High-Level Energy Estimation for Sub-Micrometric TSV Arrays

Lennart Bamberg, Student Member, IEEE, and Alberto Garcia-Ortiz, Member, IEEE

Abstract—Three-dimensional integration using through silicon vias (TSVs) is one of the most promising approaches to overcome the interconnect delay problem of current CMOS technologies. Nevertheless, the TSV energy consumption is not negligible due to the high capacitive coupling. This work presents an abstract and yet accurate model to estimate the pattern dependent energy consumption in arrays of TSVs; it is the first high-level model including the effects of the voltage dependent MOS capacitances surrounding each TSV and a possible temporal misalignment between the input signals. We propose a regression method to estimate the dynamic size of the coupling capacitances as a function of the bit probabilities. Experimental results for real and synthetic data streams, a submicron 9-bit TSV array and a 65 nm technology show that the presented TSV energy model exhibits a maximum error of 5.53 %, while the traditional high-level model shows errors of up to 79.77 %. Further more, the new insights provided by our model reveal a possibility to easily boost the efficiency of existing low-power codes for TSV structures by over 10 % without affecting the coding efficiency for the planar metal-wires or the encoder complexity.

Index Terms—through silicon vias (TSVs), 3D integration, 3D ICs, high-level power/energy estimation, low-power coding

I. INTRODUCTION

With shrinking technologies, the interconnects become the bottleneck of a traditional system on chip, due to the increasing delay and energy gap between the interconnects and the transistors [1]. The interconnect delay problem can be overcome by exploiting the integration into the third dimension [2]. Furthermore, a 3D integrated circuit (3D IC) enables a heterogeneous integration and a reduction of the overall system footprint. The interconnect structure between the several dies of a 3D IC typically employs through silicon vias (TSVs), as they provide a high reliability and a short delay. TSVs are suitable for ultra high frequency applications because of the small TSV resistance per unit-length [3]. However, this low-resistance does not help to reduce the energy consumption. The energy consumption of TSVs is mainly determined by the capacitive coupling [4], which is not negligible because of the large TSV dimensions and the increased number of aggressors compared to the traditional metal-wires [5]. Therefore, the interconnects still have a significant impact on the energy consumption of a 3D IC, which means that the need for methods to reduce their capacitive coupling is not alleviated.

Two complementary approaches exist to reduce capacitive coupling and consequently the energy consumption. The first one, addressing the physical level, improves the layout and the electrical characteristics of the TSVs. In [6] the authors present a layer ordering algorithm which drastically reduces the total number of TSVs in a 3D IC. Consequently, it reduces the overall energy consumption caused by the TSVs. Hexagonal TSV bundles, proposed in [7], reduce the energy and peak power consumption as well as the area of a TSV array. Already known methods for the optimization of 2D wires, like spacing and shielding, have been proposed for TSV structures [8], [9]. Both methods lead to a huge area overhead and are therefore not suitable for large TSV arrays, which already have an enormous footprint. Additionally, passive shielding only reduces the peak power and not the energy consumption.

A second approach addresses the higher levels of abstraction. Since coupling is a pattern dependent phenomena, it can be reduced by modifying the transmitted data, using a coding process [10]. 2D circuit design shows that the coding approach is usually more efficient, in terms of coupling reduction and area overhead, than the first approach [10]. An abstract and yet accurate model, accessible at the high levels of abstraction, mapping the pattern properties to the energy consumption, is required for the derivation and a fast evaluation of these coding schemes. Partially due to the lack of such a model, no fully satisfying low-power coding approach for TSVs exists in the reported literature. Existing encoding schemes [11]–[13] are designed to reduce the maximum coupling (reduces i.a. peak power consumption) per TSV, but, due to their overhead, they even increase the average TSV energy consumption.

Many previous works present formulas to extract the TSV parasitics [5], [14]–[21]. Many of these works [14]–[17] consider arrangements of two TSVs and are consequently not valid for TSV arrays due to the E-field sharing effect [5]. Although, existing formulas to extract the parasitic elements in a TSV array [5], [18]–[21] enable simulating the pattern dependent TSV energy consumption with circuit simulators, the need for a high-level model is also inevitable for system simulators, typically implemented in high-level languages (e.g. SystemC) [22], [23]. The interconnect architecture of a modern processor system has a throughput of several GB/s. This data rate makes it almost impossible to perform a simulation of the application specific energy requirements of such a system at the circuit level. For the Spectre circuit simulation and the transmission of only 3500 patterns with a data rate of 1.1 GB/s over a single modern 3×3 TSV array we obtain a computation time of ca. 1130 s on a Linux Xeon E5 machine with 512 GB of RAM. In contrast a high-level estimation only takes fractions of a second, independent of pattern count and data rate. For system simulations with data sets, containing millions of patterns, this means a speed up of more than 10⁶.
In this work, we present the first accurate high-level model to estimate the energy consumption in TSV arrays. The high-level model used to validate previous coding schemes [11], [12] considers the dependency of the effective capacitance values on the direction of the switching on the adjacent lines, but assumes that the physical capacitances in a TSV array are constant. Especially for TSV arrays in the submicron range, this is imprecise, because of the voltage dependent dynamic MOS capacitance surrounding each TSV [17], [24]. Furthermore, the traditional model for the TSV energy consumption does not consider a possible temporal misalignment between the input signals of the TSVs, which affects the pattern dependency of the energy consumption significantly, as shown in this work. Therefore, coding schemes derived by the previous model are often impractical for real applications. The model presented in this work considers the temporal misalignment effect and uses either a quadratic or linear regression to model the dynamic behavior of the coupling capacitances as a function of the bit probabilities, while keeping the model complexity low. Thus, the resulting model discloses a new optimization possibility for TSV coding schemes: The logical bit probabilities.

