Capacitive Crosstalk Effects on On-Chip Interconnect Latencies and Data-Rates

Master’s thesis performed in Electronic Devices
by

Rebecca Källsten

LiTH-ISY-EX-3617-2005
Linköping 2005
Capacitive Crosstalk Effects on On-Chip Interconnect Latencies and Data-Rates

Master’s thesis
performed in Electronic Devices,
Dept. of Electrical Engineering,
at Linköping University

Rebecca Källsten
LiTH-ISY-EX-3617-2005-SE

Supervisor: Professor Christer Svensson
Examiner: Professor Christer Svensson
Linköping 2005-01-21
Capacitive Crosstalk Effects on On-Chip Interconnect Latencies and Data-Rates

Rebecca Källsten

This thesis work investigates the effects of crosstalk on on-chip interconnects. We use optimal repeater insertion as a reference and derive analytical expressions for signal latency, maximum data-rate and power consumption. Through calculations and simulations we show that despite large uncertainties in arrival time of a signal that is subject to crosstalk, we are able to make predictions about the maximum data-rate on a bus. We also show that data-rates can exceed the classical limit of the latency inverted by using wave pipelining. To increase the data-rate, we can increase the number of repeaters to a limit, at the cost of additional latency and power. Savings in power consumption can be achieved by using fewer repeaters, paying in latency and data-rate. Through fewer repeaters, the top metal layer shows better performance in all investigated aspects.

Keyword: on-chip interconnect, crosstalk, data-rate, latency, power consumption, Miller capacitance
This thesis work investigates the effects of crosstalk on on-chip interconnects. We use optimal repeater insertion as a reference and derive analytical expressions for signal latency, maximum data-rate and power consumption. Through calculations and simulations we show that despite large uncertainties in arrival time of a signal that is subject to crosstalk, we are able to make predictions about the maximum data-rate on a bus. We also show that data-rates can exceed the classical limit of the latency inverted by using wave pipelining. To increase the data-rate, we can increase the number of repeaters to a limit, at the cost of additional latency and power. Savings in power consumption can be achieved by using fewer repeaters, paying in latency and data-rate. Through fewer repeaters, the top metal layer shows better performance in all investigated aspects.
Acknowledgements

I would like to take this opportunity to thank the people who have helped and supported my work on this master’s thesis. First of all, I thank my examiner and supervisor Professor Christer Svensson for his valuable advice, ideas, support, and contributions to this report.

Of all the people of the Electronic Devices research group at Linköping University, there are a few I want to thank in particular: Lic. Eng. Peter Caputa for discussions, good advice, and for allowing me to use his wire modelling scripts, Professor Atila Alvandpour for introducing me to researching and for interesting discussions, and Anna Folkesson for her help with the bar mouse.

Though not directly involved in this thesis, I want to thank my wonderful significant other, Fredrik Orinius, and my parents, for their encouragement during this thesis work and always.

Last but not least, I thank Junaid Aslam for opposing this thesis work.
# Table of Contents

Abstract .............................................................................. i

Acknowledgements ................................................................. iii

Table of Contents ................................................................. v

1 Introduction ........................................................................... 1

2 Basic Theory ......................................................................... 3
   2.1 Wire Models .................................................................. 3
      2.1.1 Model for Simulations ............................................. 3
      2.1.2 Simplified Model for Calculations .......................... 3
   2.2 Elmore Delay ................................................................. 4
   2.3 Wire Section with Repeater ........................................... 4
   2.4 Optimizing a Wire for Fast Signaling .............................. 5
   2.5 Crosstalk .................................................................... 6
   2.6 Data-rate .................................................................... 7
      2.6.1 Classical View of Data-rate .................................... 7
      2.6.2 Data-rate in Relation to $\delta$ ................................. 8
   2.7 Power and Energy .......................................................... 8

3 Simulations ........................................................................... 11
   3.1 Setup .......................................................................... 11
   3.2 Measurements on Simulated Bus Components .......... 12
      3.2.1 Transistor Specification ........................................ 12
      3.2.2 Delay and Energy ................................................ 13
   3.3 Wire Characteristics for Different Metal Layers .......... 14

4 Results ................................................................................. 15
   4.1 Delay vs. Number of Sections ...................................... 15
   4.2 Bandwidth and Wave Pipelining .................................. 17
   4.3 Energy vs. Number of Sections ................................... 20
   4.4 Power vs. Data-rate ..................................................... 22

5 Discussion ............................................................................ 23
   5.1 Energy vs. Repeater Size ............................................. 23
   5.2 Sources of Error .......................................................... 23
   5.3 The Impact of Inductance ............................................. 24

6 Future work .......................................................................... 25
7 Conclusions ................................................................. 27
Appendix A - Simulation Data .......................................... 29
References ................................................................. 33
Chapter 1

1 Introduction

As chip manufacturing techniques go towards an increasing level of complexity and chip features continue to shrink, gate speeds are increasing and systems-on-chip have higher requirements to go faster. At the same time, wires get worse performance for each chip generation. Wire resistances go up when they are made thinner compared to their length, as do inter-wire capacitances which may lead to severe crosstalk effects. Signalling over on-chip wires is a bottleneck for increasing speed in future technologies.

This thesis work is carried out in a 180 nm technology with six metal layers. We have investigated the effect of crosstalk on latency, maximum data-rate and power consumption of busses with repeaters. For simplicity we have chosen a fixed bus length of 6 mm, and compared a few metal layers, M1, M2, and M6. Our results can be generalized to any bus with repeaters in any process, by generating parameters for that process and entering new values in our formulas.

Initially, we present the basic theory that is needed to understand wire behavior and simulation results. Then we show the setup for our measurements, after which we present our predictions and the simulated results and discuss the accuracy of our calculations.
2 Basic Theory

To analytically investigate crosstalk on wires, we present models for wires and derive expressions for wire performance.

