Leveraging Symbiotic On-Die Decoupling Capacitance

Michael Sotman
michael.sotman@intel.com
Phone: +972-4-8656539
Fax: +972-4-8655999
Intel Corporation
Haifa, Israel

Mikhail Popovich
nhlover@ece.rochester.edu
Phone: 585 275-1022
Fax: 585 275-1606
University of Rochester,
Rochester, 14627 New York

Avinoam Kolodny
kolodny@ee.technion.ac.il
Phone: +972-4-8294764
Fax: +972-4-8295757
Technion
32000 Haifa, Israel

Eby Friedman
friedman@ece.rochester.edu
Phone: 585 275-1022
Fax: 585 275-1606
University of Rochester,
Rochester, 14627 New York

Abstract— **Estimates of symbiotic on-die decoupling capacitance are provided for well-junction, interconnect, and quiescent circuits. The available symbiotic capacitance is derived, and the intentional capacitance required to obtain a desired supply voltage noise target is determined.**

I. INTRODUCTION[1]

CURRENT consumption is constantly rising with increasing clock frequencies in modern VLSI circuits. The Power Delivery Network (PDN) is required to have a low-impedance resonant-free profile over a wide frequency range. This requirement is achieved by adding decoupling capacitances at the board, package, and die levels. The impedance at low frequencies is associated with board-level components, at middle frequencies with the package, and at high frequencies with on-die elements [1, 2]. The duration of a single instruction execution in a processor clocked at 3 GHz with a pipe depth of 30 stages is

$$d = \frac{pipe\_stages\_\#}{f} = \frac{30}{3*10^9} = 10ns \tag{1}$$

which corresponds to a current frequency component of 100 MHz. Package decoupling capacitors, however, are typically effective up to 10-20MHz, so these capacitances cannot filter out the frequency component at 100 MHz. Hence, on-die decoupling capacitance is essential.

There are two kinds of on-die decoupling capacitance: intentional and symbiotic. Intentional decoupling is created by specially-designed structures which are placed on the die and connected between the power and ground rails. For example, a MOS transistor gate capacitance can be employed, but it requires die area and increases leakage current. Thus, intentional capacitance usage is expensive and undesirable. The other kind is symbiotic capacitance which always exists on chip [3, 4, 5]. This symbiotic capacitance is comprised of transistor, interconnect, and well-to-substrate capacitances. As the activity factors in digital circuits are quite low, the symbiotic on-die transistor and interconnect capacitance at any given cycle is provided by idle circuits. Moreover, in typical circuits it takes less than 10% of the clock cycle for a cell to change its logical state. The on-die power grid provides an electrical connection between an active (switching) circuit which consumes current and idle (non-switching) circuits which temporarily provide symbiotic decoupling capacitance as shown in Figure 1.

A simplified equivalent circuit is presented in Figure 2. Here, the idle circuit is represented by an effective decoupling capacitance C_{eff}, the on-die power grid is assumed to be an ideal short-circuit, and the off-die PDN is modeled by an inductance L_{PDN}. Assume that L_{PDN} prevents current passing from the battery to the current-consuming switching circuit at high frequencies. Hence, the decoupling capacitance supplies current by releasing a stored charge ΔQ, with a corresponding voltage drop ΔV. The supplied charge quantity is

$$\Delta Q = C_{eff} * \Delta V. \tag{2}$$

Information characterizing the consumer circuit current consumption permits the required charge to be determined from

$$\Delta Q_{req} = \int I(t)dt. \tag{3}$$

To prevent circuit failure, ΔV must be limited. A typical design target is about 5-10% of the nominal supply voltage. The required effective decoupling capacitance can be determined to satisfy the design target.

The nature of symbiotic on-die capacitances is analyzed in this paper and an approach to estimate the capacitance is presented. An analysis of symbiotic capacitance is provided in section II, and the amount of intentional capacitance required to achieve a desired supply voltage noise target is discussed in section III.

