

On-Package Decoupling Optimization with Package Macromodels[†]

Hui Zheng¹, Byron Krauter², Lawrence Pileggi¹

¹Carnegie Mellon University
Department of Electrical and Computer Engineering
Pittsburgh, PA
{hzheng, pileggi}@ece.cmu.edu

²IBM Corporation
Austin, TX
krauter@us.ibm.com

ABSTRACT

Suppressing clock-gating-induced noise at the chip-package interface is one of the most challenging power distribution integrity issues. In this paper, accurate and efficient assessment of the effectiveness of on-package decoupling is facilitated by package macromodels which compactly represent packaging parasitics among multiple on-package decoupling and on-chip ports. Based on such assessment, a simulated-annealing-based optimization procedure is developed with the goal of finding the most cost-effective on-package decoupling while meeting the noise budget.

1 Introduction

As on-chip switching current and power consumption skyrockets due to the increasing speed and density of VLSI chips, it has become increasingly important to accurately model and analyze the combined chip-package interfaces of power/ground distribution for the following reasons. Firstly, aggressive clock gating schemes, which have been widely adopted to reduce chip power, generate large current transients. Due to the wide spectral distributions of those current transients, it is imperative to maintain a low target impedance throughout the mid-frequency spectrum at the chip-package interface to meet the noise budget. To achieve this objective, on-package decoupling optimization has proven to be an effective approach [2]. However, accurate assessment of the effects of the on-package decoupling capacitors must be based on accurate package models. Secondly, the increase in signal pin requirements is outpacing the relatively slow advances in packaging technologies such that relatively fewer pin and wiring resources are available for on-package power/ground distribution. It has come to a point where the package-to-decoupling-capacitor parasitics are becoming comparable to those of the decoupling capacitors themselves. Thus, there is an increasing need to replace the traditional simple models [1] -- which ignore or crudely model the packaging parasitics -- with multiport models that accurately represent the parasitics for the chip-package interface.

In this paper, a methodology that we have developed for accurately modeling packages is briefly described in Section 2. This methodology tackles the modeling and anal-

[†]This work was supported by Semiconductor Research Corporation (SRC) under contract 2000-TJ-778 and International Business Machines Corporation (IBM).

ysis complexities for on-package power/ground distribution via window-based susceptance (inverse of inductance) extraction [4, 5] and subsequent model order reduction (MOR) [6, 7]. The resulting macromodels, derived from the detailed extracted RCS (S representing susceptance) circuits, provide not only the accurate multiport representation but also the efficiency that is necessary for design optimization [3]. Using these package macromodels, we can then study how combinations of on-package decoupling capacitors can affect the frequency-domain behavior at the chip-package interface, which is reported in Section 3. Furthermore, the package macromodels with multiple on-chip ports allow us to address the impact of on-chip non-uniform switching on clock-gating-induced noise. A simple metric for assessing these effects is given in Section 4. In Section 5, with the ability to quickly evaluate different deployments of on-package decoupling capacitors, a simulated-annealing-based optimization procedure is developed with the goal of reducing decoupling capacitor cost while achieving the target impedance. Throughout sections 2, 3, 4 and 5, a C4-technology-based industrial package example is modeled and analyzed for demonstration purposes.

2 Package Macromodeling

The main difficulty in efficient modeling of on-package power/ground distributions lies in the complexity of modeling and analyzing massive 3D magnetic couplings among hundreds of thousands of conductors. In our methodology which is described fully in [3], the complexity problem is tackled in the following two-step procedure. Firstly, we adopt the susceptance concept (inverse of inductance) which enables more robust sparsification of the magnetic couplings than the inductance [4, 5]. This concept, coupled with several advanced implementation techniques, has made it tractable to build detailed RCS (S representing susceptance) circuit models for some ASIC packages in a reasonable amount of time. For example, the statistics of

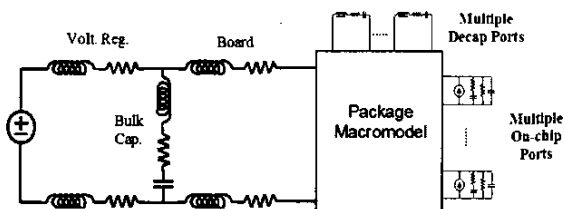


Fig. 1: Package Macromodel for Chip-Package Interface

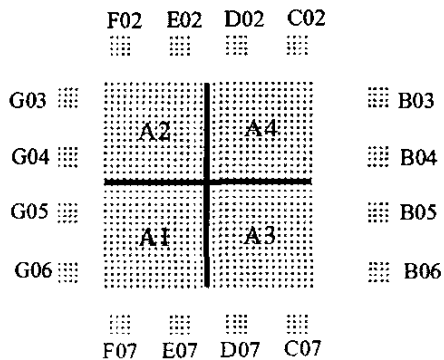


Fig. 2: Top view: package to chip and decap pins

applying our package extractor to the power/ground distribution in the testbench package is tabulated in Table 1 (Per-

No. of Conductors	No. of Mutuals	No. of Circuit Nodes	Extraction Runtime
79,770	536,953	125,722	102 minutes

Table 1: Extraction Statistics

formed on an IBM RS/6000 44p Model 270 Machine).

