RC23670 (W0507-139) July 19, 2005
Electrical Engineering

IBM Research Report

Decoupling Capacitance Modeling

Sani Nassif, Kanak Agarwal, Emrah Acar
IBM Research Division
Austin Research Laboratory
11501 Burnet Road
Austin, TX 78758

Research Division
Almaden - Austin - Beijing - Haifa - India - T. J. Watson - Tokyo - Zurich

Decoupling Capacitance Modeling

Sani Nassif, Kanak Agarwal, Emrah Acar

IBM Research, Austin

{nassif, kbagarwa, emrah}@us.ibm.com

Abstract: Technology scaling and the push for ever increased performance has resulted in the rapid increase of integrated circuit power dissipation. We are already in the era of the 100 Watt IC [1]. This necessitates the detailed modeling and analysis of the on-chip power distribution for robustness and reliability [2, 3, 4]. An important component of this model is the decoupling capacitance of the design which includes dedicated decoupling capacitors as well as the capacitance of non-switching circuits. This paper describes a technique for modeling the decoupling capacitance of circuits. An exact simulation-based method is outlined, and fast but accurate analytical models are proposed.

1. Introduction

The scaling of process technology in the nanometer regime has resulted in smaller feature sizes and increased levels of device and interconnect integration. This trend has enabled faster switching speeds and higher current densities but it has also caused various noise and signal integrity issues such as power supply noise [5]. Supply noise is caused when large instantaneous current is drawn from the power supply due to the simultaneous switching of a large number of devices. The large current peaks can cause significant IR drop and LdI/dt noise in the power grid.

Decoupling capacitors (commonly referred to as decaps) are an effective solution for mitigating power grid related noise. In the simple circuit in Figure 1, which is a canonical model of a power grid, L is the package inductance, R_g models the power grid, R_d and C_d model the decoupling capacitance, and I_{load} models the time dependent current waveform for the load, which we model as a triangular waveform with slope μ and peak time t_p. It was shown in [6] the maximum voltage drop V_{max} is well approximated by:

$$V_{\max} = \mu(L + R_g(t_p - C_d R_g(1 - e^{\frac{-t_p}{(R_g + R_d)C_d}}))) \quad (1)$$

Given a modular design with fixed power grid, operating frequency and power consumption, hence a fixed current waveform (μ and t_p), a module designer's options for local control of noise are focused entirely on the local decoupling capacitance C_d and how *close* it is to the load, which is modeled via the series resistance R_d. This is because package inductance and overall the power grid (i.e. L and R_g) are determined early in the design cycle since they are difficult to change without significant design rework [7, 8, 9, 10]. This explains the need to carefully analyze local decoupling capacitance to insure robustness against power grid noise [11, 12, 13].

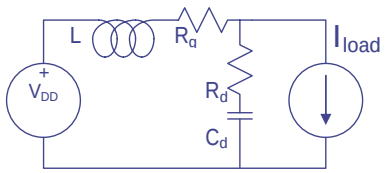

Fig 1: Canonical Power Circuit

Thus the accurate estimation of circuit decoupling capacitance is important for accurate modeling, analysis and optimization of power distribution and interconnect electromigration reliability. On-chip decoupling capacitance comes from the following four sources:

Tub Capacitance: This is the capacitance contributed by the deep tub implants in a CMOS bulk (i.e. non-SOI) process. Since the tubs are typically held at fixed potentials, this capacitance is constant and easy to estimate from the area and perimeter of the tubs, and we will not discuss it further in this paper.

Dedicated Decoupling Capacitance: This is the capacitance contributed by special-purpose decoupling capacitance structures. This capacitance is also constant and easy to characterize and thus we will not discuss it further either.

Device Capacitance: This is the capacitance contributed by the individual devices that compose the circuit. Since these devices change state (i.e. maybe on or off depending on their terminal voltages) this capacitance is variable. This work focuses on estimating this component of the decoupling capacitance. To our knowledge, this is the first work in this direction.

Interconnect Capacitance: This is the capacitance associated with the wires connecting the various components, and -to a lesser extent- the wires that comprise the power grid. The methods developed in this work extend naturally to handle this capacitance and thus we will not single out further.

