Impact of Stochastic Dopant Variation and the Copper Size Effect on Gigascale Integration

Stefan Schwantes\textsuperscript{1}, Ralf Götsche\textsuperscript{2}, Wolfgang Krautschneider\textsuperscript{1} and Ulrich Ramacher\textsuperscript{2}

\textsuperscript{1}) Technical University of Hamburg-Harburg, Microelectronics
Eissendorfer Strasse 38, 21071 Hamburg, Germany
Schwantes@tu-harburg.de
\textsuperscript{2}) Infineon Technologies, Corporate Research
Otto-Hahn-Ring 6, 81739 Munich, Germany

Abstract

This paper discusses the impact of the parasitic effects on the performance of CMOS circuits. The influences have been studied on a logical unit, an on-chip interconnect scheme and on memory. Minimum feature size for MOSFETs and for interconnects with scaled ranges from 100 nm down to 25 nm have been chosen. Our results show that new parasitic effects, the size effect for copper and the threshold voltage variations due to statistical dopant fluctuations, will partly compensate the performance gain obtained by scaling down the feature size. Below the 50nm-Technology-Node, these effects have to be considered for minimum feature sizes.

1. Introduction

According to the ITRS Roadmap [1] the line width of on-chip interconnects will be scaled down to 25 nm. When the interconnect line width will be shrunk below 100 nm, the size effect on Cu resistance and process induced irregularities will increase the resistance of interconnects. Furthermore, downscaling the channel area results in an increased spread of the threshold voltage due to the statistical dopant fluctuation. This results in a large variation of the threshold voltage over the whole chip. This paper discusses based on simulations the impact of these parasitic effects on the performance of future ULSI systems consisting of logical units, long on-chip interconnects, and memory.

As a representative for a pure digital part, the logical unit, we used a 64 bit adder which is a representative part of every microprocessor. It contains almost 12000 transistors and several registers. For the mixed signal part we used a 1 Mbit SRAM representing the on chip cache memory. Both are connected via on-chip interconnect systems.

The CADENCE design software [2] was used for generating the layout of the 64 bit adder. We developed several libraries with tentative design rules for future technologies with feature sizes between 100 nm and 25 nm based on the ITRS Roadmap.

The parasitic capacitances and resistances of our layouts were extracted by DIVA and the two-dimensional field simulator COEFFGEN [2]. This allows a simulation of distributed RC lines. The circuit simulations were carried out by SPECTRE, a circuit simulator based on SPICE. We used optimized BSIM3 parameters for the circuit simulations. It was therefore possible to extend the range of validity of the BSIM3 model.

The parameters of the submicron devices were determined by extensive device simulations with ATLAS [3] based on the hydrodynamic model. By modifying the mobility parameters, we could achieve a good fit to earlier simulations of 25 nm CMOS transistors [4]. Fig. 1 shows the design flow.

2. Technology Parameters

Table 1 gives an overview of the used technology parameters. The pitch of the local interconnects and their aspect ratios are according to the ITRS Roadmap. We assume that an alternative gate dielectric is available.
This makes an aggressive downscaling of the equivalent gate oxide thickness possible. Otherwise, a downscaling of the oxide thickness below 1 nm will endanger the functionality of the logic [5]. The decreasing $I_{ON}$ to $I_{OFF}$ ratio for smaller channel lengths allows a sufficiently high effective gate voltage even for MOSFETs in the nanometer region.

### Table 1. Technology parameters.

<table>
<thead>
<tr>
<th>Technology</th>
<th>Local Pitch</th>
<th>Aspect Ratio</th>
<th>Supply Voltage</th>
<th>$I_{ON}$ to $I_{OFF}$</th>
<th>Gate Oxide</th>
</tr>
</thead>
<tbody>
<tr>
<td>100 nm</td>
<td>250 nm</td>
<td>2.3</td>
<td>1.0 V</td>
<td>6E4</td>
<td>1.3 nm</td>
</tr>
<tr>
<td>50 nm</td>
<td>120 nm</td>
<td>2.5</td>
<td>0.8 V</td>
<td>1.6E4</td>
<td>1.0 nm</td>
</tr>
<tr>
<td>25 nm</td>
<td>60 nm</td>
<td>2.7</td>
<td>0.6 V</td>
<td>8.7E3</td>
<td>0.8 nm</td>
</tr>
</tbody>
</table>

In the interconnect hierarchy four metal layers for local, three metal layers for intermediate, and two metal layers for global interconnects have been chosen. The simulations were performed with air gap structures as an intermetal dielectricum with a $k$ of 1.7 [6] which is very close to the theoretical limit. Copper has been used as the interconnect metal.

