5. Properties and modeling of on-chip Power Distribution Networks

- Decoupling capacitance
Electrical properties of on-chip PDN: capacitance

Capacitance associated with PDN:
- Gates
- Interconnects
- Well (in standard CMOS)

Effects
- Provide current to close switching gates
- Reduce the loop area where the switching current flows
- As a consequence: reduce the peak current in the power/ground pins and reduce PSN

However the intrinsic capacitance is not enough to reduce PSN to the required level

Intentionally added Decoupling capacitors (Decaps)
Intra-block decaps

• Intra-block decaps have constrained dimensions
  – For example, they are implemented in the standard-cell template

• They are usually implemented as MOS capacitors
MOS decaps

“tie-off” cell (Veendrick00)
MOS decaps

- N-well
- V_{DD}
- P diffusion
- N diffusion
- GND
- row
- height

Diagram showing the layout of MOS decaps with labels for the diffusion areas and connections labeled as V_{DD} and GND.
Decap Model

NMOS decap:

Model characteristics:
- Distributed RGC model to take into account HF effects
- Gate leakage modeled by a voltage-dependent current source
Decap Model (cnt’d)

Model simplification:
- Exploit symmetry of the decap
- Poly gate resistance ($r_G$) << channel resistance ($r$)
- Channel-to-substrate capacitance ($C_B$) neglected
Model of a PMOS decap

- PMOS decap has the same model as the NMOS decap
- N-well decap in parallel; two decap structures with different parameters, coupled through the boundary
Electrical model of MOS decaps

\[ \frac{\partial^2 v}{\partial x^2} = r_c \frac{\partial v}{\partial t} + (I_0 + g v) r \]

Channel term

Gate leakage term

- Voltage (linear scale)
- Gate current (log. scale)
- Gate current (linear scale)

\( i(v) \) is linearized around \( V_{DD} \)

\( i(v) \approx I_0 + g v \)
**Example: response of a MOS decap to DC bias + AC excitation**

In steady state, the solution can be separated into DC solution + AC solution.

**PDE:**
\[
\frac{\partial^2 v}{\partial x^2} = r_c \frac{\partial v}{\partial t} + (I_0 + g v) r
\]

**BC:**
\[
v(-L, t) = v(+L, t) = V_{DD} + V_M e^{j\omega t}
\]

**DC Solution:**
\[
v_{DC}(x) = \left( V_{DD} + \frac{I_0}{g} \right) \frac{\cosh(x\sqrt{gr})}{\cosh(l\sqrt{gr})} - \frac{I_0}{g}
\]

**AC Solution:**
\[
v_{AC}(x, t) = \text{Re} \left\{ V_M \frac{\cosh(x\sqrt{gr} + j\omega rc)}{\cosh(l\sqrt{gr} + j\omega rc)} e^{j\omega t} \right\}
\]
Example DC response: Impact of technology scaling

Drop voltage along the channel increases as $t_{ox}$ is reduced

Drop voltage along the channel increases with channel length

Normalized voltage along the channel

Normalized distance along a half of channel length length

$L = 10\mu m, W = 3\mu m,$

$L = 10\mu m$

$W = 3\mu m, 65\text{nm technology}$

$W = 3\mu m, 65\text{nm technology}$

$L = 20\mu m$
DC response: Impact of technology scaling

Loss of charge due to gate leakage

WIDTH = 3\mu m, L = \{ 5\mu m, 10\mu m, 15\mu m, 20\mu m \}

\[
\frac{Q_L}{Q_{\text{non-L}}} = \begin{cases} 
0.95 & \text{for technology } 45 \text{ nm} \\
0.9 & \text{for technology } 65 \text{ nm} \\
0.85 & \text{for technology } 90 \text{ nm} 
\end{cases}
\]
Example AC response. No leakage case

\[ v_{AC}(x,t) = A(x) \cos(\omega t + \phi) \]

Amplitude \( A(x) \) changes along the channel
It depends on \( r \) and \( c \) as well as \( \omega \)

at low \( \omega \): \( A(x) \rightarrow V_M \Rightarrow \) decap
behaves as a lumped element
Example AC response. No leakage case

\[ v_{AC}(x, t) = A(x) \cos(\omega t + \phi) \]

