Abstract

A semi-analytical technique for computation of the frequency-behavior of silicon substrates is demonstrated. The technique uses a boundary element approach, that utilizes the complex substrate Green Function and the two-dimensional Fast Fourier Transform. The resultant dense system matrix is sparsified by application of orthogonal transform operators on the matrix representing the system. Three transform operators are evaluated for this purpose - the Discrete Cosine Transform (DCT), the Discrete Wavelet Transform (DWT) and the Discrete Hadamard Transform (DHT). The application of any one of these operators provides a rigorous sparsification technique, which significantly reduces the computation time.

The Green Function is computed in the two layers at the top of the substrate. This is done so that contacts in the oxide layer can be included in the substrate model, along with contacts in the silicon substrate. Hence, substrate loss terms in metal interconnect lines and in line-to-line interaction models, can be evaluated using this technique. Extraction of a simple circuit-simulator compatible model from frequency-domain data is discussed.

1: Introduction

Increasing frequencies of operation of silicon circuits have exacerbated the problem of isolation in integrated circuits. An accurate determination of circuit parasitics, including finite substrate impedance, is essential for meaningful circuit simulation. Silicon substrates have been treated as purely resistive in the past [1], [2], [3]. The resistive approximation is valid only when the dielectric relaxation times of the layers composing the substrate cause an insignificant reactance at the highest frequencies of interest. This may not be the case in general, especially above a GHz, where silicon circuits are finding increasing application. Harmonics generated by circuit non-linearities can also exist in the frequency range in which reactive effects are of importance. Modeling of bond-pad losses, substrate losses of metal interconnect lines and substrate losses in line-to-line coupling are also important, especially in RF circuit design. Numerical techniques for frequency-dependent substrate modeling have been presented in the past [4], [5]. However the requirement for fine meshes makes these techniques slow for medium and large size problems.

The substrate is modeled as a layered lossy dielectric (Fig. 1), in the technique presented here, with each dielectric layer characterized by a complex dielectric constant.

\[ \varepsilon = \varepsilon_r - j\varepsilon_i \] (1)

The loss term equals the conductivity of the layer divided by the frequency of operation. The boundary conditions assumed on each surface are shown in Fig. 1.

The theory underlying the method is outlined in the next section. Extensive reference is made here to the technique presented in [1], which was also based on a boundary element formulation. In [1], substrates were considered purely resistive. Contacts were restricted to the top-most layer (Fig. 1) of the substrate, and were of zero-thickness. In the work presented here, the top-most layer is assumed to be oxide dielectric. Hence substrate contacts are in the second layer from the surface. The finite thickness of the contacts is also modeled, as discussed in the next section.

One of the problems with the application of the Boundary Element Technique is that the method yields a dense system of equations. Inversion of the resultant matrices is an O(N^3) process, for NxN matrices. Thus, treating large problems becomes progressively difficult with increasing N. Heuristics have been presented in [1], [2] and [3] for solving this system, where contacts at a sufficient distance from a particular contact are lumped together. However, a more rigorous approach is to solve the system of equation in a transform domain, where some of the off-diagonal matrix-elements are made negligibly small by application of the transform operator. Application of such techniques in image processing for compression are well known [6]. The operators considered in our work include the DCT, the DWT and the DHT. These techniques are discussed in Section 3.

Contacts are treated as equipotential regions. The contact-to-contact complex impedance is extracted at certain discrete frequencies and appropriate passive element models are computed from the complex data, as shown in Sec-
2: Theory: The Substrate Green Function and Coefficients of Potential

The technique uses the Green function of the medium which provides the potential at point $r_1$ due to a point charge at $r_2$. An arbitrary point $r$ is defined by its coordinates in the cartesian system, i.e. $(x,y,z)$. The lateral extent of the coordinates is given by the range $[0,a]$ and $[0,b]$, in the $X$ and $Y$ directions respectively. The vertical extent of $r$ is limited to any location in the top two layers of the substrate.

Following the methodology of [1], in order to compute the contact-to-contact impedance matrix, a coefficient-of-potential matrix is generated at first, using the substrate Green function in a convolution integral. The matrix is then inverted to obtain a coefficient of induction matrix. As in [1], the coefficient of potential matrix is denoted by $[P]$, where each element $p_{ij}$ represents the average potential induced on contact-$i$ due to a unit charge on contact-$j$. In the remainder of this section, the computation and properties of the complex ratio $p_{ij}$ will be discussed.

Three formulations are described for $p_{ij}$: The case when both contacts are in silicon or one contact is in silicon and the other contact is in oxide is described in section 2.1. The case when both contacts are in oxide is similar to the case described in [1] and is discussed in section 2.2.

