POWER SUPPLIES FOR ULTRA LOW POWER APPLICATIONS

by

ABRAM P. DANCY

S.B., Massachusetts Institute of Technology (1996)

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCE
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE
OF

MASTER OF ENGINEERING

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

August 1996

© Abram P. Dancy, 1996. All rights reserved.

Author . . . .
Department of Electrical Engineering and Computer Science
August 30, 1996

Certified by . . . .
Anantha Chandrakasan
Assistant Professor
Thesis Supervisor

Accepted by . . . .
Frederic R. Morgenthaler
Chairman, Department Committee on Graduate Theses

Power Supplies for Ultra Low Power Applications

by

Abram P. Dancy

Abstract

This document presents several design issues pertinent to power supplies specifically designed for battery operated ultra low power ($< 50mW$) applications. Optimizations are derived for the sizing of switches in switched capacitor converters, the relative volumes of a power supply and a battery, and the selection of the inductor, capacitor, and switching frequency of an inductor-based switching power supply. The task of optimizing a converter maintaining a fixed volume is stressed as a means to provide maximum system run time and for comparing the efficacy of different converter types. Circuits are developed to accomplish the control of a standard down-converter using digital devices and consuming very little power. These circuits could either utilize measuring the output supply voltage or some other metric, such as the load circuit delay. The use of a single bit feedback signal is proposed as an option which trades performance for lower power dissipation in the control. Results are presented from a fabricated integrated circuit which implements a digital controller with a tapped delay line PWM. The chip demonstrates that a single bit control circuit can result in a stable output, and provides the pulse width modulation function in as little as $15\mu W$ at $250kHz$. Finally, a micromechanical structure which could create a fully integrated power supply is described, and a sample design which processes $1mW$ of power in $10mm^2$ is derived.

Thesis Supervisor: Anantha Chandrakasan
Title: Assistant Professor

I dedicate this work to my two children, Tobin and Sarai. I hope Leigh and I can provide them with as many opportunities as my parents have given to me.

Acknowledgments

I would like to acknowledge the many people who have assisted me over the course of this project. Their help has been invaluable, and I could not have finished without them.

- My advisor Prof. Anantha Chandrakasan provided the enthusiasm and motivation required to lift this research from theory to a fabricated chip. He continuously focuses our research on more novel ideas, which keeps life interesting.

- Prof. Marty Schlecht advised me in the early stages of this project. Marty's influence directed me toward the statements I make about volume as one of the most important design parameters. Marty has also helped me gain an intuition for DC-DC converter design and control over the last five years.

- Fellow graduate student Rajeevan Amirtharajah is to be credited with the design of the dynamic comparator implemented on the test chip. Raj and I have worked together to try to understand the effects of quantization in the digital control on power supply performance.

- Jim Goodman, Michael Perrott, and Jennifer Lloyd suffered through the process of preparing our project chip for fabrication with me. I could not have gotten my chip fabricated without their persistence to decode obscure instructions, tame the CAD tools, and steal licenses. Ted Tewksbury of Analog Devices was also extremely helpful in making sure we spoke to the right people, and then making sure they listened.

- Duke Xanthopolous has taught me nearly everything I know about IC layout, although unfortunately that may not give him enough credit. Duke and Vadim Gutnik answered many of my questions about our CAD tools and surviving in a Unix environment. Vadim also provided insight related to the minimization of the switched capacitor power dissipation equations.

- I would like to that my parents, Chip and Marie Dancy, for their love and support. I wouldn't be here without them!

- Most importantly, I would like to recognize my wonderful wife Leigh Dancy, who has given me more love and happiness than I could ask for. She has been quite tolerant of the writing period of this thesis, and has done much more than her share of the chores for far too long now.

Contents

Chapter 1

Introduction

Designing digital and analog signal circuitry for low power and low supply voltage applications is an area of increasing interest. This trend is driven by the growing number of mobile systems, the desire to add more functionality and operating life to a system with a given battery, and increasing problems with dissipating heat generated by high performance chips. The developments in new technologies and circuit designs to reduce the power of the signal or computational circuitry must be met with advances in the circuitry to provide power to these circuits, which may require extremely low power supply voltages (such as $0.5V$ and below), variable supply voltages [6], or even AC supply voltages [35].

In mobile systems, electronic circuits can usually be designed to operate over most of the range of the voltages supplied by a typical battery over its discharge cycle. Adding some form of power electronics to regulate the supply voltage can give the important advantage of being able to operate the circuitry at an optimal supply voltage (which may be lower or higher than the battery voltage) from the perspective of power consumption and throughput. This flexibility may greatly increase battery life. Power regulation also enables more exotic forms reducing system power, such as variable supply voltages or AC supply voltages as mentioned above. In any case, designers of power circuits for very low power systems must weigh the benefits regulation provides against the increased volume and cost inherent in creating a more complex system.

Low power mobile or distributed systems have the property that they include distinctly different types of subsystems. A portable computer may have radio frequency components for wireless communication, a display and possibly a CCD camera, audio output, mechanical subsystems such as disk drives, and of course digital logic circuitry. Each different type of circuitry will be optimized with a different supply voltage or power delivery scheme. Radio frequency circuits may require 5 to 7 Volt inputs to achieve the required speed, an LCD will generally require large bias voltages, as well as lower signaling voltages, the CCD may benefit from an AC voltage delivery to recover

9

energy from its large capacitive load, and digital logic circuitry may consume the lowest power with supply voltages lower than 1V or with variable supply voltages. As a result, the design of the power supplies must be done in an integrated fashion with the entire system, so that power consumption is minimized across the system, rather than simply within individual blocks. The development of a power supply for a multimedia portable terminal is discussed in [24]. An optimization in a single subsystem that requires a new output voltage may incur overhead or additional volume in the power supply which outweighs the advantage of that optimization. It is helpful to create power circuits which lend themselves to multiple outputs without requiring large amounts of additional volume, as well as control circuits which present very low barriers to additional outputs (low quiescent power and minimal area).

1.1   Low Power Supply Regulators

The types of power converters used for low power applications are the conventional buck, boost, and buck/boost DC-DC converters, switched capacitance voltage dividers and multipliers, and simple linear regulation. The second two items, switched capacitor circuits and linear regulators are generally not used at higher powers due to the poor power density of these converters at high currents. However, the power density required for some low power applications does not preclude the use of these alternate types of converters, and such solutions can be competitive with the more common (for higher powers) inductor based DC-DC converters.

Power density is an important metric for power supplies. Without a statement of the volume of a converter, the power handling efficiency of the converter reveals little about the novelty of the design. The basic tradeoffs for a DC-DC converter are those between volume, efficiency, and performance. Therefore, this thesis attempts to develop theory to evaluate the power density of the various types of power converters. A reasonable way to compare different topologies is to measure what the increase in system run time (for a battery powered system) is for a fixed volume.

1.2   Specific Requirements for Low Power Digital Circuits

The power dissipation of CMOS digital circuits is governed by the equation

$$P_{diss} = \alpha f_{sw} C V_{DD}^2, \tag{1.1}$$

where C is the physical capacitance of the circuit, and α is a measure of the fraction of the capacitance which undergoes transitions at each clock cycle. Designing CMOS digital circuits for low power operation involves decreasing α, by avoiding unnecessary switching, decreasing f_{sw}, sometimes carried out by parallelizing computations, reducing the physical capacitance, accomplished by

scaling devices smaller, and finally, reducing the power supply voltage. Reducing the power supply voltage is not without its detrimental effects; lower supply voltages create longer circuit delays. This can be expressed to the first order by [2]

$$t_d = \frac{CV_{DD}}{k' \frac{W}{2L} (V_{DD} - V_T)^2}.$$

(1.2)

The desire to reduce V_{DD} has several effects on the power supply design. The first and most obvious effect is that power supplies must be designed for lower output voltages. Supply voltage levels will fall as low as tenths of a volt, for technologies such as silicon-on-insulator or reduced V_T bulk processes. Output voltages as low as this are not common among power supplies, and this shift to decreasing voltages will force designers to modify their power supply designs. A second effect on power supply design caused by the need to reduce V_{DD} is a shift toward variable power supplies. Since lowering the supply voltage causes a decrease in circuit speed, it is an attractive proposition to lower the supply voltage when the circuit is to experience a reduced computational load, and raise the voltage again when shorter circuit delays are required. For example, a DSP chip operating on a compressed video stream will have very little load when processing frames which only contain the difference between the previous image and the current image, but the same system could see a large computational load when it must process a full frame. Again, such requirements on power supply outputs for digital circuits were not common in the past. Typically output voltage, ripple, and transient response is specified, without any need to dynamically change the output voltage. Finally, power reduction techniques are creating a push to build power supply systems which do not directly control their output voltage, but some metric based on the speed requirements of their load. For low power design, there is a need for power supplies to interact with their loads in more profound ways than simply supplying a fixed voltage to their power buses.

1.3 Driving Application for Current Research

This research is motivated in part by the MIT Ultra Low Power Wireless Sensor Project (Figure 1-1). The goal of this project is to design, build, and test a sensor platform capable of operating over a wide range of data rates, consuming as little power as possible. Operation will be demonstrated with an imager, a low rate temperature sensor, and perhaps a CD-quality audio data stream. Components on the system include a reconfigurable A/D converter, a DSP core to provide compression and encryption, a $2GHz$ transmitter, and a low data rate control receiver. Data rates vary from tens of bits per second to 1 million bits per second, and system power consumption varies from $50mW$ to $100\mu W$. The total system volume is approximately one cubic inch, of which one-third is allowed for the power system (including battery). The project will help to demonstrate optimization for
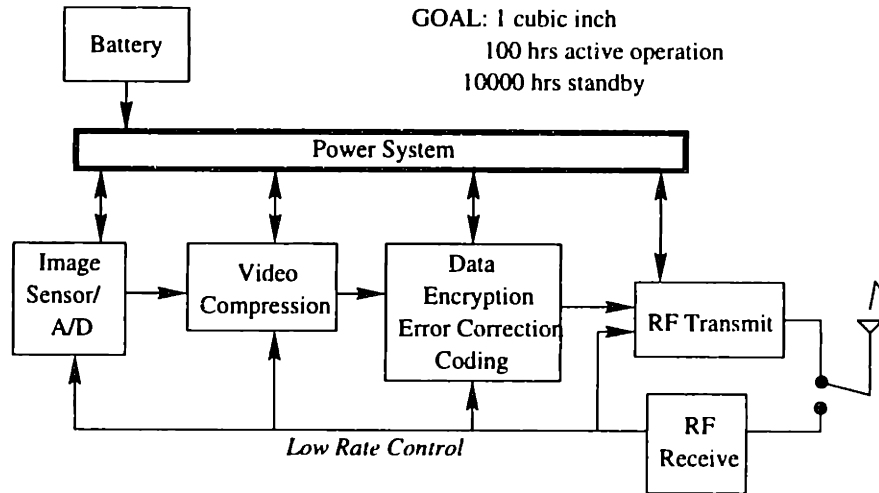
Figure 1-1: Block diagram of MIT Ultra Low Power Wireless Sensor Project

low power at the system level by negotiating the interfaces between components. This negotiation includes the choice of voltage supplies for each of the blocks, as well as the possibility of implementing a variable supply voltage scheme for certain blocks.

1.4 Thesis Scope

This document addresses the optimization of switched capacitor and inductor based voltage converters (frequency selection, filter sizing). Equations for making the tradeoff between converter volume and battery volume are described, and a method for evaluating run time improvements for arbitrary battery discharge characteristics is derived. Design issues related to inductor based switching converters are addressed qualitatively and quantitatively, such as switch sizing and techniques for decreasing gate drive power. Circuit designs for low power digital control circuitry for switching converters are proposed and discussed, including circuits which accomplish feedback on state variables besides output voltage (such as circuit delay). Results from a fabricated chip implementing some digital control techniques are presented. Finally, the possibility of using micromechanical structures for fully integrated voltage conversion, is explained, and a sample design of this type of converter is developed.

Chapter 2

Techniques for Voltage Regulation

The ability to regulate output voltage can increase battery life by allowing the optimization of the supply voltage to the function of a circuit. Regulation also allows us (in some cases) to decouple the choice of the battery technology from the choice of the supply voltages of the electronics. (A qualification is added because if linear regulators are used, the battery voltage must be greater than the output voltage.) The purpose of this chapter is to provide an overview of the available techniques for voltage regulation. Detailed discussion of the design of the more complex options, inductor based switching regulators and micromechanical switching regulators will be relegated to later sections.

Power supply voltage regulation is typically achieved by either linear regulation, switched capacitor voltage boosters or dividers, or switching regulators utilizing inductors and capacitors. A power converter design of any of these types makes trade-offs between often conflicting goals of high efficiency, low cost, small size, and high quality output power. For very low power mobile applications, efficiency is an issue because of the desire to increase running time with a given power source. Reducing costs for low power applications is often manifested as a need to reduce part count and simplify construction. Small size for these application is not a strict $Watts/m^3$ metric, but the goal of integrating the entire power converter into a single, small, cheap package with no external parts. The quality of the output power is generally measured by the output ripple and the transient response. For low power applications, the output ripple requirements are getting tighter as voltage supplies decrease to $1V$ and below, and transient responses are becoming more important as techniques such as variable voltage supplies and selective powering down are used.

For battery powered applications, isolating the input power source from the output load or system package is usually not an issue, since the cell voltages are not dangerously large. Furthermore, although separate ground paths will be required by sensitive subsystems, there is generally no need to provide isolated output voltages. Because isolation is not required, the power supply design is somewhat simplified, and techniques such as linear regulation are feasible.
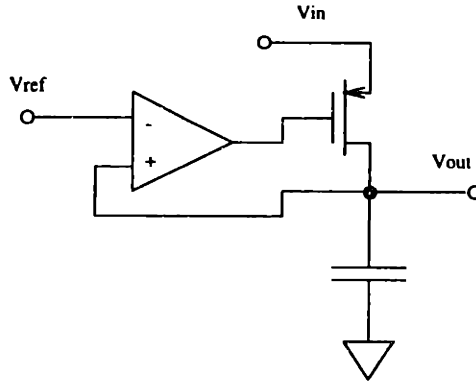
Figure 2-1: A Linear Regulator

2.1 Linear Regulation

A linear regulator is simply a device which maintains a fixed output voltage with a varying input voltage by controlling the V_{DS} (or V_{CE}) of a transistor with the gate (or base) input. The difference between the desired output voltage and the available input voltage is dropped across the transistor in the linear regulator. As such, this circuit is inherently lossy, its power conversion efficiency is merely $\frac{V_{out}}{V_{in}}$. Of course, linear regulators are limited to applications where the power source is at a higher voltage than the load. The voltage difference between the input and the output must be greater than the drop-out voltage, the saturation voltage of the power transistor.

At high output powers (greater than $100mW$), linear regulators may be highly undesirable due to the necessity to provide heat sinking sufficient for the worst case conditions. At fairly low powers, on the other hand, heat dissipation is not difficult, and linear regulators are extremely simple to implement. They are readily available in a single chip, with no external components. Furthermore, for certain battery and load conditions, a linear regulator may provide a significant increase in battery run time. This increase could be quite attractive relative to the minor cost and size of the circuit. Linear regulators can also provide variable voltage supplies with fast response times quite easily.

Linear regulation can provide an increase in run time if the battery voltage is larger than the minimum voltage required by the load, and the load current increases with supply voltage. Consider a system consisting of a battery with a discharge characteristic which varies from $3.5V$ to $2V$, and a CMOS load (modeled as a resistor) which only requires $2V$ for proper operation. A plot of the hypothetical discharge characteristic is shown in Figure 2-2. (Note that the proposed discharge curve models a battery as a linear capacitor which stops operating below a certain voltage.) If the $Amp \cdot hour$ capacity of the battery is Q, the running time of the system with no linear regulator is $\frac{RQ}{V_{max}-V_{min}} ln \frac{V_{max}}{V_{min}}$, or $0.37RQ$. The running time of the system with a linear regulator is given by RQ/V_{min}, or $0.5RQ$. For this example, adding a linear regulator (which will dissipate an appreciable fraction of the total battery energy, 27%) extends the system run time by 34%. A more detailed analysis of this effect is included in Section 2.5.1 (page 24), and some consideration of this may also
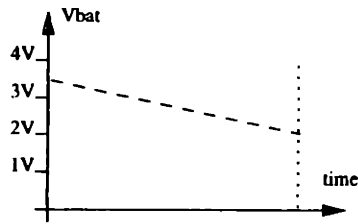
Figure 2-2: Idealized battery discharge curve under constant current load.
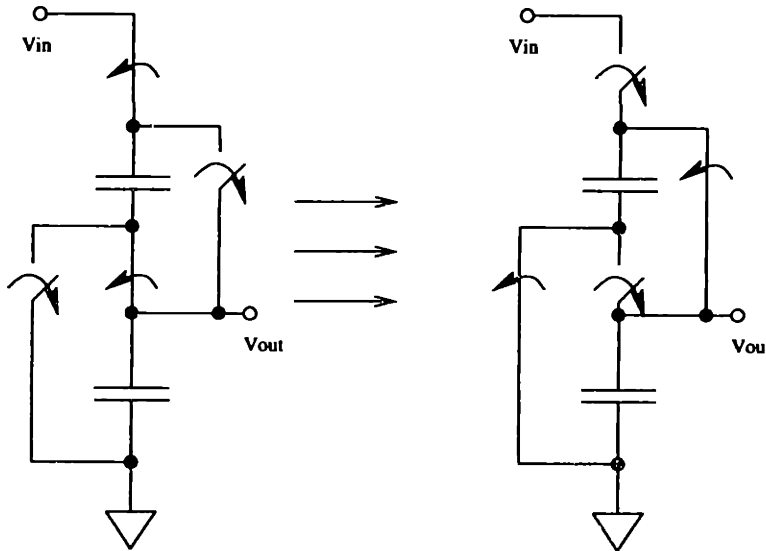
Figure 2-3: A Switched Capacitor Voltage Divider

be found elsewhere [2].

Due to the truly minimal cost of a low power linear regulator, the increases in battery run time due to any other proposed power circuit must be compared to the benefits of a linear regulator, not simply the case where no power conversion is done.

2.2 Switched Capacitor Circuits

Switched capacitor circuits operate by alternately connecting capacitors in parallel and in series, as shown in Figure 2-3. Two capacitors connected in series to an input voltage can then be arranged in parallel with a load to create an output voltage which is nominally one-half of the input voltage. When the capacitors are placed in series, they are each charged to a fraction of the input voltage, and when the capacitors are in parallel, that fractional voltage is presented to the load. These circuits are found in very low power applications such as watches. Basic discussions of their operation and losses can be found in [23, 21].

Switched capacitor techniques can theoretically be used to create any rational voltage ratio, although creating ratios besides ÷2, ×2, ÷3, and ×3 may be of little value because of the extra losses due to the number of switches such ratios require. Boosters and dividers have the disadvan-

15
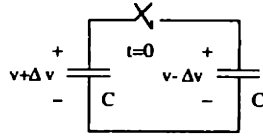
Figure 2-4: Two capacitors charged to different values, shorted together, create a net loss of energy.

tage that zero voltage switching is impossible, which limits our ability to decrease losses in these converters. Zero voltage switching is accomplished by exploiting lossless inductor–capacitor ringing, and activating a switch when the voltage across it reaches zero. There are no inductive elements in a switched capacitor circuit, so this technique is not available. Another problem with switched capacitor circuits is that they have an inherent output resistance, which compounds the problem of precisely regulating the output voltage. This output resistance is related to the frequency of operation and the size of the capacitors used.

Switched capacitor voltage boosters and dividers are more complex than linear regulators, but given that they only use switches and capacitors, it is still possible to integrate a low power converter on a single integrated circuit. The boosters and dividers have the advantage (over linear regulators) that they can have fairly high efficiency without limiting $V_{out} \sim V_{in}$. On the other hand, capacitive converters do not allow tight voltage regulation. Some degree of regulation is possible by dynamically varying the output resistance (by changing the switching frequency), but this will also change the supply voltage ripple. Combining a variable ratio switched capacitor circuit with a voltage regulator is one possible, full-featured voltage converter, whose merits should be evaluated. Such a combination is the basis for some commercial low power converters. The primary issues for analysis in switched capacitor circuits are power density and losses.

2.2.1 Losses in ideal Switch Capacitor Converter

A switched capacitor converter is not an ideally lossless power converter. This means even if the power required to drive the switches and the resistance of the switches were zero, the converter would still dissipate a finite amount of power. This can be demonstrated by considering the losses involved in shorting together two capacitors charged to different levels, as shown in Figure 2-4.

Capacitors are charge storing elements, so the total charge on the positive plates of the capacitors of Figure 2-4 will be the same before the switch is closed and for all time after the switch is closed. If the initial charge on one capacitor is $v + \Delta v$, the initial charge of the second capacitor is $v - \Delta v$, and the final equilibrium voltage is v_{eq}, we can write the charge conservation as

$$C(v + \Delta v) + C(v - \Delta v) = 2Cv_{eq}.$$ (2.1)

We find then that:

$$v = v_{eq}. \tag{2.2}$$

Given the initial and final values of the voltages on the capacitors, an accounting of the energy in the system will show the loss of energy caused by shorting the two capacitors together. Before the switch is closed, the energy in the system is:

$$E_{initial} = \frac{1}{2}C\left[(v + \Delta v)^2 + (v - \Delta v)^2\right] = \frac{1}{2}C\left(2v^2 + 2\Delta v^2\right). \tag{2.3}$$

After the switch is closed and equilibrium is reached, the energy in the system is:

$$E_{final} = \frac{1}{2}2Cv^2 \tag{2.4}$$

The net loss of energy in the system is $C\Delta v^2$. This result can also be shown by calculating the power dissipation in a switch of finite on-state resistance for the same circuit. The dissipation in the switch will be $C\Delta v^2$, independent of the resistance of the switch. Hence, in the limit as the resistance approaches zero, the circuit dissipation is still $C\Delta v^2$.

2.2.2 Power Density of Switched Capacitor Converters

In order to compare the usefulness of switched capacitor converters relative to other types of power converters, a detailed analysis must be made of the losses in such converters. An appropriate metric (as discussed elsewhere in this thesis), is a function which describes the minimum power dissipation over a range of converter volumes for a given set of load conditions and device characteristics. This function will allow making trade-offs between the system battery and the converter volume, and identifying what type of converter is suitable for different levels of miniaturization.

In my investigation of the dependencies of the power dissipation in a switched capacitor converter, I analyzed a $\div 2$ converter, where the two capacitors were of equal value, and the load presented is a current source. Setting the two capacitors to be of equal value simplifies the analysis. A constant current load also simplifies the analysis and is an accurate representation of the load in a low output voltage ripple condition. The switches are modeled as either resistors or open circuits, where the gate capacitance which must be driven to switch states is proportional to $\frac{1}{R_{sw}}$. The simplest analysis of switched capacitor converters utilizes ideal switches to demonstrate that the circuit is not lossless (see Section 2.2.1). This analysis ignores additional losses caused by the load current flowing through the switches, as well the effects caused by the condition that the time constant with which the capacitors charge and discharge may not be much shorter than the switching interval.

The two topologies which the circuit realizes in a switching cycle are shown in Figure 2-5. R represents the series combination of resistances of two switches in their on state ($R = 2R_{sw}$). T is

17
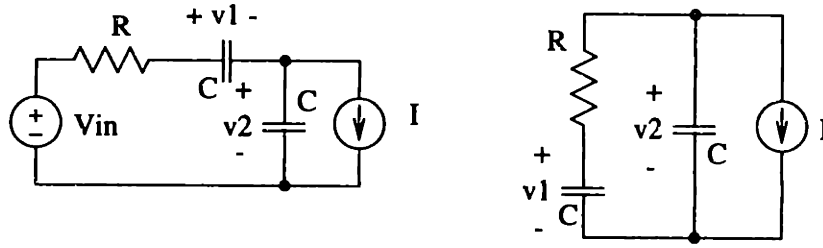
Figure 2-5: Two topologies of switched capacitance converter

the time the circuit remains in each configuration, so the actual switching frequency of each of the four necessary switches is $\frac{1}{2T}$. For $0 + 2nT \le t \le T + 2nT$, the circuit is configured as shown on the left in Figure 2-5, and for $T + 2nT \le t \le 2nT$, the circuit is configured as shown on the right in Figure 2-5, where n is an integer.

Rather than deriving the solutions to the differential equations of the two circuits shown in Figure 2-5 with the state variables v_1 and v_2, using differential and common mode state variables somewhat simplifies the task. The new variables are defined by:

$$v_c = v_1 + v_2 \tag{2.5}$$

$$v_d = v_1 - v_2 \tag{2.6}$$

The resulting differential equations are:

$$
\begin{cases}
\frac{C}{2}\frac{dv_c}{dt} = -\frac{v_c}{R} - \frac{I_l}{2} + \frac{V_{in}}{R} & 0 \le t \le T \\
C\frac{dv_d}{dt} = I_l \\
\\
C\frac{dv_c}{dt} = -I_l & T \le t \le 2T \\
\frac{C}{2}\frac{dv_d}{dt} = -\frac{v_d}{R} + \frac{I_l}{2}
\end{cases}
\tag{2.7}
$$

These equations lead to the following time-domain solutions for v_c and v_d:

$$
\begin{cases}
v_c(t) = v_c(t=0) + \left(V_{in} - \frac{I_l R}{2} - v_c(t=0)\right)\left(1 - e^{-\frac{t}{\tau}}\right) & 0 \le t \le T \\
v_d(t) = v_d(t=0) + \frac{I_l}{C}t \\
\\
v_c(t) = v_c(t=T) - \frac{I_l}{C}(t-T) & T \le t \le 2T \\
v_d(t) = v_d(t=T) + \left(\frac{I_l R}{2} - v_d(t=T)\right)\left(1 - e^{-\frac{(t-T)}{\tau}}\right)
\end{cases}
\tag{2.8}
$$

where τ is given by $\frac{RC}{2}$. In steady state, by definition, $v_c(t=0) = v_c(t=2T)$ and $v_d(t=0) =$

18

$v_d(t = 2T)$. Using the above equations and these two relationships, we can find the values of v_d and v_c at times $t = 0, T, 2T$. These results are:

$$v_c(t = 0) = v_c(t = 2T) = V_{in} - \frac{I_l R}{2} - \frac{I_l T}{C} \frac{1}{1 - e^{-\frac{T}{\tau}}}$$

$$v_c(t = T) = V_{in} - \frac{I_l R}{2} - \frac{I_l T}{C} \frac{e^{-\frac{T}{\tau}}}{1 - e^{-\frac{T}{\tau}}}$$

$$v_d(t = 0) = v_d(t = 2T) = \frac{I_l R}{2} + \frac{I_l T}{C} \frac{e^{-\frac{T}{\tau}}}{1 - e^{-\frac{T}{\tau}}}$$ (2.9)

$$v_d(t = T) = \frac{I_l R}{2} + \frac{I_l T}{C} \frac{1}{1 - e^{-\frac{T}{\tau}}}$$

The behavior of the output voltage can be derived from these results, combined with Equation 2.5 and Equation 2.6. The following results characterize the output voltage, v_2.

$$v_2(t = 0) = v_2(t = T) = v_2(t = 2T) = \frac{V_{in}}{2} - \frac{I_l R}{2} - \frac{I_l T}{2C} \left(\frac{1 + e^{-\frac{T}{\tau}}}{1 - e^{-\frac{T}{\tau}}} \right)$$ (2.10)

$$\overline{v_2} = \frac{V_{in}}{2} - \frac{3}{4} I_l R - \frac{I_l T}{4C} \left(\frac{1 + e^{-\frac{T}{\tau}}}{1 - e^{-\frac{T}{\tau}}} \right)$$ (2.11)

Note that Equation 2.10 also represents the minimum of the output voltage in steady-state.

The above information can be used to construct a design which is optimized to reduce power consumption. First, however, an appropriate metric must be chosen. Because a switched capacitor converter can not regulate its output voltage, it is often desirable to process the output power with a linear regulator. (As discussed elsewhere, this is particularly helpful if the load acts resistive, such as CMOS logic.) In the case where the output of the switched capacitor converter receives no further passive filtering, and is followed by a linear regulator, the minimum output voltage is an important characteristic. For such a situation, the best case power dissipation for the combination of the capacitive converter and the linear regulator is $\left(\frac{V_{in}}{2} - v_{2(min)} \right) I_l + P_{switches}$, assuming a zero dropout voltage linear regulator. (Note that it can be shown that the input power of the above converter is $\frac{V_{in}}{2} I_l$.) This is the metric minimized below. Minimization of the power dissipated in only the capacitive converter would be accomplished with the metric $\left(\frac{V_{in}}{2} - \overline{v_2} \right) I_l + P_{switches}$. A comparison of Equation 2.10 and Equation 2.11 reveals that these expressions differ only in the values of two constants; the minimizations will be qualitatively identical.

In an extremely low power system, the actual die size of the switches used will be very small relative to the size of the load die and the die packaging. Therefore, it is convenient to assume the volume of the capacitors of the switched capacitor converter dominate the total volume of the power circuit. Because we seek to minimize the power consumption for a given power supply volume, I will assume that the capacitors are chosen to take up the entire given volume, and the design variables are the size of the switches and the switching frequency selected. The switching losses in the switches

19

are assumed to be inversely proportional to the size of the switches, as prescribed by the formula

$$P_{switches} = \frac{C'}{2R_{sw}}V_g^2 Nf. \tag{2.12}$$

For the circuit described above, Nf is $\frac{4}{T}$. V_g is the MOSFET switch gate voltage. C' is a property of the semiconductor technology used to fabricated the switch. Although the notation implies that only gate losses are counted, other parasitic capacitive losses could also be accounted for in the above expression, as long the capacitance in question is proportional to switch device area. The total losses which we would like to minimize are:

$$P_{diss}(R_{sw}, T) = I_l^2 R_{sw} + \frac{I_l^2 T}{2C}\left(\frac{1 + e^{-\frac{T}{\tau}}}{1 - e^{-\frac{T}{\tau}}}\right) + \frac{NC'V_g^2}{2R_{sw}T} \tag{2.13}$$

R_{sw} is the resistance of a single switch, τ is the time constant from the differential equations, $R_{sw}C$, N is the number of switch transitions in the time T, and C' and V_g are switch parameters. The first term represents the dissipation due to the load current flowing through the switches, the second term represents the dissipation due to current flowing as the charge on the capacitors redistributes, and the third term represents the losses incurred during switch transitions.