In the remainder of the paper, we discuss techniques for modeling the decoupling capacitance of non-switching circuit blocks. The paper is organized as follows. In the following section, we analyze the equivalent capacitance of a MOS device. In Section 3, we propose first a pattern-dependent simulation-based and second an analytical model order reduction based method for decoupling capacitance estimation. In Section 4, we show that for large circuit blocks, decoupling capacitance is fairly insensitive to both pattern and process variations. Based on these observations, we develop a fast but accurate static decoupling capacitance estimation method in Section 5. Finally, our conclusions are summarized in Section 6.

2. MOSFET Device Capacitance

In this section, we model a non-switching MOS device with a network of linear resistances and capacitances. The resulting RC network is then simplified to obtain equivalent resistance and capacitance of a MOS transistor.

The various capacitances and resistances associated with a MOSFET device can be represented by an RC network shown in Figure 2 [14]. The RC network includes: 1) C_{GD} – the gate to drain capacitance 2) C_{GS} – the gate to source capacitance 3) C_{GB} – the gate to body capacitance 4) C_{DB} – the drain to body capacitance 5) C_{SB} – the source to body capacitance and 6) R_{DS} – the drain to source resistance. These components depend on the terminal voltages and can vary by orders of magnitude. To understand the relative contribution of the various components and their dependence on voltages, we computed the NMOS device capacitance and resistances for different steady state terminal voltages. Table I shows the normalized NMOS parameters in an industrial 130nm, 1.2V bulk process for different gate, source and drain voltages. The body is tied to the ground. The table shows the five possible NMOS terminal voltage configurations under the assumption that all node voltages can either be at V_{DD} (logic level 1) or ground (logic level 0), and that no steady state current is flowing through the device – with the exception of subthreshold leakage current.

To develop an insight into the equivalent resistance and capacitance of a MOSFET, we reduce the capacitive network of Figure 2 to a single capacitance. Figure 3 shows the stepwise reduction process which results in the parallel combination of a resistance and a capacitance. Table I also shows the equivalent capacitance as computed by this reduction method. We observe a significant variation in capacitance and resistance values based on device terminal voltages. Hence, any decoupling capacitance modeling approach should take this effect into account.

3. Pattern-Dependent Decap Model

In Section 2, we showed that the capacitance of a MOSFET varies significantly with terminal voltages. This leads us to believe that the equivalent decoupling capacitance of a non-switching circuit block is dependent on input pattern. In this section, we propose first a pattern-dependent simulation-based method, then an analytical model order reduction based method for decoupling capacitance estimation.

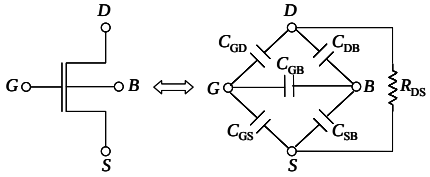

Fig 2: Resistance and Capacitances associated with a MOSFET

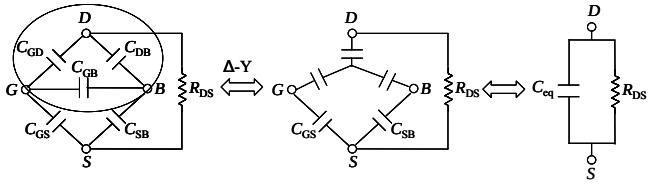

Fig 3: Equivalent RC of a MOSFET

Table I: Normalized NMOS RC parameters for different terminal voltages

G-S-D	C_{GD}	C_{GS}	C_{GB}	C_{DB}	C_{SB}	R_{DS} (Ω)	C_{eq}
0-0-0	1	1	0.366	1.083	1.083	1.08×10^{5}	1.04
0-0-1	1	1	0.366	0.608	1.083	6.05×10^{5}	0.89
0-1-1	1	1	0.366	0.608	0.608	1.71×10^{21}	0.80
1-0-0	1.86	1.86	0.005	1.22	1.22	1.21	1.54
1-1-1	1	1	0.233	0.608	0.608	1.56×10^{7}	0.80

