

APPLIED NUMERICAL ELECTROMAGNETIC ANALYSIS FOR PLANAR HIGH-FREQUENCY CIRCUITS

JAMES C. RAUTIO
Sonnet Software, Inc. Phoe
Phoenix, New York

1. INTRODUCTION

Both commercial and military wireless applications push limits for size and performance. In the 1970s troops had to make educated guesses about the battlefield situation. Today, frontline troops can have a real-time display showing where every one is, even including the enemy. Cellphones of the 1980s were the size of a book. Today, they have greater coverage and are so small that they get lost in a pocket. Technology drives these advances. A critical part of this technology is electromagnetic (EM) analysis.

Wireless technology achieves these dramatic advances by shifting functionality from printed circuit boards (PCB) and discrete components (e.g., inductors, capacitors, resistors, transistors) to monolithic integrated circuits. These radiofrequency integrated circuits (RFICs) can be on silicon (Si) or gallium arsenide (GaAs). Another choice, midway between RFIC and PCB, is low-temperature cofired ceramic (LTCC). For example, if you were to open your cellphone, you would see a tiny piece of ceramic connected to the antenna. That little piece of ceramic is an LTCC circuit that replaces a PCB and numerous discrete elements formerly occupying a volume almost as large as your entire modern cellphone.

Other portions of your cellphone might include Si RFICs. Formerly restricted to digital/computer applications, silicon integrated circuits can now be used to shrink entire PCB assemblies down to the size of a few grains of sand. For example, silicon RFICs are widely used for radiofrequency identification (RFID) tags. Costing pennies and the size of a credit card, RFID tags use an embedded Si RFIC to transmit information whenever they receive an appropriate radio signal. You might use this when you buy gasoline or pay a toll. Demanding military applications, where high frequency and high power are important, can make use of the superior, but more expensive, GaAs.

Figure 1 shows a photomicrograph of a GaAs MMIC (monolithic microwave integrated circuit) wafer. During the design phase it is common to place multiple designs on a single GaAs wafer. The circuit in the center is a two-stage 1.2–3.2-GHz 1-W GaAs amplifier with 20 dB small-signal gain. It was fabricated on a 75- μm -thick wafer and is 2.98 mm on each side. Electromagnetic analysis is absolutely critical in achieving success on first fabrication for circuits like this.

In all of these applications, it is important to squeeze every gram of performance out of the selected technology, and to do that with the smallest possible size and the shortest possible design and fabrication time.

The high-frequency design process widely used prior to the 1990s began with an approximate design, then the circuit was built. The technology to complete a design with

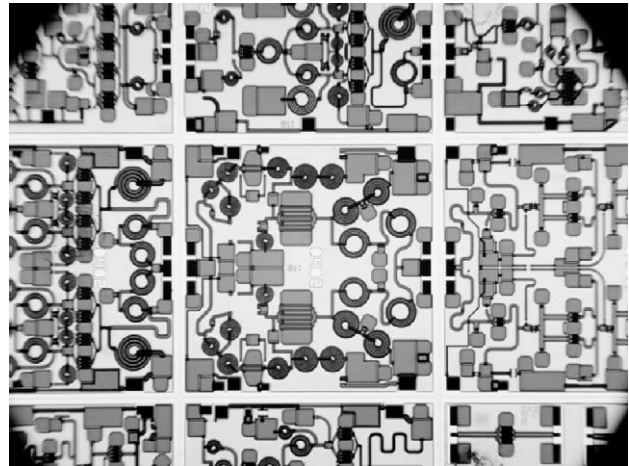


Figure 1. During the design phase, multiple circuits are usually placed on a single GaAs wafer. Here, in the center of this photomicrograph, is a two-stage 1.2–3.2-GHz 1-W GaAs MMIC amplifier with 20 dB small-signal gain. (Image courtesy of M/A-COM, Inc., A Tyco Electronics Company.)

one fabrication simply did not exist. Thus, after building the circuit using these early design approaches, measurements would invariably show that its performance did not meet requirements. For example, an amplifier might not have enough gain, or a filter might be set for the wrong frequency. This was fully expected.

After the initial failure, the designer would review exactly how the circuit failed to meet expectations and then incrementally tune the circuit (a process called “tweaking”), or possibly redesign and refabricate the circuit with the hope that it would perform better the second time around. This redesign–refabricate cycle might be exercised half a dozen times before the desired performance was achieved.

Then, in the 1980s, high-frequency designers started making RF circuits on silicon and gallium arsenide. Now, instead of a few days for fabrication, GaAs RFICs required 3–6 months for fabrication and the cost for fabrication was around \$50,000 each time. Tweaking a circuit was no longer possible. Five or six redesign–refabricate iterations were too expensive and took far too long. This is where the technology of numerical electromagnetics found its first widespread application.

2. APPLICATION EXAMPLE

Prior to the advent of RFICs, components like inductors were literally a simple coil of wire. The approximate inductance was calculated using simple formulas. Then with the inductor in the circuit, it could be tweaked, if desired, to the exact inductance by bending the wire or tuning a small ferrite slug.

Figure 2 shows a spiral inductor on silicon. The venerable coil of wire has been flattened into a spiral so that it can be placed on the flat surface of a silicon IC. This results in several non-inductor-like characteristics. First, recall that the magnetic field in the old coil of wire is just like that of the classical bar magnet, with all the magnetic

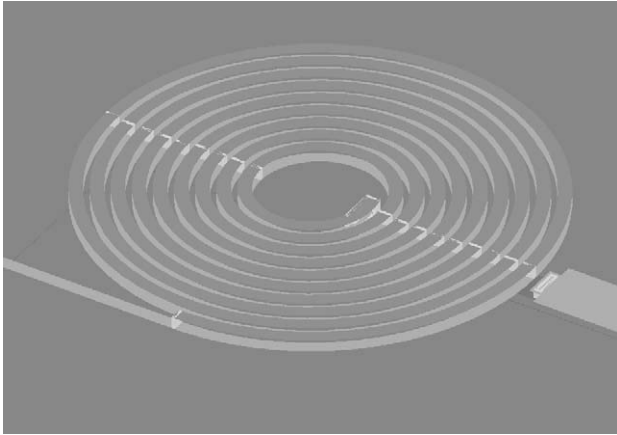


Figure 2. An 8.25-turn spiral inductor on silicon has thick conductors that must be modeled with two sheets of conductor to accurately include the skin effect and current crowding on inductance and loss.

field going through the center of the coil and then loosely spreading around the outside and finding its way back to the other end.

In a spiral inductor, some of the returning magnetic field pushes back through in between the spiral turns. The magnetic field that penetrates the plane of the spiral pushes and pulls the current in the spiral turns, the result is called “current crowding.” This increases the already high resistive loss of the tiny conductors in the inductor. In addition, while the spiral inductor is on top of the silicon substrate, there may be a conducting ground plane on the bottom side of the silicon. There is extra capacitance to this ground plane that can be significant. Making things more difficult, the semiconducting silicon substrate itself allows current to flow, further modifying this extra capacitance and increasing loss.

