SUBMISSION UNDER REVIEW, PLEASE DO NOT DISTRIBUTE

Field Programmable Multi-Vdd FPGA with Stacked Power Supplies *

Changbo Long, Fei Li, Sasank Reddy and Lei He
Electrical Engineering Department, UCLA

ABSTRACT

Multiple supply voltage levels (multi-Vdd) are effective to reduce power at the system level, and field programmability of multiple Vdd domains is a must for designing a single FPGA chip architecture for a spectrum of applications. This paper develops circuits and architectures with field programmable multiple Vdd domains for system-on-an-FPGA. The sizes and locations of these domains are configurable. The voltage levels of these domains are field programmed by a stacked power supply, which consists of an off-chip pulse width modulation (PWM) buck converter to supply majority of the power and a distributed array of on-chip pulse frequency modulation (PFM) buck converters to obtain supply voltage levels that may be individually programmed during mapping-time and run-time for different regions over the chip. The current is recycled between stacked Vdd domains, and total supply current from outside of the chip is reduced to 1/2 and 1/3 in the two- and three-level stacked power supplies, respectively. The reduced current proportionally decreases the number of power pins and alleviates signal integrity concerns. Experiments show that the proposed stacked power supply has over 90% power efficiency in voltage conversion. Finally, we also discuss CAD needs to map system applications to the proposed multi-Vdd FPGA.

1. INTRODUCTION

Benefiting from the continuous scaling of semiconductor technology, modern FPGAs have multi-million gate capacity and high speed clock frequency approaching 400MHz which makes a *system-on-an-FPGA* become realizable. Multiple soft and application specific processor cores (in addition to hard cores for general-purpose) can be implemented on one single FPGA. Power consumption is a large limitation for such a system-on-an-FPGA, and the existing power minimization techniques for a system-on-a-chip should also be examined for a system-on-an-FPGA. A system-on-a-chip, mostly in the form of ASIC, has multiple cores or modules on a single chip. These multiple cores either communicate with each other through system level pipelines [1] or have little interaction between each other, but each of them has individual performance constraints. The technique of multiple supply voltages has been studied to reduce power consumption of such a system [2–6]. For example, [2,3] applied different voltages to different stages of a system level pipeline and removed the temporary "idleness" in the stage computation due to unbalanced stage delay, but still guaranteed the pipeline latency requirement. [4] studied system-level application partitioning for multiple voltage processor cores and minimized power under quality of service (QoS) constraint for each application. To support multi-Vdd, multiple supply voltages could be generated off-chip and distributed to cores in the chip by multiple sets of Power/Ground (P/G) networks. Alternatively, DC/DC converters can be integrated on chip and convert the input voltage to different voltage levels internally [7–9].

However, there are unique challenges in applying multi-Vdd to a system-on-an-FPGA. First, unlike ASICs where the multi-Vdd layout can be customized for different applications, FPGAs have to use the same fabric and layout to accommodate a spectrum of applications. Previous studies have shown that a fixed multi-Vdd layout does not produce a satisfactory power and performance trade off, and field programmable dual-Vdd [10–12] is needed to obtain a good trade off. While the programmable Vdd in [10–12] is fine-grained, the core level multi-Vdd in [2–6] can be viewed as coarse-grained. Similar to the fine-grained designs, we believe that power domains in coarse-grained multi-Vdd FPGAs should be able to accommodate (i.e., be reconfigurable for) different requirements on voltage levels, domain sizes, and domain locations by various applications in order to achieve the desired power reduction without performance loss. In other words, *field configurable* multi-voltage domains should be developed to reduce the power of a system-on-an-FPGA. Secondly, generating and delivering multiple Vdd levels from off-chip with large amount of current require more power pins, which is not preferred in FPGA because FPGA is often pin limited.

In this paper, we develop FPGA circuits and architectures that support field programmable multi-Vdd for system-on-an-FPGA. The enabling technique is an innovative stacked power supply system, containing an off-chip pulse width modulation (PWM) buck converter to supply majority of the power and a distributed array of on-chip pulse frequency

*This research is partially supported by NSF CCR-0401682 and CCR-0306682 and UC MICRO sponsored by Actel, Intel and Mindspeed.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
Copyright 200X ACM X-XXXXX-XX-X/XX/XX ...$5.00.

modulation (PFM) buck converters to obtain supply voltage levels that may be individually programmed during mapping-time and run-time for different regions over the chip. Power efficiency, defined as the ratio of output power delivered and input power, is over 90% compared to 78% for the charge-recycling linear regulator based stacked power supply that has been studied in the literature [7]. Because the current is recycled between stacked voltage domains, the total supply current from outside of the chip is reduced to 1/2 and 1/3 in the two- and three-level stacked power supplies, respectively. The reduced current decreases the number of power pins. According to [13], 257 out of 684 pins are used as power pins in the BF957 package for Virtex-II. A reduction in supply current roughly decreases the number of power pins linearly and also alleviates signal integrity concerns. For example, the IR drop [14, 15] through P/G network segments reduces proportionally as the supply current decreases. In contrast to fine-grained programmable Vdd in [10–12] with pre-defined Vdd levels and patterns, the proposed FPGA in this paper can configure the Vdd levels and patterns (i.e., Vdd domain locations and sizes) when mapping an application to the FPGA. Also, the proposed FPGA allows dynamic voltage scaling during runtime[1].

The rest of the paper is organized as follows. In Section 2 we introduce circuits and architectures for field programmable mutli-Vdd FPGA. We use the simplest stacked power supply based on a linear regulator to illustrate the concepts. In Section 3 we describe a stacked power supply system that consists of a PWM and multiple PFMs with power efficiency much higher than linear regulator. In Section 5 we discuss the necessity of CAD algorithms to support the design of stacked power supplies, and we conclude the paper in Section 6.

2. FPGA FABRIC FOR PROGRAMMABLE STACKED POWER DOMAINS

2.1 A primitive example of stacked voltage supplies

2.1.1 Circuit design

The first power supply system described which implements stacked voltage domains is a linear regulator. Fig. 1 (a) shows the schematic of a conventional linear regulator. The variable resistor is implemented as a power transistor controlled by a feedback amplifier and the intermediate voltage level $VddL$, which is converted from the original supply voltage $VddH$, is kept almost constant. Although this linear regulator does not require off-chip components, its energy efficiency is limited to $VddL/VddH$, and a significant portion of the total power is wasted on the power transistor.

