An Energy Optimal Power Supply for Digital Circuits

by

Yogesh Kumar Ramadass

B.Tech. Electronics and Electrical Communication Engineering
Indian Institute of Technology, Kharagpur, 2004

Submitted to the Department of Electrical Engineering and Computer Science in partial fulfillment of the requirements for the degree of

Master of Science in Electrical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2006

Author.....
Department of Electrical Engineering and Computer Science
May 24, 2006

Certified by.....
Anantha Chandrakasan
Professor of Electrical Engineering
Thesis Supervisor

Accepted by.
Arthur C. Smith
Chairman, Department Committee on Graduate Students

An Energy Optimal Power Supply for Digital Circuits

by

Yogesh Kumar Ramadass

Submitted to the Department of Electrical Engineering and Computer Science on May 24, 2006, in partial fulfillment of the requirements for the degree of Master of Science in Electrical Engineering

Abstract

The energy efficiency of digital circuits continues to be a major factor in determining the size and weight of battery-operated electronics. Integration of more functionality in a single system has made battery longevity a major problem. Operating circuits at their minimum energy operating voltage (MEP) has been proposed as a solution for energy critical applications where performance is not a key constraint. This thesis explores the sensitivity of the MEP to operating conditions and motivates the need for continuous minimum energy tracking based on the energy savings possible. A circuit that can dynamically track the MEP of a digital circuit with varying load conditions and temperature is presented. A low power, voltage scalable DC-DC converter is also embedded within the chip.

The proposed minimum energy tracking algorithm uses a novel approach to sense the energy consumed per operation. The energy sensing circuitry does not use high-resolution Analog-to-Digital converters or high gain amplifiers. The energy estimate is used in a slope tracking algorithm to track the minimum energy operating voltage.

The minimum energy tracking loop along with a low-voltage DC-DC converter and test circuitry were fabricated in a 65nm CMOS process. The circuits are powered from an external 1.2V supply. The digital test circuitry was capable of operation at voltages as low as 0.25V. The tracking of the minimum energy operating voltage with change in workload and temperature was observed. The DC-DC converter was able to deliver load voltages between 0.25V and 0.7V with an efficiency > 78% at load power levels of the order of 10μW and above.

Thesis Supervisor: Anantha P. Chandrakasan
Title: Professor of Electrical Engineering

Acknowledgements

Firstly, I would like to thank my parents and my sister for their love, support and constant encouragement.

My special thanks to Prof. Anantha Chandrakasan, for truly being a great advisor.

Thanks to members of Ananthagroup for being a wonderful group of individuals, and for creating a work atmosphere that is the best I have come across. I would like to specially thank Brian Ginsburg for all the help with CAD tools and for having solved innumerable layout problems. I would also like to thank Dr. Benton Calhoun for having encouraged me to work in this problem.

I am indebted to Prof. Nirmal B. Chakrabarti of IIT-Kharagpur for always being my source of inspiration and for having got me to MIT.

Finally, I would like to thank Texas Instruments for providing fabrication facilities and valuable technical support.

Contents

List of Figures

Chapter 1

Introduction

The energy efficiency of digital circuits continues to be a major factor in determining the size and weight of battery-operated electronics. Sensor networks deployed with finite energy resources, for example, will greatly benefit from energy-conscious design techniques. A key requirement in the design of sensor systems is constraining the power dissipation of the system below 100μW [1], [2] which will allow operation strictly using scavenged energy. To achieve this goal, the power dissipation must be drastically reduced from current approaches. Sub-threshold operation (i.e., operating the circuits with a supply voltage below the threshold voltage of the transistors) is a technique that will allow energy dissipation to be reduced by an order of magnitude over current approaches. However, it comes at the cost of circuit speed. In many applications such as sensors, the performance requirements are quite relaxed and sub-threshold operation is a feasible approach.

However, there are several challenges in sub-threshold operation. First, the circuits are very sensitive to process variations as the delay is exponentially dependent on the operating voltage. Second, the optimum energy point (the operating voltage at which the

energy consumed per operation is minimum) is sensitive to operating conditions such as temperature, load, and data dependencies. Third, robust operation of memory circuits is particularly challenging across process corners. Delivering subthreshold voltages to the digital circuit from a standard 1.2V Ni-MH battery poses new challenges in itself. Getting high efficiencies when operating in the low power regimes of subthreshold voltages needs careful circuit design to identify and minimize leakage paths and the power consumed by the control circuitry in the DC-DC converter.

1.1 Minimum Energy Operating Voltage

The variation of the energy consumed per operation (E/op.) with the operating voltage for an FIR filter is shown in Figure 1-1. The energy consumed per operation decreases as the voltage is reduced from the supply voltage of 1.2V. It continues decreasing till it hits a minimum beyond which it increases as the voltage is decreased. This voltage at which the energy consumed is minimal to complete a particular operation is called the minimum energy operating voltage [3]. Substantial savings in the energy consumed by a digital circuit can be obtained by operating the circuit at the minimum energy operating voltage.

The E/op. curve is dynamic in nature and changes with temperature, workload of the circuit, nature of operations performed by the circuit and data handled. The optimum energy point shifts widely as the curve changes and this necessitates a circuit to track the optimum energy point with changing conditions, to truly minimize energy consumption of the digital circuit over varying operating conditions. The optimal supply voltage for minimum energy operation usually falls into the subthreshold region of operation of digital circuits, i.e. below the threshold voltages of the transistors used in the circuit. At

these voltages, the circuits operate substantially slower. This is not a handicap in energy driven class of circuits like the sensor nodes where circuit speed is not a key issue. The primary constraint in these types of circuits is the energy consumed to perform a few digital operations. If we can devise a way to track the minimum energy operating voltage over varying operating conditions, we can operate the sensor node of these voltages and minimize its total energy consumed. Even performance driven circuits operate in the high-performance mode only for a small fraction of the time and idle the rest of the time running some basic operations which are not very demanding in terms of speed. We can use the minimum energy feedback loop when the circuit is operating at a low frequency to gain savings in total energy consumed.

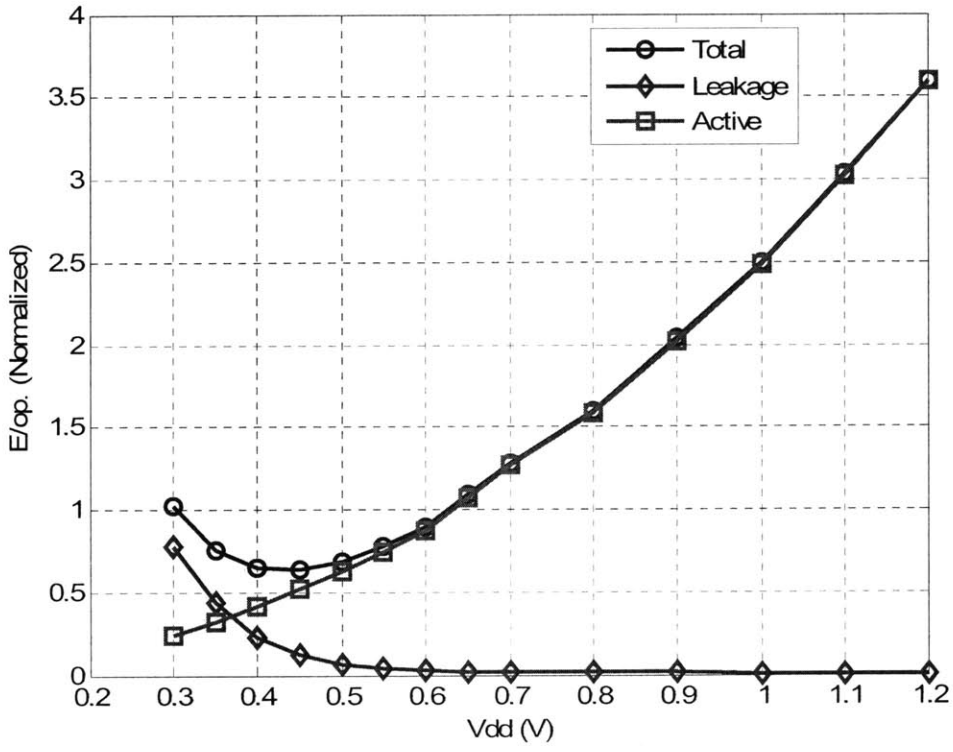

Figure 1-1: Active, leakage and total energy per operation curves showing the minimum energy point (0.4V) for an FIR filter implemented in 65nm CMOS

The minimum energy tracking circuitry should be non-intrusive in the sense that the digital circuit should continue normal operation while the energy sensing and tracking operations is going on. Further, the energy sensing and tracking circuitry should be simple and consume very little energy overhead in order not to eat away into the energy savings obtained from operating in the subthreshold regime.

1.2 Thesis Organization

Chapter 2 deals with the different operating scenarios encountered by a digital system and explores how digital operations can be completed with the minimum possible energy for the different scenarios. An analytical expression for the minimum energy operating voltage is derived. The energy savings that can be obtained by continuous tracking of the minimum energy point is discussed and the need for a minimum energy tracking loop is put forward. The chapter concludes with some measurement results of the energy savings that can be obtained from an FIR filter test circuit.

Chapter 3 describes the design and implementation of a low power subthreshold DC-DC converter. The various converter topologies are explored and the design issues related to low power subthreshold DC-DC converter design are dealt with. Circuit techniques and tradeoffs involved to achieve > 80% efficiencies at the low power operating points are discussed. A synchronous rectifier buck converter is implemented in a 65nm CMOS process and measured results of the DC-DC converter are provided.

Chapter 4 deals with sensing the energy consumed per operation at a particular operating voltage and having done that, tracking the minimum energy operating voltage. It dissects the individual circuits needed for energy sensing and tracking and explores the

tradeoffs involved in designing each one of them. An analysis of errors introduced in E/op. sensing is provided and an estimate of the minimum difference in energy levels that can be identified is given.

The goal of this thesis is to analyze and demonstrate the energy savings that can be obtained by continuously tracking the minimum energy operating voltage. In particular, a new low power way to sense the energy per operation at a particular operating voltage is presented along with a low voltage DC-DC converter to provide that voltage at the required power to the digital circuit.

Chapter 2

Minimum Energy Analysis – System Scenarios and Energy Contours

A digital system can operate in different modes depending on its expected performance output. The broad range of operation of a digital circuit can be classified into three separate modes: idle, performance-critical, and performance-non-critical. The classification is based on the performance output required of the circuit at any instant of time. Minimizing energy for a system scenario that consists of a combination of these modes simply requires applying the proper technique for each mode. When in idle mode, circuits perform no switching operations. In contrast, active mode circuits perform switching operations, and they can be further divided into two different modes based on the time it takes to complete an operation at the minimum energy operating voltage. The terms minimum energy operating voltage and minimum energy point (MEP) have been used interchangeably in this thesis and they both refer to the voltage at which the energy consumed to complete an operation is the minimum. Let the time required to complete an

operation at the MEP be T_{minE}. When the allowable delay for completing an operation which is set by the user or a system at a higher level is less than T_{minE}, the system is operating in its performance-critical mode. Performance-non-critical mode occurs when the allowable delay exceeds T_{minE}.

Table 2.1: Operating scenarios of a digital system

Operating scenarios		Description
Idle		Performs no switching operations
Active	Performance Critical	Performs switching operations with allowable delay per operation < T_{minE}
	Performance-non-Critical	Performs switching operations with allowable delay per operation > T_{minE}

2.1 Operating Scenarios

2.1.1 Idle Mode

An idle circuit block doesn't perform any active switching operations and keeps consuming leakage energy. If the amount of time that the block remains idle depends on some external factor (such as incoming data rate), then the time for which the idle circuit block consumes leakage power is fixed. Here, the standby leakage energy is minimized by minimizing leakage *power*. In this context, the term *shutdown* can be defined as a state

in which the leakage power of an idle block (system) is reduced relative to the leakage power of the same block (system) during normal idle mode.

For a shutdown mechanism that incurs no energy or time overhead, idle blocks should always enter shutdown, unless the state of the system is to be preserved. In reality, the time and energy overhead associated with shutdown impose a break-even time constraint, which is the minimum length of time a block must remain in shutdown in order for the energy savings to compensate for overhead. Substantial previous work has analyzed the idle mode power reduction problem for multiple shutdown modes, differing idle times, etc. (e.g. [4]).

2.1.2 Performance-critical Mode

In a performance-critical mode, the allowable delay for a given operation is less than T_{minE}, so the system is performance constrained. The most common approach in this scenario is to use the well-known dynamic voltage scaling (DVS) method. In this approach, the frequency and V_{DD} are increased together until the block barely meets the performance requirement. This approach corresponds to tracing upwards along the energy per operation curve (in and out of subthreshold if the performance requirement is high).

Another approach to increasing performance involves architectural changes such as parallelization or pipelining. These approaches have been well-studied for above-threshold operation (e.g. [5], [6]). Shallow pipelining can increase the performance in subthreshold linearly with the number of stages at the same time as reducing energy per operation, but deeper pipelining increases overall energy because of the overhead. Similarly, the opportunity to use parallelization in subthreshold is more limited than

above-threshold. The exponential increase in delay with lower supply voltages means that many more parallel paths are necessary to achieve a given speed-up. If the energy and timing overhead are negligible, a large number of parallel paths operating at the MEP can provide the lowest energy per operation for any required performance. However, the large area penalty and the significance of overhead make subthreshold parallelization impractical for many cases. Therefore, this approach should be evaluated on a system-by-system basis.

2.1.3 Performance-non-critical Mode

In a performance-non-critical scenario, the allowable delay for a given operation exceeds T_{minE}, so the system has more time than it needs to perform the operation at the MEP. From an energy standpoint, this scenario is ideal because the block can consume the minimum possible energy while completing a given operation. Once it finishes processing the data, it enters idle mode and can shutdown during the remaining timing slack.

Performance-non-critical circuits that cannot have a shutdown mode must handle this scenario differently. An example of such a system is an embedded memory capable of operating in subthreshold. The data retention ability of the bitcells limits the memory array supply voltage to a few hundred millivolts [7], which cannot be reduced further even when the memory is idle. Since the time when the system is on is the same irrespective of voltage, minimizing leakage energy equates to minimizing leakage power. For this type of block, total energy is minimized when active energy and leakage power are minimized together, both of which can be attained by operating at the lowest functional V_{DD}.

2.1.4 System Scenarios

A given system will face operating scenarios that consist of combinations of the three modes that have been described. Generally, each scenario can be decomposed into blocks in time corresponding to the different modes and then dealt with according to the procedures outlined in the preceding sections.

For example, consider a DSP running a video decoding algorithm. As the DSP decodes frames of video data, it operates in performance-constrained mode. The processor saves energy when the workload decreases by standard DVS. If the video frames arrive as bursts, then the DSP alternates between operating in performance-constrained mode and shutting down in idle mode when the idle time exceeds the break-even time. Interestingly, many such systems that require high performance operation are still performance-non-critical for a fraction of their active operation. Suppose our DSP completes its video task and needs to perform some system checking. This performance-non-critical task can run at the MEP, and then the system can shutdown.

Micro-sensor nodes provide an excellent example of systems that regularly operate in performance-non-critical mode. Since micro-sensors must operate for long lifetimes with limited energy provided by a non-replaceable battery, they need to conserve as much energy as possible [2]. Also, they spend the majority of their lifetimes in shutdown mode. Consider a micro-sensor node that awakens once per second to monitor an environmental parameter. Even at the MEP, the node can complete its processing within a few milliseconds, so it can then return to shutdown. In the relatively rare case when performance becomes critical, the node increases its supply voltage and frequency to complete the necessary operations and then returns to its standard operating scenario.

2.2 Analytical Expressions for the Minimum Energy Point

The drain-to-source current in the subthreshold regime is exponential with the gate-to-source voltage and can be modeled in a simple way as

$$I_{DS,sub} = I_o e^{\frac{V_{GS}-V_T}{nV_{th}}} \tag{2.1}$$

where I_o is the drain current when $V_{GS}= V_T$, $n > 1$ is the subthreshold slope factor [8] (practically always below 1.6) which arises due to the capacitive divider action between the depletion capacitance and oxide capacitance in a MOSFET and $V_{th} = kT/q$ is the thermal voltage. The expression for the current in equation 2.1 is simplified and does not include effects due to low V_{DS} roll-off and Drain Induced Barrier Lowering (DIBL). A more complete expression for the subthreshold drain to source current incorporating DIBL and low V_{DS} roll-off can be given as

$$I_{DS,sub} = I_o e^{\frac{V_{GS}-V_T+\eta V_{DS}}{nV_{th}}} (1-e^{-\frac{V_{DS}}{V_{th}}}) \tag{2.2}$$

where $I_o = \mu_o C_{ox} \frac{W}{L_{eff}} V_{th}^{\ 2}(n-1)e^{1.8}$ and η is the linearized DIBL coefficient.

In the following section, a closed form expression for the minimum energy operating voltage [9] is derived assuming it falls in the subthreshold regime ($V_{DD} < V_T$). Equation 2.3 gives the propagation delay for a characteristic inverter with total output capacitance C_g in subthreshold:

$$t_d = \frac{KC_g V_{DD}}{I_{o,g} e^{\frac{V_{gs}-V_{T,s}}{nV_{th}}}} \tag{2.3}$$

where K is a delay fitting parameter. The expression for current in the denominator of equation 2.3 models the on current of the characteristic inverter, so it accounts for transitions through both NMOS and PMOS devices. Thus, the terms $I_{o,g}$ and $V_{T,g}$ are fitted parameters that do not correspond exactly with the MOSFET parameters of the same name. Assuming that the critical path consists of L_{DP} number of characteristic inverters, the frequency of operation is given by

$$f = \frac{1}{t_d L_{DP}} \tag{2.4}$$

The dynamic energy consumed per switching cycle assuming rail to rail swing is given by:

$$E_{DYN} = C_{eff} V_{DD}^{\ 2} \tag{2.5}$$

where C_{eff} is the average effective switched capacitance per switching cycle. The leakage energy consumed per cycle is:

$$E_L = W_{L,eff} I_{o,g} e^{\frac{V_{gs}-V_{T,s}}{nV_{th}}} V_{DD} t_d L_{DP} = W_{L,eff} K C_g L_{DP} V_{DD}^{\ 2} e^{\frac{V_{DD}}{nV_{th}}} \tag{2.6}$$

where $W_{L,eff}$ is the average effective width of the leaking transistors. Adding equations 2.5 and 2.6, we get the total energy consumed per switching cycle as:

$$E_{TOT} = V_{DD}^{\ 2} \left[C_{eff} + W_{L,eff} K C_g L_{DP} e^{\frac{V_{DD}}{nV_{th}}} \right] \tag{2.7}$$

Equations 2.5 – 2.7 extend the expressions for current and delay of an inverter to arbitrarily larger circuits. This extension sacrifices accuracy for simplicity since the fitted parameters cannot account for all the details of every circuit. Thus, C_{eff} is the average total switched capacitance of the entire circuit and it includes within it a measure of the

activity of the circuit and the dependency of switching energy on the data being processed by the circuit. Similarly, $W_{L,eff}$ estimates the average total width that contributed to leakage current and includes within it the state dependence of leakage. E_{TOT} is a non-linear function of V_{DD} and can be minimized with respect to V_{DD} by differentiating 2.7 and equating it to 0. The value of V_{DD} that gives the minimum for E_{TOT} has been derived in [9] and is given by:

$$V_{\min E} = nV_{th}(2 - lambertW(\beta)), \tag{2.8}$$

$$\beta = \frac{-2C_{eff}}{KW_{L,eff}C_g L_{DP}} e^2 > e^{-1} \tag{2.9}$$

The *lambertW* function, *W=lambertW(x)*, gives the solution to the equation $We^W = x$. For a more detailed description of Lambert function, the reader is directed to [9].

2.3 Sensitivity of the Minimum Energy Point

This section deals with the sensitivity of the MEP to circuit operating conditions. The analytical solution for the minimum energy operating voltage V_{minE}, has been derived in the previous section. The key parameters that V_{minE} depends on are i) C_{eff}, the average total switched capacitance per cycle of operation of the entire circuit, ii) $W_{L,eff}$, the effective total width that contributes to leakage current within that cycle, iii) L_{DP}, the number of characteristic inverter delays required to complete an operation and iv) the operating temperature. Any relative increase in the active component of energy per cycle by an increase in C_{eff} will push V_{minE} lower. This is because active energy decreases as V_{DD} is lowered. Likewise, any increase in the leakage component through an increase either in $W_{L,eff}$, L_{DP} or temperature as will be described below will push V_{minE} higher. This

happens because an increase in V_{DD} speeds up the operation of the circuit and reduces the time for which the circuit leaks thereby reducing the leakage energy.

