Leakage Optimized DECAP Design for FPGAs

Balaji Vaidyanathan, Suresh Srinivasan, Yuan Xie, Narayanan Vijaykrishnan
Department of Computer Science and Engineering
Pennsylvania State University
PA, USA
{bvaidyan, yuanxie, vijay}@cse.psu.edu

Luo Rong
EE Department
Tsinghua University
Beijing, P.R. China
luorong@tsinghua.edu.cn

Abstract— **On-chip decoupling capacitors (Decaps) are widely used to mitigate on-chip power supply noise. At and below 100nm on-chip decaps face leakage and area overhead problems associated with it and is estimated to increase with technology scaling. A recent work has proposed a gated decap structure to reduce leakage in decaps. Their work analyzes leakage saving obtained by implementing gated decap structure in a pipelined super scalar core. FPGAs on the otherhand face similar leakage problem associated with decaps in their unmapped regions. We analyze here the leakage saving due to gated decap structure in FPGAs. With the on-chip gated decap structure we do uniform placement of decaps that achieves decap leakage savings of 7-60% with 39% on an average for various MCNC benchmarks mapped on to the FPGA device.**

I. Introduction

Similar to ASICs, FPGAs are moving to-wards increased integration with rapid technology scaling. Due to this technology scaling current consumption is predicted to follow an exponential trend. This increasing current is attributed to higher density of device integration and higher static power consumption aggravated by process variation [1] as technology scales below 100nm. Consequently power grid design is a growing concern due to this increased current consumption and parasitics associated with the power grid as they aggravate the IR drop and Ldi/dt noise.

Power supply integrity is becoming an issue in FPGAs [8] as a byproduct of reducing supply voltage with the shrinking process geometries. Currently Packaging solutions like placement of off-chip decoupling capacitors is one of the post-manufacture solutions which can mitigate supply noise for FPGAs. But with off-chip decap the impedance response of the power grid is found to be linearly related to frequency [2]. Xilinx ASMBL architecture [8] supports internal power and ground pads which provides uniform power distribution across the FPGA device. Another solution popular in ASIC domain is the addition of on-chip decap to keep the power noise within limits at higher frequency of operation. These on-chip decaps have to be manufactured into the chip, so its placement has to be decided at design time.

There has been work in the ASIC domain which analyzes the current consumption profile at design time and selectively places decaps on-chip [3]. FPGAs in contrast have diverse designs mapped onto them which can lead to loads varying temporarily and spatially across designs. Hence the on-chip decap needs to be placed uniformly over the chip to overcome supply noise. This ad-hoc placement of decaps in FPGA can have area and leakage power overheads. Area overhead is due to excessive placement of decaps that takes care of supply noise for any arbitrarily mapped design. Leakage power overhead is due to the presence of unused on-chip decaps in part of the chip where no design is mapped. We propose here a run-time on-chip decap activation to cater dynamically varying transient loads and also reduce leakage power overheads associated with having unused decaps on-chip.

This paper is organized as follows. Section II reviews related work and touches upon the direction of our work. Models for power grid, decap and current consumption in FPGA are explained in section III. Section IV explains the methodology and experimental setup used. Results are provided in section V. Section VI concludes with possible future work.

II. Related Work

There are two popular static design solutions to mitigate supply drop. They are supply wire thickening in places where IR drop is expected to be high [13] and insertion of on-chip decaps providing local supply of charge [10] to lessen dynamic voltage fluctuation.

As Current density increases with technology scaling an intelligent placement of decaps is necessary in places where activity is high. Decaps placed far away from noisy nodes lead to insufficient noise reduction. An optimal placement of minimum-sized decap is proposed for power noise reduction in standard-cell based ASIC design [14]. On-chip decaps are typically designed using MOS based thin gate oxide transistors which have excessive gate leakage currents for gate oxide thickness below 20Å.

Fu et al. [11] proposed a power grid optimization algorithm for power supply noise reduction taking decap leakage and routing area as additional optimization criteria. Their work assumes that the on-chip decap placement is known at design time so that the wire sizing can be done in places where the decap leakage is high. FPGAs have current consumption profile which is unknown at design time. The mapped designs decides the current consumption profile of the chip at configuration time. Hence a large number of on-chip decaps are necessary to achieve robust power grid design in FPGA catering to different designs mapped onto it. Consequently there will be more unused on-chip decaps in that part of the FPGA where a design is not mapped.