Recently, charge-recycling voltage regulator has been proposed to improve the energy efficiency of on-chip regulators [7]. The idea is illustrated in Fig. 1 (b). Logic circuits in different Vdd domains can be concatenated in series or stacked between $VddH$ and Gnd, which are provided externally. $VddL$ is an intermediate voltage level due to the stack structure. The top logic operates between $VddH$ and $VddL$, and the bottom logic operates between $VddL$ and Gnd. The lost energy for providing $VddL$ in a conventional linear regulator is now employed in the computation of the top logic. The charge/currents used in the top logic is "recycled" in the computation of the bottom logic. By regulating $VddL$ to a specified voltage level with a push-pull power transistor pair, we can control the effective supply voltages of different power domains. The two power domains in one single stack structure are called *sibling power domains*. To reduce the regulation current i_{reg} drawn from the regulator and therefore ensure a high energy-efficiency, we need to balance the charge demand between the two sibling domains. Specifically, the current i_{top} and i_{bot} should match each other within certain range of runtime variations.

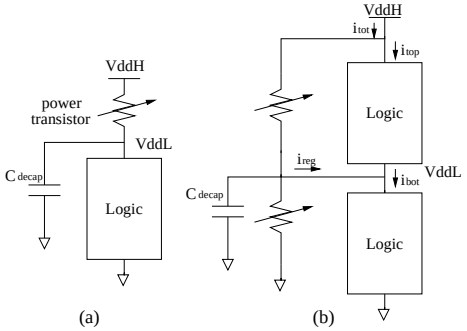

Figure 1: Design concepts: (a) Conventional linear regulator; (b) Charge-recycling voltage regulator.

We use the circuits for a charge-recycling regulator in Fig. 2, which is similar to that [7]. The push-pull output stage is comprised of two power transistors $M1$ and $M2$. Two single stage amplifiers, $Amp1$ and $Amp2$, provide the negative feedback control for the power transistors respectively. The output stage is a pair of source-follower type transistors. The voltage of intermediate node $VddL$ stabilizes at the reference voltage V_{ref}. V_{ref} is on-chip programmable and can be generated by a simple resistance ladder or switch capacitor (SC) voltage divider controlled by two-phase non-overlapping clock signals. An auxiliary decoupling capacitor may be attached to the regulator output to provide additional regulation capability for a very quick change of the charge demand. Overall, this linear regulator has a limited power efficiency, and a more power-efficient power supply will be presented in Section 3.

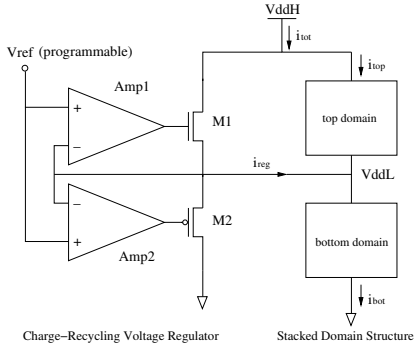

Figure 2: Circuit design for charge-recycling voltage regulator.

[1]Nevertheless, the fine-grained programmable Vdd [10–12] and the coarse-grained one proposed here are orthogonal and can be applied simultaneously.

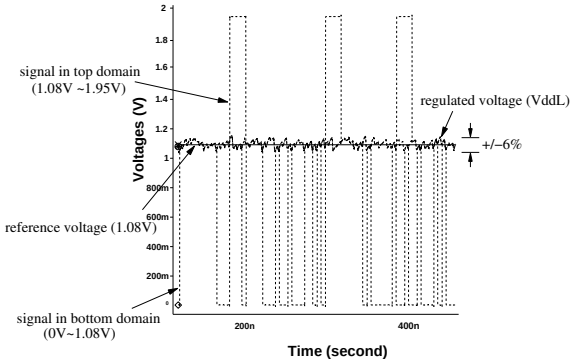

Figure 3: SPICE simulation results for the charge-recycling voltage regulator.

2.1.2 Simulation Results

The functionality of the voltage regulator is validated by SPICE simulation using FPGA circuits. Two MCNC benchmark circuits of similar scales, *b1* and *parity*, are first mapped to a conventional FPGA fabric with lookup table (LUT) size 4 and cluster size 4. LUT size is the number of inputs to an LUT, and cluster size is the number of LUTs inside a logic cluster[2]. We assign circuits *b*1 and *parity* to the top and bottom domains, respectively, to form a stacked structure. The externally generated supply voltage $VddH$ is 1.95V, and the reference voltage V_{ref} is programmed to 1.08V. The sizes of transistors M1 and M2 in Fig. 2 are 3330X and 6660X minimum width, respectively. The two amplifiers $Amp1$ and $Amp2$ operate between 2.0V and 0V (Gnd). We do not attach a decoupling capacitor to the intermediate node $VddL$ in this validation experiment. This allows us to test the regulation capability of the regulator itself.

We apply different input vectors to the top and bottom domains, and the simulation results are shown in Fig. 3. Comparing the waveforms in Fig. 3 and the stand-alone waveforms of the top and bottom domains, it is clear that both power domains function correctly in their own voltage ranges. The operating voltage ranges of the two sibling power domains do not overlap with each other. The output voltage of the regulator is also shown in the figure and it is within +/-6% away from the reference voltage. This amount of voltage oscillation does not cause circuit malfunction and only has a little impact on the circuit delay. The energy efficiency observed in a simulation period of 1320ns is 78%. A simple linear regualtor to produce the same voltage level only has an energy efficiency of 55% (=1.08V/1.95V).

2.2 Inter-domain voltage level shifters

Vdd level conversion circuits are needed as the interface between two voltage domains. Traditional level shifters [17–20] require that the operating voltage range of one domain be covered by the operating range of another, and they do not work for our stacked power domains. We use the sense amplifier based flip-flop (SAFF) [21] as our synchronous level shifter. The circuit is shown in Fig. 4. Signals D and $D\prime$ are the complementary inputs of the level shifter and operate either between the $VddH$ and $VddL$ or between $VddL$ and Gnd. The sense amplifier goes from pre-charge state to the evaluation state at the rising edge of the clock signal clk. The sensing nodes $S\prime$ and $R\prime$ further drive a RS flip-flop to generate rail-to-rail output signals. The full-swing signals Q and $Q\prime$ converted from the level shifter inputs can drive any other power domain, regardless of its operating voltage range. Therefore, the same level shifter can be used when either the top domain drives the bottom domain or vice versa. Fig. 5 shows the SPICE simulation results for the level shifter. We present the waveforms of three signals: level shifter input D, level shifter output Q, and clock signal Clk. We label the operating voltage range beside each signal waveform. The level shifter output signal always makes transitions at the rising edge of the Clk. It is clear that the input signal D is correctly latched into the synchronous level shifter.

[2]Refer to [16] for more detailed descriptions of FPGA architectures.

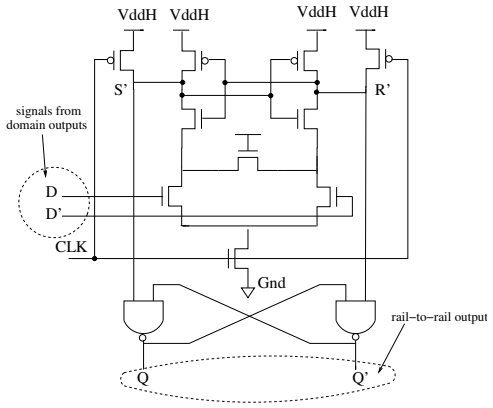

Figure 4: Sense amplifier based level shifter in our stacked power domains.