However, in an optimization scenario where multiple evaluations are needed, it becomes cumbersome to apply detailed RCS circuits in a simulation engine. For example, it consumes about 1 Gigabyte memory and 20 minutes to do a transient simulation of 400 time steps on the extracted circuit from this package. An AC sweep analysis of this package is expected to take more time and space since it involves complex matrix computations. Therefore, we take the second step to make the detailed models compact through some model order reduction (MOR) techniques which have been extended to handle RCS circuits with high efficiency and accuracy [6, 7]. The resulting package macromodels can significantly reduce the simulation complexity. More importantly, the multiport nature of these macromodels allows accurate characterization of the parasitics among multiple on-package decoupling capacitors and different on-chip regions, as shown in Fig. 1. In the testbench package, the top view of which is shown in Fig. 2, there are 16 decoupling positions and the chip is partitioned into 4 switching regions. Adding the bottom port, a 21-port 105th-order macromodel was constructed through our MOR program. The MOR statistics is shown in Table 2 (Performed on an IBM 7017-s85 AIX system). We

No. of State Variables in Original System	No. of State Variables in Reduced-Order System	MOR Runtime
205,492	105	664 seconds

Table 2: MOR Statistics

choose the order of the reduced system to be 105 since our experiments show that there is no significant accuracy improvement for orders above 105.

3 Effectiveness of Mixing Decoupling Capacitors

To reduce the power supply noise caused by current surges, it has been a common practice to place decoupling capacitors at all packaging levels to serve as charge reservoirs and suppliers, as shown in Fig. 1. However, the inherent inductance in the power/ground distribution, combined with those decoupling capacitors and associated parasitics, can generate resonance peaks in the frequency-domain characteristics. This phenomenon can magnify the noise problem especially when some clock-gating-induced current transients contain some considerable components at frequencies close to the resonant frequencies. Therefore, a good package design must ensure that those resonance peaks be suppressed below a certain target impedance.

Due to its inherent parasitics, a realistic on-package decoupling capacitor is modeled as a series connection of three components: an equivalent series capacitor (ESC), an equivalent series resistor (ESR) and an equivalent series inductor (ESL). Generally, for the same amount of ESC, a decoupling capacitor with a lower ESL is more expensive. The 4 types of decoupling capacitors used in our experiments are characterized in Table 3. All of them have the same ESR and ESC except different ESL. The actual price information is not available, so some numbers correlated with quality are

	No C	C1d	C1c	C1b	C1a
ESC	-	50 nF	100 nF	50 nF	100 nF
ESR	-	0.060 ohm	0.060 ohm	0.030 ohm	0.030 ohm
ESL	-	100 pH	100 pH	40 pH	40 pH
Price	0	1	2	2	4

Table 3: Decoupling Capacitors

used instead. Since we also have the choice of not populating a decoupling port, a fifth choice, "No C", is listed in the table with zero cost.

To study the efficacy of the decoupling capacitor in the context of packaging parasitics, we conduct the following experiment. First, a 18-port 90th-order macromodel is built to represent the testbench package. The 18 ports include one on-chip port, 16 on-package decoupling ports and 1 bottom port.

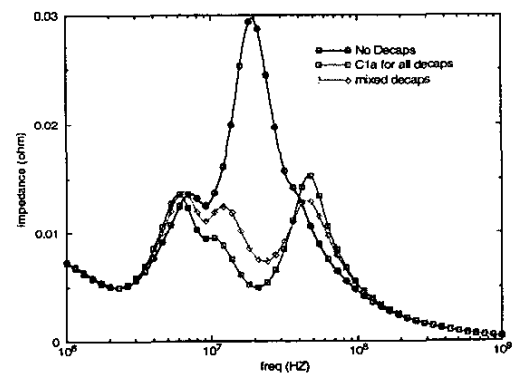


Fig. 3: Comparison of Decoupling Effectiveness

We use one on-chip port since uniform switching across the chip is assumed (The impact of non-uniform switching will be discussed in the next section). Again, the order of 90 is chosen by experiment. Then for each deployment of decoupling