### 3. The Size Effect in Narrow Interconnects

As the interconnect line width shrinks below 100 nm, the size effect of metal interconnects becomes important. The metal line widths approach the mean free path of conducting electrons. This results in a growing influence of surface scattering and grain boundaries which increases the metal resistance.

Fig. 2 Calculated resistance versus line width for copper. The size effect will become more important for smaller feature sizes.

This size effect was theoretically predicted by [7] and experimentally verified by [8,9]. The influences of the rough surfaces with measured ranges of several nanometers and the effects of grain sizes and boundaries on the copper resistance have been proved [9]. According to [9] the combined surface and boundary scattering effect can be summarised as:

$$
\rho_{line} = \rho_{bulk} \left(1 + \frac{(50 \text{nm}) S}{d} + \frac{(25 \text{nm})}{g}\right)
$$

(1)

$\rho_{bulk}$ is the intrinsic sheet resistance of the metal line. $S$ stands for the surface roughness, $d$ for the line width and $g$ for the grain size. Fig. 2 is a plot of equation (1) for copper interconnects with a mean free electron path of 40 nm and a grain size as large as the interconnect width. A significant increase of the resistance for interconnect widths below 50 nm is observed and its effect on the performance of future deep submicron CMOS circuits must be investigated.

### 4. Threshold Voltage Variation

Random microscopic fluctuations of the number and location of dopant atoms in the channel depletion region of a MOSFET induce stochastic variations of the threshold voltage. The standard deviation of the threshold voltage increases as the minimum feature size is scaled down [10]. By utilising two-dimensional device simulations we obtained the standard deviation of the threshold voltage for feature sizes down to 10 nm as can be seen in Fig. 3 [11]. A large increase of the threshold voltage variation is expected and its impact on future circuits must therefore be investigated.

Fig. 3. Standard threshold voltage deviation versus feature size.

### 5. Circuit Simulation

#### 5.1 Digital logic

We first started with the investigation of the two parasitic effects on the logic unit. For a large number of MOS transistors, the $6\sigma$ deviation has to be considered. The threshold voltage variation was applied to the input flip flops of the 64bit adder. This results in an increased delay of the input registers.

Fig. 4. Delay time of the 64bit adder with copper interconnects and air gap structures. Simulations were performed with copper size effect and threshold voltage fluctuation (squares) and without these effects (circles).

Fig. 4 plots the delay time of the 64bit adder versus the minimum feature size. The simulation was performed once with and once without the parasitic effects. If we neglect the parasitic effects a linear increase in
performance can be achieved by scaling down the minimum feature size if a low-k intermetal dielectrium is used. If we take the threshold voltage variation and the size effect into account the performance increase starts to slow down around 50 nm. But a feasible performance gain can still be achieved for pure digital logic by scaling down the minimum feature size.

Fig 5. Performance loss for the 64bit adder versus minimum feature size due to the size effect of small copper interconnects (circles) and the threshold voltage variation (squares).

Fig. 5 plots the speed loss due to the size effect and the threshold voltage variation. Even with an increase of the interconnect resistance by a factor of 3 for the 25 nm generation due to the size effect (Fig. 2), the speed loss is kept within reasonable limits.

The resistance for local interconnects is several ten Ohm. This is much lower than the output resistance $U_{DS}/I_{DS}$ of the MOSFET which is around 1kΩ for all simulated technologies. Therefore, the output resistance of the MOSFET is much more important than the interconnect resistance [12] and this will reduce the effect of parasitic interconnect resistance increase. The size effect must be taken into account but it is not a road block.

The effect of the threshold voltage variation is negligible for logical units with 10000’s of transistors because the variation apply only to a small number of them.

5.2 Mixed circuits

The impact of the parasitic effects on mixed analog and digital systems is investigated at a 1Mbit SRAM. Fig 6. shows the schematic of the simulated SRAM. It is organized into 64 macro blocks each containing 256 rows and 256 columns of cells. We used a cross-coupled inverter as a sense amplifier. We neglected all details of address generation. The 1σ threshold voltage variation was applied as a mismatch parameter to the sense amplifier.