Amplitude \( A(x) \) changes along the channel
It depends on \( r \) and \( c \) as well as \( \omega \)

\[ v(x, t) \]

at higher \( \omega \Rightarrow A(x) \approx V_M e^{-\delta \sqrt{\frac{\omega r c}{2}}} \], where \( \delta = L - x \)

\[ \Rightarrow \delta = \sqrt{\frac{2}{\omega r c}} \]

the amplitude decreases \( e^{-1} \) times

"skin depth"

If \( \delta \leq L \) (at \( \omega \geq \omega_c \)), decap AC performance decreases.
Example AC response. No leakage case

\[ v_{AC}(x,t) = A(x) \cos(\omega t + \phi) \]

Amplitude \( A(x) \) changes along the channel
It depends on \( r \) and \( c \) as well as \( \omega \)

If \( \delta \ll L \) (at \( \omega \gg \omega_c \)), decap AC performance is very low.
Example AC response. Leakage case

\[ v_{AC}(x,t) = A(x) \cos(\omega t + \phi) \]

Amplitude \(A(x)\) changes along the channel
It depends on \(r, c\) and \(g\) as well as \(\omega\)

At low \(\omega\) is only dimmed by \(g\) and \(r\)
Example AC response. Leakage case

\[ v_{AC}(x,t) = A(x) \cos(\omega t + \phi) \]

Amplitude \( A(x) \) changes along the channel. It depends on \( r, c \) and \( g \) as well as \( \omega \).

Now \( \delta_L \) can be approximated by

\[ \delta_L \approx \delta \sqrt{1 + \left(\frac{g}{\omega c}\right) \times \frac{1}{2} \left(\frac{g}{\omega c}\right)^2} < \delta \]
AC response: Impact of technology scaling

Solid blue line: no leakage
Dashed red line: leakage

Decap channel half length $L = 20\mu m$
Frequency response
“Skin depth” (NMOS $W = 3 \, \mu m$, $L = 2.8 \, \mu m$)

Dots: HSPICE  
Lines: equations  
0.35 $\mu m$ technology
Input Impedance of a MOS Decap

\[ Z_{IN} = Z_0 \coth(\gamma l) \]

\[ Z_0 = \sqrt{\frac{r}{g + j\omega c}} \]
\[ \gamma = \sqrt{r(g + j\omega c)} \]

Critical frequency at \( l = \delta \)

- The frequency that separates lumped and distributed behaviour
- Lower critical frequency in case of gate oxide leakage

**NO LEAKAGE:** \( f_C = \frac{1}{\pi rcl^2} \)

**LEAKAGE:** \( f_C^L = \frac{\sqrt{1 - grl^2}}{\pi rcl^2} \)
Normalized $R$ and $C$ as a function of frequency

$$R_{\text{EQU}} = \text{Re}\{Z_{\text{IN}}\}$$
$$C_{\text{EQU}} = \frac{-1}{\omega \text{Im}\{Z_{\text{IN}}\}}$$

normalized $R = \frac{R_{\text{EQU}}}{l_r}$
normalized $C = \frac{C_{\text{EQU}}}{l_c}$
Conclusions

• At high frequency, decaps must be modeled as a distributed structure

• Model based on physical grounds
  • Relevant parameters for each technology node are easily obtained

• Methodology to analyze the decap response
  • Example: analysis of the DC and AC response
  • Critical frequency/dimension expressions define the border between full and reduced decap performance w/wout gate leakage

• Gate leakage decreases the decap performance at low as well as at high frequencies
Allocation of intra-block decaps

Simple strategy (Smith94)

Decap at block $i$ should be able to provide the charge $\delta Q_i = I_i^{avg} / f_{CLK}$ during one clock cycle as the power voltage level varies as $\delta V_{DD}$

$$C_{i \text{decap}} = \frac{\delta Q_i}{\delta V_{DD}} = \frac{I_i^{avg}}{f_{CLK} \delta V_{DD}}$$

Assumptions:

a) during the switching event, the decap is effectively disconnected from the power supply. It provides all the required charge.

b) The decap is fully recharged before the beginning of the next clock cycle.

