K, the optimal threshold voltage stays very close to 20 mV, peaking at about 21.7 mV at about 250°K.

Now, given these values of r which yield minimal total energy dissipation, we finally have a basis for comparing CMOS's minimum energy dissipation at a given temperature to SCRL's. But first, let's plug in our chosen values of the parameters $\alpha_{\mathrm{sw}} = 0.5$ and $n_{\mathrm{dd}} = 4$ to make eq. 93 more explicit:

$$E_{\mathrm{tot}} = \left(\frac{9}{2}r^2\phi'^2_{\mathrm{t}} + \frac{80}{9}\frac{I_0}{k}e^{-r}\right)C_{\mathrm{L}} \qquad (107)$$

[now do similar for SCRL]

5 Conclusion

Duh....

References

[1] S. G. Younis. *Asymptotically Zero Energy Computing Using Split-Level Charge Recovery Logic.* PhD thesis, MIT Artificial Intelligence Laboratory, 1994.

[2] S. G. Younis and T. F. Knight, Jr. Asymptotically zero energy split-level charge recovery logic. In *International Workshop on Low Power Design*, pages 177–182, 1994.

Appendix A: Worst-case V_{T0} for SCRL

In this section we derive and analyze the implications of a formula for the V_{T0} that leads to the maximum energy dissipation for SCRL circuits whose speed is adjusted for minimum energy at the given threshold. *I.e.*, we are finding the maximum point of eq. 64 illustrated in fig. 7.

From eq. 64 we can easily derive the value of V_{T0} that maximizes SCRL's minimum energy, by setting the derivative of E_{min} with respect to V_{T0} equal to zero.

$$\begin{aligned}
0 &= \frac{dE_{\mathrm{min}}}{dV_{\mathrm{T0}}} && (108)\\
&= \frac{d}{dV_{\mathrm{T0}}}\left(sC_{\mathrm{L}}v_{\mathrm{c}}V_{\mathrm{T0}}e^{-\frac{1}{2}V_{\mathrm{T0}}/\phi_{\mathrm{t}}'}\right) && (109)\\
&= sC_{\mathrm{L}}v_{\mathrm{c}}\frac{d}{dV_{\mathrm{T0}}}\left(V_{\mathrm{T0}}e^{-\frac{1}{2}V_{\mathrm{T0}}/\phi_{\mathrm{t}}'}\right) && (110)\\
&= \frac{d}{dV_{\mathrm{T0}}}\left(V_{\mathrm{T0}}e^{-\frac{1}{2}V_{\mathrm{T0}}/\phi_{\mathrm{t}}'}\right) && (111)\\
&= V_{\mathrm{T0}}\frac{de^{-\frac{1}{2}V_{\mathrm{T0}}/\phi_{\mathrm{t}}'}}{dV_{\mathrm{T0}}} + e^{-\frac{1}{2}V_{\mathrm{T0}}/\phi_{\mathrm{t}}'}\frac{dV_{\mathrm{T0}}}{dV_{\mathrm{T0}}} && (112)\\
&= V_{\mathrm{T0}}\frac{de^{-\frac{1}{2}V_{\mathrm{T0}}/\phi_{\mathrm{t}}'}}{dV_{\mathrm{T0}}} + e^{-\frac{1}{2}V_{\mathrm{T0}}/\phi_{\mathrm{t}}'} && (113)
\end{aligned}$$

$$\begin{aligned}
-e^{-\frac{1}{2}V_{\mathrm{T0}}/\phi_{\mathrm{t}}'} &= V_{\mathrm{T0}}\frac{de^{-\frac{1}{2}V_{\mathrm{T0}}/\phi_{\mathrm{t}}'}}{dV_{\mathrm{T0}}} && (114)\\
&= V_{\mathrm{T0}}e^{-\frac{1}{2}V_{\mathrm{T0}}/\phi_{\mathrm{t}}'}\frac{d(-\frac{1}{2}V_{\mathrm{T0}}/\phi_{\mathrm{t}}')}{dV_{\mathrm{T0}}} && (115)\\
&= V_{\mathrm{T0}}e^{-\frac{1}{2}V_{\mathrm{T0}}/\phi_{\mathrm{t}}'}(-1/2\phi_{\mathrm{t}}') && (116)
\end{aligned}$$