II. ANALYSIS OF SYMBIOTIC CAPACITANCE

A first-order evaluation of symbiotic on-die capacitances based on an inverter chain was discussed in[3]. The value of a "symbiotic bypass capacitance" was estimated there as half the load capacitance of a quiescent circuit. A refinement to the model is discussed in this section. A detailed analysis of the current paths within a simple inverter chain is presented in subsection A, and an estimate of the effective decoupling capacitance is described. Considering the resistive elements of each current path, the derived capacitance is useful over the whole frequency range of interest, as discussed in subsection B. The current path analysis extended by generalizing the inverter chain circuit to a logic gate network, the average decoupling capacitance is evaluated using a statistical approach on industrial circuits in subsection C. The result is about 21% of the total gate capacitance. Other symbiotic decoupling capacitances elements due to well-to-substrate junction and metal interconnect are analyzed subsequently in subsection D.

A. Decoupling current paths in an inverter chain

A simple model of a quiescent inverter chain is shown in Figure 3, where part of an idle circuit operates with stable inputs. In order to evaluate the effectiveness as a decoupling capacitance, a small step change in the supply voltage is applied, and the resulting currents are observed. The currents between the power and ground nets inject some charge which was stored in the circuit.

The current paths which exist in the circuit are noted in Figure 3. In path #1 current flows through C_{gs} of the ON-state P-channel transistor P_i and R_{on} of the previous stage ON-state N-channel transistor N_{i-1}. Path #2 is similar to the first path, but uses R_{on} of P_i and C_{gs} of N_{i+1}. In path #3 current flows through the same C_{gs} of P_i as in path #1, but continues to ground through C_{gb} of the OFF-state transistor N_i. Recall that in the ON-state, the transistor gate capacitance consists mainly of C_{gs} while in the OFF-state this capacitance is connected to the bulk as C_{gb} [3]. Hence, path #3 consists of two capacitors in series, so the effective capacitance is lower than C_{gs}. Paths #1 and #2 together provide a total effective capacitance of C_{gs} (from both the N and P transistors), but the path between the power and ground rails includes R_{on} of the opposite-type ON-state transistor in

---

the previous logic stage. The next subsection examines the speed at which these paths can provide current in response to a change in the supply voltage.

B. Time and frequency domain simulation of decoupling current paths

Normalized simulated current waveforms for the three described paths are shown in Figures 4 and 5, where 500 ps and 5 ps fall times are used in the stimulus ramp function. The stimulus voltage is depicted by a dotted line. Note that path #3 does not play a significant role. Primary paths are path #1 and #2. The time constant $\tau = RC$ of these paths is defined by R_{on} and C_{gs} corresponding to each path. The fast ramp results are presented in Figure 5a and show the currents through paths #1, #2 and #3. A more detailed analysis exhibits a resistive part in this current path due to the polysilicon contact of the transistor gate and the substrate well. The behavior of path #3 allows a faster response to voltage fluctuations.

The simulation performed in the frequency domain uses a similar simulation setup. The step-function time domain voltage source is replaced by a small amplitude AC voltage source in series with a DC voltage source, which is required to maintain a stable operating point. The simulation results are presented in Figure 6. At high frequencies the current in path #3 (C_{gs} of P_i and C_{gb} of N_i) is dominant. The maximal high-frequency component in typical on-die transients can be approximated based on the fastest on-die rise time, which is about 10 ps. The well known empirical formula

$$f_{max} = {0.35}/{t_{rise}} \tag{4}$$

yields an f_{max} of 35GHz. This value has the same order of magnitude as the crossing point of curves shown in Figure 6, therefore both paths types are relevant for C_{eff}. Process scaling won't change this relation since transistor switching time is scaled as well as the capacitances.