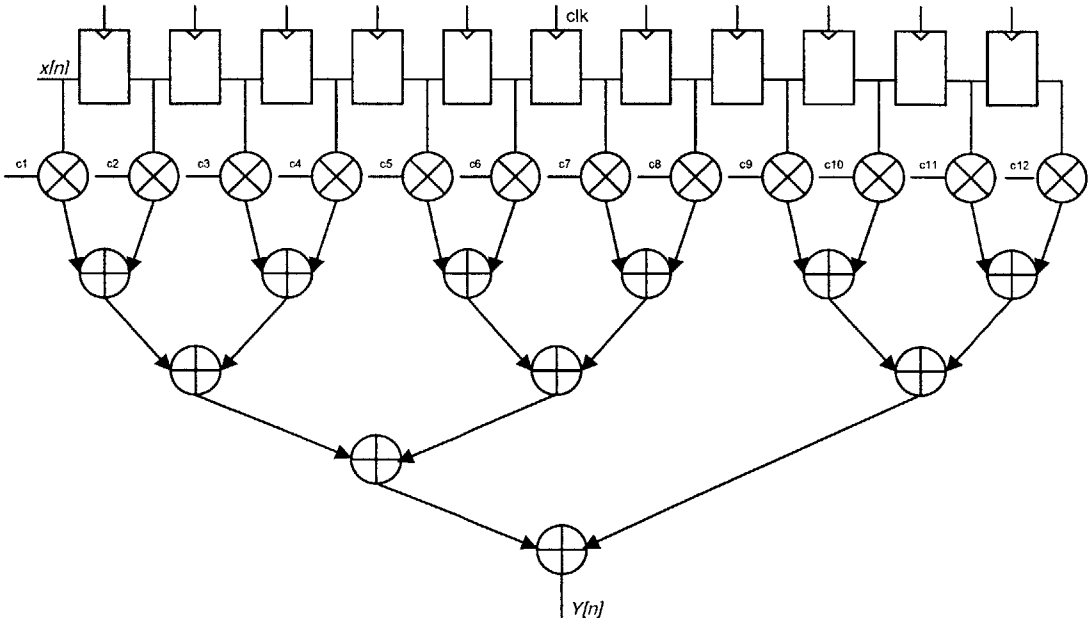

Figure 2-1: Block diagram of a 12-tap FIR filter. c_1, c_2,..., c_{12} are the tap coefficients which can be set to zero if the tap is not used. The registers can be clock gated independent of each other

The movement of the MEP with changes in workload in the circuit is demonstrated with the help of a 65nm CMOS 12-tap FIR filter in Figure 2-2 (a). The workload of the filter changes depending on the number of taps in use. The block diagram of the 12-tap FIR filter is shown in Figure 2-1. To vary the number of taps, the coefficients corresponding to the unused taps are set to zero and the corresponding registers clock gated. $W_{L,eff}$ remains constant as the number of taps is changed, as the unused taps still contribute to leakage. L_{DP} also stays the same because the time taken to finish an operation does not change by much with change in the number of taps while C_{eff} which is a direct measure of the workload in the circuit, increases in value with the number of taps. In this scenario, the active energy per operation goes up as the number of taps in use

is increased while the leakage energy per operation stays constant. This relative increase of the active energy with respect to the leakage energy causes the MEP to decrease in voltage. V_{minE} is 0.45V when the number of taps is 1 and can go as low as 0.25V as the number of taps is increased to 12. It is possible to power gate the unused taps to reduce their leakage power in processes that allow multiple threshold transistors, but that is not considered in this analysis.

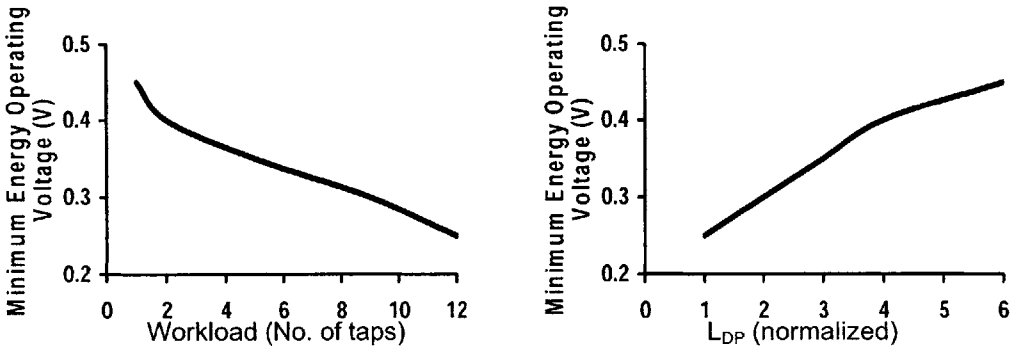

Figure 2-2: Variation of the minimum energy point, (a) with workload and (b) with L_{DP} normalized to its minimum value of a 65nm CMOS FIR filter

The MEP can also shift when there is a relative increase or decrease of the leakage energy with respect to active energy. This can occur when the amount of time required by the circuit to finish an operation is determined by external slower blocks, for instance when waiting for data from the memory. So, even though the circuit could have performed the switching operation in a much smaller amount of time, it is forced to wait. The switching energy consumed per operation stays the same but the leakage energy varies depending on the time the circuit has to wait. While C_{eff} and $W_{L,eff}$ remain fairly constant here, L_{DP} varies as the time taken to complete an operation changes.

The movement of the MEP with change in L_{DP} normalized to its minimum value is shown in Figure 2-2 (b). The change in L_{DP} was simulated by changing the time period of

the clock used by the circuit. Doubling the time period of the clock is equivalent to doubling L_{DP} and it corresponds to the circuit being forced to wait twice as longer to finish an operation. Any increase in L_{DP} increases the leakage energy per operation relative to the active energy. This causes the MEP to increase in voltage. It can be seen that the MEP moves from 0.25V to 0.45V as the normalized L_{DP} varies from 1 to 6.

A change in temperature can also affect the leakage energy per operation. The way in which temperature impacts leakage energy is dependent on the operating voltage. When the operating voltage falls close to the threshold voltage, the delay of the circuit is primarily determined by active saturation currents. Active currents increase in value with increase in temperature mainly because of the decrease in the threshold voltage of the transistors. However, this increase is small compared to the exponential increase of leakage currents with temperature. Hence, the leakage energy per operation, which is the product of the leakage current, operating voltage and the delay per operation, increases in value with increase in temperature. The active energy per operation which is responsible for charging the internal nodes of the circuit does not change in value with temperature. When the operating voltage is below the threshold voltage, it can be seen from equation 2.8 that the minimum energy operating voltage is directly proportional to the thermal voltage V_{th} which is equal to kT/q. Thus, below threshold, the MEP increases directly with temperature. Thus, irrespective of the operating voltage, the leakage energy per operation goes up with temperature while the active energy per operation stays nearly the same. This relative increase in leakage energy per operation compared to the active energy raises the value of the minimum energy operating voltage with increase in temperature as can be seen from Figure 2-4.

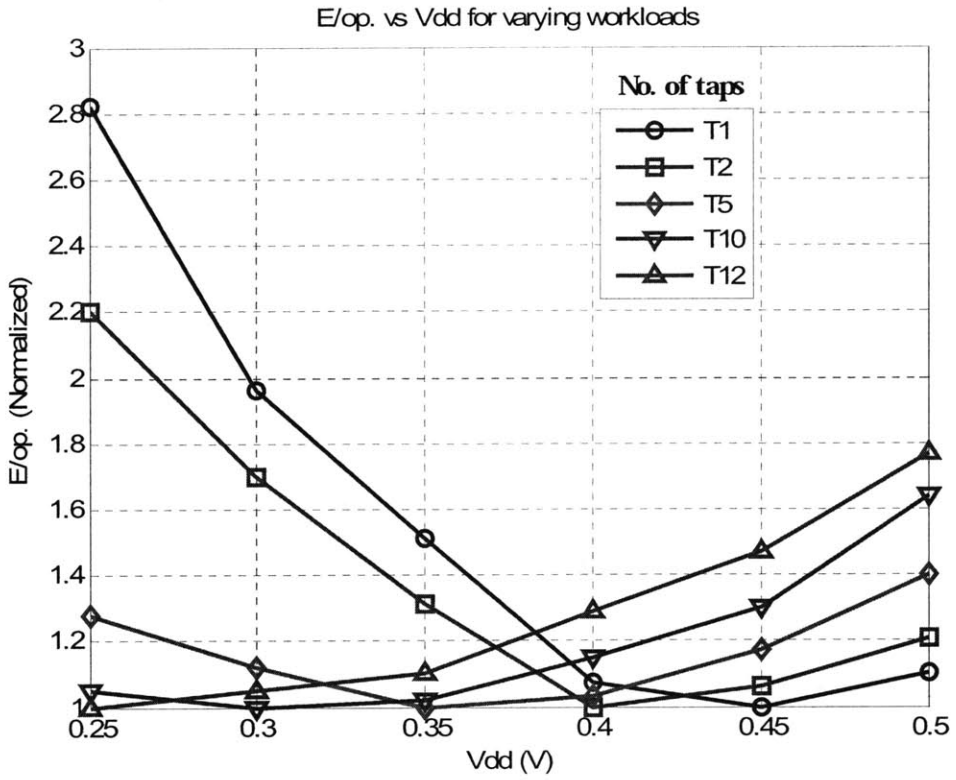

Figure 2-3: Simulated Energy/operation contours with change in number of taps of a 12-tap FIR filter at 27°c

2.4 Need for Continuous Minimum Energy Point Tracking

The shift in the MEP with operating conditions causes the energy/operation contours to change. Figure 2-3 shows the energy/operation curves, normalized to the corresponding minimum energy consumed, with change in the number of taps of the 12-tap FIR filter. While a particular voltage corresponds to the minimum energy for a specific workload, the circuit begins to consume more than the minimum energy at this same voltage with change in workload. For instance, if 0.4V was chosen as the global MEP and the operating voltage was set to 0.4V irrespective of the workload of the circuit, the circuit would consume 29% more energy than the actual minimum when 12-taps of the FIR filter

were in use. This number increases to 47% if the circuit always operated at 0.45V, to 51% at 0.35V, and can be as high as 182% at 0.25V.

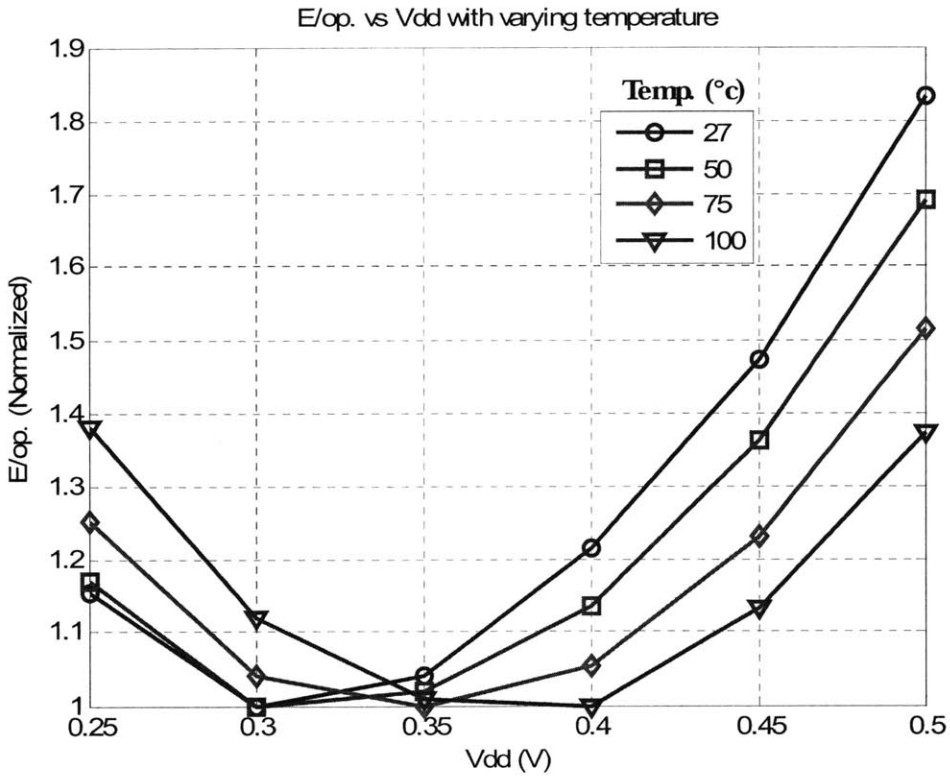

Figure 2-4: Simulated Energy/operation contours with change in temperature

Figure 2-4 shows the energy/operation curves of the FIR filter with 7 taps in use, with change in temperature. It can be seen that the MEP increases in voltage with temperature, the reasons for which were explained in the preceding section. Here again, a savings in energy amounting to 12% had the circuit always operated at 0.3V and 22% had it always operated at 0.4V can be obtained by tracking the MEP. Though it appears from the figure that only 6% extra energy would have been consumed in the worst case had the circuit always operated at 0.35V, it should be noted that even when the circuit is operating at a particular temperature, the energy contours might change depending on the current workload being processed. Hence, to truly minimize energy over varying operating

conditions, it becomes necessary to continuously track the MEP as opposed to setting the operating voltage to one particular value.

Having a continuous minimum energy tracking loop also helps in system-on-chip scenarios which consist of multiple blocks that can be turned on or off while other blocks are in operation. Depending on the application being executed, some blocks might be functioning at one point of time while others are in shutdown mode.

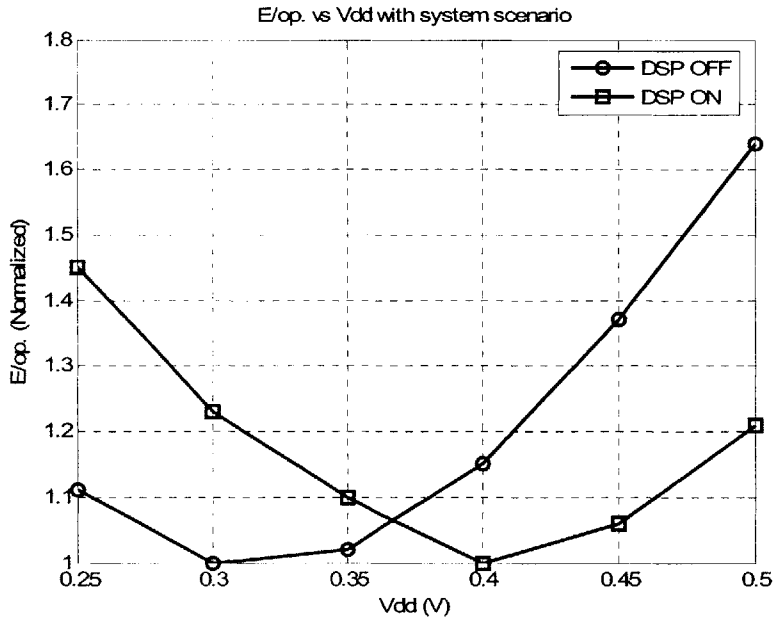

Figure 2-5: Change in energy contours when the DSP is switched ON or OFF in a FFT + DSP + Memory system

The energy contours with V_{dd} vary depending on which blocks are ON, and hence the MEP might be different at different points in time depending on the application being executed. Consider a hypothetical example of an FFT block working alongside a DSP and a memory block. Let us assume that the FFT and memory blocks always function irrespective of the application as long as the system is not in shutdown, and that the DSP might be switched ON as required. Assuming that the C_{eff} and $W_{L,eff}$ of the DSP is 3 times as large as the FFT and L_{DP} becomes thrice as large when the DSP is functioning, the

energy contours for the application when the DSP is ON and the one when it is not, are shown in Figure 2-5. It can be seen that the MEP shifts to the right as the DSP is turned ON because of a relative increase of the leakage energy compared to active energy per operation. A savings in energy amounting to 15% had the circuit always operated at 0.4V and 24% had it always operated at 0.3V can be obtained by tracking the MEP. Though it appears from the figure, that only 6% extra energy would have been consumed had the circuit always operated at 0.37V, it should be noted that even within each scenario the energy contours might change depending on the current workload being processed and the current temperature. Thus, a multi-block system on chip scenario requires a continuous minimum energy tracking system to provide energy savings over a wide range of operating conditions irrespective of the type of application being processed by the system.

The changes in workload, active cycle time and temperature described here are slowly varying in the sense that they remain fairly constant over a large number of operations before changing in value again. C_{eff}, $W_{L,eff}$ and L_{DP} can also change between cycles due to data dependency and the type of instruction being executed in processors. This data dependent change is too fast to determine the MEP and the overhead in tracking the MEP for such short term changes would overrule any potential energy savings that can be obtained. The minimum energy point tracking should be activated only when it is known that a particular load phenomenon is going to exist for some critical number of operations called the break-even number of operations, which is in principle similar to the break-even time for entering shutdown after completing operation at MEP, as explained in section II.

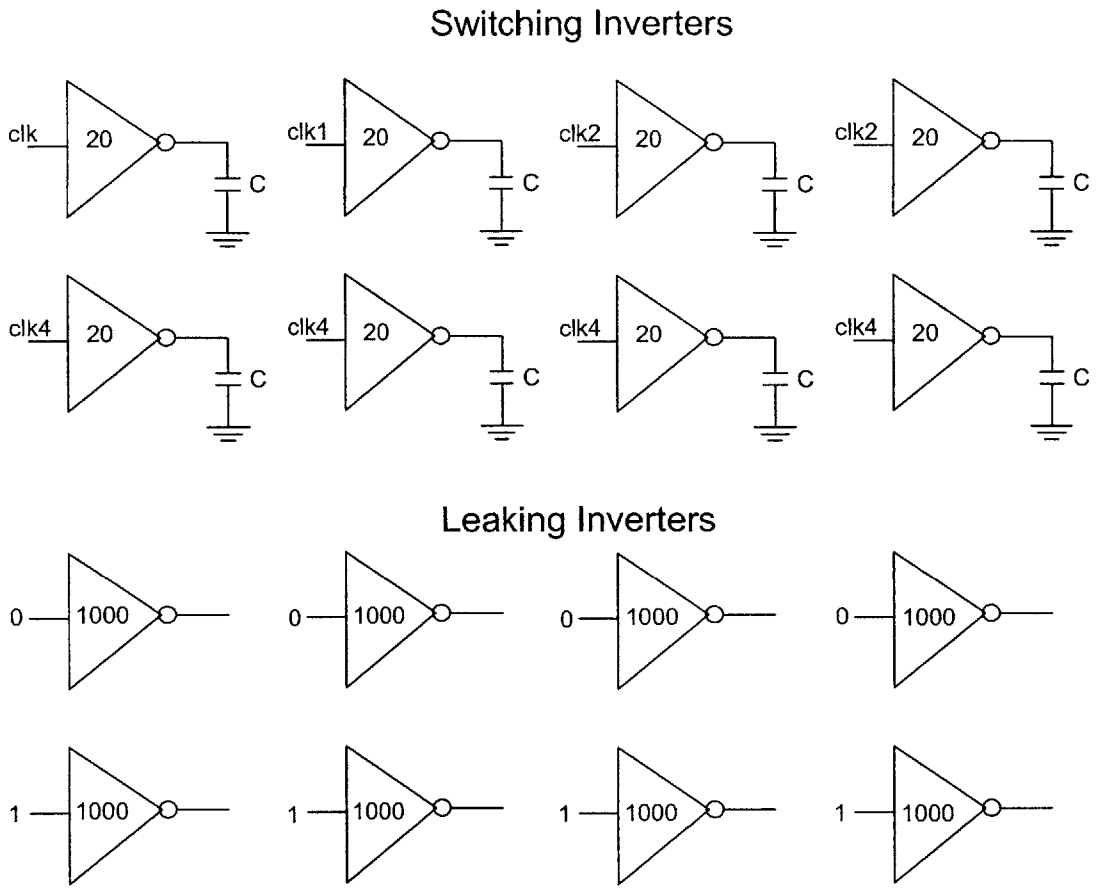

Figure 2-6: Bank of switching and leaking inverters

2.5 Measured Results

A 7-tap FIR filter was implemented on-chip in a 65nm CMOS process. The FIR filter uses the same architecture as shown in Figure 2-1 except that the number of taps is reduced to 7. The chip also has a bank of switching and leaking inverters for test purposes as shown in Figure 2-6. The numbers within the inverter symbol indicate their relative strength. Some of the leaking inverters have their inputs tied to logic 0 while others have logic 1 as their input. Only the NMOS is made stronger for the leaking inverters whose input is tied to logic 0 while only the PMOS is made stronger in the other

case. By changing the number of switching inverters through suitably enabling the different clocks, the change in activity or workload of an actual circuit can be simulated. *clk1*, *clk2* and *clk4* are all obtained from the system clock *clk*. Figure 2-7 shows the measured curves of the energy per operation changing with V_{dd} as the workload of the circuit is changed. The workload (C_{eff}) is changed by switching in more inverters while the total amount of leaking inverters ($W_{L,eff}$) remains the same. It can be seen that the MEP moves from 340mV to 400mV as the workload is increased 8 times.

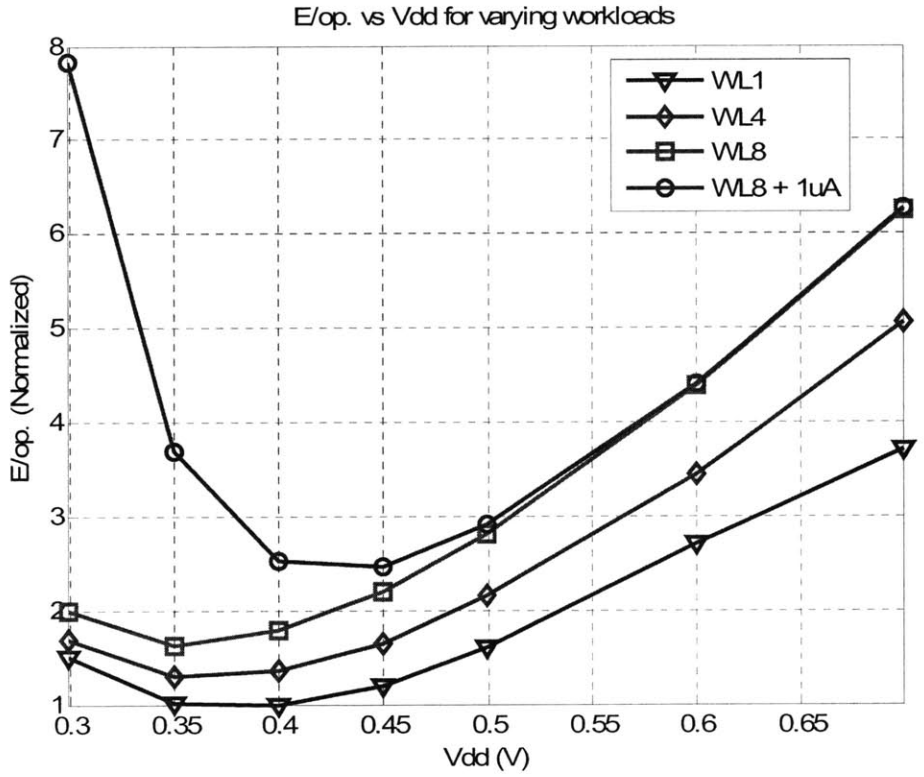

Figure 2-7: Measured Energy/operation contours with change in workload at 27°c

The WL8+1uA curve corresponds to the case where at the maximum workload, an additional 1μA of leakage current is added. This is similar to the change in system scenario case and we see that due to a relative increase in leakage energy, the MEP

moves to 440mV. All the curves in this figure are normalized to the energy per operation consumed at the MEP at a workload WL1.

Figure 2-8 shows the measured curves of the energy per operation changing with V_{dd} as temperature is changed. As predicted by equation 2-8 and as explained in Section 2.3, the MEP increases with increase in temperature. Here again, the curves are normalized to the energy per operation consumed at the MEP when the temperature is 0°c.