A bus in our case consists of three wires. Each wire is divided into N equal sections of one repeater and a wire section, where the wire section is 1/N times the length of the total wire. For simplicity, all repeaters have the same size and we generalize analytical calculations on one repeater and a wire section to be approximately valid for all sections.

2.1 Wire Models

We use different models for analytical calculations and simulations, since we will both estimate performance by hand calculations and run simulations on a computer. Both models uses wire parameters from HSPICE, but the model for calculations is very simplified.

2.1.1 Model for Simulations

A wire section for simulations consist of an inverter and a LRC-model of the wire [1]. For the inverter, we use transistor models given by the process manufacturer ST. The LRC-model contains a series of lumped sections with capacitances and inductances between each pair of parallel wires as well as inductance, capacitance towards the substrate, and resistance for each wire.

2.1.2 Simplified Model for Calculations

A wire in the lower metal layers on a chip can be viewed as a distributed RC chain. For calculations, we simplify this to a lumped RC-π-section, see figure 1. Inter-wire capacitances are accounted for by adjusting the value of the capacitances. Inductances are neglected.

A repeater is simplified to an output resistance and capacitance when placed before a wire section, and a capacitive load when it’s connected to the end of a wire. See figure 2 on page 4.

Further analytical discussions will be based on this simple model.
2.2 Elmore Delay

Delay is a measure of performance that is traditionally of great importance. We may for example have requirements on the delay to be less than a certain fraction of the frequency at which we operate to ensure that data is processed in one clock cycle.

The Elmore delay is a first order approximation of the time constant $\tau$ in a RC network. For a network with $N$ nodes, it is given by (1), where $C_j$ is the total capacitance of node $j$, and $R_{ij}$ is the resistance that the paths from node 0 to nodes $j$ and $i$ have in common [2].

$$\tau = \sum_{j=1}^{N} \sum_{i=1}^{j} R_{ij} \times C_j$$

(1)

If the step response is given by (2), $\tau$ gives a lot of information about the behavior of our network. However, since the Elmore delay only gives the dominant pole of our system, we use it with some correction for the neglected time constants. We may also use $\tau$ for other things, such as calculating rise and fall times. We can also derive optimal sizes for drivers and optimum wire delay.

$$S(t) = 1 - e^{-t/\tau}$$

(2)

2.3 Wire Section with Repeater

Figure 2. Model of wire section.
We model a wire section between two repeaters as an inverter output resistance and capacitance, a wire modeled as a lumped π-element, and an inverter input capacitance. See figure 2 for details. The Elmore delay for a repeater section is:

$$\tau = R_d \times \left( C_d + \frac{C_w}{2} \right) + \left( R_d + R_w \right) \times \left( \frac{C_w}{2} + C_g \right)$$  \hspace{1cm} (3)$$

where

$$R_d = \frac{r_d}{w_t}$$
$$C_d = c_d \times w_t$$
$$C_g = c_g \times w_t$$
$$R_w = r_w \times L/N$$
$$C_w = c_w \times L/N$$  \hspace{1cm} (4)$$

$R_d$, $C_d$, $R_w$, and $C_w$ are the resistance and capacitance of the driver and the wire, respectively. $C_g$ is the input load of the next repeater/wire section, $L$ is the wire length, $N$ is the number of sections we have, and $w_t$ is the repeater NMOS transistor width.

From this we can derive the delay $t_d$ of a wire with $N$ sections to be given by (5), by collecting the lumped parts in the first term and the distributed wire in the second term [3].

$$t_d = 0.69 \times \left( r_d \times c_d \times N + \frac{r_d \times c_w}{w_t} \times L + r_d \times c_g \times N + r_w \times L \times c_g \times w_t \right) + 0.38 \times r_w \times c_w \times \frac{L^2}{N}$$  \hspace{1cm} (5)$$

### 2.4 Optimizing a Wire for Fast Signaling

From (5) on page 5, we can derive formulas for optimized driver width, section length, and number of sections. Optimal driver width, for example, is calculated by setting the derivative with respect to $w_t$ equal to 0, see (6).

$$\frac{dt_d}{dw_t} = 0 \Rightarrow w_t = \sqrt[2]{\frac{r_d \times c_w}{r_w \times c_g}}$$  \hspace{1cm} (6)$$

In the same way, we derive expressions for the optimal section length and number of sections in (7) and (8).

$$L_{opt} = \sqrt[2]{\frac{0.69 \times r_d \times (c_d + c_g)}{0.38 \times r_w \times c_w}}$$  \hspace{1cm} (7)$$

$$N_{opt} = \frac{L_{tot}}{L_{opt}}$$  \hspace{1cm} (8)$$
2.5 Crosstalk

Propagation delay through a wire depends highly on the capacitive load $c_w$ of the wire, as can be understood from (5) on page 5. For a bus, this load varies with the pattern sent on the bus, introducing a spread $\delta$ in the arrival time at the far end of the wire. This is because of the different amounts of charge that has to be transported to achieve the change in voltage level. For the worst case, when two wires switch in opposite directions, we have to transfer twice as much charge to the coupling capacitor compared to the case when one of the wires is quiet. We could also say that one wire experiences the capacitive load towards its neighbor to be doubled. This is called the Miller effect. With the same reasoning, we see that if two wires switch in the same direction, no charge has to be transferred to the coupling capacitor and we have a minimum case of experienced interwire capacitance.

