Return-Limited Inductances: A Practical Approach to On-Chip Inductance Extraction

Kenneth L. Shepard, Member, IEEE and Zhong Tian

Abstract—Decreasing slew rates and efforts to reduce the resistance–capacitance (RC) delays of on-chip interconnect through design and technology have resulted in the growing importance of inductance in analyzing interconnect response for timing and noise analysis. In this paper, we consider a practical approach for extracting approximate inductances of on-chip interconnect. This approach, which we call the method of return-limited inductances, is based on performing the inductance modeling of signal lines and power-ground lines independently and on taking advantage of the power and ground distribution of the chip to localize inductive coupling. A set of simple geometry-based matrix decomposition rules guide sparsification in these extractions.

Index Terms—Inductance, parasitic extraction, signal integrity.

I. INTRODUCTION

W ith technology scaling, chips consist of more interconnect wires of smaller cross sections packed closer together. As a result, resistance–capacitance (RC) delays have become an important performance limitation, and capacitive coupling is becoming a significant source of on-chip noise. These trends have resulted in significant effort to model and extract interconnect as coupled RC networks [1]. At the same time, increasing chip densities have put a much heavier burden on the power and ground distribution to carry current to switching circuits, and the delta-I noise caused by the transient demands of switching circuits are becoming ever more problematic [2].

Much practical effort is being focused on reducing the resistance and capacitance of interconnect through technology and design. Next-generation technologies can be expected to make increasing use of low-resistivity metals (copper) and low-dielectric-constant insulators. Designers routinely apply wide, thick upper-level metal for power and ground distribution and for long signal runs to reduce resistance. The inductance of the on-chip power and ground lines along with the package inductance and on- and off-chip decoupling capacitance have become important in determining the power supply integrity. At the same time, inductance and inductive coupling have become important in the timing and noise analysis of signal lines in three important ways.

• Inductance must be included to accurately predict rise and fall times and delays in timing analysis. Without considering inductance, one might believe that one can reduce

the propagation delay of a wire indefinitely by lowering its associated RC delay, when there is actually a lower bound on this response given by the speed of light in the interconnect dielectric (66 ps/cm in silicon dioxide).

• If an inductive net is overdriven, an underdamped ringing response can be observed. This can happen easily if consideration is not taken of the inductive properties of a line and the driver is tuned to a capacitive load to achieve an aggressive slew. Ringing noise, which is worst under fast process, low temperature, and high-supply voltage conditions, can potentially result in functional failure in receiving circuits.

• Inductive coupling, along with capacitive coupling, can be a significant source of noise on quiet nets due to the switching of nearby perpetrators [3].

To ascertain if inductance is significant in determining the electrical response of a given on-chip signal line, one must compare the characteristic impedance of the line, its resistive loss, and the effective strength of the driver. We know that for the case of quasi-TEM propagation in a lossy transmission line, inductance is important if [4]

\[
R_{\text{driver}}, R_{\text{line}} < Z_0 = \sqrt{\mathcal{L}/C}
\]  

(1)

where

- \(R_{\text{driver}}\) is the driver’s effective resistance;
- \(R_{\text{line}}\) is the line’s total resistance;
- \(Z_0\) is the line’s characteristic impedance;
- \(\mathcal{L}\) and \(C\) are the inductance and capacitance per unit length.

While (1) is not directly applicable to the very nonuniform transmission line properties that can characterize on-chip interconnects [5], [6], it does explain qualitatively why inductance is becoming important for an increasing number of nets on chip. \(R_{\text{line}}\) is driven lower by the use of thick, upper-level interconnect and copper wires. \(Z_0\) is generally increasing due to lower wire capacitances. \(R_{\text{driver}}\) is decreasing because of more aggressive slew times to match increasing clock frequencies.

Printed-circuit board (PCB) and package designers have long worried about return paths and inductance and included explicit ground and supply planes to control signal-line impedance. Because inductance has not been of concern until recently for on-chip wires in digital applications, designers have largely viewed on-chip signal wires on these chips as single-ended distributed RC lines without concern for current returns. Power and ground distributions have been designed independently to meet power-supply integrity requirements. The growing importance of inductance in high-frequency digital circuits, however, breaks this division, and the design of the power and
The complicated interconnect geometries of digital integrated circuits (IC's), in general, require frequency-dependent resistances and inductances to describe their electrical response. This frequency dependence comes from two effects. The first is the skin and proximity effect in each conductor, which causes the crowding of time-varying currents near the surfaces of conductors with increasing frequency. Frequency-dependence also results from the dependence of the current distribution which characterizes the return path on frequency. At low frequencies, current returns favor low-resistance paths, sometimes very far from the driven wire or through the package. At high frequencies, current returns favor the lower-inductance paths, usually the closest power or ground line, but sometimes a neighboring signal line or collection of signal lines. While power and ground lines are always-available paths for high-frequency current returns in digital CMOS integrated circuits (assuming good power-ground design and adequate decoupling), when signal lines constitute part of the return path, they usually do so in a way that depends, in part, on the switched state of driving and receiving circuits. Displacement currents capacitively coupled into the (resistive) substrate find their way onto power and ground lines based on the proximity of substrate and nwell plugs.