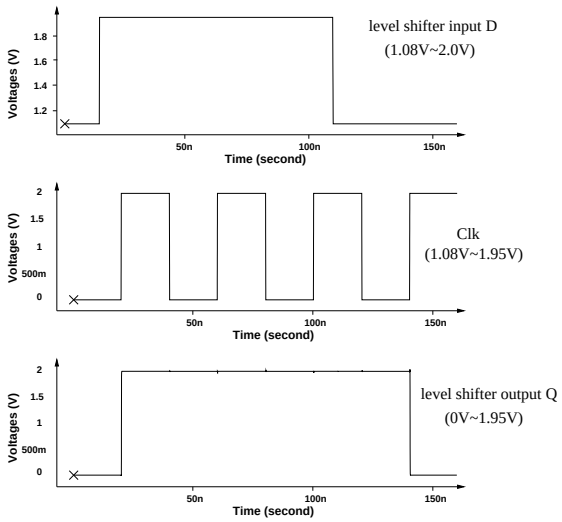

Figure 5: SPICE simulation results for the sense amplifier based level shifter.

2.3 Fabric of configurable stacked power domains

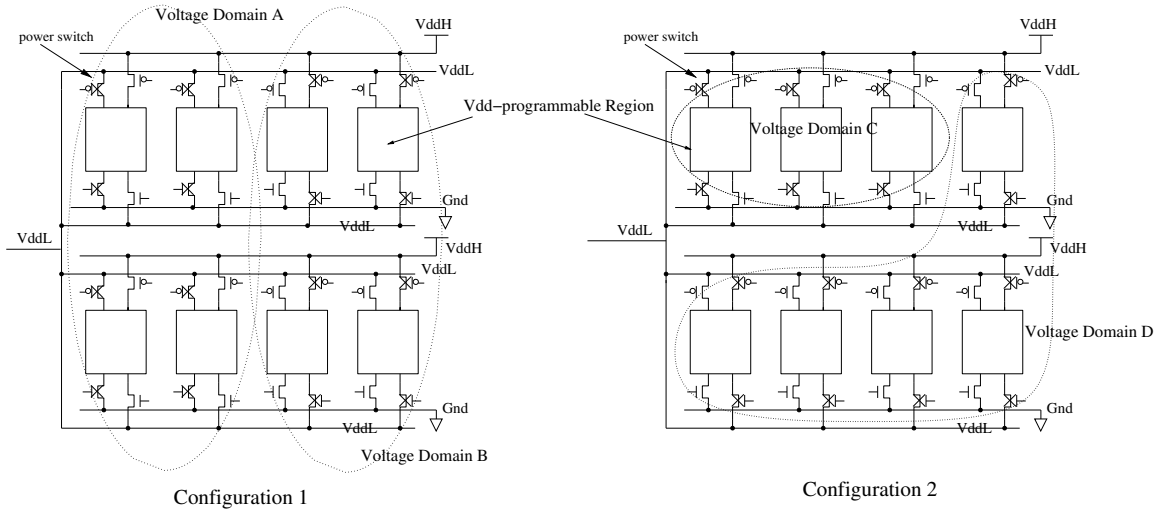

Figure 6: An FPGA fabric with configurable voltage domains. VddH and VddL are generated on-chip and can be programmable.

We show the FPGA fabric to implement configurable stacked power domains in Fig. 6. A Vdd-programmable region is the smallest granularity for coarse-grained Vdd change and contains Vdd-programmable logic blocks and interconnects. In order to make the location and size of a voltage domain software-configurable, power switches are inserted between Vdd-programmable regions and power supplies as shown in Fig. 6. Specifically, we insert PMOS power switches between the Vdd port of a Vdd-programmable region and the supply node $VddH$ as well as between the Vdd port and intermediate node $VddL$. We also insert NMOS power switches between the ground port of a Vdd-programmable region and intermediate node $VddL$ as well as between the ground port and the real ground node Gnd. The power switches are controlled by configuration SRAM cells, and therefore each Vdd-programmable region can be field programmed to operate either between $VddH$ and $VddL$ or between $VddL$ and Gnd. Note that the voltage levels of these regions are programmable on-chip. As shown in Section 3, an array of PFM DC/DC converters will be used to customize Vdd levels for different regions over the chip.

All Vdd-programmable regions configured to the same operating voltage range naturally form a voltage domain and therefore the domain size and location can be easily controlled. For example, in Fig. 6, the same circuit can be configured as two horizontal voltage domains A and B in configuration 1 or two irregular domains C and D in configuration 2. This novel fabric provides similar Vdd programmability originally proposed in the previous work [10–12], but the programmable voltage region is far more coarse-grained (several logic blocks) and the area for power switches is much smaller. Because the logic blocks consist of LUTs and flip-flops, we use the sense amplifier based flip-flop for all the flip-flops in a Vdd-programmable region. The boundary logic blocks in a Vdd-programmable region can be configured as I/O blocks and their flip-flops may serve as the level shifters between different power domains. The sense-amplifier based flip-flops also function correctly as an intra-domain connections. Conventional flip-flops without level shifting function may also be used for intra-domain connections to reduce the circuit area.

3. PWM/PFM STACKED SUPPLIES

Due to the fact that linear regulator based power supply systems have downfalls, especially in terms of power efficiency, we introduce DC/DC converters that are based on inductor-capacitor (LC) low-pass filters, which extract a DC output from an input pulse signal. If the output voltage is less than input, such DC/DC converters are called buck converters. In this section, we first introduce pulse width modulation (PWM) and pulse frequency modulation (PFM) buck converters and then propose a PWM/PFM stacked power supply system based on these two types of converters.

3.1 Pulse width modulation buck converters

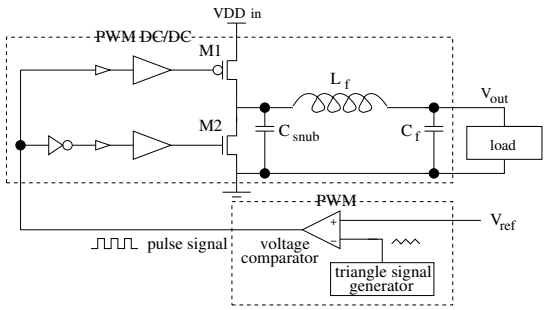

Figure 7: Schematic of a PWM buck converter.

The schematic of a PWM buck converter is shown in Fig. 7. It consists of two power transistors, $M1$ and $M2$, with their drivers, a low-pass LC filter consisting of L_f and C_f, a snubber capacitor C_{snub}, and a pulse width modulator. The basic operation of the PWM can be summarized as follows: The voltage comparator converts a source voltage to a pulse signal with an average magnitude equal to a given reference voltage. The low-pass filter is used to obtain the desired DC component. Furthermore, the power transistors' switching

patterns are used to dictate the duty cycle and in effect the output voltage.