$\delta$ can be calculated as the difference between $t_{d,max}$ and $t_{d,min}$ given by (5) on page 5. This leads us to (9), where $c_{wmax}$ is the experienced Miller capacitance when the aggressor wires are driven in the opposite direction to the victim wire and $c_{wmin}$ is the case when wire-to-wire capacitances are discarded as the wires are driven in the same direction.

$$\delta = N \times \left( 0.69 \times \frac{r_d \times c_{wmax}}{w_t} \times L + 0.38 \times r_w \times c_{wmax} \times \frac{L^2}{N} \right)$$

Another aspect of crosstalk is that it may introduce bit errors. For a risetime greater than zero, we expect an influence from the neighboring wire according to (10) [1]. For safe communication, we require a $\Delta V_B$ of less than $0.5V_{dd}$. Since the crosstalk affects all wires and our bus consists of three equal wires, $t_{rA}$ is also a function of $R_0$, $C_c$, and $C_0$ in figure 3, that presents the setup for (10).

$$\Delta V_B = \frac{C_C}{C_0 + C_C} \times \Delta V_A \times \left( \frac{R_0 \times (C_C + C_0)}{t_{rA}} \right) \times \left( 1 - e^{-\frac{t_{rA}}{R_0 \times (C_C + C_0)}} \right)$$

A complicating factor when considering crosstalk is that it is very bit pattern dependent. On a bus with more than two bits, crosstalk affects different wires in different ways, leading to different propagation delays on neighboring wires. This leads to coupling effects that vary along the length of a wire. In a worst case scenario for the middle wire on a three bit bus for example, we might only have the actual worst case on the near end of the first segment of the wire. On the far end, the aggressor wire signals may arrive before the victim signal, leading to less
severe crosstalk effects as experienced by the victim wire. Neither delay effects nor voltage glitches are constant along the length of a wire. This leads to errors in our calculations. We will overestimate the delay spread since the maximum delay is smaller than we think. This will give an estimated bandwidth that is smaller than possible in reality. When it comes to voltage glitches, we can also note that the aggressor signal risetime decreases along a wire section, and relieves the crosstalk effects.

In this presentation of the problem, we neglect these effects and are satisfied with a theoretical solution that is slightly worse than simulations.

2.6 Data-rate

When chip device speeds are increasing, it becomes increasingly important not to look only at the latency of an interconnect, but we also have to consider data throughput [4]. If we optimize an interconnect for short delays only, the possible data-rates on the bus may be too low to keep up with the much faster logic. As we will see, we have a trade-off between delay and data-rate.

Eye diagrams are used to show detectability of data. They consist of possible waveforms that appear where data is detected, at the far end of the wire in our case. These waveforms form an ‘eye’ that are given specific requirements. The basic requirement is that the eye must be open or else we have no means for discriminating between our logical signaling levels.

2.6.1 Classical View of Data-rate

The traditional way of calculating on achievable data-rates is to look at the step response $S(t)$ on a wire, set the symbol time to $T$ and require that $S(T)$ must meet a specification of the eye opening, $e_o$ (see figure 4). This reasoning gives us (11), which is analytically solvable for $S(t)$ according to (2) on page 4, but has to be solved numerically for a more accurate approximation of the step response.

$$e_o = 2 \times S(T) - 1$$  (11)

Figure 4. Eye opening.

This requirement on the eye opening is given in normalized signal levels. As we will soon see, such calculations become partially invalid when we consider crosstalk.

(11) can only be used on linear systems. As soon as we insert a repeater, we cannot approximate the wire to a linear system due to the unlinear properties of transistors. For a wire with inverters, we no longer have an analytical expression for possible data-rates.
2.6.2 Data-rate in Relation to $\delta$

As already stated in chapter 2.5, crosstalk introduces an uncertainty in the arrival time of a symbol. If we have a bus with a random pattern, this gives us an interval of length $\delta$ during which the symbol will appear at the far end of a wire. As mentioned in chapter 2.6.1, we normally set requirements for how open the eye must be, measured in volts, to be able to detect a symbol sent on the bus. However, with this uncertainty in when the symbol will appear on the far end of the wire, we have another limitation on the eye diagram. In this work, we will assume that we have a restriction that the eye diagram must have an opening of 50% of both the signal swing and the symbol time. With these specifications of the eye opening, we can calculate the bandwidth $B$ according to (12), which holds as long as the worst case step response leads to an eye opening that meet our requirements.

$$B = \frac{1}{2 \times \delta}$$  \hspace{1cm} (12)

2.7 Power and Energy

The power dissipated when switching a wire section is given by (13), where $\alpha$ is the switching activity, $f$ is the frequency, and $E$ is the average energy dissipated per transition given by (14). $C_L$ is the total load on the driving inverter and is composed of the wire capacitance $C_w$, the output capacitance $C_d$, and the input capacitance of the next section, $C_g$. Since power depends on data activity and frequency, which we know nothing about at this point, it makes more sense to consider the switching energy. By looking at energy rather than power, we can more accurately compare energy efficiency for different bus configurations without considering possible data-rates. Power will only be considered under the assumption that the bus is switched at the highest possible rate.

The energy given by (14) is an approximation that does not take into account the non-linear effects of $C_d$ and $C_g$, nor does it account for the short-circuit currents during the switching.

$$P = \alpha \times f \times E$$  \hspace{1cm} (13)

$$E = \frac{C_L \times V_{dd}^2}{2}$$  \hspace{1cm} (14)

If we inspect the probability for the worst case scenario, that the aggressor wires switch in the opposite direction compared to the victim wire, we see that such a transition statistically only takes place in 2 out of 64 cases. This would suggest that the worst case scenario overestimates any actual application quite a bit. As an attempt to estimate the signaling energy better, we therefore calculate the average experienced capacitative load on the victim wire under the assumption that all 64 possible transitions are as likely. This yields $\alpha = 0.5$. Table 1 shows coefficients for calculating an approximation of the experienced wire capacitance of the middle wire for different signalling cases, according to (15). It also shows statistics of occurrences of the different cases. The experienced capacitance depends on the inter-wire capacitance $c_{in}$ to one neighbor, and the wire-to-substrate capacitance $c_{sub}$, when applicable (we have a transition on the middle wire).
When calculating the energy $E$, $C_L$ is now given by (16), where $c_{w,avg}$ is the weighted average of $c_w$ and the weights are the relative occurrences we can calculate from Table 1.
Chapter 3

3 Simulations

3.1 Setup

The basic testbench setup is presented in figure 5. It consists of a pattern generator, wire sections (inverter + wire) that act as transmitter and receiver, and a wire that is 6 mm long and split up in N sections. We measure delay from before the first section to after the last section of the wire. Since the wire is driven and terminated by wire sections, we have the same waveform at both the driver and receiver ends, which is practical during measurements.