Note that this inductor is actually modeled as two thin sheets very close to each other. This is because the thickness of the metal is about the same as the gap between the spiral turns. In this case, both sheets are needed for the highest accuracy at high frequency because the current actually flows close to the surface of the metal (“skin effect”). Since there are two sheets of skin effect current in the actual inductor (one flowing on top of the thick metal, the other on the bottom), two sheets of current are used to model the inductor. So, while it initially looks like just one inductor, we actually have two very tightly coupled inductors, one directly on top of the other, and that is the way it is analyzed. This simple inductor is not so simple.

The Z axis is magnified for the inductor of Fig. 2, the metal thickness is $2.8\ \mu\text{m}$, and the gap between lines is $3.2\ \mu\text{m}$. No vias connecting the edges of the sheets, and no additional sheets are used as they are not needed. Testing to see if edge vias or additional sheets are needed is easily done by simply adding them, reanalyzing, and comparing the results. Typically we find that this simple two-sheet model is sufficient when the gap is on the order of the thickness, as we have here.

Measured versus calculated data for this inductor are shown in Fig. 3. The quality factor (Q) is related to resis-

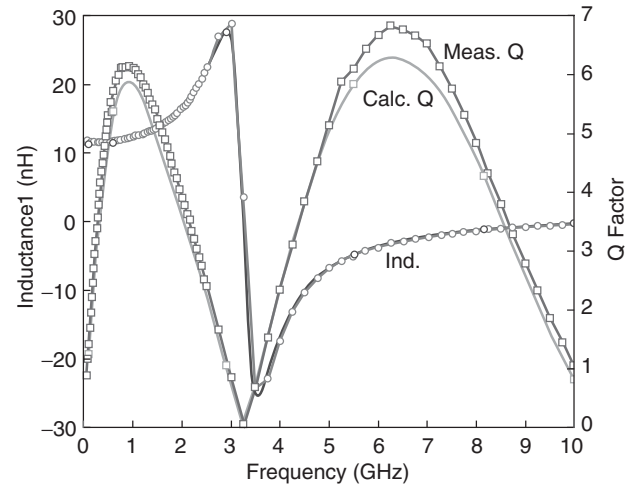


Figure 3. Measured (thin lines) versus calculated (thick lines) inductance and Q for the spiral inductor of Fig. 2.

tive loss; higher Q indicates lower loss. These results are typical of the kind of agreement usually seen with EM analysis. With this kind of capability, the high-frequency designer can now modify the design on the computer to achieve exactly the desired response, and then build the circuit once. Once a rare event, achieving complete success on the first fabrication for even the most complicated circuits is now common.

Figure 4 shows the current distribution on the inductor. First notice that there is high current on the edges. This edge concentration of current is characteristic of planar circuits and causes increased loss. Inclusion of this high edge current requires use of a very small subsection size. If a large subsection size is used, the high edge current cannot be included in the analysis and the loss is underestimated.

Note also that the high edge current is present on only one edge of some of the interior turns of the inductor. This is the current crowding referred to above, where the inductor magnetic field penetrating the plane of the induc-

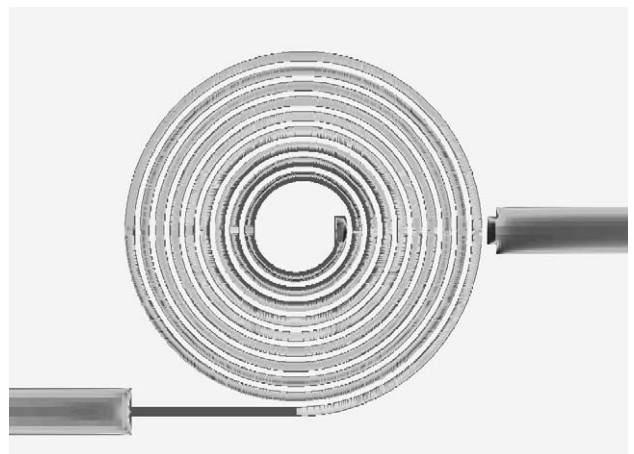


Figure 4. High current on the edges of planar conductors significantly increases loss. The effect of current crowding causes current to switch from side to side, further increasing loss. Current on the bottom side of the thick metal is shown.

tor pushes and pulls the current on the spiral turns to one edge or the other. This additionally increases loss and if not included in the analysis, results in an optimistic estimate of loss. Current distribution visualization is an important diagnostic tool.

This spiral analysis uses a new type of meshing, “conformal meshing,” which allows accurate and fast analysis. This spiral was analyzed in 5 min 9 s per frequency on a 3-GHz Pentium. Conformal meshing is discussed in detail below.

3. CIRCUIT THEORY ANALYSIS OF RFICS

Prior to the 1980s, most RF design was based on approximate circuit theory. The performance of any electrical circuit can be described in terms of voltage [measured in volts (V)] and current [amperes (A)] at a specified frequency [hertz (Hz)]. For example, a wall outlet might be described as providing 115 V at up to 15 A at 60 Hz.

This is true for RF circuits, only the frequency is much higher. For example, the radio signal going to the antenna of your cellphone might be 1 V at 20 mA at 900 MHz. This is a different voltage, different current, and different frequency, but basically the same idea.

The interplay between voltage and current (assuming a single fixed frequency) are mathematically described in terms of Ohm’s law, Kirchhoff’s current law, and Kirchhoff’s voltage law. We won’t detail these laws here, other than to point out that they are easily implemented in a computer. This was done just as soon as computers started to become widely available in the 1970s.

With the combination of computers and circuit theory, RF designers could quickly analyze their proposed designs consisting of inductors, capacitors, resistors, and transistors. This worked well, up to a point. At higher frequencies, the inductor was no longer just an inductor. Capacitance and resistance become important as well. In addition, there might be unwanted coupling between two different inductors. All this was not too much of a problem, because the designers could build their circuit, then tweak the design with a soldering iron, screwdriver, and pliers to achieve the desired performance.

Then, in the 1980s, circuits started getting smaller. Si and GaAs RFICs generated substantial interest. In the 1990s LTCC (with a dozen or more layers in one tiny ceramic block) saw development. As things became smaller and packed more closely together, tweaking a circuit simply was not possible. Effects that circuit theory did not include, such as stray coupling, became important. When a design did not work the first time, it was a complete re-fabrication. Designers could not wait years for the completion of multiple redesign–refabricate iterations.