The output voltage level V_{out} is the DC component of the pulse signal generated by the PWM, and it is

$$V_{out} = Vdd_{in} \cdot D, \tag{1}$$

where D is the duty cycle of the pulse signal, which is controlled by V_{ref} as an input of the PWM. In fact,

$$D = \frac{V_{ref}}{Vdd_{in}}. \tag{2}$$

Therefore, we have

$$V_{out} = V_{ref}. \tag{3}$$

As shown in [22], the output voltage ripple of a PWM buck converter can be expressed as

$$\Delta V_{out} = \frac{Vdd_{in}(1-D)}{8L_f C_f f^2}, \tag{4}$$

where L_f and C_f are the inductance and capacitance of the LC filter and f is the frequency of the pulse signal. f is also called the operation frequency of the buck converter.

Equation (4) shows that to keep ΔV_{out} at a low level, L_f and/or C_f has to be large if the operation frequency f is low. In other words, an effective way to reduce the area of the LC filter in the buck converter is to use a high operation frequency [8]. However, a high operation frequency leads to a high switching power loss. To reduce the switching power loss, a technique called zero voltage switch (ZVS) has been widely adopted. As shown in [23], ZVS ensures that both power transistors switch under a zero voltage drop between source and drain.

Assuming ZVS, the conduct loss of power transistors is given by

$$P_{conduct} = \frac{i_{rms}^2 R_0}{W}, \tag{5}$$

and the switching loss is given by

$$P_{switching} = E_0 f_s W, \tag{6}$$

where i_{rms} is the RMS current passing through the power transistor, R_0 is the equivalent series resistance of a unit size transistor in the ohmic region, W is the size of the transistor, E_0 is the energy of one switch for a unit size transistor, and f_s is the operation frequency of the DC/DC converter. The optimum width to minimize the total power loss is given by

$$W_{opt} = \sqrt{\frac{R_0 i_{rms}^2}{f_s E_0}}, \tag{7}$$

and the total power loss could be expressed as

$$P_{min} = 2\sqrt{R_0 i_{rms}^2 f_s E_0}. \tag{8}$$

3.2 Pulse frequency modulation buck converters

The schematic of a synchronous pulse frequency modulation (PFM) buck converter is shown in Fig. 8 [24]. In the PFM buck converter, V_{out} is compared with a reference voltage level V_{ref}. If $V_{out} < V_{ref}$, the output of the voltage comparator opens power transistor M_1 for a short time period T. During this time period, currents are charged into C_f from Vdd_{in} passing through M_1 and L_f. Afterward, M_1 is closed, and the edge detector opens the NMOS

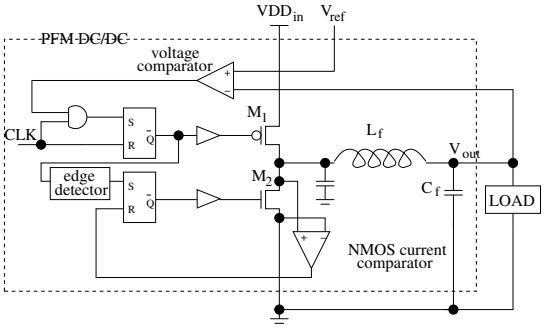

Figure 8: Schematic of a synchronous PFM buck converter.

power transistor M_2 to maintain a continuous current passing through L_f. This current is monitored by the NMOS current comparator, which signals to turn off M_2 once the current decreases to zero. In summary, the PFM buck converter provides a current pulse when V_{out} becomes smaller than V_{ref}.

PFM DC/DC converters have a higher power efficiency than PWM converters when the loading burden is low because power transistors only open when V_{out} drops below V_{ref}, which significantly reduces both conductive and switching power loss. In contrast, the power transistors in a PWM converter always switch and power loss is constant regardless of the loading.

When M_1 is open, the current of L_f is given by,

$$L_f \frac{di}{dt} = Vdd_{in} - V_{out} \tag{9}$$

$$i(t) = \frac{Vdd_{in} - V_{out}}{L_f} t \tag{10}$$

Assume the opening time period of M_1 is T, the peak current of L_f is given by,

$$i(T) = \frac{Vdd_{in} - V_{out}}{L_f} T. \tag{11}$$

Assuming the current of L_f is triangle, we can obtain that this current pulse raises the output voltage by,

$$\Delta v = \frac{Vdd_{in} - V_{out}}{LC \cdot D} T^2 \tag{12}$$

Therefore, the maximum loading power that the DC-DC converter can provide without lowering the output voltage level is

$$P_{max} = \frac{1}{2} \cdot V_{out} \cdot i(T) \tag{13}$$

$$= \frac{1}{2} \frac{(Vdd_{in} - V_{out}) \cdot V_{out}}{L_f} \cdot T \tag{14}$$

According to [24], the switching energy loss for a pulse is equal to

$$E_{switch} = \frac{1}{2} \cdot C_x (Vdd_{in} - V_{out})^2 + \frac{1}{2} \cdot C_x Vdd_{in}^2 + \frac{1}{2} C_x V_{out}^2$$

The conduct power loss of the PMOS power transistors is

$$P_{conduct} = \frac{1}{6} i(T)^2 \cdot (R_{PMOS} + R_{NMOS}) \quad (15)$$

$$= \frac{1}{6} (R_{PMOS} + R_{NMOS}) \cdot (\frac{Vdd_{in} - V_{out}}{L_f} \cdot T)^2 \quad (16)$$

The total energy delivered to the output capacitor is

$$E_{deliver} = \frac{1}{2} i(T) \cdot V_{out} \cdot T \quad (17)$$

Therefore, the power efficiency is

$$\eta = 1 - \frac{E_{switch} + P_{conduct} * T * D}{E_{deliver}} \quad (18)$$

Note that when T is comparable to the delay of the NMOS current comparator, the power efficiency could be significantly degraded. This is because the power of the ground bounce caused by the late close of M_2 and eventually dissipated on the M_2 channel resistor becomes more significant.

3.3 PWM/PFM stacked power supply

3.3.1 Principles

In this paper, we propose a PWM/PFM stacked power supply. The schematic is shown in Fig. 9, where (a) and (b) shows a two- and three-level stacked power supply system, respectively. A two-level stacked power supply consists of an off-chip (in package) PWM DC/DC converter and an on-chip PFM converter. In a three-level stacked power supply, there are two on-chip PFM converters.

The FPGA fabric scheme proposed in Fig. 6 could be fully adopted for the two-level DC power supply. More power switches are needed to support the three-level or more stacking level supplies. In general, an array of PFM DC/DC converters is used to customize Vdd levels for different regions over the chip. These converters are all stacked with the off-chip PWM converter similar to the two- and three-level cases in Fig. 9.