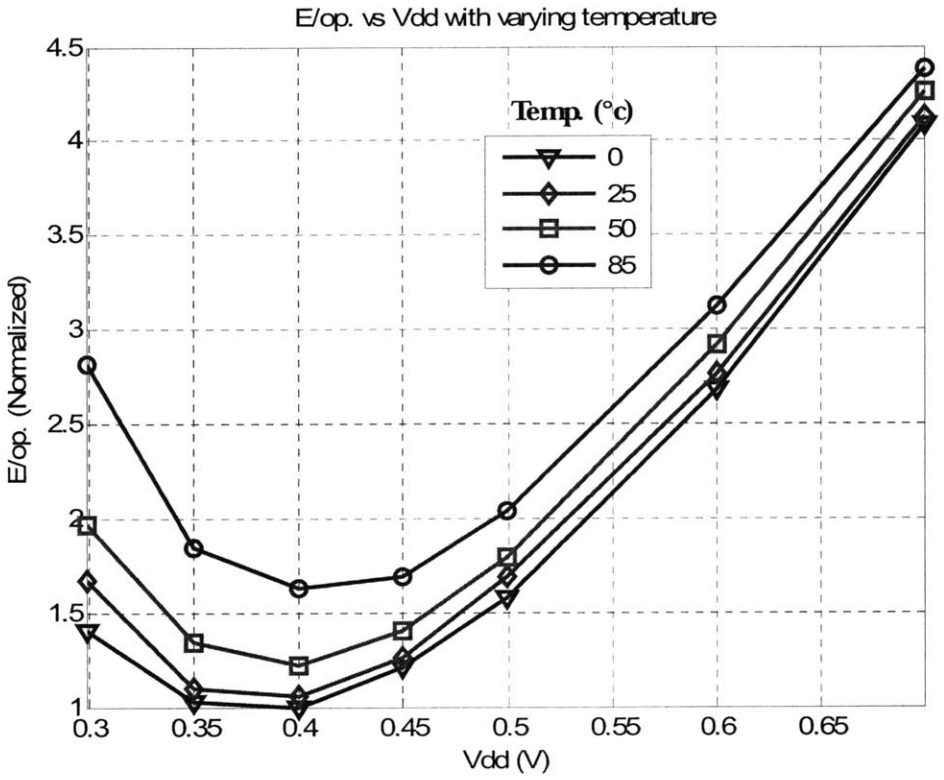

Figure 2-8: Measured Energy/operation contours with change in temperature

Chapter 3

Low Power Subthreshold DC-DC Converter

Operating digital circuits at subthreshold voltages necessitates a low power highly efficient DC-DC converter that is capable of delivering subthreshold voltages from a standard 1.2V Ni-MH battery. The DC-DC converter for the minimum energy tracking circuitry should be able to provide stable voltages from 300mV all the way up to 700mV. This is the region within which the minimum energy operating voltage falls and the DC-DC converter has to be able to deliver these voltages at high efficiencies. A DC-DC converter takes as its input a reference voltage which can be set in an analog or digital fashion. It is then expected to provide an output voltage that is very close to the desired value within an acceptable ripple voltage level. The primary metrics for the DC-DC converter are its efficiency, the chip area and board volume that it occupies and the ripple to which it can stabilize the output voltage. There are various converter topologies that trade one of these metrics to gain the other. These different topologies and the various tradeoffs involved will be discussed in the next few sections.

3.1 Converter Topologies

Traditionally DC-DC conversion has been performed using voltage regulators. Power supply voltage regulation can be achieved through a variety of converter topologies namely linear regulators, switched capacitor voltage dividers or switching regulators. Since power supply regulation for integrated circuits in sub-micron technologies usually involves an off-chip battery that delivers the maximum voltage the circuits can operate at, only voltage dividers would be considered in this thesis. A whole other class of converters for voltage boosting has been used for long in power applications [11].

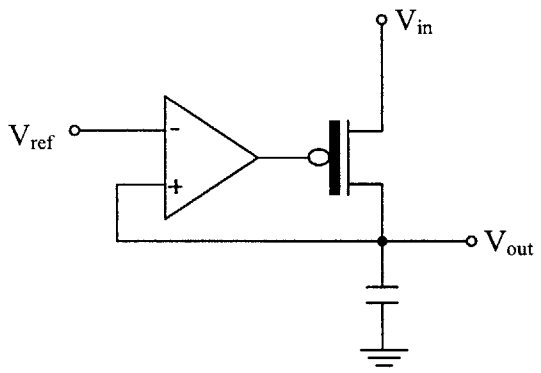

Figure 3-1: A Linear Regulator

3.1.1 Linear Regulators

Linear regulators, a conceptual illustration of which is shown in Figure 3-1, are limited by two constraints. The output voltage V_{out} must be less than the input voltage V_{in}. This is not a primary constraint in the application addressed in this thesis as subthreshold voltages are much lower than the minimum voltage the battery can droop down to. The main problem with linear regulators is that their efficiency η can never be greater than

V_{out} / V_{in}. This essentially prohibits the use of linear regulators for subthreshold applications. For example consider delivering 0.3V to the digital circuits from a battery voltage of 1.2V. This would limit the efficiency of the linear regulator to 0.3/1.2 = 0.25. This low efficiency is unacceptable for battery powered applications and negates the potential advantages of going subthreshold. The main advantage of linear regulators is that they require only a few or in some cases no reactive components, and hence can be made really simple and small. Though this is a major advantage when it comes to portable applications where the converter volume is a primary design constraint, the extremely low efficiencies that linear regulators provide at subthreshold voltages limits their usage.

Linear regulators can be efficient in applications where the output voltage is only slightly below the input voltage. An example of this scenario would be in cases where a linear regulator is used after a switched capacitor converter. Since, a switched capacitor converter cannot inherently regulate the output voltage, a linear regulator often follows a switched capacitor converter to provide a smooth low ripple output voltage. In this case, since the output voltage of the switched capacitor converter is very close to the final desired output voltage a linear regulator provides regulation at high efficiencies. More on linear regulation and its usage in high power applications can be found in [11].

3.1.2 Switched Capacitor Converters

Switched capacitor converters (charge pumps) are widely used in applications where a voltage higher than, or of the opposite polarity to, the input voltage is needed. A switched capacitor converter comprises only of capacitors and hence doesn't need the bulky magnetic storage elements needed in buck converters. This advantage of switched

capacitor converters also acts to their disadvantage. This is because of the inherent loss in energy when transferring charge from a capacitor at a higher voltage to one at a lower voltage. In the absence of any intermediate storage elements, some of the energy transferred would always be lost in the connecting switches no matter how large the switches are made. This energy lost is proportional to the difference in voltage between the charging capacitor and the one that is being charged.

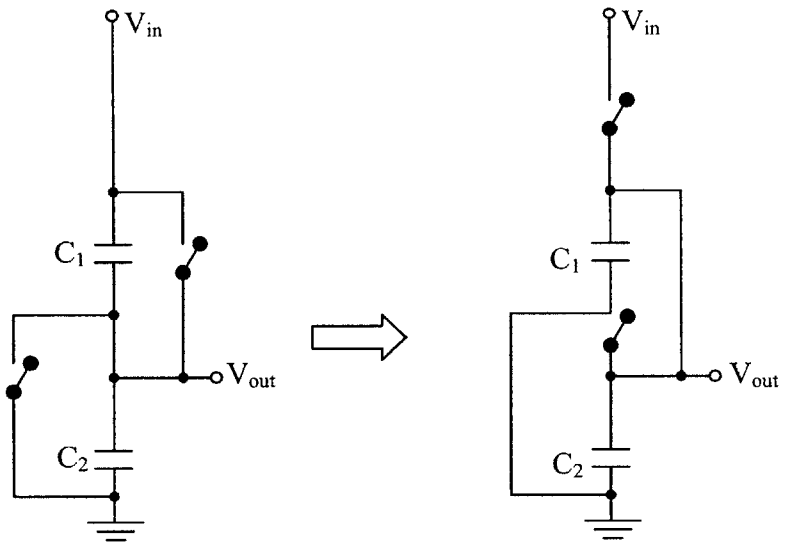

Figure 3-2: A Switched Capacitor voltage divider

Figure 3-2 illustrates the basic principle of operation of a switched capacitor voltage divider. There are 4 switches, 2 of which are closed during φ_1 while the other 2 are open and vice-versa. During the φ_1 phase, the capacitors C_1 and C_2 get charged to half the input voltage assuming that the capacitors are equal and the bottom-plate parasitics are small. Then the φ_1 switches are opened and the φ_2 switches are closed. This places C_1 which was charged to $V_{in}/2$ in parallel with the similarly charged C_2 and together they charge the output load capacitor. This cycle repeats and maintains the output voltage near $V_{in}/2$.

Switched capacitor converters can be used to produce any rational conversion ratio, for example by stepping up the voltage by an integer ratio and then stepping it down by another integer ratio.

While switched capacitors are very good in converting voltages, they are not very efficient voltage regulators [12]. This combined with the inherent losses in a switched capacitor circuit limits the usage of these converters in practical circuits. In the case of the low power subthreshold DC-DC converters, it is the inherent loss within the switched capacitor converters which prevents them from being used as very tight voltage regulation is not a rigid necessity for ultra low power applications because of the flat nature of the E/op. vs. V_{dd} curve near the minimum energy operating voltage.

3.1.3 Switching Regulators

Switching regulators are converters that can be used to create precise output voltages, above or below their input voltage, and have the additional advantage that they are ideally lossless converters. This means that if the switches and the filter passives used in the converter are ideal with no parasitics and the external control circuitry consumes negligible power, then theoretically switching regulators can be made 100% efficient. Unlike linear regulators, the switching converter is a sampled rather than a continuous time system. Depending on the topology of the switching and the filter elements, a switching regulator can be used to step-up or step-down a voltage, with the same or opposite polarity. One of the topologies that is most commonly used to deliver low voltages on an integrated circuit called the step-down buck converter [13] would be discussed in detail in the following section.

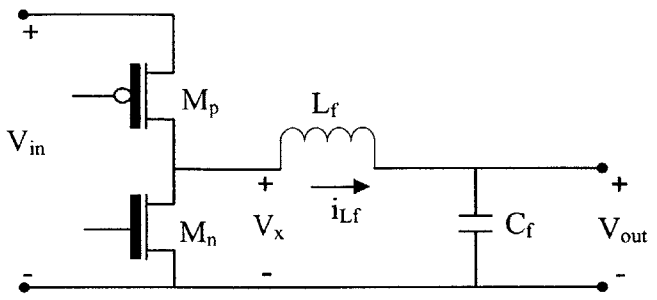

Figure 3-3: A synchronous rectifier buck converter

The buck converter, a typical implementation of which is shown in Figure 3-3, consists of the power transistors M_p and M_n, and the filter elements L_f and C_f in its core. Traditionally, a fly back diode [11] has been used for high voltage applications to provide continuous inductor current, but in on-chip applications where the voltages are usually low, the 0.7V drop across a diode can be debilitating in terms of efficiency and hence the diode is replaced by an NMOS transistor which is switched appropriately. This type of implementation is called the synchronous rectifier buck converter. The buck converter is a direct converter in that there is a direct connection between the input source and the output for a part of the switching cycle. It operates by storing some part of the energy from the battery in the inductor during the PMOS ON cycle and when the NMOS is turned ON, the energy stored in the inductor is delivered to the load.

Buck converters can be operated in two modes based on how the power transistor's switching is controlled. The first mode is called the Pulse Width Modulation mode [14] in which the power transistors chop the battery input voltage to reduce the average voltage that is output. This produces a square wave of variable duty cycle D before the filter at the node V_x which is then filtered to give an average output voltage of $V_{out} = D.V_{in}$. The transistors are switched at a constant frequency f_s. A typical periodic steady-state waveform of the buck converter operating in the PWM mode is shown in Figure 3-4.

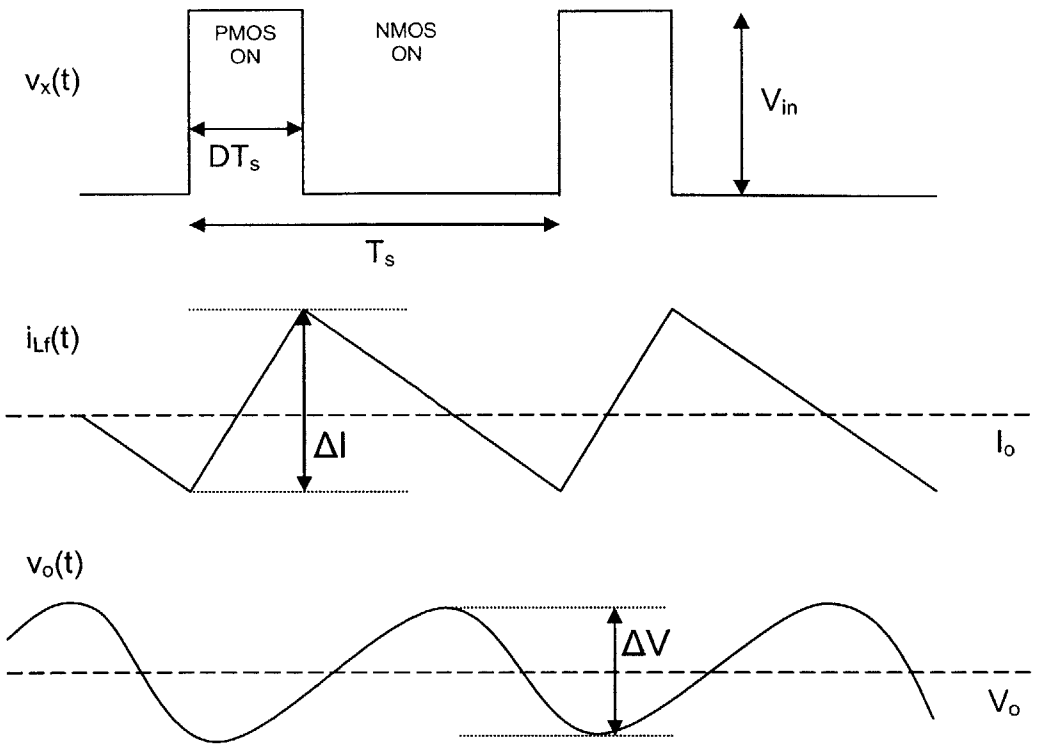

Figure 3-4: Nominal steady-state waveforms of the buck converter operating in PWM mode

When the PMOS is ON, the voltage at node V_x charges up to V_{in} and stays there. At this time the inductor current i_{Lf} ramps up with a slope $(V_{in} - V_o)/L$. When the PMOS turns OFF and NMOS turns ON, the voltage at V_x falls to zero and the inductor current ramps down with a slope V_o/L. When the inductor current is positive (negative), the output voltage increases (decreases). This cycle repeats itself every T_s seconds and the output voltage is held close to $D.V_{in}$, with a ripple of ΔV.

A different mode of operation of the buck converter is called the Pulse Frequency Modulation (PFM) mode where the transistors are switched ON only when the output voltage falls below the desired reference voltage V_{REF}. Every switching cycle, the PMOS switch is turned ON first ramping up the inductor current. The PMOS switch is turned OFF after a specified time τ_{PMOS} and the NMOS switch is turned ON. The NMOS switch

remains ON ramping down the inductor current till it hits zero. The time for which the NMOS is ON is dependent on V_{in}, V_o and τ_{PMOS}. During the time when the inductor current is positive, the output load voltage increases. When both the transistors are OFF, the load current ramps down the voltage across the output load capacitor, till it falls below the reference voltage and another switching cycle begins. Depending on the power the load consumes, the frequency of switching can be widely different especially at subthreshold voltages where the load power consumed falls exponentially with voltage. A typical periodic steady-state waveform of the buck converter operating in the PFM mode is shown in Figure 3-5.

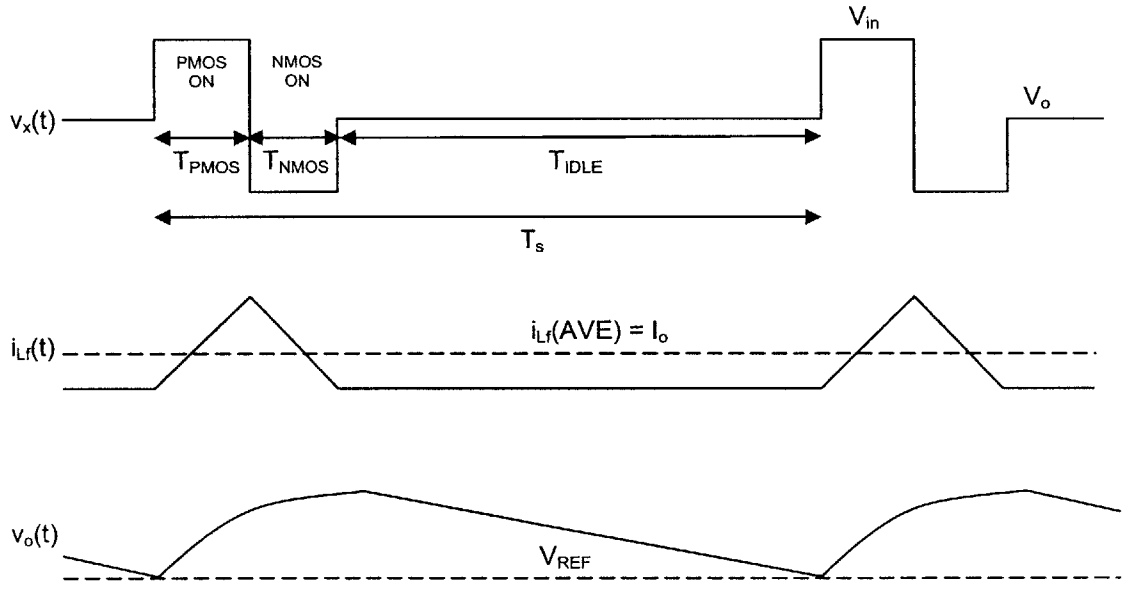

Figure 3-5: Nominal steady-state waveforms of the buck converter operating in PFM mode

3.2 Efficiency of Buck Converters

The various sources of power loss that brings down the efficiency of the buck converter in both the PWM and PFM modes of operation are discussed below.

3.2.1 Conduction Loss

Conduction loss is common to both the PFM and PWM modes of operation. Current flow through non-ideal power transistors, filter elements, and interconnections results in dissipation in each component. The power dissipated in any component through conduction in PWM mode can be given by

$$P_{cond,m} = i_{rms,m}{}^2 R_m \tag{3.1}$$

where $i_{rms,m}{}^2 = i_{rms,mDC}{}^2 + i_{rms,mAC}{}^2$ is the rms current through that component and R_m is its resistance. The DC component of the rms current is proportional to the average load current I_o, and the AC component is proportional to the current ripple ΔI. Thus, while the DC conduction loss scales quadratically with the load current, AC conduction loss is a fixed quantity which depends on the frequency of switching, the output voltage and the value of the filter inductor. This is detrimental to low output load power operations as the constant AC conduction loss may degrade the efficiency substantially.

Since, in PFM mode the current through the inductor is not continuous and the transistors are OFF for a major part of a charge cycle, it is more convenient to discuss the efficiency of the PFM mode controller by talking about energy losses in the converter as opposed to power losses. The energy dissipated in one charge cycle due to conduction loss in the power transistors, non-ideal passives and interconnects can be given by,

$$E_{cond} = \int_0^{T_{pulse}} i(t)^2 R dt \tag{3.2}$$

where T_{pulse} is the duration during which current flows through the inductor and is equal to the sum of τ_{PMOS} and τ_{NMOS}, the times for which the PMOS and NMOS power transistors are ON respectively. R includes the parasitic resistance of the passives and the on-state drain-to-source resistance of the power transistors. While calculating the integral, the first part of integration from 0 to τ_{PMOS} should include the resistance of the PMOS transistor and the second part from τ_{NMOS} to T_{pulse} should include the resistance of the NMOS transistor. During their conduction intervals, the power transistors operate in the triode region where their resistances can be modeled as being inversely proportional to their width W.

3.2.2 Switching Loss

Charging up the gate capacitance of the buck converter's transistors is not a lossless process and involves considerable amount of power consumption. In the PWM mode the power loss due to conduction is given by

$$P_{sw} = E_{sw} f_s \tag{3.3}$$

where E_{sw} comprises of the energy needed to switch ON and OFF the PMOS and the NMOS transistors once in a cycle. The gate capacitance includes not only the gate-to-source capacitance of the transistors but also additional capacitance due to Miller effect on the gate-to-drain capacitance. The energy consumed for switching in one cycle of operation of the PFM mode converter is given by E_{sw}. This energy is proportional to the

width of the switching transistors and hence can be written as $E_{sw} = E_{g0} * W$, where E_{g0} is a constant of proportionality and includes C_{ox} and the minimum channel length L.

3.2.3 Loss due to timing errors

Timing losses arise due to non-optimal switching of the NMOS and PMOS turn-on pulses. The different timing loss mechanisms are:

i) Short-circuit loss: This can occur when there is no dead time between the PMOS turning OFF and NMOS turning ON. This creates a short circuit path and the two devices conduct simultaneously gorging power.

ii) NMOS Body-diode conduction: If the dead time is too large, the body-diode of the NMOS transistor may be forced to pick up the inductor current for a fraction of each cycle. Since there is a voltage drop across the diode when it is conducting, there is power lost each time the diode burdens a part of the inductor current.

iii) Capacitive-switching loss: Each time the PMOS is switched ON, it not only charges the load capacitance and the inductor but also parasitic capacitances along the drain of the device. Once the PMOS is turned OFF, this parasitic capacitance feeds the inductor current till either the NMOS is turned ON or till it discharges completely wherein the body diode conduction begins. Thus ideally the dead time should be equal to the time it takes the parasitic capacitance to discharge. This is referred to as zero-voltage switching (ZVS) [12]. If the NMOS switches ON before this time, the energy stored in the parasitic capacitance is lost and this adds on to the other losses to bring down the efficiency.

iv) PMOS Body-diode conduction: If the NMOS is turned ON for a time longer than is necessary for the inductor current to get to zero, the inductor current reverses direction and goes negative. Now, when the NMOS is turned OFF, the PMOS body-diode starts conducting and adds to further losses.

The overall loss due to timing errors can be accounted for in the efficiency equation by assuming that P_{timing} is the power lost due to timing errors which corresponds to E_{timing} energy loss per switching cycle.