We validate the presented abstract model by comparing it to reference results obtained through electro magnetic (EM) simulations in combination with a circuit simulator. Focusing on a modern 9-bit TSV array, we consider 256 synthetic data streams, all with different stochastic properties, and we cover the cases of signal alignment and misalignment. For this wide set of experiments, the resulting normalized root-mean-square derivation (NRMSD) of the high-level estimation is always between 1% and 2%, independent of the used regression method for the capacitance values. The resulting maximum error for both regression methods is approximately 5%.

Additionally, a case study is presented to highlight the practical value of our model. It presents a modification, revealed by the presented model, to improve existing low-power coding architectures for TSV interconnects by up to 11%, without increasing the coder complexity or decreasing the coding gain for 2D interconnects. Therefore, the modification makes the coding schemes suitable for 2D and TSV interconnects simultaneously, without having any drawbacks.

The rest of this work is organized as follows: Section II includes the preliminaries for the derivation of the high-level model which is subsequently presented in Section III. The model is analyzed by simulation results in Section IV, which show the accuracy of the capacitance and the resulting energy model, as well as the need to consider the effect of a temporal misalignment. Subsequently, in Section V the case study is presented. Finally, a conclusion is drawn in Section VI.

II. PRELIMINARIES

In this section, we introduce the considered TSV array structure. An exemplary $2 \times 2$ TSV array is depicted in Fig. 1 (a). The TSVs are homogeneously placed in the $M \times N$ array and indexed as TSV$_i$, where $m_i = \text{mod}(i, N)$ and $n_i = \text{ceil}(i/N)$ are the column and row location, respectively. The distance between the centers of direct adjacent vias is constant and denoted by $d_{\text{min}}$. Thus, the distance between any diagonal adjacent vias is $\sqrt{2}d_{\text{min}}$. The cylindrical or conical TSVs of length $l_{\text{tsv}}$ and radius $r_{\text{tsv}}$, are made up of a low-impedance material like copper. The TSVs mainly traverse through the doped silicon substrate. Due to the metal bonding, they also traverse thin SiO$_2$ dielectric layers. For DC insulation, a TSV is surrounded by a SiO$_2$ dielectric. The resulting cross sections of a TSV crossing a substrate and a metal layer are illustrated in Fig. 1 (b) and (c), respectively.

TSV, dielectric and substrate form a metal oxide semiconductor (MOS) junction (Fig. 1 (b)). Therefore, in the substrate, a TSV is surrounded by a depletion region, if the average voltage between the TSV and the substrate exceeds the Flatband voltage [17]. This flatband voltage is typically negative [16]. Thus, for an average voltage of 0 V, a depletion region already exists. The width of the depletion region surrounding TSV$_i$ depends on the average voltage between the TSV$_i$ and the substrate, and can be represented in an equivalent circuit by a dynamic capacitance between the oxide and the substrate [15]. For the typical p-doped substrate, this capacitance, and as a consequence the connected coupling capacitances, decrease with an increasing TSV voltage. For an n-doped silicon, the capacitances increase with an increasing TSV voltage. The serial combination of the oxide and the depletion region capacitance is denoted as the (dynamic) MOS capacitance. These MOS capacitances, in combination with the substrate admittances, cause the capacitive coupling of the TSVs. The coupling/crosstalk between non-adjacent TSVs can be neglected because of the Faraday cage effect [18]. Therefore, as shown in Fig. 2, a TSV has up to eight connected coupling capacitances [7]. A coupling capacitance between diagonal adjacent TSVs (hereinafter denoted as diagonal capacitance) is always smaller than a coupling capacitance between direct adjacent TSVs (hereinafter denoted as direct capacitance), because of the different inter-TSV pitches.

![Fig. 1. (a) Fragment of a TSV array in 3D perspective, (b) TSV through substrate cross section and (c) TSV through metalization cross section](image-url)
calculated using the 1-bit probabilities $p_i$ as constant. Therefore, the average TSV voltages can be seen while transmitting one given data stream, can be seen if the input signal switching and the capacitance values. Finally, we combine these formulas with the TSV capacitance model to obtain the formulas for the TSV energy estimation.

A. Capacitance Modeling

Previously used TSV coupling models [11]–[13], assume that the coupling capacitances are constant and that the ratio between a diagonal capacitance and a direct capacitance is $1/4$. In general, the ratio between the capacitance values depends on the geometry and technology parameters. Moreover, the size of a coupling capacitance depends on the depletion region widths. In fact, as shown in the later analysis, the TSV voltages can influence the coupling capacitance sizes in modern TSV structures by more than 40%.

In principle, the average TSV voltage for the calculation of a depletion region width has to be calculated over a sliding window, and the formulas to find the correct window duration show a huge complexity. In this work, we assume that all data streams are stationary. In this particular case, the average TSV voltages $V_i$ can be calculated using the 1-bit probabilities $p_i$:

$$V_i = p_i V_{dd}.$$  
(1)

The resulting dependency between the size of the coupling capacitances and the probabilities is quite complex. To keep the model compact, we propose to approximate the dependency by a polynomial of the first or second degree. These polynomial coefficients, called $\vec{\beta}$, quantify how much the capacitance values are reduced as a function of the 1-bit probabilities of the transmitted data stream. We define the average 1-bit probabilities on two TSVs $i$ and $j$ as $p_{i,j} = (p_i + p_j)/2$.