In a stacked power supply, circuit modules (load1, load2, and load3 in Fig. 9) are all stacked and their currents are recycled. In turn, the total current supplied to the chip is reduced. For example, if we assume that the switching currents of $load1$ and $load2$ have the same value of I under a supply voltage level of V. When they are parallely connected in a traditional design, the total chip supply current is $2I$. However, in a stacked power supply system as in Fig. 9 (a), the total chip supply current is only I if $V_{out1} = V$, and $V_{out2} = 2V$. Similarly, a three-level stacked power supply could reduce total supply current by a factor of 3.

In the stacked power supplies, majority of the power is supplied by the off-chip PWM converter, and the on-chip PFM converters are used to control and stabilize the programmable voltage levels. Taking the two-level stacked power supply system as an example, the power supplied by the off-chip PWM and on-chip PFM converter are

$$V_{out2} \cdot I_{load1} \quad (19)$$

and

$$(V_{out2} - V_{out1}) \cdot (I_{load2} - I_{load1}), \quad (20)$$

respectively. When $I_{load2} < I_{load1}$, the power of the PFM converter becomes negative by (20), meaning that there is a current $(I_{load1} - I_{load2})$ charged to the output capacitor C_f of the PFM converter. The consequence is that the voltage level of the output capacitor C_f in the PFM converter, which is the same as the voltage level of $load2$ V_{load2}, raises above V_{ref2}. In turn, the voltage level of $load1$ V_{load1} drops below the desired voltage level $V_{ref1} - V_{ref2}$. This problem can be solved by always using the one with larger average current as $load2$, and adopting a large enough C_f in the PFM converter to tolerate short-time negative current charging.

When $I_{load2} \geq I_{load1}$, the power supplied by the PFM converter is non-negative. Once the output voltage level of the PFM converter drops below V_{ref2}, the PMOS power transistor M_1 opens and a pulse of current charge is injected into C_f in the PFM converter. In this way, the voltage level of $load2$ and $load1$ are maintained at V_{ref2} and $V_{ref1} - V_{ref2}$, respectively.

In the ideal case when $I_{load2} = I_{load1}$, the power supplied by the PFM converter is zero. Practically we choose $load2$ with a current slightly larger than $load1$ to reduce the power drawn from PFM, which helps to reduce the output voltage ripple and area usage of the on-chip PFM converter.

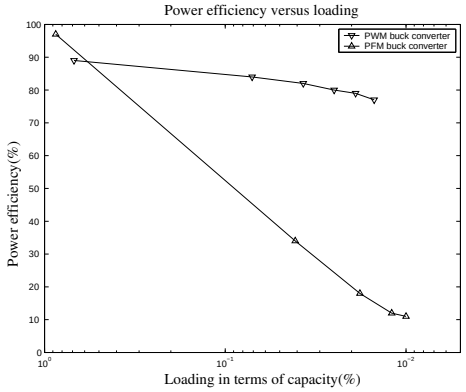

Figure 10: Power efficiency of PWM and PFM buck converters versus loading.

We justify the usage of PFM instead of PWM converters as the on-chip converters in the stacked power supplies by showing the power efficiency of both PWM and PFM converters under different loading levels in Fig. 10. The X axis of Fig. 10 is the loading in terms of capacity where 100% stands for full capacity, and the Y axis is the power efficiency of the PWM and PFM buck converters. As shown in the figure, the power efficiency of PWM converters drops dramatically while the power efficiency of the PFM converters stays high as the loading becomes low. This behavior is due to the fact that the power transistors of PWM converters are always switching regardless of the loading while those of PFM converters only open when the output voltage level drops below the reference voltage level. Note that most power loss of these types of converters comes from the conductive and switching power loss of the power transistors.

3.3.2 Analysis

We first study the output voltage ripple and power efficiency of a two-level stacked power supply and then extend our study to the three-level case.

In a two-level stacked power supply system, the voltage ripple of V_{out2} is mainly determined by the off-chip PWM

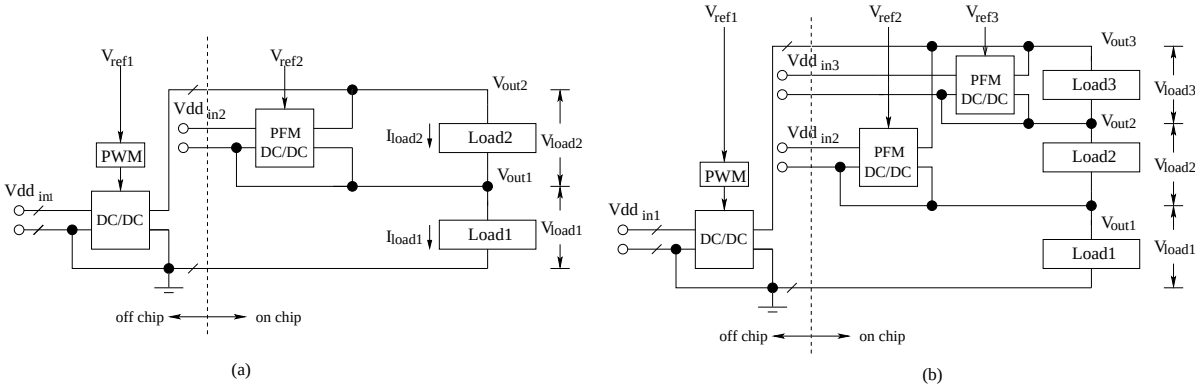

Figure 9: Schematic of PWM/PFM stacked power supply systems.

DC/DC converter because 1) the output capacitance of the PWM is much larger than that of the on-chip PFM DC/DC converter 2) the voltage level of V_{out2} is the same as that of the output capacitance of the PWM converter. I.e, the voltage ripple of V_{out2} is

$$\Delta V_{out2} = \Delta V_{PWM} = \frac{Vdd_{in}(1-D)}{8L_f C_f f^2}. \quad (21)$$

Also, as the voltage level of *load*2 is the voltage level of the output capacitor of the PFM DC/DC converter, the output voltage ripple is the same as the PFM DC/DC converter. I.e., the voltage ripple of *load*2 is

$$\Delta V_{load2} = \Delta V_{PFM} = \frac{Vdd_{in} - V_{out}}{LC \cdot D} T^2. \quad (22)$$

Therefore, the voltage ripple of *load*1, i.e., V_{out1} is given as

$$\Delta V_{load1} = \Delta V_{out1} = \Delta V_{out2} + \Delta V_{load2} = \Delta V_{PWM} + \Delta V_{PFM}. \quad (23)$$

In a two-level stacked power supply system, the power efficiency is a function of the total time that the PFM DC/DC converter is open, which depends on the loading specifications. Assuming that the ratio to open the PFM DC/DC converter is R, the total power loss of the two-level stacked power supply system is

$$P_{loss} = 2\sqrt{R_0 i_{rms}^2 f_s E_0} + (\frac{E_{switch}}{T \cdot D} + P_{conduct}) \cdot R. \quad (24)$$

