Chandrakasan & Brodersen
Kluwer

[Williams88] J. Williams, "Achieving Microamp Quiescent Current in Switching Regulators," *Linear Technology Design Notes*, DN 11(11), June 1988.

[Yachi91] T. Yachi, M. Mino, A. Tago, and K. Yanagisawa, "A new planar microtransformer for use in micro-switching-converters," *22nd Annual Power Electronics Specialists Conference*, pages 1003-1010, June 1991.

[Yachi92] T. Yachi, M. Mino, A. Tago, and K. Yanagisawa, "A new planar microtransformer for use in micro-switching-converters," *IEEE Trans. on Magnetics*, 28(4):1969-73, 1992.

[Yamaguchi93a] K. Yamaguchi, E. Sugawara, O. Nakajima, and H. Matsuki, "Load characteristics of a spiral coil type thin film microtransformer," *IEEE Trans. on Magnetics*, 29(6):3207-3209, 1993.

[Yamaguchi93b] K. Yamaguchi, S. Ohnuma, T. Imagawa, J. Toriu, H. Matsuki, and K. Murakami, "Characteristics of a thin film microtransformer with spiral coils," *IEEE Trans. on Magnetics*, 29(5):2232-2237, 1993.

[Yamaguchi93c] M. Yamaguchi, S. Arakawa, H. Ohzeki, Y. Hayashi, and K. Arai, "Characteristics and analysis for a thin film inductor with closed magnetic circuit structure," *IEEE Trans. on Magnetics*, 29(5), 1993.

[Yamasawa90] K. Yamasawa, K. Maruyama, I. Hirohama, and P. Birringer, "High-frequency operation of a planar-type microtransformer and its application to multilayered switching regulators," *IEEE Trans. on Magnetics*, 26(3):1204-1209, May 1990.

# 6

# Adiabatic Switching

*Lars Svensson*
*University of Southern California, Information Sciences Institute*

The fundamental cause of CMOS dynamic power dissipation is the organization of the energy transport in the circuit. Charging a node with a node capacitance, $C$, to a voltage, $V$, means storing a signal energy, $E_{sig} = CV^2/2$, on the node. In a level-restoring CMOS circuit with rail-to-rail swing[1], the signal charge, $Q = CV$, is drawn from the power supply at a constant voltage, $V$. Thus, as pointed out in Chapter 3, an energy $E_{inj} = QV = CV^2$ is injected into the circuit from the power supply. The injected energy is twice the signal energy; half of it is dissipated for the other half to be delivered to its destination. When the node is pulled low, the charge is drained from the node to ground, and the other half of the injected energy is consequently dissipated. Thus, all energy drawn from the supply is used only once before being discarded.

To reduce the dissipation, the designer must minimize the number of charging events, reduce the node capacitance, decrease the voltage swing, or apply some combination of these methods. Depending on the throughput requirements of the circuit and on the power-delay product, or switching energy, of the available logic style and MOS devices, the energy to carry out a certain operation in a certain time will vary, but the $CV^2$ barrier is impossible to circumvent as long as the overall approach of using energy only once is retained.

1. In this chapter, such logic styles are collectively called "conventional" styles.

Adiabatic switching allows the recycling of energy to reduce the total energy drawn from the power supply. This chapter discusses the limits to energy recycling achievable with a given silicon technology and a given switching speed and voltage swing. Logic circuits have to be reorganized, sometimes radically, to make signal energy recycling possible. The efficiency of some example circuits of different styles is analyzed in this chapter.

The term "adiabatic" is usually used to describe thermodynamic processes that exchange no heat with the environment. Here, the "process" is the transfer of electric charge between nodes in a circuit. Any energy dissipated as a result of the transfer is spread into the environment as heat. As is the typical use of the term in thermodynamics, fully adiabatic (and thus completely dissipation-less) operation is an ideal condition that may be asymptotically approached as the process is slowed down. A dissipation decrease with increased switching time is therefore the defining property of adiabatic switching. In many practical cases, the dissipation for a certain charge transfer is composed of an adiabatic component and a non-adiabatic one, and only the former will decrease when the process is slowed down. Such a partially-adiabatic system will not operate completely without dissipation, regardless of how slowly it is operated.

When adiabatic switching is compared to the voltage-scaling approach described in Chapter 4, both similarities and differences come to light:

- Both approaches address mainly the dynamic dissipation, so reducing the switching frequency and the driven capacitance is beneficial whether the driving circuit uses conventional or adiabatic switching.
- Adiabatic switching allows dissipation to be reduced further without changing the voltage swing, something which cannot be done with conventional circuits.
- In both cases, the dissipation improvement comes at the cost of slower switching speed, so the amount of hardware has to be increased for constant throughput, as described in Chapter 4.
- The "exchange rate" between dissipation and switching speed is different for the two approaches. Typically, voltage scaling initially gives more dissipation improvement for a certain speed reduction. Adiabatic charging is, however, more scalable and may be able to ultimately offer less dissipation at low speeds.



- Voltage scaling is conceptually simpler, requiring no radical re-thinking of the way the logic circuits are organized. There is no consensus yet on the best ways to construct logic systems that use adiabatic switching.

The choice between adiabatic switching and voltage scaling is determined by the circumstances. The conceptual simplicity and high payback of voltage scaling make it preferable in many cases. In situations when the voltage swing cannot be scaled down, adiabatic switching is the only known way to trade speed for power dissipation. Most commonly, these situations arise in interfaces between different parts of a system, where the logic operations performed are rarely more complex than simple power amplification or buffering. Examples include off-chip bus drivers, where the voltage swing is set by industry standards (such as 0–5 or 0–3.3 volts), and drivers for transducers and micromechanical devices, where the necessary swing is determined by some physical phenomenon employed by the device. Additionally, noise immunity may set the lower limit on practical signal energies, especially in environments rich in noise sources (such as commutators, electromagnetic actuators, cosmic rays, and radio transmitters).

When no "external" limits such as those mentioned in the previous paragraph must be taken into account, the voltage swing may be decreased further, but not indefinitely. The switching speed of conventional restoring CMOS gates quickly decreases when the supply voltage approaches the sum of the threshold voltages for the PMOS and NMOS devices, as described in Chapter 4. Gates in other styles of logic may be functional down to one threshold voltage. Beyond this, any decrease in the supply voltage will cause a disproportionally large increase in switching time. Further voltage scaling is practical only if the threshold voltages are scaled as well (cf. Section 4.5). The threshold voltages, in turn, are limited by the desire to keep a sufficient ratio of the on- and off-resistances of the devices. Estimates of the minimum practically useful threshold voltages vary from 50 mV to 200 mV [Mead] [Burr91] [Liu93]. It appears that a radical reorganization of the logic or new kinds of switching elements will be necessary to decrease the dissipation beyond what is possible with voltage scaling methods only. In this chapter, we address the former of these alternatives.

It must finally be noted that the bulk of the research into adiabatically-switching digital circuits has taken place in the last two years. The ideas and techniques described in this chapter are thus necessarily less mature than the voltage-scaling approach that is the main theme of this book. The field is, however, very active, and new results are published frequently.

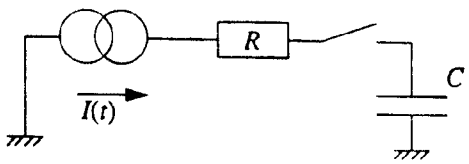



Figure 6.1: Current source charging a capacitance through a switch with a certain on-resistance.

## 6.1 Adiabatic Charging

We first consider a very simple circuit, illustrated in Figure 6.1. A time-dependent current source, $I(t)$, is used to charge a capacitance, $C$, through a switch with a resistance, $R$. This circuit is used as a model of a CMOS circuit with a certain output resistance driving a capacitive load. Note the similarity to the conventional model, where a load capacitance is charged through a switch resistance from a *voltage* source. As in the conventional case, the source models the power supply.

The load capacitance is discharged at time 0. The capacitance voltage as a function of time, $V_C(t)$, is then given by:

$$V_C(t) = \frac{1}{C}\int_0^t I(\theta)d\theta = \frac{1}{C}\bar{I}(t)t \tag{170}$$

$\bar{I}(t)$ is the average current from 0 to $t$:

$$\bar{I}(t) = \frac{C \cdot V_C(t)}{t} \tag{171}$$

The energy dissipation in $R$ from 0 to $t = T$ is then given by:

$$E_{diss} = R\int_0^T I(\theta)^2 d\theta \geq R\int_0^T \bar{I}(T)^2 d\theta = R\bar{I}(T)^2 T = \frac{RC}{T}CV_C(T)^2 \tag{172}$$

with equality when $I(t) \equiv \bar{I}(T)$, that is, when the current is constant. Other distributions of the current over time give higher dissipation. The influence of the current waveform may be quantified in a shape factor, $\xi$:



$$\xi = \frac{\int_0^T I(\theta)^2 d\theta}{\bar{I}(T)^2 T} \geq 1 \tag{173}$$

Then:

$$E_{diss} = \xi\frac{RC}{T}CV_C(T)^2 \tag{174}$$

This expression was introduced by Seitz and co-workers [Seitz85] (who apparently assumed constant-current charging and therefore omitted $\xi$).

The voltage on the output of the current generator in Figure 6.1, $V_I$, is connected to $I$ (and thereby to $V_C$) by the following differential equation:

$$V_I = RI + V_C = RC\frac{d}{dt}V_C + V_C \tag{175}$$

A constant $V_I$ (the "conventional" charging case described in Chapter 3) corresponds to an exponential current waveform with the time constant $RC$. A constant current corresponds to a linear voltage ramp.