B. Dynamic Energy Consumption

Finally, we obtain the following compact formula to estimate the size of the coupling capacitances in a TSV array:

$$C_{i,j} = \begin{cases} C_r(1 - \beta_n^T \cdot \vec{p}_{i,j}) & \text{for } d_{i,j} = d_{min} \\ C_d(1 - \beta_d^T \cdot \vec{p}_{i,j}) & \text{for } d_{i,j} = \sqrt{2}d_{min} \\ 0 & \text{else} \end{cases}$$  
(2)

where $\vec{p}_{i,j}$ is equal to $[p_{i,j} p_{i,j}^T]^T$, for the quadratic regression, and to $[p_{i,j}]$, for the linear regression. Further on, $d_{i,j}$ is the distance between the TSVs $i$ and $j$, while $C_r$ and $C_d$ are the sizes of a direct capacitance and diagonal capacitance for an average voltage of 0 V on all TSVs, respectively.

Besides the coupling capacitances, we have to consider the TSV-ground capacitances as well. A TSV-ground capacitance is composed of the load capacitance and the capacitance between the TSV and the bulk. In general, the TSV-bulk capacitance $C_{i,BULK}$ is significantly smaller than the load capacitance $C_l$. Therefore, the TSV-ground capacitance is approximately equal to the load capacitance.

$$C_{i,0} = C_{i,BULK} + C_l \approx C_l.$$  
(3)

In this work, we assume that the load capacitances of all TSVs are equal. Hence, we model all TSV ground capacitances by one constant value.

Formulas for modeling the dynamic power consumption in 2D interconnects are well known (e.g. [4], [10]). However, they consider only two aggressors and neglect the effect of misalignments among signals. In this subsection, we develop a model that overcomes those limitations. To derive these general formulas, we consider Fig. 3. Neglecting the leakage, the energy extracted from driver $r$ in the $n^{th}$ clock cycle is:

$$E_{e,r} = \int_{T_{clk}}^{T_{clk}} V_{dd} i_r(t) dt$$  
(4)

where $V_{dd}$ is the supply voltage, $i_r$ is the current flow through the driver $r$, and $T_{clk}$ is the clock period. If the binary input $b_r$ is zero, the current $i_r$, and therefore also the
energy consumption of the supply, is zero. Since $i_r = \sum_k i_{r,k}$ (Kirchhoff’s law), we can analyze the contribution of each capacitance individually.

Let us first assume that there is no temporal misalignment between the signal edges at the input ports of the drivers $r$ and $s$. In this case, by considering the current voltage relation of a capacitance, the energy extracted from driver $r$ due to the coupling capacitance between the interconnects $r$ and $s$ can be calculated as follows:

$$E_{e,r,s} = b_r^+ \int_{t=(n-1)T_{cik}}^{nT_{cik}} V_{dd} C_{r,s} \frac{\partial}{\partial t}(v_r(t) - v_s(t))dt$$

$$= V_{dd} C_{r,s} b_r^+ (V_{dd} b_r^+ - V_{dd} b_r^- + V_{dd} b_s^-)$$

$$= V_{dd} C_{r,s} b_r^+ (\Delta b_r - \Delta b_s)$$  \hspace{1cm} (5)

where $v_i$ is the voltage between node $i$ and the ground, which is equal to the binary value on interconnect $i$ times $V_{dd}$; $b_r^+$ and $b_r^-$ are the binary values on interconnect $i$ in the actual and the previous clock cycle, respectively. $\Delta b_i$ is equal to $b_i^+ - b_i^-$. Analogously, we can calculate the energy extracted from driver $s$ because of $C_{r,s}$:

$$E_{e,s,r} = V_{dd} C_{r,s} b_s^+ (\Delta b_s - \Delta b_r).$$  \hspace{1cm} (6)

Consequently, the total extracted energy because of $C_{r,s}$ is:

$$E_{e,r,s} = E_{e,r,s} + E_{e,s,r} = V_{dd} C_{r,s} (b_r^+ - b_s^+)(\Delta b_r - \Delta b_s).$$  \hspace{1cm} (7)

Eq. (7) is not valid when a temporal misalignment between the two input signals occurs. Let us assume that the signal edges on interconnect $r$ are delayed by $\delta_r$ against the signal edges on interconnect $s$. If $\delta_r$ is greater than the rise/fall time of the input drivers, plus the maximum delay over one of the two interconnects, Eq. (5) has to be modified to:

$$E_{e,r,s} = b_r^+ \int_{t=(n-1)T_{cik}+\delta_r}^{nT_{cik}} V_{dd} C_{r,s} \frac{\partial}{\partial t}(v_r(t) - v_s(t))dt$$

$$+ b_r^+ \int_{t=(n-1)T_{cik}}^{nT_{cik}} V_{dd} C_{r,s} \frac{\partial}{\partial t}(v_r(t) - v_s(t))dt$$

$$= V_{dd} C_{r,s} b_r^+ (\Delta b_r - \Delta b_s).$$  \hspace{1cm} (8)

$E_{e,s,r}$ can still be calculated with Eq. (6) if the input signal of interconnect $r$ is delayed. Thus, the total extracted energy because of $C_{r,s}$ is:

$$E_{e,r,s} = V_{dd} C_{r,s} (\Delta b_r (b_r^+ - b_s^+) + \Delta b_s (b_s^+ - b_r^-)).$$  \hspace{1cm} (9)