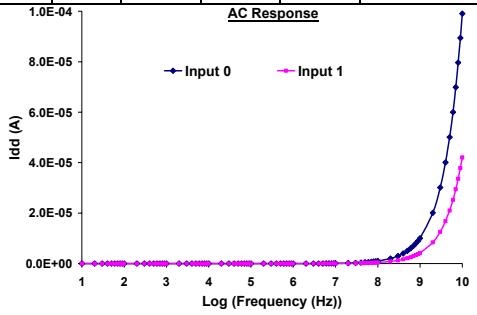

Fig 4: AC Response of a CMOS inverter

Table II: Pattern-dependent equivalent RC computed using our simulation based and analytical methods

Cell-i/p	Simulation		Analytical	
	R(MΩ)	C(fF)	R(MΩ)	C(fF)
Inv-0	0.7736	1.581	0.8085	1.582
Inv-1	10.55	0.6689	1.110	0.6690
Nand-00	0.3868	3.162	0.4043	3.165
Nand-01	0.2906	2.605	0.2963	2.609
Nand-10	0.2981	2.501	0.3102	2.505
Nand-11	0.2381	1.336	0.2542	1.336

3.1 Simulation-Based Decap Model

In order to motivate an equivalent simple model for decoupling capacitance we perform an AC analysis of a CMOS inverter and measure the power supply current I_{dd} as a function of frequency. This is motivated by the fact that power supply voltage fluctuations are expected to be small (~10%) and thus the linearization assumed in the AC analysis is valid. Figure 4 shows the magnitude of I_{dd} vs. frequency for the inverter input at zero and at one, with a unit 1 V AC excitation in series with the power supply. The response indicates that a simple one-pole high-pass circuit would suffice to model the inverter. For the prediction of power-supply induced noise, one can dispense with the resistor since its current is typically much smaller than that drawn by other switching components. The high-pass response of inverter further supports the RC model of the MOSFET discussed in Section 2 and shown in Figure 3.

Once AC analysis is performed, the parameters of the equivalent parallel RC model can easily derived by setting $y = I_{dd}^2$ and $x = (2\pi f)^2$ then performing a linear regression fit of y vs. x, i.e. $y \approx a\,x + b$ from which we find that the coefficient $a = C^2$ and $b = (1/R^2)$. Table II shows R and C values corresponding to the waveforms in Figure 4 and includes similar results for a 2-input NAND gate. Both R and C vary significantly with circuit state (R ≈ 8.2×, C ≈ 2.3×).

3.2 Analytical Decap Model

The simulation-based method outlined above would be useful in situations where a detailed and accurate model is desired, and is easily integrated into a standard cell characterization flow. In this section, we propose a methodology for the analytical modeling of decoupling capacitance. We showed in Section 2 that a MOSFET can be represented by an RC network. So for our analytical modeling methodology, we simply replace each transistor with its equivalent linear RC network. The resulting RC circuit can then be reduced to obtain effective decoupling capacitance of a circuit block.

The capacitive and resistive parameters in the equivalent RC network are functions of device terminal voltages. Hence, in order to have an accurate equivalent RC representation of a device, we need to estimate the terminal voltages of the devices in the circuit accurately. This is typically done by switch level simulation which allows us to compute the internal node voltages in a circuit for a given input vector in an efficient manner [15]. In order to simplify the process, we further assume that all internal nodes attain full logic levels and in a stack, the entire voltage drop is in the uppermost "off" device. Under these assumptions, each NMOS and PMOS device in a circuit can have five possible node voltage configurations. The five configurations of the NMOS device were shown in Table I. PMOS devices have a similar set of five states which can be obtained by reversing all the node voltages.

For analytical modeling of decoupling capacitance, the per-unit width resistance and capacitance parameters of NMOS and PMOS devices for the five possible terminal configurations can be pre-characterized in a manner similar to Table I. All capacitors increase linearly with the width of a device while the resistance varies inversely with the width. Hence, the RC parameters of an arbitrary width device can be easily computed by scaling the per-unit width parameters. Now, for a given circuit and a given input vector, a switch-level simulation is performed to determine the internal node voltages and thereby the state of each transistor in the circuit. Based on the state of the device, each transistor is replaced by its corresponding RC network. The resulting network is linear and hence can be easily reduced by any of a number of model-order reduction techniques [16, 17, 18, 19] to obtain the equivalent R and

C values. Figure 5 shows this methodology applied to an inverter with its input at zero.