The conventional approach to power and supply design is to create an on-chip grid as shown in Fig. 1(a), [8]. The inductance of critical nets, such as clocks, can be controlled by routing them close to the power and ground lines of the grid [7], [9]–[11]. To provide even more control and predictability in signal-line and power-ground inductance, the Digital Alpha 21264 design devoted two metal layers exclusively for the distribution of power and ground [12] as shown in Fig. 1(b). Openings are made in these ground and supply planes to via though signal line connections in a manner similar to PCB design. This approach is generally considered to be very expensive because wiring resources are always scarce in very large scale integration (VLSI) chips. Therefore, the grid design is preferable.

To allow equivalent circuits to be developed in complex integrated circuit environments in which return paths are not known, the concept of partial inductances is traditionally used. The partial inductance technique, which assigns portions of the loop inductance to segments along the loop, appeared in work by Rosa in the early 1900’s [13]. The book by Grover [14] provides a comprehensive tabulation of formulas for partial inductances and partial mutual inductances for different geometries. Applying partial inductances to the modeling of complex multiconductor geometries was formalized by Ruehli with the development of partial-element equivalent circuits (PEEC) [15]–[17]. Because partial inductances obey the same branch constitutive relations as closed-loop inductances, they can be conveniently applied in the context of modified nodal analysis (MNA) and used internally in circuit simulators such as SPICE. The linear state-space representations that result from MNA are also the starting point for reduced-order modeling techniques for linear interconnect [18]–[20]. The partial-inductance technique can even be used to model the skin effect by dividing a conductor up into multiple longitudinal filaments [21]–[23].

In the partial inductance approach, the signal lines and supply and ground lines are treated equivalently, resulting in a large, densely coupled network representation. One approach to make the inductance matrix sparse is simply to discard those terms of the inductance matrix which are below a certain threshold. This approach, however, does not guarantee the positive semidefiniteness of the resulting inductance matrix. In particular, by truncation, it is possible to have mutual inductive coupling between two segments through intermediate segments that are not directly coupled to each other. Reference [24] shows an example of an instability created by a truncated nonpositive-definite inductance matrix. As an alternative to simple truncation, another proposal [25], [26] to render the inductance matrix sparse is to associate with every segment a distributed current return out to a shell $r_0$ away. Segments spaced more than $r_0$ apart have no inductive coupling. The shell technique has also formed the basis for approximate loop-inductance bounds [27]. The value of $r_0$ that must be used to achieve a desired accuracy depends on the interconnect topology, the connectivity of the nets in question, and the particular net or coupling being considered. This leads to complicated schemes in which the value of $r_0$ is dynamically chosen based on accuracy in a particular transfer function [24] calculation. Moreover, this approach does not work well for long wires broken into many segments. For a value of $r_0$ sufficiently small to render coupling in the transverse direction sparse, correctly modeling the loop inductance of these wires may require including mutual inductances between collinear segments more than $r_0$ away.

III. RETURN-LIMITED INDUCTANCE EXTRACTION

In this paper, we present an approximate inductance extraction technique which guarantees a positive-semidefinite inductance matrix for the on-chip signal lines, while rendering the inductance network sparse in two important ways.

- The inductance of the signal lines are extracted independently from the power and ground wires; therefore, there
are no explicit mutual inductances between power and ground lines and signal lines. This technique is consistent with the traditional approach in digital IC design of performing independent signal and power-grid analysis and takes advantage of the known availability of power and ground as high-frequency current returns.

- Mutual inductances between signal lines are further restricted by a set of simple geometry-based decomposition rules, which we refer to as halo rules.

The technique, which we call return-limited inductance extraction, can be applied to complex geometries with the same generality as a partial inductance formulation. The cost of this simplification is some loss of accuracy in the inductance calculation which can be easily tolerated within the performance and noise margins typical of digital IC’s. Throughout this paper, the accuracy of this technique for different geometries is clarified with many comparisons to the results of detailed extraction using Fasthenry [22], [23].