These unused decaps have leakage overheads associated with them and the leakage power for these decaps is expected

[0]This work was supported in parts by grants from NSF 0454123 and GSRC

to rise exponentially with reduction in oxide thickness. We use MOS based decap in this study as they occupy less area compared to other types of on chip capacitors that are used [11]. To overcome the leakage power consumption in unused decaps we propose a dynamic mechanism to turn off decaps.

III. Model

To model the problem of finding the number of decaps and propose decap leakage optimizations for an FPGA device, we need to have detailed dynamic current profiles and appropriate models for interconnect and the decaps. Each of the models used by us are described in detail below.

A. Interconnect model

We assume a mesh structure to build the power grid for the FPGA as shown in Figure 1. The device and interconnect layers are modeled as a RC network. Package and power contact parasitics are modeled as RLC elements. The intrinsic capacitance of the circuit and the extrinsic on-chip decaps are added to the power grid. The VDD nodes are assumed to be uniformly distributed over the power grid in a flip chip packaging. We use 65nm BPTM [5] interconnect parameters to calculate the parasitics associated with the power grid. The supply wires are appropriately sized so that the IR drop at nodes are kept minimum.

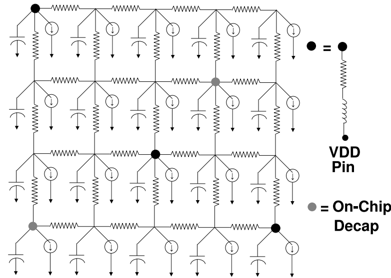


Fig. 1. Power Grid Layout with VDD nodes, current sources and On-chip decaps.

B. Decap Power Models

Thin-oxide MOS-based decaps are NMOS or PMOS transistor with their source, body and drain tied together [11] where, the gate node provides the necessary capacitance. These transistors typically have long and wide channels with large capacitance to supply huge transient currents. With wide gate area and thin gate oxide (1nm for sub-100nm technology [9]) the gate leakage current of these devices has become significantly large. Further, in advanced technologies the leakage current of thin oxide MOS are found to exponentially vary with small variations in oxide thickness [15]. The decap gate leakage is formulated as shown in Eq 1. The decap capacitance is given by Eq 2, which is similar to the one mentioned in Fu et al. [11].

$$I_{gate} = \alpha \times \epsilon_{ox} \times e^{\frac{-\beta}{\epsilon_{ox}}} \times W_n, \quad (1)$$

Where $\epsilon_{ox} = \frac{V_{ox}}{T_{ox}}$ and the constants α and β are technology dependent parameters. It is evident that with the reduction in oxide thickness the gate leakage consumption of the decaps increase. However, using a thicker oxide for the design of decaps requires increasing the width and thereby the area of the decap in the same proportion to achieve the required nodal capacitance as evident from equation 2, which impacts the on-chip integration of decaps.

$$C_d = \gamma \times \frac{\epsilon \times \epsilon_0}{T_{ox}} \times S, \quad (2)$$

Where W_n, T_{ox}, V_{ox}, and S are width, oxide thickness, gate voltage, and area of the decap respectively. The parameter γ depends upon the shape of the transistor used and to remove such a dependence we assume a value of $\gamma = 1$ for our analysis as in [11]. To obtain the values of parameters α and β we simulated a NMOS decap of $W_n = 0.2\mu m$, effective gate length $L_eff = 0.038\mu m$, $T_{ox} = 10$Å, $V_{ox} = 1.0$, and $S = W_n \times L_eff$ in 65nm technology using BSIM4 model [5]. The values of $\alpha = 2.7916e^{-11}$, and $\beta = 2.6146e^{+07}$ are obtained using such simulations.

C. Current Profile of the FPGA

We assumed an island style FPGA architecture for our implementation as depicted in Fig. 2. Since, for sub-100nm technologies apart from the dynamic power, the static power becomes a significant component we estimate both the components of power consumption accurately for each of the components.

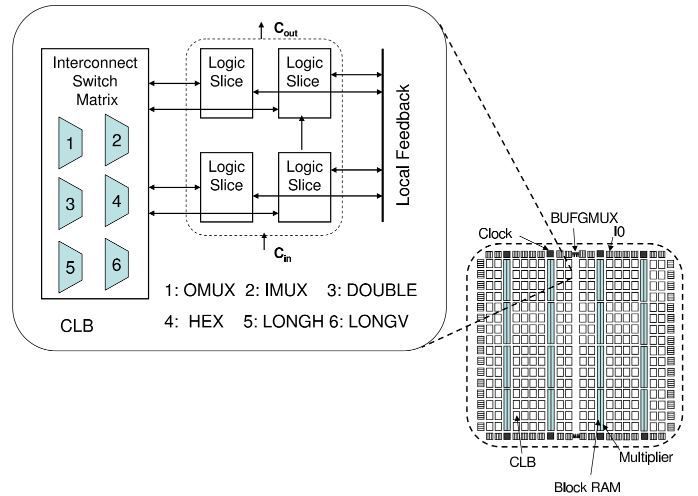


Fig. 2. Virtex-II FPGA architecture.