Similar conclusion can be drawn for three-level stacked power supply system, the voltage ripple of *load*3, *load*2 and *load*1 are

$$\Delta V_{load3} = \Delta V_{PFM1}, \quad (25)$$

$$\Delta V_{load2} = \Delta V_{PFM2} + \Delta V_{PFM1}, \quad (26)$$

and

$$\Delta V_{load1} = \Delta V_{PWM} + \Delta V_{PFM2}, \quad (27)$$

respectively, where $PFM1$ and $PFM2$ stand for the right and left on-chip PFM converters in Fig. 9 (b), respectively. The total power loss of the three-level stacked power supply system is

$$P_{loss} = 2\sqrt{R_0 i_{rms}^2 f_s E_0} + (\frac{E_{switch1}}{T_1 \cdot D_1} + P_{conduct_1}) \cdot R_1 \quad (28)$$

$$+ (\frac{E_{switch2}}{T_2 \cdot D_2} + P_{conduct_2}) \cdot R_2 \quad (29)$$

3.3.3 *Design considerations*

In order to design the PWM to achieve maximum efficiency, several tuning factors including L_f, C_f, C_{snub}, and the size of the power transistors need to be determined so that the circuit operates in ZVS mode. L_f and C_f were tuned to achieve an optimum level for the current across the inductor so that charging and discharging of all capacitances occurs in a lossless manner. The size of C_{snub} was adjusted so that the transition time of the switching is ideal for operating under ZVS mode. Also, the dead times, where neither of the power transistors conduct, must be equal to the time it takes for the inverter output to transition. Finally, the size of the power transistors was manipulated to account for the load so that the voltage drop from Vdd to ground occurred optimally. All these values were tuned while still considering the impact of the changes on the overall area usage.

One of the down sides with the current tuning scheme is that it has to be reconfigured when a different load is used. [23] has introduced a technique to adjust the gate signals of the power transistors so that a proper dead time is insured for optimal ZVS operation for any load. Integrating this technique into stacked power supplies will be our future work.

The design of on-chip PFM buck converters needs to balance maximum loading power capability, output voltage ripple, and efficiency. The key parameter to find the balancing point is the opening time of the PMOS power transistor T. By (14) and (11), the maximum loading power and current of L_f are both proportional to T. According to (12), the output voltage ripple is proportional to T^2. It is also shown that the efficiency of the PFM increases as T increases.

To obtain high efficiency and high loading power capability, a larger T is used. This requires a large L_f and C_f to limit the output voltage ripple to a low level. As shown in Section 4, this case puts the inductance in package due to the area footprint. On the other hand, to integrate the inductance on-chip while keeping output voltage ripple low, a small value T is adopted and power efficiency is sacrificed. As shown in 4, this scheme is suitable for the situation where the loading current is more balanced among voltage domains and the power drawn from the PFM converters is small.

4. EXPERIMENT RESULTS

4.1 Vdd programmability

We have implemented prototype designs for two-level and three-level PWM/PFM stacked power supplies in 100nm technology and simulated our designs by SPICE under Berkeley predictive technology models [25]. To obtain realistic loading circuits for demonstrating the functionality and performance of the proposed stacked power supplies, we use macro-models. We first map three MCNC benchmark circuits of similar scale $b1$, $cm138a$, and $parity$ to a conventional FPGA fabric with LUT size 4. We then simulate these circuits by SPICE and obtain the current waveforms drawn by these circuits from an ideal voltage source. These current waveforms are amplified 1000 times to mimic a real total chip current drawn by a typical FPGA chip[3]. To consider the factor that the total chip current usually has much smaller fluctuation than small circuits, we average the current value every 10ns. In other words, every 10ns we obtain a data point which is the average current value of the original current waveform, and then link all these data points together to build a new PWL current source. These current sources are used as the loading in our experiments.

In the experiments for two-level stacked power supplies, we target the voltage level of V_{out2} at 2.3 volt and V_{out1} at 1.0 volt (refer to Fig. 9 (a).). For three-level, we target the voltage level of V_{out3}, V_{out2} and V_{out1} at 2.7, 1.9 and 0.9 volt, respectively. The output voltage waveform from the SPICE simulation of our prototype designs are shown in Fig. 11. It can be seen clearly that after certain setup time ($8u$ for two-level and $4u$ for three-level), the output voltage are stabilized at the desired levels. The overall voltage ripple for all these output voltages is less than 15%.

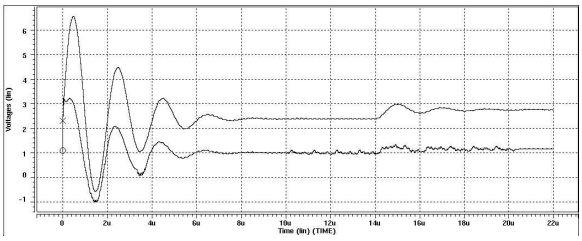

Figure 13: Dynamic voltage scaling from (1.0V, 2.3V) to (1.2V, 2.7V) in the two-level stacked power supply.

The output voltage levels, such as V_{out1} and V_{out2} can be dynamically adjusted during run-time by changing the reference voltage levels. Fig. 13 shows an example. The reference voltage levels of V_{ref1} and V_{ref2} are initially set at 2.3 and 1.3 volt, respectively. Under these reference voltage levels, V_{out1} and V_{out2} are programmed at 1.0 and 2.3 volt, respectively. At $14us$, V_{ref2} and V_{ref1} are reset to be 2.7 and 1.5 volt, which re-programs V_{out2} and V_{out1} to 2.7 and 1.2 volt, respectively. It is clearly shown in the figure that after a $6u$ transition time, these two voltages settle down to the desired levels.

4.2 Power efficiency versus area

One of the main design considerations when creating a power supply system is the area associated with such a design. In our experiment, we estimate the chip area by the total current of the circuit divided by a current density of $1A/cm^2$, representing the typical cases of FPGA chips. For on-chip capacitance and inductance densities, we use $7.94nF/mm^2$ [8] and $0.33nH/mm^2$, respectively. Note that the on-chip inductance area is evaluated based on thin film technology [9]. Also, technology has been developed to place inductor for DC/DC converters both inside and outside the CPU die [9] and it can be applied to this paper too. Table 1 shows the area for two- and three-level stacked power supplies. With inductance (L_f in the PFM converters) off-chip (in package), the introduction of two-level and three-level stacked power supplies can be achieved with an area for capacitance that is less than 10% and 20%, respectively. Furthermore, with the introduction of inductance on-chip, the area is slightly higher at levels of 15% and 35% for the two and three voltage domains.