In subsequent discussions, we assume that the geometries on the chip are Manhattan; that is, all edges of all shapes are either horizontal or vertical and these two directions are orthogonal. While this is not a necessary condition, it simplifies the implementation and is satisfied by nearly all VLSI interconnect layouts. Geometries which are not Manhattan can be approximated by “staircase” constructions. We introduce an orthogonal coordinate system in which horizontal refers to the \( x \) direction and vertical refers to the \( y \) direction. The \( z \) direction is perpendicular to the surface of the chip.

We begin with a few definitions. A wire segment is a rectangular parallelepiped defined by coordinates \((x_{\text{min}}, y_{\text{min}}, z_{\text{min}})\) and \((x_{\text{max}}, y_{\text{max}}, z_{\text{max}})\). A horizontal segment is one in which the current flow is known to be horizontal (i.e., in the \( x \) direction), while a vertical segment is one in which the current is known to be vertical (i.e., in the \( y \) direction). Moreover, a mixed segment is one in which the current direction is not consistently horizontal or vertical. The halo of a segment consists of the six semiinfinite subregions shown in Fig. 2. The horizontal halo consists only of regions \( R_3, R_4, R_5, \) and \( R_6 \), while the vertical halo consists only of regions \( R_1, R_2, R_3, \) and \( R_4 \).

The halo rules as applied to signal-line extraction are given as follows.

- Horizontal and vertical signal line segments are treated independently since they do not inductively couple to each other. Segments with horizontal currents can only couple inductively with other segments with horizontal currents. Similarly, vertical segments can only couple inductively with other vertical segments.\(^1\)
- Horizontal halos of power and ground are “blocked” by horizontal signal segments while vertical halos of power and ground are “blocked” by vertical signal segments. If the halos are viewed as columnated beams emanating orthogonally from each face of a segment, then blocking occurs whenever these beams are interrupted by another segment. Fig. 3(a) shows the \( R_2 \) portion of the vertical halo.

\(^1\)This means that the extraction can really proceed in two separate “passes”: one to extract the horizontal segments and one to extract the vertical segments.
of a ground segment blocked by the segment of a neighboring signal wire A.

- Inductive coupling between two horizontal segments is nonzero if and only if it is possible to connect two segments by a path which does not cross the horizontal halo of any ground or supply line. Similarly, inductive coupling between two vertical segments is nonzero if and only if it is possible to connect two segments by a path which does not cross the vertical halo of any supply or ground line.\(^2\)

In Fig. 3(b), signal segments A and B inductively couple, for example, because it is possible to connect them by a path (shown as a dotted line) which does not cross a halo. A and C, however, do not couple because all paths [such as the one shown as the dotted line in Fig. 3(b)] between them must cross either halo region \(R_1\) or \(R_5\) of \(gnd2\).

These halo rules divide the chip interconnect into a collection of disjoint horizontal interaction regions defined by the nonblocked horizontal halos of the power and ground distribution. Horizontal segments must be contained within the same horizontal interaction region to inductively couple. Independently, the chip is also divided into a collection of disjoint vertical interaction regions defined by the nonblocked vertical halos of the power and ground distribution. Similarly, vertical segments must be contained within the same vertical interaction region to inductively couple. For example, in Fig. 3(b), the halo rules result in the definition of two vertical interaction regions, labeled \(I\) and \(II\). Region \(I\) contains signal segments A and B, while region \(II\) contains signal segments C and D. For a given segment, using the halos of same direction power or ground to define an interaction region is equivalent to assuming that current returns are zero beyond the nearest same-direction power or ground lines. Since power and ground lines are always available as current returns, this assures that every signal line has an associated fail-safe current return. Allowing other same direction segments to block the halos recognizes the fact that coupling to these signal lines allows segments on either side of the overlapped ground or supply line to interact. Defining interaction region boundaries with the halos of orthogonally directed power or ground depends primarily on the ability with this localization to preserve enough inductive coupling between collinear segments to accurately predict the inductance of long wires. We will discuss the limitations of this approach later in this section.

In Fig. 4, we consider the halo rules as applied to a more complex three-dimensional (3-D) interconnect topologies in a five-layer-metal process (layers labeled from the top—M5, M4, M3, M2, and M1). In this example, we are looking down on a wire topology in layers M3, M4, and M5. The M4 and M5 shapes are all associated with power or ground. Two signal lines are routed on M3. The interaction region shown in cross-hatch acts as a vertical interaction region for signal segments 1 through 4. Note that segments 1 and 4 couple (and, therefore, belong to the same vertical interaction region), for example, because the vertical halo of the intervening ground line is blocked.