3.2.4 Leakage Loss in the power transistors

Owing to the extremely large power transistors in use in a DC-DC converter and the extremely low load power levels at subthreshold voltages, the leakage power loss in the power transistors is a major component that brings down the efficiency of the overall converter, when the converter is delivering low load powers. This is especially a factor in scaled technologies like 65nm CMOS, where leakage currents are significant. Assuming I_{leak} is the average off-state leakage current through the PMOS and NMOS power transistors, the power lost due to subthreshold leakage is given by $P_{leak} = V_{in} * I_{leak}$, where V_{in} is the nominal battery voltage. The energy lost due to leakage of the power transistors in one switching cycle of the PFM converter is given by $E_{leak} = V_{in} * I_{leak} * T_s$. The leakage current through the power transistors is directly proportional to their width W and hence the leakage loss scales proportionally with the width of the power transistors.

3.2.5 Loss in control circuitry

While the losses mentioned above occur in the DC-DC converter core, there are other losses in the control and pulse generation circuitry which severely hamper efficiency at

ultra low load power levels. The control circuitry for the PFM mode converter consists of PMOS and NMOS pulse generators, a comparator which regulates the output voltage and a level converter. The PWM mode controller has a Pulse Frequency Detector in place of the voltage comparator. The functioning of each of the control elements used in the PFM mode controller is explained later in this chapter. Each of these circuit elements consumes some amount of switching energy during each charge cycle. Apart from that they also consume considerable leakage power when idle. This is of specific concern at subthreshold voltages where the load power is orders of magnitude lesser than at above threshold voltages as shown in Figure 3-6.

3.2.6 Overall Efficiency

Considering all the losses encountered within the DC-DC converter core and the external control circuitry, the overall efficiency (η) for the PWM mode buck converter can be expressed as:

$$\eta = \frac{Power\ delivered\ \ to\ load}{Power\ drawn\ from\ the\ supply} = \frac{P_{load}}{P_{load} + P_{cond} + P_{sw} + P_{leak} + P_{timing} + P_{control}} \tag{3.4}$$

Since it is easier to talk in terms of energy delivered per charge cycle in PFM mode as the switching frequency is not fixed, the efficiency equation for the PFM mode converter can be expressed as:

$$\eta = \frac{Energy/cycle\ delivered\ \ to\ load}{Energy/cycle\ drawn\ from\ the\ supply} = \frac{E_{load}}{E_{load} + E_{cond} + E_{sw} + E_{leak} + E_{timing} + E_{control}} \tag{3.5}$$

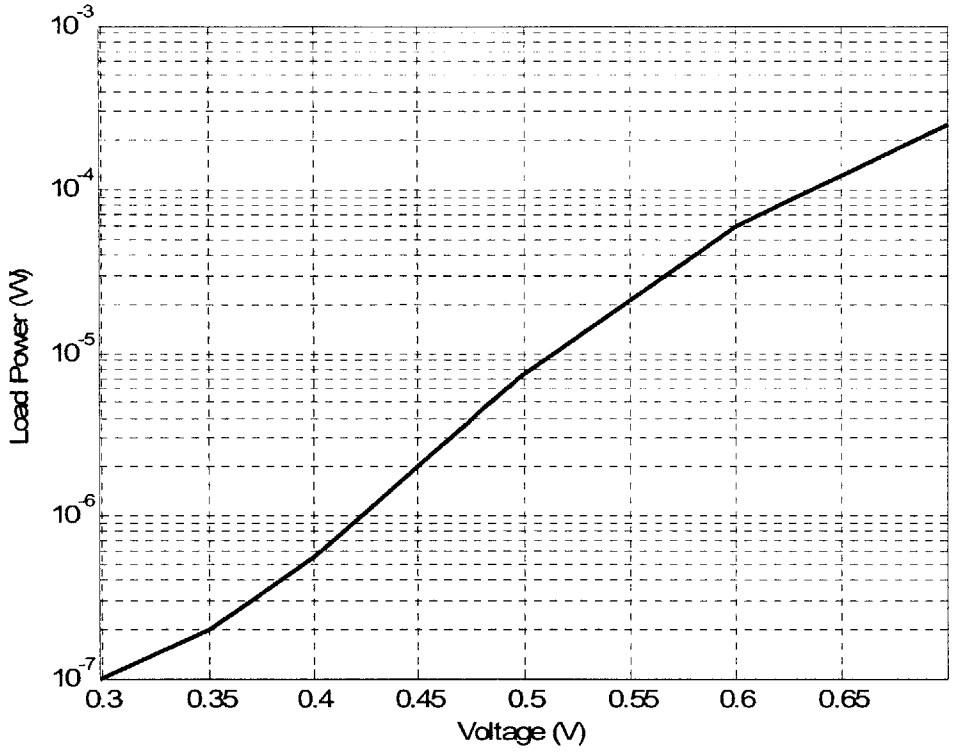

Figure 3-6: Plot showing the variation of load power with V_{dd} for a 7-tap FIR filter as the load

3.3 Converter Implementation

The DC-DC converter to be used as a part of the energy minimization loop has to supply a load whose power consumption varies from around .1μW to around 250μW as shown in Figure 3-6. While a PWM DC-DC converter can be made highly efficient at full load which is of the order of milliwatts, many of its losses are independent of load current, and therefore, a significant portion of the output power is dissipated within the converter. As the load scales downward, AC conduction loss, switching loss and PWM control power become increasingly significant and total dissipation in the converter asymptotes to fixed minimum power dissipation. Since, at subthreshold voltages where the minimum energy operating voltage usually falls, the load power is extremely small, the alternative

control scheme to achieve high efficiency is the PFM mode. The advantage with PFM mode of operation is that the power transistors are switched ON only when necessary, i.e. when the load voltage falls below the reference voltage. Most of the losses in the PFM mode converter are load power dependent and hence, the losses in PFM mode scale with the load power. Also, the control for a PFM mode converter which is essentially an on-off mode control is very simple. This brings down the power consumption of the control circuitry which is load independent.

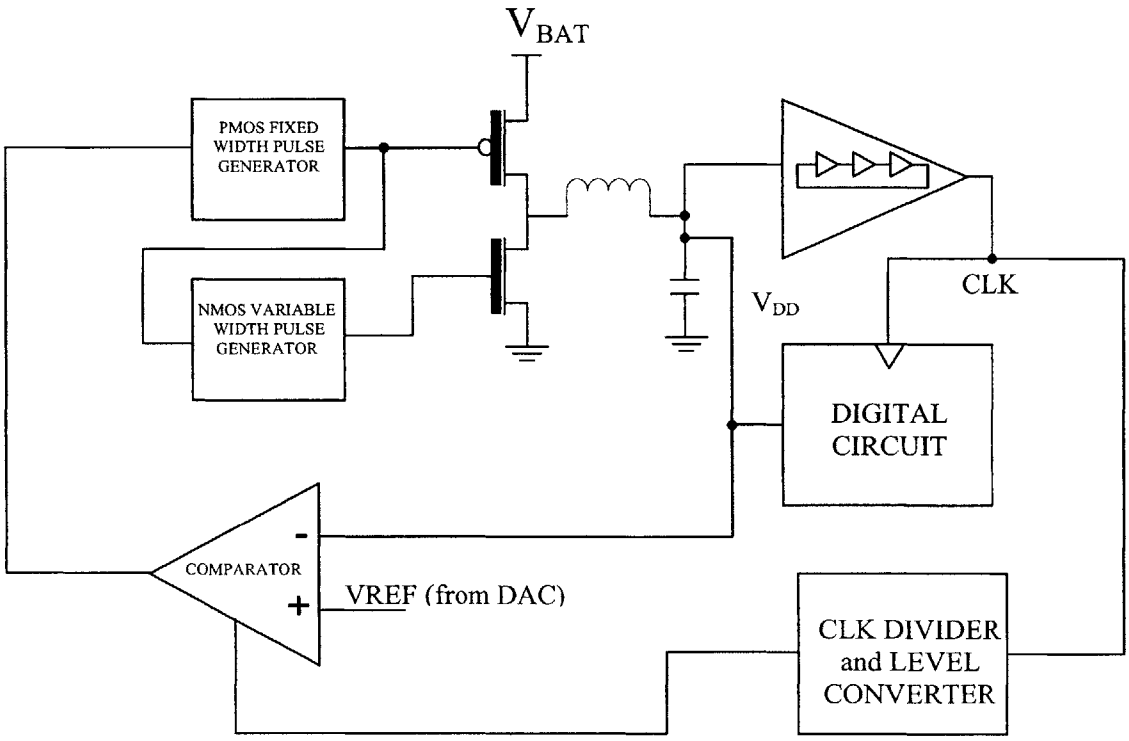

Figure 3-7: Block diagram of the PFM mode buck converter

The key blocks that comprise the PFM mode DC-DC buck converter (Figure 3-7) include the power transistors, the filter elements, pulse generators for the PMOS and NMOS power transistors, the pulse width adjusting circuitry, the reference voltage

comparator, and the clock divider and level converter which feeds a variable frequency clock to the comparator. Apart from the filter inductor and capacitor which are off-chip blocks, the rest of the components of the buck converter are made on-chip. Each of these blocks needs to be carefully designed with appropriate tradeoffs to build a highly efficient power converter. The design of these individual blocks is discussed below.

3.3.1 Power Transistors

The power transistor design trades off conduction loss with switching and leakage loss. As was explained in the previous section, the switching loss and leakage loss increase proportional to W while the conduction loss decreases with W. During their conduction intervals, the power transistors operate exclusively in the triode region, where their ON resistance is inversely proportional to their width W. Thus the energy lost due to the finite non-zero resistance of the transistors can be modeled as:

$$E_{cond,M} = \frac{i^2{}_{d,rms} R_0 T_{pulse}}{W} \tag{3.6}$$

where R_0 is a constant of proportionality, and $i_{d,rms}$ is the rms drain-to-source current flowing through the transistors, when they are ON.

The device parasitics scale linearly with W and hence the energy lost in driving the gate of these transistors increases with W. Also, the subthreshold leakage current and hence the leakage power increases linearly with W. The energy lost due to switching and subthreshold leakage in a switching cycle can be modeled as:

$$E_{total} = \left(E_{g0} + I_{l0} V_{bat} T_s\right) W \tag{3.7}$$

where I_{l0} is the leakage current in a unit width transistor at the nominal bias conditions. Since, the transistors leak even when the switching pulses stop, the leakage power has to be integrated over the whole switching cycle and hence, the leakage power is multiplied by T_s, the time period of the switching cycle.

The conduction and leakage losses are dependent on the operating point, i.e. the load voltage and load power. This is because $i_{d,rms}$, T_{pulse} and T_s all depend on the operating point. Hence, minimizing the total energy lost at the most critical operating point, we get the optimum width of the transistors as:

$$W_{opt} = \sqrt{\frac{i^2{}_{d,rms} R_0 T_{pulse}}{E_{g0} + I_{l0} V_{bat} T_s}} \tag{3.8}$$

3.3.2 Output Filter Design

The filter elements are decided on the basis of the maximum load current that can be supported and the output voltage ripple allowable. To support a maximum load current, $I_{0(max)}$, and a peak-to-peak output voltage ripple ΔV ,the inductor and capacitor in the filter need to be:

$$L_f = \frac{T_{PMOS}(V_{in} - V_o)}{2I_{0,\max}} \tag{3.9}$$

$$C_f = \frac{Q_L}{\Delta V} \tag{3.10}$$

where $Q_L = \frac{T_{PMOS}{}^2 (V_{in} - V_o) V_{in}}{2 V_o L_f}$ is the total charge delivered to the capacitor in one charge cycle.

3.3.3 Pulse Generation Circuitry

The pulse generation circuitry is implemented digitally with the help of delay lines. The width of the PMOS switch ON pulse is fixed at a width τ_{PMOS}. The delay is obtained using cascaded inverters, the length of whose transistors are made much more than their width to obtain large delays. Fixing τ_{PMOS} as the controlling variable, the width of the NMOS pulse can be obtained for a given operating voltage V_o as:

$$\tau_{NMOS} = \frac{(V_{in} - V_o)}{V_o} . \tau_{PMOS} \tag{3.11}$$

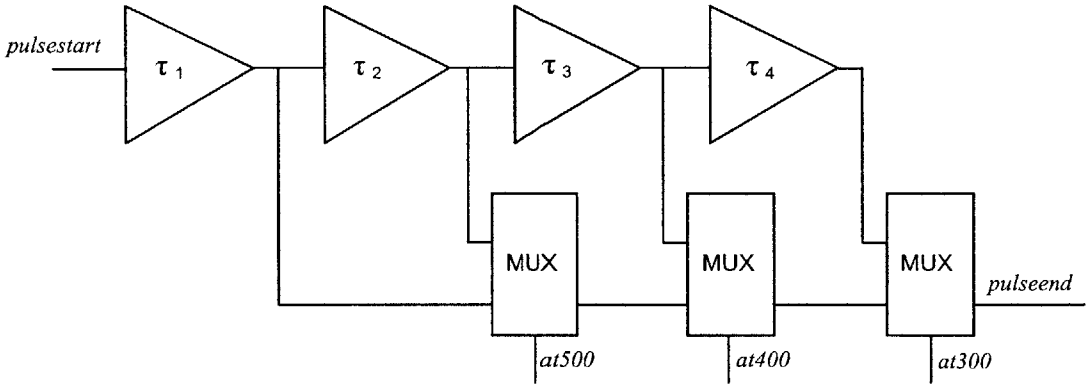

Figure 3-8: NMOS pulse width determining circuitry

The operating voltage V_0 when the converter is in steady state operation would be very close to the reference voltage V_{ref}, assuming that the ripple is small which is usually the case for a reasonable sized load capacitor. Since, we have the digital representation of V_{ref}, we can choose the delay that corresponds to an operating voltage and just multiplex the required delay. Figure 3-8 shows a block diagrammatic representation of the pulse width determining circuitry. τ_1, $(\tau_1+\tau_2)$, $(\tau_1+\tau_2+\tau_3)$ and $(\tau_1+\tau_2+\tau_3+\tau_4)$ correspond to the NMOS pulse widths required at an operating voltage of 600mV, 500mV, 400mV and 300mV

respectively. These can be calculated from equation 3.11. The ratio of the required NMOS pulse width to the PMOS pulse width with change in operating voltage (V_o) is shown in Figure 3-9. The blocks corresponding to the individual delays are made up of similar cascaded inverters as those used to construct the PMOS pulse. Also, all the delay elements used to construct both the PMOS and NMOS pulses are laid out close together. Since the absolute value of the delay doesn't matter as long as the ratios stay the same, the advantage with the above practice is that process corner variations will not have a harmful effect on the delay ratios. Also, since these delay blocks operate above threshold, random dopant variations which affect the threshold voltage of the transistors do not play a significant role in changing the delays.

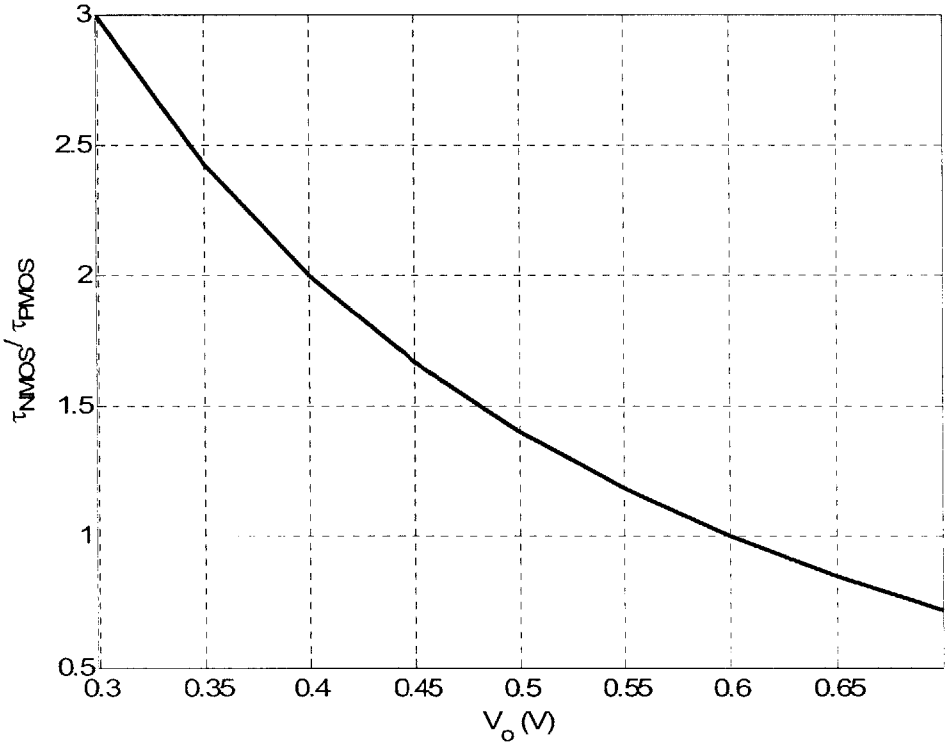

Figure 3-9: Ratio of the required NMOS pulse width to PMOS pulse width with operating voltage

The signals *at300*, *at400* and *at500* are obtained from the digital representation of V_{ref}. Since, the NMOS pulse width is determined only at specific operating points, the pulse width obtained would be different from the exact pulse width needed from equation 3.11 for operating voltages that are not the same as the ones used to calculate the delays. For example, if the operating voltage V_0 is 360mV; in the above circuit *at400* would go high and the NMOS pulse width corresponding to 400mV would be switched in. Thus, the pulse width obtained would be lesser from the actual pulse width needed by a factor,

$$\frac{(1.2-0.36)*0.4}{(1.2-0.4)*0.36}=1.17$$

This error in the NMOS pulse width manifests itself as a timing error and leads to a decrease in efficiency as explained in section 3.2.3. The error can be minimized by incorporating multiple operating points in the pulse width determining circuitry at the cost of increasing complexity of the multiplexing and decoding circuitry.

3.3.4 Level Converter Circuitry

The reference voltage comparator is clocked by a divided version of the clock generated by the critical path ring oscillator. Depending on the voltage the circuit is operating of, the voltage level of the clock varies. This needs to be converted to 1.2V in order to clock the comparator. The level conversion circuitry that is used to perform this conversion is shown in Figure 3-10.

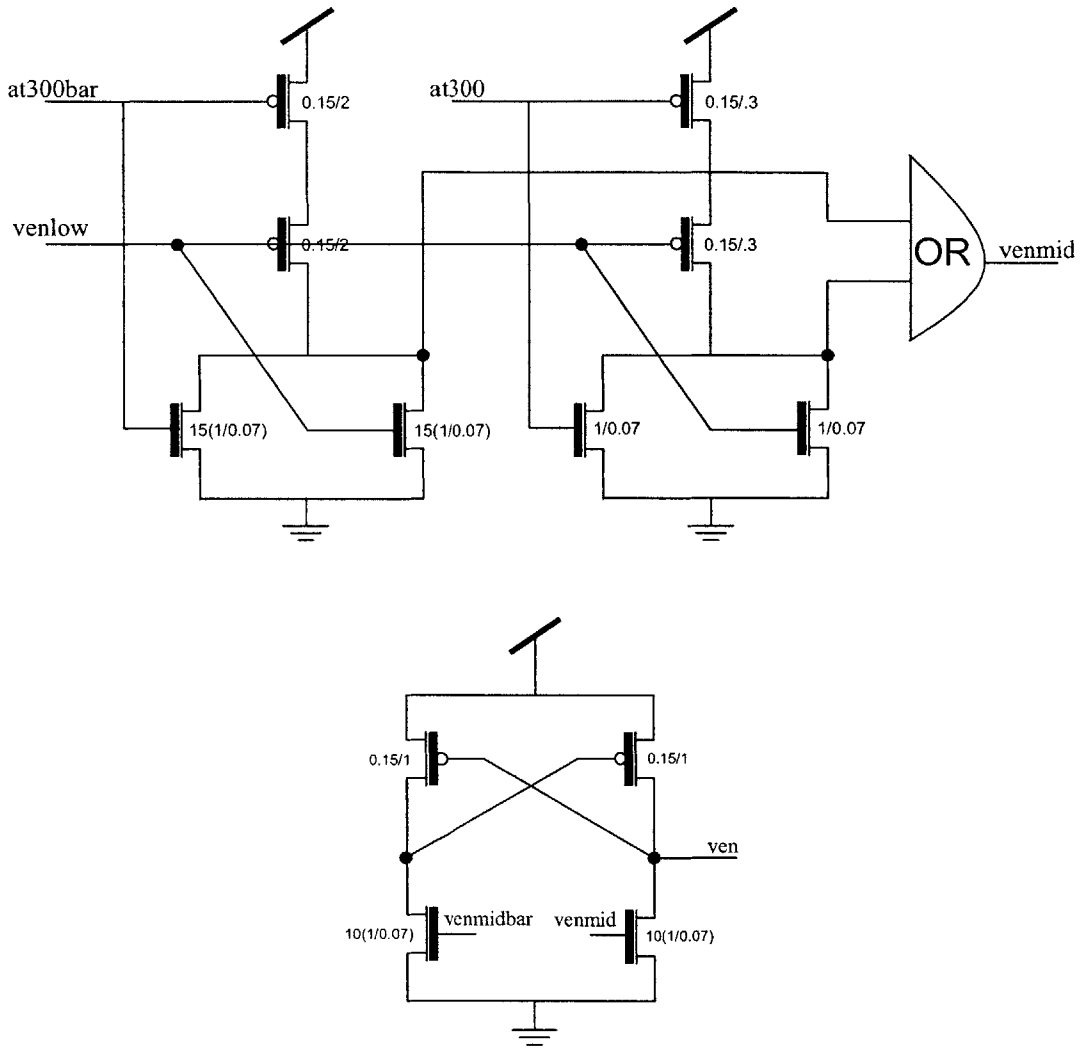

Figure 3-10: Level Conversion circuitry

The level conversion is done in two steps. In the first step, the clock is converted to 0.6V. The second step takes this clock with a voltage of 0.6V and converts it to a 1.2V clock. Since at 0.3V, the NMOS drive currents are exponentially smaller, the devices corresponding to 0.3 –> 0.6V conversion need to be sized accordingly. This particular sizing when used to convert other voltages, consumes a lot of switching energy and hence, two different conversion branches are used. The outputs of these branches of which only

one has the level converted clock are ORed together and fed to the 0.6 –> 1.2V converter. The sizing for this converter is different from the other two owing to above threshold voltage conversion.