More elaborate strategy (Zhao02)

1. The circuit without any decap is analyzed and the worst case power noise is determined.
2. Blocks with power noise $V_{noise}$ below the target margin $\delta V$ do not receive additional decaps.
3. In blocks where $V_{noise} > \delta V$, an additional decoupling capacitance is added

$$C_{\text{decap}} = \frac{V_{noise} - \delta V}{V_{noise}} \left( \frac{\delta Q}{\delta V} \right)$$

The additional charge provided by the decap is proportional to the difference $V_{noise} - \delta V$
Allocation of intra-block decaps

A further refined strategy (Su03)

1. For each circuit node determine the value of the following metric:

\[ M_j = \int_0^T \max \left[ \left( V_j(t) - \delta V \right), 0 \right] dt \]

2. Calculate the overall power supply quality as:

\[ M = \sum_j M_j \]

3. Add decaps at the places where \( M_j > 0 \). \( M \) become zero when the power supply margin is satisfied at all times and nodes.
An on-chip voltage regulator (Ang00)
Gated decaps, active decaps (Roy05, Gu06)
Application of the model: inter-block decaps (Rius06)

- Intra-block decaps have constrained dimensions
- Inter-block decaps do not suffer from this constraint
  - Typically, used for EMI reduction purposes
Example Inter-Block Decap: Gate length must be limited
Example Inter-Block Decap: Fingers and Stripes
Inter-Block Decap Model Parameters

- Model parameters are defined to be independent of length and width

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_{\square}$</td>
<td>Channel sheet resistance [$\Omega/\square$]</td>
</tr>
<tr>
<td>$C_{OX}$</td>
<td>Gate capacitance per unit area [F/m$^2$]</td>
</tr>
<tr>
<td>$g_{OX}$</td>
<td>Gate oxide conductance per unit area [S/m$^2$]</td>
</tr>
<tr>
<td>$I_{OX}$</td>
<td>Gate current density per unit area [A/m$^2$]</td>
</tr>
<tr>
<td>$WL$</td>
<td>Total gate area [m$^2$]</td>
</tr>
</tbody>
</table>

Critical frequency

$$f_{C} = \frac{\sqrt{1 - g_{OX} r_{\square} l^2}}{\pi r_{\square} C_{OX} l^2}$$

Total gate-oxide leakage

$$I_{LEAK} = WL \left( g_{OX} V_{DD} + I_{OX} \right)$$
Procedure for Optimum Inter-Block Decap Design

1. Define the total decoupling capacitance \( C_{DEC} \) to be included in the IC

2. Determine the effective total area as

\[
A_C = \frac{C_{DEC}}{C_{OX}}
\]

3. Obtain the gate length of a finger \( L_{F0} \)
   to get the maximum frequency \( f_C \) for
   which the decap needs to perform

\[
f_C^L = \frac{\sqrt{1 - g_{ox} r L_{F0}^2}}{\pi r C_{OX} L_{F0}^2}
\]

4. Define the number of gate fingers as

\[
n = \left\lceil \frac{A_C}{L_{F0}} \right\rceil
\]

5. Obtain the gate length of a single decap as

\[
L_F = \sqrt{A_C} / n
\]

6. Define the number of stripes as
   where \( W_{MAX} \) is the maximum allowed gate width

\[
m = \left\lceil \frac{A_C}{W_{MAX}} \right\rceil
\]

7. Obtain the gate width of a single decap as

\[
W_S = \sqrt{A_C} / m
\]

8. Calculate total leakage

\[
I_{LEAK} = A_C \left( g_{ox} V_{DD} + I_{OX} \right)
\]
Example

- 90nm GP technology
- Required decap $C_{DEC} = 1\text{nF}$
- Three gate-oxide thicknesses

Results:
- Area factor = 1.01 to 1.23
- Total leakage current $I_{LEAK} = 1.1\text{ mA}$