\(^2\)Since when doing vertical signal line extraction, for example, we only need to consider vertical halos of ground or supply lines and these halos can only be blocked by vertical signal segments, horizontal (or mixed) signal segments do not have to be considered at all.

---

Fig. 4. Application of halo rules in three dimensions. All of the M4 and M5 shapes correspond to power and ground lines, while two signal lines are routed on M3. Cross-hatching denotes the vertical interaction region by which segments 1-4 interact. All of the fractures necessary for the signal line extraction are shown. All of the power-ground line segments are labeled as \(G_1-G_{15}\). Segment 1 forms return-limited loop inductances with \(G_2\), \(G_6\), and \(G_{10}\). Segments 3 and 4 form return-limited loop inductances with \(G_8\), \(G_9\), and \(G_{12}\). Segment 6 forms return-limited loop inductances with \(G_{14}\) and \(G_{15}\), while segment 5 forms return-limited loop inductances with \(G_{12}\) and \(G_{13}\).

---

Fig. 5. Static fracturing and direction setting. Shaded interconnect is on metal-2 while the remaining interconnect is on metal-1. Vias connect rectangle A on metal-1 with rectangle K on metal-2. Segments with an unambiguously defined horizontal or vertical current flow are noted with arrows. Segments B and C are external contacts.

After the interaction regions are defined, extraction begins by fracturing each wire into a set of segments. Fracturing is itself a two-stage process—static fracturing to establish rectangular segments and current direction, followed by dynamic fracturing to define the geometries for inductance extraction. The static fracturing approach employed is very similar to that of the INDEX [28] extractor. If a given rectangle has two edges that are contacted by a neighbor, we determine if the current flow through the segment is horizontal, vertical, or mixed. An example is shown in Fig. 5. In particular, if the contacted edges are the top and bottom, then the current flow is vertical (as for segments F and G) while if the contacted edges are left and right, then the current flow is horizontal (as for segments H, I, J, and L). In other cases, the current flow direction is mixed, as for
Fig. 6. Signal-line resistance and inductance extractions for Fig. 4. (a) Full return-limited loop inductance representation. Each inductor is coupled in general to all of the others in the interaction region. (b) Equivalent return-limited inductance representation.

rectangles that have only one neighbor, corner segments (as for segments D and E), segments connected to vias (as for segments A and K), and segments associated with external contacts (as for segments B and C). Naturally, static fracturing seeks to minimize the wire area represented by mixed segments, since inductances are only extracted for horizontal and vertical segments.

Dynamic fracturing is used to create additional fractures in the signal, ground, and supply wires required for the inductance calculation. Dynamic fractures must be generated whenever the interaction environment as defined by the power or ground lines of the interaction region changes; that is, static fractures in the ground and supply lines must be projected onto the signal lines. Additional signal line fractures may sometimes be necessary for long uniform wire runs to ensure a reasonable approximation of the distributed nature of the resistance and inductance. (The fracture between segments 2 and 3 in Fig. 4 might be such a case.) Also, signal segments cannot span multiple interaction regions; therefore, new fractures must be created to split segments between interaction regions. Once all of these signal line fractures have been defined, these fractures, both static and dynamic, must be, in turn, projected onto the parallel adjacent ground and supply lines. All of the fractures necessary for signal line extraction are shown in the example of Fig. 4.

The model for each signal segment which results from this dynamic fracturing is a resistance (dc value for the signal segment, discussed in detail later in this section) in series with a parallel combination of (open) loop inductances, where each loop inductance is defined by a signal line returning through one of the return segments defined on the parallel, adjacent supply or ground lines in the interaction region. We call these inductance values return-limited loop inductances. In the example of Fig. 4, the segment 1 has three loop inductances associated with it as defined by the current returns through \( G_{14} \), \( G_{15} \), and \( G_{16} \). Segment 2 has current returns through \( G_{23} \), \( G_{24} \), and \( G_{26} \), while segments 3 and 4 have current returns through \( G_{34} \), \( G_{37} \), and \( G_{11} \). Segment 5 returns through \( G_{12} \) and \( G_{13} \), while segment 6 returns through \( G_{14} \) and \( G_{15} \). The corresponding circuit representations for \( \text{net1} \) and \( \text{net2} \) are shown in Fig. 6(a). The current \( I_1 \), for example, is the current flowing through segment 1 and returning through segment \( G_{1} \). \( I_2 \) is the current flowing through segment 1 and returning through segment \( G_{5} \), while \( I_3 \) is the current flowing through segment 1 and returning through segment \( G_{8} \).