A change of variables in Equation 2.13 can show that the value of $\frac{T}{\tau}$ at the minimum of this equation is a constant, independent of the other parameters in the equation. The new variables are $x = \frac{T}{\tau} = \frac{T}{R_{sw}C}$ and $y = R_{sw}T$. Rewriting Equation 2.13 in the new variables yields

$$P_{diss}(x, y) = I_l^2\sqrt{\frac{y}{xC}}\left(1 + \frac{x(1 + e^{-x})}{2 - 2e^{-x}}\right) + \frac{NC'V_g^2}{2y}. \tag{2.14}$$

In order to find the minimum of this equation with respect to x, we look for the roots of the partial derivative in x; that is,

$$-\frac{I_l^2\sqrt{y}\left(2 - 4e^{-x} + 2e^{-2x} - x + xe^{-2x} + 4x^2 e^{-x}\right)}{4x^{3/2}\sqrt{C}\left(-1 + e^{-x}\right)^2} = 0, \tag{2.15}$$

or

$$2 - x + 4e^{-x}\left(-1 + x^2\right) + e^{-2x}\left(2 + x\right) = 0. \tag{2.16}$$

Notice that the values of x at the roots of this equation are not dependent on N, C', C, V_g, or I_l. Equation 2.16 can be solved numerically using Newton's Method. The positive value of x which corresponds to a minimum of Equation 2.14 is 3.41037445342282 ...

The variable x gives a relationship between R_{sw} and T at the minimum of Equation 2.13, so this equation can then be rewritten in terms of one of these variables alone. For notational convenience,

20
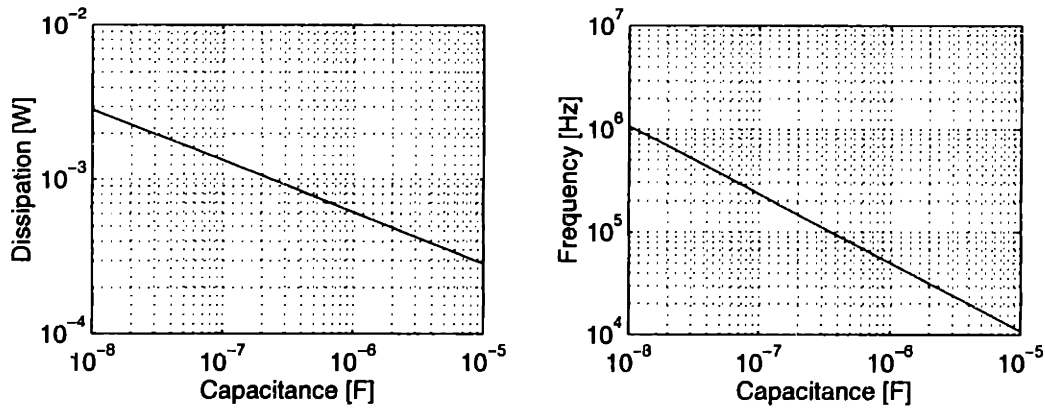
Figure 2-6: Minimum dissipation and desired operating frequency of capacitive halving voltage converter, using results of Equation 2.20 and Equation 2.21. The output current for this calculation is $5mA$, and the switch parameters roughly correspond to those found in a standard CMOS technology. The capacitance of the horizontal axis refers to the capacitance of each of the two capacitors in the converter.

set

$$\alpha = \frac{T}{\tau} = x_{min} \tag{2.17}$$

and

$$\gamma = \left(\frac{1 + e^{-\frac{x}{\tau}}}{1 - e^{-\frac{x}{\tau}}} \right) . \tag{2.18}$$

R_{sw} can now be expressed as a function of T,

$$R_{sw} = \frac{\tau}{C} = \frac{T}{\alpha C} \tag{2.19}$$

and the minimization becomes a function of T alone. The resulting minimization is

$$T = \left(\frac{2C'NV_g^2 C^2}{I_l^2} \frac{\alpha^2}{2 + \alpha\gamma} \right)^{\frac{1}{3}} \tag{2.20}$$

$$P_{diss} = \frac{3}{4} \left(\frac{I_l^4 C'NV_g^2}{C} \frac{2(2 + \alpha\gamma)^2}{\alpha} \right)^{\frac{1}{3}} , \tag{2.21}$$

where α is a constant equal to $3.410374\ldots$, and γ is about 1.06831399. Plots of Equation 2.20 and Equation 2.21 over a range of capacitances are shown in Figure 2-6.

Besides serving as simply a design guide for switched capacitor converters, Equation 2.21 gives the order of the relationships between the power dissipation and the capacitance used (which is proportional to the converter volume). If we consider how the capacitance needed scales with the load current, at constant output voltage and efficiency, we find

$$C \propto I_l, \tag{2.22}$$

21

which indicates that volume scales linearly with output current. Another interesting relationship to note is that between the switch parameters $C'V_g^2$ and necessary converter capacitance. At fixed load and efficiency

$$C \propto C'V_g^2, \tag{2.23}$$

meaning that a decrease in the switching losses by a factor of two (due to improved semiconductors) allows a reduction in the volume of the converter by a factor of two, for the same amount of losses.

2.3   Inductor Based Switching Regulators

Inductor based switching regulators can be used to create precise output voltages, above or below their input voltage, and have the additional advantage that they are ideally lossless power converters. Inductor based power supplies can completely decouple the battery voltage (and its droop characteristics) from the load voltage. The load is presented with a fixed supply voltage, while the regulator imposes a constant power drain (assuming the load draws a constant power) to the source. Such converters are also readily adaptable to variable voltage outputs.

Scaling suggests that the conventional down, up/down, and up converters do not have problems scaling to very low powers, and this can be verified with actual low power converters. The essential challenges with such designs are finding ways to construct and package such devices in a way which limits volume and cost, and developing designs which will operate over a wide range of conditions efficiently. A detailed description of the appropriate tradeoffs for minimum volume design and techniques to enhance efficiency over broad ranges of operating conditions is reserved for a later chapter (Chapter 3). This thesis only considers non-resonant converter topologies. Resonant converters are typically designed to save some of the switching losses, at the expense of more complex control. Whether such tradeoffs are worthwhile for very low output powers remains to be evaluated; it may be that the added control complexity requires more power consumption than could be conserved through soft-switching transitions.

Switching power supplies topologies can be considered derivatives of a single canonical switching cell [10, 9]. The canonical cell and the three common topologies are shown in Figure 2-7. The steady state relationship between the input voltage and the output voltage for these topologies is shown in Table 2.3. The reader is referred elsewhere [9] for a detailed explanation of how these converters

Converter Type	V_{out}/V_{in}
Down	D
Up	$\frac{1}{D}$
Up/Down	$-\frac{D}{1-D}$

Table 2.1: Steady state relationships between V_{in} and V_{out} for standard DC-DC converter topologies.
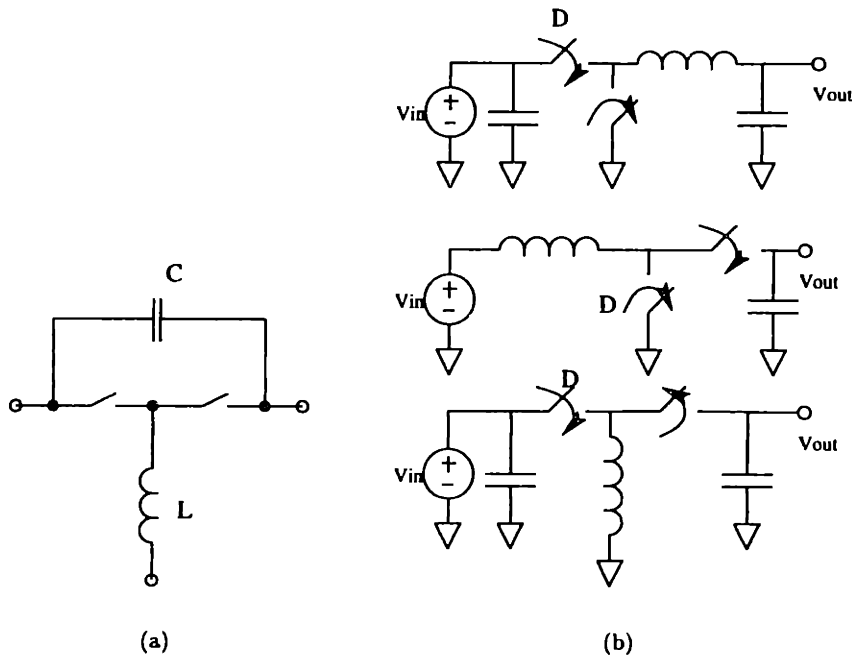
Figure 2-7: (a) Canonical cell for DC-DC converters. At most one switch is closed at any time. (b) Standard DC-DC power converters: down, up, and up/down. Active switch labeled with on-time duty cycle D.

operate.

2.4 Electromechanical Power Converters

It is currently very difficult to construct acceptable inductors on silicon wafers which incorporate other circuitry (some work in this area can be found in [26, 25, 1]). As a result, inductor based power converters are not fully integratable. A topology with the advantages of inductor based regulators but does not require an inductor clearly would be extremely useful.

Switching power converters can process power between two voltage sources by storing energy in the magnetic field of an inductor. Since we cannot losslessly transfer power between voltage sources with only capacitors, we must find identify another, (non-electrical) type of energy storage if we are to replace inductors. It has been proposed [17] that storing energy in a resonant mode of a mechanical diaphragm, and using the motion of the diaphragm to create variable capacitance capacitors could allow the construction of an inductor-less DC-DC converter.

The mechanically coupled capacitors formed on a resonant diaphragm can be used to transfer energy losslessly between two voltage sources. Such a structure has the capability to precisely control the conversion voltage ratio and to convert power from high voltages to lower voltages or the other way around. The proposed structure (Figure 5-1) should be easier to fabricate in a manner

compatible with standard devices than the inductors which have been integrated to date, offering the possibility of a single chip solution.

Although no functional power converter based on a micromechanical resonator has been reported in the literature, the possibility of such devices is intriguing. The paper design of such a converter is presented in Chapter 5.

2.5 Comparing Designs

The goal of the converters described above is to provide a controlled output voltage, despite changing input or load conditions. As a result, we cannot quantify the performance of a converter by a single quantity such as the "efficiency." The converters described each perform best within certain regions of operation. Which converter performs best within a given system is a function of which regions of operation are exercised by the system.

A natural metric for comparison for a battery powered application is the run time of the entire system for a given initial battery charge. Of course, the discharge cycle of a battery for a given system will not be the same as the workload will vary. This can be quantified by comparing performance for various discharge characteristics.

2.5.1 Effect of battery and load characteristics

The discharge characteristic of the system battery will change the relative merits of the different types of voltage regulation (no regulation, linear regulation, and ideal voltage regulation). This was mentioned previously in Section 2.1, to demonstrate how linear regulation can improve system running time, under the assumption of a linear discharge curve. The run time extension given a linear discharge curve is also discussed in [2]. A more rigorous, general analysis of the effect of the discharge characteristic on the type of regulation is possible. The analysis proposed below makes the assumption that the amount energy of which can be used in a battery is independent of the loading of that battery during discharge. Certainly this is not the case [13]; in reality a fast discharge rate will result in less energy output due to losses internal to the battery, and an extremely slow discharge rate will result in lower energy output due to internal battery leakage. However, the assumption that energy is independent of loading will be true if the load current is not varied a great deal from the reference loading, where the reference point is the reference loading is that used to measure the battery discharge characteristic. This assumption allows us to make approximate calculations without using a complicated battery model.

Consider the function $q(t)$ to be the charge removed from a battery up until time t. When the battery is drained $q(t_{drained}) = Q_o$, where Q_o is the $A \cdot hr$ capacity of the battery. The function

Load Type	Discharge Function $i(q)$	t_o for $v(q) = V_{max} - \frac{V_{max}-V_{min}}{Q_o}q$
Resistive	$\frac{v(q(t))}{R_{load}}$	$\frac{RQ_o}{V_{max}-V_{min}}ln\frac{V_{min}}{V_{max}}$
Linear Regulator (Constant Current)	I_{load}	$\frac{Q_o}{I_{load}}$
Ideal Voltage Regulator (Constant Power)	$\frac{P_{load}}{v(q(t))}$	$\frac{Q_o}{P_{load}}\frac{V_{max}+V_{min}}{2}$

Table 2.2: Appropriate values for $i(q)$ with different load types.

$q(t)$ is given by the integral

$$q(t) = \int_0^t i(q(t'))dt' \tag{2.24}$$

where $i(q)$ is given by the battery discharge curve and the type of load. The battery discharge curve is given by $v(q)$, which shows the drop in cell voltage over time given a constant current load (e.g. Figure 2-2). This is the typical way to measure battery discharge, and for this case, $q = I_{load}t$. The proper equations for $i(q(t))$ are shown in Table 2.2.

The integral of Equation 2.24 can be converted to a differential equation by differentiating both sides with respect to t,

$$\frac{dq(t)}{dt} = i(q(t)) \tag{2.25}$$

which may be solvable analytically (depending on the function $i(q)$), but could certainly be solved numerically. After $q(t)$ has been found, the system running time is found by solving for t_o in the equation

$$q(t_o) = Q_o. \tag{2.26}$$

In fact, it is not necessary to solve the differential equation for $q(t)$ to find t_o, as shown in an example below. For the three load types shown in Table 2.2, solutions for t_o are also given, for the case of a linear discharge curve $v(q)$.

As an example of the application of the method described above, consider the discharge equation

$$v(q) = b - aq^{\frac{1}{2}}. \tag{2.27}$$

When a battery described by this equation is loaded with a regulating power supply with some fixed efficiency, the formulation of Equation 2.25 is

$$\frac{dq(t)}{dt} = \frac{P_{load}}{\eta\left(b - aq(t)^{\frac{1}{2}}\right)}. \tag{2.28}$$
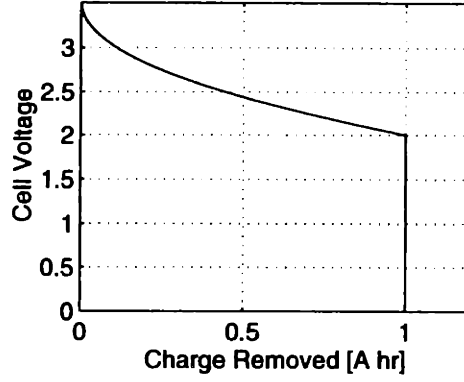
Figure 2-8: Discharge curve $v(q) = b - aq^{1/2}$, with $b = 3.5$ and $a = 0.025$.

Isolation of the variables of this equation yields the implicit integral

$$\left(b - aq^{\frac{1}{2}}\right) dq = \frac{P_{load}}{\eta} dt,$$ (2.29)

which evaluates to

$$bQ_o - \frac{2 Q_o^{3/2} a}{3} = \frac{P_{load} t_o}{\eta}$$ (2.30)

over the ranges $q = [0, Q_o]$ and $t = [0, t_o]$. From this equation, it is trivial to find the value of t_o,

$$t_{opow-sup} = \frac{\eta}{P_{load}} \left(bQ_o - \frac{2 Q_o^{3/2} a}{3}\right).$$ (2.31)

For the same discharge characteristic, if the battery is loaded with a linear regulator, it is straightforward to find t_o without using Equation 2.25, since the discharge curve is measured under the constant current state, and we have made the assumption that the total battery charge is constant regardless of the output loading. The run time for the linear regulator loading case is

$$t_{olin-reg} = \frac{Q_o}{I_{load}}.$$ (2.32)

Now consider $a = 0.025 (V/s)^{\frac{1}{2}}$, $b = 3.5V$, and $Q_o = 3600As$. This represents the case of a $1A \cdot hr$ battery whose voltage sags from a peak of $3.5V$ to a minimum of $2V$, as shown in Figure 2-8. Also consider the load for this comparison to be equivalent to a $1k\Omega$ resistor, which requires a minimum supply voltage of $V_{DD} = 1.8V$. In this case, $P_{load} = 3.24mW$ and $I_{load} = 1.8mA$. Under this set of conditions, we find

$$t_{opow-sup} = \eta \cdot 2.78 \times 10^6 s$$

and

$$t_{olin-reg} = 2.00 \times 10^6 s.$$

26

This analysis predicts that a power regulator with at least 72% efficiency will provide a longer run time than a linear regulator. Under the set of conditions described, forgoing any sort of voltage regulation (connecting the $1k\Omega$ resistor to the battery) would result in a run time

$$t_o = 1.47 \times 10^6 s,$$

based on the numerical integration of the appropriate equations. Linear regulation would provide a 36% extension of run time, while lossless regulation would provide an 89% extension of run time.

2.5.2 Trading off efficiency and volume

The designers of a portable system are free to choose the nature of their energy source and their power converter. This flexibility enables the designers to make trade-offs between volume of the battery and the volume of the converter in order to maximize the actual system running time. Power converters will generally increase in efficiency with increases in volume. Assuming a limited volume for the entire system, there is a point of diminishing returns, where increases in the volume of the power converter will actually decrease the system run time because the battery volume must decrease to keep the total volume constant. This relationship can be expressed quantitatively, which assists in optimization at the system level.

The choice of the relative volumes of the system battery and the power supply should be dictated by a maximization of the total system running time. If we assume that the energy density, S, of the battery is a constant with respect to volume, and the total volume (including the battery and the power supply) is fixed, we can write:

$$T_{run} = SV_{battery} \frac{1}{P_{system} + P_{diss}(V_{supply})} \tag{2.33}$$

or,

$$T_{run} = S(V_{total} - V_{supply}) \frac{1}{P_{system} + P_{diss}(V_{supply})} \tag{2.34}$$

where P_{system} is the average system power dissipation and $P_{diss}(V_{supply})$ is the average power supply power dissipation as a function of its volume. This expression can then be maximized with respect to V_{supply}. At the maximum, we find that

$$0 = \frac{dT_{run}}{dV_{supply}} = \frac{-S}{P_{system} + P_{diss}(V_{supply})} - \frac{S(V_{total} - V_{supply})}{(P_{system} + P_{diss}(V_{supply}))^2} \frac{dP_{diss}(V_{supply})}{dV_{supply}}. \tag{2.35}$$

This can be rewritten as

$$\frac{P_{system} + P_{diss}(V_{supply})}{V_{battery}} = -\frac{dP_{diss}}{dV_{supply}}. \tag{2.36}$$

A useful interpretation of Equation 2.36 is that the optimal battery size occurs when the power
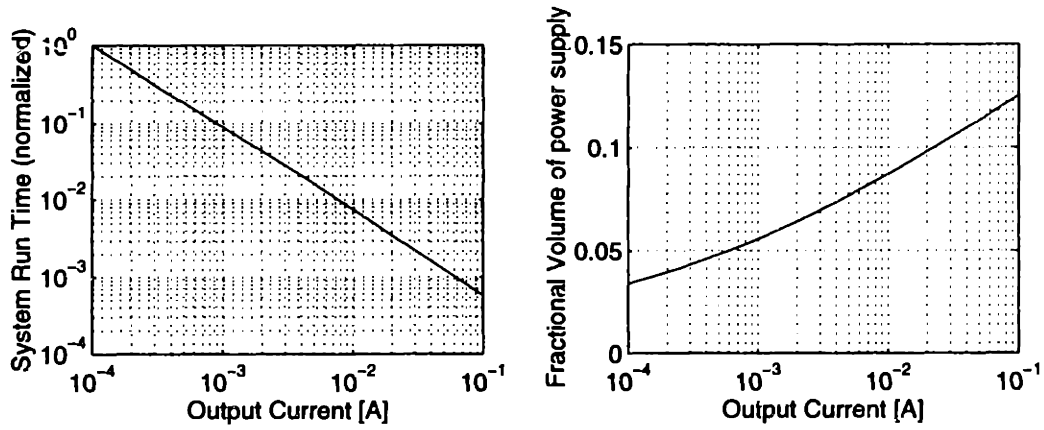
Figure 2-9: Numerical solutions to Equation 2.36 over a range of output currents for a switched capacitance converter as discussed in Section 2.2.2. On the graph on the right side, the volume of the power supply is expressed as a fraction of the total volume of the battery and the power supply. For the calculations shown, the output voltage was chosen to be $2V$. Run time decreases faster than output power increases, due to reduced efficiency as the power density increases. Relative volume of the power supply increases with output power to offset this effect.

density of the battery is matched by the incremental decrease in converter dissipation per unit volume. The system running time will be the longest for a given battery, power supply, and volume if the battery density is xW/m^3 and the dissipation of the power supply will be reduced xW per unit increase in volume.

For the example of the switched capacitor converter considered in Section 2.2.2, Equation 2.21 allows a basis for making the trade-off between battery size and converter size. Assuming that the volume of the capacitor C is proportional to its capacitance, this equation can be restated more simply as:

$$P_{diss} = \beta V_{supply}^{-\frac{1}{3}}. \tag{2.37}$$

Utilizing this relationship in Equation 2.36 yields

$$\frac{P_{system} + \beta V_{supply}^{-\frac{1}{3}}}{V_{battery}} = \frac{1}{3}\left(\beta V_{supply}^{-\frac{4}{3}}\right) \tag{2.38}$$

which simplifies to the equation

$$\frac{3P_{system}}{\beta} = V_{total}V_{supply}^{-\frac{4}{3}} - 4V_{supply}^{-\frac{1}{3}}. \tag{2.39}$$

This equation can be solved numerically under some specific set of conditions to assist in the design of the relative volumes of the battery and power supply. Figure 2-9 shows a set of such solutions. Notice that as the output power increases, the relative volume of the power supply should also increase, to maintain efficiency at the expense of some battery volume.

28

2.5.3 Providing Multiple Outputs

As pointed out earlier, mobile systems quite often require several different voltage outputs. Therefore, it is desirable to develop topologies for each of the converter types which are amenable to having multiple outputs, preferably with independent controls. In many cases simply having parallel converters with different outputs voltages will be the best we can do. Some information about the system is helpful in the determination of the multiple output solution. Switched capacitor converters can be exploited if one of the output voltages is an integral multiple of another, or an integral multiple of a divided value (that is, if a $\times \frac{3}{4}$ ratio is used, a $\times \frac{1}{2}$ output could be added easily). As is the case in any switched capacitor converter, regulation of the output voltage directly is impossible.

Multiple outputs can be added to an up/down converter by putting additional coils on the magnetic core. Unfortunately, all of the output voltages are determined by the voltage on just one of the outputs. Additional regulation in the form of linear regulators or extra filtering (more inductors) is required to attain independently controlled outputs. If an up/down converter is not otherwise required by the system, this solution also will incur the additional losses associated with the higher current and voltage stresses inherent in the up/down converter.

A second way to provide multiple outputs in the inductor based topologies is to time-multiplex a single inductor among multiple outputs. Although this has the advantage of using a single inductor, the volume required by this inductor to avoid core saturation will exceed the sum of the volumes of the inductors used if completely independent converters were used. Additional drawbacks of this approach are that the peak current stresses in the power switches will be significantly increased, and the required output capacitances will also be increased.

An advantage of a micromechanical converter, as discussed in Chapter 5 is that multiple outputs can be readily accommodated. The multiple output converter can be implemented as either a single input stage with multiple output stages, or several parallel power stages. The necessary area for either of these approaches is the same, and the preferred implementation probably depends on more subtle design issues. The problem with parallel inductor based supplies, namely that it is difficult and expensive to produce and connect several inductors, does not plague the micromechanical converter because all components can be integrated.

Chapter 3

Inductor Based Power Supplies

Power converters using inductors to provide ideally lossless DC-DC conversion are the mainstay of power supply design for output powers in the range of hundreds of milliwatts and above. Extending the usefulness of these devices to extremely low powers (from $100mW$ to $100\mu W$ and below) requires a careful examination of the scaling of the components and the design of the control circuitry. For high output power applications, the design of the control circuitry does not need to be focused on reducing its power consumption. However the consumption of these traditional controllers can easily exceed the output powers for the low power systems of interest.

This chapter will address output filter design, power switch sizing and driving, control techniques, and various circuits to implement the control. Anywhere that there is difference in the relevant analysis among the converter topologies the analysis is carried out for a down converter. This choice reflects the need to create supply voltages for low voltage digital logic, where the output voltage is exclusively lower than the battery voltage. Despite the loss in generality, the fundamental methods and conclusions are applicable to all converter topologies.

3.1 Filter Sizing for Inductor-Based Converters

For any active power regulator, the operating frequency dictates the necessary cycle to cycle energy storage, which in turn controls the volume necessary for the passive energy storage elements. Both capacitors and inductors have a limited energy density; for capacitors this is determined by the breakdown field, and for inductors this is determined by the saturation field in the magnetic material. (Saturation occurs when all magnetic domains have been aligned with the magnetic field.) For a power converter at a certain frequency, there is a minimum volume for the power filter (dependent on the quality of the capacitors and magnetic material chosen).

The choice of operating frequency is a balance between efficiency and power density. Low frequencies yield larger filters and higher efficiency, while higher operating frequencies allow the use

31

of smaller filters at the expense of a lower efficiency. A good method for choosing the operating frequency and output filter characteristics is to attempt to maximize converter efficiency at a given power supply volume. The result of this exercise is a function (either analytical or numerical) which relates power supply filter volume to the efficiency which can be achieved at that volume. This function can be used to make a trade-off between converter volume and battery volume, assuming a fixed volume for the total power supply system.

It can be generally stated that at a given power supply volume, the losses will be minimized when the switching frequency is chosen to be as low as possible. This assumption allows us to reduce the task of optimizing the efficiency to one of finding the lowest achievable switching frequency for a given supply volume. The power supply volume is composed of the inductor and capacitor volumes, and the volume of the power switches and control circuitry. For the power levels and systems of interest, we can make the simplifying assumption that the volume of the power switches and control circuitry is constant. Although the optimal area of the power switches will increase as the switching frequency decreases, for such low powers, the total size of the packaged die containing the control circuitry and the power switches will vary little, if at all. Our goal then, is to find the lowest switching frequency achievable with a given output filter volume.

The optimization of the relative sizing of the inductor and capacitor to maximize efficiency is presented below. First the design of the inductor is chosen, then the relative sizing of the inductor and capacitor (as a function of frequency) are selected to minimize the peak flux density in the inductor core. Finally, the minimum switching frequency possible is chosen, based on the saturation flux density of the magnetic material selected.

3.1.1 Inductor Design

The inductors considered here are exclusively those with high permeance, ferrite, toroidal cores. High permeance magnetic materials limit where the magnetic flux flows, and provide a large value of inductance per turn. Ferrite cores, such as Mn-Zn and Ni-Zn are not conductive, so they do not require laminations, and are appropriate for high frequency ($100kHz$ to $10MHz$ operation). Toroidal cores afford the simplicity of being specified with only two geometry parameters, and are available in small sizes (circumference on the order of $1cm$).

A typical magnetic material can be modeled as being linearly magnetizable, with some fixed saturation flux density. We can assume that a magnetic material will provide an enhanced value of μ, provided that the peak magnetic flux density does not exceed B_{max}. The peak magnetic flux density is computed by:

$$B_{peak} = \frac{\mu N I_{peak}}{l_{mag}} \qquad (3.1)$$

where I_{peak} is the peak current in the inductor, N is the number of turns on the inductor, and l_{mag}
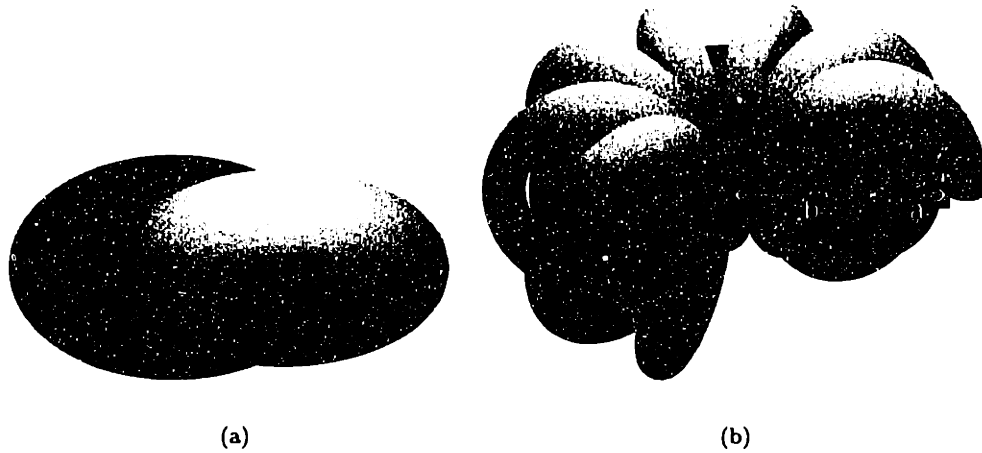
Figure 3-1: Diagram of toroid inductors. (a) shows simplified model, where the conductor is modeled as a single toroid, which may consist of N turns. (b) shows a more practical inductor, with the major and minor radii labeled as b and a, respectively.

is the mean magnetic path length.

The losses in a ferrite core can be modeled with the equation

$$P_{core}[W/m^3] = K_1 f^p B^q \tag{3.2}$$

were $K1$, p, and q are found empirically in the region of operation [14, 22, 4]. Typical values for these constants are shown in [22]; for ferrites p will vary between 1 and 3, while q varies from 2 to 3.