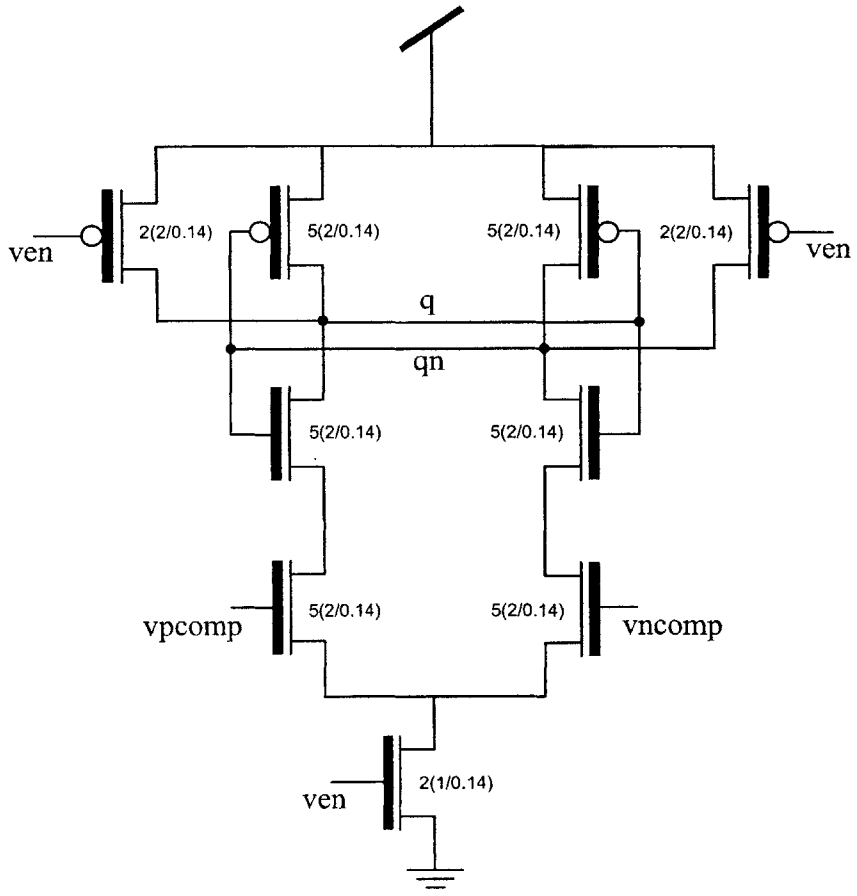

Figure 3-11: Reference Voltage Comparator

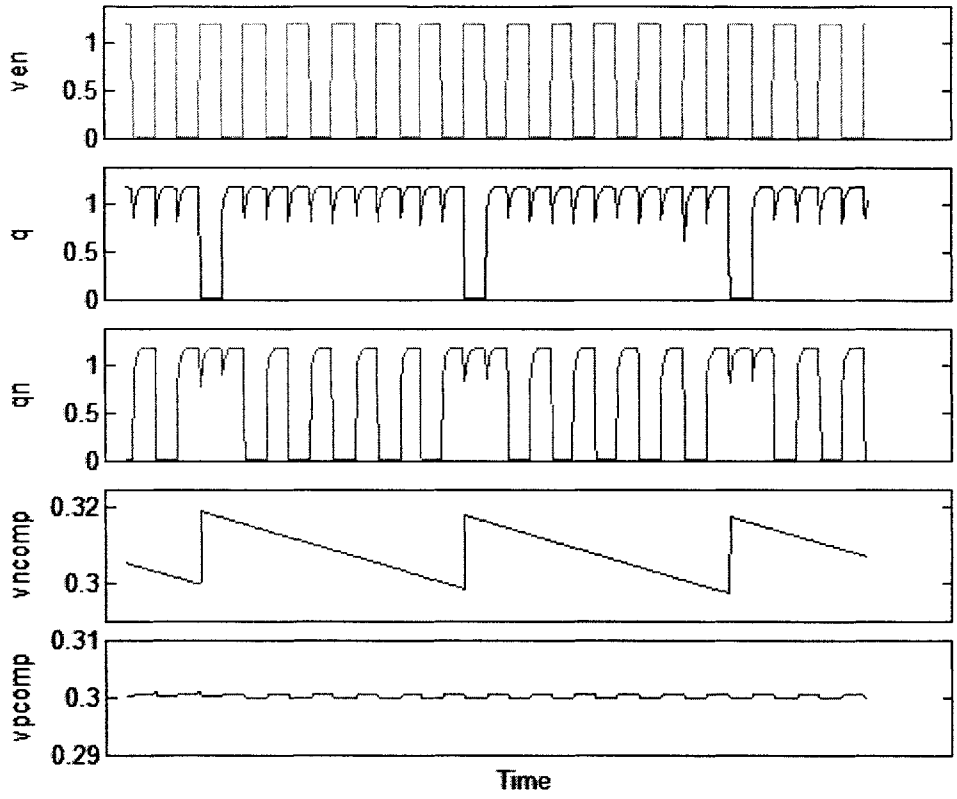

Figure 3-12: Simulated Waveforms showing the reference voltage comparator in operation

3.3.5 Reference Voltage Comparator

The reference voltage comparator as with the other blocks has to be designed so that it consumes no quiescent power in order to maximize efficiency at low load power levels. The comparator topology chosen here was the clocked comparator based on the StrongARM latch architecture. The schematic of the comparator used is shown in Figure 3-11. The reference voltage to the comparator that feeds into the *vpcomp* terminal can be set between 0.3V to 0.7V and is fed by the output of a Digital-to-Analog converter which will be explained in the next chapter.

The operation of the comparator is shown in Figure 3-12. Here the reference voltage which is fed to the *vpcomp* terminal is set to 0.3V. *ven* is the clock that feeds the comparator and is obtained from a divided version of the system clock after level conversion. *vncomp* is the output voltage of the load capacitor. When *ven* goes low, *q* and *qn* go high. At the instance when *ven* goes high, depending on the instantaneous voltage at *vpcomp* and *vncomp*, one of the branches of the comparator sinks more current. This triggers a positive feedback effect and the comparator approaches one of its bistable operating points. If *vpcomp* is high, *q* is low and vice-versa. When *q* goes low, a switching cycle of the DC-DC converter is initiated and a charge quantum is delivered to the load capacitor, thereby raising its voltage above the reference voltage.

3.4 Stability of the DC-DC converter

The DC-DC converter that has been constructed as part of the energy minimization loop functions in the PFM mode of operation all the time. In the PFM mode, the converter operates discontinuously and pulses in inductor current to transfer discrete quantities of charge to the load capacitor. This is shown in Figure 3-13, which shows the inductor current I_{Lf} along with the power transistors' gate voltage and the voltage at the common drain node of the power transistors. The PMOS switch turns ON first and stays ON for a time τ_{PMOS}. Once the PMOS is turned OFF, the NMOS turns ON and stays ON till the inductor current returns to zero. The node V_x stays close to V_{bat} when the PMOS is ON and close to GND when the NMOS is ON. Once, the NMOS is turned OFF, the inductor resonates with the parasitic capacitance at the node V_x and charges it up to the load voltage, such that the average voltage across the inductor in steady state remains

zero. Since, the inductor current I_{Lf} returns to zero at the end of each charging pulse, the inductor's current flow is not continuous, and the two-pole LC filter reduces to a single dominant pole which is set by the load capacitor and the effective load resistance [15]. This makes the loop stable as long as the frequency at which the reference voltage comparator is clocked at is high enough compared to the loop bandwidth across varying V_{dd} and load current regimes.

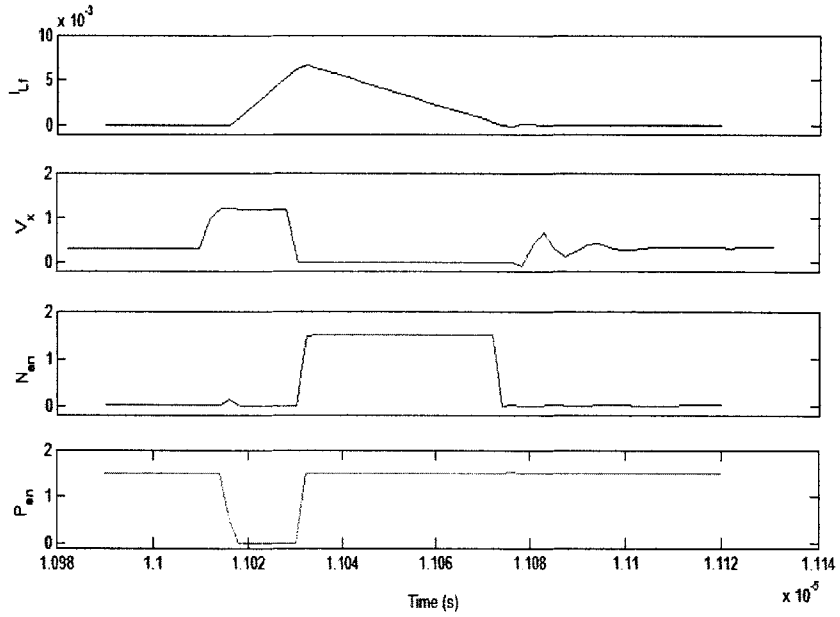

Figure 3-13: Simulated waveforms of the PFM buck converter

3.5 Measured Results

The DC-DC converter was implemented in a 65nm CMOS process as part of the energy minimization loop. All elements of the converter except for the filter inductor and the load capacitor were implemented on-chip. The reference voltage to the comparator is set by the energy minimization loop circuitry as would be explained in the next chapter.

Figure 3-14 shows the output load voltage when the reference voltage is set at 420mV. The characteristic ripple of PFM operation can be seen on the V_{dd} curve.

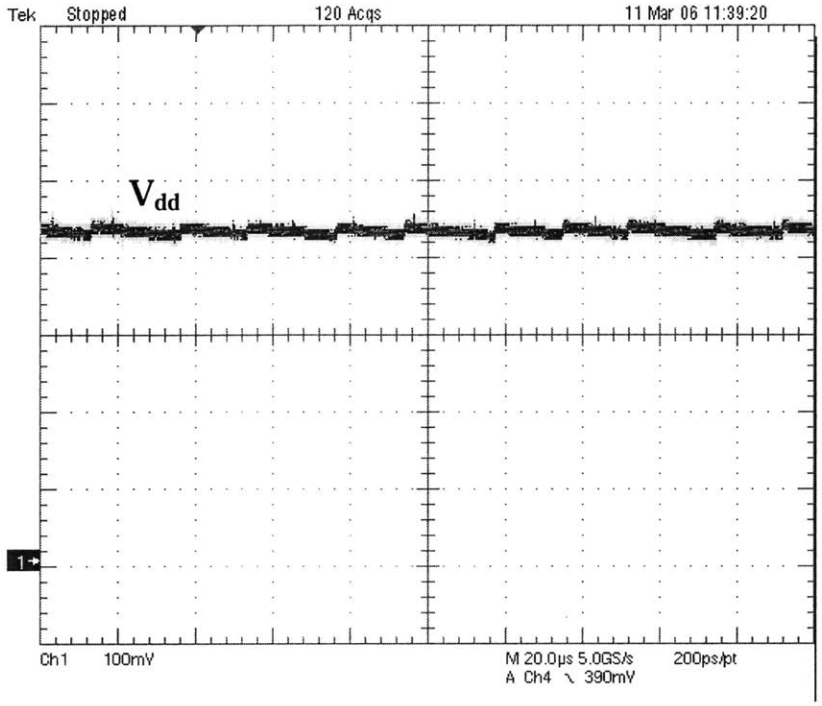

Figure 3-14: Measured load voltage of the PFM buck converter at a reference voltage of 420mV

Figure 3-15 shows a plot of the simulated and measured efficiency of the DC-DC converter with load power. The measured efficiency is not as good as the simulated efficiency curve. The primary reason for this was that the layout of the converter wasn't optimal. This increased the conduction losses in the converter and thereby affected the efficiency. The efficiencies at low load power levels are also pretty low primarily due to the power consumed by the control circuitry blocks. An improved design of the converter with better layout and clock gating of unused control elements was made. Figure 3-16 shows the post-layout, extracted efficiency curve of the improved converter design. We see that we can get better efficiencies even at low load power levels by clock gating the control elements when they are not in use.

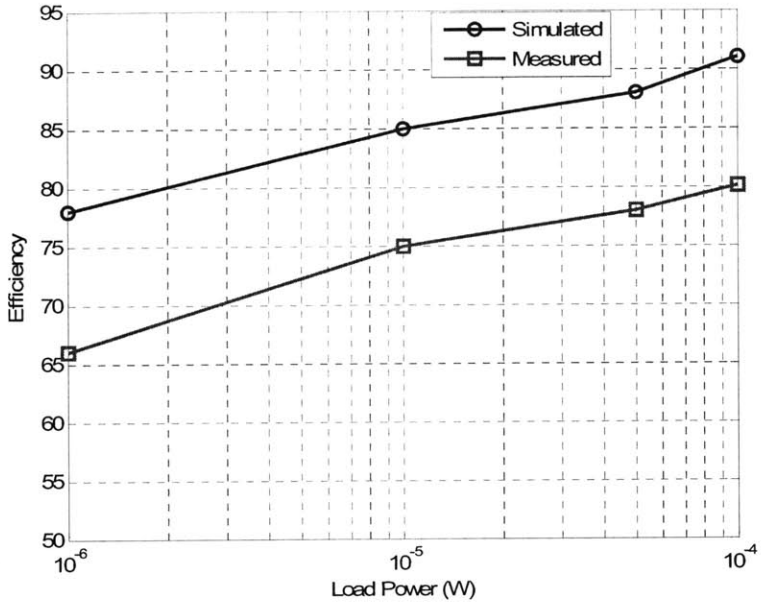

Figure 3-15: Simulated and Measured efficiency plots of the DC-DC converter

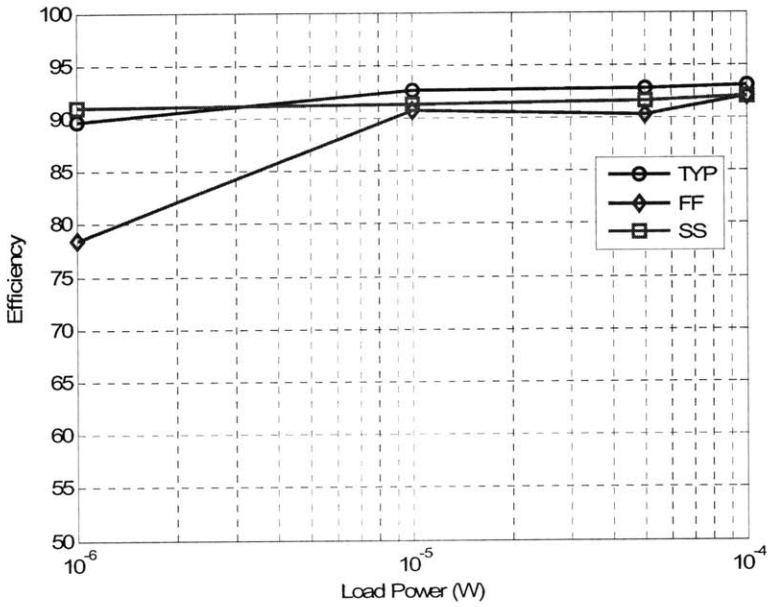

Figure 3-16: Post-layout extracted efficiency plot of an improved DC-DC converter

Chapter 4

Energy Sensing and Minimum Energy Tracking – Principles and Circuitry

This chapter focuses on the principles and circuitry involved in sensing the energy per operation consumed by the digital circuit at a particular operating voltage and tracking the minimum energy operating voltage. The first step in tracking the minimum energy operating voltage is to sense the energy consumed per operation. The energy sensing circuitry should be able to provide a normalized digital representation of the actual energy consumed per operation by the digital circuit. It should also be non-intrusive in the sense that the digital circuit should continue normal operation while the sensing operation is going on. Further, the sensing circuitry should be simple and consume minimal energy in order not to eat away into the energy savings obtained from going subthreshold. The minimum energy tracking circuitry should make use of the E/op. obtained and arrive at the minimum energy operating voltage with very little overhead energy.

4.1 Energy Sensing Technique

There are different options available to sense the Energy consumed per operation. One popular way to measure load power is to measure the voltage drop across some parasitic resistance along the power delivery path. Here, the parasitic resistance of the inductor is usually used to measure the current flowing through it. Some novel techniques also use the package resistance for really high power applications of the order of 10's of watts. The problem with using these techniques to measure load currents of the order of μA's is that owing to the small value of this parasitic resistance, we need power hungry amplifiers to get an accurate value of the current. This is definitely not ideal when we are trying to measure energies of the order of picojoules per operation.

Another way to sense energy is to digitize the voltage across the storage capacitance before and after completing some fixed number *(N)* of operations. The energy per operation can then be obtained by getting an estimate of the energy lost by the capacitor from the end-point voltages. In order to keep the load voltage ripple to an acceptable level, the drop in voltage of the storage capacitor C_{sto} is usually limited to be around 10mV's. This being the case, the analog to digital converter used to digitize the end-point voltages, needs to have a precision close to one mV to accurately get a measure of $(V_1 - V_2)$, where V_1 and V_2 are the voltages across C_{sto} before and after completion of the N operations. Thus from a 1.2V supply, an 11-bit ADC would be required to get this precision. Designing a low power 11-bit ADC is not energy efficient and the power consumption of the ADC would far outweigh the savings from going subthreshold. This theme is repetitive and one of the major challenges with designing a control loop that operates in subthreshold voltages is to have the energy consumption of these control

elements comparable to the energy consumed by the digital circuitry. Circuits that consume no quiescent current are the most suitable for low energy applications.

The method to sense energy adopted by the circuit described below makes use of the availability of the digital value of the reference voltage. Assuming that the voltage across the storage capacitor falls from V_1 to V_2 in the course of N operations, the energy consumed per operation is proportional to $C_{sto} * (V_1^2 - V_2^2)$. In the ADC method mentioned above this value was calculated digitally using the digital equivalents of V_1 and V_2. Using a digital multiplier this value could be squared and subtracted to get a measure of E/op. An elegant and energy efficient way to do this is to observe that V_1 is very close to V_2. Thus, $V_1^2 - V_2^2$ can be simplified as $2V_1 * (V_1 - V_2)$ within an acceptable error. Since, we already have the digital value of V_1 which is nothing but the reference voltage to the DC-DC converter, all we need to find to get an estimate of E/op. is to get the digital value of $V_1 - V_2$. A way to efficiently get $V_1 - V_2$ is described in the following section.

The voltage difference measuring circuitry consists of two 5pF capacitors C_1 and C_2, a fixed frequency clock generator, a counter, current sink and a comparator. The algorithm to estimate $V_1 - V_2$ goes as follows:

i) Before starting the first operation of the N operation cycle, the PFM mode buck converter is disabled. This is done to allow C_{sto} to droop in voltage over the course of N operations.

ii) The voltage across C_{sto} is sampled across the small capacitor C_1. Let us call this voltage V_1.

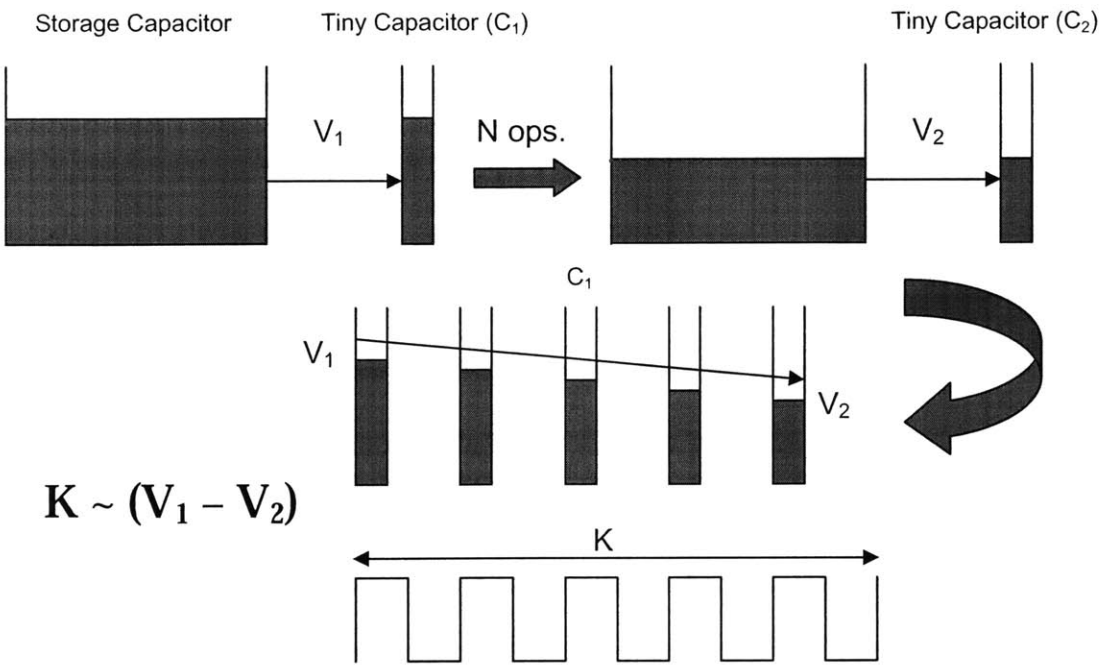

Figure 4-1: Graphic showing how a digital representation of $V_1 - V_2$ is obtained

iii) The digital circuit is allowed to run for N operations and the voltage across C_{sto} droops to some value $V_2 < V_1$.

iv) This voltage V_2 is sampled across the other small capacitor C_2 at the end of the N^{th} cycle.

v) The PFM mode buck converter is enabled and normal operation of the digital circuit continues

vi) Next, a current sink connected across C_1 is turned ON. At the same time, a fixed frequency clock is turned ON and a counter starts counting the number of clock cycles of the fixed frequency clock.

vii) As the voltage across C_1 keeps falling, the counter waits for the comparator to toggle when the voltage falls below V_2.

viii) As the comparator toggles, the count is stored in a register, the comparator, clock and current sink circuitry are switched OFF and remain idle till the next energy sense cycle.