We may now make several observations:

- Of all possible distributions of charging current over time, a constant current causes the least dissipation. In this sense, constant-current charging is the most efficient way to charge a capacitance through a resistance to a certain voltage in a certain time.
- When the exponential current waveform of the conventional case is substituted in Equation 172, $R$ and $T$ cancel out, and the dissipation is once again given by $E_{diss} = (1/2)\,CV_C(T)^2$.
- The dissipation is lower than for the conventional case if the current is constant and $T > 2RC$.
- The dissipation may be made *arbitrarily small* by further extending the charging time: $E_{diss} \sim T^{-1}$.
- A smaller $R$ also brings a lower dissipation. Again, this is in contrast to the conventional case, where dissipation depends only on the capacitance and the voltage swing.

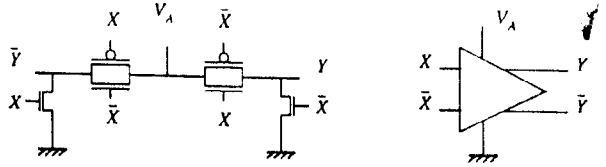



Figure 6.2: Adiabatic amplifier; circuit schematic and logic symbol.

We have efficiently moved energy from the power supply onto a load capacitance by using slow, constant-current charging. Reversing the current source will cause the energy to flow from the load capacitance back into the supply. Thus, in marked contrast to the conventional case, energy is not discarded (dissipated) after being used only once. The power supply must be designed to be able to retrieve the energy fed back to it; otherwise, only half of the potential benefit is realized.

It is clear from Equation 175 that to achieve the non-exponential current waveform desirable for low dissipation, it is necessary to provide a non-constant supply voltage. Adiabatic-switching circuits thus require non-standard power supplies with time-varying voltage and current. We will sometimes refer to these as "pulsed-power supplies." A low overall dissipation can be achieved only if the voltage and current waveforms of the supply both allow power-frugal logic circuits to be built and are possible to generate efficiently. A compromise between the conventional constant-voltage case and the ideal constant-current case may be reached by using sinusoidal waveforms, which can be efficiently generated with inductor-based pulsed-power supplies (as described in Section 6.5).

## 6.2 Adiabatic Amplification

We will now investigate a simple adiabatic-switching buffer or amplifier circuit. We will analyze its efficiency for different charging times and voltage swings and compare it to that of conventional buffers. The exposition is similar to that of Athas *et al.* [Athas94a]. The example amplifier circuit analyzed here has also been implemented; it should be noted that other circuit topologies are possible. The analyses for other cases will differ in the details, but the overall approach should be generally applicable.

Figure 6.2 shows a simple adiabatic amplifier for capacitive loads. It consists of two transmission gates (T-gates) and two NMOS clamps. The input is



dual-rail encoded, since both signal polarities are needed to control the T-gates. The output is also dual-rail encoded, which is required when other T-gates (such as those in other amplifiers) are to be controlled by the output signal. Also, dual-rail signalling keeps the capacitive load on the power supply data-independent, which simplifies the power supply design (cf. Section 6.5).

The operation of the amplifier is straightforward. First, the input is set to a valid value: $X$ and $\bar{X}$ cannot be equal. Next, the amplifier is "energized" by applying to $V_A$ a slow voltage ramp from 0 to $V_{dd}$. The load capacitance connected to one of the outputs is adiabatically charged to $V_{dd}$ through one of the T-gates, while the other output is clamped to ground. When charging is complete, the output signal pair is valid and can be used as an input to other circuits. Next, the amplifier is de-energized by ramping the voltage on $V_A$ back to 0. The signal energy that was stored on the load capacitance flows back into the power supply connected to $V_A$. The input signal pair must be stable throughout the process.

The dissipation caused by the operation of the amplifier is easy to determine. As shown in Section 6.1, the dissipation caused by adiabatically charging and discharging a capacitance depends on the on-resistance of the switch. The analysis assumed that the resistance was linear. MOS devices are highly non-linear, but a T-gate can be linearized to a first approximation by carefully selecting the device widths, as is shown next.

A T-gate is turned on with minimal on-resistance when the gate of the PMOS device is grounded and the gate of the NMOS device is connected to $V_{dd}$. Both devices are in the triode region when the voltage drop across the T-gate is small, which is the intended region for adiabatic circuits. Following Mead and Conway [Mead80], we model the conductance of the NMOS device, $G_n$, as:

$$G_n = \frac{C_n}{K_n}(V_{dd} - V_{cA} - V_{th}) \tag{176}$$

$$K_n = \frac{L^2}{\mu_n} \tag{177}$$

$V_{cA}$ is the average channel voltage, $V_{th}$ is the threshold voltage, and $C_n$ is the gate capacitance of the device. $K_n$ is a process constant that combines mobility,

$\mu_n$, and channel length, $L$ (the minimum channel length allowed in the process is used for all devices). Likewise, for the PMOS device, we get:

$$G_p = \frac{C_p}{K_p}(V_{ch} - V_{th}) \tag{178}$$

These equations do not take into account body effects nor the difference in threshold voltage between NMOS and PMOS devices. Accuracy is therefore limited to within a factor of two.

The sum of the two conductances may be simplified by selecting the widths of the two MOS devices such that $C_n/K_n = C_p/K_p$:

$$G_p + G_n = \frac{C_n}{K_n}(V_{dd} - V_{ch} - V_{th} + V_{ch} - V_{th}) = \frac{C_n}{K_n}(V_{dd} - 2V_{th}) \tag{179}$$

The on-resistance of the T-gate is then independent of the channel voltage:

$$R_{TG} = \frac{K_n}{C_n(V_{dd} - 2V_{th})} \tag{180}$$

Equations 179 and 180 are valid only when both devices are conducting, which is not the case when $V_{ch}$ is within one threshold voltage of either supply rail. At these extremes, the on-resistance will be less than the value given by Equation 180, so the formulation can be used as an estimate of the upper bound on the resistance. In practice, the body effect increases the resistance further.

The energy efficiency of the amplifier may now be analyzed with the help of Equations 174 and 180. The dissipation in the amplifier caused by first charging and then discharging one load capacitance, $C_L$, is:

$$E_{load} = 2\xi\frac{R_{TG}C_L}{T}C_L V_{dd}^2 = \frac{2\xi}{T}\frac{K_n}{C_n(V_{dd} - 2V_{th})}C_L^2 V_{dd}^2 \tag{181}$$

Additionally, parasitic effects such as diffusion capacitance of the T-gates and the clamp NMOS devices will contribute to the total load capacitance. Terms modeling these effects may be easily added to Equation 181. For simplicity, parasitics are neglected in this analysis.



Let $V_{dd} = m \cdot V_{th}$, and collect all process constants in one parameter, $\tau_n = K_n/V_{th}$. Then:

$$E_{load} = \left(2\xi\frac{\tau_n}{T}\frac{1}{(m-2)}\frac{C_L}{C_n}\right)C_L V_{dd}^2 = \left(2\xi\frac{\tau_n}{T}\frac{m^2}{(m-2)}\frac{C_L}{C_n}\right)C_L V_{th}^2 \tag{182}$$

We see that the dissipation decreases linearly with increasing $T$. Also, since $V_{dd} = m \cdot V_{th}$, the dissipation increases only linearly with the voltage swing, as opposed to the $V_{dd}^2$ dependence of the conventional case.

Energy is also dissipated to drive the input capacitances. When this energy is taken into account, the dependence on $T$ and $V_{dd}$ is affected, as is shown in the following sections.

### 6.2.1 One-Stage Adiabatic Buffer in Conventional System

We first analyze the case when the inputs are driven conventionally to $V_{dd}$, so that all the input energy, $E_{in}$, is dissipated. We assume that the total input capacitance of each of the input lines is proportional to the gate capacitance of the T-gate NMOS device, with a constant of proportionality, $\alpha$.[2] As an example, if no parasitic capacitances are considered and the gate capacitances of the clamp devices are neglected, $\alpha = (C_n + C_p)/C_n$. The dissipation caused by driving the inputs of one of the T-gates is given by:

$$E_{in} = \alpha C_n V_{dd}^2 \tag{183}$$

The total energy dissipated per cycle for one driven output is then:

$$E_{total} = E_{in} + E_{load} = \alpha C_n V_{dd}^2 + \left(2\xi\frac{\tau_n}{T}\frac{1}{(m-2)}\frac{C_L}{C_n}\right)C_L V_{dd}^2 \tag{184}$$

Equation 184 defines an important trade-off in adiabatic CMOS circuits with conventionally-driven inputs. When the channel width, and thereby the input capacitance, is increased, the energy dissipated in charging the load decreases, but the energy dissipated in charging the inputs increases proportionately. The

2. This $\alpha$ should not be confused with the activity factor of the previous chapters.

minimal energy dissipation is achieved by choosing the device sizes such that the two terms of Equation 184 are equal. The optimum NMOS gate capacitance is:

$$C_{n_{opt}} = \sqrt{\frac{2\xi}{\alpha}\frac{\tau_n}{T}\frac{1}{(m-2)}}C_L \tag{185}$$

Inserting Equation 185 into Equation 184 yields the following expression for the minimum dissipation:

$$E_{total_{min}} = \sqrt{8\xi\alpha\frac{\tau_n}{T}\frac{1}{(m-2)}}C_L V_{dd}^2 = \sqrt{8\xi\alpha\frac{\tau_n}{T}\frac{m^4}{(m-2)}}C_L V_{th}^2 \tag{186}$$

Equation 186 illustrates an important limitation for adiabatic charging with conventionally-driven control signals: because of the complete dissipation of the controlling gate energy, the total switching energy will only scale as $T^{-1/2}$, as opposed to the $T^{-1}$ scaling of Equation 174.