4. ELECTROMAGNETIC ANALYSIS

In the 1980s, with the introduction of the IBM PC, serious computer power became available on the desktop at reasonable cost. Although initially starting with a clock rate under 4 MHz, the PC became steadily faster achieving multi-GHz clock rates with the turn of the century. It was

the advent of fast inexpensive desktop computing that allowed the next step in RFIC design to take place.

Clearly, something more precise than circuit theory with its volts and amperes was required. This turned out to be electromagnetics, which describes circuits in terms of electric field and magnetic field. We are all familiar with the static (i.e., unchanging) electric field surrounding the head of a longhaired youngster touching a Van de Graph generator. We are also familiar with the static magnetic field surrounding a bar magnet.

The interplay between electric field and magnetic field is governed by Maxwell’s equations. For a description of Maxwell’s equations appropriate for the college-bound high school senior, see Ref. 1. Maxwell’s equations deal with more than just static electric and magnetic fields. They completely describe how changing (i.e., dynamic) electric and magnetic fields behave. For a quick explanation, Maxwell’s equations state that a changing electric field generates a magnetic field. It also works in reverse; a changing magnetic field generates an electric field.

Remember that cellphone transmitting a signal at 900 MHz? Its signal, which consists of changing electric and magnetic fields interwoven together and each completely dependent on the other for their existence, changes direction and then back again 900 million times a second. How they do this and what happens as a result is completely covered by Maxwell’s equations.

The mathematical details are discussed later in this article; and while the top-level concepts are relatively simple, the detailed math is complicated. To appreciate the complexity, note that both the electric and magnetic fields are vector fields. In other words, any computer analysis must determine both the direction and magnitude for both electric and magnetic fields everywhere in the circuit being analyzed. This is not a trivial problem.

Additionally, if the analysis is for a specific frequency (i.e., a frequency-domain analysis), the phase of each field at each point in the circuit must be determined. In an alternative approach, the EM analysis generates signals at all frequencies simultaneously. This is EM analysis in the time domain, and now the magnitudes and directions of both fields must be determined as a function of time everywhere in the circuit.

4.1. Volume Meshing EM Analysis

There is tremendous complexity within Maxwell’s equations. This means that there is also a huge diversity in methods of solution. Four general approaches have seen wide application in high-frequency design. All the approaches rely on meshing the high-frequency circuit into many small subsections. Maxwell’s equations are then applied to find the electric and magnetic fields for each subsection in the entire mesh.

The most general class of techniques relies on meshing the entire volume of the problem. The approach called “finite elements” meshes the entire volume into small tetrahedral cells. A technique known as finite-difference time domain (FDTD), or a related technique known as finite integration technique (FIT), meshes the entire volume into tiny rectangular cells. For both techniques, when a

more accurate solution is required; simply use more and smaller cells or tetrahedra.

Of course, a finer mesh means the analysis time increases. As with all EM analyses, analysis time can easily exceed tens of hours. The analysis time is strongly related to the number of cells or subsections. Thus, there is a practical upper limit on the complexity of the circuit that can be analyzed using electromagnetics. Where this upper limit occurs depends on the specific EM analysis used and the type of circuit being analyzed.

The different EM analysis techniques all have their relative advantages and disadvantages. This is important to understand as using an EM analysis on an inappropriate type of problem can result in analysis times 10–1000 times longer, if indeed the problem can be done at all.

Volume meshers are ideal for 3D arbitrary structures that can have any shape and structure whatsoever. A classic example is a cellphone held close to a human head (Fig. 5).

A disadvantage of meshing a problem using tetrahedra is that linear current flow is not well represented. This is easily seen by viewing a finite-element current distribution. The current distribution can lack the smoothness that is characteristic of a true current distribution. In addition, the high current that is naturally present at any sharp metal edges can be indistinct, if present at all.

Fortunately, the response of a circuit is not strongly dependent on the exact current distribution. Thus, useful data can easily result even if the current distribution is not exactly calculated. However, an extremely high accuracy solution might be elusive.

The finite-element method works in the frequency domain. In other words, the analysis assumes that there is a signal at only one specific frequency at any one time. In order to obtain data at 100 frequencies, 100 analyses must be performed. There are interpolation approaches that can provide a spectrally rich dataset after completing a full analysis at fewer frequencies as discussed below.

FDTD and FIT are time-domain approaches that assume that all frequencies are present in the circuit at the same time. This is done by exciting the circuit with an

impulse (actually a narrow Gaussian pulse, to avoid numerical problems). The impulse contains energy at all frequencies, and the circuit response is analyzed as a function of time. After the time impulse response is calculated, it is transformed to the frequency domain. In this way, a spectrally rich dataset is generated at all frequencies present in the original impulse input without need for interpolation.

The advantage of a time-domain analysis is that the circuit response at all frequencies is determined in a single analysis. The disadvantage is that the analysis can handle an impulse input on only one port (i.e., input or output terminal/connector) at a time. Thus, in order to completely characterize a circuit with, say, 15 ports, 15 complete time-domain analyses are required. The finite-element approach also suffers from this same problem.

While both finite-element and the time-domain analyses can, and are, used for planar circuit analysis, they are usually much slower than tools specifically designed for planar analysis. Thus, volume meshing tools should typically be used for planar circuits only when a specialized planar tool is not available, or if there is an important nonplanar aspect to the circuit. In addition, the volume meshing tools can be used to double-check the results of a planar tool, assuming that the circuit is not too complicated.

4.2. Surface Meshing EM Analysis

Silicon and GaAs RFICs, and LTCC circuits are all planar circuits. In general, use of a volume meshing tool to analyze these, and other planar circuits, is inappropriate. Both faster analysis times and higher accuracy are realized by using EM tools made specifically for planar circuits. Such tools take advantage of the fact that these kinds of circuits are mostly planar (with short vias connecting circuit on different layers) and embedded in layered dielectric. In addition, each layer of dielectric must be uniform within itself, with the possible exception of very small volumes (e.g., a small hole in a layer of dielectric). These specializations allow these tools to subsection only the surface of the metal in the circuit. This is far more efficient and accurate than subsectioning the entire volume of a planar circuit.

There are two basic kinds of surface meshing planar EM tools: (1) those intended for unshielded circuit analysis and (2) those used for shielded circuit analysis (i.e., a circuit contained in a conducting box). In both cases, the circuit metal (and only the circuit metal) is meshed into a set of small subsections. As with the volume meshers, smaller subsections mean higher accuracy, but at the cost of increased analysis time. And also as with the volume meshers, at some point the number of subsections increases the analysis time to tens of hours, setting an upper limit to the level of circuit complexity that can be analyzed. Exactly where this upper level is can vary widely depending on the type of circuit and specific tool being used. No single tool is superior for all planar circuits.