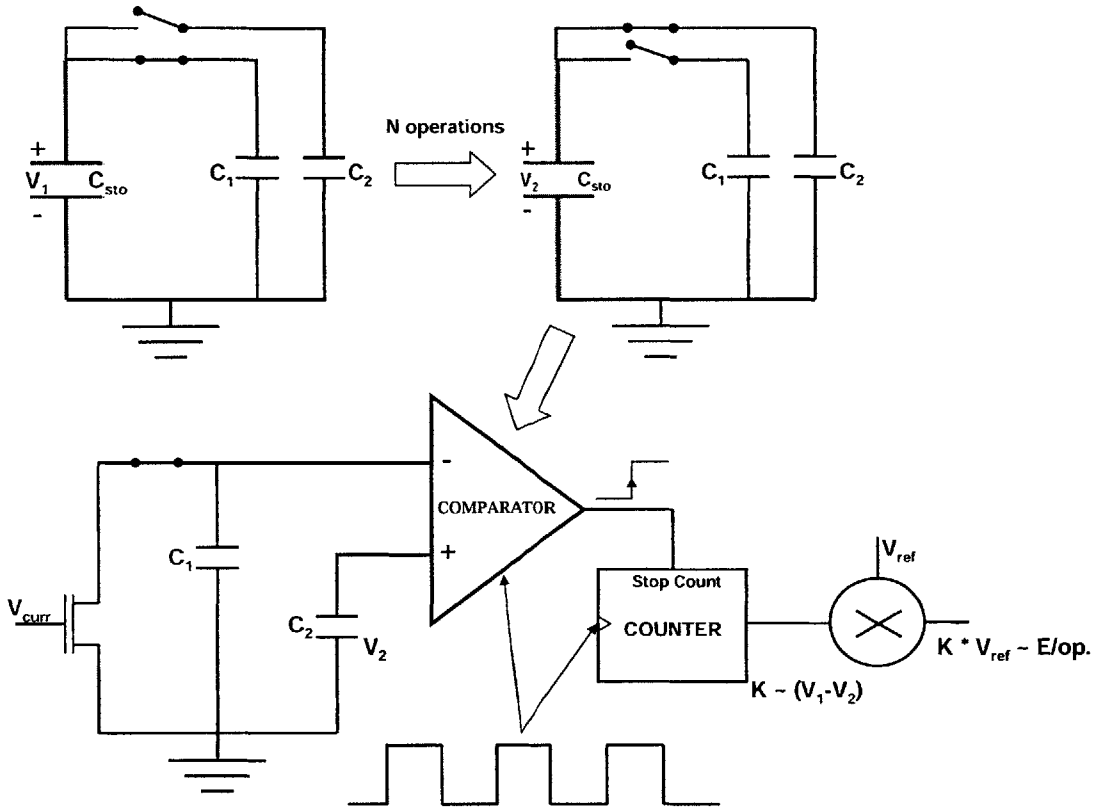

Figure 4-2: Block diagram depicting the way in which E/op. is obtained

The algorithm is shown in a graphical way in Figure 4-1. Once the value of $V_1 - V_2$ is obtained digitally, it is multiplied with V_1 to get an estimate of $V_1^2 - V_2^2$ which is directly proportional to the energy consumed per operation. This proportionality is described in the following equation:

$$E/op. = \frac{1}{2 \times N} C_{sto}\left(V_1^2 - V_2^2\right) = \frac{1}{2 \times N} C_{sto}\left(V_1 + V_2\right)\left(V_1 - V_2\right) \approx \frac{1}{N} C_{sto} V_1\left(V_1 - V_2\right) \propto V_1\left(V_1 - V_2\right) \quad (4.1)$$

where C_{sto} is the storage capacitance that feeds the digital circuit and it is assumed that the digital circuit performed N operations between the two sampling instants. The block diagram of the circuit used to obtain E/op. is shown in Figure 4-2 and the waveforms obtained from actual circuits that are used to measure $V_1 - V_2$ are shown in Figure 4-3. In the simulation, V_1 starts at 300mV and V_2 is at 290mV. The moment $Start_{cnt}$ goes down, $Clock_{cnt}$ is activated and the counter begins counting the number of clock cycles of $Clock_{cnt}$. As long as V_1 is greater than V_2, $Comparator_{out}$ stays low, and the counting continues. When V_1 falls just below V_2 (assuming the offset in the comparator is low), the comparator trips and $Stop_{cnt}$ goes down. The counter stops when $Stop_{cnt}$ goes down and the count registered in the counter at this point in time is proportional to $V_1 - V_2$. This value of the count is multiplied with the digital reference voltage to get a measure of E/op. The comparator and the preamplifiers that are used to calculate E/op. are described below.

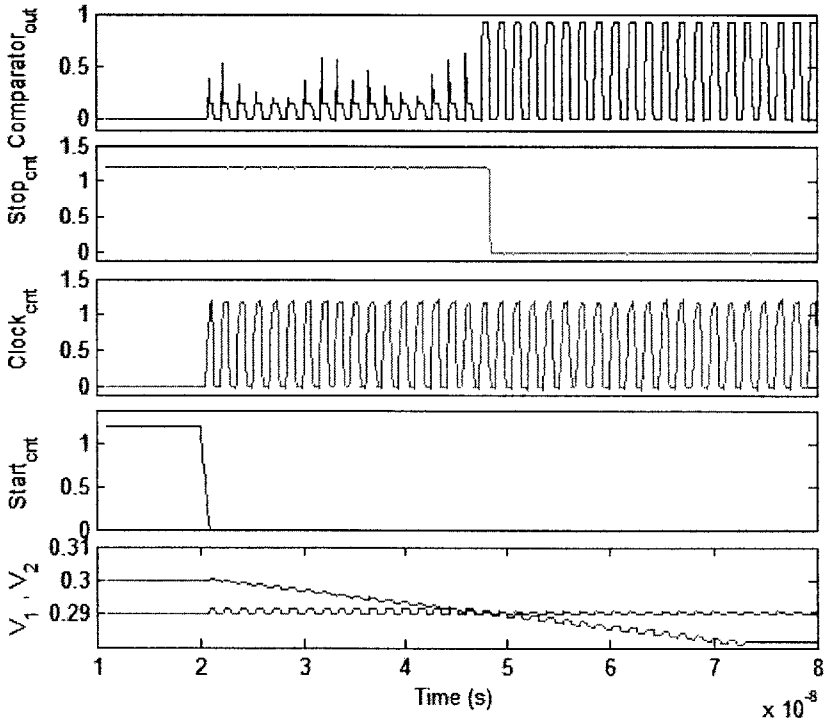

Figure 4-3: Plots showing the calculation of $V_1 - V_2$

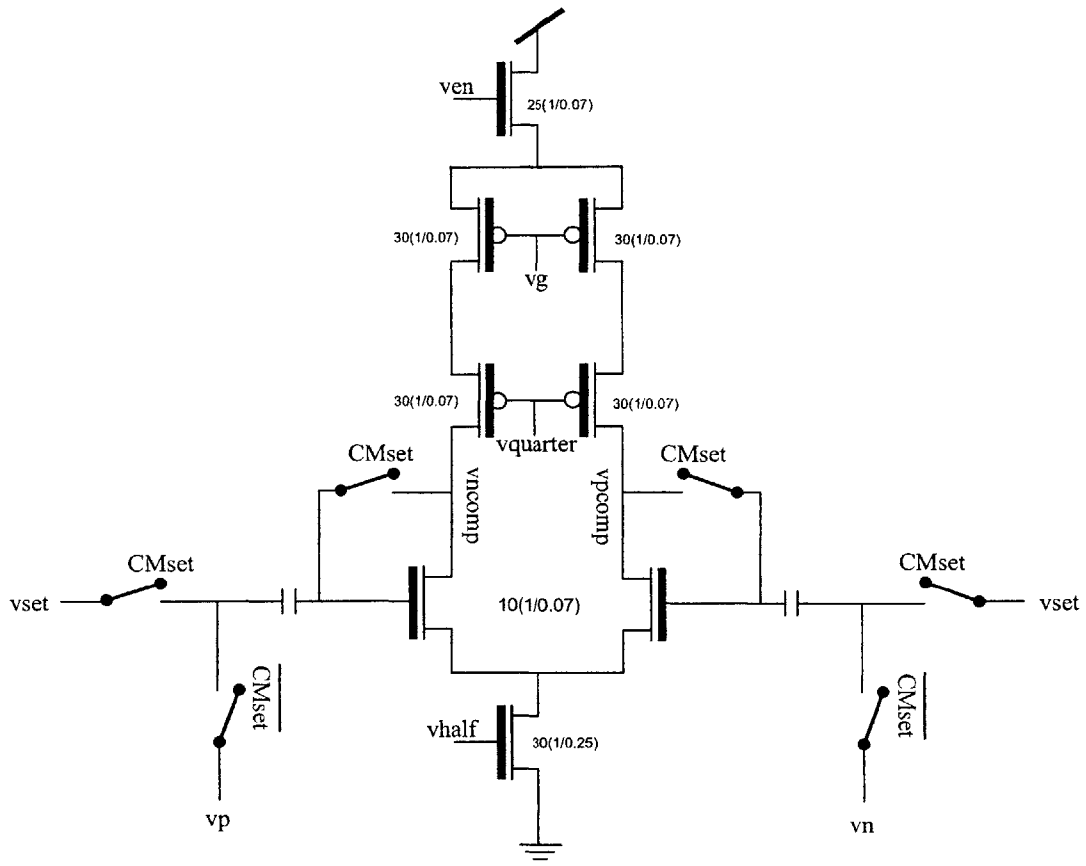

Figure 4-4: Schematic of the pre-amplifier

4.1.1 Pre-amplifier

The comparator that is used to compare the voltages across C_1 and C_2 needs to have a very small offset so that it can calculate an accurate measure of E/op. This necessitates a pre-amplifier to be used in front of the comparator to reduce its offset. The schematic of the pre-amplifier that precedes the comparator is shown in Figure 4-4. The pre-amplifier is a fully differential one that takes as its input, the voltages from the 5pF capacitors C_1 and C_2. It then amplifies the inputs differentially and feeds the outputs to the comparator. The inputs to the pre-amplifier can be anywhere between 0.3V to 0.7V. Thus, it becomes necessary to set the common-mode voltage at the output of the pre-amplifier for effective

comparison down-stream. The switched-capacitor common mode feedback circuit [17] that is used to set the common mode is shown in Figure 4-5. The advantage of using switched-capacitor common feedback circuitry is the absence of large resistors and the additional amplifier that is needed in a continuous time feedback circuit employing resistors. The pre-amplifier is needed only when we calculate the energy/operation and during that time, the common-mode needs to be set only once and not continuously since the voltages do not deviate in a large way. This makes possible the use of switched-capacitor common mode feedback. The common mode is set using *vset* which is the voltage across C_{sto}, just before the N operation cycle begins. Once, the common mode is set, *vp, vn*, the voltages across the small capacitors C_1 and C_2 are amplified and the amplified signals *vpcomp* and *vncomp* are fed to the comparator.

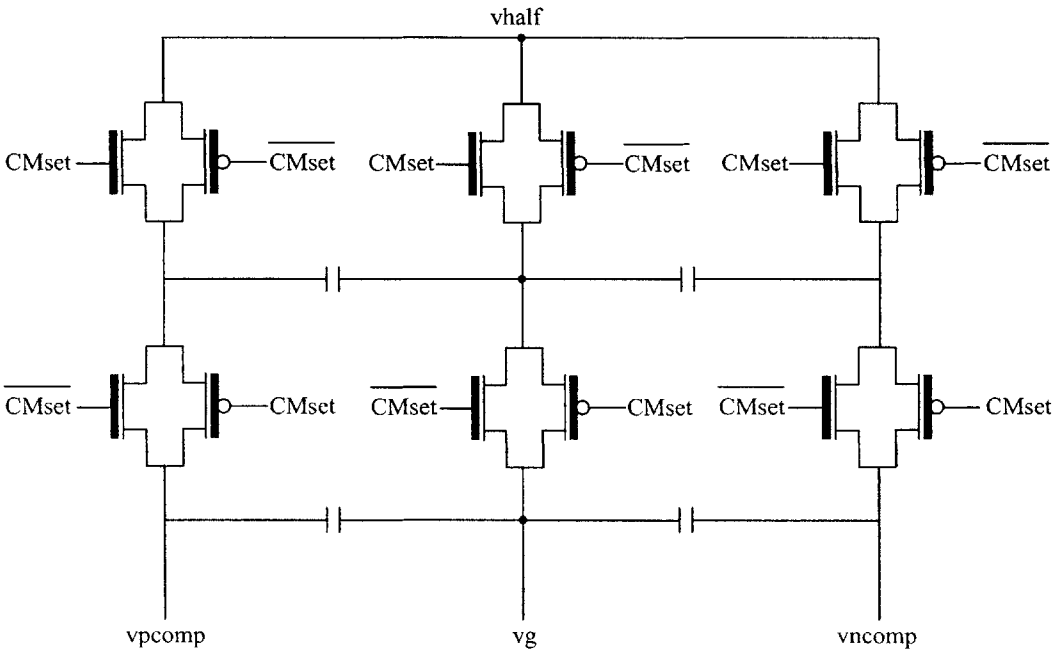

Figure 4-5: Schematic of the switched capacitor common mode feedback circuit

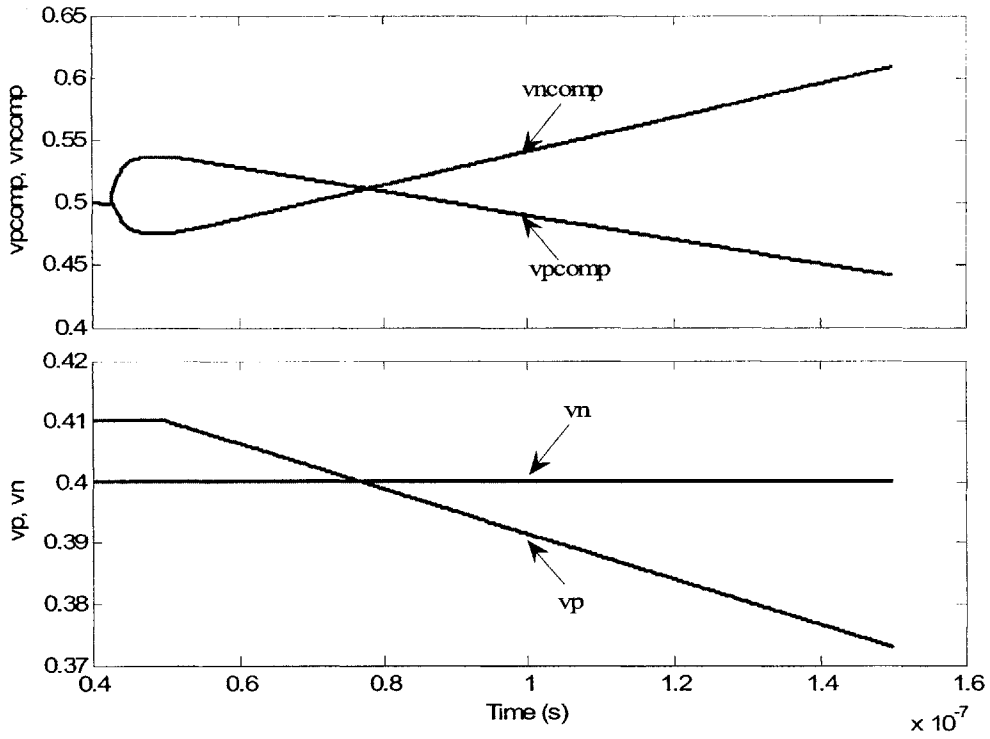

Figure 4-6: Simulated waveform showing the pre-amplifier in operation

In Figure 4-4, the signals *vhalf* and *vquarter* are voltages that correspond to one-half and one-quarter of the nominal battery voltage. These are bias voltages to the pre-amplifier and can be obtained from the battery voltage using voltage dividers. Figure 4-6 shows the pre-amplifier in operation. The common mode is set first using *vset* which is 0.41V in the above figure. *vset* is equal to the value where *vp* starts of. Also, when the pre-amplifier is in operation, *vp* and *vn* do not vary by a large way from *vset*. This enables us to set the common-mode only once using *vset* before amplifier operation starts. The pre-amplifier provides a gain of 6.5. The common mode voltage is set close to 0.5V, and the amplified signals feed the dynamic comparator.

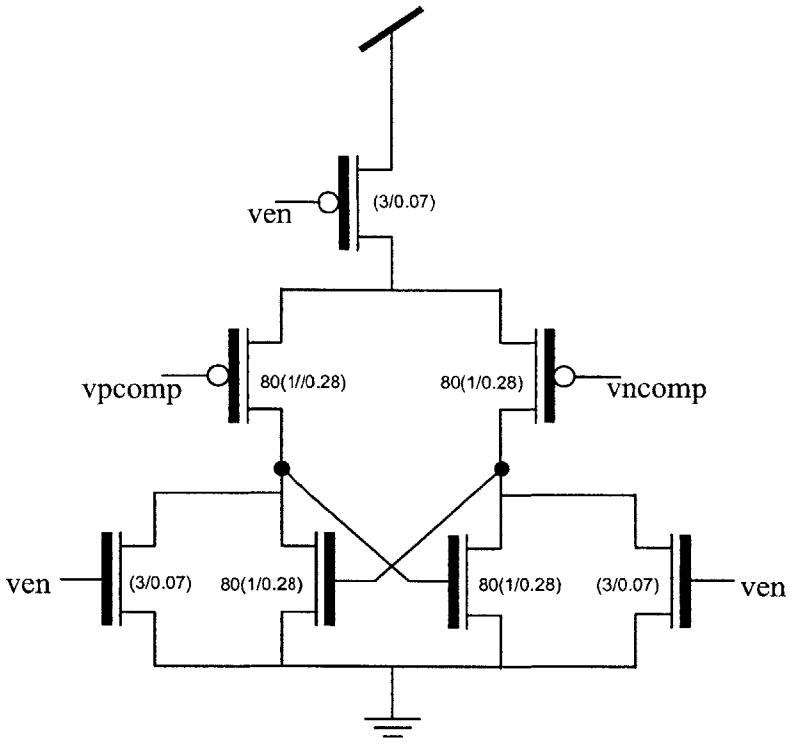

Figure 4-7: Schematic of the dynamic comparator

4.1.2 Dynamic Comparator

The dynamic comparator shown in Figure 4-7 is used to compare the voltages across the small capacitors that have been pre-amplified. When the enable signal *ven* is high, the output nodes are discharged and brought down to ground level. As *ven* goes low, one arm of the comparator which has a lower input voltage draws more current. This is followed by positive feedback kicking in, ramping up the voltage on one side of the comparator to 1.2V while the other arm stays at 0V. The devices in the comparator are sized up in order to keep the offset due to mismatch low [18]. A Monte-Carlo simulation performed on this comparator with 3σ mismatch models for the component transistors shows that the offset of the comparator is 4.4mV. The pre-amplifier described in the previous section provides a gain of 6.5. Thus, the overall offset of the pre-amplifier, comparator system is less than

1mV. Since the comparator is dynamic, it does not consume static current. Also, the comparator and pre-amplifier are turned ON only when the loop needs to calculate $V_1 - V_2$, thereby minimizing the energy overhead of the preamplifier-comparator block.

4.2 Errors in calculating Energy / operation

The approximations introduced in the above algorithm and the non-idealities of the comparator and current sink introduce errors in the calculated energy per operation. The different sources of error are analyzed below and an estimate of the energy difference that can be resolved is provided.

4.2.1 Error due to voltage approximation

The actual value of energy lost in the storage capacitor when the voltage across it falls from V_1 to V_2 is proportional to $V_1^2 - V_2^2$. The error introduced in approximating V_2 to be equal to V_1 is given by:

$$\delta En_{app.} = CV_1(V_1 - V_2) - \frac{1}{2}C(V_1^2 - V_2^2) = \frac{1}{2}C(V_1 - V_2)^2 \qquad (4.2)$$

The relative error introduced in the calculation of the energy is given by:

$$\frac{\delta En_{app.}}{En} = \frac{\frac{1}{2}C(V_1 - V_2)^2}{\frac{1}{2}C(V_1^2 - V_2^2)} = \frac{V_1 - V_2}{V_1 + V_2} \qquad (4.3)$$

The relative error is worst when V_1 is small, i.e. when we are calculating the energy per operation at the lowest voltage at which we allow operation. Since, the lowest voltage the circuit is allowed to operate at is 300mV and assuming V_1 falls from 300mV to 290mV in the course of N operations, the relative error introduced due to approximation is equal to

10 / 590 which is equal to 1.7%. Thus, when we calculate the E/op. at 300mV we overestimate this number by 1.7%. To get an estimate of the minimum energy difference that can be resolved, assuming that E/op. is the same at both 300mV and 400mV. A 10mV drop at 300mV has the same energy consumption as an 8mV drop at 400mV. An 8mV drop at 400mV corresponds to an overestimate of 8/792 which is 1.01%. Thus, 0.69% of the energy consumed at 300mV or 400mV is the minimum that can be resolved correctly. This value is small enough to not considerably affect the accuracy of the minimum energy tracking circuitry, but it has been arrived at by only considering the error introduced due to voltage approximation.