### 6.2.2 Two-Stage Adiabatic Buffer in Conventional System

A variation on the solution described above uses two amplifier stages, where a smaller adiabatic amplifier is used to drive the input signals of the final stage, as shown in Figure 6.3. Let $C_{n1}$ and $C_{n2}$ be the gate capacitances of the NMOS devices of the T-gates of the first and second amplifier, respectively. The input capacitance of the second amplifier, $\alpha C_{n2}$, then constitutes the load of the first amplifier. Allotting half of the total charging time to each stage, we get:

$$E_{total} = \alpha C_{n1} V_{dd}^2 + 4\xi\frac{\tau_n}{T}\frac{1}{(m-2)}\left(\frac{\alpha^2 C_{n2}^2}{C_{n1}} + \frac{C_L^2}{C_{n2}}\right)V_{dd}^2 \tag{187}$$

With both $C_{n1}$ and $C_{n2}$ as free variables, the minimum total energy dissipation is:

$$E_{total_{min}} = 8\left(\alpha\xi\frac{\tau_n}{T}\frac{1}{(m-2)}\right)^{3/4}C_L V_{dd}^2 = 8m^2\left(\alpha\xi\frac{\tau_n}{T}\frac{1}{(m-2)}\right)^{3/4}C_L V_{th}^2 \tag{188}$$



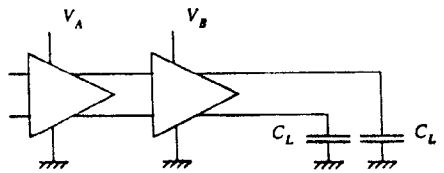



Figure 6.3: Two cascaded adiabatic amplifiers driving capacitive loads. $V_A$ and $V_B$ are two pulsed-power supply voltages, chosen to guarantee that the output of the first amplifier is valid when the second amplifier is energized. The input of the first amplifier is driven conventionally.

From Equation 186 to Equation 188, the energy scaling improves from $T^{-1/2}$ to $T^{-3/4}$. It can be shown that for a cascade of $n$ such amplifiers, the energy dissipation scales as:

$$E_{total_{min}} \sim T^{2^{-n}-1} \tag{189}$$

This result is not very useful in practice, since parasitics can no longer be neglected when the capacitances of the final gate-drive MOS devices become comparable to that of the load. Also, if the switching is to be performed within a constant time interval, the ramp time for each of the steps must decrease at least linearly with the number of cascaded amplifier stages.

### 6.2.3 Fully-Adiabatic System

As shown above, conventionally driven control signals cause a sub-linear dependence of energy on time. True $T^{-1}$ scaling can be achieved only if the system uses adiabatic charging throughout. It is not straightforward to build such a system (some issues are outlined in Section 6.3), but the resulting dissipation for charging a certain capacitance is easy to calculate. Assume a cascaded arrangement like that shown in Figure 6.3, but with the inputs of the first amplifier adiabatically driven, its device dimensions fixed, and a charging time $T$ allowable for each amplifier. Then, the dissipation for driving the load and the inputs of the second amplifier is:

$$E_{total_{min}} = \left(2\xi\frac{\tau_n}{T}\frac{1}{(m-2)}\right)\left(\frac{\alpha^2 C_{n2}^2}{C_{n1}} + \frac{C_L^2}{C_{n2}}\right)V_{dd}^2 \quad (190)$$

The $C_{n2}$ that minimizes the dissipation is independent of $m$ and of $\tau_n/T$, and the dissipation is linear in $T$:

$$C_{n2_{opt}} = \sqrt[3]{\frac{C_L^2 C_{n1}}{2}} \quad (191)$$

$$E_{total} = \left(2\xi\frac{\tau_n}{T}\frac{1}{(m-2)}\left(\sqrt[3]{2} + \frac{1}{\sqrt[3]{4}}\right)\sqrt[3]{\frac{C_L}{C_{n1}}}\right)C_L V_{dd}^2 \quad (192)$$

### 6.2.4 Comparison with Conventional Buffer

We can now compare the dissipations of the adiabatic buffers with that of a conventional one, for different speeds and voltage swings. To drive the input and load of a conventional buffer with mobility-scaled device sizes (giving the same total input capacitance as for the T-gate) causes a dissipation given by:

$$E_{total} = (\alpha C_n + C_L)\, V_{dd}^2 \quad (193)$$

By rewriting Equation 90 of Chapter 3, the charging time of a conventional buffer may be approximated as:

$$T = \frac{2K_n V_{dd}}{(V_{dd} - V_{th})^2}\cdot\frac{C_L}{C_n} = 2\tau_n\frac{m}{(m-1)^2}\cdot\frac{C_L}{C_n} \quad (194)$$

Thus:

$$E_{total} = \left(1 + 2\alpha\frac{\tau_n}{T}\frac{m}{(m-1)^2}\right)C_L V_{th}^2 \quad (195)$$

We now plot Equations 186, 188, 192, and 195 as functions of the switching speed for different values of $m$. These plots are shown in Figures 6.4 and 6.5. For each plot, the dissipation is normalized to $C_L V_{dd}^2$. The time scale is normalized to the delay time of a conventional buffer when $C_L = 20C_n$, as given by Equation 194. The variants of the adiabatic buffer improve with increased supply voltage and charging time, whereas the conventional buffer gets little dissipation benefit from increased switching time. With a lower supply voltage, the



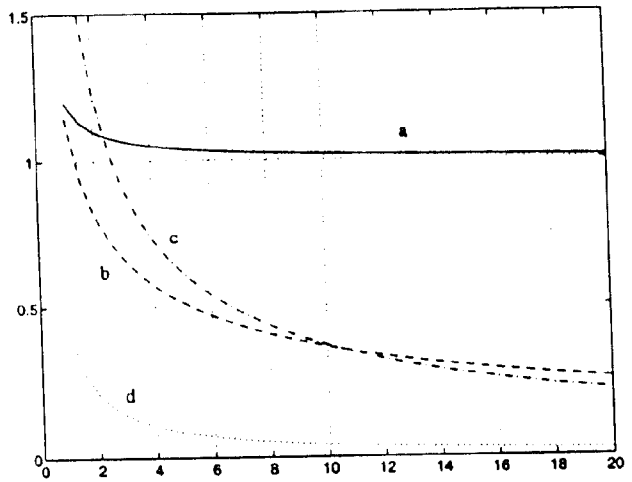



Figure 6.4: Comparison of dissipation as a function of the transition time for a) conventional driver, b) one-stage adiabatic driver, c) two-stage adiabatic driver, and d) driver in fully adiabatic system. The voltage swing is fixed to three times the threshold voltage ($m = 3$). Dissipation is normalized to $C_L V_{dd}^2$. Transition time is normalized to that of a conventional driver when the load capacitance is 20 times the gate capacitance of the NMOS device of the driver.

cross-over point where the adiabatic approach is preferable occurs at a longer switching time.

Care must be taken when these diagrams are interpreted. First, in all cases the dissipation for driving *one* line conventionally is compared to that of driving *one* line adiabatically; but the design of the adiabatic amplifier presupposes dual-rail signalling. Second, the dissipation in the power supply (which generates the voltage ramps) is not taken into account; it is non-trivial to minimize this dissipation. Third, the use of T-gates in the buffers causes a lower limit on the usable supply voltage: if $m < 2$, there will be a region in which neither device is conducting, and thus the resistance (and, according to the model, the dissipation)

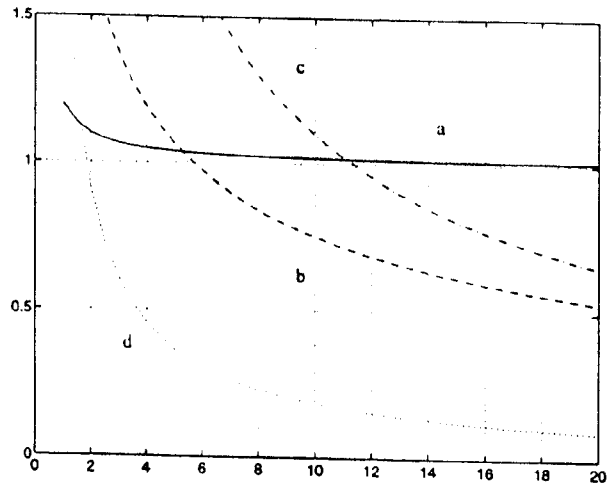



Figure 6.5: As Figure 6.4, but for a lower supply voltage: $m = 2.1$. With the lower supply voltage, the transition time has to be increased more over that of the conventional buffer for the adiabatic approach to yield lower dissipation.

goes to infinity. The latter limitation may be avoided with other types of adiabatic circuits (cf. Section 6.3).

To investigate how well the theory reflects reality, an adiabatic amplifier for stand-alone use as a line driver for the address bus of a memory board was designed, implemented, and tested [Athas94b]. The most significant deviation from the theoretical results was caused by dissipation in the power supply (which was similar to those described in Section 6.2.5). The non-unity shape factor and additional parasitics further reduced the dissipation gain from a theoretical factor of 20 (according to Equation 181) to a factor of 6.3 (when the dissipation in the power supply is included).



### 6.2.5 Supply Voltage Influence

In Section 6.2.4, we compared the dissipations of the different adiabatic amplifiers with that of a conventional buffer at several fixed supply voltages. If we are free to choose the voltage swing for minimum dissipation, we might try to scale down signal energies (by decreasing $V_{dd}$) as far as the application permits and then apply adiabatic-charging techniques to reduce switching energies further. Such an approach does, however, not necessarily lead to minimal dissipation for adiabatic circuits, since the reduced gate drive at lower supply voltages increases the on-resistance of the switches. For a T-gate-based driver, the resistance increase counteracts the dissipation gain caused by the reduced signal energies, leading to a process-dependent optimal voltage swing which yields the lowest overall dissipation.