We implemented the above analytical decoupling capacitance estimation approach and we compare the results with the simulation based method discussed in Section 3.1. Table II shows the comparison for an inverter and a 2-input NAND gate for all the relevant input patterns. For this experiment, the RC network was reduced using HSPICE simulations but, as stated above, any model-reduction technique can be used. Figure 6 compares the AC response of original NAND-00 gate with that of the equivalent RC network. The figure shows that two curves are practically identical which shows the accuracy of this analytical modeling method.

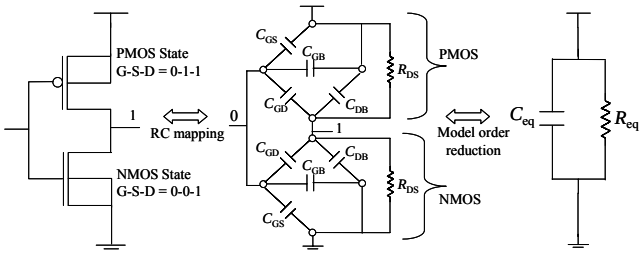

Fig 5: Analytical decoupling capacitance estimation for an inverter at input 0

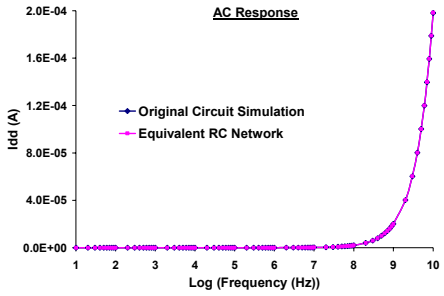

Fig 6: Comparison of AC response of original NAND00 gate with the response of equivalent RC network

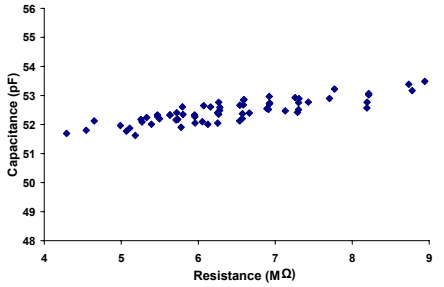

Fig 7: Decap resistance and capacitance for the C1355 circuit for different input patterns

4. Circuit State and Process Dependence

Often, it is more useful to have a method for estimating a state-independent decoupling capacitance for a circuit since the number of possible states grows exponentially with the size of the circuit. In order to explore this aspect, we performed the AC-based characterization outlined in Section 3.1 on the ISCAS 85 combinational benchmark circuits for 64 unique random input vectors each, mapped using a 0.18µm static CMOS cell library. For each of the vectors we computed the equivalent decoupling R and C. Figure 7 shows a pair-plot of the resulting resistance and capacitance values for the C1355 benchmark circuit. Note that while the resistance R varies by about a factor of two, the capacitance C varies by less than 1% which is a strong indication that a pattern-independent model for the capacitance is possible. Note also that the results may be function of the circuit family used to map the benchmark circuit, and that an implementation using different techniques (e.g. dynamic CMOS) may not show these same trends. The ubiquity of static CMOS, however, makes these results of general interest nevertheless.

We did a similar analysis for all 10 of the ISCAS 85 combinational benchmark circuits. For each circuit, we computed the coefficient of variance (COV, defined as the standard deviation divided by the mean and expressed as a percentage) of the equivalent R and C for the 64 unique random input vectors. Table IV shows those coefficients and confirms that the decoupling capacitance is substantially constant, with a coefficient of variation at least an order of magnitude less than that for the resistor. Recall that we are not very interested in the value of the resistance since it does not impact power supply noise.

We have seen that decoupling capacitance for large circuit blocks is fairly insensitive to input pattern. Now we study the sensitivity of decoupling capacitance to process variation. To check this sensitivity, we computed the decoupling capacitance of C1355 benchmark circuit at 58 different process corners. Figure 8 shows the resulting capacitance values (averaged over the same 64 random input vectors) for the different process corners. Figure 8 shows that –somewhat surprisingly- the decoupling capacitance is also fairly independent of technology variations.