4.2.2 Error due to comparator offset

Another major source of error in the E/op. calculation is the finite offset errors in the dynamic comparator used to compare voltages across the small capacitors C_1 and C_2. The pre-amplifier gain before the comparator is made large enough to minimize the offsets in the comparator. Further, the comparator is made from larger than minimum sized devices to reduce the effect of variations in a MOSFET that decrease proportional to the increase in area of the MOSFET. Even after taking these measures, the offset in the comparator is of the order of one mV. Assuming the offset is ΔV, which may be positive or negative, the error introduced in the measure of E/op. is given by:

$$\delta En_{off} = CV_1(V_1 - V_2) - CV_1(V_1 - V_2 - \Delta V) = CV_1\Delta V \qquad (4.4)$$

The relative error introduced in the calculation of the energy due to offset error in the comparator is given by:

$$\frac{\delta En_{off}}{En} = \frac{CV_1\Delta V}{CV_1(V_1 - V_2)} = \frac{\Delta V}{V_1 - V_2} \tag{4.5}$$

This relative error will be large when we are trying to calculate $V_1 - V_2$, which might be of the order of 10mV. An offset of 1mV leads to a 10% relative error in the calculation of the energy consumed per operation. But this 10% relative error doesn't transform directly into an error when comparing the E/op. at two different voltages. The reason behind this is that, the comparator offset is a static error and hence the same type of error occurs when we calculate the energy per operation at a different voltage. Thus, the offset in voltage creates a minimum difference in energy below which the wrong decision would be taken. This difference in energy can be calculated as follows:

Let us assume that during the course of N operations, the storage capacitor droops from V_1 to V_2 when V_1 is the operating voltage and from V_3 to V_4 when V_3 is the operating voltage. Therefore, the actual difference in energy without offset errors is proportional to $V_1 * (V_1 - V_2) - V_3 * (V_3 - V_4)$. With the offset errors in the comparator, the obtained difference in energy is given by $V_1 * (V_1 - V_2 - \Delta V) - V_3 * (V_3 - V_4 - \Delta V)$. The error in the energy difference is $(V_3 - V_1) * \Delta V$. The relative error in energy difference is given by

$$\frac{\left|\delta En, diff_{off}\right|}{En} = \frac{(V_1 - V_3)\Delta V}{V_1(V_1 - V_2)} \tag{4.6}$$

An estimate of this error can be obtained by plugging in some actual numbers. $(V_1 - V_3)$ is usually of the order of a 100mV, ΔV is 1mV, V_1 can go as low as 300mV and $(V_1 - V_2)$ is around 10mV. This gives a relative error of 3.3%. This sets a lower limit on how fine we can track the minimum energy operating voltage. Between voltages where the difference in energy consumption is lesser than 3.3% of each other, the loop cannot find

the actual minimum. Since, this value is very low, we don't lose much by operating at the non-optimal point in terms of total energy expended.

4.2.3 Error due to leakage in the small capacitors C_1 and C_2

Leakage from the small capacitors C_1 and C_2 can lead to errors in calculating $V_1 - V_2$. When the digital circuit performs the N operations during which time C_{sto} droops from V_1 to V_2, C_1 discharges because of leakage. The situation is worst when the circuit is operating at the lower voltages because the time taken to finish the N operations is longer. But, the leakage current across C_1 also goes down exponentially with the operating voltage and hence, the total charge lost doesn't go up by much. When the operating voltage is 300mV, the leakage current across C_1 which is 5pF is around 20pA. The corresponding time required to finish N operations is close to 200μs. In this time, C_1 droops from 300mV to 299.2mV. Leakage from C_2 isn't a big concern because as soon as V_2 is sampled across C_2, the comparator starts and the comparison process gets over within 100ns. Traditional methods of calibration can be used here, to calibrate out the offset that occurs due to leakage and the overall error can be minimized significantly.

4.2.4 Error due to finite output impedance of current sink

All along we have assumed that the current sink is an ideal one. Any change in the current drawn with voltage would cause errors in the E/op. calculated. Thus, the current sink has to be of high output impedance. The current sink used in the minimum energy tracking circuitry has an output impedance of 11.5MΩ. This is high enough not to introduce any significant errors in the energy calculation. The high output impedance was achieved using a cascode structure. Since, the current that is sunk by the current sink is

small, it is biased in the subthreshold region and hence the cascode structure stays saturated over the operating range of the current sink.

4.2.5 Error due to quantizing the clock cycles

Quantization of clock cycles might also lead to errors. This can be minimized by reducing the current of the current sink and increasing the frequency of the clock. This way the gain of the voltage to number of clock cycles conversion would be high and error due to clock quantization would be minimized.

4.3 Minimum Energy Tracking

Once the energy per operation consumed at a particular operating voltage is calculated, the minimum energy tracking circuitry takes over and moves the operating voltage in the direction of the minimum energy point. The minimum energy tracking circuitry consists of a multiplier, a digital comparator which is nothing more than a subtractor and based on the decision of the comparator, circuitry that adds or subtracts steps of voltage from the reference voltage. The operation of the circuit blocks and the algorithm used to get to the minimum energy operating voltage are described below.

The algorithm for tracking the minimum energy point goes as follows:

(i) The reference voltage is initially set at an arbitrary voltage. This voltage is usually set to 400mV around which the minimum energy operating voltage often falls for normal operating conditions.

(ii) The DC-DC converter when ON brings the operating voltage of the digital circuit to the reference voltage, which is set through a digital-to-analog converter. Now, the energy minimization loop is switched ON.

(iii) The energy per operation consumed at the reference voltage is calculated and this is stored in the minimum energy register.

(iv) The loop then increments the reference voltage by a voltage step which is usually 100mV but can go as low as 50mV for finer tracking.

(v) The DC-DC converter now changes the operating voltage of the digital circuit to the new reference voltage. Once, the voltage settles down near the new reference, the E/op. consumed at this new voltage is calculated.

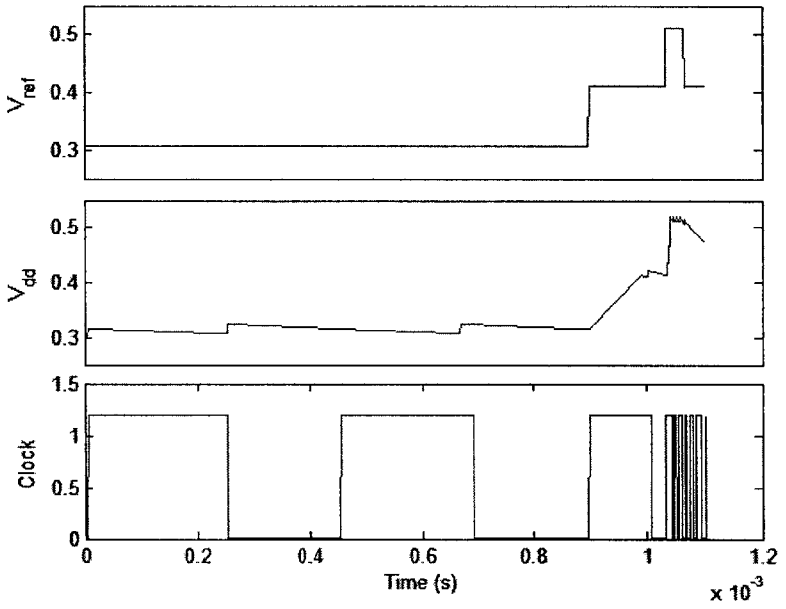

Figure 4-8: Simulation waveform of the minimum energy tracking loop in operation

(vi) This value is compared with the one stored in the minimum energy register. If smaller, this value overwrites the one in the minimum energy register and the loop continues in the same direction. If larger, step viii is executed.

(vii) Repeat steps iv, v and vi till the E/op. calculated is more than the one stored in the minimum energy register. Then, stop the loop, decrement the reference

voltage by one voltage step and shutdown the minimum energy tracking circuitry. Go to step ix.

(viii) Decrement the reference voltage by one voltage step and continue in this direction till the E/op. calculated is more than the one stored in the minimum energy register. Now, stop the loop, increment the reference voltage by one voltage step and shutdown the minimum energy tracking circuitry.

(ix) The final reference voltage obtained at the end of this algorithm is the minimum energy operating voltage.

Figure 4-8 shows the minimum energy tracking algorithm in operation. The loop starts of at an operating voltage of 300mV. The energy consumed per operation is calculated there and stored in the minimum energy register. Then, the reference voltage is incremented by 100mV. The DC-DC converter ramps up V_{dd} to 400mV and after making sure that the operating voltage is settled around 400mV, the energy per operation consumed at this new operating voltage is calculated. This is found to be lower than the value stored in the minimum energy register. Hence, the newly calculated E/op. is stored in the minimum energy register. The algorithm then forces the loop to continue in the same direction and increments the reference voltage by another voltage step. The DC-DC converter now ramps up V_{dd} to 500mV and the energy per operation consumed here is calculated. This is found to be more than the value stored in the minimum energy register. The loop stops here, decrements V_{dd} by one voltage step to get to the minimum energy operating voltage, shuts OFF the energy sensing and tracking circuitry and lays idle till it is called into service again.

The minimum energy tracking loop works by estimating the energy consumed per operation at a particular operating voltage and changing the voltage suitably to minimize the overall energy consumed. Thus, it will work equally well even if the minimum energy operating voltage ends up being above the threshold voltage. The primary concern is how much the voltage across C_{sto} droops to in the course of the N operations. For efficient functioning of the minimum energy loop, it is recommended to keep $V_1 - V_2$ to around 10mV. If the energy consumed per operation for the digital circuit that is being controlled is very large, then we might have to suitably reduce N to keep $V_1 - V_2$ to within an acceptable value. The other option is to increase the size of the storage capacitor C_{sto}.
The circuit blocks which are used to implement the minimum energy tracking algorithm are described below.

4.3.1 Multiplier and Comparator

The multiplier takes as its input the 7-bit digital representation of the reference voltage and the 6-bit digital count output by the counter that counts the number of clock cycles taken to fall from V_1 to V_2 as shown in Figure 4-9. This 7 * 6 multiplier feeds a 13-bit output to the digital comparator (subtractor). The comparator then compares the multiplier output which is an estimate of the E/op. at the current reference voltage with the value stored in the minimum energy register which holds the value of the minimum E/op. so far estimated. The 13-bit subtractor is implemented using ripple carries. The sign bit of the subtractor output is used as the comparator output and a delayed version of the sign bit is fed to a pulse generation circuit that produces a pulse on the rising edge of the sign bit. The delay is introduced to avoid errors due to glitching in the subtractor. This

pulse is used to clock in the current output of the multiplier and this is stored as the new minimum energy estimate.

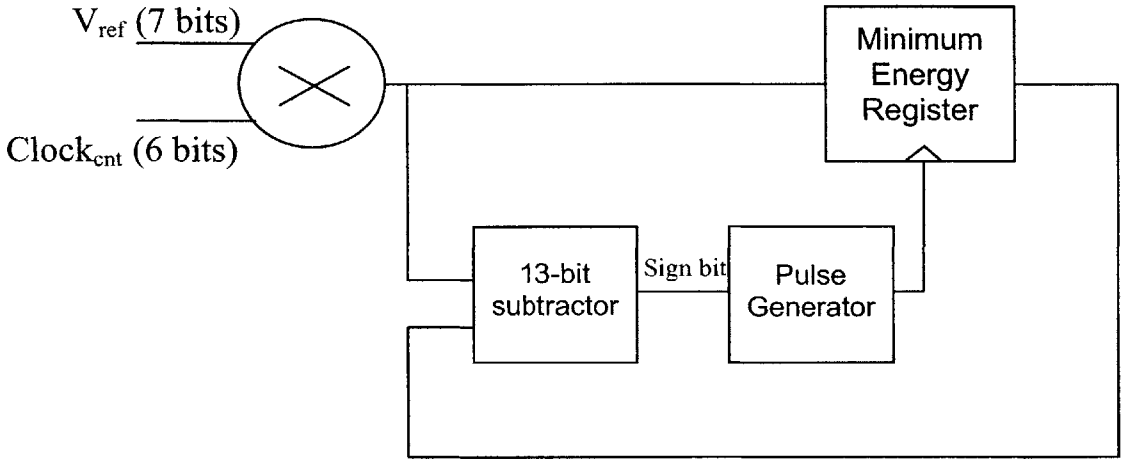

Figure 4-9: Block diagram of the multiplier and comparator circuitry that determine the minimum energy estimate

4.3.2 Reference voltage setting circuitry

The reference voltage setting circuitry consists of the part that adds or subtracts a voltage step to the reference voltage, the register that stores the value of the reference voltage and the digital-to-analog converter (DAC) which feeds the positive terminal of the reference voltage comparator described in section 3.3.5. The block diagram of this part of the circuitry is shown in Figure 4-10. The voltage step is set at the digital equivalent of 100mV. It is the voltage step that determines how fine we can track the minimum energy operating voltage. Setting the voltage step to anything lesser than 50mV doesn't bring any significant advantages in terms of energy savings. Also, setting the voltage step too low might cause the loop to stop at a non-optimal point if the energy difference between the operating voltages is very small. This small difference might be

swamped by the errors introduced in the estimation of E/op. (section 4.2) and cause the loop to misbehave.

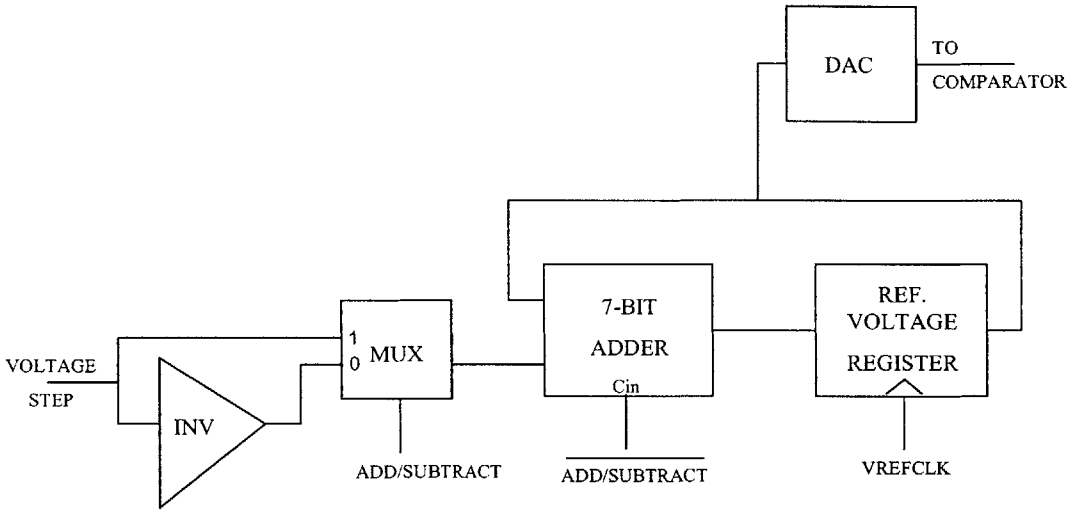

Figure 4-10: Block diagram of the reference voltage setting circuitry

The inverter sub-block is used to create the negative of the voltage step. This when fed to the adder with its carry in set to 1 can be used to subtract one voltage step from the reference voltage. This new added or subtracted version of the reference voltage is clocked in to the reference voltage register by the minimum energy tracking algorithm. This clock can be derived from the same pulse that clocked the minimum energy register shown in Figure 4-9. The ADD/SUBTRACT terminal of the multiplexer is fed by that part of the algorithm which determines the direction in which the loop should be traversed. If the E/op. obtained at the end of the second energy sensing cycle is more than the one stored in the minimum energy register, it indicates that the MEP is lower than or equal to the voltage where the loop started from and hence the loop changes direction. Since, the loop always starts with ADD/SUBTRACT being set high, on changing

direction ADD/SUBTRACT becomes low. A block diagrammatic representation of this circuit is shown in Figure 4-11.

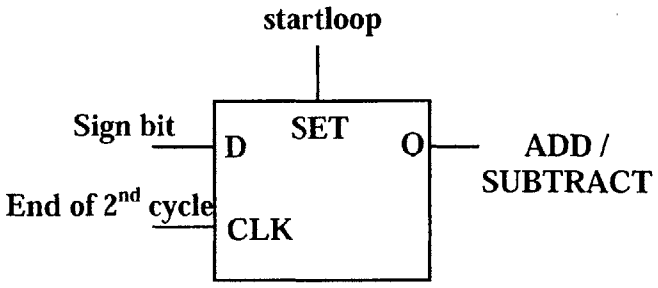

Figure 4-11: Block diagram of the circuit that changes the direction in which the loop is traversed

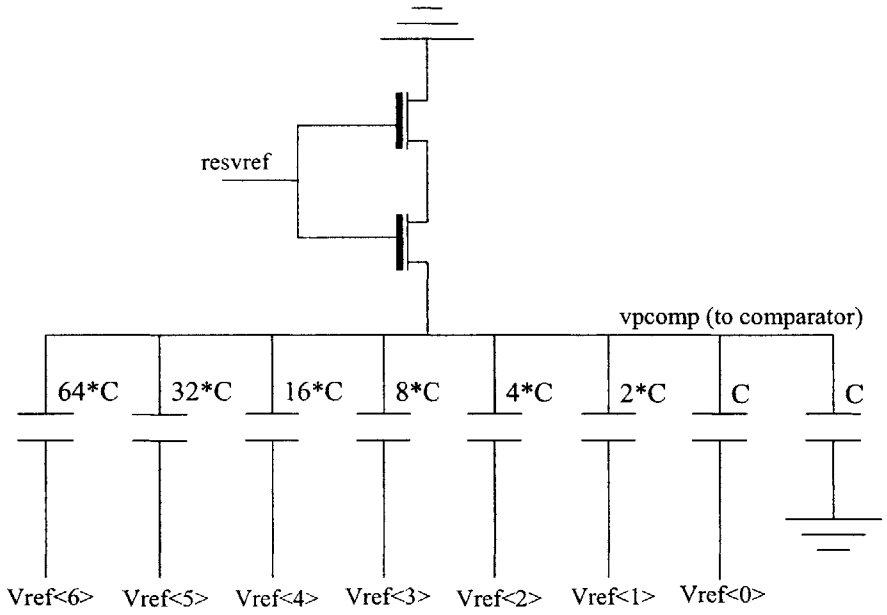

Figure 4-12: Reference Voltage Digital-to-Analog Converter

The digital-to-analog converter used to convert the 7-bit digital reference voltage to an analog value that can be fed to the comparator is shown in Figure 4-12. The DAC should consume very low power since it is always ON when the converter is functioning. This means that its quiescent power consumption should be negligible. The charge redistribution DAC is well suited for such an application. It consumes no static current and very little dynamic current. 7 bits of resolution for the DAC can give reference

voltages spaced at $1.2 / 2^7 \approx 10$mV. Since, no significant advantage can be gained from having the minimum energy operating voltage resolution smaller than 50mV, a 7-bit DAC is a good compromise between finer voltages and ease of implementation. The DAC capacitors' top and bottom plates are grounded before start of operation. This is done by setting *resvref* to go high and setting *Vref<0>* to *Vref<6>* to 0V. Once the capacitors are zeroed, *resvref* is bought down to 0V and normal operation begins. The output of the DAC feeds the positive terminal of the reference voltage comparator. The operation of the DAC is shown in Figure 4-13. The DAC capacitors are zeroed initially and then *vref* is set to '0110101' which corresponds to 494mV of a 1.2V supply. *vref* is then changed to '0100000' which corresponds to 300mV and then to '1000000' which corresponds to 600mV. The change in DAC output voltage with change in *vref* can be seen in the figure. The capacitor sizes are chosen such that the effect of parasitics is minimized while at the same time minimizing the overall area consumed.

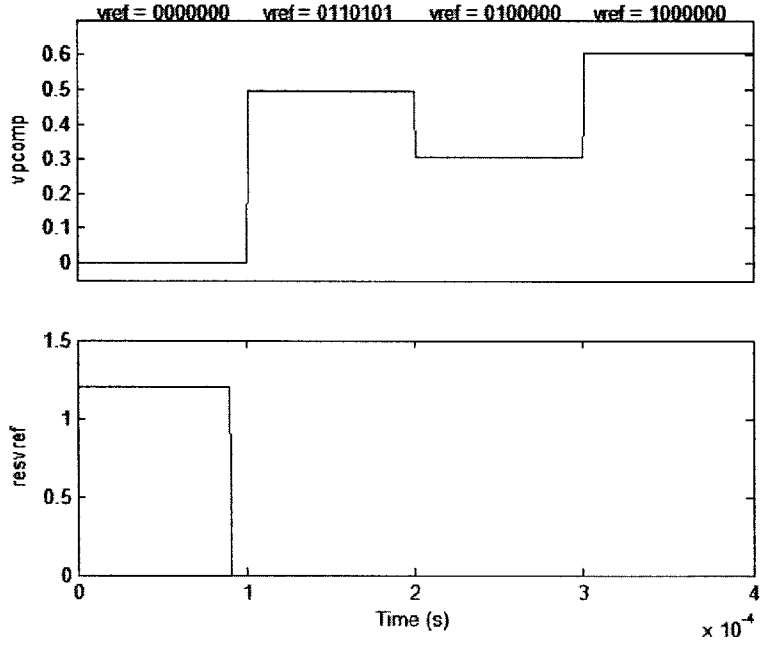

Figure 4-13: Waveforms showing operation of the DAC

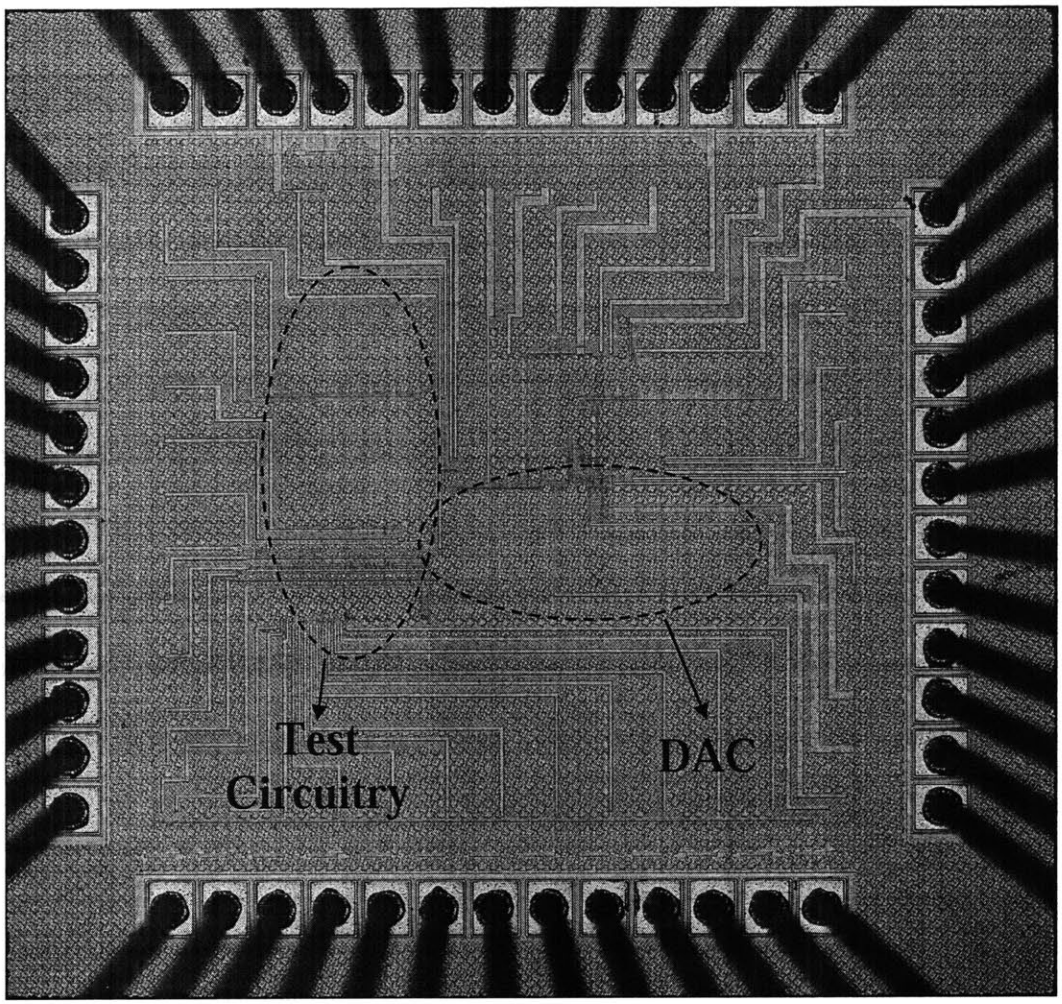

Figure 4-14: Die photo of the minimum energy point tracking test-chip

4.4 Simulation and Measured Waveforms

The minimum energy tracking loop along with the DC-DC converter and the control circuits explained in chapters 3 and 4 was implemented in a 65nm CMOS process. The fabrication was done by Texas Instruments. The chip measured 1.05mm * 1.12 mm in dimension including the pads. The chip was pad-limited with the active circuits consuming just 0.45mm * 0.3mm of the total area. There are 50 pins in the chip for testing the minimum energy loop, the DC-DC converter, providing inputs and reading

outputs of the digital test-circuitry on-chip. Figure 4-14 shows the die photo of the fabricated chip highlighting the portions of the chip occupied by the DAC and the test circuitry. Figures 4-15 and 4-16 show simulated waveforms of the minimum energy tracking loop in operation. The minimum energy point of the digital circuit in this case is around 350mV. In Figure 4-15, the loop starts of at a V_{dd} of 400mV and the voltage step is set to 100mV. The loop then measures the E/op. at 400mV. It then goes on to measure the same at 500mV and finding that the E/op. at 500mV is higher than that at 400mV, changes direction. The loop then measures the E/op. at 300mV. It finds that even this is higher than the E/op. at 400mV and so it settles at 400mV as the minimum energy operating voltage. The actual minimum energy point was at 350mV for the digital circuit at this operating condition. With a voltage step of 100mV and starting of at 400mV, we would never be able to hit the actual minimum.