Power supply efficiency plays a critical role in evaluating a power delivery system. From our experiments as shown in Fig. 12, we were able to achieve power efficiencies consistently over 90% for both the two-level and three-level stacked power supplies with inductance placed in package. When the inductors are placed on-chip, the inductance of the inductors has to be small due to the area of the on-chip inductors. As shown in Section 3.2, a small inductance leads to a small opening time of the PMOS power transistor M_1 to limit the output voltage ripple and therefore a lower power efficiency due to relatively more power loss during the ground bounce caused by late close of power transistor M_2. Because the power efficiency of PFM converters with on-chip inductance is lower, the total efficiency of stacked power supplies with on-chip inductance is lower than that with off-chip inductance. This trend is clearly shown in Fig. 12 especially when the currents among loading are unbalanced. As shown by (20), the power supplied by PFM converters are increased when the loading current has more unbalance. In conclusion, we obtain a high power conversion efficiency of over 90% for both two-level and three-level stacked power supplies when inductors are placed off-chip, and when the loading current are balanced with inductors placed on-chip. In contrast, the power efficiency of the charge-recycling linear regulator example in Section 2 has a power efficiency of 78%.

4.3 Trade-off of stacking levels

Adding additional stacking levels does not come without trade offs. To demonstrate the trade offs, we compare the metrics of two- and three-level stacked power supplies in Table 1. Two PFM converters are placed on-chip in three-level stacked power supplies, which by default doubles the area usage of three-level stacked power supplies as compared to two-level ones. However, to keep the voltage ripples less than 15%, 35% of the chip area is consumed when the inductors of PFM converters are placed on-chip. Note that the additional area usage helps to improve the average power efficiency by 4.6% (83.8% versus 79.2%) as shown in the table. While area usage is increased with additional stacking levels, it provides more voltage level options and reduces the total supply current. As shown in Table 1, ideally three-level stacking would reduce the total supply current by 2/3 while two-level stacking reduce the supply current by 1/2. Also, the reduced supply current decreases the power pin number proportionally and alleviates signal integrity concerns such

[3]It takes hours to simulate these "simple" circuits and it is unrealistic to simulate 1000x larger circuits in SPICE.

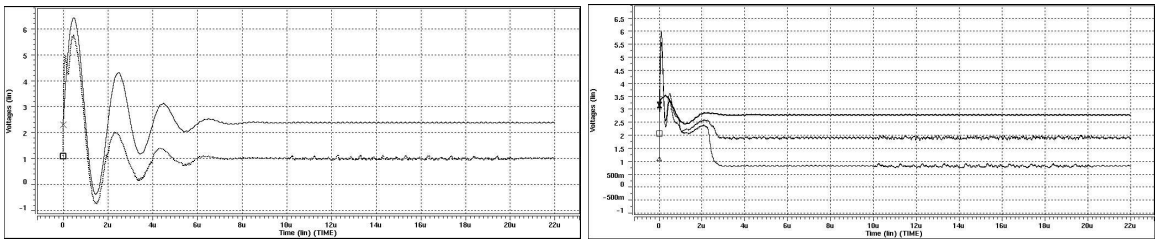

Figure 11: Output voltage levels for two- and three-level stacked power supplies.

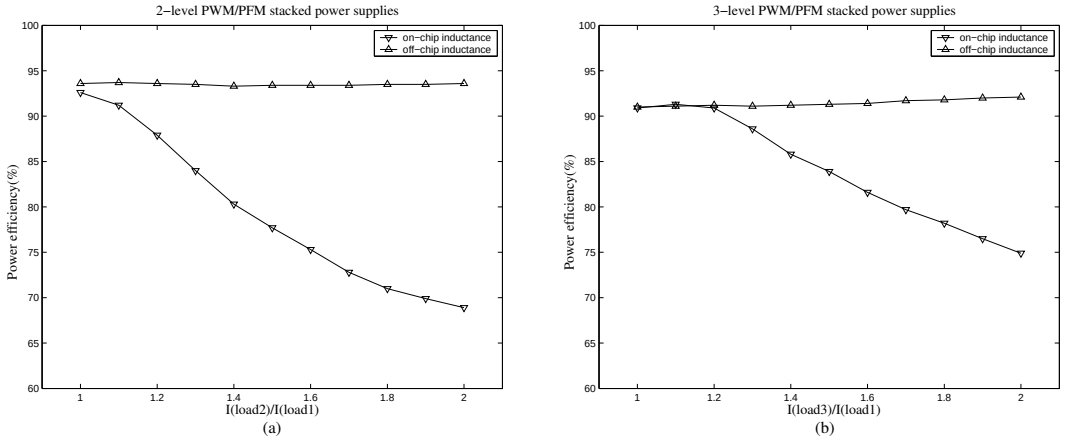

Figure 12: Power efficiency versus current imbalance in (a) two-level and (b) three-level PWM/PFM stacked power supplies.

as IR drop.

5. DISCUSSIONS ON CAD ALGORITHMS

The proposed power delivery system with field configurable stacked power domains can be readily used for hard cores implemented inside FPGA. For example, the two PowerPC cores in Vertix can be placed in two stacked power domains. For application to programmable fabrics, we need to optimize architectural parameters such as granularity and organization of Vdd-programmable regions and distribution of on-chip and off-chip DC/DC converters in addition to the conventional FPGA architectural parameters such as LUT size and logic block size. An architecture study similar to that for island style FPGA architectures [26–29] needs to be conducted, and design automation techniques to map a system to the multi-voltage FPGA need to be developed. The CAD components may include:

Clustering Algorithm: The objective of the clustering algorithm is to group modules with similar voltage requirements and to minimize the connections between different clusters. It therefore minimizes the potential interface cost between different voltage domains.

Vdd Level Assignment: In this design stage, Vdd levels for each cluster are determined, where clusters on critical paths use high Vdd levels and clusters with loose timing constraint can use low Vdd levels. Other constraints to consider during this assignment process include number of voltage regulators available, capacity of each voltage domain cluster, data-latching property of domain boundaries, and current matching of power domains.

Voltage Domain Pairing: The efficiency of the charge-recycling voltage regulator depends on the current match level of the two sibling voltage domains. Different logic functions mapped to the same programmable-Vdd region can exhibit dramatically different current behavior. One must ensure that voltage domains be paired up to minimize the mismatch of current and maximize the utilization of voltage domain clusters.

Placement of Voltage Domains: The location of Vdd-programmable regions for voltage domains affect the interconnect power and system performance. Placement should be developed to minimize the interconnect power and meet timing constraints.