![Figure 5. Testbench setup, N=3.](image)

The pattern generator is a verilog-A coded block that takes a clock as an input and generates a bit pattern that includes the best and worst cases as well as the case where we have no crosstalk.

For the wires, a discrete LRC model is used [5]. The model uses the manufacturer's specifications of wire geometry and physical parameters. Each wire is sectioned in 0.50 µm pieces, where one piece represents one π-element. The wires that are used in Cadence simulations are verilog-A code blocks generated with a Perl script that in it’s turn calls for HSPICE circuit simulator to generate numerical values for the LRC model, given our geometry. Numerical values on wire parameters are taken directly from this model when they are needed for calculations.

The bus configurations used in this work are all minimum sized with respect to width and signal separation. Moreover, they are 3 bit busses where midle wire is the one we take interest to since it’s the one most affected by interwire phenomena. The wires have separate power supplies so that we can measure current drawn by the middle wire only.
3.2 Measurements on Simulated Bus Components

We use simulations to determine numerical values to enter in our analytical expressions, and to compare theory with simulation.

For our inverters, the PMOS transistor width is 2.59 times the width of the NMOS transistor.

3.2.1 Transistor Specification

In our theoretical formulas, we have the quantities $r_d$, $c_d$, and $c_g$, which we need to determine values of through simulations. We want an approximation that is valid for a realistic case, where we spend most time just holding a ‘1’ or a ‘0’ and the switch is quite fast.

An approximative value of an inverter’s output resistance, output capacitance, and input capacitance, is calculated from (17). Figure 6 a) - c) shows the simulation setup. The inverter is given a step input, and we measure the time it takes to change the output by 50%.

$$t_d = 0.69 \times R \times C$$

(17)

![Figure 6. Setup for measuring inverter properties.](image)

For measuring $R_d$, we put a large load on the inverter and solve (17) for $R = R_d$ and $C = C_{load}$, see figure 6 a). When we want to know the output capacitance $C_d$, we remove the output load, see figure 6 b), and solve $R_d C_{load} t_{da} = R_d C_d / t_{db}$, where $C_{load}$ is sufficiently much larger than $C_d$ for this to be a valid approximation. For the input load, $C_g$, of the inverter, we load the inverter with an equal inverter, see figure 6 c), and solve (17) for $C = C_g + C_d$ and $R = R_d$. 


The parameters \( r_d, c_g, \) and \( c_d \) are now given by (4) on page 5 and presented in table 2.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( r_d )</td>
<td>2.18m( \Omega )m</td>
</tr>
<tr>
<td>( c_g )</td>
<td>6.2nF/m</td>
</tr>
<tr>
<td>( c_d )</td>
<td>10.1nF/m</td>
</tr>
</tbody>
</table>

Table 2. Simulated values on \( r_d, c_g, \) and \( c_d \).

### 3.2.2 Delay and Energy

The measurements we are interested in are the nominal, maximum, and minimal delay, the average energy drawn from \( V_{dd} \) per transition, and the peak crosstalk at the far end of the wire.

For accurate latency and energy measurements, we use a low frequency that allows transient effects to settle and the signals to reach full swing.

Maximum delay is measured when the bit pattern sent on the bus is 010->101->010.

The nominal delay is measured for the bit pattern 000->010->000 and 111->101->111. For every configuration with an odd number of sections, it may make a slight difference if we only consider one of the cases when the bus is quiet.

Minimum delay is measured for the bit pattern 000->111->000.

For the average switching energy, some assumption has to be made about the probability that current is drawn from the power supply. We assume that all transitions are as likely, so the average current drawn from power supply is measured during the 64 possible transitions. This assumption gives a switching activity \( \alpha = 0.5 \), which is high, and should overestimate the actual switching activity for any application that doesn’t send a clock (\( \alpha = 2 \)) on the bus. We measure the average switching energy by coding a bus so that it makes every possible transition once and then measure the average current drawn from power supply. The energy is calculated from the average current \( I_{avg} \) according to (18). \( I_{avg} \) is the average current drawn from supply during a time interval of \( t \) and \#symb is the number of symbols that are transmitted during this time.

\[
E = \frac{I_{avg} \times V_{dd} \times t}{\#symb / 2} \quad (18)
\]
3.3 Wire Characteristics for Different Metal Layers

Table 3 presents an overview of wire characteristics and bus configurations for optimal (nominal) delay. The bus is of minimum allowed size in this process, which is a 180 nm process, leading to maximum crosstalk. The wire sections add up to a length of 6 mm. Nominal wire capacitance, \( c_{w,nom} \), for the middle wire is \( 2c_{iw} + c_{sub} \). This is the capacitance we optimize our bus for.