The geometry of a toroidal core is specified by a major radius and a minor radius (see Figure 3-1). For an inductor (core and conductor) of a given volume, the values of these radii specify the relative volumes of the core and the conductor. These relative volumes can be chosen based on the maximization of L/R, the ratio of the inductance achieved and the parasitic resistance of the inductor. This calculation is most easily carried out using the model shown in Figure 3-1(a). The core and the conductor are interlocked toroids, and the design of the inductor geometry is the choice of the values of the toroids' minor radii.

The inductor core is characterized by l_{mag}, the effective magnetic path length, and A_{mag}, the cross-sectional area of the core. An approximation for l_{mag} is $2\pi b$, where b is the major radius of the core. The conductor is also characterized by a length and an area, l_c and A_c. Note that $l_c = l_{mag}$ (since the major radii of the two cores coincide), and both quantities will henceforth be referred to as l. The inductance is specified by

$$L = \frac{\mu N^2 A_{mag}}{l}, \tag{3.3}$$

where N is the number of turns within the conductor, and μ is the core permeability. The resistance

33

of the conductor is

$$R = \frac{N}{A_c} \frac{Nl}{\sigma} \tag{3.4}$$

where σ is the conductance of the conductor material. There is additional power dissipation in the core material, which is given by Equation 3.2. Expressing the flux in terms of the current in the core, the total power dissipation is

$$P_{ind} = \frac{N^2 l}{\sigma A_c} I_{rms}^2 + K_1 \left(\frac{\mu N I_{rms}}{l}\right)^q f^p A_{mag} l. \tag{3.5}$$

We now wish to maximize the quantity L/P_{ind}, by varying A_c and A_{mag}. This can be done by defining the quantities $x = \sqrt{A_{mag}}$ and $w = \sqrt{A_c} + x$. The minor radii of the two toroids are proportional to x and $w - x$. Differentiating our metric with respect to x reveals

$$\frac{d(L/P_{ind})}{dx} = \frac{2\mu N^4 x (w-x)\sigma I_{rms}^2 (w - 2x)}{\left[N^2 l I_{rms}^2 + \sigma K_1 \left(\frac{\mu N I_{rms}}{l}\right)^q f^p x^2 (w-x)^2\right]^2}. \tag{3.6}$$

The metric has a minimum at $x = w/2$, showing that the best choice for A_c and A_{mag} are $A_c = A_{mag}$, so the volumes of the conductor and core are equal. This result is independent of the permeability of the core and the conductance of the conductor.

When the inductor is designed according to the optimization above, it is possible to calculate the number of turns on the inductor given the inductor volume. The volume of the rectangular solid containing the interlocked toroids is

$$V_{ind} = 4b \times 3b \times 3b = 36b^3 = \frac{36l^3}{8\pi^3}. \tag{3.7}$$

The number of turns on an inductor of given inductance and volume can now be stated as

$$N = 4\sqrt{\frac{L}{\mu}} \sqrt[6]{\frac{9}{2V_{ind}}}. \tag{3.8}$$

In order to avoid saturation of the core, this inductor is limited by

$$I_{peak} < \pi B_{max} \sqrt{\frac{V_{ind}}{72\mu L}}. \tag{3.9}$$

3.1.2   Relative Sizing of Inductor and Capacitor

With a second order L–C output filter for a step-down DC-DC converter, we can find the necessary capacitance value for any choice of inductance, if frequency and output ripple are specified. Capacitance is found by $C = \frac{V_{OUT}}{V_{Ripple}} \frac{1}{L} \frac{1}{(2\pi f)^2}$ in order to meet the specified output ripple. Here this is done by setting the filter frequency, $\frac{1}{2\pi\sqrt{LC}}$ to be a certain fraction, $\sqrt{\frac{V_{ripple}}{V_{OUT}}}$ of the switching fre-
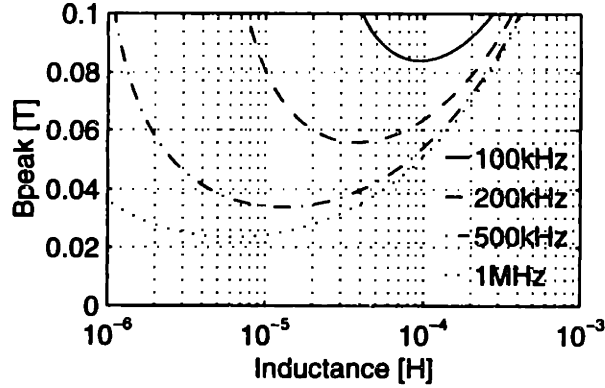
34

Figure 3-2: Peak flux as inductance is varied, for several frequencies.

quency. The fundamental component of the switching waveform is attenuated by the filter an amount $(f_{sw}/f_{filter})^2$. The volume of this capacitor can then be determined based on the capacitance per unit volume of readily available capacitors. For switching frequencies in the range of hundreds to thousands of kilohertz, this capacitor must be ceramic so that the characteristic impedance is capacitive, not resistive, at the switching frequency. From this derived value of capacitor volume, we can find the volume remaining for the inductor (this exercise is carried out for a fixed filter volume).

At this point, we can consider B_{peak} in the case where we fix the output filter volume and the switching frequency, but vary the value of the inductance. As the value of the inductance varies, the capacitance required changes (to maintain a fixed filter frequency), changing the volume available for the inductor. Since the inductor volume and the inductance are given, l_{mag} and N of the inductor can be found, assuming it is designed according to the optimization shown in the previous chapter. Under the stated conditions (inductance is varied, switching frequency is fixed, capacitance varies to fix filter frequency, N and l_{mag} vary as inductance and inductor volume change), B_{peak} has a minimum. If the value of the inductance is extremely large, there are a large number of turns and B_{peak} increases. If the value of the inductor is very small, there are large ripple currents in the inductor, so I_{peak} increases, and B_{peak} will become very large. A sample graph of B_{peak} is shown in Figure 3-2. Note that designs which result in a value of B_{peak} larger than B_{max} will saturate the magnetic core and are therefore not useful. Also, note that decreasing the switching frequency (at a constant volume) causes the B_{peak} curve to shift upward; the inductor must store more energy to provide the output current over a longer period.

The peak flux in the core, B_{peak} is specified by Equation 3.1. The variables l_{mag}, N, and I_{peak} may be specified in terms of the inductor volume, operating frequency, and the inductance.

$$I_{peak} = I_o + \frac{1}{2}\Delta I \tag{3.10}$$

$$\Delta I = \frac{(V_{in} - V_{out})D}{f_{sw}L} \tag{3.11}$$

35

$$N = 4\sqrt{\frac{L}{\mu}} \sqrt[6]{\frac{9}{2V_{ind}}} \tag{3.12}$$

$$l_{mag} = \sqrt[3]{\frac{2\pi^3 V_{ind}}{9}} \tag{3.13}$$

The inductor volume itself can be expressed in terms of the total volume allotted for the filter, the desired output ripple, and the specific capacitance available.

$$C = \frac{V_{out}}{V_{ripple}} \frac{1}{(2\pi f_{sw})^2 L} \tag{3.14}$$

$$V_{cap} = C\mathcal{V}_c \tag{3.15}$$

$$V_{ind} = V_{total} - V_{cap} \tag{3.16}$$

Finally, it is possible to express the B_{peak} in terms of the inductance, switching frequency, total filter volume, and constant parameters.

$$B_{peak} = \frac{6\sqrt{2\mu V_{ripple}}\,(2\,I_o\,f_{sw}\,L + D(V_{in} - V_{out}))}{\sqrt{4\,V_{total}\,V_{ripple}\,\pi^2 f_{sw}{}^2 L - V_{out}\,\mathcal{V}_c}} \tag{3.17}$$

The inductance can now be chosen as a function of the filter volume and the switching frequency, by minimizing B_{peak} with respect to L. The rationale for this is that it allows the selection of the lowest switching frequency possible for a given filter volume. Minimizing B_{peak} over L gives

$$L = \frac{I_o\,V_{out}\,\mathcal{V}_c + V_{ripple}\,V_{total}\,\pi^2 f_{sw}\,D(V_{in} - V_{out})}{2\,V_{total}\,V_{ripple}\,\pi^2 f sw^2 I_o}, \tag{3.18}$$

and the minimum value for B_{peak} is

$$B_{peak} = \frac{6\sqrt{2\mu I_o}\sqrt{I_o\,V_{out}\,\mathcal{V}_c + 2\,V_{ripple}\,V_{total}\,\pi^2 f_{sw}\,D(V_{in} - V_{out})}}{\pi^2 f_{sw} V_{total}\sqrt{V_{ripple}}}. \tag{3.19}$$

3.1.3 Switching Frequency

In order to find the lowest possible frequency for a given output filter volume, we must find the frequency for which the B_{peak} curve has its minimum at B_{max}. This is accomplished by solving the equation

$$B_{peak} = B_{max} \tag{3.20}$$

over f_{sw}, where B_{peak} is given above in Equation 3.19. This task yields

$$f_{sw} = \frac{72\mu D I_o}{\pi^2 V_{total} B_{max}^2} \left[(V_{in} - V_{out}) + \sqrt{(V_{in} - V_{out})^2 + \frac{B_{max}^2 \mathcal{V}_c V_{out}}{72\mu V_{ripple} D^2}} \right]. \tag{3.21}$$

As expected, this optimal f_{sw} shows that operating frequency is proportional to output power density. This quantitative expression for f_{sw} can now be used to find the efficiency of the power supply as a function of filter volume, V_{total}. The function $\eta(V_{total})$ then allows the tradeoff to be made between the filter volume and the battery volume (per Section 2.5.2).

3.2 Design of Switches

The most dominant loss mechanisms in a high frequency power supply are the switching losses associated with the switches, the conduction in the switches, and energy lost in parasitic elements. (Dominant parasitics are inductive at high output currents and capacitive at low output currents.) Thus, achieving a high efficiency depends on the proper design of the active devices. Williams, Blattner, and Mohandes [34] detail the loss mechanisms involved in DMOSFET switches, and address designing switches for high efficiency operation. Although they consider vertical DMOSFET devices exclusively, the physics and analysis presented are applicable to MOSFET switches in general. One important point they make is that since any converter must operate over a range of conditions, no single switch design will be ideal for all circumstances. Also, they note that at high frequencies, optimum sizing becomes more difficult, since efficiency will fall dramatically as the operating point moves away from the conditions for which the switch was optimized. Techniques which can increase the efficiency of a switch over a broader range of operating points are presented below.

3.2.1 Selecting Switch Size

The key design choice for power switches in any power supply (beyond the selection of the type of device) is the selection of the size of the switch. This choice negotiates the static losses incurred by activating the switch with the losses caused by the load current flowing through the switch. For a MOSFET switch operated in its linear (resistive) region, this trade-off can be formulated as an exchange between the resistance of the switch and the gate and junction capacitance of the switch.

The resistance of a MOSFET switch in its linear region is governed by the equation

$$R_{sw} = \frac{1}{\mu C_{ox} \frac{W}{L}(V_{GS} - V_T)},$$ (3.22)

where μ is the mobility of carriers in the switch, C_{ox} is the gate capacitance per unit area of the gate, W and L are the dimensions of the gate, V_{GS} is the gate to source voltage, and V_T is the switch threshold. The energy which is dissipated during a single switching cycle (due to switching capacitance alone) can be expressed by

$$E_{sw} = WL\left(V_{GS}^2 C_{ox} + V_{DS}^2 C_j\right),$$ (3.23)

37

where V_{DS} is the voltage switched at the drain node, and C_j is the junction capacitance. For simplicity, we can rewrite Equation 3.23,

$$E_{sw} = Q_o V_{GS} \tag{3.24}$$

$$Q_o = W \cdot L \cdot V_{GS} C_x \tag{3.25}$$

where C_x is a linear combination of C_{ox} and C_j. Because the total energy dissipated in the output switches is proportional to R_{sw} and Q_o, a useful metric for the characterization of switches in a semiconductor technology is then

$$Q' = R_{sw} Q_o = \frac{V_{GS} C_x L^2}{(V_{GS} - V_T) C_{ox} \mu}. \tag{3.26}$$

L is specified by the minimum gate length possible in that technology (or the minimum for the desired switch drain voltage), and the other parameters of Q' are specified by the properties of the process and the conditions under which the switch is to be used. (Note that a related metric, $C' = Q'/V_{GS}$ was introduced in Equation 2.12.) Semiconductor technologies which offer the lowest values of Q' will provide the most ideal (lossless) switches.

The power dissipation in a power switch can now be expressed with the equation

$$P_{sw} = I_{rms}^2 R_{sw} + \frac{f_{sw} Q' V_{GS}}{R_{sw}}, \tag{3.27}$$

which under minimization yields

$$R_{sw\,opt} = \sqrt{\frac{f Q' V_{GS}}{I_{rms}^2}}. \tag{3.28}$$

I_{rms}^2 is the root mean square current in the switch at issue. For a down converter in continuous mode conduction, this corresponds to $D \cdot I_{ind\,rms}{}^2$ for the high-side switch, and $(1 - D) \cdot I_{ind\,rms}{}^2$ for the low-side switch. The minimum power dissipation (for the specified load and semiconductor technology) is

$$P_{sw\,min} = 2 I_{rms} \sqrt{f_{sw} Q' V_{GS}}. \tag{3.29}$$

The transistor width for this minimization is given by

$$W = R_o \sqrt{\frac{I_{rms}^2}{f_{sw} Q' V_{GS}}}, \tag{3.30}$$

where R_o, the resistance per unit width, is
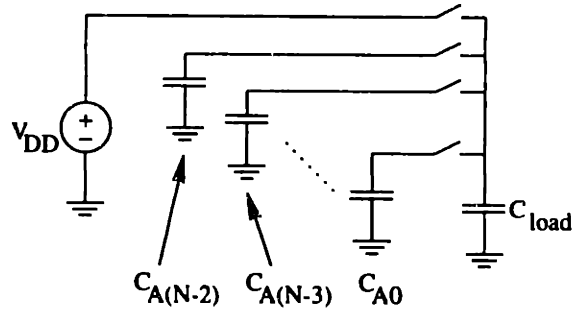
$$R_o = \frac{L}{\mu C_{ox} (V_{GS} - V_T)}. \tag{3.31}$$

Figure 3-3: N-step stepwise charger using a single voltage source and a bank of $N-1$ capacitors.

3.2.2 Decreasing Switch Drive Dissipation

If the gate of a MOSFET is driven high and low from a voltage source through switches, the energy that is placed on the gate capacitor each cycle is lost. Decreasing the amount of energy lost by using alternate charging techniques will increase the efficiency of the associated power supply. Effectively, Q' will be decreased, and it will be possible to use a larger switch for any given operating frequency. Two charging techniques which achieve such gains are stepwise charging [2, 27] and adiabatic inductive charging [2, 12, 28, 32].

Stepwise Charging

When any capacitor (such as the capacitive gate of a MOSFET switch) is charged through a switch from a voltage source V_{DD}, the energy $\frac{1}{2}CV_{DD}^2$ is dissipated. One way to reduce this fixed loss is to charge the capacitor through a sequence of voltage sources. Consider the case where there are N voltage sources V_{si}, of values $V_{DD}\frac{i}{N}$. If the capacitor is charged to V_{si-1}, and shorted to voltage source V_{si}, the energy dissipated is only $\frac{1}{2}CV_{DD}^2/N^2$. When the capacitor is charged to V_{DD} by connecting it in sequence to V_{s0}, V_{s1}, ..., the total power dissipated is only $\frac{1}{2N}CV_{DD}^2$, a significant savings.

In Svensson and Koller [27], the stepwise charging principle is explained, and an optimum value for N is derived, accounting for the additional losses due to the extra switches involved. They also mention that the sequence of voltage sources could be implemented with a one voltage source, V_{DD}, and $N-1$ identical capacitors, C_{A0} through $C_{A(N-2)}$ (See Figure 3-3). In steady state, the capacitors will reach the desired steady state values, V_{si}. If all the capacitors begin with no voltage on them, the load capacitor will act to transport charge from the supply to the capacitor bank capacitors. The voltage on each capacitor will increase until it is half-way between the voltages on the capacitors above and below it.

Unfortunately, treating the capacitors as voltage sources is only accurate if the capacitance C_A is much larger than C_{load}. If these capacitors are to be integrated on chip, this means that the

39

Capacitor	q_{Ai0}	q_{Ai1}
C_{A0}	a	$a\frac{C_A}{C_A+C_{load}}$
C_{A1}	$a+b$	$\left(b+a+a\frac{C_{load}}{C_A+C_{load}}\right)\frac{C_A}{C_A+C_{load}}$
C_{Ai}	$a+bi$	$\left(a+bi+q_{A(i-1)1}\frac{C_{load}}{C_A}\right)\frac{C_A}{C_A+C_{load}}$

Table 3.1: Charge on bank capacitors in load charged up state and load charged down state.

total area of the stepwise charging capacitors will be much larger than the area needed for the output switches. If the capacitors are external, the power supply will require a significant number of additional pins, and the capacitors may occupy a considerable volume.

Whether it is worthwhile to use stepwise charging will depend on the power density of the system and the volume required by the additional capacitors. This can be evaluated with the expression of Eq. on 2.36, by comparing the power dissipation savings, additional volume, and the battery power density. We would also like to quantify how the power savings of stepwise charging varies when C_A is not much greater than C_{load}; this is discussed below.

Consider the bank of capacitors to be labeled C_{A0} through $C_{A(N-2)}$, each identified with a charge q_{Ai}. As the load capacitor is charged up and down, the charge on each capacitor will decrease and increase. In steady state, there are essentially two states; when the load capacitor has zero volts on it, the bank capacitors are charged to $q_{Ai} = q_{Ai0}$, and when the load capacitor is at V_{DD}, the bank capacitors are charged to $q_{Ai} = q_{Ai1}$. In steady state, we also know that the voltage difference between adjacent capacitors (C_{Ai} and $C_{A(i+1)}$) is the same for all capacitors. Thus, we can express q_{Ai0} as $a + bi$, where a is the charge on the first capacitor, and b is the charge difference on adjacent capacitors. Furthermore, using the values of q_{Ai0}, the values of q_{Ai1} can be derived based on the charge sharing that occurs as the load is connected to each of the bank capacitors. Table 3.1 summarizes the charges on the bank capacitors in the two states. The charge on capacitor C_{Ai} when the load is at V_{DD}, is given by

$$q_{Ai1} = \left(a + bi + q_{A(i-1)1}\frac{C_{load}}{C_A}\right)\frac{C_A}{C_A + C_{load}}.$$ (3.32)

This expression for q_{Ai1} can be expressed as a summation (removing the recursive definition),

$$q_{Ai1} = \frac{C_A}{C_A + C_{load}}\sum_{n=0}^{i}(a + b(i-n))\left(\frac{C_{load}}{C_A + C_{load}}\right)^n.$$ (3.33)

Due to the symmetry of the situations when the load is charged to $0V$ and V_{DD}, it is apparent that q_{Ai1} can be expressed by $V_{DD} - a - b((N-2) - i)$ (that is, the charge difference between adjacent capacitors is still b, and the voltage drop between the most positive capacitor and V_{DD} is

40

Event	C_1	q_1	C_2	q_2
1	C_{A0}	q_{A00}	C_{load}	0
2	C_{A1}	q_{A10}	C_{load}	$q_{A01}\frac{C_{load}}{C_A}$
3	C_{A2}	q_{A20}	C_{load}	$q_{A11}\frac{C_{load}}{C_A}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
i+1	C_{Ai}	q_{Ai0}	C_{load}	$q_{A(i-1)1}\frac{C_{load}}{C_A}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
N-1	$C_{A(N-2)}$	q_{A0}	C_{load}	$q_{A(N-3)1}\frac{C_{load}}{C_A}$
N	C_{load} shorted to voltage source V_{DD}			

Table 3.2: Set of capacitor-shorting events which comprise the charging process.

a/C_A). A value for b can now be derived by solving $b = q_{A(i+1)_1} - q_{Ai_1}$. This results in the relation,

$$b = \frac{C_A}{C_A + C_{load}}a. \qquad (3.34)$$

In addition, the equation $a/C_A = V_{DD} - q_{A(N-2)_1}/C_A$ allows us to derive a value for a,

$$a = \frac{V_{DD}C_A^2}{C_{load} + NC_A}. \qquad (3.35)$$

At this point, we can write simplified expressions for q_{Ai0} and q_{Ai1}:

$$q_{Ai0} = \frac{V_{DD}C_A \left(C_A(i+1) + C_{load}\right)}{C_{load} + NC_A} \qquad (3.36)$$

$$q_{Ai1} = \frac{V_{DD}C_A C_{load}i}{C_{load} + NC_A} \qquad (3.37)$$

We now have sufficient information to derive the energy dissipation involved in charging the load capacitor from 0 to V_{DD}. When two capacitors, C_1 and C_2, with charges q_1 and q_2 are shorted together, the energy dissipation is

$$E_{diss} = \frac{(q_1 C_2 - q_2 C_1)^2}{2C_1 C_2(C_1 + C_2)}. \qquad (3.38)$$

The charging process can then be considered a set of N such actions, where q_1, q_2, C_1, and C_2 are given in Table 3.2. The total power dissipation is the sum of the dissipations from each of these N events,

$$E_{diss} = \frac{V_{DD}^2 C_{load}(C_A + C_{load})}{2(C_{load} + NC_A)}. \qquad (3.39)$$

The dissipation is the same when the load is discharged from V_{DD} to 0. It can be easily seen that this expression approaches the limit put forth in [27], $\frac{V_{DD}^2 C_{load}}{2N}$ when $C_{load} \ll C_A$. Figure 3-4 shows a plot of this expression for various values of N.

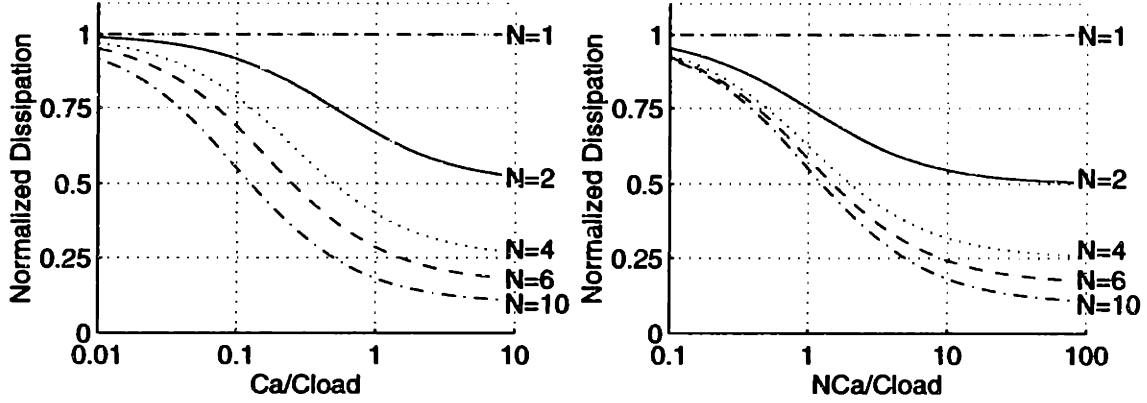Considering the losses incurred by additional switches, Svensson and Koller [27] derive an optimal

Figure 3-4: Energy dissipated during stepwise charging of capacitor, with bank of capacitors of value C_A. Right hand figure plots relative to the ratio of the total capacitance in bank to the load capacitance. Neither plot includes the power dissipated driving the gates of the charging devices.

value for N. This optimum will change when under the condition that C_A is not much greater than C_{load}. When $C_A N \gg C_{load}$, this optimum is approximately

$$N_{opt} = \sqrt[3]{\frac{T}{4m\bar{\rho}} \frac{C_A + C_{load}}{C_A}} \tag{3.40}$$

using the notation of [27]. T is the total charging time, m is the number of RC time constants spent on each charging step, and $\bar{\rho}$ is a weighted metric quantifying the quality of the switch (similar to Q' defined in Equation 3.26). The optimum presented in [27] lacks the ratio $\frac{C_A + C_{load}}{C_A}$ under the radical. This value of N_{opt} will correspond to an energy dissipation of approximately

$$E_{opt} = \frac{3}{2} \sqrt[3]{\frac{4m\bar{\rho}}{T} \left(\frac{C_A + C_{load}}{C_A}\right)^2} C_{load} V_{DD}^2. \tag{3.41}$$

Adiabatic Inductive Charging

Although the word "adiabatic" has the connotation of a slow process, strictly speaking it means "without loss." The loss of energy inherent in charging and discharging a gate through a switch can be avoided by using the resonant ring between an inductor and the gate capacitor to add and remove energy from the gate. The inductor is connected between a voltage source and the capacitor; when the voltage source turns on, the inductor current builds up, and the capacitor begins to accumulate charge. After the gate voltage surpasses the input voltage, the inductor current will begin to decrease. When the inductor current reaches zero, the circuit is broken with a switch. If there were no parasitic resistance, the gate voltage would be twice the input voltage, and no energy would have been dissipated during the charging process.

The clear disadvantage of resonant gate drives is that they require the addition of an inductor,
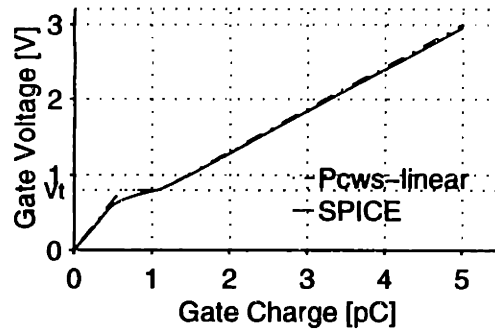
Figure 3-5: Typical MOS gate charge curve, with linear approximation and SPICE simulation. Gate is charged with a current source, while the drain is connected to the supply through a resistor.

which will necessarily be an external (off chip) component. The advantage of conserving the gate power is quickly lost for a low power, low volume power supply; simply driving the off chip signal to activate the inductor may consume an undesirable amount of power. The additional components will cost money, as will the additional package pins and board space.

3.2.3 Decreasing Drive Requirements at Light Loads

When the volume of a power supply must be very small, the switching frequency will be increased to improve the power density of the converter. Higher switching frequencies cause higher power dissipation in the switches (Equation 3.27), and therefore lower efficiency. Since a portion of the additional dissipation is static (for example, the amount of energy dissipated charging and discharging the gates of the output switches), and not proportional to the load current, the drop in efficiency will be exacerbated when the load current is reduced to a level below the optimized operating point. Decreasing the static dissipation assists in the goal of creating a power supply which is efficient over a broad range of operating points. The trade off between switch resistance and gate capacitance can be made dynamically, by not charging the entire gate, dynamically modifying the threshold of the device, or by not charging the gate to the full supply potential.

When the gate of a MOSFET switch is charged, there are three distinct regions (Figure 3-5) [34]. In cutoff, below threshold, the gate can be treated as a linear capacitor, C_{ISS}. (C_{ISS} is a common name for the incremental gate capacitance for a gate voltage of $0V$.) The capacitor C_{ISS} is a nonlinear capacitor, however, it is treated here in a piecewise linear fashion. At threshold, as there is a zone of high incremental capacitance, caused by the discharge of the gate-drain overlap capacitor. Consider the charge delivered to the gate in this region to be Q_{G_m}. Above threshold, when the device is in the linear region, the gate appears to be a capacitor of value C_{lin}. C_{lin} is typically several times larger (2 to 5) than C_{ISS} [34]. It can be shown that the total energy dissipated

43

by charging a MOSFET gate high then low from a voltage source through a resistor is

$$E_{gate} = V_{DD}\left(C_{ISS}V_T + Q_{G_m} + C_{lin}(V_{DD} - V_T)\right)$$

$$E_{gate} = V_{DD}Q_{total}. \tag{3.42}$$

The total power in the output switches is

$$P_{sw} = \frac{I_{rms}^2}{k\frac{W}{L}(V_{DD} - V_T)} + f_{sw}V_{DD}\left(C_{ISS}V_T + Q_{G_m} + C_{lin}(V_{DD} - V_T)\right), \tag{3.43}$$

where k is μC_{ox}. The capacitor C_{ISS} is smaller than C_{lin} because C_{ISS} corresponds to the creation of the depletion region, whereas C_{lin} corresponds to the growth of the channel.

When the switch size is already specified for optimal operation at a higher load current, three ways to reduce the switch losses accounted for in Equation 3.43 are: activating only parts of the gate, reducing the voltage the gate is charged to, and increasing the threshold voltage (possible in a dual-gate SOI process). Each of these techniques trades switch resistance for the necessary gate energy. The dependencies on V_{DD} and V_T are stated explicitly in Equation 3.43; note that each of the terms in Q_{total} scale linearly with the fraction of the gate which is energized.

Of the three techniques suggested, clearly partial activation of the gate is the most practical. Changing the gate voltage and the threshold voltage both require the generation of a new voltage signal, either V_{DD} or the back gate voltage. Partial activation of the gate is also a more effective solution at very low output powers. Consider the limit as $R_{sw} \to \infty$. In this case, when using partial gate activation, $E_{gate} \to 0$. However, if V_{DD} is changed dynamically, $E_{gate} \to V_T\left(C_{ISS}V_T + Q_{G_m}\right)$, and if the threshold is varied dynamically, $E_{gate} \to V_{DD}\left(C_{ISS}V_{DD} + Q_{G_m}\right)$. Partial gate activation is the most practical method of improving efficiency over wide ranges of output current, and it is unlikely that it would be worthwhile to combine two or three of the techniques suggested. Other techniques to reduce power at light loads not considered here are pulse frequency modulation or switching frequency reduction.