$$\begin{aligned}
V_{\mathrm{T0}} &= \frac{-e^{-\frac{1}{2}V_{\mathrm{T0}}/\phi_{\mathrm{t}}'}}{e^{-\frac{1}{2}V_{\mathrm{T0}}/\phi_{\mathrm{t}}'}(-1/2\phi_{\mathrm{t}}')} && (117)\\
&= 2\phi_{\mathrm{t}}' && (118)
\end{aligned}$$

$$\boxed{V_{\mathrm{T0}} = 2(1+\alpha)k_{\mathrm{B}}T/q} \qquad (119)$$

A fascinating result! The minimal energy dissipation of SCRL circuits (with respect to speed) is maximized (with respect to threshold voltage) *when the device threshold is equal to exactly twice the thermal voltage* $k_B T/q$, when corrected by the technology-dependent fudge factor $(1+\alpha)$. Note that this result does not depend on what kind of SCRL cycle we are using, or on the load capacitance C_L, or on how large the transistors are!

Just for fun, let's now see what we get when we plug eq. 119 back into eq. 64, and call the result E_{mm} since it is the "maximum minimum" energy, being maximized with respect to cycle time and minimized with respect to threshold voltage.

$$E_{min} = sC_L v_c V_{T0} e^{-\frac{1}{2}V_{T0}/\phi'_t} \quad (120)$$

$$E_{mm} = sC_L v_c (2\phi_t) e^{-1} \quad (121)$$

$$s' \equiv 2s/e \quad (122)$$

$$\boxed{E_{mm} = s' C_L v_c \phi_t} \quad (123)$$

As an example, I estimate the v_c for a simple inverter in the HP14 process to be around 70 mV[4], and s' for 2-phase SCRL with $n_{dd} = 4$ and a reasonable body-effect fudge factor to be about 10. At room temperature, ϕ_t is 26 mV. Therefore, the maximum possible energy dissipation of SCRL circuits, in a process like HP14 but with an worst-case threshold voltage, when the speed is adjusted for minimum energy, comes out to be 1.8 fJ/pF, *i.e.*, two femto-Joules per pico-Farad of load capacitance, per complete SCRL cycle of operation (from one input to the next). Note that this worst-case energy is proportional to temperature.

Let's go back now a little bit and figure out what the speed is when operating at the worst-case threshold. From earlier (see eqs. 43, 52, 57, and 119), we had

$$t_r = \left(\sqrt{\frac{2c_{dd}}{n_t}}\right) C_L \sqrt{\frac{V_{T0}}{kP_{leak}}} \quad (124)$$

$$P_{leak} = I_0 e^{-V_{T0}/((1+\alpha)k_B T/q)} n_{dd} V_{T0} \quad (125)$$

$$V_{T0} = 2(1+\alpha)k_B T/q. \quad (126)$$

If we expand the first occurrence of V_{T0} in eq. 125, we get

$$P_{leak} = I_0 e^{-2} n_{dd} V_{T0}, \quad (127)$$

and if we then plug this value back into eq. 124, we get

$$t_r = \left(\sqrt{\frac{2c_{dd}}{n_t}}\right) C_L \sqrt{\frac{e^2}{kI_0 n_{dd}}} \quad (128)$$

$$= \left(e\sqrt{\frac{2c_{dd}}{n_t n_{dd}}}\right) \frac{C_L}{\sqrt{kI_0}} \quad (129)$$

which comes out, for an HP14-like process, according to my calculations, as 18.6 ns per pico-Farad of load capacitance, for a line that is driven by a minimum-sized NFET. This seems pretty reasonable. For example, the 170fF load capacitance per signal I estimated for my Billiard Ball chip comes out to 3 ns per edge, which comes out to 57 ns/cycle or 17 MHz if I used 2-phase clocking (which I didn't, actually; it would be a little worse for 3-phase). This seems pretty good, considering the power comes out to only 0.12 μW per cell of my chip, or only 0.5 mW for a whole $(8\text{mm})^2$, 4000-cell chip—even at the V_{T0} yielding worst-case energy. If the chip were implementing a better-designed architecture that performed 1 MIPS/MHz, rather than the relatively inefficient billiard-ball model, this would be a MIPS/Watt ratio of about 40,000, which is 100 times better than the DEC StrongARM.