The FPGA circuit comprises of the logic blocks or look-up-tables (LUT), routing muxes, SRAM cells, Flip-flops, and BRAM cells as shown in Figure 2. We custom designed the 4-input LUT multiplexer using a 2-level transmission gate design as shown in Figure 3a. The routing fabric in FPGAs typically comprises of multiplexers of different sizes described by Anderson et al. [12]. Figure 3b shows the circuit for a 16 input routing multiplexer circuit design used by us for our analysis.

All the circuits were simulated using BSIM4 model [5] for 65nm technology. The devices were operated at 1V with clock period of 1GHz and 0.1ns rise and fall time. The dynamic (static) power consumption of a component depends on the switching (static) probabilities of its inputs. The static and switching probabilities of each of the nets were obtained using the placement and routing information provided by

the Xilinx Development Language (XDL) file, which is a text file comprising information about the placement, routing and configuration of the each of the slices. We developed a tool FPGA Probability Estimator (FPEST) in java which models complete hierarchy of the netlist and obtained the static and the transition probabilities of each of the driven nets of the design purely using the XDL file. The input signal probabilities at the inputs pins in our case were assumed to be 0.5.

Using these probability numbers at each of the nets along with SPICE models for FPGA circuits like the logic blocks (CLBs) and the routing fabric (routing muxes), static and dynamic current consumption of each component is obtained. In this work we demonstrate the results for VDD grid only and similar experiments can be done for GND planes with minor changes.

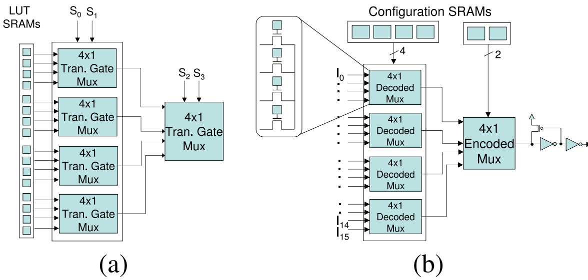


Fig. 3. (a) 4-Input LUT. (b) 16x1 Router Mux.

IV. METHODOLOGY

To reduce the on-chip decap leakage power consumption we propose a modified decap circuit in FPGA context which may be turned off whenever possible. Consequently, to obtain the leakage benefits, we only turn on the decaps which are essential to maintain the noise margins based upon the current drawn from the power grid. We determine such a turn off policy based on an iterative scheme, which operates depending on the noise margin threshold requirements. The new decap circuit design for FPGAs and the methodology to determine the number of decaps is presented in the following sections.

A. Leakage Control in On-chip Decap

The on-chip decap leakage power overhead associated with unused on-chip decaps in FPGA was introduced in section II. A simple solution to reduce the leakage power overhead associated with these unused decaps is to turn them off. The decap structure for FPGAs with a capability of turning off is shown in Figure 4. This is a gated decap structure introduced by Yiren et al. [4].

This gated decap has a NMOS decap connected to GND node though a NMOS sleep transistor S. The gate input to the sleep transistor s comes from a SRAM configuration bit. SRAM configuration bit is set to ’0’ in the sleep mode and ’1’ in the normal operational mode. In the sleep mode the decap structure has negligible leakage due to increased voltage at node G. In the normal operational mode the decap structure operates as a normal decap with the voltage at G being pulled down to GND. More details on decap leakage components in sleep and normal operational modes can be got from [4].

[4] reported minimal area and energy overhead (in operational mode) of the gated decap as 6.78% and 1.65% compared to the normal un-gated decap for 90nm. The leakage power savings obtained by gated decap in sleep mode is 99.40% compared to the un-gated decap. The gating period of the gated decap can be theoretically assumed to be ∞ in FPGA as we switch all the required decaps only at the time of configuration. This type of configuration thus eliminates the frequent turning-on and turning-off energy overhead associated with the sleep transitor S.