	No C	C1d	C1c	C1b	C1a	Peak Z (mohm)	Decap Cost
No Decaps	All					29.8	0
Priciest					All	15.3	64
Mixed	G03, F02, E07, B04	G06, F07, B06	G05, E02, B05	G04, D02, D07	C02, C07, B03	13.7	27

Table 4: Experiment with three deployments

capacitors in Table 4 at the 16 ports, an AC sweep analysis with the package macromodel is performed to determine the impedance seen by the chip port as shown in Fig. 3. It can be seen that placing the priciest capacitor at all the decoupling ports does not always lead to the lowest peak impedance. In this case, the most cost-effective way to achieve a target impedance of 14 mohm is to use a mixed combination of capacitors and missing capacitors. This observation motivates our optimization approach in Section 5, the goal of which is find the right deployment of decoupling capacitors while striving for the minimum cost.

4 Addressing On-Chip Non-uniform Switching

The discussion in Section 5 is based on the assumption that all circuit blocks on a chip have the same switching density. However, this does not hold true in modern modular VLSI designs, i.e., some parts of a chip are always hotter. Therefore, to accurately capture the noise behaviors and interactions between different parts of a chip, multiple on-chip ports should be included in the package macromodel.

Now, suppose there are N on-chip ports. With the package macromodel, an $N \times N$ impedance matrix $Z(j\omega)$ can be easily calculated at a certain frequency. Given the clock-gating-induced currents at on-chip port, the frequency-domain response at port i is:

$$V_i(j\omega) = \sum_{j=1}^N Z_{ij}(j\omega) I_j(j\omega) \quad (1)$$

Since the phases of the currents are nearly arbitrary due to the decentralized nature of aggressive clock gating, it is very likely that the noise magnitude takes the upper bound:

$$|V_i(j\omega)| = \left| \sum_{j=1}^N Z_{ij}(j\omega) I_j(j\omega) \right| \leq \sum_{j=1}^N |Z_{ij}(j\omega)| |I_j(j\omega)| \quad (2)$$

Suppose the current at port j is a portion of the total current: $I_j = \mathcal{H}_j \cdot I_{total}$, where \mathcal{H}_j represents the relative switching intensity of partition j , we can define an effective impedance for port i as:

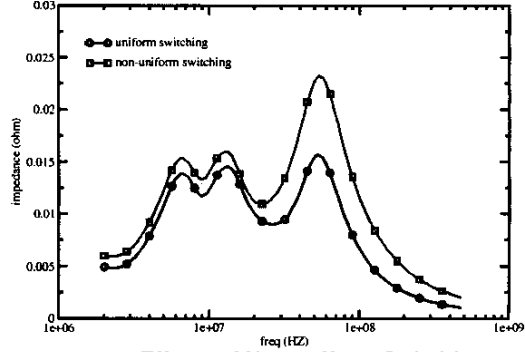


Fig. 4: Effects of Non-uniform Switching

$$|Z_i(j\omega)|_{eff} = \sum_{j=1}^N |Z_{ij}(j\omega)| \mathcal{H}_j \quad (3)$$

Due to the symmetry of the impedance matrix, finding all the peak effective impedances at N on-chip ports only requires N AC sweep analyses. The biggest peak effective impedance among them should be below a target impedance in order to meet the noise budget.

To demonstrate the non-uniform switching effects, we divide the testbench chip into 4 parts (as shown in Fig. 2), and perform AC analyses with two switching scenarios: 1) uniform: $\mathcal{H}_1 = \mathcal{H}_2 = \mathcal{H}_3 = \mathcal{H}_4 = 0.25$; 2) non-uniform: $\mathcal{H}_1 = 0.5$, $\mathcal{H}_2 = 0.2$, $\mathcal{H}_3 = 0.2$, and $\mathcal{H}_4 = 0.1$. In both cases, all decoupling ports use C1d. As shown in Fig. 4, the worst-case port peak impedance (23.2 mohm) with non-uniform switching is 47% higher than that (15.7 mohm) with uniform switching.

5 On-Package Decoupling Optimization

Based on the discussion in section 3 and 4, we can formulate the following on-package decoupling optimization problem: given a set of candidate decoupling capacitors and a user-defined target impedance, find the set of capacitors with the lowest possible cost that ensures the impedance peak is below the target impedance. Expressed mathematically, it is:

$$\begin{aligned} \min \quad & \sum_{i=1}^n P_i \\ \text{subject to} \quad & \max(Z^j_{peak}) \leq Z_{target} \end{aligned} \quad (4)$$

where n is the number of decoupling capacitor ports and P_i is the price of the decoupling capacitor placed at port i . Z^j_{peak} is the peak of the effective impedance at on-chip port j as defined in (3).