Table IV: Coefficient of variance for R and C of ISCAS 85 benchmark circuits for 64 random input vectors

Circuit	COV (R)	COV (C)
C432	27.99	1.12
C499	13.84	1.16
C880	25.23	1.60
C1355	16.46	0.77
C1908	11.30	0.38
C2670	9.73	0.45
C3540	14.31	0.60
C5315	10.51	0.61
C6288	24.23	0.69
C7552	4.73	0.34

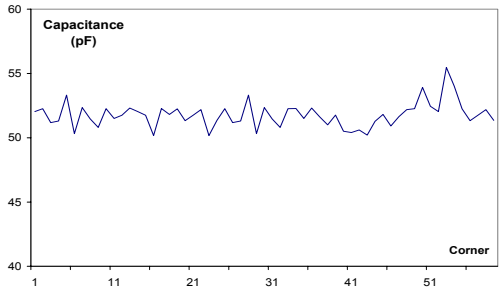

Fig 8: Decap resistance and capacitance for the C1355 circuit for 58 different process corners

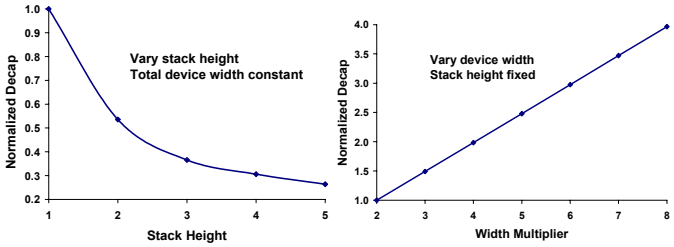

Fig 9: Dependence of decoupling capacitance on stack height and channel width

5. Static Decoupling Capacitance Model

In previous section, we showed that decoupling capacitance is not very sensitive to either circuit state or process variations. This leads us to believe that we can develop a simple state-independent (static) model that can be used to quickly estimate the decoupling capacitance of non-switching circuit blocks. In order to motivate a simple first order model for the decoupling capacitance of the circuits, we make two observations:

- All the capacitance components associated with a MOSFET are proportional to the channel width of the MOSFET.
- The capacitance of identical MOSFETs stacked in series is expected to be inversely proportional to the number of MOSFETs (i.e. the stack height).

To verify our claim that the decoupling capacitance is directly proportional to device width and inversely proportional to stack height, we perform a simple simulation experiment. We consider a stack of NMOS transistors with all the gate terminals tied together. The source terminal of the bottom-most device is connected to ground while the drain of the uppermost device is connected to an AC source. We vary stack height and device width and compute the equivalent decoupling capacitance using the simulation based method described in section 3. Figure 9 shows the results of the experiment. The figure shows that the decoupling capacitance has a linear dependence on total device width and an inverse dependence on stack height.

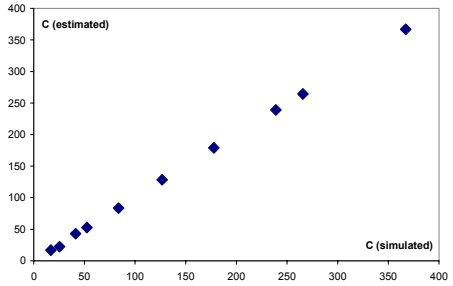

Fig 9: Fit of static decap model for ISCAS benchmarks

Table V: Results of synthetic benchmark circuits

Circuit	W_N (µm)	W_P (µm)	C (pF)	% Error
C1355	7458	17478	52.45	-2.0
Variant 1	10503	17163	60.9	4.08
Variant 2	10599	17235	61.7	3.5
Variant 3	10383	17073	60.5	3.85

Based on the two observations above, we propose expressing the equivalent decoupling capacitance of a CMOS as:

$$C = \alpha_N \sum_{n \in N} \frac{W_N}{H_N} + \alpha_P \sum_{n \in P} \frac{W_P}{H_P} \qquad (3)$$

Where N and P are the set of all N-channel and P-channel devices, respectively, W_d is the width of device d and H_d is the height of the stack in which the device is instantiated.

We applied Equation 3 to the ISCAS benchmark combinational circuits simulated previously and found that -for this technology and cell library- the model parameters $\alpha_N = 4.02$ *fF/µ* and $\alpha_P = 1.21$ *fF/µ*. The resulting fit is illustrated in Figure 9.