The resulting formulas get simpler if we consider the dissipated energy instead of the extracted energy. On average the two energy quantities are equal; thus it is irrelevant which quantity is used. The dissipated energy is the difference between the differential stored energy and the extracted energy:

$$E_{r,s} = E_{e,r,s} - \Delta E_{stor,r,s}.$$

The difference between the energies stored in $C_{r,s}$ after and before the transitions is:

$$\Delta E_{stor,r,s} = \frac{1}{2} C_{r,s} (V_{r,s}^+ - V_{r,s}^-)$$

$$= \frac{V_{dd}^2 C_{r,s}}{2} ( (b_r^+ - b_r^-)^2 - (b_r^- - b_r^+)^2 )$$

$$= \frac{V_{dd}^2 C_{r,s}}{2} ( b_r^+ - b_r^- + b_r^- - b_r^+ ) (\Delta b_r - \Delta b_s)$$  \hspace{1cm} (11 b)

where $V_{r,s}^+$ and $V_{r,s}^-$ are the voltage drops over $C_{r,s}$ before and after the transitions, respectively. Eq. (11 c) can be derived from Eq. (11 b) by applying the third binomial formula. For perfectly aligned signals, the resulting formula to calculate the dissipated energy due to $C_{r,s}$ is:

$$E_{r,s} = \frac{V_{dd}^2 C_{r,s}}{2} (\Delta b_r - \Delta b_s)^2.$$  \hspace{1cm} (12)

For the delayed input signal on interconnect $r$, we obtain for the dissipated energy:

$$E_{r,s} = \frac{V_{dd}^2 C_{r,s}}{2} (\Delta b_r^2 + \Delta b_s^2).$$  \hspace{1cm} (13)

By analyzing Eq. (12) and Eq. (13), we see that for the dissipated energy, it is irrelevant which interconnect changes its level first. The only important issue is whether the edges show a sufficient temporal misalignment or not. Another advantage of using the dissipated energy, is that we do not have to differentiate between a charging and a discharging capacitor. Therefore, the formulas for the dissipated energy exhibit a lower complexity.

By substituting $s=0$ and $\Delta b_s=0$ in Eq. (12) or (13), we obtain for the energy dissipated by the capacitance between node $r$ and the ground:

$$E_{r,0} = \frac{V_{dd}^2 C_{r,0}}{2} \Delta b_r^2.$$  \hspace{1cm} (14)

Summarized, the overall dissipated energy of an interconnect architecture in a clock cycle, when all edges are aligned, is:

$$E_{al} = \frac{V_{dd}^2}{2} \left( \sum_{i,j} C_{i,j} (\Delta b_i - \Delta b_j)^2 + \sum_i C_{i,0} \Delta b_i^2 \right).$$ \hspace{1cm} (15)

For a sufficient misalignment between all adjacent interconnects, the dissipated energy is equal to:

$$E_{mis} = \frac{V_{dd}^2}{2} \left( \sum_{i,j} C_{i,j} (\Delta b_i^2 + \Delta b_j^2) + \sum_i C_{i,0} \Delta b_i^2 \right).$$  \hspace{1cm} (16)

Until now, the derived formulas are valid for a 2D metal-wire as well as a 3D TSV interconnect structure. In a traditional metal-wire structure a coupling capacitance $C_{c}$ exists between all adjacent wires and each wire has a ground capacitance $C_{0} = 10$. With $\lambda_{2D} = C_{0}/C_{c}$ we obtain the following formula for the energy dissipation of an $N$-bit metal-wire interconnect structure and aligned input signals:

$$E_{2D,al} = \frac{V_{dd}^2 C_{c}}{2} \left( \sum_{i=1}^{N-1} (\Delta b_{i} - \Delta b_{i+1})^2 + \lambda_{2D} N \sum_{i=1}^{N} \Delta b_i^2 \right).$$  \hspace{1cm} (17)
For misaligned signals on the adjacent metal-wires, we obtain:

\[
E_{2D,\text{mis}} = \frac{V_d^2 C}{2} \left( \sum_{i=1}^{N-1} (\Delta b_i^2 + \Delta b_{i+1}^2) + \lambda_2 D \sum_{i=1}^{N} \Delta b_i^2 \right) .
\] (18)

To derive a formula for the TSV energy dissipation, we combine Eq. (15) and Eq. (16) with the TSV capacitance model presented in Subsection A.

Under the assumption of aligned input signals, we obtain:

\[
E_{3D,\text{all}} = \frac{V_d^2 C_n}{2} \left( \sum_{i,j} (\Delta b_i - \Delta b_j)^2 (1 - \beta_d^T \cdot \bar{p}_{i,j}) + \lambda_1 \sum_{i,j} (\Delta b_i - \Delta b_j)^2 (1 - \beta_n^T \cdot \bar{p}_{i,j}) + \lambda_2 \sum_{i=1}^{N} \Delta b_i^2 \right) .
\] (19)

For misaligned signals on adjacent TSVs, the formula to estimate the TSV energy dissipation is:

\[
E_{3D,\text{mis}} = \frac{V_d^2 C_n}{2} \left( \sum_{i,j} (\Delta b_i^2 + \Delta b_j^2)(1 - \beta_d^T \cdot \bar{p}_{i,j}) + \lambda_1 \sum_{i,j} (\Delta b_i^2 + \Delta b_j^2)(1 - \beta_n^T \cdot \bar{p}_{i,j}) + \lambda_2 \sum_{i=1}^{N} \Delta b_i^2 \right) .
\] (20)

In Eq. (19) and Eq. (20) \( \lambda_1 \) is \( C_d/C_n \) and \( \lambda_2 \) is \( C_l/C_n \). \( S_{\text{dir}} \) is the set of direct adjacent TSV pairs, and \( S_{\text{diag}} \) is the set of diagonal adjacent TSV pairs.