Fig 7. plots the delay time of the SRAM circuit for different feature sizes. The simulations were performed including the parasitic effect and neglecting them. It can be seen in Fig. 7 that, contrary to the pure digital system, the effect of threshold voltage variations is more detrimental for the delay time than the size effect.

The delay time of the sense amplifier is the major part of the total SRAM delay time. It is determined by two aspects, the load of the bitline and the sensing properties of the sense amplifier. The load of the bitline is dominated by the input capacities of the SRAM cells, therefore the increasing resistive load due to the size effect does not contribute too much to the total load. But the threshold voltage mismatch will strongly affect the sensing properties of the sense amplifiers (Fig. 8). Nevertheless, even with large threshold voltage variations scaling down the minimum feature size results in a remarkable speed improvement.

Fig. 6 Schematic of the 1Mbit SRAM circuit.

Fig 7. Delay time of a 1Mbit SRAM. Simulations were performed with copper size effect and threshold voltage fluctuation (squares) and without these effects (circles). The right axis shows speed loss due to the threshold voltage variation (triangles) and the size effect (squares).

5.3 On-Chip Interconnects

So far, the parasitic effects have been investigated on digital and mixed systems with local interconnects. But the performance for future systems is strongly affected by the on-chip interconnects. Therefore the impact of the parasitic effects has also been investigated for on-chip interconnects. Fig 9 shows a schematic of the simulated interconnect system. It consists of a cascaded driver and a receiver stage. We assumed the requirement that the interconnect is able to transmit a signal within
the cycle time of the 64bit adder, so that the on-chip interconnect will not slow down the whole circuit. Table 2 shows the necessary frequencies of the interconnect.

![Schematic of the on-chip interconnects](image)

**Fig. 9. Schematic of the on-chip interconnects**

<table>
<thead>
<tr>
<th>Feature Size [nm]</th>
<th>100</th>
<th>50</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freq.</td>
<td>2 GHz</td>
<td>3 GHz</td>
<td>3.6 GHz</td>
</tr>
<tr>
<td>Chip Size [mm²]</td>
<td>118</td>
<td>181</td>
<td>238</td>
</tr>
</tbody>
</table>

**Table 2. On-chip interconnect parameters**

![Plot of the maximum drivable length](image)

**Fig 10. Plot of the maximum drivable length according to table 2. Simulations were perform with parasitic effect (circles) and without them (squares).**

Fig. 10 plots the maximum on-chip interconnect length that can be driven within the cycle time of the 64bit adder. Due to the interconnect resistance and capacitance a dramatic reduction in the maximum length has been observed. This length is further reduced by the $V_{TH}$ fluctuations and the size effect. Fig. 11 shows the loss in maximum on-chip interconnect length due to the size effect and the $V_{TH}$ variation. The $V_{TH}$ variation will slow down the receiver stage but its effect will become negligible for smaller feature sizes due to the increasing effect of interconnect resistance and capacitance. The main problem is the size effect. It can be seen that a reduction of the maximum length by almost 50 percent is expected for the 25 nm generation. The delay time of long on-chip interconnects is dominated by their large resistance. Any further parasitic increase of the interconnect resistance like the size effect will have significant impact. It must be taken into consideration that the chip size will increase with scaling down the minimum feature size. Table 2 shows the chip size as it is predicted by the ITRS Roadmap. With increasing chip size the distance between the single components of a system will increase. But this is opposite to the driving capability of future interconnects. Further investigations of the impact of the size effect on on-chip interconnects are necessary. Otherwise this may become a serious problem in gigascale integration.

6. Conclusions

We have shown that the threshold voltage variation and the increasing parasitic resistance due to the copper size effect must be taken into account for circuits with gigascale integration. They will reduce the performance gain obtained by scaling down the minimum feature size of local systems but they will not cancel it out.

But, for intermediate on-chip interconnects the copper size effect is detrimental. Further investigations are needed to remove this road block. The effect of repeater insertion and the optimum aspect ratios of long interconnects must be investigated to overcome the impact of the size effect.

**References**


[2] [http://www.cadence.com](http://www.cadence.com)