<table>
<thead>
<tr>
<th>metal layer</th>
<th>wire width and separation [µm]</th>
<th>( c_{iw} ) [pF/m]</th>
<th>( c_{sub} ) [pF/m]</th>
<th>( c_{w,nom} ) [pF/m]</th>
<th>( r_w ) [MΩ/m]</th>
<th>( l_w ) [µH/m]</th>
<th>( w_t ) [µm]</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.32</td>
<td>93.2</td>
<td>24.9</td>
<td>211.3</td>
<td>0.15</td>
<td>1.12</td>
<td>22.3</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>0.32</td>
<td>95.5</td>
<td>18.8</td>
<td>209.8</td>
<td>0.15</td>
<td>1.43</td>
<td>22.2</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>0.64</td>
<td>92.7</td>
<td>16.2</td>
<td>199.5</td>
<td>0.04</td>
<td>1.65</td>
<td>41.6</td>
<td>2</td>
</tr>
</tbody>
</table>

*Table 3. Design rules and wire parameters.*
Chapter 4

4 Results

We will here go more into detail about the theory established in chapter 2. For our calculations, we use a model of a three bit bus in a 180nm process.

There are two main fears of how metal wires on chips will limit performance in technologies of today and in the future:

- They will degrade system speeds because the maximum delay \( t_{d,\text{max}} \) is increasing, assuming that the system speed is reversely proportional to wire latency.
- Increased crosstalk will make it harder to detect symbols, as we have a large uncertainty in arrival times.

With the fundamentals covered in chapter 2, we will now discuss and deal with these obstacles.

4.1 Delay vs. Number of Sections

The propagation delay through a wire is, as stated in chapter 2.4, related to the number of sections used. The optimal section length is given by (7) on page 5. Any change from this optimum section length or transistor widths degrades the delay aspect of the wire performance for the case which we optimized for.

Normally, one wants to optimize the worst case performance. Here, all calculations of optimal values are based on the case where we don’t consider crosstalk at all. Due to this, we get a worst case delay that isn’t optimized, as we can see in figure 7, which shows the simulated delays for the nominal and worst cases in metal 1.
Figure 7. Simulated delay as a function of the number of sections in metal 1.

We see here that the minimum worst case delay indicates that we could decrease the worst case delay by reoptimizing for the worst case scenario, since the nominal and worst case delays have slightly different optimal numbers of sections. Another interesting point is that the difference between nominal and maximal delay decreases with an increasing number of sections. The worst case delay can be as much as 53% longer than the nominal delay, but with repeaters, a more general estimation is that the difference is in the order of 20%.

Minimum delay differs from nominal delay by roughly 20%, but can differ as much as 56% for a case without repeaters.

Figure 8. Nominal delay in metal 1, simulated and calculated.

If we compare our simulated results with predictions made by (5) on page 5, we see that the trends for the delay dependency of N is accurate (figure 8). The error is
approximately 10%. Due to errors in estimating the impact of Miller effects, the error in calculating minimum and maximum latencies is about 15%.

Figure 9 provides a comparison for the three investigated metal layers. Metal 6 is superior to metals 1 and 2 mainly because of the much lower resistance. The difference between metals 1 and 2 is very small and comes from the 27% larger inductance of metal 2. The difference in wire capacitance between metal 1 and 2 is less than 1%.

![Figure 9. Simulated nominal delay vs. N for metals 1, 2, and 6.](image)

### 4.2 Bandwidth and Wave Pipelining

In chapter 2.6.2, we discussed the possibility of using the spread in delay as a new measure of bandwidth. Figure 10 shows the eye diagram for a metal 1 wire with 10 sections, where the symbol time is twice the delay spread; in this case 309 ps. The eye is open for 50% of the symbol time and all signals reach full swing; the traditional requirement on eye openings is valid.

If we choose a data-rate of 8 Gbps and look at eye-diagrams with and without crosstalk, it becomes obvious that this data-rate leads to a completely closed eye and no detectability of data when we have crosstalk (figure 11 b)), even though the case without crosstalk shows an open eye (figure 11 a)). This was our fear when it comes to detectability of data on a bus, but we have shown that our new requirement on the eye opening gives a good prediction for the possible data-rate.

Figure 12 shows the eye diagram of a metal 1 wire with only one section. In this case, both restrictions on the eye opening leads to a data-rate of 1 Gbps. If we had an even longer wire with one section, (11) on page 7 would give our possible bandwidth.
Figure 10. Metal 1 wire with N=10, driven in 3.23 Gbps.

Figure 11. Metal 1 wire with N=10, driven in 8 Gbps. a) no crosstalk, b) including crosstalk effects.
Figure 12. Metal 1 wire with N=1, driven in 1 Gbps.

The trend for bandwidth (figure 13) is that the shorter the sections get, the less crosstalk we experience (decrease in the factor $L^2/N$ in (9) on page 6) and hence less spread in the delay. This is because the performance of a wire section is dominated by the transistor capacitances rather than the wire capacitance for short wire sections. Bitrate versus number of sections do not show an optimum, but rather approaches a maximum value asymptotically.

Figure 13. Simulated bandwidth vs. N for M1, M2, and M6.

Our example in figure 10 on page 18 has a possible symbol time of 309 ps, while the worst case delay is 493 ps. This means that we may have more than one symbol on the bus at any time and it’s called wave pipelining [4]. A close look at the bandwidth and possible symbol times according to (12) on page 8 (figure 14), reveals that wave pipelining is possible above a certain limit in the number of sections. Wave pipelining facilitates data-rates above the classical limit $1/t_{d,max}$. 

4.3 Energy vs. Number of Sections

Assume that we want to decrease the power and that we can do this by decreasing the number of inverters, motivated by the assumption that the relation between energy and number of sections is linear when short-circuit currents are neglected. In figure 15, the expected average energy dissipation per bit is plotted for metal 1, along with the simulated results. It is clear that we cannot neglect short-circuit effects of the inverters when our goal is to decrease power, since the simulated energy curve shows an optimum number of sections. If we decrease the number of sections further, we increase the energy per bit. For an N smaller than optimum, short-circuit currents can give a substantial contribution to the total consumed energy since the load of each repeater is large. A large load yields long rise and fall times, which leads to a long time interval during which both transistors of the inverter are conducting and hence a lot of current is drawn directly from supply to ground.