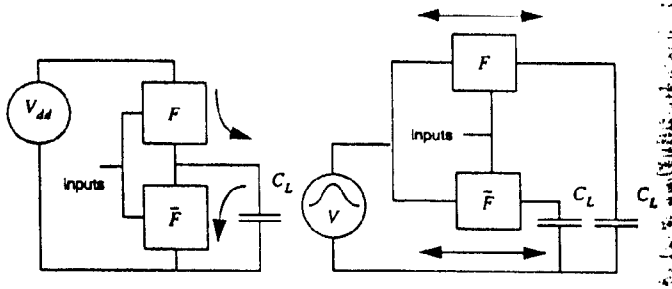
It is easy to derive the optimal voltage swing for the T-gate-based buffers analyzed above. For the one-stage buffer, we find the voltage by rewriting Equation 186:

$$E_{total_{min}} = \sqrt{8\xi_3\alpha\frac{\tau_n}{T}C_L V_{th}^2}\frac{m^2}{\sqrt{m-2}} \tag{196}$$

The energy is minimized when $m = 8/3$, and thus $V_{dd} = (8/3)\,V_{th}$. Similar derivations for the two-stage buffer (described by Equation 188) and the fully adiabatic system (described by Equation 192) give the values $m = 16/5$ and $m = 4$, respectively. Note that these optima are based on a simple model for the on-resistance of the T-gate. The body effect and the regions where only one FET device is conducting are not taken into account. SPICE simulations and more detailed analyses (that take body effects, subthreshold regions, and waveform shapes into account) all show that the minima are shallow and that the values given above yield close to the minimal dissipation.

## 6.3 Adiabatic Logic Gates

A straightforward extension of the adiabatic amplifier of Section 6.2 allows the implementation of arbitrary combinational logic functions that use adiabatic




(a) (b)

Figure 6.6: (a) A conventional CMOS logic gate. (b) A corresponding adiabatic gate may be constructed by reorganizing the switch networks as shown, replacing the PMOS and NMOS devices of the conventional gate with T-gates. The DC supply voltage must be replaced by a pulsed-power source. The arrows indicate the charge flow when the load capacitances are charged and discharged.

switching. As shown in Figure 6.6, a conventional logic gate may be transformed into an adiabatically-switched counterpart by replacing each of the PMOS and NMOS devices in the pull-up and pull-down networks with T-gates, and by using the expanded pull-up network to drive the true output and the expanded pull-down network for the complementary output. Both networks in the transformed circuit are used both to charge and discharge the output capacitances. The DC $V_{dd}$ source of the original circuit must be replaced by a pulsed-power source with varying voltage to allow adiabatic operation. The optimal sizes of the T-gates of the networks can be determined from the equations of Section 6.2. The resulting combinational circuits can switch fully adiabatically, but their device count is twice that of the conventional counterpart, and their input capacitance is relatively large because of the T-gates.

As with conventional CMOS logic, it is desirable for reasons of performance and complexity management to partition a large block of logic into smaller ones and then compose them to implement the original larger function. However, non-adiabatic flow of energy will occur if values are allowed to ripple through a



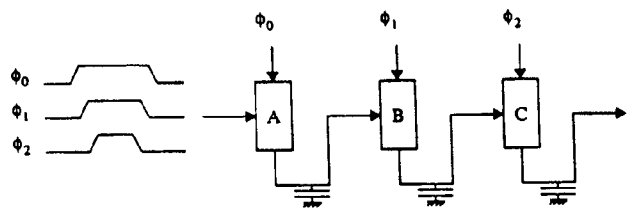



Figure 6.7: Retractile cascade of adiabatic-switching logic gates. The supply voltages are ramped up in order, energizing the gates and charging the load capacitances. The voltages must then be ramped down in reverse order to ensure that the energy stored on the capacitances is recovered. The pulsed-power voltage source waveforms are shown on the left.

chain of cascaded adiabatic logic gates. Adiabatic operation is possible only if the inputs of every gate are held stable while the gate is energized.

Hall observed that a cascade of initially de-energized circuits may be adiabatically energized in succession [Hall92]. The gates must then be de-energized in reverse order before the input values to the cascade may change (Figure 6.7). These "retractile cascades" are impractical for several reasons: they require a large and possibly indeterminate number of supply voltage waveforms; these waveforms all have different pulse widths; and since an $N$-stage cascade requires time proportional to $N$ to produce each result, the latency is proportional to $N$ and the throughput is proportional to $1/N$.

Pipelining can be used to improve the throughput of the system. By using latches to hold the inputs of a retractile stage, we may circumvent the requirement that the preceding stage stays energized until the current stage has been de-energized. Conventional latches, however, cause dissipation independent of the charging time and therefore the unfavorable sub-linear dependence on charging time described in Sections 6.2.1 and 6.2.2. It is possible to avoid using conventional latches, but at a considerable complexity cost, as will be seen in Section 6.3.1 below. An alternative is to use latches with a small, fixed dissipation and allow the dissipation of the combinational parts to scale as $T^{-1/2}$, which allows the use of logic styles with less overhead. One such style is described in Sect

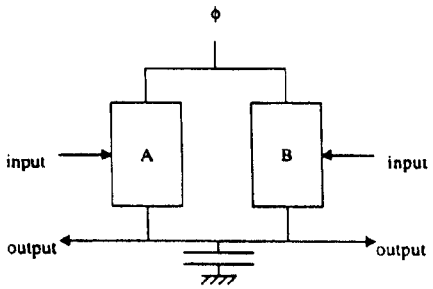



Figure 6.8: Pipelinable adiabatic gate. The load capacitance may be charged through one function network, A, and discharged through another, B. The inputs to the first network must be valid only during the charging phase, and the preceding stage can compute new values during the discharge of the load capacitance.

### 6.3.1 Fully-Adiabatic Sequential Circuits

The requirement for latches to hold the inputs of a function network arises from the need to provide stable control signals throughout the full charge-discharge cycle of the output capacitances. If the retractile stage is expanded to include an explicit discharge path, as shown in Figure 6.8, the controlling inputs of the energizing path need not be kept stable to handle de-energizing. A means for pipelining that does not use latches then exists, and $T^{-1}$ scaling can be retained.

The control signals for the de-energizing path must be stable throughout the discharge to ensure adiabatic operation. Therefore, they cannot be derived directly from the outputs of the stage being de-energized [Koller93]. Any attempt at schemes based on the same idea is bound to fail for thermodynamic reasons [Landauer61]. The de-energizing path must instead be controlled by signals derived from the output of the *following* stage in the pipeline. A pure copy of the signal is sufficient, but this method only defers the problem and is not a practical solution when resources are limited. The ability to derive control signals for the de-energizing path is guaranteed if all logic blocks implement functions that are invertible, so that their inputs may be recomputed from their output.



It is possible to assemble a fully adiabatic pipeline by constructing all of the logic stages according to Figure 6.8 and restricting the function blocks to invertible functions only [Athas94a][Younis93]. The exact logic style chosen for the pipeline determines the timing and sequence requirements for the supply-voltage waveforms. Since these not only power the computations but also pace them, they may equally well be thought of as clock signals, similar to the "hot clocks" of Seitz [Seitz85]. The use of exclusively invertible functions makes the pipeline reversible: if the clocks are reversed in time, the pipeline runs backwards. Reversible logic has been studied in theoretical physics since Bennett showed that computations need not destroy information [Bennett73]. Since then, a large body of theory has been developed [Bennett85][Bennett89][Merkle93].

The complexities of two mutually inverse function blocks are usually similar, so the typical circuit overhead for the separate discharge path is a factor of two, not counting the cost associated with the restriction to invertible functions. The latter cost can be quite prohibitive [Athas94c] and is likely to limit the use of fully-adiabatic solutions to very small circuits with simple functionality, and cases where very slow switching is acceptable as long as the energy dissipation per transition is low enough.

### 6.3.2 Partially-Adiabatic Sequential Circuits

It is possible to enjoy the benefits of adiabatic switching without the complications of reversible logic. Less complex circuitry may compensate for the unfavorable dissipation scaling with time. Several such schemes have been suggested[3] [Seitz85] [Kramer94] [Denker94] [Gabara94].

In this section, we will describe a partially-adiabatic circuit style recently proposed by Denker [Denker94], based on an idea by Koller [Koller93]. We will refer to it as the "flip-flop" style. This scheme has some desirable properties: its device count is low, approaching that of precharged dynamic logic; and a pipeline of these gates can be operated with a four-phase clock, a relatively simple requirement compared to some other suggestions. It is quite different from the adiabatic amplifiers described in Section 6.2, in that no T-gates are used. The analysis in Section 6.2 is therefore not immediately applicable.

[3] Other styles that achieve sub-$C_L V_{dd}^2$ dissipation but get no benefit from slower switching have also been described [Hinman93]; regardless of their other qualities, such schemes cannot really be considered adiabatic, based on the definition on page 182.

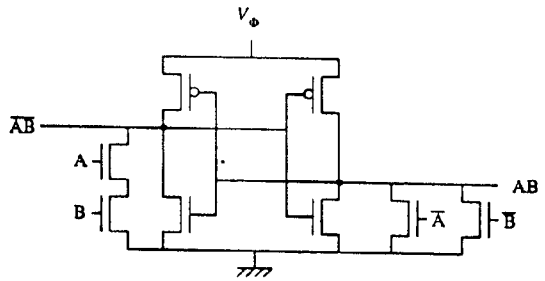



Figure 6.9: Flip-flop-style adiabatic charging logic gate. The example shows a gate implementing an AND/NAND function. Both inputs and outputs are dual-rail-encoded. Load capacitances are not shown. $V_\Phi$ is connected to the pulsed-power supply.