We denote the partial inductance matrix by \( \mathcal{L} \). In terms of these partial inductances, the return-limited loop inductances for the vertical segments of Fig. 6(a) are given by

\[
\mathcal{L}_{\text{vertical}} = \begin{pmatrix}
L'_{11} & L'_{12} & L'_{13} & L'_{14} \\
L'_{21} & L'_{22} & L'_{23} & L'_{24} \\
L'_{31} & L'_{32} & L'_{33} & L'_{34} \\
L'_{41} & L'_{42} & L'_{43} & L'_{44}
\end{pmatrix}
\]

\( L'_{ij} \) are the matrices of return-limited loop inductances associated with each signal line segment. For example, as shown in (2) at the bottom of the next page, where the tilde indices refer to the power-ground segment (e.g., \( \mathcal{L}_{11} \) is the partial self-inductance of segment \( G_1 \)).

Both the partial inductance and return-limited loop inductance matrices are symmetric and positive definite. Since the interaction regions do not couple in any way, when ordered appropriately, the return-limited loop inductance matrix of the entire chip is block diagonal, with each block representing the horizontal or vertical signal segments of a single interaction region. The return-limited loop inductance of one segment is densely coupled to the return-limited loop inductance of the other segments in the same interaction region. We note that currents circulating in the inductive loops of Fig. 6(a) correspond to current configurations in which the current is flowing through the
supply and ground lines and not in the signal lines. One can collapse the circuit of Fig. 6(a) into the circuit of Fig. 6(b) by ignoring these circulating currents. The inductances in Fig. 6(b) are the return-limited inductances. Let \( n \) be the number of return-limited loop inductances in the interaction region, and let \( m \) be the number of return-limited inductances in the interaction region, as determined by the number of signal segments. \( L' \) is an \( n \times n \) matrix. We then form the \( n \times m \) matrix \( B \), where the \( i \)th column of \( B \) is all zero except for ones in the rows corresponding to the return-limited loop inductances associated with the given return-limited inductance. One finds

\[
sI = B^T L'^{-1} BV
\]

where \( I \) and \( V \) are the currents and voltages as shown in Fig. 6(a). The return-limited inductance matrix of the interaction region is then given by

\[
L = (B^T L'^{-1} B)^{-1}.
\]

The main computational cost here is two LU factorizations, one of the \( L' \) matrix associated with each horizontal and vertical interaction region and the second to find the final return-limited inductance matrix for the given interaction region (the \( L \) matrix). If the number of segments in a typical interaction region is on the order of several thousand, direct LU decomposition is most efficient. Iterative techniques could, of course, be explored for larger interaction regions. The \( m \times m \) return-limited inductance matrix, like the other inductance matrices, is symmetric and positive-definite.

Simplified partial inductance formulas from Grover [14] are used to calculate the return-limited loop inductances. We begin by defining the geometric mean distance (GMD) between two segments \( i \) and \( j \) of cross-sectional areas \( a_i \) and \( a_j \). Let \( z \) and \( z' \) be the transverse coordinates of a point in \( a_i \) and \( x \) and \( x' \) be the transverse coordinates of a point in \( a_j \) (in this case, the segments are directed in the \( y \) direction). Then the GMD \( r_{ij} \) between \( i \) and \( j \) is given by

\[
\log_e r_{ij} = \frac{1}{2 a_i a_j} \int dx \int dz \int dx' \int dz' \log_e \left(\sqrt{(x-x')^2 + (z-z')^2}\right).
\]

It also makes sense to define the geometric mean distance \( r_{ii} \) between a segment \( i \) and itself. With these definitions, we have the approximate formulas for the partial self inductance of a segment \( i \) of length \( l \)

\[
\mathcal{L}_{ii} \approx 2 \times 10^{-13} \left[ \log_e \left( \frac{2l}{r_{ii}} \right) - 1 \right].
\]

Furthermore, if their ends are touching, then

\[
\mathcal{L}_{jj} \approx 1 \times 10^{-13} \left[ (l + m + \delta) \log_e (l + m + \delta) - (l + \delta) \log_e (l + \delta) - m \log_e m + \delta \log_e \delta \right].
\]

To find the GMD for the various cross sectional geometries encountered in extraction, we employ a summation technique [14]. For each metal layer we define a core filament of width \( \Delta \) such that we are satisfied expressing every wire width as a multiple of \( \Delta \) as shown in Fig. 7. For example, the geometric mean distance of the whole segment 1 from \( 1' \), \( r_{11'} \), is given by:

\[
\log_e r_{11'} = (\log_e r_{1'1'} + \log_e r_{1'y'} + \log_e r_{y'1'})/3.
\]