A plausible reason for the error in our calculations, compared to simulations, in the case of many repeaters is that there are less Miller effects due to unsynched signals, as discussed in chapter 2.5. The overall discrepancy is roughly 20%.
Investigating the differences in energy consumption between the different metal layers, we find that metals 1 and 2 are relatively similar. Metal 6 appears to be generally more expensive, but if we look back at figure 9 on page 17, we see that any specification we may have on wire latency in metal 1 or metal 2 is met even with no repeaters in metal 6. With an optimum section number for metals 1 and 6 respectively (table 3), we consume approximately 10% less energy in metal 1 than in metal 6. Through fewer repeaters, metal 6 shows considerably better performance in latency, data-rate and power consumption. The major drawback for metal 6 wires with many repeaters is their large repeaters which are almost twice the size of those in metal 1.
4.4 Power vs. Data-rate

If we drive a bus at the maximum rate, the power consumption can be calculated from (13) on page 8. Figure 17 shows the relation between power and data-rate, where the power measure is calculated from (13) on page 8. The energy and data-rates are simulated results, inserted in (13) on page 8. The different data-rates corresponds to different number of repeaters. We see an exponential increase in power consumption as we increase the data-rate.

*Figure 17. Power vs. data-rate in M1.*
Chapter 5

5 Discussion

5.1 Energy vs. Repeater Size

By decreasing the size of the repeaters, it should be possible to save power in a similar way to using fewer repeaters. The cost is increased latency, decreased maximum data-rates and a higher risk of data errors due to crosstalk. The smallest possible repeater size for a given bus configuration is analytically given by (10) on page 6 and our requirement that crosstalk glitches must be less than 0.5xVdd.

5.2 Sources of Error

In chapter 2.5, we noted that we will overestimate the worst case delay. This gives an error in calculated bit rate, illustrated in figure 18. We also have errors from the formula used, since we only look at first-order effects and neglect wire inductance. In all, the error is 30% for data-rate predictions in metals 1 and 2, but as much as 70% for metal 6.

Figure 18. Simulated and calculated bit rates vs. N for metal 1.
For nominal delay, we have no Miller effects. For long sections, the dominant term in the error compared to simulations originates in neglected inductance. For short wire sections, the larger part of the error comes from errors in $r_d$, $c_g$, and $c_d$.

### 5.3 The Impact of Inductance

The inductance is responsible for crosstalk and performance degradation as it works against changes in current. But, it also compensates somewhat for capacitive effects by (partially) cancelling the forward-travelling wave introduced by cross-talk on transmission lines [1].

In metal 6, where the inductance has most influence, we have a slightly larger relative error in our predictions for delays, than in metal 1. We tend to overestimate the maximum latency even more, which tells us that even though we neglected inductance in our calculations, we are still on the safe side in our estimations. A bus implemented on a chip, with specifications that are met with our calculations, would work. However, even this small relative difference in error between metals 1 and 6 has a large impact on data-rate calculations. If we want more accurate estimations of the capability of a bus in metal 6, we need to include inductance.
Chapter 6

6 Future work

In chapter 2.6.1, we claimed that there is no analytical expression for determining the eye opening in the traditional way of setting requirements on an eye diagram, if we have repeaters on our wire. It would be useful to find such an expression for cases where we need that requirement to analyse detectability.

The inductance we neglect in this work is dealt with in many articles on repeater insertion and wire performance [6][7], but the investigation of data-rates and energy consumption can be improved by looking deeper into Miller effects and adjusting the formulas accordingly. Energy and power can also be estimated more accurately if we include some measure of short-circuit currents.
Chapter 7

7 Conclusions

We have shown that the two major obstacles related to crosstalk can be overcome by setting new requirements on eye diagrams and by using wave pipelining.

We have investigated the effects of crosstalk on latency, maximum data-rate and power consumption of buses in a few different metal layers. To simplify our comparisons, we used an optimal repeater insertion (discarding crosstalk) as reference, and assumed random data on the bus. Through simulations and analytical expressions we found:

1. With crosstalk bus latency spread 20% from the nominal minimum latency value.
2. Minimum worst case latency occurs for a slightly larger repeater number than the nominal optimum number.
3. Crosstalk limits the maximum data-rate through its effect on the data eye, but still data-rates larger than the latency inverted can often be used.
4. By increasing the number of repeaters along the bus, the maximum data-rate can be increased by wavepipelining to a limit, at the cost of larger latency and increased power consumption.
5. By decreasing the number of repeaters or decreasing the size of the repeaters, it is possible to save power at the cost of increased latency and decreased maximum data-rate.
6. Metal 1 and metal 2 are relatively similar, but metal 6 show considerably better performance in latency, data-rate and power consumption (through fewer repeaters).