2.1: Contacts defined in Silicon

Consider two contacts $i$ and $j$ with corner XY coordinates given by $(a_1,b_1);(a_2,b_2)$ and $(a_3,b_3);(a_4,b_4)$ respectively. Also assume that contact-$i$ is at a depth of $z_i$ from the surface and contact-$j$ is at a depth of $z_2$ from the surface. It can be shown, using a method similar to [2], that the coefficient $p_{ij}$ defined above is given by a two-dimensional series

$$ p_{ij} = p^0_{ij} + \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} C^m_{mn} \cdot p^1_{mn}(\epsilon_{z_1}z_2)(\eta_i \cdot \eta_j) $$  \hspace{1cm}(2) \\

where $NI = N-I$ and $C^m_{nn} = 4$ for $(m,n) > 0$, $C^m_{nn} = 2$ for $(m=0$ or $n=0)$ and $C^m_{nn} = 0$ for $(m=n=0)$. The term for $m=n=0$, i.e. $p^0_{ij}$ is given by

$$ p^0_{ij} = (\beta^0_{N1} \cdot z_i + \Gamma^0_{N1})/(ab\beta^0_{N1}\epsilon_{N1}) $$  \hspace{1cm}(3) \\

$\beta^0_{N1}$ and $\Gamma^0_{N1}$ are computed iteratively as shown below for $k=N1$.

$$ \begin{bmatrix} \beta_k \\ \Gamma_k \end{bmatrix} = \begin{bmatrix} \epsilon_{k-1} - \epsilon_k \\ 0 \end{bmatrix} \begin{bmatrix} \beta_{k-1} \\ \Gamma_{k-1} \end{bmatrix} $$  \hspace{1cm}(4) \\

where $\beta_0 = 1.0$, $\Gamma_0 = d$ and $k \in [1, N]$. The other terms in (2) are defined below.

$$ \eta_i = ab \left( \sin(\delta a_2) - \sin(\delta a_1) \right) \left( \sin(\zeta b_2) - \sin(\zeta b_1) \right) $$

$$ \eta_j = ab \left( \sin(\delta a_3) - \sin(\delta a_4) \right) \left( \sin(\zeta b_4) - \sin(\zeta b_3) \right) $$

where $\delta = m\pi/a$, $\zeta = n\pi/b$.

$$ k^N_{mn}(z_1z_2) = \frac{\beta^N_{N1}\cos(\gamma(z_1 + d_N)) + \sin(\gamma(z_1 + d_N))}{(ab\gamma\epsilon_{N1}) \left( 1 + g^N_{mn} \frac{\epsilon_{N1}}{\epsilon_{N1}} \tanh(\gamma d_N) \right) + \left( \cosh(\gamma z_1 + d_N) \right) - \left( \frac{\epsilon_{N1}}{\epsilon_{N1}} \sin(\gamma z_1 + d_N) \tanh(\gamma d_N) \right) \right) $$  \hspace{1cm}(5) \\

where $\gamma = \sqrt{\delta^2 + \zeta^2}$, $z_1 = \max(z_1,z_2)$ and $z_j = \min(z_1,z_2)$, and

$$ g^N_{mn} = \frac{\beta^N_{N1}\tanh(\gamma(d - d_N)) + \Gamma^N_{N1}}{\beta^N_{N1} + \Gamma^N_{N1}\tanh(\gamma(d - d_N))} $$  \hspace{1cm}(6) \\

$\beta^N_{N1}$ and $\Gamma^N_{N1}$ are derived as shown below, for $k=N1$.

$$ \begin{bmatrix} \beta_k \\ \Gamma_k \end{bmatrix} = \begin{bmatrix} \epsilon_{k-1} - \epsilon_k \\ 0 \end{bmatrix} \begin{bmatrix} \beta_{k-1} \\ \Gamma_{k-1} \end{bmatrix} $$  \hspace{1cm}(7) \\

where $\theta_k = \tanh(\gamma_{mn} \times (d - d_k))$, $\theta_0 = 1.0$, $\Gamma_0 = 0$.

Equation (2) is computed by using the Discrete Cosine Transform as discussed in [1]. The DCT has complex coefficients in the case discussed here.

FIGURE 1: Cross-Section of Substrate
The form for \( p_{ij} \) discussed here is used when all contacts are in silicon, and the finite thickness of the contacts is to be taken into account. This form is useful in a majority of the cases, when the substrate parasitic is required for devices. Using the finite-contact thickness is important, especially for contacts that are in close proximity.