Figure 6.9 shows an example of a flip-flop gate. It consists of two inverters cross-coupled in a flip-flop configuration, and two pull-down networks, which implement two complementary logic functions. Both inputs and outputs are dual-rail-encoded.

The gate works as follows. Assume that the load capacitances (which represent inputs of other gates and wiring capacitances) and the supply/clock line are all initially at 0 V. First, the inputs are given a set of valid values. This will connect exactly one of the output lines to ground, since the networks implement complementary functions. Next, the supply line is ramped from 0 to $V_{dd}$. Initially, no significant current flows, since both the PMOS devices are turned off. When the supply voltage reaches $V_{th}$ (the PMOS threshold voltage), the device whose gate is grounded by a pull-down network will turn on, and the corresponding load capacitance will begin to charge up. This causes an initial, non-adiabatic dissipation of approximately $(1/2)\,C_L V_{th}^2$. The remainder of the charging, up to $V_{dd}$, of the chosen output capacitance will be adiabatic, and its dissipation will decrease with increased charging time. When the charging is complete, the output values are valid and the high inputs are allowed to go low. This disconnects all pull-down networks, but the feedback through the cross-coupled inverters will preserve the output values. Other gates, which use the outputs of the current gate as inputs, may be energized while the inputs of the current gate are ramped down.



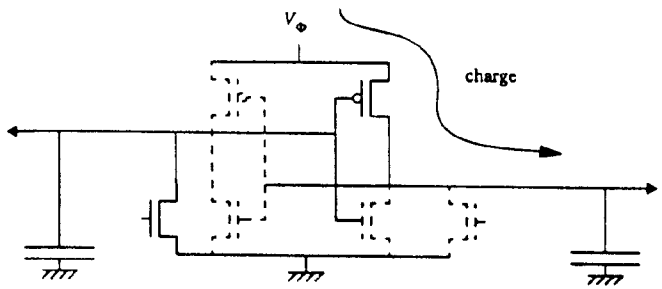



Figure 6.10: Charging an output capacitance of a buffer, implemented in the style of Figure 6.9. The pull-down device on the left grounds the gate of the PMOS device, creating a path from the source to the load. Devices drawn dashed are not conducting.

When the output values have been sampled, the gate may be de-energized by ramping down the supply voltage. The charge and energy on the load capacitance flows back into the power supply through the PMOS device. When the supply voltage reaches $V_{th}$, the PMOS device is turned off, and no further energy recovery will take place.

The charge left on the load capacitance at the end of the discharge will leak away slowly through the turned-off devices. If, however, the supply voltage is ramped up again before much charge has leaked away, and the same pull-down network is tied, the charging will resume close to $V_{th}$. Thus, the non-adiabatic dissipation will be less than $(1/2)\,C_L V_{th}^2$; ideally, with very low leakage, it will be close to 0. In case the other pull-down network is tied, both load capacitances will change values non-adiabatically, resulting in an energy dissipation of $2 \cdot (1/2) \cdot C_L V_{th}^2$. Non-adiabatic dynamic dissipation therefore occurs mainly when the output value changes, which is also the case for conventional logic.

Using the same style of analysis as in Section 6.2, we now calculate the efficiency of a buffer implemented in this style when charging a capacitive load. Figure 6.10 shows the active parts of such a buffer during the charging phase. Initially, assume that the load capacitance is charged up to $V_{th}$ (the PMOS threshold voltage). The supply voltage is ramped up from 0 to $V_{dd}$; the charging device will turn on when the supply voltage passes $V_{th}$, and the load will be charged from

$V_{th}$ to $V_{dd}$. According to Equation 178, the resistance of the charging device can be modelled as:

$$R_p(V_{ch}) = \frac{K_p}{C_p(V_{ch} - V_{th})} \tag{197}$$

Again, this does not take the body effect into account. The resistance depends on the channel voltage in a highly non-linear fashion (it grows without bounds when $V_{ch}$ approaches $V_{th}$). It is therefore not possible to break out the dependence on the current waveform shape into a $\xi$ factor, as in Equation 174. Instead, we calculate the dissipation for one specific case: that of a sinusoidal current, as would be generated by an inductive power supply with very low output impedance (cf. Section 6.5). Charging from $V_{th}$ to $V_{dd}$ will then require a current given by:

$$I(t) = \frac{\pi}{2}\frac{V_{th}(m-1)C_L}{T} \cdot \sin\left(\frac{\pi t}{T}\right) \tag{198}$$

where $V_{dd} = m \cdot V_{th}$. The voltage drop across the device is small when the charging time is long, so the channel voltage is approximately the same as the voltage on the load capacitance:

$$V_{ch} \approx V_C = V_{th} + \int_0^t I(\theta)d\theta = V_{th}\left(1 + (m-1)\sin^2\left(\frac{\pi t}{2T}\right)\right) \tag{199}$$

The resistance as a function of time is then:

$$R_p(t) = \frac{\tau_p}{C_p(m-1)}\csc^2\left(\frac{\pi t}{2T}\right) \tag{200}$$

where $\tau_p = K_p/V_{th}$. The instantaneous power dissipation during the charging is given by:

$$P(t) = I(t)^2 R_p(t) = \frac{\pi^2}{4}(m-1)\frac{\tau_p}{T^2}\frac{V_{th}^2 C_L^2}{C_p} \cdot \sin^2\left(\frac{\pi t}{T}\right)\csc^2\left(\frac{\pi t}{2T}\right) \tag{201}$$

The total adiabatic energy dissipation is:

$$E_R = \int_0^T P(t)dt = 4(m-1)\left(\frac{\pi^2}{8}\frac{\tau_p}{T}\frac{C_L^2}{C_p}V_{th}^2\right) \tag{202}$$



The adiabatic part of the dissipation for the flip-flop driver grows linearly with the voltage swing above the threshold voltage. There is no "optimum" voltage, as for the charging through a T-gate (described by Equation 182); a lower swing yields a lower dissipation. The price paid is the non-adiabatic $C_L V_{th}^2$ dissipated when the output value changes. In the T-gate driver, this dissipation is avoided through the use of an NMOS device which ensures low resistance and adiabatic behavior throughout the charging, regardless of the previous output value.

According to Equation 202, the dissipation is proportional to $\tau_p$. This is since the flip-flop driver, as described here, charges and discharges its load capacitances through PMOS devices. Since the carrier mobility is larger for NMOS devices (and therefore $K_n > K_p$ and $\tau_p > \tau_n$), the dissipation may be decreased by "turning the circuit upside down" so that the load capacitances are charged through NMOS devices instead.

The assumptions used for the derivation of Equation 202 are largely the same as those used in Section 6.2 (the load capacitance is large compared to the gate capacitance of the driver; the voltage drop across the driver device is small; the body effect is neglected; etc.). An additional assumption is that the current will stay sinusoidal despite the severe resistance variations. This is an obvious oversimplification: with an inductive supply, the real current waveform depends on the behavior of the whole ensemble of circuits driven by the same voltage phase.

It must finally be stressed that the comparison of the dissipation of two buffers does not tell the whole story about the total dissipation of systems using these buffers. The simplicity and low device count of the gate style outlined here makes it possible to build small and appealingly simple circuits. This in turn lowers the overall driven capacitances and contributes to low dissipation even for relatively large values of $m$ and $V_{dd}$.

## 6.4 Stepwise Charging

As we have seen, the key to low dissipation in an adiabatic-switching circuit is to lower the average voltage drop traversed by the charge that flows onto the load capacitance. The ideal case, yielding the smallest dissipation for a given charging time, is when the charging current can be kept constant. This requires that the power supply be able to generate linear voltage ramps. Practical supplies have

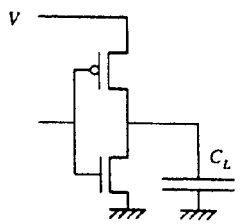



Figure 6.11: Conventional CMOS signal driver with capacitive load.

used resonant inductor circuits (as outlined in Section 6.5) to approximate the constant currents and linear ramps with sinusoidal signals.

Inductor circuits have several practical drawbacks. High-Q inductors cannot be integrated on a chip together with other circuits and thus complicate the board-level design. The efficiency of the inductive solutions may rely on accurately timed control pulses. While small timing errors can cause ringing and efficiency loss, larger errors may cause damaging voltage spikes. The problem of designing an efficient power supply for variable capacitive loads has led some workers towards dual-rail logic, which approximately doubles the amount of circuitry.

This section describes *stepwise charging*, an alternative way to generate approximate voltage ramps [Svensson94]. Being inductor-less, it suffers none of the drawbacks mentioned above. In addition, it requires no changes in the power supply and may therefore be included in existing designs with relative ease. On the other hand, the circuit solutions involved are more complex and require more silicon area. The method is likely to be useful to drive those few nodes in a circuit that cause a large part of the dissipation, such as output pads and busses. It is not readily extensible to logic gates.

Again, consider the conventional CMOS driver, repeated in Figure 6.11. All charge delivered to the load is injected at the supply voltage, and thus the dissipation caused by charging $C_L$ to $V$ is $(1/2)\,C_L V^2$, regardless of the speed with which the charging takes place. A full cycle causes twice this dissipation.