In our current approach we make a decision of turning the configuration bit on or off based upon the current consumption profile we obtain before configuration, using the models developed by us. Since we use one SRAM bit per decap, it introduces negligible area overhead, specifically when compared to other schemes like using thicker gate oxides. Furthermore, since SRAM cells in the FPGA are usually optimized for dynamic and static power, the power overhead associated with the extra SRAM configuration bits are negligible [7].

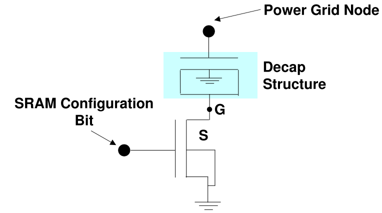


Fig. 4. Leakage Control in Decap using sleep transistor.

B. On-chip Decap Placement Structure

We consider a uniform decap placement strategy where the on-chip decaps are placed uniformly over the chip area. Note that such a choice is motivated by the primary objective of symmetry which is in coherence with the symmetric FPGA device fabrication philosophy. However, we determine the number of decaps that may be used(turned ON) for a given device using the current traces obtained from different applications as discussed further.

C. On-chip Decap Selection and VDD Noise Analysis Flow

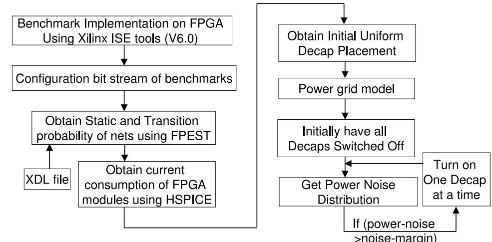


Fig. 5. On-chip decap selection and VDD drop analysis flow for FPGA.

Figure 5 describes our scheme for selecting the on-chip decaps to be turned ON and our analysis to obtain VDD profile for different benchmarks. We mapped 9 MCNC benchmarks onto the minimum sized device [6] and generated the bit streams using Xilinx ISE tools(ver 6.0). Table I provides information about the benchmarks that are implemented and the corresponding devices. The power consumption profile

is gathered by performing HSPICE simulation of FPGA circuits using transition and static probabilities computed as explained in section III using the in-house probability estimator for applications implemented on Xilinx FPGAs.

The power grid model integrated with on-chip decap placement information is simulated to get the VDD noise and the number of decaps required are obtained using an iterative process of incremental decap insertion and checking the power to be within the noise margin. We increase the number of on-chip decaps required to be turned ON in steps until power noise is within the noise margin. Figure 6 shows the VDD profile over the chip for the MCNC benchmarks apex4 and ex1010 using the minimum number of decaps turned ON while the decaps are uniformly placed over the whole device.

Benchmarks	Slices Used	IO Pins	Device
alu4	252	22	xc2v40-cs144
apex2	204	42	xc2v40-cs144
apex4	549	28	xc2v250-fg456
ex1010	567	20	xc2v250-fg456
ex5p	286	21	xc2v80-cs144
misex3	131	28	xc2v40-cs144
pdc	308	76	xc2v80-cs144
seq	605	62	xc2v250-fg456
spla	116	56	xc2v40-cs144

TABLE I

MCNC BENCHMARK DESCRIPTION AND FPGA DEVICES TO WHICH THEY ARE MAPPED TO.

V. Experiments and Results

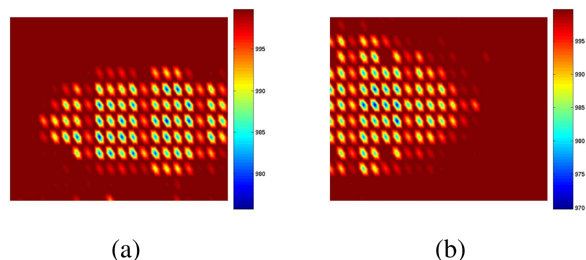
Using the aforementioned methodology of decap placement, leakage estimation in decaps and fine grain control by turning on and off of the decaps we obtain the leakage savings in the FPGAs with uniform decap placement. We use a capacitance value of 0.1nf for all the on-chip decaps placed uniformly on the chip. The decaps were switched ON to maintain the VDD noise to be within 5% of the VDD. Apart from just using the decap for VDD noise reduction we also sized the supply wires in places where the decap leakage was present to reduce the IR drop due to the on-chip decap leakage as described in Fu et al. [11].

Figure 7a shows the maximum VDD drop values observed in all the nine benchmarks mapped to the Xilinx FPGA devices. Using our proposed scheme of selective decap turning ON we demonstrate the leakage power savings can be achieved by us for different applications in figure 7b. It is evident from the graph that we obtain an average leakage power saving of 38.7% using our selective decap shut down scheme.