Both types of surface meshing approaches first calculate the coupling between every possible pair of subsections, namely, put current on one subsection and calculate

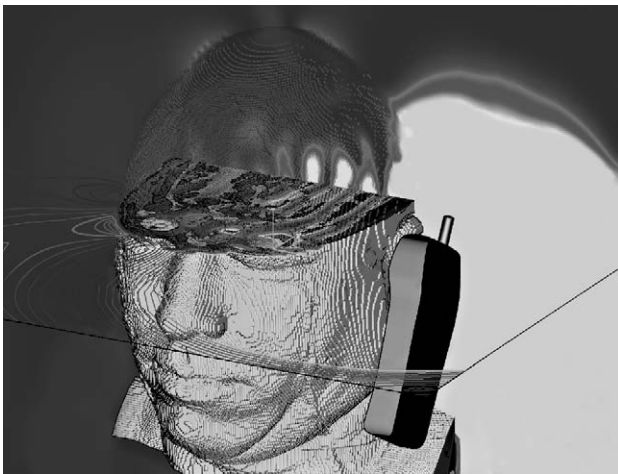


Figure 5. Volume meshing tools excel for 3D arbitrary geometries. Problems of critical importance can be solved easily.

the voltage induced in another subsection. If there are 100 subsections, then the calculated pairwise coupling is stored in a 100×100 matrix. The analysis then inverts the 100×100 matrix to yield the current distribution and total circuit response. Since this is a frequency-domain analysis, this process must be repeated at each frequency of interest. The principal difference between unshielded and shielded analysis is the means by which they calculate the coupling between subsections. Unshielded EM analysis calculates the subsection-to-subsection coupling by means of numerical integration. For shielded analysis, the coupling is calculated by using the fast Fourier transform (FFT). Each approach has advantages and disadvantages.

The advantage of the unshielded/numerical integration approach is that the subsections can be of any shape, size, or orientation (e.g., triangle or rectangle, or even polygon). This is because the numerical integration takes place over the area of the subsection. Since numerical integration is easily performed over any desired area, any desired size subsection can be used. The disadvantage of the unshielded analysis is that the numerical integration introduces numerical integration error. Special care must be taken to avoid numerical integration error in the vicinity of poles, which result from surface waves present in unbounded media. Depending on the specific circuit, design requirements, mesh, and frequency of analysis, the numerical integration error may or may not be important. Fortunately, this is typically of concern only when high accuracy is required.

Shielded analyses use a 2D FFT (fast Fourier transform) to calculate the coupling. Recall that in digital signal processing, a time signal must first be uniformly time-sampled prior to using the FFT. The same is true with FFT-based EM analysis, except that the planar surface of the circuit substrate is first uniformly space-sampled in two dimensions. This means that the shielded/FFT based analysis starts with a fine uniform underlying FFT mesh. The FFT mesh can easily be 1024×1024 cells, so the individual cell size can be about the same size as a pixel on a computer screen. Nevertheless, the principal disadvantage of shielded/FFT analysis is that, like a picture on a computer screen, the circuit outline is first pixilated so that its edges follow the fine FFT mesh.

The FFT also brings the shielded analysis its principal advantage. Because there is no numerical integration, there is no numerical integration error. The coupling between each pair of subsections is calculated to full numerical precision. Thus, the accuracy and dynamic range of shielded/FFT EM analysis is the highest that can be obtained. The accuracy of shielded EM analysis is quantitatively explored below.

Generally, if high accuracy or the effect of a shielding box might be important, a shielded/FFT analysis should be used. If accuracy is not of the highest importance or the effect of shielding sidewalls cannot be allowed, then an unshielded analysis can be used. There are additional considerations in selecting the appropriate analysis technique for a given problem, too numerous to discuss here. Ideally, a high-frequency planar circuit designer will have access to both shielded and unshielded tools and will be

able to use each as appropriate. The tradeoff between the two is very much like the tradeoff between analog cell-phones (unshielded/numerical integration) and digital cellphones (shielded/FFT). Many of the same issues arise.

5. COMPANION MODELING

The usual EM-based high-frequency design flow is

1. Design the circuit using circuit theory, optimize to meet all requirements.
2. Lay out the circuit.
3. Analyze the circuit using EM.
4. If the circuit meets requirements, build the circuit.
5. Otherwise, change (i.e., “tweak”) the layout in hopes of improving performance.
6. Return to step 3 and see if the changes worked.

This design process can be performed manually, or with the use of automated optimization algorithms. Automated optimization algorithms that drive an EM analysis should be used only if the design already yields nearly the desired response due to the typically long time required for high-accuracy EM analysis. This design flow represents a huge advance over that of the 1980s, when a fabrication rather than EM analysis, was used in step 3. This greatly increased the time and expense to design completion.

A modification of the abovementioned approach, called “companion modeling” [2], can substantially further reduce the time and expense. This process adds structure to the tweaking step, step 5, in the design flow:

- 5a. Determine a mapping between the critical parameters of the circuit theory model and the layout. For example, associate the length of a transmission line in the circuit theory model with a specific dimension in the layout. This process is called “space mapping” [3].
- 5b. Optimize the circuit theory model to match the EM analysis results.
- 5c. Note the changes in the critical parameters of the circuit theory model. Reverse those changes in the layout. For example, if the optimized circuit theory transmission line became $10 \mu\text{m}$ shorter, then lengthen that transmission line (in the layout) by $10 \mu\text{m}$.

Step 5b is counterintuitive. The circuit theory result meets all requirements. The initial EM analysis does not. So the high-frequency designer instinctively wants to optimize the layout (using EM analysis) to match the circuit theory. The problem is that any EM-based optimization is very slow. Unless the result is close, this will be a long and possibly never-ending process. Instead, step 5b optimizes the desired circuit theory result to match the not-so-good EM result. Now the optimization loop uses the much faster circuit theory. With proper selection of critical parameters, an answer is obtained quickly. Then proceeding to step 5c, if a decrease of $10 \mu\text{m}$ in a transmission-line length changes the good circuit theory response so as to match

6 APPLIED NUMERICAL ELECTROMAGNETIC ANALYSIS FOR PLANAR HIGH-FREQUENCY CIRCUITS

the not-so-good EM result, we reason that an increase of $10\ \mu\text{m}$ should make the not-so-good EM result nearly match the desired circuit theory response. Just reverse the changes, and the layout can be quickly tweaked into the desired response. The entire loop usually needs to be performed only 2 or 3 times before it is exited at step 4.

Although companion modeling (a special case of a general technique known as *space mapping*) is not presently widely known, the power of this approach means that it is likely to become, in some form, the primary design flow approach for EM-based high-frequency design.

6. DIVIDE-AND-CONQUER ANALYSIS STRATEGY

No matter what approach is used to solve a problem, for some problem size, the time required becomes too large to obtain a solution in a reasonable length of time. This can happen easily when using EM analysis tools. Analysis times of days or even weeks can be easily seen. Problems that are too large for computer memory or take too long to analyze still need a solution of some kind.