It can be concluded from these studies that in a non-switching ON-state transistor, the gate capacitance provides decoupling. Design for equal rise and fall delays leads to an N-channel to P-channel transistor width and gate capacitances ratio of 0.7. The ON-state transistor operates in the linear region, where the total gate capacitance C_{gate} is equally divided between the gate-source capacitance C_{gs} and the gate-drain capacitance C_{gd}, $C_{gs} = C_{gd} = 0.5*C_{gate}$. Decoupling capacitance coefficient K_{eff_INV} for an inverter is:

$$K_{eff\_INV} = \frac{C_{eff}}{C_{gate\_total}} = \frac{C_{gsP}*0.5 + C_{gsN}*0.5}{C_{gate\_P} + C_{gate\_N}} = \frac{0.5*C_{gate\_P}*0.5 + 0.7*0.5*C_{gate\_P}*0.5}{C_{gate\_P} + 0.7*C_{gate\_P}} = 0.25. \tag{5}$$

C. Decoupling capacitance in logic gates

An analysis of the symbiotic capacitance within more complicated gates is enabled by extending the aforementioned inverter-based model. An example of a two-input NAND gate is shown in Figure 7. The transistor sizes were chosen to maintain symmetrical rise and fall time. For inputs ab='01, two transistors conduct – P_0 and N_1. Capacitor C_{gs} of transistor P_0 is charged and can operate as a decoupling capacitor. However, C_{gs} of transistor N_1 is not charged and cannot function as a decoupling capacitor. Thus, in this case there is only one active decoupling capacitor. The normalized width of this transistor is 1 and the probability of operation with this input combination is 0.25. The same analysis is performed for all other possible input combinations and a weighted sum is determined. The expression for decoupling capacitance usage effectiveness coefficient K_{eff_NAND2} for a two-input NAND is:

$$K_{eff\_NAND2} = \frac{C_{eff}}{C_{gate\_total}} = \frac{0.5*0.25 + 1.2*0.25 + 1*0.25 + 1.4*0.25}{1*2 + 1.4*2} \cong 0.21. \tag{6}$$

The same approach can be applied to other logic gate types. The results are presented in Table 1. The general dependence is that more complicated gates produce a lower K_{eff}. The K_{eff} of any circuit can be determined as a weighted average from knowledge of the probability of each gate type in the circuit netlist. Such a calculation was performed based on statistics of 74 independently synthesized functional blocks from an industrial processor circuit in 65 nm technology, with a total gate count above 40000 gates. Results of this statistical calculation together with decoupling capacitance effectiveness are presented in Table 2. The final result is slightly above 21% of total circuit gate capacitance.

This statistical approach can be further developed. A statistical circuit analysis is required to provide signal probabilities for each specific net to operate in '0 or '1 logic state [6]. As activity factors are often quite low in modern VLSI circuits [7], each net is in a stable state most of the time and this state can be set to increase the decoupling capacitance effectiveness. However, other design considerations such as leakage current may be affected.

D. Other decoupling capacitance components

N-well and P-well areas are typically patterned for P-channel and N-channel transistors as shown in Figure 8. The P-substrate and P-well are connected to VSS while N-well is connected to VCC, a reverse bias across the junction between P-substrate and N-well. The capacitance of this P-N junction is connected between VCC and VSS and plays a decoupling role. The capacitance consists of two main components – an area capacitance and a perimeter capacitance. Area capacitance is significantly higher than the perimeter. The P-well has no substrate capacitance. The main factor is the N-well area which can be calculated easily from the N-well map of the die as shown in Figure 9. The N-well and P-well areas are typically designed as interleaved rows and the areas are equal, so the overall N-well area is approximately half of the total area,

$$C_{well} \cong C_{well\_unit\_area} * {A_{die}}/{2}. \tag{7}$$

The interconnect decoupling capacitance is treated as a decoupling capacitance because each line is connected to VCC or VSS either directly or via
an N-channel or P-channel transistor. Thus a path between VCC and VSS always exists for any charged interconnect capacitor. All of the possible interconnect capacitor cases are presented in Figure 10. Note that capacitor $C_{VCC\text{-}VSS}$ is always charged and can behave as an active decoupling capacitance. Other capacitances shown in Figure 10 are between signal line S0 and the other nets in the system. $C_{S0\text{-}VCC}$ and $C_{S0\text{-}VSS}$ have the same value and equal probability to be charged to either '0 or '1. Therefore only one of these capacitors has to be considered. A similar situation exists with the other capacitance pair – $C_{S0\text{-}S1}$ and $C_{S0\text{-}S2}$. Assuming equal probability of the signal leads to

$$C_{interconnect} = C_{VCC\_VSS} + {C_{signal\_PDN}}/{2} + {C_{signal\_signal}}/{2}. \tag{8}$$