3.3   Control

The output of any regulating power supply must be controlled by a feedback loop in some fashion. There are several reasons that closed loop operation is necessary. If a precise output voltage is needed, it is unlikely that an accurate output voltage could be created without feedback. It is much more economical (in terms of design time and cost) to create an accurate gain between an input reference and output supply voltage with feedback than without. Even when a precise output voltage is not needed, feedback allows looser bounds on the characteristics of the power supply components (the PWM, the power switches, and the inductor). With feedback, moderate variation

of the performance of these components is unlikely to have a noticeable effect on the output voltage. Aside from isolating the output from the variations of the circuits, feedback allows the output to be constant over some variation in the input source. This is particularly important when the power source is a lithium battery, which often has a large droop over battery life. Besides its effect on the DC output of the power supply, control circuitry is used to modify the dynamic response of the power supply. Closed loop operation can decrease the output impedance of the power supply at frequencies below the resonances of the output filter, preventing low frequency oscillations in load conditions from affecting the supply (and hence, other circuitry). (Note that attenuation at higher frequencies must be done passively.)

3.3.1 Standard Power Supply Control Design

The tools of the designer of a power supply controller are circuit analysis, state averaging, linearization, and of course, control theory. The application of these tools to power circuits is detailed in [9]; the treatment here is meant to outline the typical use of these methods for the specific case of DC-DC down converters. It is assumed that the reader is familiar with the methods of analysis employed.

The output filter frequency and switching frequency of a power supply is usually chosen to satisfy the system requirements on power density and efficiency. The control loop is then designed to provide good DC voltage regulation and improve the time it takes for the output to return to its nominal value after load or input variations. Stability of the power supply must be maintained over the possible operating conditions. The attenuation from the switching node voltage to the output (the ripple voltage) and the lower bound on the filter response time is determined by the filter characteristics. It is theoretically possible to improve the response time further if active damping is employed, but this is not a practical endeavor, particularly when achieving a high power density and efficiency.

The analysis presented here is for a down converter with a second order output filter. Power supply output filters are generally at least second order filters, and it is likely that the filter for extremely low power system will be limited to second order, due to size and cost constraints.

Modeling

The dynamics of switched power supplies are commonly analyzed using averaged circuit models. These models can accurately capture the behavior of the circuit at frequencies well below the switching frequency. This methodology relies on the characteristic rate of change of the averaged state variables of the supply to be significantly slower than the switching period. The averaged state
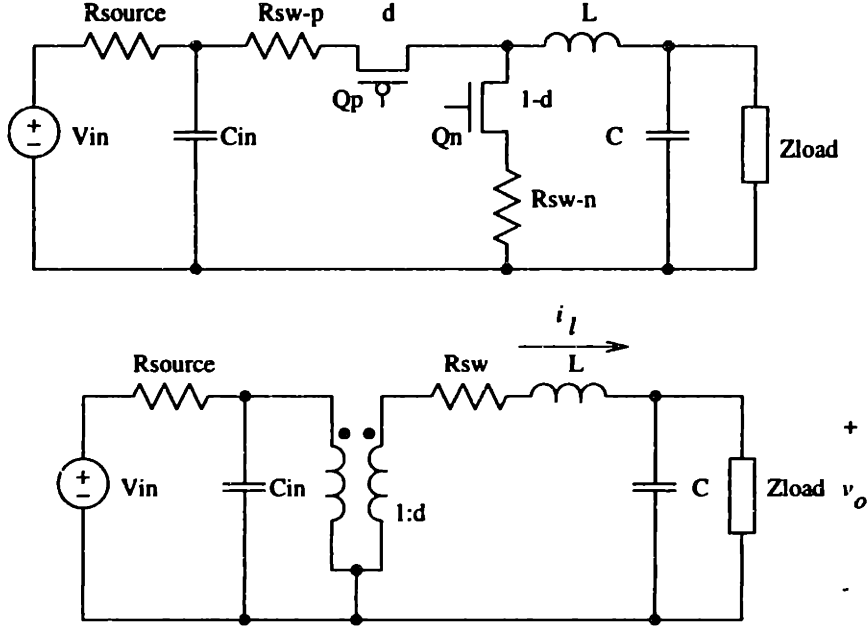
Figure 3-6: Circuit schematic and averaged model for a low power down converter with synchronous rectification. (Valid only for continuous-mode conduction; $i_{ind}(t) > 0$) $R_{sw} = (1-d)R_{sw-n} + dR_{sw-p}$

variables are defined by

$$\bar{x} = \frac{1}{T} \int_{t-T}^{t} x(\tau) d\tau \tag{3.44}$$

where T is set to the switching period of the circuit. (Note that this choice of $T = T_{sw}$ is convenient, but it is not strictly necessary.) Analyzing the behavior of the circuit through the averaged circuit variables gives a very good approximation for the dynamic behavior of the circuit, without requiring analysis of the complexity introduced by the switches [9]. In the case of resonant converters, it is necessary to also monitor the magnitude of the first fundamental of the state variables ($\bar{x}$ is 0) [8].

A circuit schematic and the corresponding averaged circuit model are shown in Figure 3-6 for the down converter topology. Switch resistance is included in the model to examine its effect on filter resonance. (Effect of switch resistance is usually negligible at higher output powers.) It is assumed that the inductance between the source and the supply will be negligible for a small, low power system. The next simplification is to monitor the incremental changes in the circuit states, which linearizes the circuit, removing such non-linear dependencies such as $v_{in} \cdot d$, where both v_{in} and d could be variables. The incremental model substitutes v_{in}, d, i_l, and v_o:

$$
\begin{aligned}
v_{in} &= V_{IN} + \hat{v}_{in} \\
d &= D + \hat{d} \\
i_l &= I_L + \hat{i}_l \\
v_o &= V_O + \hat{v}_o,
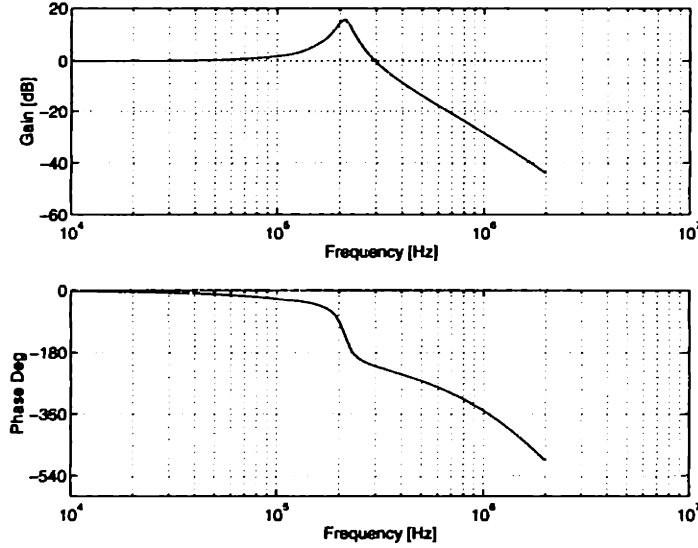\end{aligned}
\tag{3.45}
$$

Figure 3-7: Plot of transfer function $G(s)$ for $f_{sw} = 1.34MHz$, $V_{in} = 3V$, $V_{out} = 1.2V$, $Z_{load} = 19.2\Omega$, $L = 6.8\mu H$, $C = 0.25\mu F$, $R_{sw} = 0.305\Omega$, $R_{source} = 0.5\Omega$, and $C_{in} = 0.01\mu F$. These are practical values for a 75mW converter. Phase shift includes a $370ns$ delay (1/2 the switching period).

where capital letters indicate operating point values, and hat variables represent variations from the operating point.

In order to determine the loop gain (to analyze stability), we need the transfer function between $\hat{d}$ and $\hat{v}_o$. For the system described, this is given by:

$$G(s) = \frac{\hat{v}_o(s)}{\hat{d}(s)} = \frac{Z_{load}\left(LCZ_{load}s^2 + (L + R_{sw}CZ_{load})s + Z_{load} + R_{sw}\right)}{LC\,Z_{load}s^2 + (L + R_{sw}CZ_{load})s + 2Z_{load} + R_{sw}} \times$$

$$\frac{1}{K_1s^3 + K_2s^2 + K_3s + Z_{load} + R_{sw} + R_{source}\,D^2}$$

$$K_1 = R_{source}\,C_{in}\,LCZ_{load}$$

$$K2 = R_{source}\,C_{in}(L + R_{sw}CZ_{load}) + LCZ_{load}$$

$$K3 = R_{source}\,C_{in}(R_{sw} + Z_{load}) + CZ_{load}(R_{source}\,D^2 + R_{sw}) + L \qquad (3.46)$$

This gain function could be further complicated by attempts to model the higher frequency dynamics of the control circuitry. One model for a nonlinearity introduced by the PWM circuitry is a time delay, since the PWM signal output by the power circuit can be updated at most once per cycle. A delay of one half a period approximates the typical delay between choosing the new duty cycle and actually creating that duty cycle at the output. A plot of the transfer function given in Equation 3.46 is given in Figure 3-7 for a specific set of values. This plot includes a time delay equal to one half the period.

47

Control Law

For the down converter with one feedback variable, $\hat{v}_o$, the compensation function, $F(s)$ describes how $\hat{d}$ is driven based on $\hat{v}_o$. The control law of the power supply is $\hat{d}(s) = F(s)\hat{v}_o$. In the actual circuit, $\hat{v}_o$ will usually be $v_o - v_{ref}$, where v_{ref} is a voltage reference giving the desired output voltage.

Given the transfer function $G(s)$, the task of the control designer is to construct a compensation function $F(s)$ that will provide the highest crossover frequency possible while maintaining a reasonable margin of stability. The crossover frequency is the frequency at which the loop gain, $G(s)F(s)$ is equal to one. Higher crossover frequencies correspond to faster transient responses. Stability can be measured with the gain margin and the phase margin. These metrics are based on the fact that a system will be unstable if it has a gain of -1, shown on the Bode plot as a gain of $0dB$ with a phase of $\pm 180^o$. The phase margin is the difference between the phase of $G(s_o)F(s_o)$ and -180°, where s_o is given by $2\pi j f_{c-o}$, and f_{c-o} is the crossover frequency. Small phase margins correspond to more oscillatory responses, and give little room for error in case more phase is present than anticipated. Gain margin is a measure of the amount the gain is below unity at the frequency where the phase crosses -180°. The gain margin shows the amount the gain could increase before the system becomes unstable. The gain or phase could be different than that anticipated in the Bode diagram due to modeling errors, a change in the operating point, or a variation in the filter characteristics due to component variation.

In order to achieve good DC voltage regulation, the control law will typically contain an integrator, or a pole at zero frequency. This sets the DC gain to infinity, forcing the output voltage in steady state to reach the reference value. Additional poles or zeros may be added to $F(s)$ to shape the loop gain as desired, and a linear gain term can be used to adjust the crossover frequency.

The crossover frequency is usually selected to occur at some frequency below the filter resonance frequency. The poles of the filter add 180° of phase delay and there are additional poles and phase delays at frequencies not far above the filter resonance. This makes it is very difficult to set the crossover frequency above the resonant frequency and still achieve reasonable stability margins, particularly since the poles, zeros, and delays at higher frequencies are difficult to model accurately.

When the crossover frequency is chosen below the filter resonant frequency, the magnitude of the resonance acts to limit the gain margin, forcing the designer to choose a lower f_{c-o} than otherwise possible. Figure 3-8 illustrates a sample loop gain for the filter shown in Figure 3-7. Here, $F(s)$ consists of an integrator and a pole somewhat below the filter resonance. The pole (at $80kHz$) trades some of the phase margin for additional gain margin, making it possible to set the f_{c-o} at a higher frequency than would be possible without the additional pole. The pole adds 45° phase shift at $80kHz$, but reduces the gain at the filter resonance by $8.6dB$. For the filter and control law shown, the gain margin is $12dB$ and the phase margin is 70°.
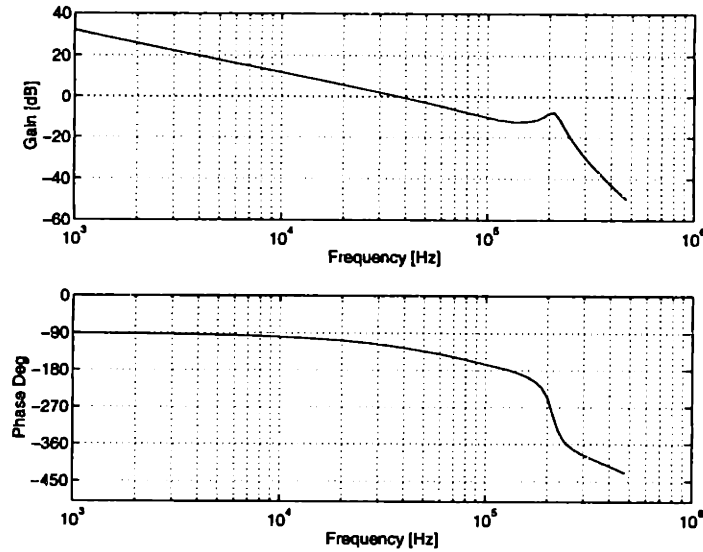
Figure 3-8: Loop gain for filter of Figure 3-7. $F(s)$ contains an integrator and a pole at $80kHz$. Crossover is at $36kHz$, phase margin is $70°$, and gain margin is $12dB$.

Damping Filter Resonance

The resonance of the output filter typically limits the speed of the closed-loop system, as shown in Figure 3-8. If the output filter were more damped, the crossover frequency could be pushed higher (although it still must be kept below the filter resonant frequency). The filter can be damped passively by either a resistance in series with the inductor or in parallel with the output capacitor. A series resistor will dissipate power, as it will always see the current through the inductor. It is not desirable to add such a loss mechanism, but for extremely low powers the switch resistance may become large enough to provide significant damping. The damping resistor presented by the switches is R_{sw}, derived earlier as a function of the switch resistances and the duty cycle. A resistor in parallel with the output capacitance will also dissipate power, but what can be done is to add to the output a resistor in series with a large capacitor. This could take the form of an electrolytic or tantalum capacitor with its internal resistance. At the high frequency of the filter resonance, the impedance of the large capacitor will be only its inherent series resistance, which will damp the output filter, without presenting a DC power dissipation. The large output capacitor will also decrease the output voltage deviation due to load transients in the time before the feedback network begins to respond.

A resistive load will also provide load damping; this is in fact the primary damping mechanism for the filter of Figure 3-7. Relying on the load to provide damping can cause problems, since the damping will change appreciably with changes in the load. It is possible that the system will become unstable under the low load conditions. It is also possible that the load will have a high impedance if is loaded with current sources, such as an analog integrated circuit, as opposed to a resistive digital
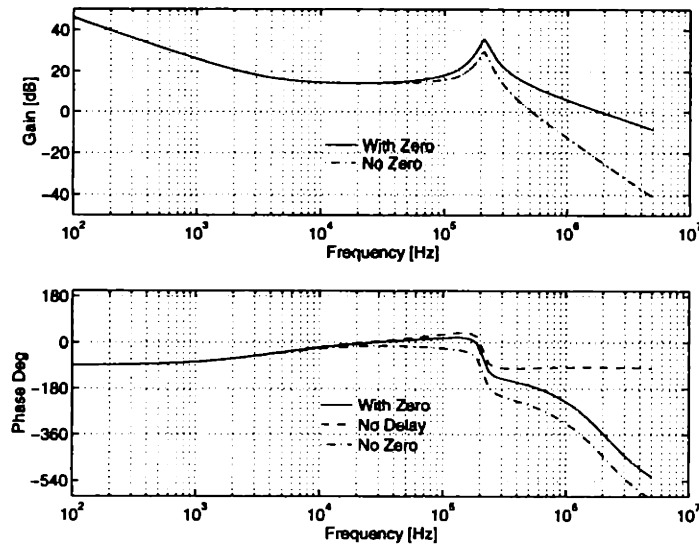
Figure 3-9: Loop gain under an attempt to provide active damping. A zero is added near resonance to reduce phase shift, and plots are given for the system with and without this zero. Time delay due to the switched nature of the system makes it impossible to have any phase margin at crossover.

circuit. In any case, as output power level decreases, the load resistor increases. If the characteristic filter impedance, $\sqrt{\frac{L}{C}}$, does not increase at the same rate as the load (which is very likely), loads will provide less damping at lower powers.

Active Damping

Active damping is an attempt to dampen the filter resonance by placing the crossover frequency above the filter resonance frequency. A gain greater than one at the filter resonance will increase the bandwidth of the system response, so that the supply will respond to transients faster than the natural time constants of the filter itself. The need to add large capacitors with series resistance is avoided. A fast response such as this is desirable for variable voltage systems where the output voltage is changed very quickly [5], perhaps on a computation by computation basis.

Active damping is not a technique that is used often in power supply design. The problem with active damping is that it is difficult to have the crossover occur between the filter resonant frequency and the switching frequency while maintaining adequate phase margin. For the system considered above, loop gains for the system under an attempt at active damping are shown in Figure 3-9. A zero is added at a low frequency ($40kHz$) to remove phase added from the integrator. Another zero is added at the filter resonance, so that the phase above resonance is only -90°, not -180°. Unfortunately, the phase added by the time delay makes it impossible to crossover very much above the resonant frequency with comfortable phase margin. A line in Figure 3-9 shows what the phase would look like without the time delay. If this were the actual phase, setting a crossover frequency above resonance (around $1 - 2MHz$ in this example) could be accomplished with no reservations

50

about the stability.

The phase at crossover can be increased by increasing the switching frequency (or lowering the filter and crossover frequencies). However, these changes make serious tradeoffs with efficiency and volume. In general the switching frequency should be chosen as low as possible for the given filter size. If it makes sense (for a particular system) to sacrifice the converter efficiency enough to make active damping feasible, it is probably possible to also increase the filter resonant frequency and set the crossover *below* the filter resonance. The increase in the crossover frequency (due to a higher filter frequency) will achieve faster transient performance, without utilizing active damping. Active damping is a dangerous proposal, given the uncertainty in the modeling of the high frequency dynamics.

3.3.2 Single Bit Feedback

The above discussion on feedback stability is independent of the type of circuitry used to implement the feedback (be it an analog continuous time filter, a discrete time switched capacitance amplifier, or a quantized digital filter). Typically, digital implementations of analog filters can use the model that the quantization of the output levels corresponds to the introduction of noise into the circuit. This model, however, depends on the fact that the signal magnitude is much larger than the quantization levels. If the feedback signal is a single bit (output too high or output too low), this no longer holds. Single bit feedback circuits make use of very simple A/D converters (just comparators), and may allow the construction of compact digital controllers.

Describing functions [18] are a method of non-linear control analysis which provide a method to analyze a system with single bit feedback signal. Describing function analysis gives a way to test the stability of systems in which the loop gain has a dependence on frequency or amplitude. It is a frequency domain analysis, so the input is always considered a sinusoid at a fixed frequency, and only the fundamental harmonic of the output signal is considered. The single bit feedback signal can be considered an amplitude dependent gain; when the input signal amplitude is small, the gain is large, relatively, and when the input signal is large, the gain becomes small (since the magnitude of the output signal is constant).

The first step in describing function analysis is to separate the loop gain into a nonlinear part, N, and a linear part, $G(j\omega)$. The poles of the closed loop system are then given by

$$1 + NG(j\omega) = 0, \tag{3.47}$$

or

$$G(j\omega) = -\frac{1}{N}. \tag{3.48}$$

Stability can then be measured by comparing the loci (in the complex plane) of $G(j\omega)$ and $-1/N$,
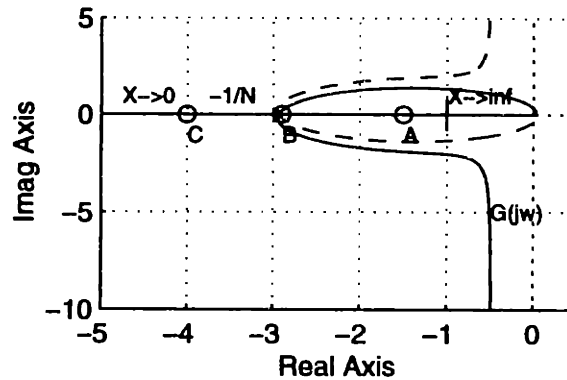
51

Figure 3-10: Nyquist plot for determining the stability of a system with a gain non-linearity such as the one caused by a single bit feedback signal. $-1/N$ is a straight line from $-\infty$ to 0.

much as the loci of $G(j\omega)$ is compared to the point -1 in linear Nyquist analysis.

The nonlinearity introduced by a single bit comparison between the output voltage and a reference can be expressed by

$$N = \frac{4}{\pi X} \tag{3.49}$$

where X is the signal amplitude of the input sinusoid, and it is assumed that the magnitude of the output signal is $+1/-1$. The loci of $-1/N$ and a sample $G(j\omega)$ are shown in Figure 3-10. Rather than examining this plot for encirclement of the $-1 + 0j$ point, we examine the encirclement of the $-1/N$ loci. Points which are encircled by $G(j\omega)$ represent stable points of operation (for these magnitudes X, the system is stable and the response will decay), and points which are not encircled represent unstable operation, where the magnitude will increase with time. For the plot shown, if the system is at Point A (corresponding to some output ripple X_A), the response is stable, so the output will decay. As the output decays, X decreases, and the operating point moves to the left, toward Point B and Point C. If the system starts from Point C, the response will be unstable, and the magnitude of the output oscillation will increase. As X increases, the system operating point moves toward Point B and Point A. Point B represents the predicted steady state limit cycle of the system. This analysis shows that the system should converge to a state where the output oscillates at a magnitude X_B, and a frequency corresponding to the value of ω where $G(j\omega)$ coincides with Point B.

The describing function analysis is an approximate technique, and the results it predicts may or may not be accurate depending on the characteristics of the system. Demonstrating a describing function analysis of a single bit feedback controlled power supply and verifying the results is not attempted here, but this would be helpful in characterizing such systems, as well as choosing the gain based on the desired magnitude of the equilibrium limit cycle.

3.3.3 Choice of Feedback Variable

Using the output voltage as the feedback state has been a ubiquitous choice in power supply designs. Although many power supplies will also feedback on the output current to control the output over-current condition, the primary mode of operation for DC-DC converters is one in which the error signal is created by comparing the output voltage to a voltage reference. However, a system with a closer relationship between the power supply and the load has the opportunity to use other states as the feedback variable to increase the efficiency of the overall system [7, 30, 16, 31, 6].

Any static CMOS logic will consume a lower energy if its supply voltage is lowered [2, 3]. Of course, this decrease in power consumption is accompanied by a decrease in speed, and if the voltage is lowered too far, the circuit will no longer compute correctly. Generally a circuit designer must choose the voltage supply level to be high enough to ensure proper operation over the range of probable process variations, environmental factors such as temperature, and input data (if the processing time is data dependent). As a result, the designer is forced to select a higher supply voltage (and hence, dissipation) than is necessary for proper operation under the most common circumstances. The ability to select a feedback variable which measures the delay of the circuit, as opposed to the value of the supply, circumvents the problem of selecting a worst case high supply voltage, allowing the supply voltage to vary dynamically in response the operating conditions.

There is no fundamental difference between selecting the output voltage as the feedback variable and selecting some metric which quantifies the performance of the circuit. The loop gain can be computed as before, except a term must be added to describe the gain from the output voltage of the power supply to the metric which is measured by the feedback loop. It is possible that this added term will also add some dynamics (and have an effect on stability), although the speed of these dynamics, and hence, their importance, depends on the nature of the digital circuitry and the performance metric of interest.

Critical Delay Length

Von Kaenel, et al. [30] propose measuring a model of the critical delay of the digital circuit to set the output voltage. An block diagram of such a system is shown in Figure 3-11. Rather than comparing the output voltage to a reference voltage, the propagation delay of a signal through a piece of circuitry is compared to the period of a reference clock. This feedback mechanism is appropriate for synchronous systems, which will maintain a fixed data rate as dictated by an external clock. The feedback loop will control the power supply voltage so that it is the minimum voltage possible to maintain the given data rate. A similar system is proposed by Horowitz [7], where the delay is measured monitoring the frequency of operation of a ring oscillator. This system also used a swtching regulator, whereas the Von Kaenel system made use of a linear regulator.

The model of the critical delay could be a chain of logic gates which match the delay seen by the
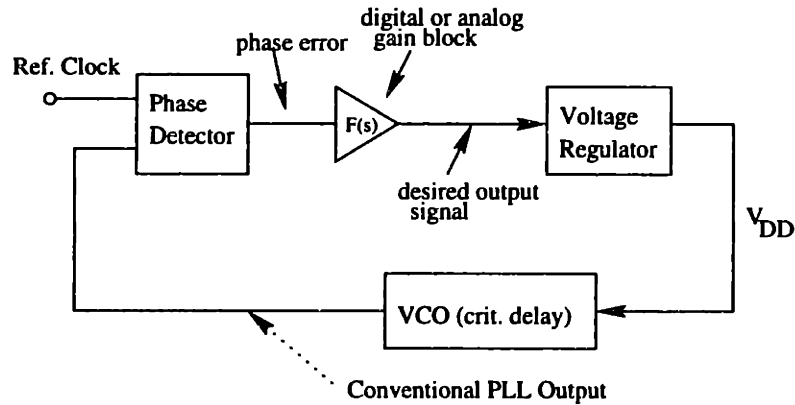
Figure 3-11: Power supply using critical delay as a feedback variable, showing resemblance of circuit to a PLL.

critical path. The critical delay could also be modeled by a rather simple set of gates, and the delay of these devices could be related to the critical delay of the circuit empirically. The issue of delay tracking between different types of gates is studied in [30]. The dynamics between the variation of the power supply voltage and the corresponding response of the delay of the logic gates will be very fast, because logic gates have no slow local state variables. In most, if not all, circumstances, it will be acceptable to consider these dynamics instantaneous, and ignore them when computing the loop gain.

If the chain of logic gates modeling the critical path is made into a ring oscillator, the resulting ring oscillator can be considered a VCO (voltage controlled oscillator), whose input control is the supply voltage. The feedback loop of the power supply is then identical to that of a phase locked loop (see Figure 3-11). When the output signal of the VCO is phase locked to a reference clock, the critical delay is matched to the clock period, and the supply voltage is at its minimum value for proper operation.

Input or Output Queue

For asynchronous (self-timed) circuits,

It is sometimes possible to use input or output FIFO buffers to measure the speed of a circuit and control the power supply, as described in [16, 6, 5]. For example, if the number of data vectors in an input FIFO increases, this means the self-timed processor is not processing data quickly enough, which should signal in increase in the power supply voltage (see Figure 3-12). It is possible then, to use the number of filled elements in the buffer as the input to the feedback compensation network. In [16], an analysis of the power savings of a self-timed circuit with a variable supply is performed, and the use of FIFOs to measure the state of the circuit is introduced. An analysis of the optimal selection of the length of the FIFOs for such circuits, and the application of this method to
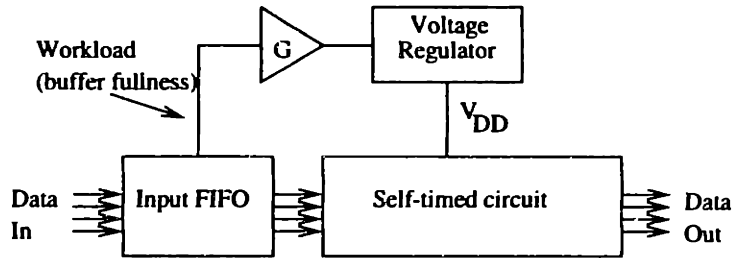
54

Figure 3-12: A circuit using queues to control the power supply. The power supply is adjusted upwards if the input FIFO begins to fill up.

synchronous circuits is presented in [5]. The dual-rail coding required by self-timed circuits creates additional dissipation, so self-timed circuits are inherently less efficient than traditional static CMOS logic.

It is quite possible that when using an input or output queue to measure the speed of the digital logic, the dynamics of the queue emptying and filling will be important to the stability of the closed loop system. If the characteristic response time of the queue is on the order of the response time of the power supply, accurate modeling of both dynamics will be necessary in the design of the feedback gain. It is most likely that the number of filled states in the input or output queue will be available as a digital word, so this method of control lends itself quite strongly toward an all digital controller.

3.4   Implementation of Control Circuitry

Essential functions of a switching power supply are selecting the desired value of the duty cycle, creating the actual pulse width modulated signal, and operating the switches in the power circuit. In a typical non-resonant analog controller, the desired duty cycle is created by using an operational amplifier with some gain, and the pulse width modulation signal is generated by using a comparator to compare the desired duty cycle signal level to a ramp signal (see Figure 3-13. The ramp signal, generated by an oscillator, often will also serve as a clock for the switching frequency.

The possibility of replacing the analog circuitry described above with digital circuits is currently being investigated [31]. There are several reasons that digital controllers are attracting interest for low power supplies. When you are comparing frequencies, the feedback signal will be readily available in a digital form. Secondly, there is hope that digital controllers might provide lower power solutions, without requiring off-chip components or trimmed components on-chip. In order to reduce the power of digital circuits, the power supplies have an increasing interaction with the actual system (for example variable voltage supplies). Since the systems they are interacting with are digital, it is useful for the controller to also be digital (in at least some aspects). This allows
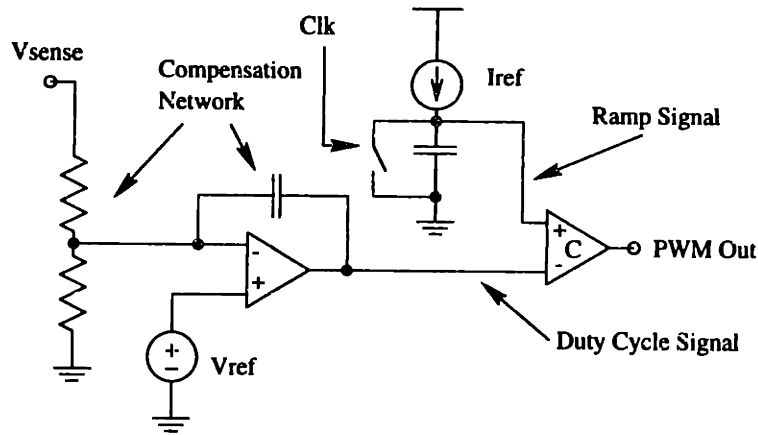
Figure 3-13: Schematic of basic analog controller for a PWM power supply.

the load to precisely specify the desired supply voltage, gate delay, or operating mode to the supply controller without D/A or A/D conversion. Finally, it may also be the case that as digital designers attempt to construct power supplies, they seek to implement the types of circuits they are familiar with, rather than the conventional analog circuits.