We further developed an analytical model describing latency, maximum data-rate and power consumption. The model always gives worse results than the simulations, but with the correct trends. The discrepancies are of the order of 15% for latencies, 30% for maximum data-rates and 20% for power consumption. The large divergence for data-rate predictions comes from overestimated Miller effects.
## Appendix A - Simulation Data

<table>
<thead>
<tr>
<th># of sections</th>
<th>$t_{d,min}$ [ps]</th>
<th>$t_{d,nom}$ [ps]</th>
<th>$t_{d,max}$ [ps]</th>
<th>$\delta$ [ps]</th>
<th>$E_{avg}$ [fJ]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>228.9</td>
<td>482.8</td>
<td>725.8</td>
<td>496.9</td>
<td>3834</td>
</tr>
<tr>
<td>2</td>
<td>206.4</td>
<td>366.5</td>
<td>512.4</td>
<td>306</td>
<td>3482</td>
</tr>
<tr>
<td>3</td>
<td>212.9</td>
<td>342</td>
<td>454.2</td>
<td>241.3</td>
<td>3587</td>
</tr>
<tr>
<td>4</td>
<td>226.6</td>
<td>339.8</td>
<td>435.7</td>
<td>209.1</td>
<td>3816</td>
</tr>
<tr>
<td>5</td>
<td>243.1</td>
<td>347.6</td>
<td>433.1</td>
<td>190</td>
<td>4096</td>
</tr>
<tr>
<td>6</td>
<td>261</td>
<td>359.7</td>
<td>438.5</td>
<td>177.5</td>
<td>4404</td>
</tr>
<tr>
<td>7</td>
<td>279.7</td>
<td>375.1</td>
<td>448.9</td>
<td>169.2</td>
<td>4731</td>
</tr>
<tr>
<td>8</td>
<td>298.9</td>
<td>391.6</td>
<td>461.6</td>
<td>162.7</td>
<td>5067</td>
</tr>
<tr>
<td>9</td>
<td>318.6</td>
<td>409.8</td>
<td>476.8</td>
<td>158.2</td>
<td>5414</td>
</tr>
<tr>
<td>10</td>
<td>338.5</td>
<td>428.4</td>
<td>493.2</td>
<td>154.7</td>
<td>5765</td>
</tr>
<tr>
<td>11</td>
<td>358.8</td>
<td>447.9</td>
<td>510.8</td>
<td>152</td>
<td>6122</td>
</tr>
<tr>
<td>12</td>
<td>379.5</td>
<td>467.8</td>
<td>529.4</td>
<td>149.9</td>
<td>6483</td>
</tr>
<tr>
<td>13</td>
<td>400.3</td>
<td>488.1</td>
<td>548.4</td>
<td>148.1</td>
<td>6846</td>
</tr>
<tr>
<td>14</td>
<td>421.4</td>
<td>508.7</td>
<td>568.1</td>
<td>146.7</td>
<td>7211</td>
</tr>
<tr>
<td>15</td>
<td>442.7</td>
<td>529.7</td>
<td>588.3</td>
<td>145.6</td>
<td>7579</td>
</tr>
<tr>
<td>16</td>
<td>464.1</td>
<td>550.6</td>
<td>608.6</td>
<td>144.5</td>
<td>7946</td>
</tr>
<tr>
<td>17</td>
<td>485.7</td>
<td>572.0</td>
<td>629.4</td>
<td>143.7</td>
<td>8316</td>
</tr>
<tr>
<td>18</td>
<td>507.5</td>
<td>593.3</td>
<td>650.3</td>
<td>142.8</td>
<td>8686</td>
</tr>
<tr>
<td>19</td>
<td>529.3</td>
<td>614.7</td>
<td>671.3</td>
<td>142.0</td>
<td>9056</td>
</tr>
<tr>
<td>20</td>
<td>551.3</td>
<td>636.5</td>
<td>692.7</td>
<td>141.4</td>
<td>9428</td>
</tr>
</tbody>
</table>

*Table 4. Simulation data for metal 1.*
<table>
<thead>
<tr>
<th># of sections</th>
<th>$t_{d, min}$ [ps]</th>
<th>$t_{d, nom}$ [ps]</th>
<th>$t_{d, max}$ [ps]</th>
<th>$\delta$ [ps]</th>
<th>$E_{avg}$ [fJ]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>203.3</td>
<td>464.5</td>
<td>711.9</td>
<td>508.6</td>
<td>3835</td>
</tr>
<tr>
<td>2</td>
<td>190.7</td>
<td>355.4</td>
<td>504.7</td>
<td>314</td>
<td>3465</td>
</tr>
<tr>
<td>3</td>
<td>200.7</td>
<td>333.3</td>
<td>448.1</td>
<td>247.4</td>
<td>3556</td>
</tr>
<tr>
<td>4</td>
<td>216.2</td>
<td>332</td>
<td>430</td>
<td>213.8</td>
<td>3774</td>
</tr>
<tr>
<td>5</td>
<td>233.6</td>
<td>340.4</td>
<td>427.4</td>
<td>193.8</td>
<td>4045</td>
</tr>
<tr>
<td>6</td>
<td>251.9</td>
<td>352.8</td>
<td>432.7</td>
<td>180.8</td>
<td>4346</td>
</tr>
<tr>
<td>7</td>
<td>270.7</td>
<td>368.5</td>
<td>442.7</td>
<td>172</td>
<td>4667</td>
</tr>
<tr>
<td>8</td>
<td>289.7</td>
<td>385.1</td>
<td>455.3</td>
<td>165.6</td>
<td>5000</td>
</tr>
<tr>
<td>9</td>
<td>309.1</td>
<td>403.5</td>
<td>470.4</td>
<td>161.3</td>
<td>5344</td>
</tr>
<tr>
<td>10</td>
<td>328.6</td>
<td>422.2</td>
<td>486.7</td>
<td>158.1</td>
<td>5694</td>
</tr>
<tr>
<td>11</td>
<td>348.4</td>
<td>441.7</td>
<td>504.3</td>
<td>155.9</td>
<td>6049</td>
</tr>
<tr>
<td>12</td>
<td>368.5</td>
<td>461.6</td>
<td>522.8</td>
<td>154.3</td>
<td>6409</td>
</tr>
<tr>
<td>13</td>
<td>388.8</td>
<td>482.1</td>
<td>542.1</td>
<td>153.3</td>
<td>6772</td>
</tr>
<tr>
<td>14</td>
<td>409.5</td>
<td>502.7</td>
<td>561.9</td>
<td>152.4</td>
<td>7137</td>
</tr>
<tr>
<td>15</td>
<td>430.3</td>
<td>523.5</td>
<td>581.9</td>
<td>151.6</td>
<td>7503</td>
</tr>
<tr>
<td>16</td>
<td>451.4</td>
<td>544.5</td>
<td>602.3</td>
<td>150.9</td>
<td>7871</td>
</tr>
<tr>
<td>17</td>
<td>472.7</td>
<td>565.9</td>
<td>623.2</td>
<td>150.5</td>
<td>8241</td>
</tr>
<tr>
<td>18</td>
<td>494.2</td>
<td>587.2</td>
<td>644.1</td>
<td>149.9</td>
<td>8611</td>
</tr>
<tr>
<td>19</td>
<td>515.9</td>
<td>608.7</td>
<td>665.2</td>
<td>149.3</td>
<td>8981</td>
</tr>
<tr>
<td>20</td>
<td>537.8</td>
<td>630.4</td>
<td>686.8</td>
<td>149</td>
<td>9353</td>
</tr>
</tbody>
</table>