In general, the key concern is not the energy dissipation in a specific clock cycle, but rather the mean energy consumption (\( \bar{E} \)) per clock cycle. The formula to estimate this quantity can be derived with the help of the expectation operator \( E\{ \} \) from the formula for the dissipated or extracted energy. For the 2D metal-wires we obtain:

\[
\bar{E}_{2D} = E\{ E_{2D} \} = \frac{V_d^2 C}{2} \left( \sum_{i=1}^{N-1} t_{c,i,i+1} + \lambda_2 D \sum_{i=1}^{N} t_{s,i} \right) .
\] (21)

And for the 3D TSVs:

\[
\bar{E}_{3D} = \frac{V_d^2 C_n}{2} \left( \sum_{i,j} t_{c,i,j} (1 - \beta_n^T \cdot \bar{p}_{i,j}) + \lambda_1 \sum_{i,j} t_{c,i,j} (1 - \beta_d^T \cdot \bar{p}_{i,j}) + \lambda_2 \sum_{i=1}^{N} t_{s,i} \right) ,
\] (22)

where \( t_{s,i} \) is the toggle activity of interconnect \( i \) (\( E\{ \Delta b_i^2 \} \)) and \( t_{c,i,j} \) is the expected value of the coupling switching between the interconnects \( i \) and \( j \). If the temporal misalignment between the two input signals is zero, \( t_{c,i,j} \) is equal to:

\[
t_{c,i,j,\text{all}} = E\{ (\Delta b_i - \Delta b_j)^2 \} .
\] (23)

For a sufficient misalignment, it is equal to:

\[
t_{c,i,j,\text{mis}} = E\{ \Delta b_i^2 \} + E\{ \Delta b_j^2 \} = t_{s,i} + t_{s,j} .
\] (24)

### IV. Simulation Results and Analysis

In this section, we analyze the accuracy of the presented model to estimate the TSV capacitance values in modern TSV structures. In addition, we report the values for the \( \beta \)- and \( \lambda \)-coefficients, as well as the required temporal misalignment to use Eq. (23) instead of Eq. (24), for the energy estimation. Finally, we determine the accuracy of the derived high-level model to estimate the TSV energy consumption under different synthetic data stream scenarios. To obtain reference values, the results are compared with the values for the traditional model used in [11], [13]. For all analyzed structures, the relative error of the electric field within the quasi-static approximation is smaller than 0.064 %, according to [25]. Even a significant increase in the TSV signal frequency to 50 GHz would keep the maximum approximation error under 1.7 %. Therefore, all analyses are performed with the help of the electromagnetic field solver, Ansys Q3D Extractor, which enables quasi-static electromagnetic parasitic extractions for different depletion region widths. For this purpose, a 3D model of the TSV array is created according to Fig. 1. In this model the TSVs are made of copper and the p-doped (Boron) substrate is biased at 0 V. In the substrate, the dopant concentration is \( 1.35\times10^{15} \text{ cm}^{-3} \), which leads to a substrate conductivity of approximately \( 10 \text{ S/m} \). We use several values for the radius \( r_{\text{tsv}} \) and minimum pitch \( d_{\text{min}} \) of the TSVs, based on the TSV dimensions projected for the 2015-2018 time frame of the ITRS 2012 [26]. A fixed TSV length, \( l_{\text{tsv}} \), of 30 \( \mu \text{m} \) is chosen, which corresponds to the reported minimum global TSV depth (20 \( \mu \text{m} - 50 \mu \text{m} \)). The ITRS did not report the expected thickness of the \( \text{SiO}_2 \) surrounding each TSV. For all dielectrics, we assume a thickness of \( r_{\text{tsv}}/5 \), which is a realistic value according to existing process nodes. The depletion region is modeled as a fully depleted area (\( \sigma = 0 \)). To determine the width of the depletion region surrounding a TSV, the exact formulas [16], based on the Poisson’s equation, are applied. The parasitic elements of a TSV array show a slight frequency dependency. For the extraction, an average input signal frequency of 6 GHz is assumed.

The high-level energy estimations and the generation of the data streams are performed with MATLAB. To obtain reference values for the energy consumption, we performed Spectre circuit simulations with the extracted parasitic elements. For these simulations, we employed standard drivers stemming from a commercial 65 nm technology at the input and the output of the TSV arrays.

#### A. Accuracy of the Capacitance Model

To determine the accuracy of the capacitance model, the coupling capacitances in the middle of a 5×5 array are considered. With the Q3D Extractor, the capacitance values are extracted for two scenarios. In the first scenario the bit probabilities on all TSVs are assumed to be equal. In the second scenario, the 1-bit probability on one of the two TSVs, which are connected by the considered capacitance, differs from the remaining 1-bit probabilities, which are equal. Consequently, we extract multiple capacitance values for several \( p_{i,j} \) values. As an example, consider an analysis where the
Fig. 4. Extracted capacitance values (Q3D) and fitted capacitance models (\(r_{TSV}=0.5\ \mu m, d_{min}=2.5\ \mu m\))

Fig. 5. Extracted capacitance values (Q3D) and fitted capacitance models (\(r_{TSV}=0.5\ \mu m, d_{min}=2.5\ \mu m\)) for equal bit probabilities on all TSVs.

bit probabilities can only be 0, 0.5 and 1. In this case, five capacitance values, for each capacitance, would be extracted for \(p_{i,j}=0.5\). Two values for a 1-bit probability of 1 on one of the two TSVs, while all remaining 1-bit probabilities are 0, two values for a 1-bit probability of 0 on one TSVs and a 1-bit probability of 1 on all other TSVs, and one value for \(p_{i,j}=0.5\) on all TSVs. For this analysis, we vary the probability values between 0 and 1 with a step size of 0.1. Thus, for each analyzed TSV array structure we extract 242 different capacitance values. 121 for the diagonal capacitances, and 121 for the direct capacitances. All extracted capacitance values are subsequently used to perform two linear and two quadratic regression analyses.