When one contact is in silicon and the other in oxide, the continuity in potential at the silicon-oxide interface provides us with a convenient way to compute the coefficients \( k_{mn} \). If we assume that \( z_j \) is in silicon and \( z_u \) is in the oxide, then we have

\[
k_{mn}(z_j, z_u) = k_{mn}(z_j, d_N) \left( \frac{\cosh(\gamma z_j)}{\cosh(\gamma d_N)} \right)
\]

where \( k_{mn}(z_j, d_N) \) is computed from (5). Eq. (8) assumes that the unit charge is placed on the contact in silicon and the potential is observed in oxide. The converse case is identical owing to the symmetry of \( p_{ij} \) under exchange of \( i \) and \( j \). The expression for \( p_{ij}^{00} \) is identical to (3).

### 3: Matrix Sparsification

As remarked earlier, the system of equations resulting from the earlier section \(([P][Q]=[V])\), is dense. As such, as the size of \([P] \) (NxN) increases, computation time goes up as \( N^3 \). Direct inversion of \([P]\) should therefore be avoided for large values of \( N \). In [1], [2] and [3], heuristics for reducing the size of the matrix have been demonstrated, by using a multipole-like approach to the problem. It is difficult to estimate in these methods, a priori, the size of the bounding box beyond which the panels may be approximated.

In solving the above system of equations, minimal computations are required if the matrix \([P]\) is diagonal. Thus, if \([P]\) can be reduced to a near diagonal matrix or a matrix with a sparse structure by some transformation, we could take advantage of this sparsity to reduce computational time and storage. For a general \([P]\), the eigen-decomposition results in a diagonal matrix. However, this requires \( O(N^3) \) computations. If \([P]\) is a circulant matrix, its Fourier transform is diagonal. This observation motivates us to try and use different orthonormal transformations to obtain a sparse structure for a general \([P]\) in the transform domain.

In this work, we solve the system of equations in a transform domain, by first applying the transformation operator to the matrix \([P]\), and \([V]\), and then solving the resultant system of equations given by

\[
[A]^T [P][A][Q] = [A]^T [V]
\]

where \([A]\) is the matrix representation of the transformation operator. The operator \([A]\) is unitary and orthonormal. The operators are chosen such that the energy is concentrated in a few elements in the transform domain, and most elements can be neglected, because their relative magnitude is small. Doing so makes the matrix sparse. The relative error in the solution vector \([Q]\) is bounded by \( \kappa ||\delta([A]^T [PA])|| ||[A]^T [PA]|| \), where \( \kappa \) is the condition number of \([P]\). The quantity \( \delta([A]^T [PA]) \) is made small by appropriate...
choice of [A] in order to minimize the error in the approximation. It should be observed that the matrix \( P \) representing a boundary value problem is generally ill-conditioned, implying a large value of \( \kappa \). Thus it is important to choose the appropriate [A]. Also, different choices of [A] leads to different sparsity structures depending on the structure of [P]. For example, if [P] is circulant, a Fourier transform diagonalizes [P], therefore for this case it is the best choice. In this paper, three operators were evaluated and compared. These are the DCT, the DWT and the DHT.

The Discrete Cosine Transform (DCT)

The DCT is used extensively for compressing speech and video information. We therefore use it in the analogous application suggested here. Any of the several forms of DCT documented in literature can be used [6].

As an example, in this paper, we consider the following matrix representation for the DCT.

\[
a_{ij} = \frac{2}{N} \cos\left(\pi i (j + \frac{1}{2}) / N\right)
\]

where \( k_i = 1/2^{0.5} \) for \( i = 0 \) or \( N \), else \( k_i = 1 \).

The Discrete Wavelet Transform (DWT)

The DWT, unlike the DCT or the Fourier transform uses basis functions which have a finite support. This implies that when a DWT is applied on a step, the discontinuity is localized to a few transform coefficients. In contrast, when a DCT is applied on a step, the effect is spread over all DCT coefficients. This property of the DWT makes it an obvious choice for our application. The DWT matrix \( \Psi \) can be written as the product

\[
[\Psi] = [\Psi_{\log_{2}(N)}][\Psi_{\log_{2}(N)-1}] \ldots [\Psi_2][\Psi_1],
\]

where each matrix \( [\Psi_k] \) can be written as

\[
[\Psi_k] = \begin{bmatrix}
I_{N/2^k} & \log_{2}(N-2^{k-1}) \\
0 & [D_{k}] \\
\end{bmatrix}
\begin{bmatrix}
[0] \\
[0] \\
\end{bmatrix},
\]