Relations between C_{gate}, C_{well} and $C_{interconnect}$ are constant in a typical circuit. All three components of the symbiotic decoupling capacitance are interrelated. Generally, the well area and interconnect properties are a function of the gate area, and linear relations of the form $C_{interconnect} = \alpha*C_{gate}$ and $C_{well} = \beta*C_{gate}$ are assumed. Typical α and β are based on extracted data for a specific technology process and design style. Typical values for a modern 65 nm process are 0.7 nF/mm^2 for the gate capacitance, 0.34 nF/mm^2 for interconnect and 0.2 nF/mm^2 for the well capacitance. The resultant coefficients are $\alpha = 0.48$ and $\beta = 0.29$. These coefficients are recalculated for each specific process or design style. For example, design corners in gate limited or interconnect limited designs have different values of α and β.

III. DESIGN IMPLICATIONS

Package and board capacitors are insufficient for high frequency decoupling operation. Their parasitic impedances place a requirement for on-die decoupling to provide current transients to maintain proper circuit operation, without producing excessive noise in the on-chip power grid. Charge required for the switching circuits is

$$Q_{switch} = af * (C_{gate} + C_{interconnect}) * V. \tag{9}$$

In this expression, *af* is the activity factor and reflects the average ratio of switching nets to the overall number of gates. Charge supplied by decoupling capacitances (including both symbiotic and added intentional decoupling capacitors) is

$$Q_{decoupling} = [(1 - \frac{af}{ispc}) * (K_{eff} * C_{gate} + C_{interconnect}) + C_{well} + C_{intentional}] * \Delta V. \tag{10}$$

ΔV in this expression is the voltage noise generated by the circuit switching activity. The gate and interconnect components of the decoupling capacitance are changed by *1-af/ispc*, where *ispc* is *inversion stages per cycle*, since a switching circuit is in a transient state only during a small fraction of the clock cycle. With $af \approx$ 1-5% and $ispc \approx$ 10, *1-af/ispc* approaches one. $C_{intentional}$ can be denoted as $C_{intentional} = \chi * C_{gate}$. Simplifying and equating these charge expressions yields

$$af * (C_{gate} + \alpha * C_{gate}) \cong (K_{eff} * C_{gate} + \alpha * C_{gate} + \beta * C_{gate}) + \chi * C_{gate}) * \Delta V / V. \tag{11}$$

Hence

$$\frac{\Delta V}{V} = \frac{af * (1 + \alpha)}{K_{eff} + \alpha + \beta + \chi}. \tag{12}$$

In order to maintain the power supply noise below ΔV_{max}, intentional on-die decoupling capacitance $\chi * C_{gate}$ is added, satisfying

$$\chi \geq \frac{af * (1 + \alpha)}{\frac{\Delta V_{max}}{V}} - (K_{eff} + \alpha + \beta). \tag{13}$$

For the case where K_{eff} = 0.21, α = 0.48 and β = 0.29 with an activity factor *af* = 5% and voltage noise target $\Delta V_{max}/V$ = 5%, the required intentional decoupling coefficient is χ ≥ 0.12. Assuming *af*= 4% leads to a negative value χ ≥ -0.176, therefore, intentional decoupling capacitance is not necessary.

IV. CONCLUSION

Symbiotic decoupling capacitance in VLSI circuit is considered in this paper. Gate, interconnect and N-well to P-substrate capacitances are addressed as the primary symbiotic decoupling capacitance components. All of these capacitances are effective over the relevant frequency range, despite series resistances. A statistical characterization of large static CMOS circuits in 65 nm technology has been performed, showing that 21% of the total circuit gate capacitance can serve as an effective decoupling capacitance, in addition to a significant fraction of the interconnect capacitances (equivalent to 48% of the total gate capacitance). The N-well to P-substrate capacitance is equivalent to 29% of the total gate capacitance. Analysis of complex gates, memory cells, domino logic structures, and other gates can be performed similarly with this approach. Based on these results, intentional decoupling capacitance required to satisfy a supply noise design target can be determined for any circuit with a known activity factor.