In order to further test the model, we created three synthetic benchmark circuit using the same technology and cell library starting from the C1355 circuit and randomly replacing gates by alternates (e.g. a 3-input NAND gate might get mapped to any other 3-input gate in the library). The decoupling capacitance predicted by the model was always within 4% of the average decoupling capacitance predicted over 64 random input vectors for each of the circuits. The detailed results are shown in Table V, where we show the original circuit and the three variations, the weighted device widths W_N and W_P, the measured capacitance and the error in the linear model of Equation 3.

6. Conclusions

Decoupling capacitance is crucial to the control of power-supply induced noise in high power/performance CMOS circuits. In this paper, we presented simulation-based and analytical methods for accurate estimation of decoupling capacitance of non-switching circuits. We showed that decoupling capacitance of circuits is insensitive to input pattern and technology variations. Based on this result, we proposed a simple first order model which can be used to generate quick estimates of decaps for large circuits.

References

[1] K. Diefendorff, "Power4 Focuses on Memory Bandwidth" *Microprocessor Report*, Oct 1999.

[2] H. Chen and D. Ling, "Power Supply Noise Analysis Methodology for Deep-Submicron VLSI Chip Design", *DAC*, pp. 638-643, 1997.

[3] A. Dharchoudhury *et al.*, "Design and Analysis of Power Distribution Networks in PowerPC Microprocessors", *DAC*, pp. 738-743, 1997.

[4] S. Nassif and J. Kozhaya, "Fast Power Grid Simulation", *DAC*, pp. 156-161, 2000.

[5] A. Mezhiba and E. Friedman, "Scaling Trends of On-Chip Power Distribution Noise", *IEEE T-VLSI*, vol. 12, pp. 386-394, Apr 2004.

[6] S. Nassif and O. Fakhouri, "Technology Trends in Power-Grid Induced Noise, "*SLIP*, pp. 55-59, 2002.

[7] C. Webb *et al.*, "A 400 MHz S/390 Microprocessor", *IEEE JSSC*, vol. 32, pp. 1665-1675, Nov 1997.

[8] D. Dobberpuhl *et al.*, "A 200-MHz 64-Bit Dual-Issue CMOS Microprocessor," *IEEE JSSC*, vol. 27, pp. 1555-1567, Nov. 1992.

[9] M. Gowan, L. Biro and D. Jackson, "Power Considerations in the Design of Alpha 21264 Microprocessor", *DAC*, pp. 726-731, 1998.

[10] V. Tiwari et al, "Reducing Power in High-Performance Microprocessors", *DAC*, pp. 732-737, 1998.

[11] H. Su, S. Sapatnekar and S. Nassif, "An Algorithm for Optimal Decoupling Capacitor Sizing and Placement for Standard Cell Layouts", *ISPD*, pp. 68-73, 2002.

[12] S. Zhao, K. Roy, C.K. Koh, Decoupling Capacitance Allocation for Power Supply Noise Suppression, *ISPD*, pp. 66-71, 2001.

[13] M. Pant, P. Pant and D. Willis, "On-chip Decoupling Capacitor Optimization using Architectural Level Prediction", *IEEE T-VLSI*, vol. 10, pp. 319 – 326, Jun 2002.

[14] B Razavi, *Design of Analog CMOS Integrated Circuits*, McGraw Hill Series, pp. 29-30, 2001.

[15] R. Bryant. "A Switch-Level Model and Simulator for MOS Digital Systems," *IEEE T-Computers*, vol. C-33, pp. 160-177, Feb 1984.

[16] L.T. Pillage and R. Rohrer, "Asymptotic Waveform Evaluation for Timing Analysis", *IEEE T-CAD*, vol. 9, pp. 352-366, April 1990.

[17] P. Feldman and R. W. Freud, "Efficient Linear Circuit Analysis by Pade Approximation via the Lanczos Process", *IEEE T-CAD*, vol. 14, pp. 639-649, May 1995.

[18] A. Odabasioglu, M. Celik, and L. Pileggi, "PRIMA: Passive Reduced-Order Interconnect Macromodeling Algorithm", *ICCAD*, pp. 58-65, 1997.

[19] C.L. Ratzlaff, N. Gopal, and L.T. Pillage, "RICE: Rapid Interconnect Circuit Evaluator", *DAC*, pp. 555-560, 1991.