Divide-and-conquer strategies are useful in many of these situations. Figure 6 shows a typical parallel-coupled line bandpass filter. While this filter is still easily analyzed as one complete filter, it serves to illustrate the divide-and-conquer strategy.

This filter is divided into eight subcircuits. Now, rather than analyzing the entire filter at once, each of the eight subcircuits is analyzed separately. Because the EM analysis time is typically proportional to the number of subsections cubed, performing eight small analyses is much faster than performing one large analysis.

After the eight subcircuits are analyzed, they are all connected together by circuit theory nodal analysis and the S parameters of the two-port filter result. In modern EM analysis software, this entire process is automatic once the user has specified the circuit subdivisions. A disadvantage of this approach is that coupling between the different sections is not included. Such coupling occurs only between line segments that are not perpendicular to the dividing line. Thus, one should not specify any horizontal circuit dividing lines for this particular circuit as such lines are parallel to the filter's coupled lines. The coupling between resonators is critical to the filter performance and must be included. Thus, all dividing lines in

this filter are perpendicular to the coupled lines of the filter.

With the circuit divided as shown, coupling between nonadjacent resonators is not included. If such coupling is important, as is the case with elliptic function filters, then this approach should not be used in a manner that causes such coupling to be eliminated from the analysis.

Extremely precise deembedding of tightly coupled ports is required for success of this approach, as described in Section 11. If there is any port discontinuity left in any of the subcircuit ports, that remaining port discontinuity is inserted in the center of the resonators of the complete circuit. This disrupts the filter response.

Manual modification of the automatically generated netlist (which is used by circuit theory to connect all eight subcircuits together) allows an even faster analysis. Note that subcircuit S_8 is identical to subcircuit S_1 (with a rotation). Thus, switching subcircuit S_8 references to subcircuit S_1 with appropriate attention to node numbering can eliminate all references in the netlist to subcircuit S_8 .

By noting other symmetries, only four of the eight subcircuits need to be analyzed electromagnetically. This modification further cuts the analysis time by half. Measured, and calculated results are shown in Fig. 7.

AU:1

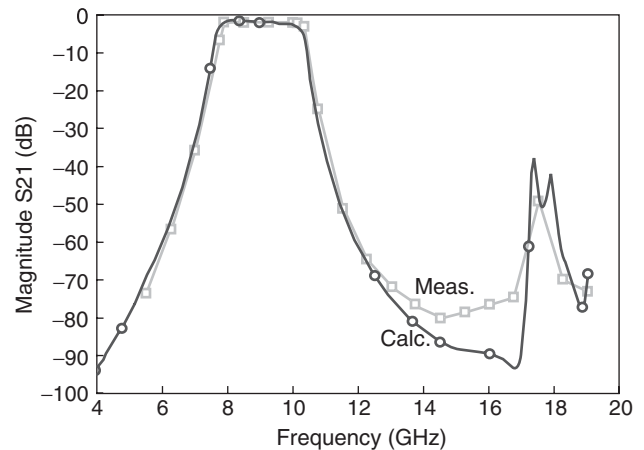


Figure 7. Measured versus calculated data show visually identical results nearly everywhere validating the divide-and-conquer approach for this filter.

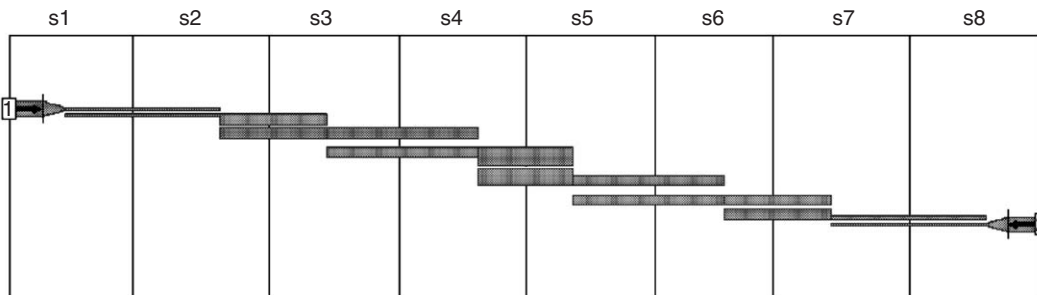


Figure 6. A parallel-coupled line bandpass filter is divided into eight smaller sections for much more rapid analysis.

7. ERROR EVALUATION

Accuracy is the primary reason in using EM analysis. If accuracy is not an issue, then the high-frequency designer should use circuit theory, which is much faster. Accuracy directly impacts whether success on first fabrication is achieved. Because of the critical importance of success on first fabrication, the issue of accuracy cannot be left to chance. All handwaving, subjective, nonquantitative claims of “high accuracy” should be ignored if such statements are made in the absence of hard quantitative data.

Because EM analysis always has some degree of error, success on first fabrication cannot be guaranteed. In a complicated circuit, there are many sources of error. Often these errors average out, so there is no problem. However, sometimes the errors add together in the same direction, sometimes by chance, and sometimes causally. When the errors add together, the result is a design failure. There is always a risk of failure. The successful high-frequency designer works aggressively to minimize that risk. In that way, rather than getting, say, 5 out of 10 designs to succeed on first fabrication, the knowledgeable designer sees 9 out of 10 designs succeed. The competitive advantage is substantial.

An attitude sometimes seen among high-frequency designers is that several percent error is not a problem. This can be true, depending on the circuit and on the requirements. Several percent error in the transmission-line characteristic impedance for a branchline coupler is not likely to be a problem. Several percent error in the velocity of propagation for a 5% bandwidth filter will require at least one additional fabrication. The informed designer will know when to, and when not to strive to realize absolute minimum analysis error.

A widely held, and usually untested, belief is that as long as the subsection size is small with respect to the wavelength, then EM analysis error is small. This is incorrect. In fact, subsection size must be small with respect to how rapidly (as a function of position) the current distribution changes. Thus, as described below, subsection width must be small with respect to linewidth in addition to being small with respect to wavelength. Specific quantitative knowledge of EM analysis error is critical to the design engineer’s success.

The most common approach to accuracy validation involves the widespread “good agreement between measured and calculated” (GABMAC) plot, examples of which are shown in Figs. 3 and 7. While certainly important as a final reality check, the GABMAC plot is of little value for quantitative determination of analysis error as differences can be additionally due to measurement and fabrication error. Determining the magnitude of the analysis error by itself in such tests is nearly impossible.

Each of the following tests are easily performed on any EM analysis. Intended to supplement, not replace, the usual GABMAC test, these tests use simple circuits that are easily analyzed on any EM tool and allow precise quantitative evaluation of analysis error.