To solve this optimization problem, we use simulated annealing, a stochastic optimization procedure which accom-

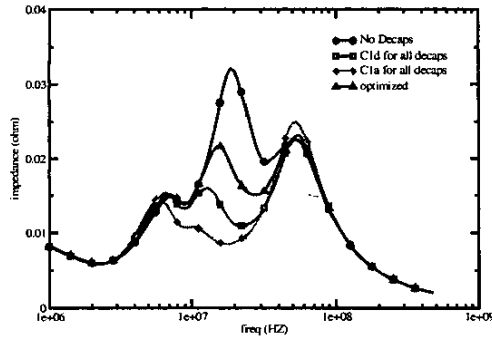


Fig. 5: Optimization Results

modates open cost functions. For our problem, the cost function is:

$$W_p \cdot \sum_{i=1}^n P_i + W_z \cdot (\max(Z^j_{peak}) - Z_{target}) \quad (5)$$

where W_p and W_z are the adjusting weights for decap cost and target impedance. In order to enforce the target impedance requirement, during the annealing process, W_z increases dynamically when the maximum port effective impedance is greater than the target impedance and becomes zero when the impedance requirement is satisfied. A drawback of simulated annealing is that it has to explore a large solution space -- more simulation points -- in order to avoid local optimum points. However, with fast analyses facilitated by our package macro-models, it becomes feasible to solve this optimization problem in a reasonable amount of time, as can be demonstrated in the following experiment.

Starting with the testbench package with 16 decoupling ports, we partition the chip into 4 switching regions as shown in Fig. 2. Assume that the chip non-uniformly switches with the weights in section 4. The decoupling capacitors are chosen from Table 3. The target impedance is set to be 22.0 mohm. The runtime for the optimization procedure is 48.6 minutes on 1 GHz Linux-Pentium III system, which includes 1920 AC sweep analyses. The impedance response for the optimized decoupling capacitor deployment is shown and compared to the responses from the "no decap", "all C1d", and "all C1a" deployments in Fig. 5. From the optimized deployment shown in Table 5, it can be seen that there are more decoupling

	No C	C1d	C1c	C1b	C1a	Peak Z (mohm)	Decap Cost
No Decaps	All					32.1	0
All C1d		All				23.2	16
All C1a					All	25.0	64
Optimized	B04, B05, B06, C02, D02, D07, E07, F02, F07, G03, G04	B03, E02, G05, G06	C07			21.1	6

Table 5: Optimization Comparison

capacitors around A1, which is the hottest region. Also, to achieve the target impedance, the optimized solution only requires five decoupling ports be populated, leaving eleven ports unused.

6 Summary

One of the biggest concerns in power distribution design is the transient noise injected by aggressive on-chip clock gating at the chip-package interface. With slow development in package design, there is an increasing need to evaluate with accurate package models the effectiveness of on-package decoupling capacitors for suppressing the noise. An efficient methodology that we have developed is able to produce package models that require little effort to analyze while retaining the information of packaging parasitics among multiple on-package decoupling and on-chip ports. Equipped with these package models, we develop a simple metric addressing the impact of non-uniform on-chip switching and further devise a decoupling optimization strategy that is demonstrated to be able to effectively prevent expensive overdesigns while meeting the noise budget.

7 Acknowledgement

The authors would like to thank Anand Haridass of IBM for providing packaging information and for the many interesting discussions on packaging issues.

References

- [1] H. B. Bakoglu, *Circuits, Interconnections, and Packaging for VLSI*, Addison-Wesley Publishing Company, 1990
- [2] L. Smith, et. al. "Power Distribution System Design Methodology and Capacitor Selection for Modern CMOS Technology," IEEE Transactions on Advanced Packaging, Vol. 22, No. 3, pp. 284-91, August 1999
- [3] H. Zheng, B. Krauter and L.T. Pileggi, "Electrical Modeling of Integrated-Package Power/Ground Distributions," IEEE Design and Test Magazine, in press.
- [4] A. Devgan, H. Ji, and W. Dai., "How to Efficiently Capture On-Chip Inductance Effects: Introducing a New Circuit Element K," IEEE/ACM Proc. ICCAD, pp. 150-155, Nov. 2000
- [5] M. Beattie and L. Pileggi, "Efficient Inductance Extraction via Windowing," IEEE/ACM Proc. 2001 DATE, pp. 430-436, March, 2001
- [6] H. Zheng and L. T. Pileggi, "Robust and Passive Model Order Reduction for Circuit Containing Susceptance Elements," to appear in Proc. IEEE International Conference on Computer Aided Design, Nov. 2002
- [7] M. Celik, H. Zheng and L. T. Pileggi, "Efficient Reduction of Susceptance-based Package Models Using PRIMA," in Proc. 11th Topical Meeting on Electrical Performance of Electronic Packaging, Oct. 2002