Similarly, \( \log_e r_{12'} = (\log_e r_{1'2'} + \log_e r_{1'y''} + \log_e r_{y'2'})/3 \) and \( \log_e r_{13'} = \log_e r_{1'y'} \). The geometric mean distance of 1 to itself is then:

\[
\log_e r_{11} = (\log_e r_{1'1} + \log_e r_{1'y} + \log_e r_{y'1})/3.
\]

Similarly, the geometric mean distance between segment 1 and 2 is given by

\[
\log_e r_{12} = (\log_e r_{1'1'} + \log_e r_{1'y'} + \log_e r_{y'1'} + \log_e r_{y'2'} + \log_e r_{2'y} + \log_e r_{2'y'})/6.
\]

Fig. 7. (a) A core filament of thickness \( t \) and width \( \Delta \). (b) Two wires, one of width \( 3\Delta \) and the other of width \( 2\Delta \).
In this way, we only need to develop lookup tables for the geometric mean distance of a core filament to itself and the geometric mean distance between two core filaments on any two metal layers as a function of distance between them.

Another important assumption of the halo rules is the discarding of mutual inductances between segments which do not fall in the same interaction region. We consider these assumptions in more detail in the example of Fig. 11 in which we consider two signal lines coplanar with two ground returns. We compare the inductances as calculated by Fasthenry (solid lines) to the return-limited inductance values (dotted lines) for three cases. In Fig. 11(b), neither of the shaded ground lines shown in Fig. 11(a) are present. We show the self and mutual inductances of the signal lines as a function of the spacing between them. As we have seen and explained in the previous examples, the return-limited technique slightly overestimates the inductances. In Fig. 11(c), we include the shaded ground line 4 μm below the plane of the signal segments shown in Fig. 11(a). In this case, the halo of this ground line puts \( \text{sig1} \) and \( \text{sig2} \) into different interaction regions and means that the return-limited mutual inductance between them is zero. This approximation grows worse in the case of Fig. 11(d), when the shaded ground line is instead 10 μm below the plane of the signal segments. Modification of the halo rules might be necessary in these extreme cases to create “super” interaction regions to preserve some of these couplings. We show an example of this in Section IV. Because we ignore the skin effect in calculating the partial inductances, the resistances and return-limited inductances extracted are frequency independent. We define the cross-over frequency \( f_c \) of a segment as the frequency at which \( R = 2\pi f_c L \), where \( R \) is the resistance of the segment and \( L \) is the return-limited self inductance. For typical cross-over frequencies, the skin depth almost always exceeds the wire thickness and width; that is, the frequency dependence of the resistance occurs at frequencies above \( f_c \), where the inductive response dominates, and, therefore, can be safely ignored in nearly all cases (for \( f = 500 \) MHz, for example, the skin depth is 3.7 μm for aluminum and 2.9 μm for copper at 300K). Additionally, we define \( f_{\text{max}} \) as the maximum frequency content of on-chip signals (one can expect the highest frequency content as determined by the fastest on-chip slew times to be 50–100 times the clock frequency—10 GHz would be reasonable for a 500-MHz clock frequency). For segments for which \( f_c > f_{\text{max}} \), the inductance of the segment itself can be ignored entirely as can its inductive coupling to other segments; that is, inductors need only be extracted where necessary. In rare cases in which the frequency dependence of the resistance might be significant for frequencies comparable to or below \( f_c \), one could consider refining the return-limited inductance approach to employ a ladder equivalent circuit [29].

As a result of the halo rules, we are restricting the current returns characterizing the frequency-dependent inductance of the given segment to same-direction lines within the interaction region. To clarify the approximation associated with this assumption, we consider a simple two-dimensional example. As shown in Fig. 9(a), a single signal line is coplanar with two ground wires where the far ends of the lines are assumed to be shorted. In Fig. 9(b) and (c), we compare the resistance and inductance of the signal wire as calculated by Fasthenry (solid curve) to the return-limited extraction result (dashed curve). Because the return-limited inductance extraction is ignoring the parallel inductance of the more distant wide-wire return and the associated skin effect, we systematically overestimate (though not significantly) the inductance as a result of this halo-rule approximation. Because the frequency-dependence of the resistance does not occur until frequencies well above \( f_c \) [noted in Fig. 9(d)], we obtain good agreement in the magnitude and phase of the impedance as shown in Fig. 9(d). We widen the signal line in Fig. 10 for the same coplanar structure to determine the error associated with a wide signal lines returning
through a narrow ground when a wider ground further away is available. In Fig. 10(b) and (c), we once again compare the resistance and inductance of the signal wire as calculated by Fasthenry (solid curve) to the return-limited extraction result (dashed curve). As in the previous example, the return-limited inductance extraction slightly overestimate the self inductance. As expected, the cross-over frequency is lower, but the strong frequency-dependence of the resistance still does not occur until frequencies above $f_c$. As a result, we still obtain good agreement in both the magnitude and phase of the impedance shown in Fig. 10(d). It is important to note that the proximity assumptions of the halo rules preserve the most problematic inductive couplings in digital designs, for example, the case of wide, simultaneous-switching parallel busses without adequate interdigitated power and ground returns [11].