First, linear and quadratic least squares fits are applied by only considering the 11 direct as the 11 diagonal capacitance values extracted for equal bit probabilities on all TSVs. Additionally, we determine the constant coupling capacitance values for the traditional model. The fitting process for \(r_{TSV}=0.5\ \mu m\) and \(d_{min}=2.5\ \mu m\) is illustrated in Fig. 4. In this scenario the quadratic fit is nearly perfect, as the absolute percentage error (APE) of the fitted model, compared to all extracted capacitance values, does not exceed 1.02%. The linear fit still shows a good accuracy, especially for the diagonal capacitances. But the traditional model reveals APE values of up to 17.65% for the direct, and 49.16% for the diagonal capacitances.

Secondly, linear and quadratic fits are applied while considering all 242 extracted capacitance values. This analysis quantifies the error caused by the averaging of the two bit probabilities \(p_i\) and \(p_j\), while all remaining bit probabilities are neglected. Fig. 5 illustrates the fitting for \(r_{TSV}=0.5\ \mu m, d_{min}=2.5\ \mu m\). Here the maximum APE values of the quadratic and the linear regression method are 6.09% and 5.23%, respectively. For the linear regression, the biggest errors occur for \(p_{i,j}=1\) and \(p_{i,j}=0\), where the variation of the capacitance values is maximal. The biggest errors for the quadratic regression occur for strongly asymmetric scenarios (e.g. \(p_i=1, p_k=0\) for all \(k \neq i\)) where the inaccuracy due to the averaging of the two bit probabilities and the neglected remaining probabilities is maximized. Even in those cases, the error is very modest.

Table I summarizes the fitted model parameters for the quadratic, the linear and the traditional model for all analyzed TSV dimensions. The goodness of the fitted models, compared to all extracted values, is included in Table II. Table II reveals that the traditional model becomes more and more inaccurate with shrinking TSV dimensions. For example: for a TSV diameter of 2.0 \(\mu m\), a minimum TSV pitch of 5.0 \(\mu m\), the maximum APE of the traditional model is 5.34% for the direct capacitances and 13.17% for the diagonal capacitances. For a diameter of 0.5 \(\mu m\) and a pitch of 2.5 \(\mu m\), these values increase to 17.65% and 49.18%. Table I reveals the two reasons for this dramatic increase. On the one hand, the assumption of an approximate 1/4 ratio between the direct capacitances and the diagonal capacitances is not accurate for TSV structures in the submicron area (tends to be rather 1/3-1/2). On the other hand, the MOS effect can reduce the size of the coupling capacitances in modern TSV structures by over 40%. For recent TSV dimensions, the linear regression shows a good
balance between complexity and accuracy, as the APE and the NRMSD do not exceed 9.79% and 4.44%, respectively. An accurate modeling of the bigger direct capacitances is more important. For these capacitance values, the APE and NRMSD of the linear regression do not exceed 6.09% and 1.84%, respectively.

In future TSV structures, the effect of the decreasing derivative of the capacitance curve with increasing probability values gains relevance, because of possible overlapping depletion regions, or depletion regions whose maximum width is reached before $V_{dd}$. For these future TSV structures and (almost) equal bit probabilities on the TSVs, a quadratic capacitance model has a significant gain over a linear model. But, for strongly varying bit probabilities on the TSVs, this improvement is rather limited, because the error of the linear model, due to the varying first order derivative.

B. Time Constants of the TSV Arrays

In this subsection, we report the time constants $\tau$ of the analyzed TSV arrays. The time constant is equal to the delay (0% to 63.2% rise/fall time) over the TSVs when the signals are completely misaligned. Therefore, the sum of the rise/fall-time of the input signal and the time constant ($\tau + t_{rise/fall}$) can be used to estimate the required misalignment to consider all connected coupling capacitances as ground capacitances (complete misalignment).

The time constants, resulting from the extracted parasitic elements, are reported in Table III. For all analyzed TSV arrays, the time constant is in the fs range, although, the sum of the capacitances connected to a TSV is bigger than 22 fF, in some cases. The reason for this, is the small TSV resistance, which in all considered scenarios does not exceed 1 Ω. In comparison, the time constants of a typical 2D metal-wire segment is several times bigger, due to its resistance, which can easily exceed 100 Ω.

Consequently, in many cases, the derived energy model for misaligned signals has to be used for TSV interconnects, while for the metal-wire interconnects, the formulas for aligned signals still show a higher accuracy.
C. Accuracy of the Energy Model

Finally, we assess the accuracy of the overall model for the TSV energy estimation. Therefore, we consider the energy consumption per transmitted bit of a 3×3 TSV array (dimensions: $r_{TSV}=0.5\ \mu m$ and $d_{min}=2.5\ \mu m$) while transmitting several synthetic data streams. All data streams have a length of 3500 patterns, and are transmitted with a bit rate of 1.1 GB/s. For the analysis, we distinguish between the odd and the even TSVs, as illustrated in Fig. 6.