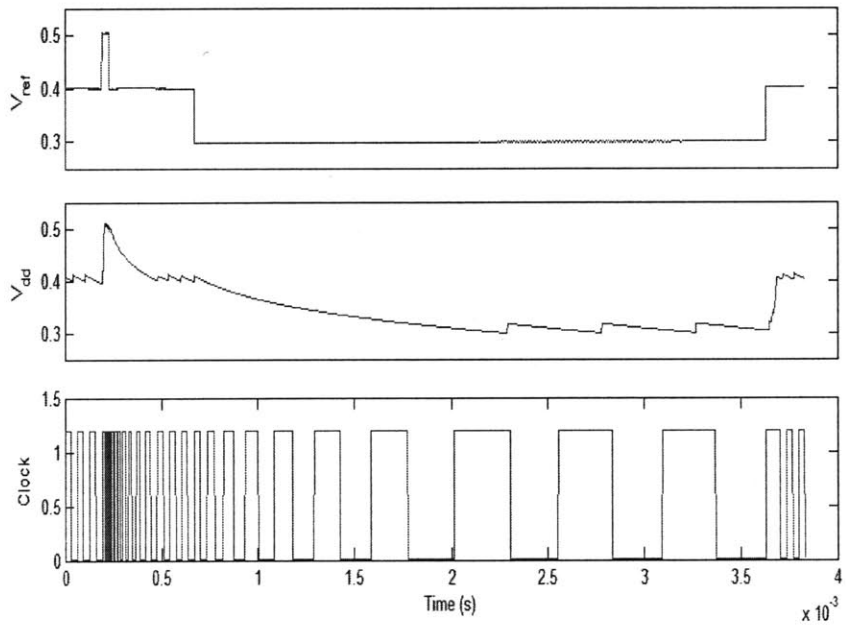

Figure 4-15: Simulation waveform showing the tracking of the MEP. Voltage step = 100mV

Figure 4-16 shows the minimum energy loop in operation for the same operating conditions except that the voltage step has been changed to 50mV. Also, the loop now starts of at 300mV. It calculates the E/op. there and then goes on to calculate E/op. at 350mV. It finds that the E/op. is lower at 350mV and hence the loop continues in the same direction to measure E/op. at 400mV. On finding that the E/op. is higher at 400mV, the loop determines the minimum energy operating voltage as 350mV, sets the reference voltage to the MEP and shuts down. We thus see that by setting the voltage step to 50mV we can hit the actual minimum.

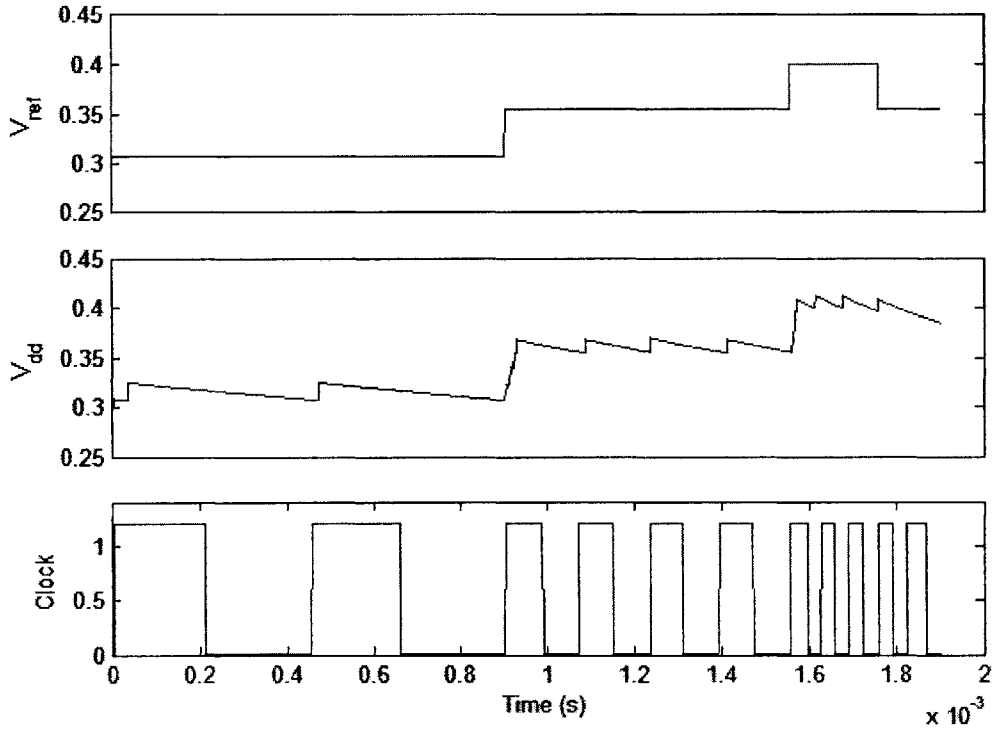

Figure 4-16: Simulation waveform showing the tracking of the MEP. Voltage step = 50mV

The measured waveforms of minimum energy tracking with change in workload are shown in Figures 4-17 and 4-18. In both the figures, the loop starts at 420mV and the voltage step is set to 50mV. In Figure 4-17, the 7-tap FIR filter test circuit is operated

with only one of its taps functional. The other 6 taps are clock gated and do not switch. At this low workload, the MEP of the circuit is at 370mV. In Figure 4-18, all 7 taps of the FIR filter are made operational and the MEP shifts to 320mV. The tracking of the minimum energy operating voltage as the workload changes can be seen from the figures.

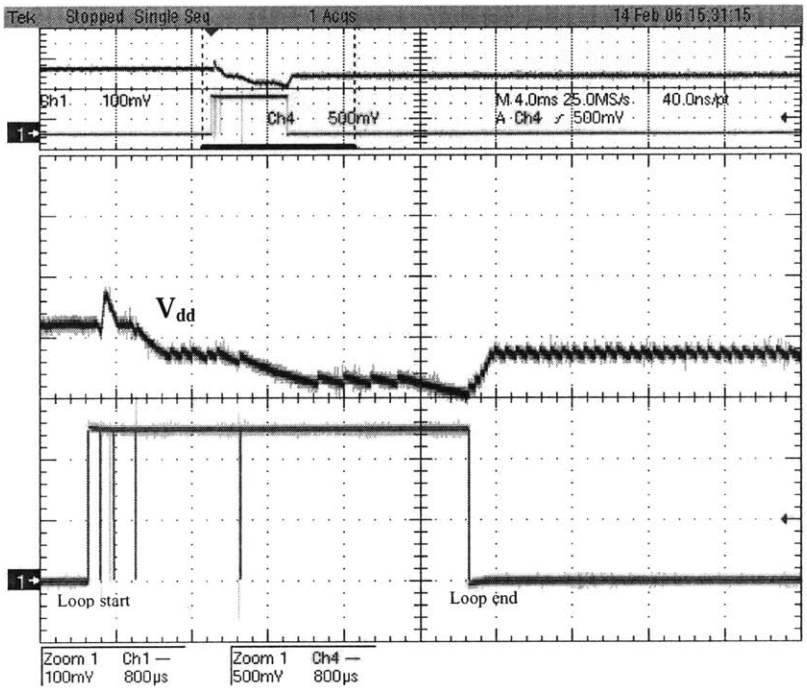

Figure 4-17: Measured waveform showing the tracking of the MEP (370mV) at low workload

Section 2.4 had talked about the change in the MEP with system scenario. In Figure 4-19 an additional 1µA of leakage current is added to the 7-tap FIR filter at full workload (all 7 taps functioning). This is done to model a scenario in which a large digital block with very little switching activity is made to operate along with the FIR filter. In this type of scenario since the total leakage energy per operation goes up, the MEP shifts up and moves to 420mV. The minimum energy tracking circuitry can be seen tracking the shifted MEP.

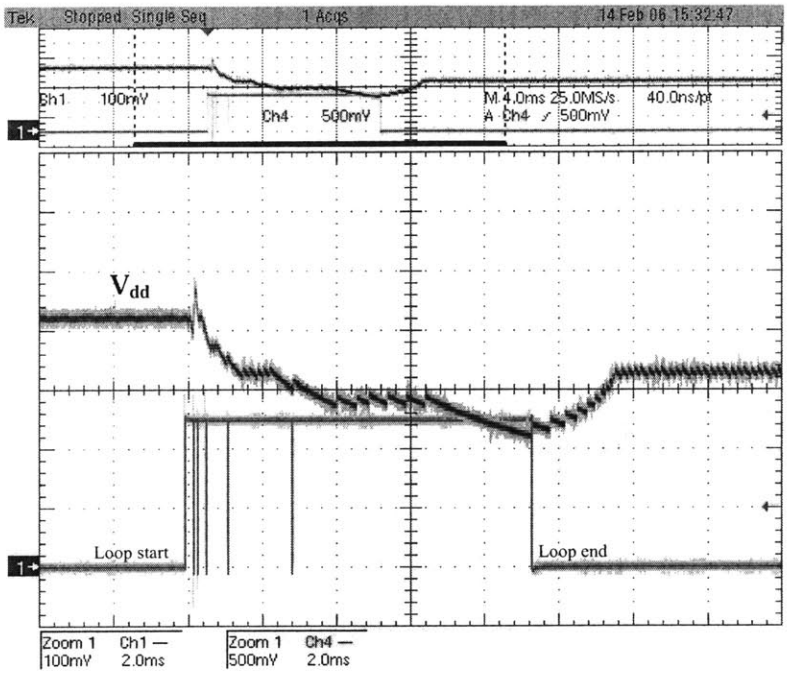

Figure 4-18: Measured waveform showing the tracking of the MEP (320mV) at high workload

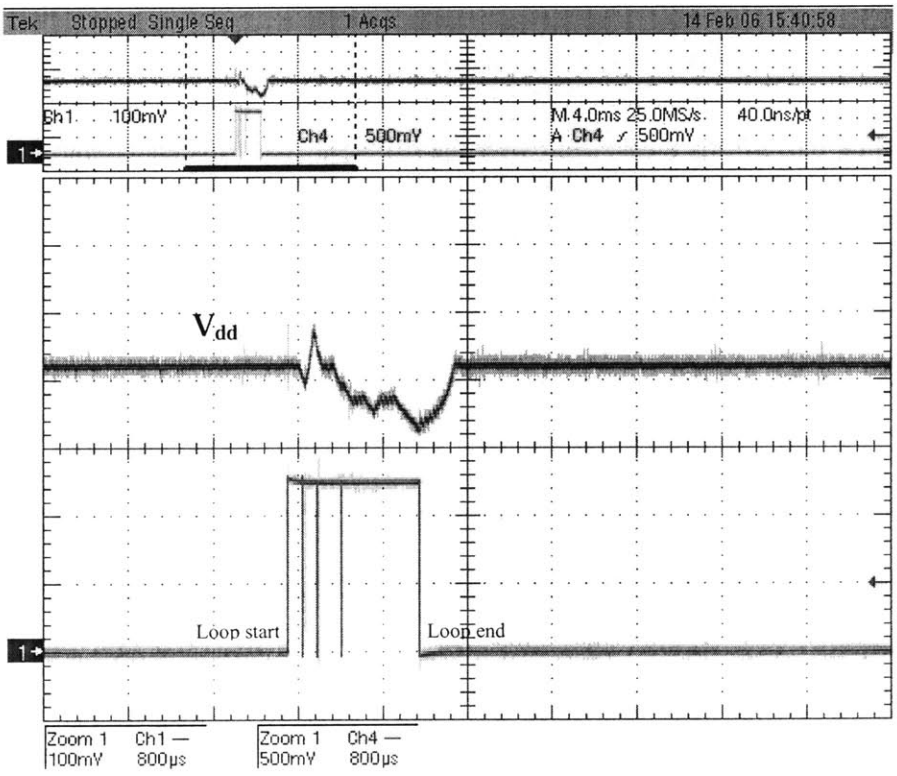

Figure 4-19: Measured waveform showing the tracking of the MEP (420mV) at high leakage

Chapter 5

Conclusion

This chapter summarizes the major contributions and results of this thesis and explores the future work and improvements possible on this current design.

5.1 Summary

The initial part of this thesis dealt with the different scenarios encountered by a digital system and explored how digital operations can be completed with the minimum possible energy for the different scenarios. An analytical expression for the minimum energy operating voltage was derived and the need for a minimum energy tracking loop was motivated. Measured results of the energy savings that can be obtained by continuously tracking the minimum energy operating voltage were presented.

Chapter 3 described the design and implementation of a low power subthreshold DC-DC converter. The various converter topologies were explored and the PFM mode synchronous rectifier buck converter was chosen to deliver power to the on-chip digital circuitry. An open loop approach to set the NMOS power transistor's turn ON pulse

width was described and it enabled to reduce the power consumption of the control circuitry. This is essential to improve the efficiencies at the low load power levels that the converter delivers. A PFM mode synchronous buck converter was implemented in a 65nm CMOS process and it achieved efficiencies > 78% at load powers of the order of 10μW and higher.

The circuitry and principles involved in sensing the energy consumed per operation at a particular operating voltage was described in Chapter 4. The digital representation of the sensed energy per operation was used in a slope tracking algorithm to track the minimum energy operating voltage. An analysis of errors introduced in E/op. sensing and an estimate of the minimum difference in energy levels that can be identified was provided. The minimum energy tracking loop along with the DC-DC converter and test circuitry were implemented in a 65nm CMOS test chip. Chapter 4 concluded with measured results demonstrating the tracking of the minimum energy operating voltage with change in workload of the digital test circuitry.

5.2 Significant contributions

i) The thesis motivated the need for a minimum energy tracking loop and looked at the energy savings that can be obtained by continuous monitoring of the minimum energy operating voltage.

ii) A novel approach to sense the energy consumed per operation of a digital circuit was presented. The approach doesn't need energy hungry amplifiers or high precision analog-to-digital converters. The method is non-intrusive and enabled continuous operation of the digital circuit while E/op. was being sensed.

The operation of the minimum energy tracking principle was demonstrated with the help of a test-chip fabricated in a 65nm CMOS process.

iii) A DC-DC converter design to deliver ultra low-load powers at high efficiencies was presented and a way to seamlessly embed the DC-DC converter into the minimum energy tracking loop was shown. An open-loop approach to set the width of the pulse width of the NMOS power transistor depending on the operating voltage was presented. The operation of the DC-DC converter was demonstrated with the help of the same 65nm test-chip and efficiencies > 65% at 1μW load power and > 78% at 10μW load power levels were obtained.

5.3 Future work

This section examines the possibilities of using the minimum energy tracking principle in performance intensive digital circuits. Further optimizations of the DC-DC converter are explored and ways to extend the output load power delivered by the converter are looked at.

5.3.1 Integrating the minimum energy tracking loop with a dynamic voltage scaling control loop

This thesis focused on how to track the minimum energy operating voltage and presented circuits and control circuitry to do the same. When digital systems work at their performance intensive mode, it was mentioned that dynamic voltage scaling (DVS) can be used to minimize energy. For systems that scale in performance, it would be beneficial from an energy point of view to integrate the minimum energy tracking loop with the well known dynamic voltage scaling loop to extend the operating region of the digital

circuitry to high performance modes. This would involve two control loops on-chip but most of the components of the loops can be shared to minimize area and control overhead.

5.3.2 Extending the load power delivered by the DC-DC converter

The DC-DC converter described in this thesis operates in the PFM mode of operation and can deliver load powers from as low as 1μW to around 200μW. If the performance of the load digital circuitry needs to be extended to the high performance end, then the power delivered by the DC-DC converter would also have to be extended. This would involve operating in the PWM mode for higher load powers. The dual mode DC-DC converter has to have separate modes of control for each mode and more complex power transistor width control.

5.3.3 Improved comparator design

The comparator that is used to measure the voltage difference $V_1 - V_2$, as mentioned earlier needs to be of very little offset. Any increase in the offset of the comparator leads to errors in calculating the energy per operation consumed. The offset of the comparator described in this work was reduced using a pre-amplifier and by sizing up the devices to minimize offset due to variations. The next step in the comparator design would be to improve its offset characteristics by having more sophisticated offset canceling schemes.

Bibliography

[1] A. Sinha and A. Chandrakasan, "Dynamic Power Management in Wireless Sensor Networks," *IEEE Design and Test of Computers*, vol. 18, no. 2, pp. 62-74, March 2001.

[2] B. H. Calhoun et al., "Design Considerations for Ultra-low Energy Wireless Microsensor Nodes," *IEEE Transactions on Computers*, vol. 54, no. 6, pp. 727-740, June 2005.

[3] A. Wang and A. Chandrakasan, "A 180-mV Subthreshold FFT Processor Using a Minimum Energy Design Methodology," *IEEE J. Solid-State Circuits*, vol. 40, no. 1, pp. 310-319, Jan. 2005.

[4] J. Liu and P. H. Chou, "Optimizing Mode Transition Sequences in Idle Intervals for Component-level and System-level Energy Minimization," *ICCAD*, pp. 21-28, 2004.

[5] A. P. Chandrakasan, S. Sheng, and R. W. Brodersen, "Low-power CMOS Digital Design," *IEEE J. Solid-State Circuits*, vol. 27, no. 4, pp. 473-484, April 1992.

[6] R. W. Brodersen, M. Horowitz, D. Markovic, B. Nikolic, V. Stojanovic, "Methods for True Power Minimization," *ICCAD*, pp. 35-42, 2002.

[7] H. Qin, Y. Cao, D. Markovic, A. Vladimirescu, and J. Rabaey, "SRAM Leakage Suppression by Minimizing Standby Supply Voltage," *ISQED*, pp. 55-60, 2004.

[8] E. Vittoz, "Micropower techniques", *Design of VLSI Circuits for Telecommunications and Signal Processing*, J. E. Franca and Y. P. Tsividis Editors, Prentice Hall, 1991

[9] B. H. Calhoun and A. Chandrakasan, "Characterizing and Modeling Minimum Energy Operation for Subthreshold Circuits," *International Symposium on Low Power Electronics and Design*, August 2004.

[10] B. Calhoun, A. Wang, A. Chandrakasan, "Device Sizing for Minimum Energy Operation in Subthreshold Circuits," *IEEE Custom Integrated Circuits Conference,* pp. 95—98, Oct. 2004.

[11] J. Kassakian, M. Schlect, and G. Verghese, "*Principles of Power Electronics*," Norwell, MA: Addison-Wesley, 1991.

[12] A. J. Stratakos, "High-Efficiency Low-Voltage DC-DC Conversion for Portable Applications," University of California, Berkeley, *Ph.D. Thesis*, 1998.

[13] A. P. Dancy, "Power Supplies for Ultra Low Power Applications," Massachusetts Institute of Technology, *M. Eng. Thesis*, August 1996.

[14] J. H. Bretz, "DC-DC converters with high efficiency over wide load ranges," Massachusetts Institute of Technology, *S. M. Thesis*, January 1999.

[15] T. D. Burd, T. A. Pering, A. J. Stratakos, R. W. Brodersen, "A dynamic voltage scaled microprocessor system," *IEEE J. Solid-State Circuits,* vol.35, no.11, pp.1571-1580, Nov. 2000

[16] J. Rabaey, A. Chandrakasan, B. Nikolic, "*Digital Integrated Circuits*," 2nd ed. Upper Saddle River, NJ: Prentice-Hall, 2003.

[17] A. N. Karanicolas, H.-S. Lee, and K. Bacrania, "A 15b 1 Ms/s Digitally Self-Calibrated Pipeline ADC," *IEEE J. Solid-State Circuits*, vol. SC-28, pp. 1207-1215, Dec. 1993.

[18] M. J. M. Pelgrom, A. C. J. Duinmaijer, and A. P. G. Welbers, "Matching properties of MOS transistors," *IEEE J. Solid-State Circuits*, vol. 24, no. 5, pp. 1433-1439, Oct. 1989.