7.1. The Stripline Standard

Exact theoretical solutions are known for a small set of planar problems. Such problems allow the precise evaluation of analysis error. Any and all differences between calculated data and the exact answer is analysis error.

Stripline is one such problem that has an exact solution and has been used to precisely evaluate the error of EM analysis [4]. For example, as applied to a shielded planar EM analysis, the following empirically determined upper bounds are found for analysis error

$$e_{z0} \leq \frac{16}{N_W} \quad (1)$$

$$e_V \leq 2 \left(\frac{16}{N_L} \right)^2 \quad (2)$$

where

e_{z0} = percent error in characteristic impedance

e_V = percent error in velocity of propagation

N_W = number of cells across width of line

N_L = number of cells per wavelength along length of line

These upper bounds are met nearly in equality for most of their range.

Figure 8 shows a performance plot where analysis error is plotted versus analysis time on a 3-GHz Pentium. For each data point, N_W is doubled. N_L is held constant at 1024 cells per wavelength. N_W is taken up to 1024, and although larger values can be evaluated, a larger value of N_L would be required in order to continue to see convergence. Even as it is, the error convergence starts to slow down for large N_W . It is important to note that the convergence to the exact correct answer is smooth, the error decreasing by about half each time N_W is doubled.

When the convergence is smooth, Richardson extrapolation [5,6] can be used to arrive at a nearly exact answer based on two lower accuracy results. For this dataset, Richardson extrapolation converges to 0.002% error, one order of magnitude better. Note that this performance plot can be viewed as a lower bound (i.e., best possible speed/

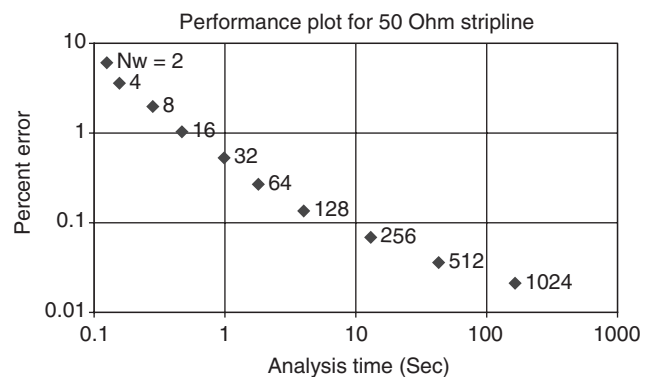


Figure 8. A performance plot shows the precise analysis error versus analysis time for an exactly known zero-thickness lossless stripline. Note the logarithmic scale on both axes.

8 APPLIED NUMERICAL ELECTROMAGNETIC ANALYSIS FOR PLANAR HIGH-FREQUENCY CIRCUITS

accuracy performance) for a given EM analysis. Most practical circuits are more complicated. If one were able to generate a performance plot for a more complicated circuit, the analysis time would be longer, moving the performance curve to the right. In addition, there would be more sources of error, thus moving the performance curve up. Thus the stripline standard performance plot represents a kind of starting line for the race to analyze a circuit both as quickly and as accurately as possible.

For metal–insulator–metal or (MIM) capacitors [7], common on RFICs, the following upper bounds are also nearly met in equality over most of their range:

$$e \cong \frac{5.12}{1/N_A + 1/N_B} \quad (3)$$

where

e = percent capacitance error
 N_A = number of cells among capacitor length
 N_B = number of cells across capacitor length

These equations may be used to evaluate and simply subtract the error from a single EM analysis of an MIM capacitor. In fact, the analysis error of the MIM capacitor is so well behaved that an EM analysis can be used to directly determine the amount of capacitance due to fringing field, usually a very small amount.

7.2. The Zero-Length Coupled Line

Another approach is to use a coupled pair of microstrip lines. There is no exact solution for such a structure; however, there is a degenerate case that is of interest for error evaluation: the zero-length coupled line.

No EM analysis can analyze a zero-length coupled line directly. Rather, they must analyze a finite length of line and then deemed to zero length. The exact answer is simply that all reflection coefficients and all coupled transmission coefficients must be zero, or $-\infty$ dB. Because all EM analyses have error, something other than $-\infty$ dB results. This result is the noise floor of the analysis.

Figure 9 shows that the noise floor for a shielded planar analysis is around -130 dB. Generally analysis results should not be trusted down to any more than 20 dB above the noise floor. The noise floor for this specific EM analysis has been found to range between -100 and -180 dB, depending on the circuit and mesh size used.

7.3. The Thick Stripline Benchmark

Another benchmark is thick stripline [8]. The characteristic impedance is known for the selected geometry (Fig. 10) to better than $\pm 0.0006\%$. This upper bound on the error is estimated by using the thick stripline equations for the line of Fig. 10, except that the thickness set to zero. That answer is then compared with the known exact solution for zero-thickness stripline. It is at zero thickness that the thick stripline equations have maximum error; thus the difference at zero thickness is an upper bound on the thick stripline equation error when using nonzero

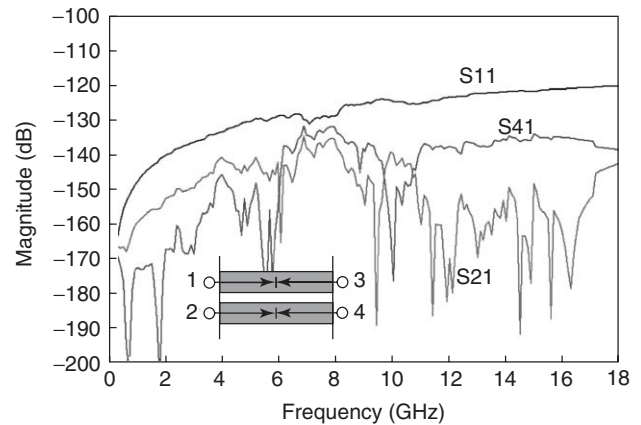


Figure 9. The exact value of S_{11} , S_{21} , and S_{41} of a zero-length coupled line should be $-\infty$ dB. The calculated values, plotted here, show that the analysis numerical noise floor is below -120 dB.

thickness. In addition, when the zero-thickness result is compared to the result with thickness, as in Fig. 10, it is found that the characteristic impedance changes by 25% between the two cases.

Thus, we have a structure whose correct answer is known very precisely and that is also strongly dependent on the parameter of interest: thickness. This is ideal for a benchmark. Thickness is modeled in a planar analysis with multiple sheets [9]. As the number of sheets increase, and as N_W increases, the result should converge to the exact answer, as seen in Fig. 11. Note once more that the smooth convergence, with the error decreasing by about half each time, allows a Richardson extrapolation [5,6] to provide a nearly exact answer.

The error convergence has been shown to be good for this specific shielded EM analysis; a similar convergence test can be performed to quantify the expected error due to thickness for an actual design situation. In practice, it has been found that multisheet models for thickness are needed when either the width of the line, or the gap between lines, is on the order of or less than the thickness of the lines. In some RFIC situations, this can be common.