6. CONCLUSION

In this paper, we have designed a novel multi-Vdd power supply system used to support an FPGA fabric that provides coarse-grained voltage programmability. With the use of off-chip pulse width modulation buck converter and multiple on-chip pulse frequency modulation buck converters, we are able to create mapping-time and run-time programmable supply voltages and obtain over 90% power efficiency during power conversion. This is a vast improvement from existing charge-recycling linear regulator based power supply systems that achieve power efficiencies in the 78% range. Our power supply system also has the added advantage of creating voltage signals with minimal noise. For instance, the output voltage ripple is less than 15%. Also, using stacked voltage domains reduces total supply current which enables us to have a solution with lower total power pins and IR

	voltage ripple	current reduction	PWM freq/L_f/C_f	PFM w/ off-chip inductance			PFM w/ on-chip inductance		
				T/L_f/C_f	area	efficiency	T/L_f/C_f	area	efficiency
two-level	$< 15\%$	$\sim 1/2$	5M/50n/2u	14n/50n/75n	$< 10\%$	93.5%	3n/1.5n/75n	< 15 %	79.2%
three-level	$< 15\%$	$\sim 2/3$	5M/30n/3u	9n/60n/85n	$< 20\%$	91.4%	3.1n/1.6n/100n	< 35 %	83.8%

Table 1: Metrics of the prototype designs for two- and three-level stacked power supplies.

drop. In the future, we will layout and fabricate the proposed field programmable multi-Vdd FPGA and develop the required CAD algorithms.

7. REFERENCES

[1] R. Wang, A. Krishnamurthy, R. Martin, T. Anderson, and D. Culler, "Modeling communication pipeline latency," vol. 26, no. 1, pp. 22–32, 1998.

[2] G. Qu and M. Potkonjak, "Techniques for energy minimization of communication pipelines," in *Proc. Intl. Conf. Computer-Aided Design*, pp. 597–600, Nov 1998.

[3] G. Qu, D. Kirovski, and M. Potkonjak, "Energy minimization of system pipelines using multiple voltages," in *Proc. Int. Sym. on Circuits and Systems*, vol. 1, pp. 362–365, June 1999.

[4] G. Qu and M. Potkonjak, "Power minimization using system-level partitioning of applications with quality of service requirements," in *Proc. Intl. Conf. Computer-Aided Design*, pp. 343–346, 1999.

[5] J. Chang and M. Pedram, "Energy minimization using multiple supply voltages," in *Proc. Intl. Symp. Low Power Electronics and Design*, pp. 157–162, 1996.

[6] A. Iyer and D. Marculescu, "Power efficiency of voltage scaling in multiple clock, multiple voltage cores," in *Proc. Intl. Conf. Computer-Aided Design*, pp. 379–387, Nov 2002.

[7] S. Rajapandian, Z. Xu, and K. Shepard, "Charge-recycling voltage domains for energy-efficient low-voltage operation of digital cmos circuits," in *Computer Design, 2003. Proceedings. 21st International Conference on*, pp. 98–102, 2003.

[8] V. Kursan, S. G. Narendra, V. K. De, and E. G. Friedman, "Analysis of buck converters for on-chip integration with a dual supply voltage microprocessor," *IEEE Trans. VLSI Syst.*, vol. 11, pp. 514–522, June 2003.

[9] G. Schrom, P. Hazucha, J.-H. Hahn, V. Kursun, D. Gardner, S. Narendra, T. Karnik, and V. De, "Feasibility of monolithic and 3d-stacked dc-dc converters for microprocessors in 90nm technology generation," in *Proc. Intl. Symp. Low Power Electronics and Design*, pp. 263–268, 2004.

[10] F. Li, Y. Lin, and L. He, "FPGA power reduction using configurable dual-vdd," in *Proc. Design Automation Conf.*, pp. 735–740, June 2004.

[11] F. Li, Y. Lin, and L. He, "Vdd programmability to reduce fpga interconnect power," in *Proc. Intl. Conf. Computer-Aided Design*, November 2004.

[12] Y. Lin, F. Li, and L. He, "Power modeling and architecture evaluation for FPGA with novel circuits for vdd programmability," in *Proc. ACM Intl. Symp. Field-Programmable Gate Arrays*, Februray 2005.

[13] Virtex-II Platform FPGAs: Complete Data Sheet.

[14] A. Pannikkat, J. Long, and J. Zhao, "Power delivery modeling and design methodology for a programmable logic device package," in *Electrical Performance of Electronic Packaging, 2004. IEEE 13th Topical Meeting on*, pp. 115–118, 2004.

[15] S. Lin and N. Chang, "Challenges in power-ground integrity," in *Proc. Intl. Conf. Computer-Aided Design*, pp. 651–654, Nov. 2001.

[16] V. Betz, J. Rose, and A. Marquardt, *Architecture and CAD for Deep-Submicron FPGAs*. Kluwer Academic Publishers, Feb 1999.

[17] K. Usami and et al, "Automated low-power technique exploiting multiple supply volgates applied to a media processor," *IEEE Journal of Solid-State Circuits*, vol. 33, no. 3, pp. 463–472, 1998.

[18] K. Usami and M. Horowitz, "Clustered voltage scaling techniques for low-power design," in *Proc. Intl. Symp. Low Power Electronics and Design*, 1995.

[19] M. Hamada and et al, "A top-down low power design technique using clustered voltage scaling with variable supply-voltage scheme," in *Proc. IEEE Custom Integrated Circuits Conf.*, pp. 495–498, 1998.

[20] F. Ishihara, F. Sheikh, and B. Nikolic, "Level conversion for dual-supply systems," in *Proc. Intl. Symp. Low Power Electronics and Design*, pp. 164 – 167, 2003.

[21] J. Montanaro and et. al., "A 160-MHz, 32-b, 0.5-w CMOS RISC microprocessor," in *IEEE Journal of Solid-State Circuits*, November 1996.

[22] D. W. Hart, ed., *Introduction to Power Electronics*. Prentice Hall, Upper Saddle River, N.J., 1997.

[23] A. J. Stratakos, S. R. Sanders, and R. W. Brodersen, "A low-voltage cmos dc-dc converter for a portable battery-operated system," in *Power Electronics Specialists Conference, PESC '94 Record., 25th Annual IEEE*, pp. 619–626, June 1994.

[24] A. J. Stratakos, "High-efficiency low-voltage dc-dc conversion for portable applications," *PhD dissertation, University of California, Berkeley*, 1998.

[25] http://www-device.eecs.berkeley.edu/ ptm/.

[26] J. Kouloheris and A. E. Gamal, "FPGA performance vs. cell granularity," in *Proc. IEEE Custom Integrated Circuits Conf.*, 6.2.1 - 6.2.4 1991.

[27] J. Kouloheris and A. E. Gamal, "FPGA area vs. cell granularity - lookup tables and PLA cells," in *First ACM Workshop on FPGA's*, Feb 1992.

[28] V. Betz and J. Rose, "Cluster-based logic blocks for FPGAs: Area-efficiency vs. input sharing and size," in *Proc. IEEE Custom Integrated Circuits Conf.*, 1997.

[29] E. Ahmed and J. Rose, "The effect of LUT and cluster size on deep-submicron FPGA performance and density," in *Proc. ACM Intl. Symp. Field-Programmable Gate Arrays*, pp. 3–12, Feb 2000.