where \( [I] \) is an \( l \times l \) identity matrix and \( [0] \) denotes a null matrix. For a dyadic DWT, the \((N/2^{k-1})x(N/2^{k-1})\) matrix \( [D_k] \) is composed of two \((N/2^{k})x(N/2^{k-1})\) sub-matrices which are obtained as follows: the first row of the first \((N/2^{k})x(N/2^{k-1})\) sub-matrix is the wavelet filter coefficient vector \([c]\) padded appropriately by zeros. Each of the remaining rows is then obtained by a circular shift of the immediate previous row by two elements. The second \((N/2^{k})x(N/2^{k-1})\) portion of \( [D_k] \) is obtained similarly, but with its first row as \([d]\) padded appropriately by zeros. The elements of \([d]\) are given as

\[
d_i = (-1)^i c_{M-l-1}, 0 \leq l \leq M - 1,
\]

where \( M \) is the length of the vector \([c]\). While \([c]\) can be chosen to correspond to any of the several different known wavelets, in our work, we use the \([c]\) corresponding to Haar wavelet due to its computational simplicity.

The Discrete Hadamard Transform (DHT)

The DHT operator to the matrix \( [P] \) is restricted to \( 2^N \), where \( N \) is an integer. Thus, the size of the matrix \( [P] \) is restricted to \( 2^N \).

Simulation experiments show that the DWT and the DHT have the highest efficiency in terms of the number of elements that are below a threshold, post-transformation. This is shown in the fill-in maps of a sample matrix \( [P] \), for a five node problem, where the size of the matrix is 512x512 (Fig. 2a-d) and the threshold is 1e-3.

Application of the DHT operator to the matrix \( [P] \) is more computationally efficient because it involves mere additions and subtractions. Generation of the DHT operator can be done as shown in (16) above, or from an FFT butterfly operation, with the butterfly coefficients set to 1, which makes the computation trivial. Also, it was seen that if each of the contacts is divided into powers of two number of panels, the DHT offers significant sparsification (Fig. 2d). Combined with its low computational complexity, the DHT therefore offers significant promise.

TABLE 1. Impedance values (powers-of-2 contact divisions)

<table>
<thead>
<tr>
<th>Z(Ω)</th>
<th>DCT</th>
<th>DWT</th>
<th>DHT</th>
<th>Full Inv</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z_{12}</td>
<td>1152-i 1.9</td>
<td>2837-i 17</td>
<td>2883-i 12.3</td>
<td>2895-i 11</td>
</tr>
<tr>
<td>Z_{13}</td>
<td>494-i 2</td>
<td>568-i 1.8</td>
<td>591-i 1.8</td>
<td>586-i 1.76</td>
</tr>
<tr>
<td>Z_{14}</td>
<td>2424-i 24</td>
<td>2712-i 17</td>
<td>2695-i 10</td>
<td>2822-i 10.9</td>
</tr>
<tr>
<td>Z_{15}</td>
<td>964-i 3.1</td>
<td>942-i 2.4</td>
<td>986-i 3.6</td>
<td>995-i 3.64</td>
</tr>
</tbody>
</table>

TABLE 2. Impedance values (non-powers-of-2 contact divisions)

<table>
<thead>
<tr>
<th>Z(Ω)</th>
<th>DCT</th>
<th>DWT</th>
<th>DHT</th>
<th>Full Inv</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z_{12}</td>
<td>694-i 11.9</td>
<td>3177-i 14</td>
<td>2520-i 79</td>
<td>2847-i 11</td>
</tr>
<tr>
<td>Z_{13}</td>
<td>1171-i 58</td>
<td>575-i 1.9</td>
<td>591-i 0.6</td>
<td>577-i 1.76</td>
</tr>
<tr>
<td>Z_{14}</td>
<td>1369-i 37</td>
<td>2442-i 27</td>
<td>1739-i 49</td>
<td>2761-i 10.5</td>
</tr>
<tr>
<td>Z_{15}</td>
<td>953-i 10.9</td>
<td>954-i 3.7</td>
<td>953-i 18</td>
<td>1003-i 3.67</td>
</tr>
</tbody>
</table>
The sparse matrices generated above using the DWT can be used directly in a Cholesky decomposition [7]. Those generated by the other transforms do not in general have a form suitable for LU decomposition. The Cuthill-McKee algorithm [8] has been used to permute the rows and columns, so as to reduce the bandwidth of the matrix. Finally, a brief comparison is shown in Tables 1 & 2 for the resistance and reactance of a five-node problem, with 512x512 divisions is shown, for all three transforms, with an error threshold of 1e-3 compared to the maximum value of the matrix. The resistance and reactance at 1GHz, connecting the first node to the other four nodes, are shown in the tables. Table 1 shows the impedance values when the contacts are divided in powers of two. Table 2 shows the impedance values when the number of divisions are not powers of 2. The substrate is two layered silicon with surface layer resistivity of 0.1Ω·cm and bulk resistivity of 1Ω·cm, with a dielectric permittivity of 11.