In the return-limited inductance extraction approach for signal lines, we have explicitly ignored package and substrate returns. While it is common to regard the substrate as a ground equipotential in extracting capacitances, the substrate is electrically far away due to its high resistivity and, therefore, large skin depth [30]. We, therefore, assume that all ac current returns will find a path onto metal interconnect, which we assume will always be electrically closer. Furthermore, the power grid is always accessible from the substrate through plugs. Whether current returns through the package can be ignored in on-chip inductance extraction, particular in the context of last metal interconnect, will depend on the details of the technology. We have assumed for the purposes of this paper that these returns (such as those through the mesh plane of a multi-chip module package) are still electrically “far” away.

**IV. Results**

To further study the efficiency and accuracy of return-limited inductance extraction, we have implemented a prototype resistance and inductance extractor which can handle complex 3-D
geometries. The extractor takes a shapes text file input with established connectivity, similar in format to the Fasthenry input file [22]. For the examples presented in this section, we work with a hypothetical five-level-metal copper process shown in the layer map of Fig. 12(a). M5 is 2 μm thick, while the other metal layers are 0.9 μm thick. (b) Power and ground are distributed in a grid spaced 100 μm apart vertically and 21.6 μm apart horizontally. $V_{DD}$ is carried on the M3 lines running horizontally at $y = 900, 700, 500, 300$, and 100 μm. Ground is routed on the remaining M3 horizontal lines. Similarly, $V_{DD}$ is carried on the M2 and M4 lines running vertically at $x = 21.6, 64.8, 108$ μm. Ground is routed on the remaining alternate M2 and M4 vertical lines. (c) Three different “ideal” power-ground decoupling assumptions are explored. In the no decoupling case, we only short the power and ground distribution at the two black rectangular pins. In the minimal decoupling case, we short at the “X” points, while in the complete decoupling case, we short at all the “O” points. For the examples presented in this section, we work with a hypothetical five-level-metal copper process shown in the layer map of Fig. 12(a). M5 is 2 μm thick, while the other metal layers are 0.9 μm thick. (b) Power and ground are distributed in a grid spaced 100 μm apart vertically and 21.6 μm apart horizontally. $V_{DD}$ is carried on the M3 lines running horizontally at $y = 900, 700, 500, 300$, and 100 μm. Ground is routed on the remaining M3 horizontal lines. Similarly, $V_{DD}$ is carried on the M2 and M4 lines running vertically at $x = 21.6, 64.8, 108$ μm. Ground is routed on the remaining alternate M2 and M4 vertical lines. We will explore extraction results with three different power-ground decoupling assumptions, shown in Fig. 12(c). In the no decoupling case, we only short the power and ground distribution at the two black rectangular pins. In the minimal decoupling case, we short at the “X” points noted in Fig. 12(c), while in the complete decoupling case, we short at all the “O” points. We choose shorts rather than capacitors to facilitate straightforward comparison to Fasthenry. Fig. 13(a) shows the first extraction example in which a single wire, beginning at A and ending at B, is routed within the power and ground distribution network of Fig. 12. Node B is shorted to the power grid at the far end. We compare the results of Fasthenry to the return-limited inductance extraction (dashed curves) for the impedance seen between A and the adjacent near-end power grid. We also assume two different decoupling assumptions in the Fasthenry calculation. The solid curves use the complete-decoupling assumption of Fig. 12(c), while the dotted curves use the no-decoupling assumption of Fig. 12(c). (b) Resistance at the near end as a function of frequency. (c) Inductance at the near end as a function of frequency. (d) Magnitude and phase of the impedance at the near end as a function of frequency.
important until frequencies at which the impedance is dominated by the inductance. As a result, we get good agreement in the near-end impedance (magnitude and phase) between the return-limited values and the Fasthenry result assuming complete decoupling, as shown in Fig. 13(c).