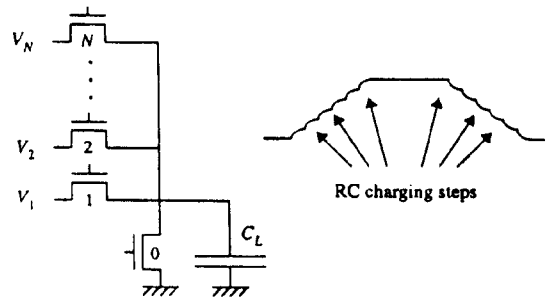



Figure 6.12: A stepwise driver for a capacitive load. Voltages are given by the formula $V_i = (i/N)\,V$. Switches are shown as NMOS devices, but some may be PMOS. The voltage waveform on the load capacitance is also shown.

If several voltage sources are available, we may save energy by carrying out the charging in several steps. Figure 6.12 shows a basic stepwise driver for a capacitive load, implemented with NMOS devices. A bank of $N$ voltage supplies with evenly distributed voltages is used. The load is charged by connecting $V_1$ through $V_N$ to the load in succession (by closing switch 1, opening switch 1 and closing switch 2, etc.). To discharge the load, $V_{N-1}$ through $V_1$ are switched in in the same way, and then switch 0 is closed, connecting the output to ground.

For each step, the dissipation is again given by the transferred charge and the average voltage drop across the switch resistance:

$$E_{step} = Q\bar{V} = C_L\frac{V}{N}\cdot\frac{V}{2N} = \frac{1}{2}C_L\frac{V^2}{N^2} \tag{203}$$

$N$ steps are used to charge $C_L$ all the way to the supply voltage $V$, so the total energy dissipation is:

$$E_{stepwise} = N\cdot E_{step} = N\cdot\frac{1}{2}C_L\frac{V^2}{N^2} = \frac{1}{2}C_L\frac{V^2}{N} \tag{204}$$

Again, a full charge-discharge cycle will cause twice the dissipation of the charging only. Thus, according to this simplified analysis, charging by $N$ steps reduces

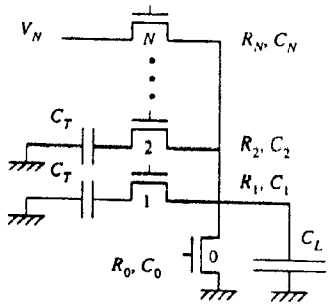



Figure 6.13: A stepwise driver with all but one of the necessary voltages supplied by tank capacitors, $C_T$. The $R_i, C_i$ pairs represent the on-resistance and the gate capacitance of the $i$th switch.

the energy dissipation per charge-discharge cycle, and thereby the total power dissipation, by a factor of $N$. The decrease comes from a reduction of the voltage drop across the switch through which the charge flows, which is now on average a factor $N$ smaller than in the conventional case.

While avoiding the inductor problems, this scheme suffers from a few other drawbacks. Most notably, multiple supply voltages are needed. Not only is the overhead of routing many supply lines to each circuit in a system undesirable; more importantly, a power supply able to efficiently generate the voltages would be complex and expensive.

These problems can be circumvented by using the circuit in Figure 6.13. It shows the same circuit as Figure 6.12, but with all supplies except one replaced with large "tank" capacitors. The tank capacitor voltages will change little during the entire charging of the load capacitance, provided that the capacitors are large enough and initially charged to the evenly distributed voltages of Figure 6.12. The behavior and dissipation will then be the same as for the circuit in Figure 6.12. No extra circuitry is required to maintain the voltages on the tank capacitor bank: it can be shown that the tank capacitor voltages converge to the desired, evenly distributed voltages. Thus, only one supply line must be routed to the chip, and the power supply is no more complicated than in the conventional case. The main cost is a larger chip area for the switches and the package pins set



aside for the tank capacitors (these would most often be kept off-chip, since they must be large compared to the driven capacitance). Note, however, that several drivers may share the capacitor bank, and hence the drivers for an entire bus may need only a few extra package pins.

Equation 204 indicates that dissipation decreases linearly with increasing $N$. $N$ cannot, however, usefully be made arbitrarily large, because each step requires that a switch be turned on and off, which itself causes dissipation. Also, the energy used to control each switch depends on the width of the device, which should be just enough to allow the charging to "complete" before the next step commences. Thus, for a given total allowable charging time, there is an optimal number of steps and a set of optimal device sizes which lead to minimal total dissipation. We find these by deriving an expression for the gate drive energy as a function of $N$, and then minimize the sum of the gate-drive energy and the dissipation caused by the charge flowing onto the load capacitance.

In the derivations that follow, we assume that the voltage swing is identical in all parts of the circuit, so that the gate of a switch device is tied either to ground or to $V$. We model the channel of switch $i$ as a linear resistance, $R_i$, and its gate as a linear capacitance, $C_i$, connected to ground. All switch devices are assumed to be minimum-length; only the width will be varied. We introduce the switch quality measure, given by:

$$\rho_i = R_i C_i \tag{205}$$

$\rho_i$ is independent of the switch width. However, it varies with $i$, since the channel voltage is different for different $i$; the voltage dependence is shown in Figure 6.14. It is also different for PMOS and NMOS devices. Finally, $\rho_i$ is obviously process-dependent. It is similar to the $\tau$ of Mead and Conway [Mead80], in that it quantifies the maximum speed attainable in the process, but it takes the behavior across the entire voltage range into account. Referring again to Figure 6.14, it is clear that a PMOS device is a better switch at high channel voltages, while an NMOS device is preferable at lower voltages.

Assume that the gates of all the switches are driven conventionally, so that the dissipation is given by $CV^2$. The total gate energy spent to charge and discharge the load capacitance once is then:

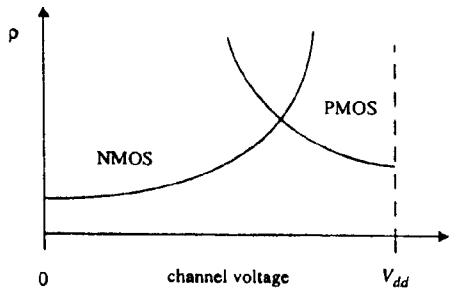



Figure 6.14: Conceptual dependence of the switch quality, ρ, on the channel voltage. For high channel voltages, a PMOS device with its gate tied to ground conducts better than an equally-sized NMOS device with its gate tied to $V_{dd}$. A lower value means a better switch.

$$E_{gates} = \left(\sum_{i=1}^{N} C_i + \sum_{i=0}^{N-1} C_i\right) V^2 \tag{206}$$

Next, we relate the switch gate capacitances, $C_i$, to $N$. Allot each step one $N$th of the total available charging time, $T$. Then:

$$\frac{T}{N} = jR_iC_L \tag{207}$$

Here, $j$ is a new parameter: the number of $RC$ time constants spent waiting for each charging step to "complete." If its value is chosen too small, there will still be a significant voltage across a switch when the next switch is to take over. Hence, there is an increase in the average voltage across each switch, and therefore a dissipation increase. If on the other hand $j$ is chosen unnecessarily large, time is wasted that could have been used to increase the number of steps. Suitable values range from 2 to 4. We see from Equation 207 that if $j$ is the same for all steps, all the switch devices should have equal on-resistance. The required gate capacitance for switch $i$ is then given by:

$$C_i = \frac{\rho_i}{R_i} = \frac{Nj\rho_i}{T}C_L \tag{208}$$

Substitute Equation 208 into Equation 206:



$$E_{gate} = \frac{Nj}{T}\left(\sum_{i=1}^{N} \rho_i + \sum_{i=0}^{N-1} \rho_i\right) C_L V^2 \tag{209}$$

Equation 209 still cannot be used immediately for optimization, since the dependence on $N$ is complicated and non-intuitive. We therefore introduce $\bar{\rho}$, a weighted average of $\rho_i$ for the different switches:

$$\bar{\rho} = \frac{1}{2N}\left(\sum_{i=1}^{N} \rho_i + \sum_{i=0}^{N-1} \rho_i\right) \tag{210}$$

$\bar{\rho}$ is close to the unweighted average of ρ over the entire voltage range if $N$ is sufficiently large. The unweighted average is independent of $N$ and may be used in place of the weighted average when finding the best $N$. We get:

$$E_{gate} = \frac{Nj}{T} \cdot 2N\bar{\rho} \cdot C_L V^2 \tag{211}$$

Combining Equations 211 and 204, we get the following expression for the total dissipation for a full charge-discharge cycle:

$$E_{tot} = \left(\frac{1}{N} + 2N^2 j\frac{\bar{\rho}}{T}\right) C_L V^2 \tag{212}$$

The $N$ that minimizes $E_{tot}$ is given by:

$$N_{opt} = \sqrt[3]{\frac{T}{4j\bar{\rho}}} \tag{213}$$

The corresponding energy dissipation is:

$$E_{opt} = \frac{3}{2}\sqrt[3]{\frac{4j\bar{\rho}}{T}} C_L V^2 \tag{214}$$

By using the number of stages given by Equation 214, the designer can minimize the power dissipation of the driver. The minimum is rather shallow, however, so a lower $N$ (as would most often be dictated by practical considerations) will still give a considerable improvement over the conventional case; $N = 2$ already gives almost 50% reduction unless $T$ is very small. Once $N$ and $j$ have been selected, the gate capacitance, and thereby the width, of each switch is given by Equation 208. The values of ρ for a certain process can be found by

circuit simulation or by measuring the on-resistances of test devices of known widths.

The accuracy of this design procedure is surprisingly good, given the many simplifications that have been made. A test driver with six steps (2μ CMOS, $T$ = 500ns) was predicted by the equations to reduce the dissipation by 80%. Circuit simulation of the extracted layout resulted in 77% improvement. Measurements on the fabricated chip yielded 73% improvement [Svensson94].