It has certainly not been demonstrated that digital controllers are the lowest power solution for low power PWM controllers. It is possible analog or hybrid controllers may win the price/performance tradeoff for low power control. The purpose here is to investigate the ways that digital controllers can be constructed and to evaluate their performance.

A digital controller must at least consist of two parts. It must in some way generate the binary word which represents the desired duty cycle, and it must then convert that binary word into a pulse width modulated signal to feed to the output switches.

3.4.1 Feedback

The methods employed to measure the error in the current output voltage differ between the various methods of control which could be used. In any case, once an error signal is produced (be it one bit or multi-bit), that signal can be subjected to a digital filter in order to shape the compensation of the feedback. The simplest such filter is a counter, which will operate as an integrator, counting up or down depending on the sign of the error. More sophisticated filters could be constructed, by summing the outputs of a counter with scaled and/or delayed versions of the error signals [31]. There is little or nothing to be gained by using a complicated filter if the input has very course resolution (one or two bits). With fine resolution input signals, the value (in terms of improved frequency response) to be gained by utilizing such a filter must be weighed carefully against the additional power consumption and area required by it.
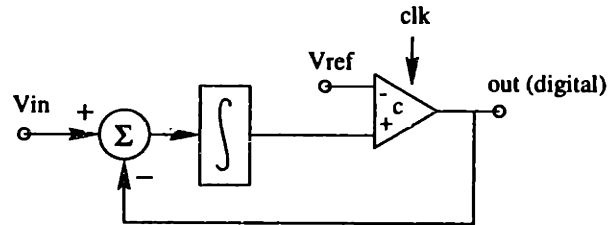
56

Figure 3-14: Sigma delta converter used to create digital stream which on average is proportional to the error $V_{in} - V_{ref}$.

Setting output voltage to a reference value

When the desired function of the control circuit is to drive the supply voltage to a reference value, some form of analog to digital conversion is required. The options available are a single bit A/D converter (comparator), a multi-bit A/D converter, or a comparator enhanced with a sigma-delta converter.

Using a comparator to create the feedback signal limits the control input information to "output too high" or "output too low." This information can then be used to increment or decrement the duty cycle word. Although providing a very low power mechanism to monitor the output (the comparator takes occasional samples and does not need to consume static power), the speed of the response is severely limited by how often the output is checked, since there is no measure of how large an error there is in the output. This method could be enhanced by added more than one comparator, perhaps creating a zone where the output is satisfactory, and regions indicating that the output is 20% high or 20% low (for example).

A sigma delta converter would add an analog integrator to the input, and generate a stream of output bits which, when integrated, would provide a signal proportional to the error. The stream of output bits could be the inputs to an up/down counter to create a digital signal proportional to the error in the output. The sigma delta converter may provide a practical tradeoff between a simple comparator and a multi-bit analog to digital converter.

Locking phase of test circuit to a reference clock

If the control is operated as a phase locked loop, where the delay of piece of circuitry is compared to the delay of a reference clock, there is again a problem of converting a signal from the continuous domain to the discrete domain. The phase error is essentially a time, representing the difference between the two delays. This signal could be treated as a single bit quantity, a measure could be taken of which delay is longer, and the duty cycle could be incremented (or decremented) accordingly. Otherwise, the phase error needs to be converted to a digital word by either low pass filtering and using an A/D converter, or by using a fast counter to measure the length of the delay error signal.
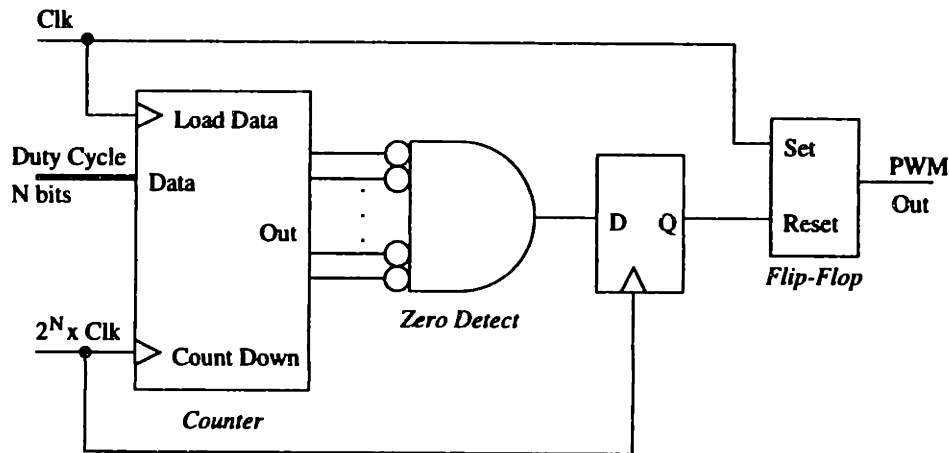
57

Figure 3-15: Creation of PWM signal with a fast clock, using a counter and a flip-flop. This technique is undesirable because it requires running a counter at a high frequency, dissipating an inordinate amount of power.

Setting frequency of test circuit

When controlling the delay of piece of reference circuitry, one can measure the frequency of that reference circuit when wired as a ring oscillator. The frequency of the oscillator can be measured by feeding its output to a counter, and counting for a fixed amount of time (the time reference might be the period of the regulator switching frequency, derived from an external clock). The value on the counter after the fixed delay could then be compared to the desired frequency to create a digital word proportional to the frequency error. Clearly such a scheme is very amenable to digital control, more so than phase locking or comparing the output to a reference voltage. A digital error signal can also be easily derived when using a self-timed circuit with an input or output FIFO; the number of items in the queue can be ascertained (as a digital value) relatively easily.

3.4.2 Pulse Width Modulation

There are several ways to create a pulse width modulated signal from an N-bit digital input word. Some examples are using fast clock (2^N times the switching frequency) and a counter, using N binary weighted fixed delays, or using a 2^N-tap delay line. Each of these options is discussed below, particularly the 2^N-tap delay line, which was implemented on the prototype chip.

Fast Clocked Counter

An N-bit digital input word can be converted to a pulse width modulated signal at frequency f if a clock at frequency $2^N f$ is available, as suggested in [6, 7] [other??], and shown in Figure 3-15. If a high speed clock is already distributed in the system, there is no issue (specific to the power supply)
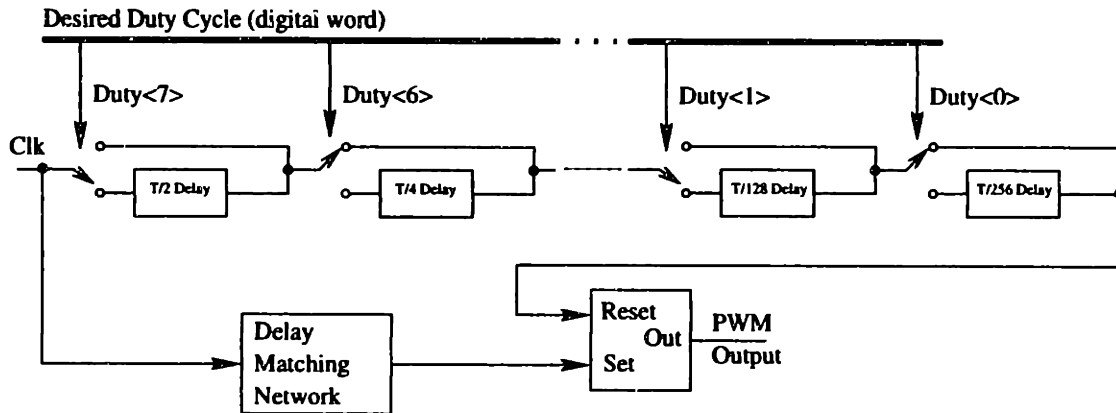
Figure 3-16: Creation of PWM signal with binary weighted fixed delays. This method of creating PWM signals is elegant, but controlling the length of the delays can be difficult, and the transfer function may be non-monotonic in the case of mismatches. A delay matching network is meant to match the fixed delays the *Reset* signal will experience.

regarding the generation of this signal. However, for the example of a 256 level duty cycle for a power supply switching at $500kHz$, the necessary clock speed is $125MHz$, which is a very fast signal to distribute among different chips in the same system. It is quite possible that in some situations there will be no need for the fast clock in the computational blocks (for example, if the system is in a low-computation mode), yet the power supply will still require the faster clock in order to achieve the desired resolution on the output voltage. Creating the fast clock locally would require a phase locked loop, with its associated power dissipation and filter components. Furthermore, the counter which keeps track of the number of fast clock pulses which have occurred must operate at the $125MHz$ rate, and time must be left for the worst case propagation of signals from the low bit of the counter to the highest bit. The constant toggling of the fast clock will consume power, and the necessary speeds for the counter may preclude operation at very low ($1V$) supply levels.

Binary Weighted Fixed Delays

Another way to create a variable pulse width from an N-bit digital input is through the use of binary weighted fixed delays, as shown in Figure 3-16. Each bit of the input digital word is used to either bypass or utilize a different fixed delay. There would be N fixed delays, each representing a binary weighted fraction of the total clock period (i.e. $2^{-1}T$, $2^{-2}T$, $2^{-3}T, \ldots 2^{-N}T$). This method avoids the need for a high speed clock, and it also does not require complicated decoding circuitry to select the output value from the input word. This may save power and chip area over the other suggestions made here.

However, the binary weighted delay method presents a challenge concerning controlling the delays of each of the N delays. The delays could be created by using ratioed current sources or ratioed capacitors. Figure 3-17 illustrates the creation of binary weighted delays using a combination of
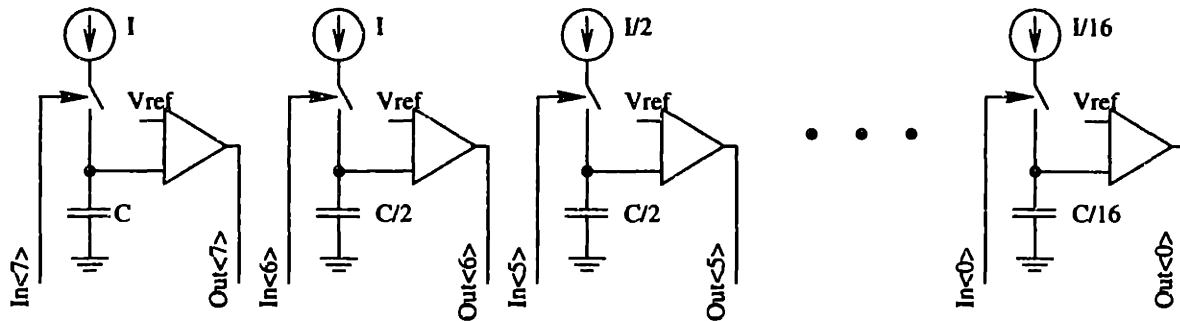
59

Figure 3-17: Creation of binary weighted delays with ratioed capacitors and current sources. Strong resemblances to traditional D/A converters are apparent.
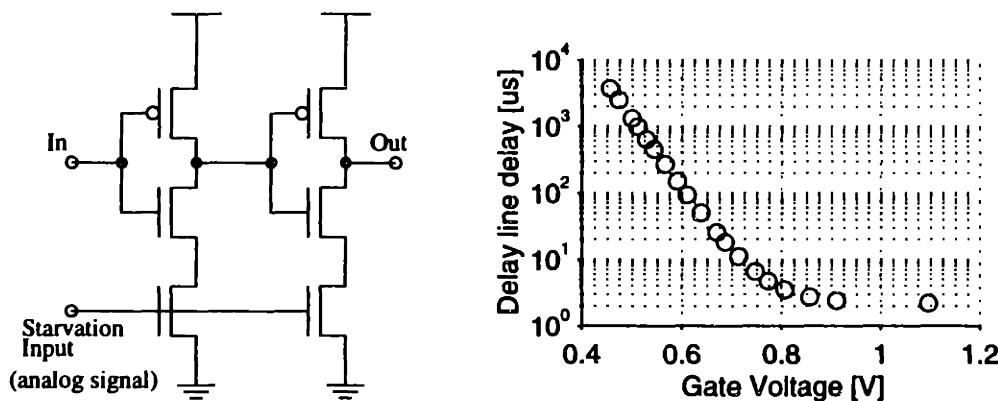

Figure 3-18: Circuit for variable delay buffer using starved inverters. The delay is controlled by the analog level on the gates of the lowest two NMOS devices. Delay of 256 stage chain shown on right. Supply voltage is $1.1V$.

ratioed current sources and capacitors. This would require integrating a reference current source, which will be associated with some static power dissipation. Avoiding mismatch problems will require more power dissipation, due to higher currents in the current sources and larger capacitors. The inherent monotonicity of other methods is not present with ratioed devices.

Binary weighted delays could also be constructed with delay lines (strings of buffers) of different lengths. The length of the delay of each line could be controlled by adding devices to the buffers to create "starved-inverters" (see Figure 3-18). If each delay line incorporated a number of devices proportional to the length of its delay, all the starved-inverter inputs could be adjusted as one. The delay could then be controlled by comparing the delay time of one string (eg. the $2^{-1}T$ chain) to the period of a reference clock.
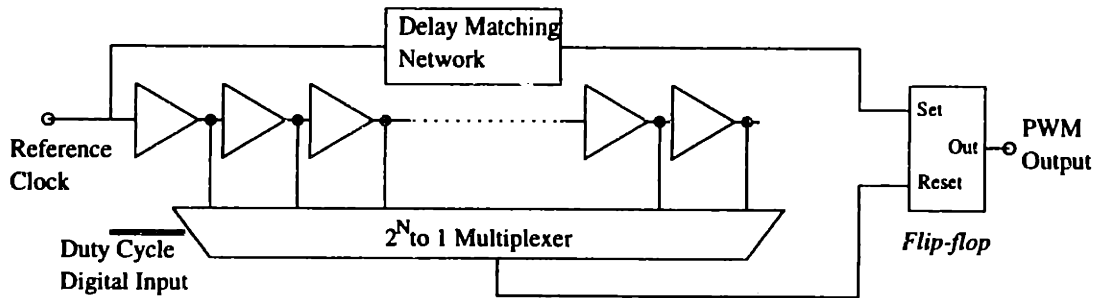
Figure 3-19: A PWM circuit created from a tapped delay line and a multiplexer. Circuit has large area requirements for large values of N, and the delay of the entire delay line must be controlled with a feedback signal. Buffers could be starved buffers shown in Figure 3-18. Delay matching network matches the delay of the multiplexer.

Tapped Delay Line

A pulse width modulated signal can also be created by using a tapped delay line. If the total delay of the delay line is controlled to be equal to the period of the switching frequency clock, an M staged delay line provides M equally spaced outputs which can be used to create a varying duty cycle. For an N bit digital word, a 2^N stage delay line is required.

The essential components of a tapped delay line PWM circuit are the delay line and a multiplexer, as shown in Figure 3-19. A pulse from a reference clock starts a cycle, and sets the PWM output to go high. The reference pulse propagates down the delay line, and when it reaches the output selected by the multiplexer, it is used to set the PWM output to become low. The total delay of the delay line must be adjusted so the total delay is not a small fraction of the reference clock period (limiting the maximum output duty cycle), and not much larger than the clock period (which would create large steps in duty cycle between consecutive code words).

This method of creating pulse width modulated signals avoids using very fast clocks, but can become area intensive, depending on the number of output levels required.

Hybrid Circuits

One way to avoid large area requirements of a full tapped delay line is to combine the tapped delay line method with the fast clock and counter method. For example, to achieve 256 level resolution, a 16 stage delay line could be used in concert with a counter running at 16 times the switching frequency (Figure 3-20). The delay line is used as a ring oscillator, while the counter monitors the number of times the input pulse propagates around the ring. After the signal had propagated the desired number of times (satisfying the most significant four bits of the input word), the output ("end pulse") signal could be selected with a 16 to 1 multiplexer.
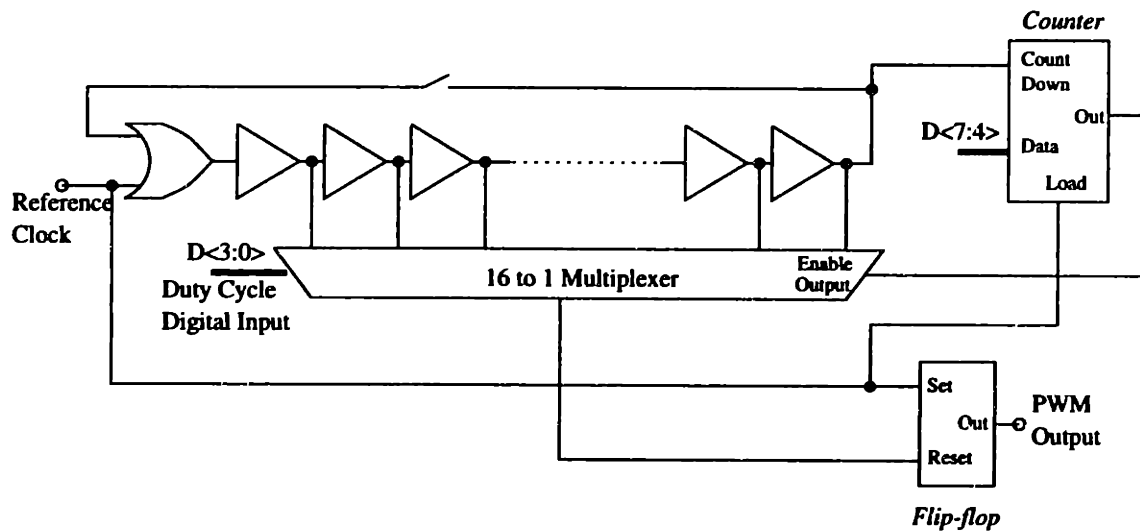
Figure 3-20: Combination of counter and tapped delay line to create a PWM signal. Circuit trades off large areas required by delay line and high speeds required by a counter alone. Counter operates at $16f_{switch}$ to create 256 levels of output.

Chapter 4

Fabricated Chip: Implementation and Results

An integrated down converter has been designed and fabricated, which includes digital control circuitry and an output stage. The fabricated chip was intended to demonstrate an efficient operation while providing the power for the digital pieces of the MIT Ultra Low Power Sensor Project, described in Chapter 1. The output voltage was chosen to be $1V$, with an output power of $5mW$, which is in line with the requirements of the DSP circuitry for the wireless sensor. The controller was designed for an input voltage of $3V$, which is a voltage typical of available Lithium batteries.

The chip implements a delay line based pulse width modulator, as well as digital sampled time feedback. Some advantages sought for the circuit were low operating power, and a wide range of usable duty cycles (10% to 95%). The goal of the chip is to develop a baseline for digital PWM circuits, and provide a basis for comparison against the conventional analog circuits.

4.1   Design of Integrated Down Converter

Rather than using an analog signal representing the desired duty cycle, the duty cycle is stored as a digital word. This word is generated by a very simple DSP circuit (a counter), replacing the typical op-amp. From a circuits perspective, the task of creating the PWM signal from this digital word is the most challenging part of designing the controller. The design of the output stage for the digital control circuit is not unique from the task faced by designers of analog controllers. Due to the low power output needed from this power supply, the output driver could be (and was) implemented on the same chip as the control circuitry.

The volume requirements (total power system including volume in approximately $0.3in^3$) led to the conclusion that an appropriate operating frequency was in the range $100kHz$ to $1MHz$.
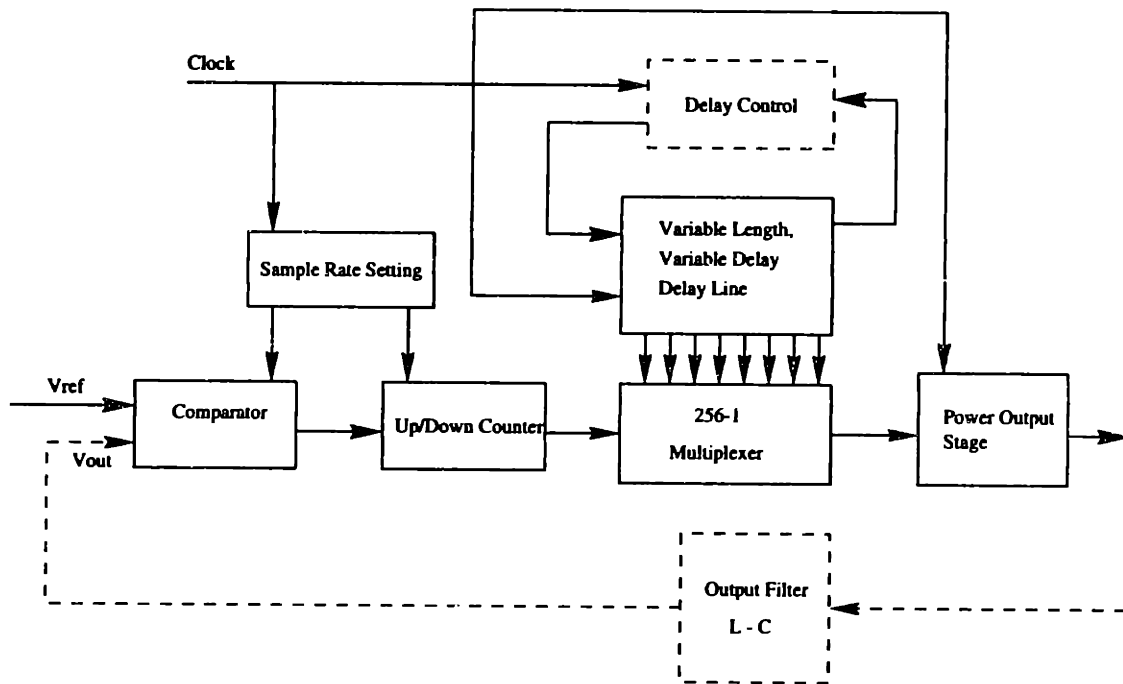
Figure 4-1: Block diagram of implemented converter. Dashed lines represent blocks not integrated on chip.

This frequency selection is based on the optimization of battery and power converter volumes, and the minimum frequencies which the resulting filter volumes correspond to. A precise number for the frequency is not stated here since this could change depending on the final conditions in the system and the types of components (capacitors and inductor cores) available. The likely operating frequency is in the range stated above, and the control circuitry should operated over this range of conditions.

4.1.1 Output Feedback Circuits

Two versions of the power supply chip were fabricated. The first version simply loads its duty cycle from input pins, and therefore the feedback loop must be closed external to the chip. The second creates a feedback signal (a digital word) by comparing the output voltage to a voltage reference (the reference is external) and incrementing a counter to adjust the duty cycle.

The comparison between the output voltage and the reference is carried out at a fraction of the switching frequency. The signal which initiates the comparison is created by a divide-by-N counter, which loads a 16-bit value for N from the external data bus.

After a comparison completes, an up/down counter increments or decrements the duty cycle depending on whether the output was higher or lower than the reference voltage. The counter value is the selection input to a multiplexer (described below) which selects a tap of a delay line. The counter value is changed only after the output of the PWM has gone low. Under this condition, the output of the PWM will not transition until the next switching clock pulse is received, so glitching

64

from the output of the multiplexer will not effect the PWM signal. The counter is also controlled so that it does not wrap around; that is, if the counter value is all ones it will not increment to all zeros, or if it is all zeros it will not decrement to all zeros.

The sample rate (N) was chosen with the help of simulations of the system. Because the output filter responds at a characteristic frequency much lower than the switching frequency, sampling the output (and adjusting the duty cycle) near the switching frequency will result in oscillations. In steady state, on any given sample, the output will read either high or low. If the output is high and the duty cycle is decreased, it is likely that on the next switching cycle, the output will still be high. Therefore, sampling the output at a rate faster than the filter can respond will result in over-compensation. For a representative system, Figure 4-2 shows the relationship between steady state ripple and dividing ratio. The system has a separation of a factor of 15 between the switching speed and the filter frequency. A sample rate at least 3 times slower than the filter frequency ($N = 45$) gives three level duty cycle ripple, and a sample rate 10 times slower ($N = 150$) gives only 2 level duty cycle ripple. Of course, lower values of N will correspond to slower transient responses.

An alternative to choosing a slow sample rate (in order to avoid oscillation) is select a high data rate (as high as the switching frequency) and dividing the value of the output counter in order to create the duty cycle command. That is, rather than dividing the sample rate by N, the duty cycle command could be divided by N. This method allows a faster transient slew rate for the same output ripple, but would correspond to somewhat more power dissipation (since there are more frequent comparisons and counter transitions). A simulation of this method is shown in Figure 4-2(d), where samples are taken every cycle, and the resulting counter value is divided by 15 to create the duty cycle command. The counter value and the duty cycle are shown on the lower plot. On the fabricated chip, the slow sample rate method (not the divided duty cycle method) was used.

4.1.2 Pulse Width Modulator

The number of output levels needed on the pulse width modulator is determined by the desired resolution on the output voltage. For a down-converter with N equally spaced output duty cycles, the resolution on the output voltage is $\frac{V_{in}}{N}$. In the case where the commanded duty cycle is stepping periodically between two adjacent values (as is expected with a 1 bit feedback signal), the magnitude of the output ripple is bounded by $\frac{3V_{in}}{N}$. If the $L - C$ filter is undamped, the maximum of the output voltage ring from a step in duty cycle of one increment will correspond to a change of two increments in the output voltage. A total ripple magnitude of three increments is realized when the duty cycle steps occur both up and down. Rough requirements of $50mV$ maximum output ripple and output voltage resolution of $20mV$ with a $3V$ input voltage dictated 256 levels of output.

It is certainly possible to choose the output duty cycles to not be equally spaced fractions of the period. This design choice trades off resolution in some regions of operation for finer resolution in the
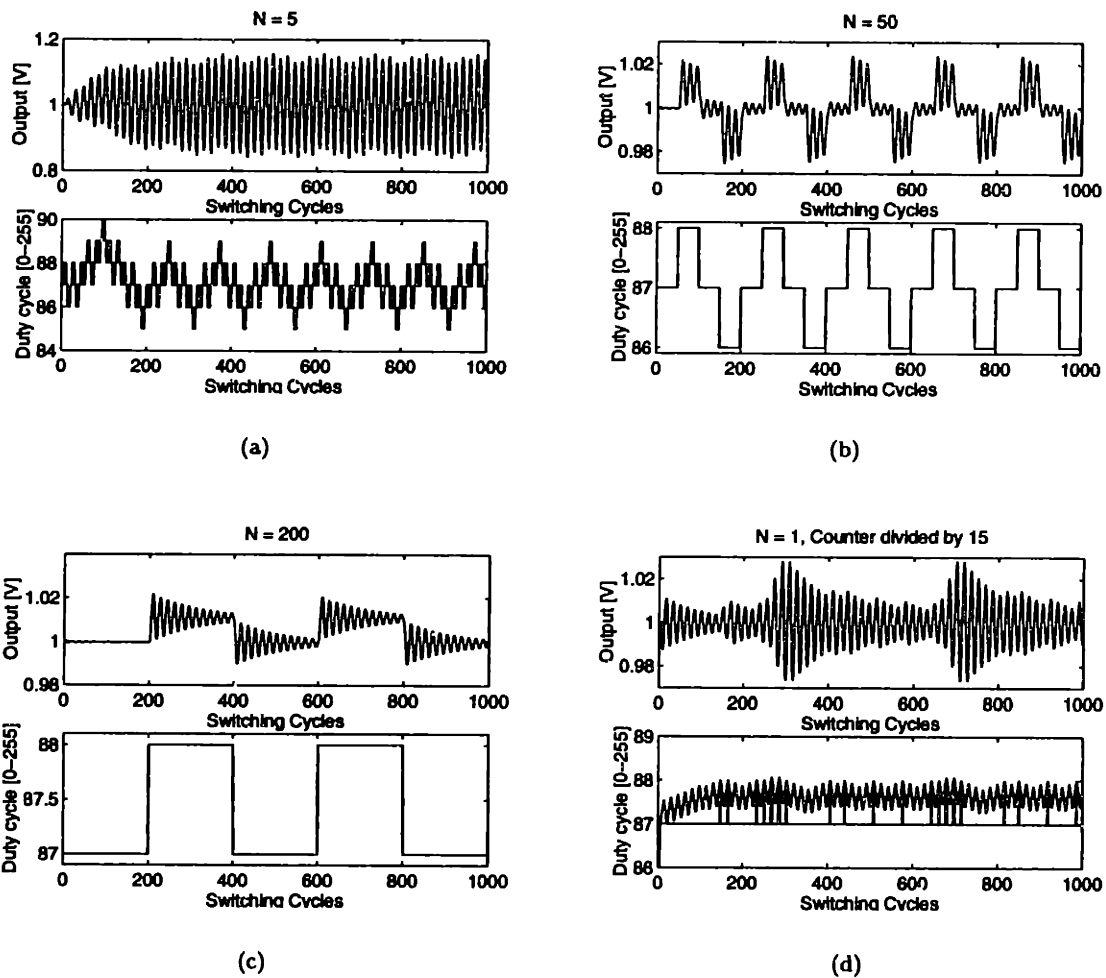
Figure 4-2: Sample rates of $f_{sw}/5$, $f_{sw}/50$, and $f_{sw}/200$ show change in steady state ripple. (d) has a sample rate of f_{sw} but the resulting counter is divided by 15. The counter value is the triangular wave while the duty cycle is flat with several spikes. Input voltage is $3V$, switching frequency is $500kHz$, and filter frequency is $30kHz$. Note that high frequency ring is the filter resonance, not the switching noise. This resonance could be damped by a large capacitance with series resistance on the output.