It is important to note that the such leakage benefits primarily depend upon the utilization of the device and the current drawn by the different components due to their switching activities. This is the reason why benchmark alu4 shows only an leakage reduction of 7% due to its high device utilization, while on the other hand we obtain leakage benefits of close to 60% for ex5p.

VI. Conclusion and Future Work

Our work also demonstrates a detailed methodology to analyze the power grid design for FPGAs considering a careful evaluation of different associated parasitics. We adopted



(a) (b)

Fig. 6. VDD drop plot for benchmark apex4 (a) and ex1010 (b) mapped onto a Virtex-II FPGA XC2V250-fg456.

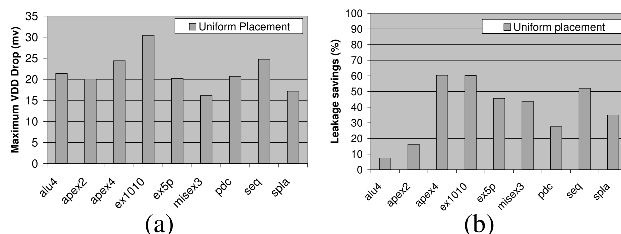


Fig. 7. Maximum VDD drop observed in all the benchmarks mapped to Xilinx FPGA devices (a) and Decap leakage savings with our selective decap shutdown (b).

a gated decap structure to reduce the leakage overhead associated with unused decaps in FPGA and proposed a design methodology to implement it on FPGAs. The proposed selective turning off scheme for the on-chip decaps shows an average power savings of 38.7% for nine MCNC benchmarks implemented on minimal sized Xilinx FPGA devices.

References

[1] I. A. Ferzli, F. N. Najm "Analysis and Verification of Power Grids Considering Process-Induced Leakage Current Variations," In *IEEE Trans. on CAD of Integ. Circuits and Systems*, 2006.

[2] H. H. Chen, D. D. Ling "Power Supply Noise Analysis Methodology for Deep-Submicron VLSI Chip Design," In *Proc. of Design Automation Conferernce*, 1997.

[3] M. D. Pant, P. Pant, D. S. Wills "On-chip decoupling capacitor optimization using architectural level prediction," In *IEEE Trans. on VLSI Systems*, Vol.10, pp.319-326, june 2002.

[4] Y. Chen, H. Li, K. Roy, C. -K. Koh "Gated Decap: Gate Leakage Control of On-chip Decoupling Capacitors in Scaled Technologies," In *Proc. of IEEE Custom Integ. Circuits Conference*, 2005.

[5] " BPTM 65nm: Berkeley Predictive Technology Model "

[6] " http://direct.xilinx.com/bvdocs/publications/ds031.pdf"

[7] " http://www.xilinx.com/bvdocs/whitepapers/wp221.pdf"

[8] " http://fpgajournal.com/articles/platform_xilinx.htm"

[9] " ITRS 2003 Report: http://public.itrs.net/ "

[10] S. Zhao, K. Roy, C. -K. Koh "Decoupling capacitance allocation and its application to power-supply noise-aware floorplanning," In *IEEE Trans. on CAD of Integ. Circuits and Systems*, 2002.

[11] J. Fu, Z. Luo, X. Hong, Y. Cai, S. X. -D. Tan, Z. Pan "VLSI On-Chip Power/Ground Network Optimization Considering Decap Leakage Currents," In *Proc. of the ASP-DAC*, 2005

[12] J. H. Anderson, F. Najm, and T. Tuan. "Active leakage power optimization for FPGAs," In *Proc. of ACM/SIGDA Intl. Symposium on FPGAs*, 2004

[13] X. -D. Tan, C. -J. R. Shi, D. Lungeanu, J. -C. Lee, L. -P. Yuan "Reliability-constrained area optimization of VLSI power/ground networks via sequence of linear programmings," In *Proc. of Design Automation Conference*, 1999

[14] H. Su, S. S. Sapatnekar, S. R. Nassif "Optimal Decoupling Capacitor Sizing and Placement for Standard-Cell Layout Designs," In *IEEE Trans. on Computer-Aided Design of Integ. Circuits and Systems*, 2003

[15] Y. -C. Yeo, T. -J. King, C. Hu "MOSFET gate leakage modeling and selection guide for alternative gate dielectrics based on leakage considerations," In *IEEE Trans. on Electron Devices*, 2003