REFERENCES

[1] A. Mezhiba and E. Friedman, *"Power Distribution Networks in High Speed Integrated Circuits"*, Kluwer Academic Publishers, 2004.
[2] A. Waizman and C.-Y. Chung, "Resonant free power network design using extended adaptive voltage positioning (EAVP) methodology", *IEEE Transactions on Advanced Packaging*, volume 24, issue 3, Aug 2001, Page(s):236-244.
[3] W. Dally and J. Poulton, "Digital systems engineering", Cambridge university press, 1998, Page(s):160-161, 244-245.
[4] H.H. Chen, J.S. Neely, M.F. Wang and G. Co, "On-chip decoupling optimization for noise and leakage reduction", *Integrated Circuits and Systems Design*, 2003, Proceedings 16th Symposium on 8-11 Sept. 2003, Page(s):251 – 255.
[5] H.H. Chen and J.S. Neely, "Interconnect and circuit modeling techniques for full-chip power supply noise analysis", Components, Packaging, and Manufacturing Technology, Part B: Advanced Packaging, *IEEE Transactions on Components, Hybrids, and Manufacturing Technology*, Volume 21, Issue 3, Aug. 1998, Page(s):209 – 215.
[6] SYNOPSYS VCS tool, http://synopsys.com/products/simulation/simulation.html.
[7] N. Magen, A. Kolodny, U. Weiser and N. Shamir, "Interconnect-power dissipation in a microprocessor", *International Workshop on System Level Interconnect Prediction (SLIP)*, February 2004, pages: 7-13.

Table 1: Effective decoupling capacitance per gate calculation

Cell type	C_{gate} eff	C_{gs} eff	Total C	$K_{eff\ gate}$
BUFF	1.7	0.85	3.4	0.25
INV	.85	0.425	1.7	0.25
NAND2	2.05	1.025	4.8	0.213542
NAND3	3.3375	1.66875	9.3	0.179435
NOR2	2.2	1.1	5.4	0.203704
NOR3	4.05	2.025	11.1	0.182432
CLK BUFF	1.7	0.85	3.4	0.25
Latch	6.025	3.0125	15.3	0.196895
FF	12.5	6.25	30.6	0.204248
Complex				0.15

Table 2: Total effective decoupling capacitance

| Cell type	Cell #	probability	$K_{eff\ gate}$	K_{eff_weight}
BUFF	2453	0.07278	0.25	0.0182
INV	11076	0.27790	0.25	0.0695
NAND2	5871	0.12830	0.213542	0.0274
NAND3	2986	0.04717	0.179435	0.0085
NOR2	4498	0.10836	0.203704	0.0221
NOR3	1785	0.03206	0.182432	0.0058
CLK BUFF	1507	0.04124	0.25	0.0103
Latch	3966	0.12911	0.196895	0.0254
FF	1231	0.02965	0.204248	0.0061
Complex	4777	0.13342	0.15	0.0200
Total	*40150*	*1*		*0.2133*
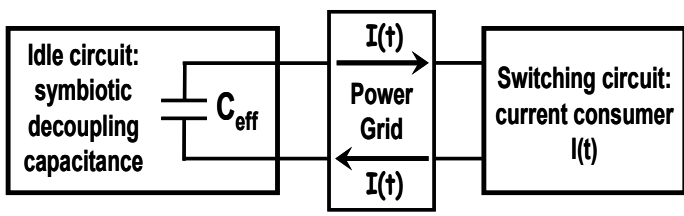
Fig. 1 High level model of on-die current supply structure
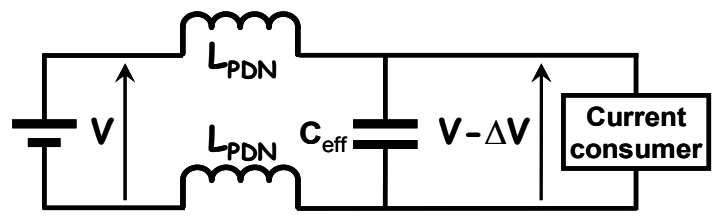