The improvement figures do not take the energy needed to generate the sequence of control signals into account. In the test chip, this was done with a state machine that had not been optimized for low power. In a practical chip, part of this overhead may be amortized over several drivers that share part of the control logic (for example, all signals on a bus would switch in synchrony).

A stepwise-charging driver may save an order of magnitude in dissipation over the conventional counterpart when the switching may be slow compared to what would be achievable with a conventional driver. It uses only capacitors and switches and may therefore be integrated (but the tank capacitances would most be kept off-chip). Finally, it does not require a special power supply, and is therefore considerably easier to design into a conventional system.

## 6.5 Pulsed-Power Supplies

The preceding sections have discussed the design of adiabatic-switching gates and circuits. The pulsed-power supply waveforms were assumed to be given. This section provides an introduction to the problems of designing a resonant inductive power supply, and gives an analysis of the attainable efficiency for a simple example.

Figure 6.15 depicts a very simple inductive pulsed-power supply. It consists of an inductor, $L$, and three switches labelled A, B, and C. The load presented to the power supply by the logic circuit is mainly capacitive (the total capacitance of all the driven MOS gates, plus wiring capacitance and other stray capacitances). Most of the capacitance is charged through device channels, which adds a resistive component to the load. In this section, we will assume that the load can be represented by a linear resistance in series with a linear capacitance.

The circuit works as follows. Assume that the load is initially fully discharged, and that only switch B is tied. To charge the load, B is cut and A is tied. A and $L$ now form an RLC circuit together with the load. Charge will flow



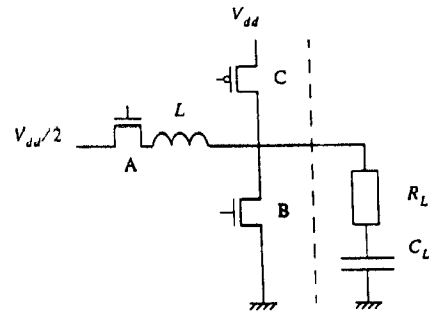



Figure 6.15: A simple inductive power supply driving a load with a resistive and a capacitive component, representing an adiabatically switching logic circuit.

through $L$ onto $C_L$. Provided that the RLC circuit is underdamped, the load voltage will swing past $V_{dd}/2$; if the circuit is highly underdamped, the voltage will peak at close to $V_{dd}$ when the inductor current goes to 0. At this instant, A must be cut and C tied to pull the output all the way to $V_{dd}$. Similarly, discharging the load is accomplished by cutting C, tying A, waiting for the load voltage to reach its lowest value, and cutting A and tying B.

The power-supply circuit is a second-order system; its dissipation during a full charge-discharge cycle is easily found using standard time-domain circuit analysis. The state of the circuit at the end of a full cycle is identical to that at the beginning, and for reasons of symmetry, the amount of charge pulled from the $V_{dd}/2$ connection during the charge phase is equal to that pushed back during the discharge phase. Thus, the dissipated energy is equal to the energy injected from $V_{dd}$ through switch C. The waveforms are given by the values of $L$, $C_L$, and $R_{total}$, which consists of $R_L$ and $R_A$, the on-resistance of switch A. Assuming a heavily underdamped circuit, we get:

$$I(t) = \frac{V_{dd}}{2\omega L} e^{-\frac{\alpha t}{2}} \sin \omega t \tag{215}$$

$$\alpha = \frac{R_{total}}{L} = \frac{R_L + R_A}{L} \tag{216}$$

$$\omega = \sqrt{\frac{1}{LC_L} - \frac{\alpha^2}{4}} \approx \frac{1}{\sqrt{LC_L}} \tag{217}$$

The maximum voltage reached at the load (at time $T = \frac{\pi}{\omega} \approx \pi\sqrt{LC_L}$) is:

$$V_{max} = V_{dd}(1-\varepsilon) \tag{218}$$

$$\varepsilon = \frac{1}{2}\left(1 - e^{-\frac{\alpha\pi}{2\omega}}\right) \approx \frac{\alpha\pi}{4\omega} \approx \frac{\pi^2}{4} \cdot \frac{R_{total}C_L}{T} \tag{219}$$

The injected energy, and therefore the dissipated energy, is given by:

$$E_{inj} = V_{dd}(V_{dd} - V_{max})\,C_L = \varepsilon \cdot C_L V_{dd}^2 \tag{220}$$

We make the following observations:

- Comparing Equations 174, 219, and 220, we see that $\xi = \pi^2/8 \approx 1.234$ for sinusoidal currents.
- Unless $C_L$ is constant from cycle to cycle, the charging time $T$ will have to vary (assuming that $L$ is held constant).
- The $V_{dd}/2$ source may be replaced with a large tank capacitance, $C_T$, charged to $V_{dd}/2$. As long as $C_T \gg C_L$, the behavior change is negligible.
- The dissipated energy may be separated into two terms, one depending on $R_A$ and the other on $R_L$:

$$E_{inj} = \frac{\pi^2}{4} \cdot \frac{R_A C_L}{T} \cdot C_L V_{dd}^2 + \frac{\pi^2}{4} \cdot \frac{R_L C_L}{T} \cdot C_L V_{dd}^2 \tag{221}$$

  The latter term depends only on the load and is independent of the power supply, as long as the voltage and current levels stay the same.

We will now size switch A to minimize the overall dissipation, including that of the circuit driving the gate of switch A. $R_A$ is given by Equation 176, with $V_{ch} = V_{dd}/2$:

$$R_A = \frac{2K_n}{C_A(V_{dd} - 2V_{th})} = \frac{2\tau_n}{C_A(m-2)} \tag{222}$$

Since $\tau_n$ is process dependent, it may be different for the power-supply switch than for the logic circuits. Substitute Equation 222 into the first term of



Equation 221, and assume that the gate of switch A is driven conventionally. We must minimize this expression with respect to $C_A$:

$$E_{supply} = C_A V_{dd}^2 + \frac{\pi^2\tau_n}{4}\frac{2}{T}\frac{C_L}{(m-2)C_A} \cdot C_L V_{dd}^2 \tag{223}$$

This expression is similar to Equation 184. The optimal value for $C_A$ is:

$$C_{A_{opt}} = \sqrt{\frac{\pi^2\tau_n}{2}\frac{1}{T}\frac{1}{(m-2)}}\,C_L \tag{224}$$

The total energy dissipation is given by:

$$E_{tot} = \left(\sqrt{2\pi^2\frac{\tau_n}{T}\frac{1}{(m-2)}} + \frac{\pi^2 R_L C_L}{4}\frac{1}{T}\right)C_L V_{dd}^2 \tag{225}$$

When the charging time is increased, the last term of Equation 225 becomes negligible due to its $T^{-1}$ dependence, and the $T^{-1/2}$ term dominates. We recognize the $\sqrt{1/T}$ dependence from Section 6.2.1. Again, the situation may be improved by driving the gate of the switch device adiabatically, which would allow us to use a wider switch. However, a very large gate capacitance also means that the drain of switch A (the right-hand side in Figure 6.15) has a large stray capacitance. When switch A is cut, this stray capacitance forms an RLC circuit with $L$ and the load. Ringing will ensue due to the voltage present across $L$ when switch A is cut. The ringing not only produces high voltages at the drain of switch A, but also causes dissipation proportional to the drain stray capacitance.

The ringing problem can be circumvented by removing the controlling switch device from the main charge path, as shown in Figure 6.16. The RLC circuit is now free-running, and the efficiency is limited mainly by the load resistance, but one degree of freedom has been lost, in that the supply voltage waveforms provided to the load are now close to sine waves. The energy dissipation per cycle is reduced if the charging time is chosen to be very long (by increasing $L$). Eventually, the limit on energy dissipation per operation will be set by the parasitic series resistances of the available and affordable inductors.

Most published papers on power supplies for adiabatic circuits describe variants on this theme [Hinman93][Younis93][Athas94b] [Athas94c] [Gabara94]. As an additional variation, only the timing of the switches must be changed to operate the circuit of Figure 6.16 as an edge-resonant circuit [Maksimovic91]

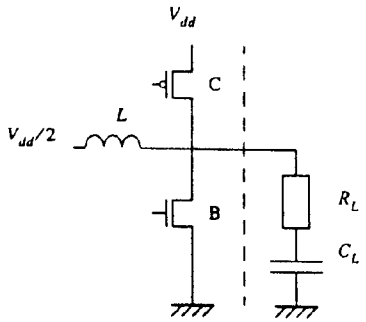



Figure 6.16: Inductive power supply like the one in Figure 6.15, but lacking switch A, which causes most of the dissipation.

[Wik94]. When the load is discharged, switch B is tied long enough to let a significant current build up in the inductor. When B is cut, the current keeps flowing and charges the load almost like a constant-current source. Switch C is tied when the load reaches $V_{dd}$ and held closed for long enough for the current to reverse.

In another variation, if several symmetrical clock waveforms are used, $C$ may be replaced by the "opposite" load capacitance with no significant change in efficiency, as shown in Figure 6.17. Such a circuit can only be used when the capacitances to be driven by the two phases are equal. Again, the switch in the main path of the charge may be removed for improved efficiency, but at the price of less controllability.

Other types of power supplies are conceivable. In addition to other kinds of inductive supplies, it would be possible to use a switched-capacitance circuit like those used as signal drivers in Section 6.4, since the load presented to the power supply is mainly capacitive; the latter approach would appear to give a rather low overall efficiency. It has also been speculated that high-Q circuits of other kinds, such as piezo-electric resonators, might become useful [Solomon94].

We finally note that until now, the power supplies required for adiabatic circuits have received considerably less attention than the logic circuits themselves. More work is clearly needed in this area for adiabatic switching to become useful at very small switching energies.