In total, we analyze 256 data streams, which are grouped in four different sets:

#1 119 data streams. For each data stream, the 1-bit probabilities are $p_{even}$ on all even TSVs and $p_{odd}$ on all odd TSVs. $p_{odd}$ and $p_{even}$ go from 0 to 1 with a step size of 0.1. All bits are spatially uncorrelated

#2 119 data streams. For each data stream, the 1-bit probabilities are $p_{even}$ on all even TSVs and $p_{odd}$ on all odd TSVs. $p_{odd}$ and $p_{even}$ go from 0 to 1 with a step size of 0.1. All odd and all even bits are perfectly spatially correlated

#3 9 data streams. For each data stream, all 1-bit probabilities are equal $p$. $p$ goes from 0.1 to 0.9, with a step size of 0.1. All bits are spatially uncorrelated

#4 9 data streams. For each data stream, all 1-bit probabilities are equal $p$. $p$ goes from 0.1 to 0.9 with a step size of 0.1. All bits are perfectly spatially correlated

Here, a perfect spatial correlation between bits means that their binary values are equal in every clock cycle. Hence, if one of the correlated bits toggles, all toggle.

The streams are defined to highlight different scenarios with temporal and spatial correlation. In the individual data sets, we only vary the bit probabilities, because this variation automatically causes a variation of the self switching ($\Delta b^2$). Besides the self switching, the coupling switching, for aligned signal edges, is dependent on the bitwise spatial correlation. This spatial correlation is zero for uniformly distributed signals. For normally distributed signals represented in the two’s complement, the LSBs are uncorrelated, while the MSBs are almost perfectly spatially correlated [27]. Therefore, the MSBs of a normally distributed signal show a sign character and they toggle with a probability dependent on the temporal pattern correlation. Hence, the analyzed data streams cover the key scenarios found during the transmission of real data streams.

In total, we perform 512 Spectre simulations, because each data stream is used for two energy analyses. In the first analysis, all input signals are perfectly aligned, and in the second analysis, the signals on all odd TSVs are delayed by $\delta_{odd}=0.1\ ns$. This results in a temporal misalignment between all direct adjacent TSV signals. For each Spectre simulation, we determine the total TSV energy consumption (including the leakage) and the difference to the estimated values, according to the presented and the traditional high-level formulas. For the four data stream sets (and for all data streams together), the maximum APE and the NRMSD of the estimates are determined for the aligned and the misaligned signals individually.

The results are reported in Table IV. For all analyzed scenarios, the presented model exhibits a significantly higher accuracy than the traditional model, except for the aligned signals and data stream set #1. In the corresponding signal streams, the coupling switching is always zero, so only the ground capacitances affect the energy consumption. Consequently, the aligned transmission of this data set leads to the least energy consumption. This results in high APE and NRMSD values for the presented energy model because the influence of the non-considered drivers (leakage) increases. For the remaining data stream sets, the size of the coupling capacitances has an influence on the energy consumption and, as a consequence, the presented model reveals a significant improvement over the traditional model. The highest energy consumption appears for the data stream set #1. For the aligned signals, the improvement in the maximum APE and NRMSD of the presented quadratic model over the traditional model is 15.59 percentage points (pp) and 7.68 pp, respectively. The linear model shows a similar improvement (15.72 pp and 7.68 pp). For the data stream set #2 and the aligned signals, the coupling switching between all diagonal adjacent TSVs is zero. In this case the error of the traditional model is only caused by neglecting the dynamic capacitance values. Nevertheless, the presented linear/quadratic model still shows a maximum APE and NRMSD improvement of ca. 9.7 pp and 3.6 pp, respectively. Further on, Table IV shows that the neglected misalignment effect in the traditional model leads to big errors for correlated data streams (sets #2 and #4). The spatial correlation leads to an expected coupling switching of zero for aligned signals, while it is $t_{s,i}+t_{s,j}$ for completely misaligned signals. This leads to an energy underestimation of up to 79.77 %, and to an NRMSD of 82.3 %, for set #4.

In total, for all 512 analyzed signals, the estimation errors of the presented model do not exceed 5.53 % and 4.79 %, respectively. The improvement of the quadratic model over the linear model is relatively low, and generally does not com-

![Fig. 6. Odd and even TSVs in a 3×3 array](image-url)
pensate for the additional model complexity. For all aligned and all misaligned signals, the presented model exhibits an NRMSD of 1 %-2 %, independent of the regression method. In comparison the traditional model shows an NRMSD of 7.89 % for the aligned signals, and 22.95 % for the misaligned signals.

To validate the model for other technology nodes, we repeat the Spectre simulations for data stream set #3 with 22 nm Predictive-Technology-Model (PTM) drivers, instead of the 65 nm drivers. The resulting NRMSD and MAE of the linear model, for aligned signal edges, are 1.01 % and 2.65 %, respectively. For the misaligned signal edges, we obtain an NRMSD value of 1.24 % and a MAE value of 2.34 %. For the traditional model, the NRMSD values are about 9 % while the MAE values are above 20 %. These values only slightly differ from the ones reported for the 65 nm technology. Consequently, the presented model to estimate the TSV energy consumption is valid, independent of the drivers used.

V. Case Study

The high-level model presented in this work reveals a possibility to easily boost the efficiency of several existing coding schemes for TSV interconnects, which we outline by means of this case study. In addition, the case study shows the suitability of the presented linear model for a fast evaluation of TSV low-power coding approaches.