Extreme cases of thickness are becoming common as decreasing loss becomes important in RFIC design. For example, a $6\ \mu\text{m}$ metal thickness combined with a $2\ \mu\text{m}$

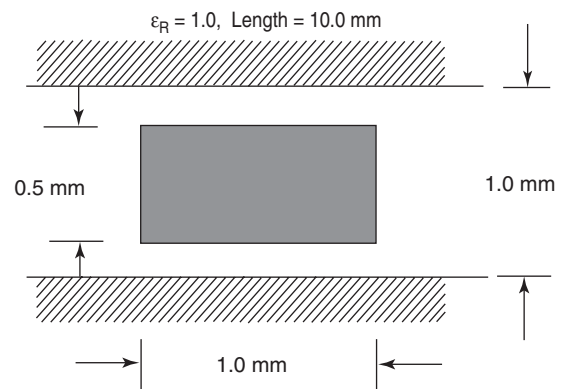


Figure 10. An extremely thick stripline is used to test the multisheet model for accuracy.

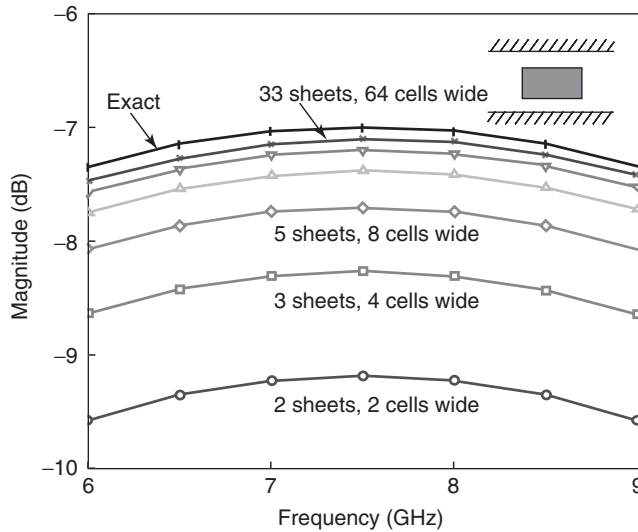


Figure 11. As the number of zero-thickness sheets used to represent this very thick stripline increases, S_{11} converges to the exact answer.

gap between lines would require perhaps half a dozen sheets for accurate analysis at high frequency. In this case, it is possible to simulate coupling between such closely spaced thick lines by using the two-sheet model and modifying the permittivity and permeability of nearby dielectric to compensate for the coupling that is otherwise underestimated by the two-sheet model.

AU:2

8. ELECTROMAGNETIC INTERPOLATION

When using a frequency-domain analysis, which includes all the planar approaches described above, a complete analysis must be performed at each frequency. If data at 100 frequencies are required, then 100 complete analyses must be performed.

In order to reduce the number of analyses required, interpolation can be used. Substantial work in this area has been accomplished since the mid-1990s, with spectacular results. A recent (at time of writing) publication in this area, including an extensive bibliography, is Ref. 10. Here, we briefly describe only a simple approach to give an idea of what can be done.

8.1. Linear Interpolation

In high school, we are introduced to linear interpolation, where we are given two data points: (x_0, y_0) and (x_1, y_1) . We then calculate the coefficients a and b of the equation

$$y = ax + b \quad (4)$$

that passes through the two points. We can then perform an interpolation by using this equation to calculate y for all desired values of x . While simple algebra is all that is required, the solution of this problem can also be cast in the form of a 2×2 matrix. The solution for a and b is obtained by inverting the matrix. For EM analysis, x is the frequency and y is the calculated data (usually S , Y , or Z

parameters), which is complex (real and imaginary, or magnitude and angle).

8.2. Cubic Spline Interpolation

At the next level, we have the cubic spline. The process is identical to that above, except that it uses an equation that is a bit more complicated:

$$y = a_3x^3 + a_2x^2 + a_1x + a_0 \quad (5)$$

Now there are four unknown a coefficients; thus we require four data points and we invert a 4×4 matrix to solve for the four coefficients. Once that task is done, we may evaluate the equation for any and all values of x (frequency) that we desire and realize many data points after having only calculated data at four frequencies.

8.3. Padé Rational Polynomial Interpolation

The cubic spline has problems for interpolating high-frequency data. For example, if impedance or admittance is being interpolated, y might need to go to infinity (i.e., a pole). At a given frequency (x), and with finite coefficients, the result of a cubic spline cannot go to infinity. This suggests that we use a ratio of two polynomials, for example

$$y = \frac{a_3x^3 + a_2x^2 + a_1x + a_0}{b_2x^2 + b_1x + 1} \quad (6)$$

where, we have four unknown coefficients in the numerator (providing for three “zeros” in the resulting function and setting the amplitude of y at $x=0$) and two unknown coefficients in the denominator (providing for two poles in the resulting function). With a total of six unknowns, we now require six data points. In a manner similar to that seen for the cubic spline, with six data points, we can write six linear equations (by multiplying both sides by the denominator polynomial), resulting in a 6×6 matrix. Solution of the matrix yields the six unknown coefficients.

The total number of terms both numerator and denominator must be less than or equal to the total number of data points taken. However, their distribution between numerator and denominator is arbitrary. More recent research has been directed toward finding a distribution that is optimum in some way for a given set of data.

This ratio of two polynomials is known as a *Padé rational polynomial*. By noting the similarity to the Laplace transform of a lumped circuit, we note that, given a sufficient number of terms, the Padé rational polynomial can exactly represent an arbitrary lumped circuit. However, it cannot exactly represent a distributed circuit. Thus, for high-frequency circuits, the Padé rational polynomial is generally bandlimited.

8.4. Applied Interpolation Issues

As typically applied to EM analysis, at least two analyses are initially performed. This yields two data points. Then several interpolations are formed, either with different numbers of data points, or with a different distribution of coefficients between numerator and denominator. The dif-

ferences between these different interpolations (all based on the same data) are treated as an estimate of the interpolation error. While the true interpolation error can be as much as 20 dB greater than the estimate, the estimated error does tend to correctly show the frequency at which the true error is highest. The next data point is then taken at this highest error frequency. The iterative algorithm proceeds until the maximum estimated error is below a user-selected threshold.

Early application of the Padé rational polynomial to high-frequency EM analysis was limited in bandwidth, dynamic range, and occasional failure to converge. However, more recent results now provide a much more robust implementation. Basically, in addition to the actual high-frequency circuit data at each frequency, the interpolation can also make use of the tremendous amount of information contained within the method-of-moments (MoM) matrix used by the EM analysis, for example, frequency derivative information.