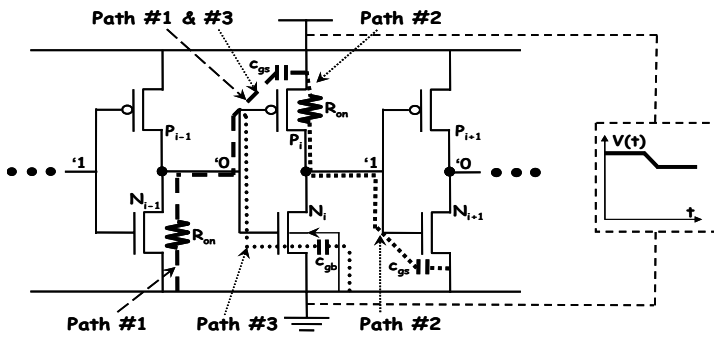
Fig. 2 Simplified power distribution network structure

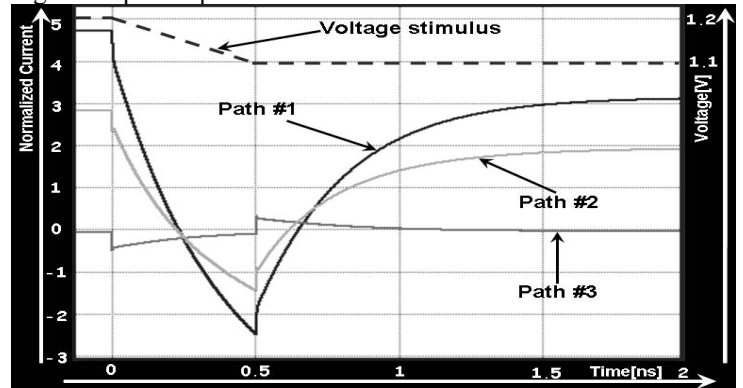Fig. 3 Current paths through symbiotic capacitance
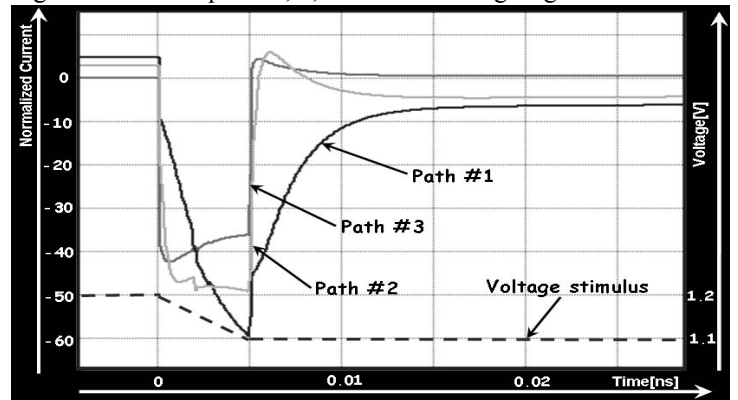