Figure 4-3: Simulated output ripple as commanded duty cycle steps between two consecutive values. Output filter is under-damped. The down converter has 100 output levels, with a $3V$ input. The total ripple is approximately $3V_{in}/N$, or $90mV$. The commanded duty cycle corresponds to a fraction of the input voltage somewhat larger than the actual output due to source resistance.

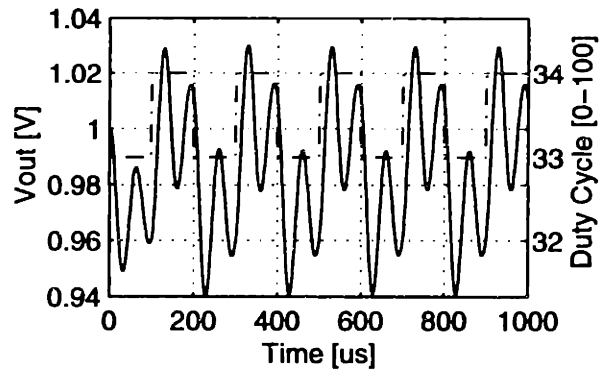region of the expected duty cycle. In the case of a delay line, the line could be constructed with many more stages than taps, with taps most concentrated near the region of anticipated operation and more spread out in the regions where operation is expected under transient conditions. Such careful consideration of the placement of taps is more suitable in situations where the input and output voltages are strictly defined. The design discussed here is meant to be general purpose converter for low voltage battery systems. Neither the input voltages or output voltages are certain.

The digital pulse width modulation circuitry was implemented as a tapped delay line in the fabricated chip. The fast clocked counter discussed in Section 3-15 was ruled out because of the power dissipation which would be caused by the excessive switching of the internal nodes of the counter (even given the assumption that the counter could be made to operate at a frequency as high as $250MHz$). The tapped delay line approach was preferred to the fixed binary weighted delay approach due to the simplicity of the circuit implementation. Although the tapped delay line is more hardware intensive, there is no worry about non-monotonicity, and the calibration of the delay of the delay line (or fixed delays) is somewhat simpler. Building a design which had a good chance of working was more important than building the most efficient circuit in terms of power dissipation and circuit area.

The delay line is a chain of 256 buffers, the outputs of which pass through a 256 to 1 multiplexer. The multiplexer selects a single output to pass, creating the variable delay. The multiplexer is designed to minimize glitching and power loss. None of the nodes internal to the multiplexer switch until the selected buffer output is activated. The chain of buffers is divided into 64 segments, each of which can either propagate the input signal to its outputs, or stop the rippling signal from continuing. The enabling inputs are arranged so that the signal rippling down the chain of buffers does not propagate beyond the multiplexer-selected output. This offers a large power savings for
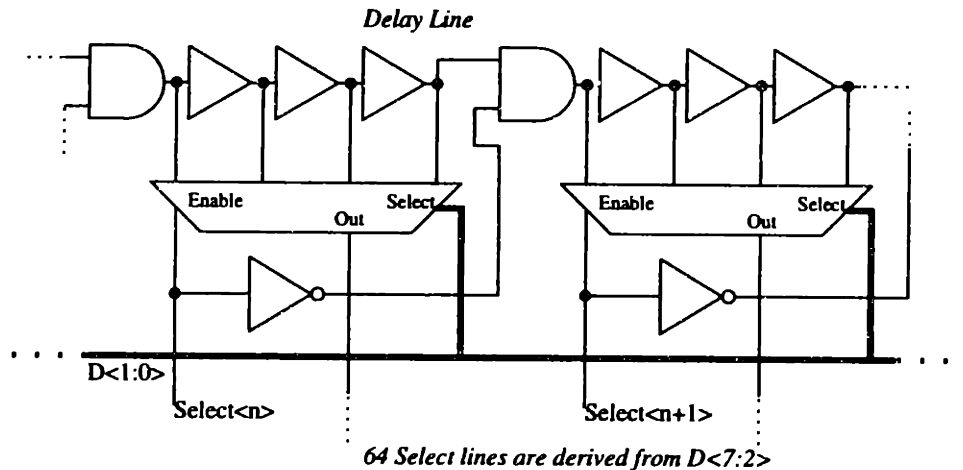
Delay Line

Enable   Select   Enable   Select

Out   Out

D<1:0>

Select<n>   Select<n+1>

64 Select lines are derived from D<7:2>

Figure 4-4: Delay line with selection logic to prevent propagation of input pulse past the selected output.

low duty cycles.

Delay Line

The buffers in the delay line have an input to control their speed; they were implemented as shown in Figure 3-18. It is apparent that both inverters in the buffer will pull high quickly, but pull down slowly, due the starvation N-MOS device in the path to ground. Any edge which is buffered by this circuit will experience both the short and the long delay, so positive and negative edges should propagate through the gate with the same delay. As a result, it is not necessary to have the cut-off transistors on the high and low branches of each inverter (as is typically done in starved inverters).

As a consequence of the design and layout of the delay line and multiplexer, selection signals are available which indicate which block of four buffers contains the tap which the multiplexer is configured to select. These selection signals take the form of 64 "one-hot" signals, each available in the proximity of their respective set of buffers. This arrangement makes it possible to gate the pulse propagating down the delay line to save power. The pulse is prevented from propagating past the set of four buffers which is selected by the multiplexer (see Figure 4-4), eliminating unnecessary switching events in the nodes downstream from the selected taps. Another signal is available to override this feature so that all pulses propagate to the end of the delay line. This is necessary to measure the total delay of the delay line, so the starvation input can be adjusted accordingly. Note that the non-uniformity of the gates in the delay line (buffers mixed with AND gates) will cause non-linearity in the mapping between commanded duty cycle and actual output pulse width. This non-uniformity could be corrected by either careful device sizing, or using AND gates with one input tied high in place of the buffers. No such attempt was made in the implemented design, as the non-linear effect was expected to be small.
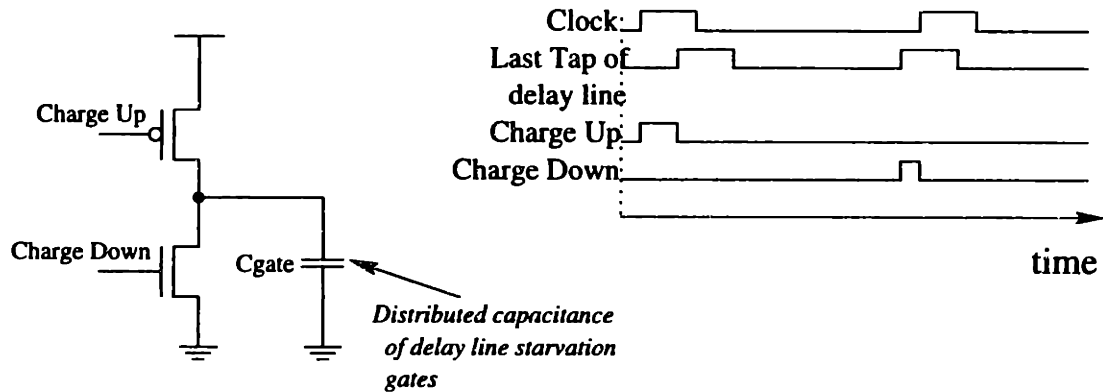
68

Figure 4-5: Circuitry to charge rate control gate of delay line, and sample timing diagram of how control would operate.

Delay Line Length Control

As mentioned previously, the delay of the delay line should be controlled to match the period of the switching clock cycle, in order to get a wide range of output duty cycles and achieve the designed resolution. The control of the delay time is the starvation gate, that is, the gate signal on cut-off n-MOS devices in the bottom of Figure 3-18. All of the starvation inputs of the buffers on the delay line are connected together, so that the delay of the entire line is adjusted at once. The starvation gate signal is an analog signal, where increasing voltage on the gate creates shorter delays through the delay line.

The starvation gate signal can be created by comparing the delay of the delay line to the period of the switching clock, and charging the gate higher or lower, depending on the which delay is longer (see Figure 4-5). In order to match the delay line delay and the clock period, the time of arrival of a pulse to the last tap of the delay line can be compared to the time of the next clock pulse. The control circuitry can determine which pulse arrived first, then charge or discharge the gate signal until the other pulse arrives. This will result in a charge proportional to the time error being added to the gate capacitors. The gain of this loop is then determined by the relative sizes of the charging switches and the total capacitance on the signal node.

Circuitry which implements the generation of the charge-up and charge-down control signals is shown in Figure 4-6. It is assumed that the length comparison is not done every cycle. This helps to avoid the problem that the delay of the delay line may be a multiple of the clock period. If the delay of the delay line is twice the clock period, pulses at the last tap may align perfectly with input clock pulses, and the control circuitry will not converge correctly. When the length comparison is not done every cycle, pulses are only allowed to propagate to the last tap of the delay line when a comparison is desired, and the threat of converging to the wrong delay length is reduced. The sample rate for the length comparison is determined by the leakage on the control gate, the tolerable
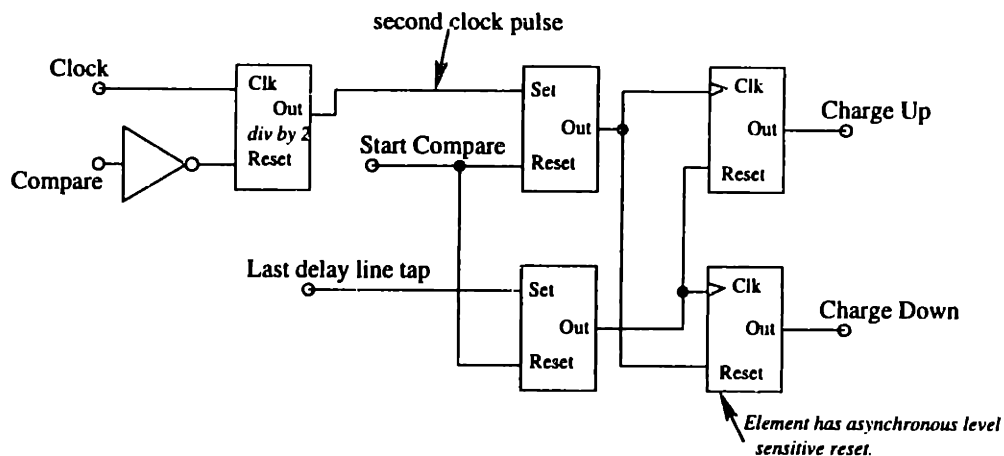
69

Figure 4-6: Generation of charge-up and charge-down signals. Comparison of clock period and delay line delay enabled with *Compare* signal, and initiated with *Compare Start* signal.

ripple of the output duty cycle, and the necessary transient response for the delay length.

It is difficult to do a complete analysis of the feedback system representing the delay line and the associated control circuitry of Figure 4-5 and Figure 4-6. The relationship between the starvation gate signal and the associated delay is exponential at gate voltages below V_T and saturates at higher voltages (see Figure 3-18). Likewise, the amount of charge deposited on the starvation gate capacitors (from the circuits in Figure 4-5) for a given on-time is a non-linear function of the supply voltage and the voltage on the gate. Despite these nonlinearities, it is still informative to perform a linearized analysis, which shows the system response given small variations about the operating point.

The delay of the delay line can be represented by the difference equation

$$t[n] = t[n-1] + K_A(T_{clk}[n] - t[n-1]) + K_B I_{lk}[n] \tag{4.1}$$

where $t[n]$ is the delay, K_A is the linearized gain between delay error and delay correction, T_{clk} is the reference clock period, K_B is the gain from the leakage current to the additional delay, and I_{lk} is the leakage current at the starvation gate node. K_A can be described by

$$K_A = \frac{W_{charge}}{C_{gate}} G(V_{gate}, V_{supply}) \tag{4.2}$$

where K_A is the width of the charging devices (two MOSFETs in Figure 4-5), C_{gate} is the starvation gate capacitance, and $G(\ldots)$ is a linearized function of the operating point. The z-transform of the
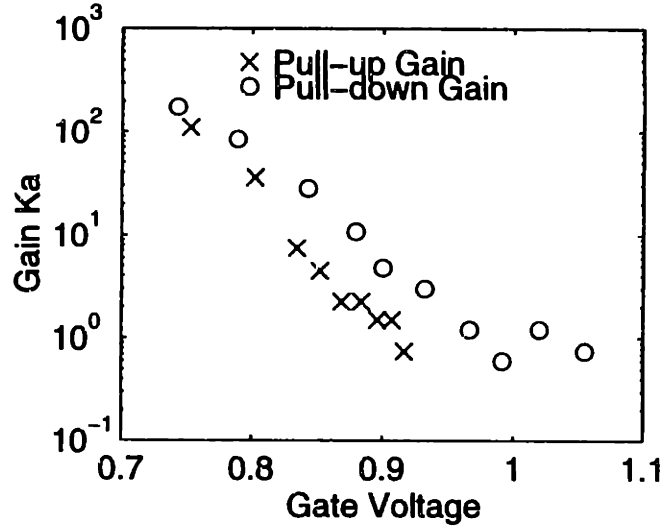
70

Figure 4-7: Nonlinear gain K_A at one set of conditions. Supply voltage is $1.1V$. Lower gate voltages (x-axis) represent longer delay times (lower switching frequencies). Range of gate voltages shown covers typical regime of operation. Two sets of points represent the gain for the pull up and pull down feedback paths.

difference equation (Equation 4.1) is given by

$$t(Z) = \frac{K_A T_{clk}(Z) + K_B I_{lk}(Z)}{1 - (1 - G)z^{-1}} \qquad (4.3)$$

yielding the impulse response (given an impulse at $T_{clk}[n]$)

$$t[n] = G(1 - G)^n u[n]. \qquad (4.4)$$

The system is stable for $0 < K_A < 2$ (determined by requiring the poles of the z-transform to be within the unit circle). The impulse response of the system is oscillatory of $K_A > 1$. The steady-state error, found by setting $t[n] = t[n-1]$, is $\frac{K_B}{K_A} I_{lk}$. (See [19] for discussion of difference equations, the Z-transform, and stability of discrete time systems.)

The non-linear gain, $K_A(V_{gate}, V_{supply})$ was measured using a SPICE simulation. Figure 4-7 shows the gain measured at various values of V_{gate}, for a fixed value of V_{supply}. There are separate curves for charging up and charging down (when the error is positive or the error is negative), since these two gains (positive and negative) are accomplished with different devices with somewhat different gains. The charging devices were $1\mu m$ wide and $1\mu m$ long, and the process was a standard $0.6\mu m$ CMOS process. The gate capacitance present (C_{gate}) is just that presented by the starvation gates. The graph shows that in order to maintain a stable system, at low operating frequencies (corresponding to low gate voltages), it will be necessary to add capacitance to the gate node. At higher operating frequencies, larger charging devices may be necessary for fast impulse responses.

The delay line length control was not integrated on chip with the other control circuitry, due to

time constraints. This portion of the control was implemented with discrete gates external to the chip. The charging devices were implemented with a tri-state output buffer. A resistor was added in series with the buffer to limit its driving ability, and capacitance was added to the starvation gate node for noise immunity (this capacitance also acts to reduce the gain and enhance stability). This external circuitry is not counted in the control circuit power consumption, but the additional power of the delay line length control will be very small. The length control circuit has very few nodes which switch, and it operates at a frequency below the switching regulator frequency.

256 to 1 Multiplexer

The design criteria for the multiplexer stage were minimal superfluous switching activity and glitch-free output. The design is realized with a flat multiplexer, where only the selected input propagates (causes switching transitions) to the outputs of the input gates. Each of the inputs to the multiplexer are fed to one input of a 3-input AND gate. The other two inputs to the gate are selection signals, one derived from the six most significant bits of the digital input word (selected duty cycle) and the second derived from the two least significant bits if the digital input word. As a result, only one of the 256 input AND gates will have both of its selection gates high, and only this gate will propagate the pulse on its input gate to its output. The outputs of the AND gates are input to a 256-way OR gate (implemented with a tree of gates), and the output of this gate is used to signal the end of the high time of PWM output pulse.

Once the selection signals have stabilized, there will be no switching on the output or internal nodes of the multiplexer (besides a node internal to each of the AND gates). The output of the multiplexer is input to the R input of an set–reset latch. The S signal on this latch is a delayed version of the clock pulse which is the input to the delay line. Therefore, after the output of the multiplexer goes high, glitching on its outputs will not effect the pulse width modulated signal as long as it occurs before the next S signal is received from the input clock. To take advantage of this, the selection signals (the digital word which determines the duty cycle) are modified after the output of the multiplexer is activated. The time it takes for the selection signals to stabilize limits the maximum duty cycle the delay line can produce, since it must be guaranteed that the next cycle does not start before the multiplexer has set up the selection signals for the next cycle.

4.1.3   Output Stage

The output stage drives the load with the pulse width modulated signal created by the delay line and multiplexer. The first part of the output stage is a level converter, which translates the voltage levels of the signal from the levels of the digital control circuitry (as low as $1V$) to the levels of the output drivers (the battery voltage, around $3V$). The level converter passes its signal to a chain of buffers of increasing size, which finally drive the output transistors. These buffers are designed

as CMOS digital inverters; no attempt is made to skew the timing of the gate drives to reduce shoot-through current.

The chain of buffers were sized so that the device widths of each buffer was approximately 3.8 times wider than the buffer before it. This roughly corresponds to the sizing suggested in [15, 33] for CMOS cascaded buffers. This sizing is meant to minimize total delay. It may be more appropriate to design this cascade of buffers according to the theory put forth by Veendrick, in [29], which derives a minimization with respect to power dissipation, not total delay. Following the recommendations of [29] would lead to a much larger ratio between subsequent stages for the modern CMOS process utilized for the test chip. A closer look at the tradeoff between total delay and total power dissipation in the drivers needs to be made.

The output is driven with a large n-MOS device and a large p-MOS device. A down-converter can use a diode in place of the n-MOS device, but given the low output voltage, a diode drop would produce very significant losses. The output devices were sized to minimize their total dissipation under full load, $5mW$, but at a duty cycle of 50%. This corresponds to an output voltage of $1.5V$, somewhat larger than the output voltage stated previously. The result of this discrepancy is that the high side device is oversized for the $1V$ application, and the driver will dissipate somewhat more power than is necessary. It was assumed that the gates of the devices would be driven the full swing of their power supply, and no provision was made to selectively activate portions of the devices under lower load conditions.

Equation 3.28, which gives the minimum switch loss under a given set of conditions, relies upon the parameter $\mathcal{Q}'$ to encapsulate the characteristics of the semiconductor technology used for the switch. For the $0.6\mu m$ CMOS process used, this parameter was extracted using Equation 3.26, restated here.

$$Q' = \frac{V_{GS}C_x L^2}{(V_{GS} - V_T)C_{ox}\mu}.$$

(4.5)

Given the technology parameters and the operating conditions, finding the switch sizes for minimum dissipation is a matter of applying Equation 3.30, restated here:

$$W = \frac{L}{\mu C_{ox}(V_{GS} - V_T)}\sqrt{\frac{I_{rms}^2}{f_{sw}Q'V_{GS}}}.$$

(4.6)

The frequency for this optimization was chosen to be $300kHz$, the geometric mean of the range of frequencies the system is designed for. The p-MOS device size fabricated was $3.8mm$, and the n-MOS size was $1.8mm$.

The layout of the power switches was designed to reduce the drain capacitance and the wiring resistance. The layout is depicted roughly in Figure 4-8. The drain side wall capacitance is reduced by enclosing the drain region with the gate. The source and drain interconnect runs perpendicular to the rectangular drain and source regions, "strapping" across them. This layout uses very few net
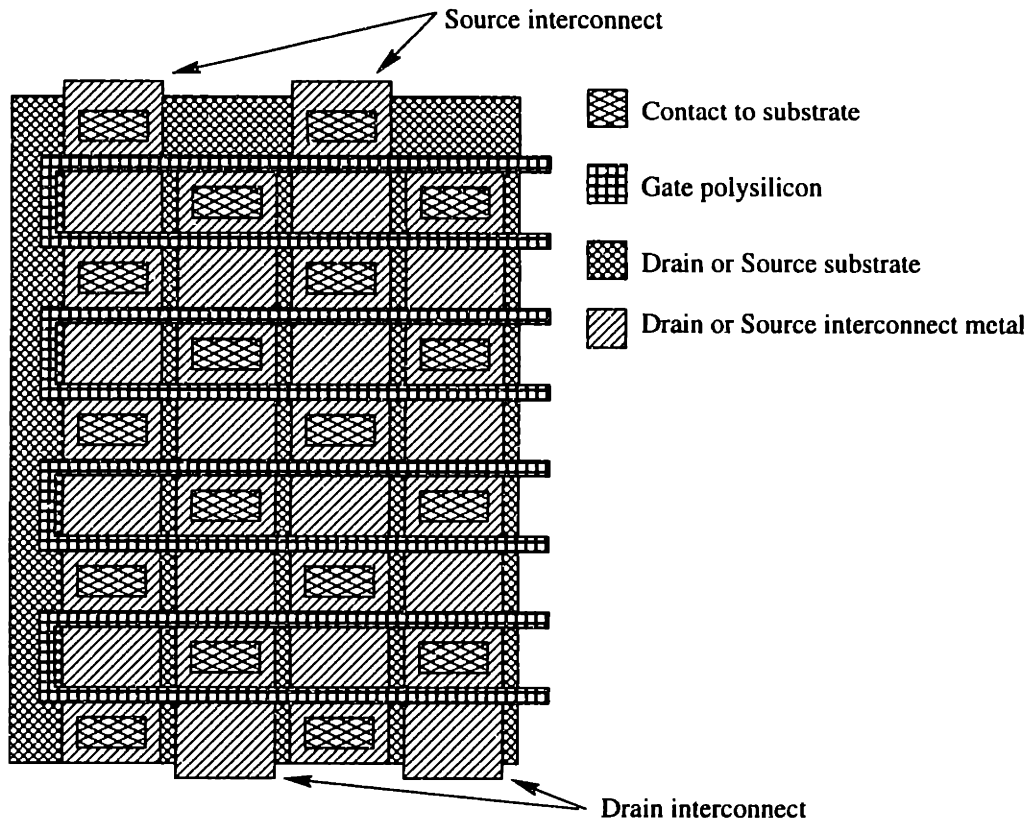
Figure 4-8: Schematic of power transistor layout.

squares of metal to create a wide transistor in a fairly dense manner.

4.1.4 Dynamic Comparator

In order to compare the filtered output voltage level to a reference voltage, a comparator is implemented on chip. The comparator is fully dynamic, which means it only carries out a comparison (consuming power) when it receives a command to do so. When the command signal to evaluate is logically low, two internal nodes charge high (precharge state). When the evaluate signal is high, those nodes are slowly discharged by FETs whose gates are connected to the two input voltages. Positive feedback (from one node to the other) detects which node is discharging faster (hence has a higher FET gate voltage) and latches the output high or low accordingly. A schematic of the circuit is shown in Figure 4-9.

4.2 Results

The die size of the fabricated chip is roughly $2.3mm$ by $2.4mm$, although the core of the chip is significantly smaller than that. One thing that is somewhat surprising is the tiny dimensions of the $5mW$ output transistors. The two output transistors occupy an area of only $125\mu m$ by $190\mu m$,

Figure 4-9: Schematic of dynamic comparator included on test chip.

which is similar to the size of an bond pad ($100\mu m$ by $100\mu m$), and dwarfed by the size of the delay line with 256 stage multiplexer, $600\mu m$ by $1100\mu m$. This demonstrates the assertion that at very low power levels, volume tradeoffs should consider only the filter size, not the semiconductor area of the switches. A die plot is shown in Figure 4-10. In general, the chip operates to voltages as low as $1.0V$. However, most tests were carried out at higher voltages, due to limitations imposed by external circuitry, such as the delay line length control.

4.2.1 Delay Line

The delay of the delay line can be varied over several decades. The measured delay is plotted in Figure 4-11 against the gate voltage on the starvation input which corresponds to this delay. With a $1.1V$ supply voltage, the lower limit on delay was $2\mu s$, while delays as long as $4ms$ were measured. With a $2.1V$ supply, the delays were measured as short as $950ns$. These delays were measured open loop, in that the delay length control was not active, and the gate voltage was controlled by a potentiometer. Measurements of delays longer than $4ms$ were not taken simply because such low operating frequencies are not of interest.

One flaw in the delay line is that the width of pulses "grow" as they propagate through the line. The low-high transitions on the input propagate faster than the high-low transitions. An observation of this phenomenon is shown in Figure 4-12. The pulse growth has an exponential relationship with gate voltage (Figure 4-13), very similar to the relationship between gate voltage and delay line delay. The pulse growth is caused in a mismatch in loading between the two inverters of the starved buffer
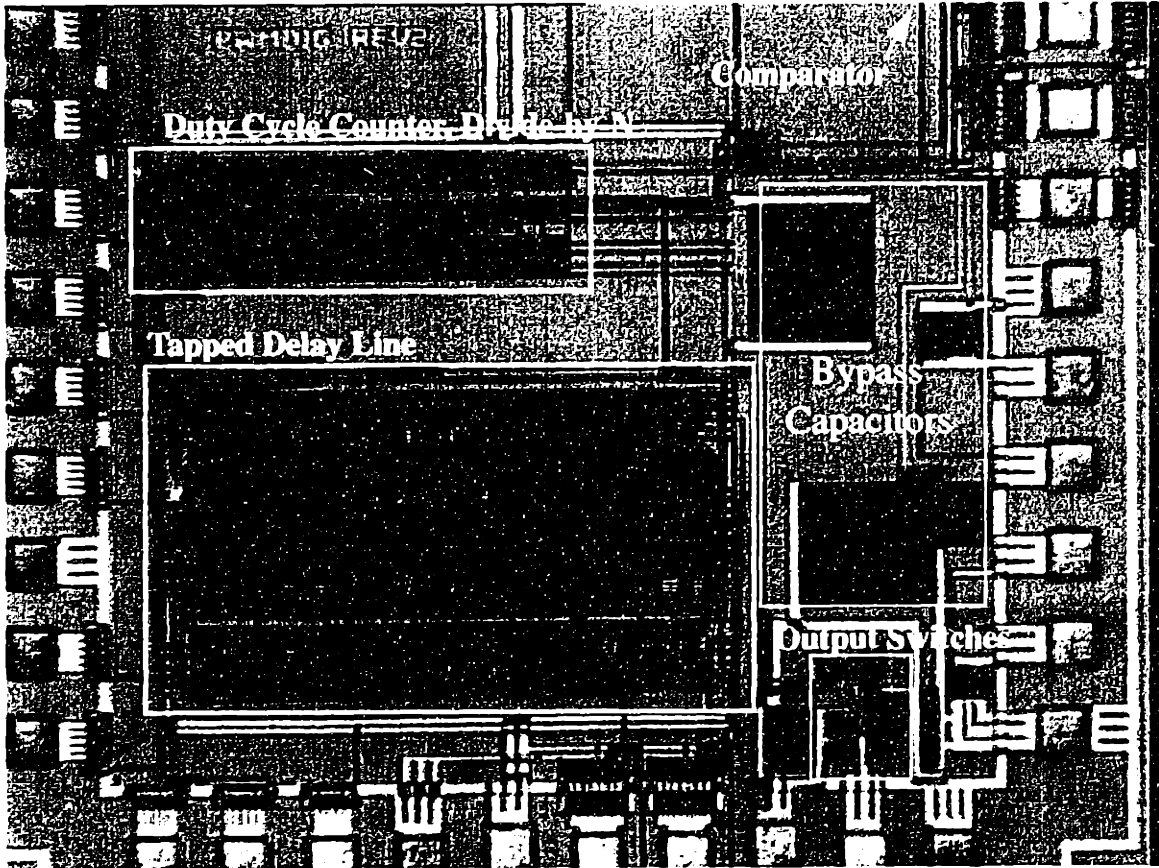
75

Figure 4-10: Photograph of fabricated chip. Includes a 256 tap delay line (lower left), a 5mW output stage (lower right), and a dynamic comparator (upper right, off of photo). The output driver is dwarfed by bypass capacitors and the delay lines.
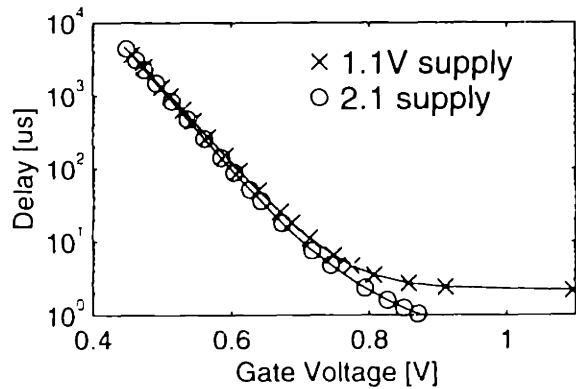

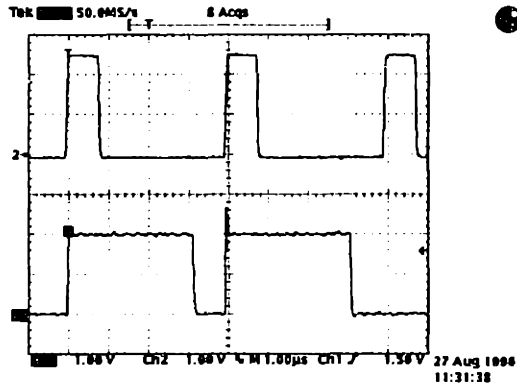Figure 4-11: Measured delay of delay chain.