[When I did the above example calculations, my treatment of leakage currents was less so-

[4] I'm not sure I estimated I_0 correctly from the HP14 manual. Need to recheck this!

phisticated than it is now. Also 2-phase clocking is not really appropriate. Need to redo the calculations.]

It is interesting to note that the above speed result is independent of temperature, if the threshold voltage is adjusted as described previously to maximize energy at the given temperature.

Another interesting thing about eq. 129 is that since k and I_0 both scale proportionally to transistor width, the worst-energy speed will also, at least until the transistors start to dominate the load capacitance, at which point additional width doesn't help the speed any further, and starts hurting the energy. This is pretty intuitive. The upshot is that we want the transistors to contribute about as much capacitance as the other, unavoidable parts of the load capacitance, such as parasitics between wires.

Finally, let's derive the power when running at a worst-case threshold voltage but running at the best-case speed.

$$
\begin{aligned}
P &= \frac{E_{\mathrm{mm}}}{n_{\mathrm{t}} t_{\mathrm{r}}} && (130)\\
&= \frac{s' C_{\mathrm{L}} \phi_{\mathrm{t}} \sqrt{I_0/k}}{n_{\mathrm{t}} e \sqrt{2c_{\mathrm{dd}}/n_{\mathrm{t}} n_{\mathrm{dd}}}\, C_{\mathrm{L}}/\sqrt{kI_0}} && (131)\\
&= \frac{s' \phi_{\mathrm{t}} \sqrt{I_0/k}\sqrt{kI_0}}{n_{\mathrm{t}} e \sqrt{2c_{\mathrm{dd}}/n_{\mathrm{t}} n_{\mathrm{dd}}}} && (132)\\
&= \frac{(2s(1+\alpha)/e)\phi_{\mathrm{t}} I_0}{n_{\mathrm{t}} e \sqrt{2c_{\mathrm{dd}}/n_{\mathrm{t}} n_{\mathrm{dd}}}} && (133)\\
&= \frac{2 \cdot 2\sqrt{2c_{\mathrm{dd}} n_{\mathrm{t}} n_{\mathrm{dd}}}(1+\alpha)\phi_{\mathrm{t}} I_0}{n_{\mathrm{t}} e^2 \sqrt{2c_{\mathrm{dd}}/n_{\mathrm{t}} n_{\mathrm{dd}}}} && (134)\\
&= \frac{4 n_{\mathrm{t}} n_{\mathrm{dd}}(1+\alpha)\phi_{\mathrm{t}} I_0}{n_{\mathrm{t}} e^2} && (135)\\
P &= \frac{4 n_{\mathrm{dd}}(1+\alpha)}{e^2}\phi_{\mathrm{t}} I_0 && (136)
\end{aligned}
$$

Interestingly, the power per SCRL signal wire in this case is always just a small constant times the thermal voltage times the effective barely-off leakage current associated with the SCRL wire. What's the deep meaning of this? I don't know.

We should make a couple of qualifying remarks about the relevance of the results presented above. The first is that the entire analysis hinged on the assumption that t_{r} was large compared to the RC time constant of our gates. However, if the fundamental I_0 of our devices (which depends on the technology being used) happens to be very large compared to the gain factor k, then t_{r} for "worst-case threshold" given by this analysis will be very short, perhaps even so short that it is comparable to RC, in which case the whole analysis will be incorrect, and the transition time predicted may not even be sufficient for correct functionality. My rough hand-calculations indicate that in the HP14 technology, this is not the case, and the times are slow enough that the analysis is fairly accurate; however, when applying these results to other technologies, one should be careful to check for this possible problem.

Another very important qualification is that the worst-case threshold voltage stated in eq. 119 may not actually be reliably achievable in a given process technology, due to uncontrollable fluctuations in dopant concentrations. Therefore, functionality may be compromised if we attempt to use the very small threshold that is worst-case for a low operating temperature. However, even if this is the case, and allowable thresholds must be restricted to be above a certain level, low-temperature operation is still favored, as can be seen from eq. 64, which gives the minimum energy achievable at a given V_{T0} and operating temperature.