Fig. 4 Currents of paths 1, 2, 3 for slow voltage signal transition

a)full view

Fig. 5 Current waveforms of paths 1, 2,3 for fast signal transition
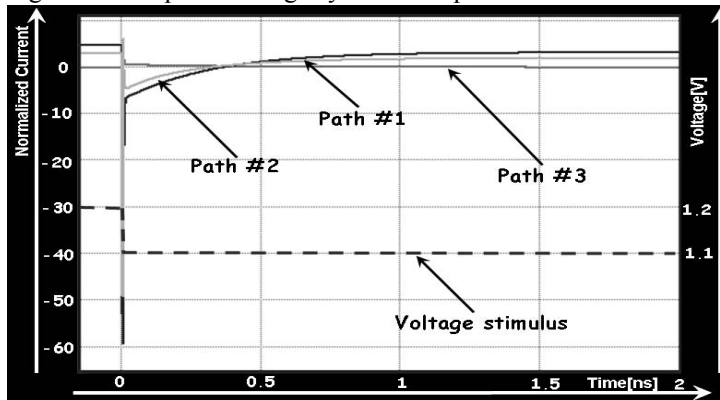
b) Closer view of currents around t=0.
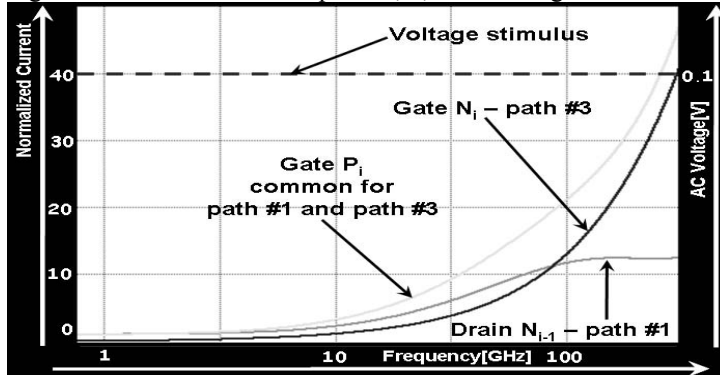
Fig. 6 Current curves of paths #1 and #3 in the frequency domain
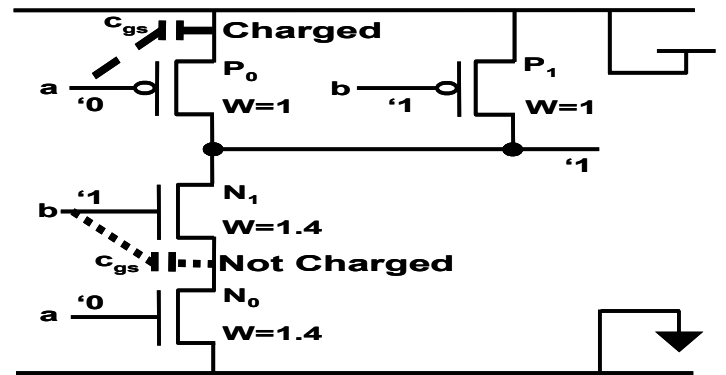
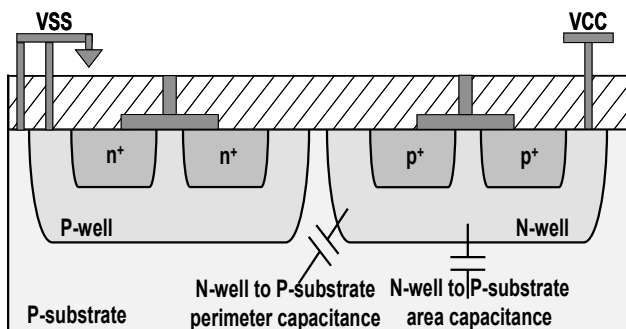Fig. 7 NAND gate with inputs ab = '01

Fig. 8 Cross-section of N-channel and P-channel transistors showing well to substrate capacitances
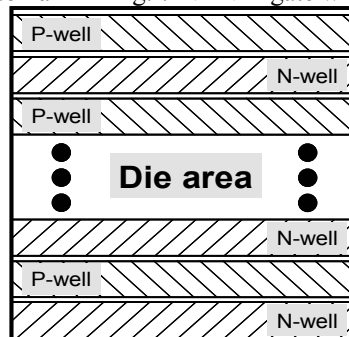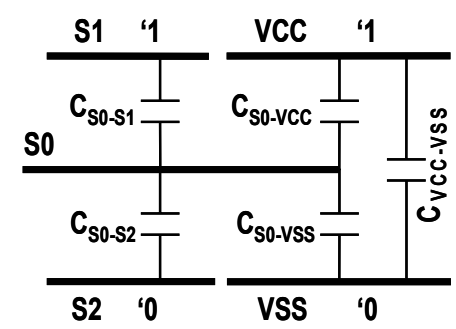

Fig. 9 View of N-wells and P-wells at die level

Fig. 10 General model of interconnect decoupling capacitances

4