In the field of low-power coding, most coding schemes for correlated and sequential data streams lead to a reduced number of 1-bits, although, their actual goal is to reduce the switching activities. Well known examples are the Inc-XOR (mainly for address data), the K0/K1-Coding (gray/sign-magnitude), the Decorrelator, the Difference-Based-Mapping (DBM) and the Value-Based-Mapping (VBM) [28], [29]. Circuit designers, being aware of the energy model presented in this work, know that these schemes are not optimal for TSVs passing the typical p-doped substrate. To make them suitable, the coding schemes need a slight modification, so that the boolean negated initial code words are obtained. In this case, the switching and the capacitances are reduced simultaneously.

Because a negation of all code words does not affect the switching, the coding schemes maintain their efficiency for the traditional 2D metal-wires. Therefore, for misaligned signals, the K0/NEGK0 coding leads to an additionally reduced coupling switching between the MSBs and the coding efficiency increases.

For an exemplary evaluation of the K0 and NEGK0 coding efficiency, we assume that the data is transmitted over TSV arrays with the same parameters as in Section IV.B. We assume the metal-wires to be in Metal 4, with a width and spacing of 0.3 µm and a segment length of 60 µm. With Synopsys Raphael, the 3π lumped RC model is extracted for a commercial 65 nm technology. The RC model is used in Spectre to obtain reference values for the metal-wire energy consumption.

In Spectre, the transmission rate is set to 1.1 GB/s.

We have performed two different analyses. For the first one, we assume that all input signals are perfectly aligned. For the second one, we assume that the signals on every second interconnect are delayed by 0.5 ns, in a way that a temporal misalignment is present between all adjacent metal-wires and all direct adjacent TSVs. The resulting energy reductions, in percentage terms, for the transmission of a male voice sample, according to the presented linear model, the traditional model and the circuit simulations, are shown in Table V.

The results show that the NEGK0 coding can reduce the dynamic TSV energy consumption by up to 10.93 %, compared to the K0 coding, without affecting the energy consumption of the metal-wires. When the input signals are misaligned, the TSV and the metal-wire energy consumption can be reduced by 33.75 % and 32.93 %, respectively. For aligned signals, these values decrease to 24.20 % and 13.78 %.

Hence, it is obvious that a low-power coding designer has to be aware of the MOS effect and possible misalignments in order to find the optimal low-power coding architecture. The model for the TSV energy estimation , presented in this work, includes these two effects in an abstract way. The energy reduction due to the coding, estimated with this model, only slightly differs (maximum 2.65 pp) from the results obtained with the circuit simulator. This shows the usability of the presented linear model for a fast evaluation of low-power coding approaches. In comparison, the previously used model, shows a prediction error of up to -14.90 pp. Furthermore, with the previous model, the circuit designer can not identify the necessity to increase the number of 1-bits in the data stream, and an unconsidered possible misalignment may lead to a significant underestimate of the coding efficiency.
VI. CONCLUSION

This work provides an accurate high-level model to estimate the energy consumption of TSV arrays, required to analyze the 3D IC energy requirements, and to develop TSV low-power coding approaches. The model includes the temporal misalignment effect and the dynamic behavior of the TSV coupling capacitances, due to the MOS effect. The dynamic behavior of the TSV coupling capacitances is modeled by a simple linear or quadratic function of the bit probabilities; in most cases the linear fit is sufficient.

The presented formulas have been validated against EM and circuit simulations, using the geometrical TSV parameters from the ITRS, for the time slot 2015-2018. For the transmission of 256 different synthetically generated data streams the model always exhibits an error of less than 3.44 % for aligned signals, and 5.53 % for misaligned signals. The traditional used high-level model shows errors of up to 19.03 % for aligned signal edges, and 79.77 % for misaligned signal edges. Compared to the traditional model, the NRMSD, obtained for all data stream scenarios, can be reduced to less than a tenth.

In addition, this work presents a case study for a 3D system, and a real audio data stream, which outlines how the presented formulas can be applied to real-life scenarios. The presented model enables the characterization of existing low-power design techniques for 3D ICs, and to develop new ones, which will be part of our future work.

REFERENCES

a 3D circuit model validated against full-wave FEM simulations and RF measurements,” in 2016 IEEE 66th Electronic Components and Technology Conference (ECTC), May 2016, pp. 966–971.


Lennart Bamberg

Lennart Bamberg received the B.Sc. and M.Sc. degree in Electrical and Information Engineering from the University of Bremen, Germany, in 2014 and 2016, respectively. He is currently working towards the Ph.D. degree in the Institute of Electrodynamics and Microelectronics at the University of Bremen, Germany. His research interests focus mainly on interconnect architectures for heterogeneous 3D integrated circuits, especially through silicon vias.

Alberto Garcia-Ortiz

Alberto Garcia-Ortiz obtained the diploma degree in Telecommunication Systems from the Polytechnic University of Valencia (Spain) in 1998. After working for two years at Newlogic in Austria, he started the Ph.D. at the Institute of Microelectronic Systems, Darmstadt University of Technology, Germany. In 2003, he received from the Department of Electrical Engineering and Information Technology of the university the Ph.D. degree with “summa cum laude.” From 2003 to 2005, he worked as a Senior Hardware Design Engineer at IBM Deutschland Development and Research in Böblingen. After that he joint the start-up AnaFocus (Spain), where he was responsible for the design and integration of AnaFocus’ next generation Vision Systems-on-Chip. He is currently full professor for the chair of integrated digital systems at the university of Bremen.

Dr. Garcia-Ortiz received the “Outstanding dissertation award” in 2004 from the European Design and Automation Association. In 2005, he received from IBM an innovation award for contributions to leakage estimation. Two patents are issued with that work. He serves as editor of JOLPE and is reviewer of several conferences, journals, and European projects.

His research include low-power design and estimation, communication-centric design, SoC integration, and variations-aware design.