It can be observed from the above tables that for a threshold of 1e-3, the DWT offers a very good match in both cases. The DHT is accurate for the case shown in Table 1. It can be seen from Fig. 2d that the fill-in for this case is the minimum. The DCT does not provide good results for either case, although the results are comparatively better for powers-of-two sub-divisions.

4: Extraction of Simulator-Compatible Models

The boundary element technique shown here is a frequency domain method. The impedances in the substrate

1. The DCT gave good results (to within 5%) with a threshold of 1e-4.
problem are evaluated at discrete frequencies. In order to include the substrate-model in a circuit simulator such as SPICE, it is necessary to model the contact-to-contact impedance by equivalent passive elements.

The substrate impedance at the surface consists of an infinite number of poles and zeroes, because of the distributed nature of the problem. A simple scheme to model the impedance is to extract the impedance at several frequencies, and fit it with the numerator and denominator polynomials, in \( \omega \), of an approximate rational function. The number of frequencies computed will determine the number of poles and zeroes included in the approximate model. Increasing the order of the approximation will, in general, result in a closer match to the correct solution. This procedure fits a reduced order polynomial to an infinite-pole problem and hence is equivalent to the Padé approximation [9]. A passive element (RCLM) model can be generated from the polynomials described above, provided that the representation of the impedance is a Positive Real Function as discussed in [10].

This technique results in good fits, but the substrate mesh size increases as \( O(M^2) \), where \( M \) is the number of substrate contacts, which is usually undesirable. Additionally the DCT of the substrate Green Function has to be computed at \( K \) frequency points, where \( K \) is the sum of the number of poles and zeroes in the impedance function, which also increases the computation time.

We therefore restrict the polynomial orders to a maximum of two poles. Typically a one pole (RC) match is developed, by considering the low-frequency impedance and the impedance near its 3dB roll-off point. This approximation yields good results as can be seen from the example below in Fig. 3. The figure shows the impedance elements \( Z_{12} \), \( Z_{10} \) and \( Z_{20} \) for two square contacts, 100\( \mu \)m to a side spaced at 50\( \mu \)m separation, where 0 signifies the backplane. The exact model is extracted by sweeping the frequency from 0.1GHz to 20GHz in steps of 0.1GHz. The substrate is a low-resistivity type [1], with a 10\( \mu \)m thick p-layer (10\( \Omega \)cm) and a 5\( \mu \)m thick oxide. The bulk layer is 100\( \mu \)m thick (0.1\( \Omega \)cm).

It was observed in our work, that at very high frequencies (> ~20GHz), in epi substrates (e.g. p on p+), the real part of the impedance \( Z_{12} \), became negative (Fig. 3). This behavior of the impedance is not unphysical, within the electrostatic approximation. At sufficiently high frequencies, the impedance in the epi-layer appears primarily capacitive, while the impedance in the bulk region remains resistive. Thus the substrate impedance can be represented by the intuitive model shown in Fig. 4. A Y-\( \Delta \) transform of the model shows that the real part of \( Z_{12} \) is indeed negative. Since the substrate extractor generates a \( \Delta \) model of the substrate, the negative real part is expected. At low frequencies, the epi-layer does not look capacitive, and therefore the real part of \( Z_{12} \) is positive.

The above behavior of \( Z_{12} \), cannot be modeled by a passive RCLM single-port impedance. This is so because the positive real condition mentioned earlier is violated. The simple model shown in Fig. 3 cannot model this behavior. This problem will be addressed later. It should be observed, however, that this issue arises at frequencies in the region of 20GHz for practical substrates. Behavior up to this frequency is well modeled by the simple model shown below. This behavior is not observed in p- on p substrates described in [1].

5: Conclusions

A technique for extraction of substrate models at high frequencies has been presented. The technique can be applied to problems of very large size, with reasonable computational cost, by use of the operator-based matrix sparsification techniques. The DWT and DHT operators were found to be very effective. Frequency domain models
consisting of passive elements were extracted. Low-order approximations are employed to avoid unnecessary increase in computation complexity in the circuit simulation, while preserving the essentials of the frequency response of the substrate. High order models can however be synthesized with an appropriate choice of test frequencies, to generate positive real impedance transfer functions.

6: Acknowledgments

Helpful comments and discussions with Kannan Krishna of TI-DSPRDC are gratefully acknowledged. Thanks are also due to Ali Niknejad of UC Berkeley for useful discussions.

7: References