Figure 4-12: Pulse on input to delay line (upper trace) "grows" inside the delay line, since low-high transitions propagate more quickly than high-low transitions. Pulses on the output of the last tap of the delay line are shown on the lower trace.
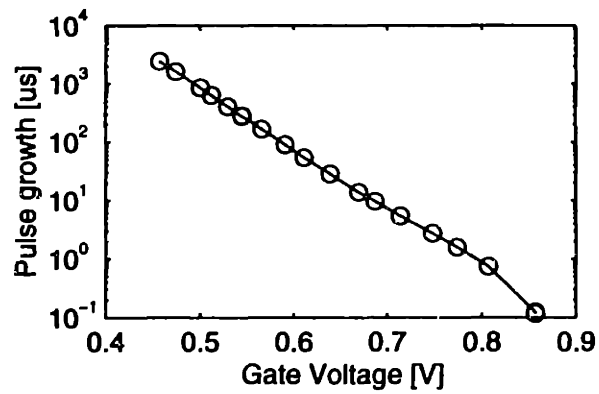

Figure 4-13: Increase in pulse size at output of delay line shown as a function of gate voltage. Supply voltage is 1.1V.
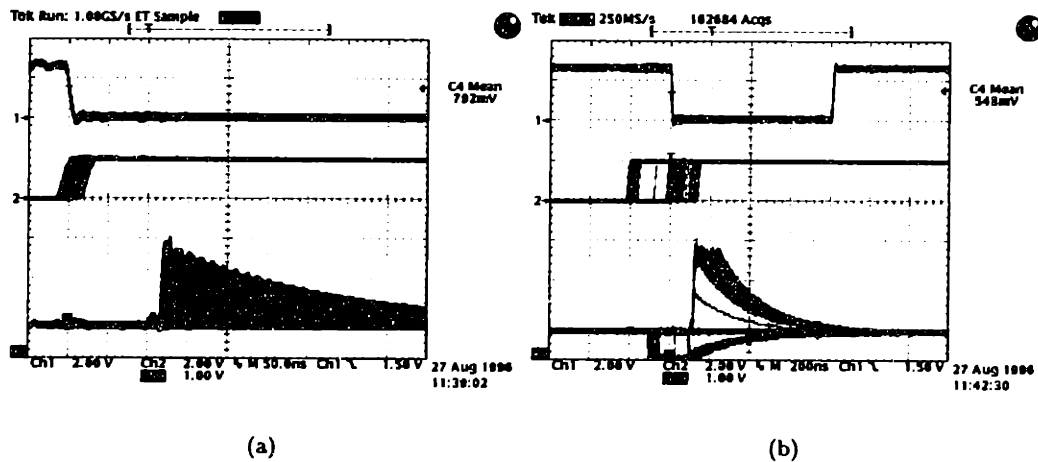
Figure 4-14: Operation of delay line length control, oscilloscope in long-persistence mode. Top trace is an inverted version of the clock pulse which the delay phase is being compared to. Second trace is the output of the last tap of the delay line. Third trace is the output of the tri-state buffer which drives the starvation gate through an R-C network. When last tap pulse is late relative to the clock, the starvation gate charges up, when it is early, the gate charges down. Clock frequency is $250kHz$ in (a) and $40kHz$ in (b). (b) shows oscillations in delay length for long delays.

(Figure 3-18). The first inverter is loaded only with the input to the next inverter, while the second inverter is loaded with the input to the next stage as well as an input to the multiplexer. As a result, high-low transitions on the second inverter are slower than high-low transitions on the first inverter. The leading edge of the input pulse experiences a high-low transition on the output of the first inverter and also a fast low-high transition (due to the p-MOS devices, which are not starved) on the output second inverter. The trailing edge experiences a fast low-high transition on the first inverter, but a slow high-low transition on the output of the second inverter. This problem could be corrected by carefully matching to loading on the outputs of the two inverters.

The delay line length control, implemented in external circuitry as described above, is able to lock the delay line delay time to a reference clock over a wide range of operation. The dependence of the gain of this circuit on operating frequency (as shown in Figure 4-7, where lower gate voltages represent lower frequencies) is observable in a qualitative sense. While the phase locking operates well at high frequencies with a certain amount of added gate capacitance, at lower frequencies, oscillations can be observed, as the charging devices over compensate at each comparison (both situations are shown in Figure 4-14). Additional gate capacitance decreases the magnitude of this oscillation.

The pulse growth presents problems for the delay line length control circuitry. The length control operates by comparing the timing between a pulse at the last stage of the delay line to a clock pulse. Pulses which entered the delay line on previous cycles should never reach the last tap, due to the mechanisms which prevent pulses from propagating past their selected output. However, when the
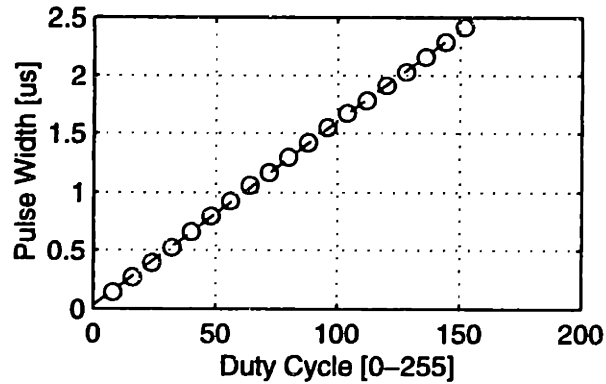
78

Figure 4-15: Pulse width for given duty cycle word, f_{sw} is $250kHz$. Least squares fit is also shown; the average (RMS) error is $0.011\mu s$, which is on the order of the measurement resolution.
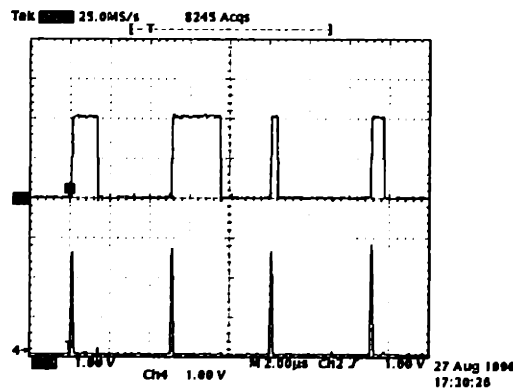

Figure 4-16: Pulse width modulated signal, with clock. Pulse width is cycled through the hexadecimal duty cycle words 10, 20, 40, 80.

pulse widths increase within the delay line, it is possible a pulse will still be in the delay line when the next cycle starts (its trailing edge has not reached the selected output). This can confuse the length control circuitry, which may interpret a pulse from a previous cycle as a fast propagating pulse originating in the current cycle. Under certain conditions with high duty cycles, the delay length control may lock onto a delay length which is some multiple of the actual clock period through this mechanism. This situation might be prevented in the future by actively resetting the delay line before the cycle in which a length comparison takes place, rather than relying on the current circuitry.

4.2.2 Pulse Width Modulation

The pulse width modulator (delay line with multiplexer) works as expected (save the unanticipated bugs in the delay line mentioned above). Figure 4-15 shows the linearity of the PWM at a high level. More precise measurements have not been taken. Figure 4-16 demonstrates changing pulse

79

widths; the pulse width is being loaded from an external function generator each cycle.

The most serious limitation on the PWM is the upper limit (somewhere between 50% and 75%, depending on operating conditions) on the duty cycle, imposed be the interaction of the pulse growth problem and the delay line feedback circuitry. A second problem which has not been fully characterized is jitter in the phase locking of the last tap and the input clock. Acceptable limits on this jitter need to be determined, and the constraints these place on the delay feedback circuitry evaluated.

The dissipation in the combined delay line and multiplexer circuit is $10\mu A$ with $2V$ supply voltage and a $250kHz$ switching frequency. Operation was observed to voltages as low as $1.1V$, however, limits imposed by external circuitry used prevented the measurement of such operation while the delay feedback circuitry was being used. The core current dropped to $7\mu A$ at a $1.5V$ supply. Switching frequency scales the power dissipation proportionally, as expected, doubling to $19\mu A$ at $500kHz$, and again to $38\mu A$ at $1MHz$. The above numbers are for the case where the duty cycle was 12.5%, and half of the pulses input to the delay line propagated to the last tap, while half stopped 12.5% down the line. When the duty cycle was increased to 50%, current draw increased to $13\mu A$ at $250kHz$ and $24\mu A$ at $500kHz$. The expected increase for this change is an increase of the ratio 4/3, which corresponds very closely to the observed increase. The fraction of pulses which need to propagate the entire length of the delay line is a function of the maximum jitter on the delay length, but this relationship has not been investigated for this system.

4.2.3   Output Feedback Circuits

Unfortunately, a logic error on the fabricated chip prevents the integrated feedback circuitry from receiving the outputs of the dynamic comparator. The divide-by-N circuitry works properly, activating the comparator every Nth cycle, but the outputs of the comparator cannot cause the duty cycle counter to increment or decrement. A small repair may be made to the chip in the future to allow characterization of these circuits. However, results are not available at this time.

Some of the functionality of the fabricated chip was implemented with external circuitry to demonstrate the performance of a single bit controller. Measurements are shown for the case that the output is sampled every 100 cycles. The external counter does not update the duty cycle until a second 100 cycles has passed; the result of this extra delay is that the steady state ripple is larger than it would be otherwise.

For the situation tested, the duty cycle stepped through four duty cycle levels. This did not improve (decrease) for lower sample rates (output monitored every 200 cycles and 300 cycles). This may be a result of the extra sample period delay introduced by the external up/down counter, but this needs to be verified. Each duty cycle level represents one-half of the actual switching frequency ripple; this fact may also account for the larger than expected limit cycle ripple. The transient
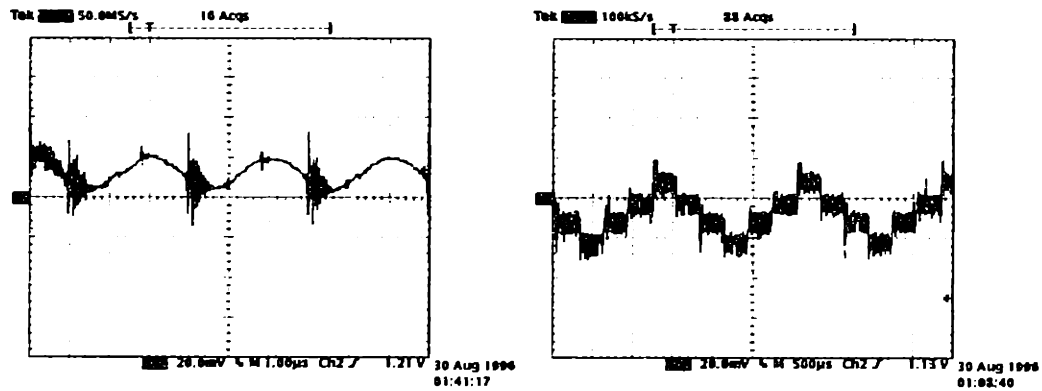
(a) Switching frequency output ripple.　　(b) Single bit feedback limit cycle ripple.

Figure 4-17: System under closed loop feedback control, duty cycle updated every 100 cycles. Clock frequency is $330kHz$, output power is $5mW$. Arrow on left indicates $1.2V$ level.

response is slew limited, as expected, due to the slow sample rate and lack of proportional feedback. Though slow, the response is certainly not unstable.

4.2.4　Output Stage

The output power stage was tested first driving only the package pin of the test chip (a 44 pin PLCC). With no connection to the test board (limiting capacitance at the pin), and a $3V$ power supply, the driver consumes $19.5\mu A$ at $250kHz$, and $38\mu A$ at $500kHz$. These values correspond almost exactly to the power required by the gate drive of the output devices (and their buffers). This indicates that the capacitance of the output pin, wire bond, and package lead are small relative to the gate drive. This also shows that the amount of shoot-through current of the output driver is limited.

The resistance of the power FETs was measured as a function of the gate voltage, and the results are shown in Figure 4-19. This data will be helpful for designing power FETs for CMOS processes in the future. The portion of the resistance which is not modulated by the gate voltage is larger than predicted in SPICE models for the process, and closer analysis of the characteristics of CMOS power transistors is warranted.

The output power stage was also tested with an L-C filter. The inductor of the filter had an inductance of $220\mu H$, and the capacitor had a capacitance of $0.22\mu F$. These values were chosen to provide a filter frequency approximately 15 times below the switching frequency of $300kHz$. The predicted inductor ripple current was $10mA$, a reasonable value for a $1V$, $5mA$ output. The capacitor was an $80mil$ by $50mil$ surface mount device, and the inductor was a $120mil$ by $100mil$ surface mount component. The magnetic path of inductor has a relatively long path through air,
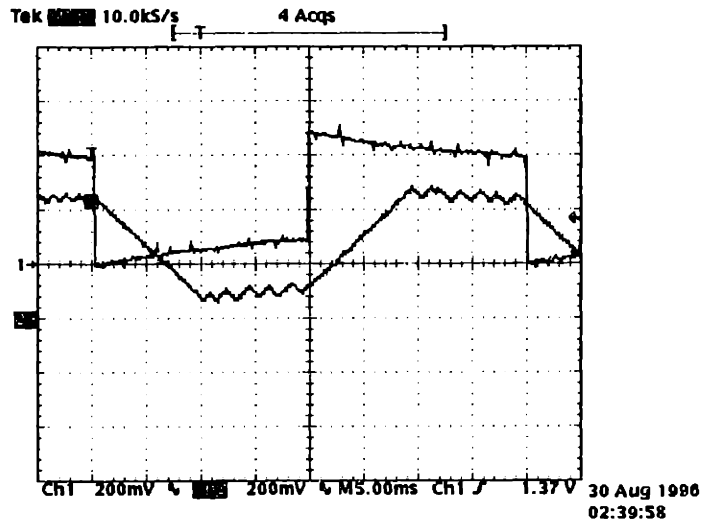
81

Figure 4-18: Transient response of closed loop power supply. Upper trace is the reference voltage, lower trace is power supply output. Traces have been separated vertically by 200mV as a visual aid. Arrow markers on left side correspond to a voltage of 1.2V.
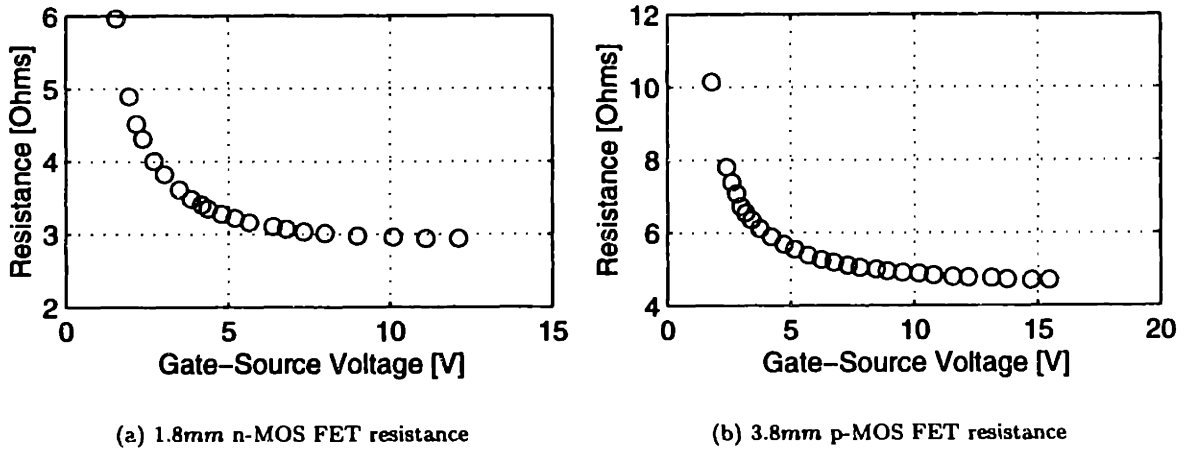

(a) 1.8mm n-MOS FET resistance

(b) 3.8mm p-MOS FET resistance

Figure 4-19: Resistance of power FETs over gate voltage. Drain current for both cases is 5mA.

Dissipation Cause	$I_{out} = 5mA$	$I_{out} = 0mA$
Gate and Drain Capacitance	$70\mu W$	$70\mu W$
DC Current in Inductor	$240\mu W$	0
DC Current in Switches	$115\mu W$	0
Ripple Current in Inductor and Switches	$120\mu W$	$120\mu W$
Other, possibly crossover loss	$185\mu W$	$210\mu W$
Total	$730\mu W$	$400\mu W$

Table 4.1: Estimated breakdown of power dissipations with L-C filter connected to output, $3V$ in, $1V$ out.

so many more turns are used than would be necessary for an inductor with a closed magnetic path through a permeable core. As a result, the inductor resistance is 9.5Ω, (two orders of magnitude larger than what could be achieved with a toroidal core).

With $3V$ input and $1V$ out, the system is 88% efficient at $5mW$ output power. At this power, the power supply dissipation is $730\mu W$. This can be attributed to the $I_o^2 R$ drops in the inductor and output switches, losses due to the ripple current, and possibly crossover losses. The dissipation caused by the DC current in the inductor and switches accounts for $350\mu W$ dissipation. With no load, the system power dissipation is $400\mu W$; of this, approximately $120\mu W$ can be attributed to the ripple current in the inductor and switches, and $70\mu W$ to the gate drive. The source of dissipation for the remaining $210\mu W$ could be crossover losses, but more investigation is required before this is shown conclusively. Certainly, $400\mu W$ dissipation at no load is too high if efficient operation over a wide range of operating conditions is required. One way this may be reduced is by using discontinuous mode operation at light loads. In the current configuration, the inductor current becomes negative under many circumstances; the output n-MOS device should be turned off when this happens to reduce current ripple.

Chapter 5

Micromechanical Power Conversion

Inductor based switching regulators can be used to create precise output voltages, above or below their input voltage, and have the additional advantage that they are ideally lossless power converters. Unfortunately, it is currently not possible to construct acceptable inductors on silicon wafers which incorporate other circuitry. Furthermore, discrete inductors are expensive to build and difficult to interconnect with other circuitry. A topology with the advantages of inductor based regulators but does not require an inductor clearly would be extremely useful.

Noworolski and Sanders [17] propose a structure consisting of two mechanically linked variable capacitors. Such a power converter stores energy in the resonant mode of a mechanical diaphragm, and uses the motion of the diaphragm to vary the capacitance of two capacitors, which can be used to transfer energy between two voltage sources. This structure has the capability to precisely control the conversion voltage ratio and to convert power from high voltages to lower voltages or the other way around.

Noworolski and Sanders treat the structure as a resonator, which could replace the inductor-capacitor tank circuit of a resonant converter. They suggest that the conversion ratio ($\frac{V_{out}}{V_{in}}$) can then be controlled by changing the frequency of the electrical excitation (as in a resonant converter) or by modifying the value of a DC bias voltage. Control similar to pulse width modulation at a fixed frequency is also possible. The primary problems with the proposed converter are low power density and difficulties designing acceptably low power control circuitry.

5.1   Operation of Circuit

The proposed structure is a conducting diaphragm, which is suspended above two electrodes, as shown in Figure 5-1. This forms two variable capacitors. The two electrodes are designed so that they are symmetric with respect to the excited mode of the diaphragm, in order to ensure that the two capacitances track each other as the diaphragm moves. The conducting diaphragm forms the
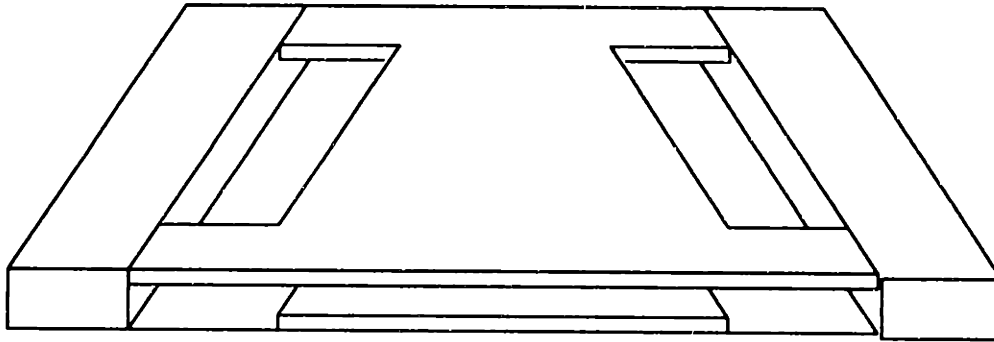
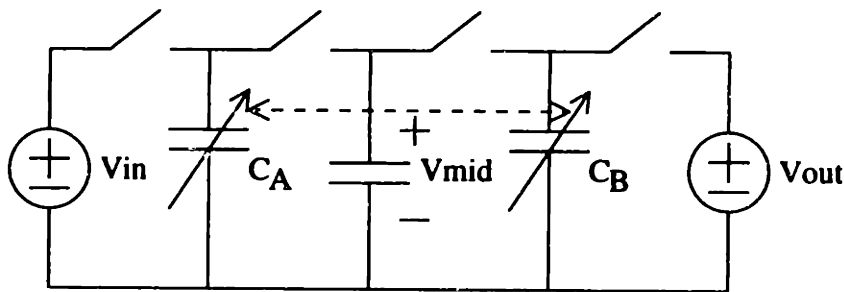Figure 5-1: Micromechanical structure for power converter

Figure 5-2: Electrical circuit for power converter

ground node of the circuit. Each electrode connects to two switches, the first connecting to either the input voltage or the output voltage, and the second connecting to an intermediate voltage level. The electrical circuit schematic is shown in Figure 5-2.

5.1.1 Theory of operation

At a frequency equal to the resonance of the mechanical diaphragm, energy is transferred from V_{in} to the mechanical system and also to the capacitor at V_{mid}. In a separate conversion cycle at the same frequency, energy is transferred from both V_{mid} and the mechanical system to the output. Two separate conversion cycles are required so that the total energy into the mechanical system is nearly zero. In steady state the magnitude of the mechanical vibration of the diaphragm should be constant, so the energy added to the mechanical system must be zero. An intermediate voltage is used so that the variable capacitors can always be charged and discharged at a constant voltage. As a result, the charging and discharging processes are ideally lossless. Note also that the intermediate voltage must be less than both the input voltage and the output voltage for it to be possible for the total energy into the mechanical system to be zero.

The operation of the power converter is best explained by examining the state of the variable
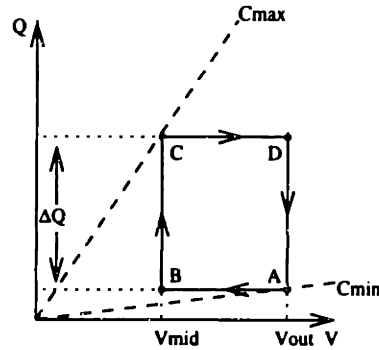
Figure 5-3: Energy conversion cycle on output capacitor

capacitors through a conversion cycle. In the charge–voltage plane, the varying capacitance is represented by a line with a slope which varies with time. Figure 5-3 shows the state of the output capacitor, C_B, through a single cycle of the converter. At the beginning of the cycle, C_B is charged up to V_{out}, the capacitor is open circuited, and the diaphragm is at the peak of its travel away from the electrodes (represented by Point A). As the diaphragm approaches the electrodes, the capacitance of C_B increases, and given the fixed charge on its plates, the voltage on the capacitor drops. When the voltage drops to V_{mid} (Point B), a switch closes, latching C_B to V_{mid}. As the capacitance continues to increase, charge now builds up on the capacitor plates and current flows from V_{mid} to C_B. When the diaphragm has reached the minimum point in its swing, the current onto C_B must drop to zero, and the switch latching the voltage to V_{mid} opens (Point C). Next, the diaphragm begins its upward swing, and the capacitor, which is again open circuited, sees an increasing voltage, due to its constant charge and decreasing capacitance. When the voltage on C_B reaches V_{out}, a switch closes, latching C_B to V_{out} (Point D). As the diaphragm continues its upward swing, charge comes off of the capacitor, and a current is delivered to the output, until the diaphragm reaches the peak of its swing and the cycle repeats itself.

By the end of the conversion cycle, an energy $\Delta Q V_{out}$ has been delivered to the output, $\Delta Q(V_{out} - V_{mid})$ has been extracted from the mechanical system, and $\Delta Q V_{mid}$ has been extracted from the intermediate voltage.

The energy conversion cycle on the input side of the converter is similar to that on the output (see Figure 5-4). The state of the input capacitor, C_A, also traverses a rectangular trajectory, but it operates to draw energy from V_{in}, and deliver energy to V_{mid} and the mechanical system. The locations of Point B' and Point D' can be the control inputs to the system. In steady state, the area $A'B'C'D'$ will be slightly larger than the area $ABCD$, so that the time average energy into the mechanical system exactly cancels the mechanical losses in the vibration. In transient operation, variations in the load are met by adding or removing energy from the mechanical system, varying the magnitude of the diaphragm vibration. Finally, note that the energy delivered to V_{mid} during
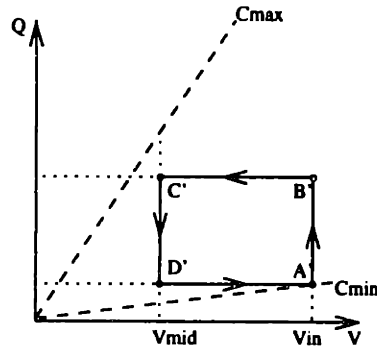
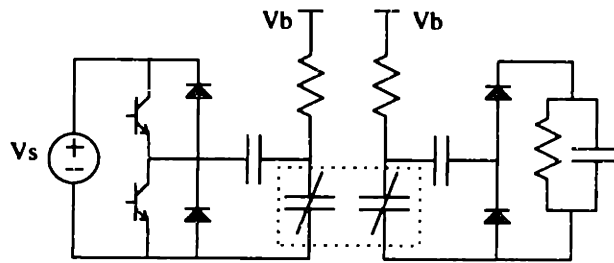Figure 5-4: Energy conversion cycle on input capacitor

Figure 5-5: Resonant micromachined converter proposed by Noworolski and Sanders

the input cycle is not equal to the energy removed from V_{mid} in the output conversion cycle. This difference in energy must be supplied to V_{mid} from either the input or the output.

5.1.2 Increasing power density with a DC bias voltage

The power transferred from the input to the output can be approximated by $f_{sw}C_{max}V_{out}^2$, under certain conditions $(C_{max} \gg C_{min})$. For reasonably low input or output voltages, low operating frequencies $(< 1MHz)$, and reasonable minimum plate separations $(\sim 0.1\mu m)$, the power density of the proposed topology is poor. The circuit proposed by Noworolski and Sanders utilizes a DC bias voltage to significantly increase the power density of this topology. The schematic of their circuit is shown in Figure 5-5. Although the circuit looks somewhat different than those of Figure 5-6, the effect of the DC bias is the same.

Consider transforming the circuit of Figure 5-2, to that of Figure 5-6. V_{out} and V_{in} have been decomposed into V_{out}', V_{in}', and V_{bias}. V_{mid} is now V_{bias}. The new state diagrams are shown in Figure 5-7. In the new system, the net energy transfered to or from V_{bias} over a cycle on the input and output sides is zero. Although there is significant reactive power through V_{bias}, no power is supplied by this source. The problem with the original circuit, that the net energy delivered to V_{mid} was non-zero, is no longer present. Furthermore, the power density is now $f_{sw}C_{max}V_{out}V_{bias}$,
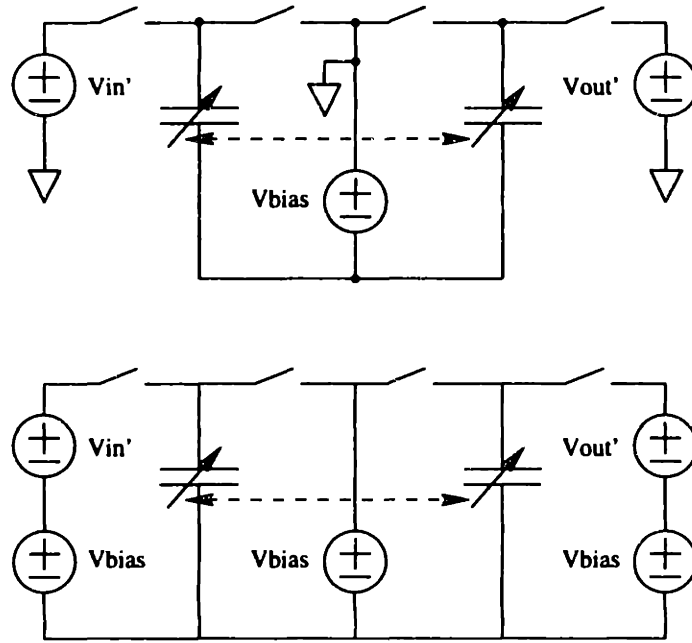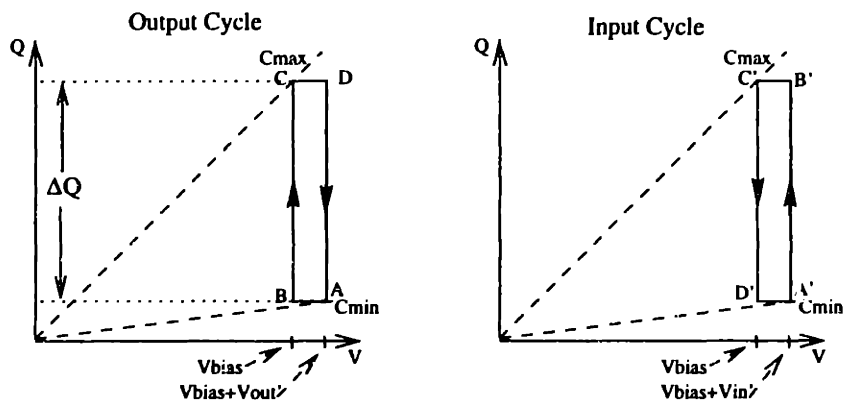
Figure 5-6: Topology transformed with bias voltage added

Figure 5-7: State diagrams including bias voltage

rather than $f_{sw}C_{max}V_{out}^2$. If V_{bias} is chosen to be large relative to V_{out}, the power density can be significantly increased.