Fig. 14(a) shows another single signal line routed through the same power and ground distribution, beginning at A and ending at B. Node B is also shorted to the power grid at the far end. Once again, we compare the results of Fasthenry (solid curve) to the return-limited inductance extraction (dashed curve). In this case, we only show the complete-decoupling-assumption results. In this geometry, because of the more distant parallel returns for the horizontal segments, the discarded forward mutual inductances are more significant. To demonstrate this, we also ran this example in which we doubled the size of the horizontal interaction regions to include more forward mutual inductances. The results are shown in the dotted curve in Fig. 14(b). Good approximation is obtained with the return-limited extraction in the near-end driving point impedance as shown in Fig. 13(c).

Once again, we favorably compare the results of Fasthenry to the return-limited inductance extraction.

In Table I, we compare the number of nonzero elements in the inductance matrix for the return-limited inductance results against the full partial inductance extraction and various applications of the shell sparsification algorithm. The number of nonzero elements directly translates into the efficiency of simulation and analysis of the resulting network. The full partial inductance extraction is shown as in the result of shell sparsification on this matrix. This compares favorably with the number of elements required in the return-limited inductance extraction. "Very big" indicates that the number of nonzero elements was too large to even practically count (>50 million).

Once again, we favorably compare the results of Fasthenry to the return-limited inductance extraction.

In Table I, we compare the number of nonzero elements in the inductance matrix for the return-limited inductance results against the full partial inductance extraction and various applications of the shell sparsification algorithm. The number of nonzero elements directly translates into the efficiency of simulation and analysis of the resulting network. The full partial inductance extraction is shown under two conditions, one in which single filament is used to represent each segment and a second in which each segment cross section is broken into 20 filaments. The Fasthenry results of Figs. 13–15 were calculated.
using the 20-filament decomposition to properly model skin-effect to 10-GHz. In Fig. 14, we shown the inductance as a function of frequency (dashed–dotted curves) for each of the shell approximations. The result at rₑ = 10 µm is far worse than the return-limited inductance result even through it requires 859 nonzero elements as compared to only 178 for the return-limited case. Good accuracy is not achieved until rₑ = 50 µm, requiring over 3000 elements. Furthermore, we have no way of knowing a priori what value of rₑ is required to achieve tolerable accuracy. The full partial inductance matrix, even with only single filament decomposition, has more than 80,000 nonzero elements.

V. CONCLUSION AND DIRECTIONS FOR FUTURE WORK

In this paper, we have presented an approximate inductance extraction approach that can be practically applied for full-chip extraction of complex integrated circuits. The approach results in a positive-semidefinite inductance matrix and achieves sparsity by breaking inductive couplings between the power-ground lines and signal lines and by using a set of geometry-based decomposition rules to discard “insignificant” coupling interactions between signal lines and between power and ground lines. By taking advantage of the fail-safe availability of power-ground lines as high-frequency current returns while preserving inductive coupling between signal lines, we have demonstrated remarkable efficiency and accuracy over other sparsification techniques.

We plan to combine this technique with capacitance extraction to produce a true full-chip RLC extraction environment. As part of this work, further refinement of the shadow rules may be necessary to achieve maximum accuracy for minimum loss of efficiency. A much closer interaction with layout will also be required in analysis and extraction in order to understand current distributions in the substrate and how this might affect interconnect response in some cases.

REFERENCES


Kenneth L. Shepard (S’85–M’81) received the B.S.E. degree from Princeton University, Princeton, NJ, in 1987 and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1988 and 1992, respectively.

From 1992 to 1997, he was a Research Staff Member and Manager in the VLSI Design Department at the IBM T. J. Watson Research Center, Yorktown Heights, NY, where he was responsible for the design methodology for IBM’s G4 S/390 microprocessors. Since 1997, he has been at Columbia University, New York, where he is now an Associate Professor. He also serves as Chief Technology Officer of CadMOS Design Technology, San Jose, CA.

Dr. Shepard received the Fannie and John Hertz Foundation Doctoral Thesis Prize in 1992. At IBM, he received Research Division Awards in 1995 and 1998. He was also the recipient of an NSF CAREER Award in 1998 and IBM Early Faculty Development Awards in 1998 and 1999. He was also awarded the 1999 Distinguished Faculty Teaching Award from the Columbia Engineering School Alumni Association. He is Associate Editor of IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS and is a program track chair for the International Conference on Computer Design.

Zhong Tian received the M.S. degree from Southeast University, Nanjing, Ching, in 1993, and the B. S. degree from University of Electronic Science and Technology, Chengdu, China, in 1990, both in electrical engineering. He is currently a Professional degree candidate at Columbia University, New York, specializing in VLSI design and CAD.