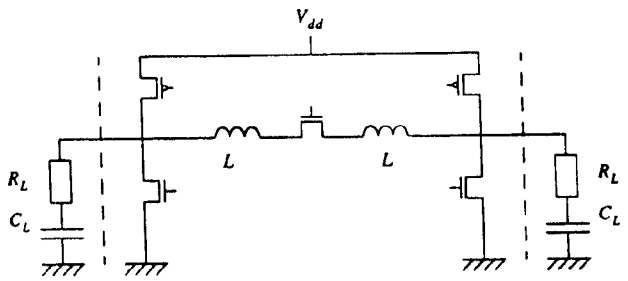



Figure 6.17: Symmetrical clock waveform generator; used to generate two clock waveforms with opposite phase. During the transition, charge flows from one load capacitance to the opposite one.

## 6.6 Summary

Most approaches to the reduction of dynamic power dissipation in electronic switching circuits start from the ubiquitous equation $P = fCV^2$. The threefold path to lower dissipation is indicated by the three factors in the expression: to reduce power, we must reduce the switching frequency, the driven capacitance, or the supply voltage, or (preferably) some combination of these. When, for whatever reason, these factors cannot be further reduced, adiabatic switching is the only known approach to further reduce dynamic dissipation.

Adiabatic circuits can clearly be used to great benefit when a small number of circuit nodes of significant capacitance must be driven to a high voltage (compared to the threshold voltage of the process). Such situations arise wherever the swing is fixed by some physical phenomenon. Examples include capacitive transducers, micromechanical devices, and LCD panels. In addition, voltage swing is sometimes determined by industry signalling standards. Then, adiabatic switching may allow a designer to reduce the interface part of the power budget by circumventing the $P = fCV^2$ barrier. The benefits increase with the voltage swing and with the switching time and are relatively easy to quantitatively assess.

Logic circuits of higher complexity than mere buffers may also benefit from adiabatic switching. The range of viable circuit styles is larger than in the driver case, and there is no consensus yet as to what schemes would be preferable in terms of circuit complexity and overall power dissipation. In part, this is since

the complexity and efficiency of the required pulsed-power supplies is difficult to factor into the equations. While buffer circuits can use relatively simple power waveforms, many of the proposed logic schemes rely on intricate timing and level relationships between several supply voltages.

The design of the pulsed-power supplies may turn out to be the most critical element to the ultimate practicality of ultra-low-dissipation adiabatic circuits. The efficiency of most pulsed-power supplies presented so far is limited by the on-resistance of a small number of switches, through which the supply currents to the logic circuit passes. Also, many current low-power systems make heavy use of gated-clock design styles, where the clock signals of an idle block are stopped to avoid unnecessary logic transitions inside the circuit. The equivalent state in an adiabatic circuit would be one where the pulsed-power supply signals are held constant for an inactive block. A fine-grain control of the clock signals would be necessary to do this selectively for the different parts of a system. It is presently unclear how to efficiently achieve this.

In adiabatic switching circuits, the on-resistance of switches determines the dissipation but has no influence on the charging times. Broadly speaking, a process improvement (increased transconductance, reduced parasitic capacitances) that increases the possible speed in a conventional circuit allows lower dissipation in an adiabatic one. Thus, the qualitative comparisons described in this chapter should stay valid while semiconductor processes evolve.

The full impact of adiabatic switching on the design of low-power switching circuits is yet to be determined. Its utility in interface circuits seems evident, but it is at this point unclear whether superior overall dissipation can be achieved for switching circuits with more complex logic functions. New logic styles which offer substantially better power-delay products may still emerge.

## 6.7 Acknowledgments

This chapter is founded on the adiabatic-switching work carried out at the Information Sciences Institute (ISI) of the University of Southern California, together with W.C. Athas, J.G. Koller, N. Tzartzanis, and E. Chou. In addition, discussions with J.S. Denker, D.J. Frank, T. Gabara, J.S. Hall, T.F. Knight, S. Mattisson, L. Peterson, P.M. Solomon, and T.R. Wik have been very helpful.

The work on adiabatic switching at ISI is supported by ARPA under contract no. DABT63-92-C0052.



REFERENCES:

[Athas94a] W.C. Athas, L."J." Svensson, J.G. Koller, N. Tzartzanis, and E. Chou, "Low-power digital systems based on adiabatic-switching principles," *IEEE Transaction on VLSI Systems*, pp. 398–407, Dec 1994.

[Athas94b] W.C. Athas, J.G. Koller, and L."J." Svensson, "An energy-efficient CMOS line driver using adiabatic switching," *Fourth Great Lakes Symposium on VLSI Design*, pp. 196–199, March 1994.

[Athas94c] W.C. Athas and L."J." Svensson, "Reversible logic issues in adiabatic computing," *IEEE Workshop on Physics and Computation*, PhysComp '94, pp 111–118, IEEE Press, Nov 1994.

[Bennett73] C.H. Bennett, "Logical reversibility of computation," *IBM J. Res. Dev.*, vol 17, pp. 525–532 (1973).

[Bennett85] C.H. Bennett and R. Landauer, "The fundamental physical limits of computation," *Scientific American*, pp. 48–56, July 1985.

[Bennett89] C.H. Bennett, "Time/space trade-offs for reversible computation," *SIAM J. Computing*, vol. 18, pages 766–776, 1989.

[Burr91] J. B. Burr and A.M. Peterson, "Energy considerations in multichip-module based multiprocessors," *IEEE International Conference on Computer Design*, pp. 593–600, 1991.

[Denker94] J. S. Denker, "A review of adiabatic computing," *IEEE Symposium on Low Power Design*, pp 94–95, 1994.

[Gabara94] T. Gabara, "Pulsed power supply CMOS - PPS CMOS," *IEEE Symposium on Low Power Design*, pp 98–99, 1994.

[Hall92] J.S. Hall, "An electroid switching model for reversible computer architectures," *Proc. ICCI'92, 4th International Conf. on Computing and Information*, 1992.

[Hinman93] R. T. Hinman and M. F. Schlecht, "Recovered energy logic: a highly efficient alternative to today's logic circuits," *IEEE Power Electronics Specialists Conference Record*, pp. 17–26, 1993.

[Koller93] J.G. Koller and W.C. Athas, "Adiabatic switching, low energy computing, and the physics of storing and erasing information," *Workshop on Physics and Computation*, PhysCmp '92, Oct. 1992; IEEE Press, 1993.

[Kramer94] A. Kramer, J.S. Denker, S.C. Avery, A.G. Dickinson, and T.R. Wik, "Adiabatic computing with the 2N-2N2D logic family," *1994 Symposium on VLSI Circuits: Digest of technical papers*, IEEE Press, June 1994.

[Landauer61] R. Landauer, "Irreversibility and heat generation in the computing process," IBM J. Res. Dev., vol 5, pp. 183–191 (1961).

[Liu93] D. Liu and C. Svensson, "Trading speed for low power by choice of supply and threshold voltages," *IEEE JSSC*, vol 28, no. 1, Jan 1993.

[Maksimovic91] D. Maksimovic, "A MOS gate drive with resonant transitions," *IEEE Power Electronics Specialists Conference*, pages 527–532. IEEE Press, 1991.

[Mead80] C.A. Mead and L. Conway, *Introduction to VLSI systems*, Addison-Wesley, Reading, 1980.

[Mead] C. Mead, "Scaling of MOS technology to submicrometer feature sizes," *Journal of VLSI Signal Processing*, 8, 9–25 (1994).

[Merkle93] R. Merkle, "Reversible electronic logic using switches," *Nanotechnology*, vol. 4, pages 21–40, 1993.

[Seitz85] C. L. Seitz, A.H. Frey, S. Mattisson, S.D. Rabin, D.A. Speck, and J.L.A. van de Snepscheut, "Hot-clock NMOS," *1985 Chapel Hill Conference on VLSI*, pp. 1–17, 1985.

[Solomon94] P.M. Solomon and D.J. Frank, "The case for reversible computation," *1994 International Workshop on Low-Power Design*.

[Svensson94] L."J." Svensson and J.G. Koller, "Driving a capacitive load without dissipating $fCV^2$," *IEEE Symposium on Low Power Design*, pp. 100–101, 1994.

[Wik94] T.R. Wik, Private communication, October 1994.

[Younis93] S.G. Younis and T.F. Knight, "Practical implementation of charge recovery asymptotically zero power CMOS," *1993 Symposium on Integrated Systems*, pp. 234–250. MIT Press, 1993.

# 7

# Minimizing Switched Capacitance

In the previous chapter, power dissipation was minimized in CMOS circuits by aggressive supply voltage scaling. Since CMOS circuits do not dissipate power if they are not switching, another approach to low power design is to reduce the switching activity to the minimal level required to perform the computation. This can range from simply powering down the complete circuit or portions of it, to more sophisticated schemes in which the clocks are gated or optimized circuit architectures are used which minimize the number of transitions. The focus of this chapter is on minimizing the switched capacitance at all levels of the design. The following sections describe a system level approach to minimize the switched capacitance which involves optimizing algorithms, architectures, logic design, circuit design, and physical design.

## 7.1 Algorithmic Optimization

The choice of algorithm is the most highly leveraged decision in meeting the power constraints. The ability for an algorithm to be parallelized will be critical and the basic complexity of the computation must be highly optimized.

### 7.1.1 Minimizing the Number of Operations

Minimizing the number of operations to perform a given function is critical to reducing the overall switching activity. To illustrate the power trade-offs that can be made at the algorithmic level, consider the problem of compressing a video data stream using the vector quantization algorithm. Vector quantization