*Table 5. Simulation data for metal 2.*
<table>
<thead>
<tr>
<th># of sections</th>
<th>$t_{d,min}$ [ps]</th>
<th>$t_{d,nom}$ [ps]</th>
<th>$t_{d,max}$ [ps]</th>
<th>$\delta$ [ps]</th>
<th>$E_{avg}$ [fJ]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>105.8</td>
<td>172.2</td>
<td>262</td>
<td>156.2</td>
<td>3211</td>
</tr>
<tr>
<td>2</td>
<td>130</td>
<td>165.2</td>
<td>239.4</td>
<td>109.4</td>
<td>3488</td>
</tr>
<tr>
<td>3</td>
<td>155.3</td>
<td>179.4</td>
<td>248.8</td>
<td>93.5</td>
<td>4042</td>
</tr>
<tr>
<td>4</td>
<td>179.4</td>
<td>197.3</td>
<td>262</td>
<td>82.6</td>
<td>4653</td>
</tr>
<tr>
<td>5</td>
<td>202.2</td>
<td>216.3</td>
<td>276.9</td>
<td>74.7</td>
<td>5287</td>
</tr>
<tr>
<td>6</td>
<td>223.9</td>
<td>236</td>
<td>292.8</td>
<td>68.9</td>
<td>5936</td>
</tr>
<tr>
<td>7</td>
<td>244.7</td>
<td>255.9</td>
<td>309.2</td>
<td>64.5</td>
<td>6597</td>
</tr>
<tr>
<td>8</td>
<td>264.8</td>
<td>276</td>
<td>325.8</td>
<td>61</td>
<td>7266</td>
</tr>
<tr>
<td>9</td>
<td>284.3</td>
<td>296.2</td>
<td>342.6</td>
<td>58.3</td>
<td>7941</td>
</tr>
<tr>
<td>10</td>
<td>303.2</td>
<td>316.4</td>
<td>359.4</td>
<td>56.2</td>
<td>8620</td>
</tr>
<tr>
<td>11</td>
<td>321.6</td>
<td>336</td>
<td>376.2</td>
<td>54.6</td>
<td>9302</td>
</tr>
<tr>
<td>12</td>
<td>339.7</td>
<td>356.9</td>
<td>393.1</td>
<td>53.4</td>
<td>9987</td>
</tr>
<tr>
<td>13</td>
<td>357.4</td>
<td>377.2</td>
<td>410.1</td>
<td>52.7</td>
<td>10670</td>
</tr>
<tr>
<td>14</td>
<td>374.7</td>
<td>397.5</td>
<td>427.1</td>
<td>52.4</td>
<td>11360</td>
</tr>
<tr>
<td>15</td>
<td>391.8</td>
<td>417.9</td>
<td>444.3</td>
<td>52.5</td>
<td>12050</td>
</tr>
</tbody>
</table>

*Table 6. Simulation data for metal 6.*
References

På svenska

Detta dokument hålls tillgängligt på Internet – eller dess framtida ersättare – under en längre tid från publiceringsdatum under förutsättning att inga extra-ordinära omständigheter uppstår.

Tillgång till dokumentet innebär tillstånd för var och en att läsa, ladda ner, skriva ut enstaka kopior för enskilt bruk och att använda det oförändrat för ickeekommerciell forskning och för undervisning. Överföring av upphovsrätten vid en senare tidpunkt kan inte upphäva detta tillstånd. All annan användning av dokumentet kräver upphovsmannens medgivande. För att garantera äktheten, säkerheten och tillgängligheten finns det lösningar av teknisk och administrativ art.

Upphovsmannens ideella rätt innefattar rätt att bli nämnd som upphovsman i den omfattning som god sed kräver vid användning av dokumentet på ovan beskrivna sätt samt skydd mot att dokumentet ändras eller presenteras i sådan form eller i sådant sammanhang som är kränkande för upphovsmannens litterära eller konstnärliga anseende eller egenart.

För ytterligare information om Linköping University Electronic Press se förlagets hemsida http://www.ep.liu.se/

In English

The publishers will keep this document online on the Internet - or its possible replacement - for a considerable time from the date of publication barring exceptional circumstances.

The online availability of the document implies a permanent permission for anyone to read, to download, to print out single copies for your own use and to use it unchanged for any non-commercial research and educational purpose. Subsequent transfers of copyright cannot revoke this permission. All other uses of the document are conditional on the consent of the copyright owner. The publisher has taken technical and administrative measures to assure authenticity, security and accessibility.

According to intellectual property law the author has the right to be mentioned when his/her work is accessed as described above and to be protected against infringement.

For additional information about the Linköping University Electronic Press and its procedures for publication and for assurance of document integrity, please refer to its WWW home page: http://www.ep.liu.se/

© Rebecca Källsten