5.2   A Sample Design

In order to demonstrate the feasibility of a micromechanical power converter, and to investigate the control of such a converter, a design is developed here which would be suitable for a low power load. The requirements for the design are $1mW$ output power, $2V$ output voltage, and a minimum input voltage of $2.5V$. The operating frequency was selected to be $100kHz$, an achievable mechanical resonant frequency that should not cause unmanageable switching losses or control bandwidth requirements. The minimum plate separation (between the electrode and the diaphragm) is assumed to be $0.1\mu m$, and that a reasonable diaphragm swing is $1.9\mu m$. Published results have boasted of swings with several tens of micrometers [36]. The creation of the bias is not addressed in detail. Some options for this are small auxiliary outputs on the diaphragm structure, or a standard capacitive voltage multiplier circuit. The power conversion is assumed to occur on a single diaphragm, rather than multiple parallelized structures. There may be complexities that arise from attempting to synchronize many parallel resonators, although this could be an attractive configuration for structural stability and operation over wide ranges of power outputs. The mechanics are modeled in a lumped parameter manner, where the diaphragm is a mass which is suspended by elastic beams. The beam is assumed to be deflected at its center by a concentrated loading force.

The output power of the converter is $f_{res}\Delta Q V_{out}$, where f_{res} is the mechanical resonant frequency and ΔQ is the difference in charge between the charged and uncharged states of C_A. Referring to Figure 5-7, we can see that

$$P_{out} = f_{res}V_{out}(V_{bias}C_{max} - (V_{bias} + V_{out})C_{min}),$$ (5.1)

which is approximately

$$P_{out} = f_{res}V_{out}V_{bias}(C_{max} - C_{min})$$ (5.2)

when $V_{out}C_{min} \ll V_{bias}C_{max}$ (this is Equation 1 in [17]). In terms of the dimensions of the capacitor, we can express

$$P_{out} = f_{res}V_{out}V_{bias}\epsilon_0 A_{C_B}\left(\frac{1}{x_{min}} - \frac{1}{x_{max}}\right)$$ (5.3)

where x_{min} and x_{max} are the minimum and maximum heights of the diaphragm above the electrodes, and A_{C_B} is the area of capacitor C_B. Given the assumptions outlined above, we find that the area of C_B is

$$A_{C_B} = \frac{P_{out}x_{max}x_{min}}{\epsilon_0 f_{res}V_{bias}V_{out}(x_{max} - x_{min})}.$$ (5.4)

The lower electrode will be coated with nitride of thickness x_{min} to prevent the two plates from touching. Typical values for breakdown fields in nitride are $5 - 10 MV/cm$ [17]. This breakdown limits the magnitude of V_{bias}. To be somewhat conservative, choose $V_{bias} = \mathcal{E}_{br} x_{min}$, with $\mathcal{E}_{br} = 2.5 MV/cm$. Assuming that $x_{max} \gg x_{min}$, we can find the area of C_B:

$$A_{C_B} = \frac{P_{out}}{\epsilon_0 f_{res} \mathcal{E}_{br} V_{out}},\tag{5.5}$$

or, for the values stated above,

$$A_{C_B} = 2.25 mm^2.\tag{5.6}$$

At steady state, we have established that the net energy into the mechanical system is nearly zero. This energy is equal to the difference of the areas of the two rectangles in Figure 5-7. This equality yields the relationship,

$$\epsilon_0 A_{C_B} V_{out} V_{bias} (\frac{1}{x_{min}} - \frac{1}{x_{max}}) = \epsilon_0 A_{C_A} V_{in} V_{bias} (\frac{1}{x_{min}} - \frac{1}{x_{max}}),\tag{5.7}$$

using the approximations $C_{min} V_{out}, C_{min} V_{in} \ll C_{max} V_{bias}$. We can then easily derive an expression relating A_{C_B} and A_{C_A},

$$\frac{A_{C_A}}{A_{C_B}} = \frac{V_{out}}{V_{in}} < \frac{2.0 V}{2.5 V}.\tag{5.8}$$

This gives the area A_{C_A} equal to $1.81 mm^2$, and the total area of the diaphragm is $4.1 mm^2$. (The resulting power density is $24 mW/cm^2$. If this is scaled by 4 to account for the effect of an input voltage of $10V$, as used in [17], and scaled by a factor of 10 to account for a higher frequency of operation, we get $1 W/cm^2$, which is on the order of the number cited in [17], $2 W/cm^2$.)

There are two important forces on the diaphragm; there is the elastic restoring force and the force of electric origin which pulls the diaphragm toward the substrate. We must ensure that the electric force will never be greater than the elastic force when the plate is at its maximum deflection (closest to the substrate). If the electric force were to exceed the elastic force, the diaphragm would stick to the nitride on the bottom electrode, and any oscillation (of the diaphragm) would cease. The condition on the design is

$$k(x - h) > \frac{1}{2} \frac{\partial C}{\partial x} V^2\tag{5.9}$$

where x is the distance from the bottom electrode to the diaphragm, and h is the rest height of the diaphragm. This expression yields the equation

$$k > \frac{2\mathcal{E}_{br}^2 x_{min}^2 A_{total} \epsilon_0}{x_{min}(x_{max} - x_{min})^2} = 125 \times 10^3 kg/s^2\tag{5.10}$$

where the maximum voltage across the capacitors has been approximated by $V_{bias} = \mathcal{E}_{br} x_{min} = 25V$, and the maximum swing of the diaphragm is assumed to be $2\mu m$.
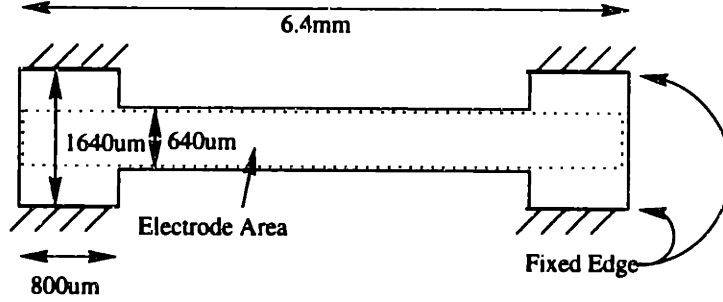
Figure 5-8: Proposed structure for 1mW down converter

The diaphragm and support beam must be designed to achieve the desired resonant frequency. The validity of the lumped parameter model has also been ensured by setting the elasticity of the diaphragm to be significantly greater than the elasticity of the support beams. As a result, the diaphragm should move parallel to the electrodes without deforming. The spring constant of these structures (for the deflection at their center as a result of a concentrated load at their center) is given by

$$k = 16\frac{Ebh^3}{l^3} \tag{5.11}$$

where E is the modulus of elasticity, b is the width, h is the height (in the direction motion), and l is the length of the beam. I used a value of $E = 170GPa$ [11].

A diaphragm for the design at issue could be constructed as shown in Figure 5-8. The structure bends along the shorter side, of length $640\mu m$. The width of the structure is $6.4mm$ to give the desired area for the variable capacitors. The thickness of the diaphragm is $34\mu m$. The spring constant of this diaphragm is then $2.6 \times 10^6 kg/s^2$, and the mass is 320×10^{-9} (utilizing the mass density of polysilicon, $2300kg/m^3$ [11]). This mass gives the desired resonant frequency for a spring constant of $150 \times 10^3 kg/s^2$. This spring constant is created by a beam of length $1.0mm$, and total width $1.6mm$.

5.3   System Modeling

The complete dynamic models for the electromechanical power converter are highly nonlinear and difficult to derive. Both the input and the output conversion cycle have four distinct regions of operation (which do not necessarily correlate with each other in time), the forces on the mechanical system from the electrical inputs are related by the square of the applied voltages, and the capacitances seen from the electrical system are proportional to the inverse of the deflection of the plates. The analysis of state space averaged equations is much more tractable. The averaged system

equations can be derived from the state diagrams shown in Figure 5-7.

The primary state of interest in the converter is the output voltage. This makes V_{out} an obvious choice for a state variable. Its state equation is given by

$$C_{out}\frac{dV_{out}}{dt} = -\frac{V_{out}}{R_{load}} + \frac{V_{bias}}{T}(C_{max} - C_{min}) \qquad (5.12)$$

where T is the period of the mechanical resonance, and the load is assumed to be a resistor in parallel with a filtering capacitor. The second term in the right side of Equation 5.12 is the total charge delivered to the output in a single cycle (divided by the period to give an average current). This term is ΔQ in Figure 5-7.

C_{max} and C_{min} in Equation 5.12 are functions of the magnitude of the deflection in the mechanical system. Clearly, we must model the state of the mechanical system. Given that the amount of energy that is added to or removed from the mechanical system each cycle is available, a natural state to measure is the total energy stored in the mechanical system. The state equation for E_{mech} is

$$T\frac{dE_{mech}}{dt} = V_{in}\Delta Q_{in} - V_{out}V_{bias}(C_{max} - C_{min}). \qquad (5.13)$$

ΔQ_{in} represents the difference in charge between the charged and uncharged states of the input capacitor, C_A. Mechanical damping has been neglected, as it is typically very small for systems such as this. Equation 5.13 is given by the difference of the areas defined by the two trajectories of the input conversion cycle and the output conversion cycle. Given E_{mech}, the magnitude of the diaphragm deflection can be found by $E_{mech} = \frac{1}{2}kx^2$, where k is the spring constant of the mechanical system, and x is the maximum diaphragm deflection.

Here, it is assumed that the system will be controlled by modifying the on-times of the input switches. In this case, the converter is operated in a fixed frequency mode. Pulse width modulation is generally a more simple control technique (compared to resonant frequency control), which may mean the control circuits will consume less power. The only inputs to the electromechanical system as described are the positions of Point B' and Point D' in Figure 5-4. All other switch transitions are governed solely by the voltages on C_A and C_B, or the currents through them. A convenient way to specify the positions of these control points is with ΔQ_{in}. The total charge through the switches could be measured by integrating current; this value could be compared to the desired ΔQ_{in} to time the switch transitions (that is, the locations of B' and D'). This is more convenient than specifying a time delay for the circuit to travel from A' to B' and from C' to D'. These two delays are not the same, and this method would require two separate control inputs.

In summary, we can write

$$\frac{dV_{out}}{dt} = -\frac{V_{out}}{R_{load}C_{out}} + \frac{V_{bias}}{TC_{out}}f(E_{mech}) \qquad (5.14)$$

and

$$\frac{dE_{mech}}{dt} = \frac{V_{in}}{T}q - \frac{f(E_{mech})V_{out}V_{bias}}{T}. \tag{5.15}$$

$f(E_{mech})$ is ΔC_B, or $C_{max} - C_{min}$, and q is the control input, ΔQ_{in}.

At this point, it is useful to linearize the state equations around the operating point. The linearized equations are:

$$\frac{d}{dt}\begin{pmatrix} \tilde{v} \\ \tilde{E} \end{pmatrix} = \begin{pmatrix} \frac{-1}{C_{out}R_{load}} & \frac{V_{bias}}{TC_{out}}\frac{df(E)}{dE} \\ \frac{-V_{bias}f(E)}{T} & \frac{-V_{out}V_{bias}}{T}\frac{df(E)}{dE} \end{pmatrix}\begin{pmatrix} \tilde{v} \\ \tilde{E} \end{pmatrix} + \begin{pmatrix} 0 \\ \frac{V_{in}}{T} \end{pmatrix}q \tag{5.16}$$

The operating point is taken to be full output power, or $R_{load} = 4k\Omega$. C_{out} was chosen to be $62.5nF$ in order to give approximately 1% output ripple. T is $10\mu S$, V_{bias} is $25V$, and V_{out} is $2V$. For this set of conditions, the eigenvalues of the linearized system are $-8.5 \times 10^3 s^{-1}$ and $-7.3 \times 10^4 s^{-1}$. These eigenvalues confirm the validity of the state space averaging, since they are both a good deal slower than the excitation frequency (and the inverse of the averaging interval), $6.3 \times 10^5 rad/s$. There was an assumption that $E_{mech} = \frac{1}{2}kx^2$, where x is the maximum deflection of the plates, which was used to find $f(E)$. This assumption rests on an expectation that E_{mech} does not vary significantly within a single conversion cycle, which is fortunately confirmed by these relatively slow eigenvalues.

Participation factor analysis can be used to gain further insight into this linearized model[20]. For the given operating condition, the participation matrix is

$$p_{ij} = \begin{pmatrix} 1.070 & -0.070 \\ -0.070 & 1.070 \end{pmatrix} \tag{5.17}$$

which indicates that the faster of the two eigenvalues is associated primarily with the second state (E_{mech}), and the slower of the two eigenvalues is associated primarily with the first state (V_{out}).

5.4 Control

The separation of time scales between the two modes (a factor of 8.6) allows us to decouple the two states. We are interested in controlling the state of V_{out}, which is the slower of the two states. Therefore, we can assume the dynamics of E_{mech} have finished when we consider the dynamics of V_{out}. That is, we set $\frac{dE_{mech}}{dt} = 0$ to find an equilibrium value of E_{mech}. This process yields

$$\frac{f(E_{mech})V_{out}V_{bias}}{T} = \frac{V_{in}}{T}q, \tag{5.18}$$

or,

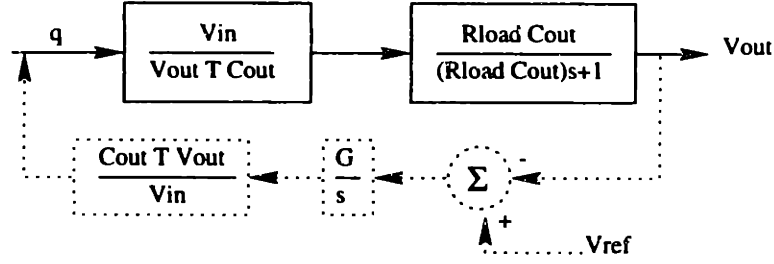$$f(E_{mech}) = \frac{V_{in}}{V_{out}V_{bias}}q. \tag{5.19}$$

Figure 5-9: Block diagram of system after time scale separation, possible feedback linearization control scheme

Inserting this result into the dynamical equation for V_{out}, we find

$$\frac{dV_{out}}{dt} = -\frac{V_{out}}{R_{load}C_{out}} + \frac{1}{TC_{out}}\frac{V_{in}}{V_{out}}q. \tag{5.20}$$

Equation 5.20 is a first order non-linear system equation. We now have a few options available for the control law. We could use a feedback linearization scheme, as shown in Figure 5-9. The primary advantage with this control law is that the resulting system is linear, and the feedback gain can be selected to set the speed of the first order response. (Of course, there is a limitation that the time-scale separation will be violated if the gain too large.) A disadvantage of this control scheme is that the circuitry to implement the non-linear operation of multiplying by $\frac{C_{out}TV_{out}}{V_{in}}$ may be difficult to design, particularly in a situation where the control circuitry should consume very little power (on the order of $10\mu W$ for the sample design presented). Another problem with feedback linearizing control is the uncertainty parameters such as V_{in}, C_{out} and T. The control must be conservative enough to maintain good performance for a reasonable range of these system parameters.

A second way to formulate a controller for the system described in Equation 5.20 is to linearize the equation about its operating point. The linearized version of this equation is:

$$\frac{d\tilde{v}_{out}}{dt} = -\frac{\tilde{v}_{out}}{R_{load}C_{out}} + \frac{V_{in}}{TC_{out}}\frac{1}{V_{OUT}}\tilde{q} - \frac{V_{in}}{TC_{out}}\frac{Q}{V_{OUT}^2}\tilde{v}_{out} \tag{5.21}$$

with the equilibrium condition that

$$0 = \frac{-1}{R_{load}C_{out}}V_{OUT} + \frac{V_{in}Q}{TC_{out}V_{OUT}}. \tag{5.22}$$

The combination of these two results gives the equation

$$\frac{d\tilde{v}_{out}}{dt} = -\frac{2}{R_{load}C_{out}}\tilde{v}_{out} + \frac{V_{in}}{TC_{out}}\frac{1}{V_{OUT}}\tilde{q}. \tag{5.23}$$

This first order linear equation could then be used to select the feedback applied to $\tilde{q}$. Again, there is

some uncertainty in the gain of the control, $\bar{q}$, since it is a product of parameters which will vary with operating condition. Also, the gain applied to the feedback must still be low enough to ensure that time-scale separation still valid. The implementation of the circuitry for this control law is simpler than that for the feedback linearizing controller, and should be more amenable to the creation of low power control circuitry. A disadvantage of the linearized controller is that the model is applicable only with some limited region about the operating point. The feedback linearizing controller had no such restriction.

A third possibility for control is to attempt to develop a control law for the linearized second order system of Equation 5.16. The primary disadvantage with this method is that E_{mech} is not available as an input to the controller. As a result, it is unclear that there is anything to be gained by considering this model.

5.5 Summary

This chapter presented the design and theory of operation of a $1mW$ micromechanical DC-DC converter. Actual fabrication of the device presented should be pursued in order to uncover problems not anticipated and to demonstrate the operation of such a device. Some issues that remain unanswered about implementation are whether there will be difficulties parallelizing many small micromechanical resonators, and what the practical limits on the frequency of such resonators are. Parallelization may be necessary to create structurally sound and robust diaphragms, and frequency scaling is important to increase power density (in terms of both the diaphragm area and the size of the capacitors required by the circuit.

This chapter also presented the modeling and control for the proposed micromechanical converter. Time scale separation was used in order to reach a first order model which could be linearized. An important thing to investigate in the future is the validity of the time scale separation. Although the separation was justified at a single operating point, for a particular design, this will not hold under other operating conditions; the linearization must be examined over the range of the input voltage and the output load. There may also be issues with the start up of the converter, when V_{out} and V_{bias} are not near their equilibrium values. Care must be taken not only to examine the relative values of the eigenvalues, but also their relationship to the excitation frequency, to check the validity of the state space averaging operation.

When the power consumption of control circuitry is important, simplifying the control law can be more critical than linearizing the behavior of the system. For this reason, it is likely that an actual implementation of a micromechanical converter for a low power system would implement a form of linear control law, as opposed to a linearizing controller.

Chapter 6

Conclusion

The most important lesson to learn about making a power supply more energy efficient is that efficiency always increases with decreasing operating frequency. The lower limits on switching frequency are set by the volume requirements and sometimes also the desired transient response. When attempting to decrease the energy dissipation of a power supply, one needs to determine the requirements on the output voltage in order to create specifications for the output filter, then optimize the filter to accommodate as low a switching frequency as possible. Other techniques to reduce power consumption, such as novel gate drive circuits and zero voltage switching should be considered as tools to be applied in concert with filter optimization and frequency reduction; these techniques are not mutually exclusive.

The goal of setting the operating frequency as low as possible for a given supply volume is a minimization of power dissipation at fixed volume. The result of this minimization is an expression of the efficiency of a certain type of power supply for different volumes. This expression, $\eta(V)$, enables the comparison of different forms of power conversion, such as inductor based converters, switched capacitance converters, linear regulators, and even micro-mechanical power converters.

As output powers drop, there is pressure to develop power supply controllers which consume very little power. Digital controllers prevent an opportunity here, offering the capability to implement more intelligent controllers which interact with their loads while consuming very little power. The experimental system demonstrated a control circuit which consumed as little as $15\mu W$, and this could easily be reduced further, as parts of the design were somewhat conservative in order to ensure the circuit was functional on this first attempt.

It is possible to reduce the complexity of the feedback circuitry by implementing single bit feedback, avoiding the power consumption of more complex A/D conversion. However, such a controller cannot be modeled as a linear system with quantization noise; methods to analyze non-linear systems need to be applied. Single bit feedback systems will always be limited in their response

times compared to what can be achieved with controllers which can measure the size of an error, not just its magnitude.

Future Work

There are many questions to be answered and problems to solve in the quest to improve the run time of low power systems. Implementing entire power supplies on a single chip is something which industry has been doing in the past few years; this should be advanced to developing power supply blocks which can be integrated as part of a larger system on a single chip. On the more theoretical side, the power densities of the various types of converters needs to be compared explicitly, perhaps showing in what regimes of operation it makes sense to use a linear regulator or switched capacitance converter instead of a more conventional switching regulator.

Integrated Systems and Control Circuits

There were problems with the performance of the fabricated system which need to be addressed. As mentioned in Chapter 4, the delay growth problem needs to be addressed in future delay line based PWMs, either by more careful control of the positive and negative edge propagation times or by actively resetting the delay line to a known state at the end of a cycle (perhaps with dynamic logic). Also, the effect of delay jitter in the delay line needs to be ascertained, and specifications developed for the delay feedback circuit.

Improvements on the controller presented include implementing all parts of the controller on chip, to create a fully integrated power supply. Also, controllers which provide multiple outputs with a single delay line can be developed to suit systems which require multiple outputs. Pin count needs to be reduced so that the size of the power supply IC package is more in line with the magnitude of its power output. A serial data input could allow the creation of an 8 pin, programmable, multiple output power supply. Further packaging improvements include the development of methods to integrate the output filter within the same plastic package as the die, as is done by a National Semiconductor power supply.

Another goal for future power supply circuits is to create a controller core with output switches which can be easily dropped into larger systems on a single chip. For this application, reducing the controller area from that in the demonstration chip discussed here could be important. This could be accomplished with a "hybrid" PWM, utilizing a smaller delay line and multiplexer along with a counter to combine the low power qualities of the delay line and the compact size of the fast-clocked counter approach.

It has not been demonstrated that digital circuits can implement PWM blocks more efficiently than analog circuits optimized for low power. Analog control circuits need to be analyzed carefully

to reduce their power consumption; in particular, areas where the performance of typical analog controllers can be traded off for lower power operation need to be identified. It is possible that a well designed analog controller could provide a lower power, lower area solution for low power regulators. Another possibility is to investigate sigma delta converts, which enable single bit A/D converters to be used with a digital controller, while not sacrificing the ability to measure signals proportional to the error.

Finally, micromechanical power supplies present an opportunity to create an efficient voltage converter without requiring off-chip inductors. Further research is called for in this area, first to investigate the feasibility of such supplies, and then to develop processes which will allow a micromechanical converter to be constructed or placed on top of a die which contains the load system.

Modeling and Optimization

The development of expressions for the power dissipation of converters given their volume provides a metric to compare the different types of voltage converters. This could lead to a conclusion that, for example, linear regulators provide the best run time extension for extremely small volumes, switched capacitor circuits are appropriate for somewhat larger volumes, and inductor-based switching converters are the most efficient choice for still larger volumes. Such an evaluation should be carried out, and it is not certain that the above statement will in fact prove true. This investigation would provide guidelines which could direct the application of power supplies for low power.

The present work has demonstrated control circuits which consume considerably less than $100\mu W$, an appropriate level for power supplies with output levels in the milliwatt range which are required to perform efficiently at powers even less than this level. More focus must now be placed on reducing the dissipation in the output stage for very low output power converters. Techniques which have been developed for this purpose at higher output powers include timing the gate drives of the output switches to accomplish zero voltage switching, and pulse squashing (dynamically reducing the switching frequency). Such techniques need to be developed with an eye toward limiting the power dissipation of the extra control circuitry to the $1 - 10\mu W$ level. Also, a circuit to provide true synchronous rectification is needed, which will prevent the inductor current from becoming negative at low output currents. This will reduce the circulating current in the filter for small output powers, and move the converter into discontinuous mode.

Single bit output feedback may be an attractive alternative to more complex A/D converters or fully analog feedback circuits. Non-linear control techniques, such as describing functions, need to be applied to the problem of single bit feedback power supply control. This will enable rigorous analysis of the non-linear system's stability, so that the gain can be set with some assurance of stability, rather than simply choosing a gain empirically that appears to result in stable behavior

under some conditions.

Bibliography

[1] W. B. Baringer and R. W. Broderson. MCMs for portable applications. In *IEEE Multi-Chip Module Conference*, 1993.

[2] Anantha P. Chandrakasan and Robert W. Brodersen. *Low Power Digital CMOS Design*. Kluwer Academic Publishers, Norwell, Massachusetts, 1995.

[3] S. Chou and C. Simonsen. Chip voltage: Why less is better. *IEEE Spectrum*, 24(4):39–43, April 1987.

[4] Alex Goldman. *Modern Ferrite Technology*. Van Nostrand Reinhold, New York, 1990.

[5] Vadim Gutnik. Variable supply voltage for low power DSP. Master's thesis, Massachusetts Institute of Technology, May 1996.

[6] Vadim Gutnik and Anantha Chandrakasan. An efficient controller for variable supply–voltage low power processing. In *Symposium on VLSI Circuits*, 1996.

[7] Mark Horowitz. Low power processor design using self-clocking. Workshop on Low-power Electronics, 1993.

[8] Clas A. Jacobson, Aleksandar M. Stankovic, and Gilead Tadmor. Design of robust controllers for resonant dc/dc converters. In *4th IEEE Conference on Control Applications*, 1995.

[9] John G. Kassakian, Martin F. Schlecht, and George C. Verghese. *Principles of Power Electronics*. Addison-Wesley Publishing Company, Inc., 1991.

[10] E. E. Landsman. A unifying derivation of switching DC-DC converter topologies. In *IEEE Power Electronics Specialists Conference*, pages 239–243, 1979.

[11] C. Linder and N. F. De Rooij. Investigations on free-standing polysilicon beams in view of their application as transducers. In *Sensors and Actuators*, volume A21-23, pages 1053–1059, 1990.

[12] D. Maksimovic. A MOS gate drive with resonant transitions. In *IEEE Power Electronics Specialists Conference*, pages 96–105, 1990.

[13] Thomas L. Martin and Daniel P. Siewiorek. A power metric for mobile systems. In *1996 International Symposium on Low Power Electronics and Design*, pages 37–42, 1996.

[14] Col. W.T. McLyman. *Magnetic Core Selection for Transformers and Inductors*. Marcel Dekker, New York, 1982.

[15] C. Mead and L. Conway. *Introduction to VLSI Systems*. Addison-Wesley, Reading, Mass, 1980.

[16] Lars S. Nielsen, Cees Niessen, Jens Sparsø, and Kees van Berkel. Low-power operation using self-timed circuits and adaptive scaling of the supply voltage. *IEEE Transactions on VLSI Systems*, 2(4):391–397, December 1994.

[17] J. Noworolski and S. Sanders. An electrostatic microresonant power conversion device. In *IEEE Power Electronics Specialists Conference*, pages 997–1002, 1992.

[18] Katsuhiko Ogata. *Modern Control Engineering*. Prentice Hall, Englewood Cliffs, NJ, second edition, 1990.

[19] Alan V. Oppenheim and Alan S. Willsky. *Signals and Systems*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1983.

[20] I. J. Perez-Arriaga, G. C. Verghese, F. L. Pagola, J. L. Sancha, and F. C. Schweppe. Developments in selective modal analysis of small-signal stability in electric power systems. *Automatica*, 26(2):215–231, 1990.

[21] John G. Ryan, Kenneth J. Carroll, and Benjamin D. Pless. A four chip implantable defibrillator/pacemaker chipset. In *Proceedings of the IEEE Custom Integrated Circuits Conference*, pages 7.6.1–7.6.4, 1989.

[22] Rudy Severns. High frequency core losses for non-sinusoidal waveforms. In *IEEE High Frequency Power Conversion Proceedings*, page 140, June 1991.

[23] S. Singer. Inductance-less up DC-DC converter. *IEEE Journal of Solid-State Circuits*, SC-17(4):778–781, August 1982.

[24] A. Stratakos, S. Sanders, and R. Broderson. A low-voltage CMOS DC-DC converter for a portable battery-operated system. In *IEEE Power Electronics Specialists Conference*, pages 619–626, 1994.

[25] Charles R. Sullivan and Seth R. Sanders. Microfabrication process for high-frequency power-conversion transformers. In *IEEE Power Electronics Specialists Conference*, volume 2, pages 658–664, 1995.

[26] Charles R. Sullivan and Seth R. Sanders. Design of microfabricated transformers and inductors for high-frequency power conversion. *IEEE Transactions on Power Electronics*, 11(2):228–238, March 1996.

[27] Lars J. Svensson and Jeff G. Koller. Driving a capacitive load without dissipating fCV^2. In *IEEE Symposium on Low Power Design*, pages 100–101, 1994.

[28] P. Theron, P. Swanepoel, J. Schoeman, J. Ferreira, and J. van Wyk. Soft switching self-oscillating FET-Based DC-DC converters. In *IEEE Power Electronics Specialists Conference*, volume 1, pages 641–648, 1992.

[29] Harry J. M. Veendrick. Short-circuit dissipation of static CMOS circuitry and its impact on the design of buffer circuits. *IEEE Journal of Solid-State Circuits*, SC-19(4):468–473, August 1984.

[30] Vincent Von Kaenel, Peter Macken, and Marc G. R. Degrauwe. A voltage reduction technique for battery-operated digital systems. *IEEE Journal of Solid-State Circuits*, 25(5):1136–1140, October 1990.

[31] Gu-Yeon Wei and Mark Horowitz. A low power switching power supply for self-clocked systems. In *1996 International Symposium on Low Power Electronics and Design*, pages 313–318, 1996.

[32] S. Weinberg. A novel lossless resonant MOSFET driver. In *IEEE Power Electronics Specialists Conference*, volume 2, pages 1002–1010, 1992.

[33] Neil H. E. Weste and Kamran Eshraghian. *Principles of CMOS VLSI Design*. Addison-Wesley Publishing Company, Reading, Massachusetts, second edition, 1993.

[34] Richard K. Williams, Robert Blattner, and Bijan E. Mohandes. Optimization of complementary power DMOSFETs for low-voltage high-frequency DC-DC conversion. In *IEEE APEC Conference Proceedings*, volume 2, pages 765–772, March 1995.

[35] S. G. Younis and T. Knight. Practical implementation of charge recovering asymptotically zero power CMOS. In *Symposium on Integrated Systems*, pages 261–264, 1993.

[36] Yafan Zhang and Kensall D. Wise. Performance of non-planar silicon diaphragms under large deflections. *Journal of Microelectromechanical Systems*, 3(2):59–68, June 1994.