The advanced techniques can easily interpolate over a $1000 \times$ bandwidth (e.g., from 0.1 to 100 GHz) with analysis at 15–20 frequencies, or over a small bandwidth with analysis at only four or five frequencies, no matter how complex the response (see Fig. 12).

Even so, today's interpolation algorithms do still exhibit failure modes of which the informed designer should be aware. In one failure mode, the interpolation fails when an attempt is made to interpolate data that are below the analysis noise floor. The noise floor depends strongly on the specific EM technique, specific circuit, specific meshing, and specific frequency. Observed noise floors range from 40 to over 180 dB down for the different types of planar analyses. Special caution should be exercised whenever interpolation is invoked on data that approach to within 20 dB of the noise floor.

The second failure mode is seen in shielded analysis. Early algorithms could fail with only one box resonance. Modern algorithms are more robust, providing accurate data even with large numbers of box resonances; however, the number of data points required for convergence can be

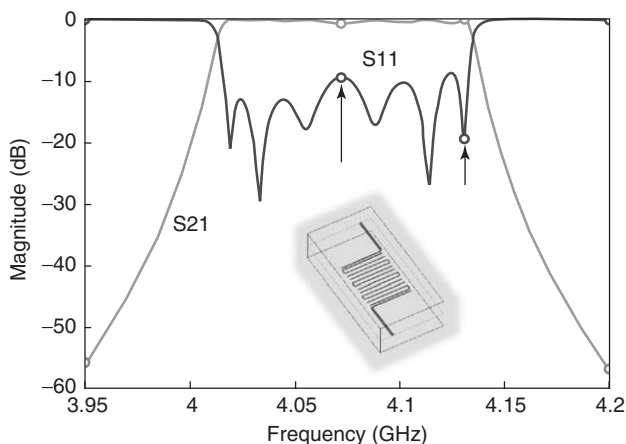


Figure 12. Only four frequencies were analyzed for this six-resonator filter: the start, stop, and the two arrowed frequencies. All the remaining data are interpolated. The interpolated data are visually identical to those from a full calculation.

excessive. Thus, even though the resulting data are still accurate, analysis time can become excessive.

9. CONFORMAL MESHING

As described above, one advantage of the unshielded EM analysis is that the required numerical integration can be performed over any arbitrary subsection as desired. In practice, arbitrary size, shape, and orientation rectangles and triangles are used. More advanced techniques additionally include arbitrary polygons.

Shielded analyses use the FFT to calculate the coupling between subsections. While the FFT provides unsurpassed accuracy and dynamic range, it requires a fine underlying uniform FFT mesh. These tiny FFT mesh cells are joined together into larger rectangular subsections.

In order to increase speed, both shielded and unshielded analyses try to maximize use of large subsections. Large subsections substantially reduce the size of the matrix, thus speeding the analysis. However, large subsections also decrease analysis accuracy, especially when the subsections are so large that the natural high edge current is not allowed to form.

The arbitrary triangles of the unshielded analysis are a distinct advantage in analyzing arbitrary smoothly curving circuits. A few triangles can easily form a piecewise linear representation of a curving edge, especially when the error introduced by ignoring the high edge current is acceptable. However, when high edge current must be included, a large number of narrow rectangular subsections must be inserted on the conductor edges, rapidly increasing the subsection count and analysis time.

Merging the small FFT cells of the shielded analysis can be severely limited by curving geometries. Even the largest rectangular subsections must still be small, thus limiting the reduction in the number of subsections. As a small compensation, the high accuracy yielded by the small subsections is still seen. To alleviate this bottleneck, conformal meshing [11] was developed. The tiny FFT cells are still present, so the accuracy and dynamic range provided by the FFT remains uncompromised. However, instead of merging the cells into larger rectangular subsections, the cells are merged into large subsections that both cover the entire width of a transmission line and curve to follow arbitrary curving edges. In this way, a subsection count reduction of 100 times or more can be achieved. Since matrix solve time increases with the subsection count cubed, a $100 \times$ reduction in subsection count realizes a $1,000,000 \times$ faster analysis.

Normally large subsections that cover the entire width of a transmission line result in decreased accuracy because high edge current is not allowed. This is not the case with conformal mesh because the mesh automatically includes the high edge current in the conformal subsections. Thus we simultaneously achieve two seemingly contradictory goals: the fast analysis of a few large subsections and the accurate analysis of many small subsections.

Figures 2–4, presented and discussed above, show a circular 8.25-turn spiral inductor that was analyzed using conformal meshing. No other form of meshing and no oth-

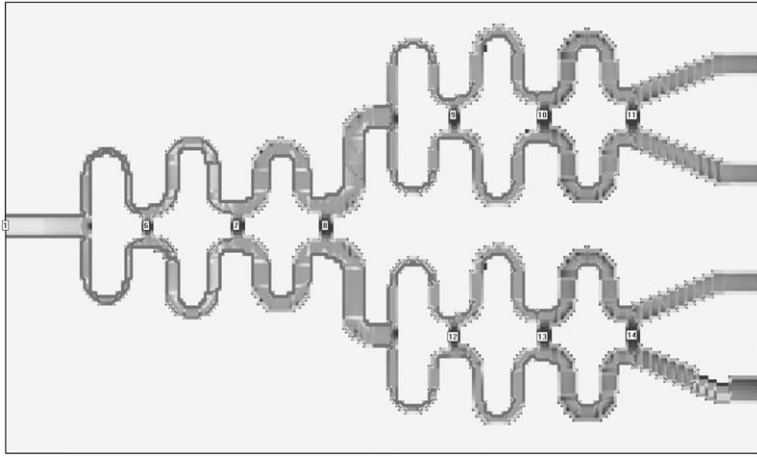


Figure 13. Conformal meshing allows analysis of this four-way power splitter including the critical high edge current in 19 s per frequency on a 3-GHz Pentium.

er technique of analysis has yet been found to successfully analyze this inductor, including the effect of conductor thickness and high edge current. While one other approach was found to be able to complete an analysis, the result was overwhelmed with numerical noise due to the large problem size. Using conformal meshing, the inductor requires 5 min 9 s per frequency on a 3-GHz Pentium. Six frequencies were analyzed and interpolated to provide the plotted data.

Figure 13 shows a four-way power splitter analyzed using conformal meshing. Note that the high edge current is well represented, even though the analysis required only 19 s per frequency on a 3-GHz Pentium.

Conformal meshing marks a major advance in the size of problem for which numerical EM analysis can achieve fast and accurate results.

10. SHIELDED ELECTROMAGNETIC ANALYSIS THEORY

Because of the complexity of Maxwell's equations, there are innumerable methods of solution, each approach having advantages and disadvantages for a given type of problem. In this section, we describe one approach to the analysis of planar circuits in a shielded environment using the method of moments [12]. This particular solution is detailed in Refs. 13 and 14.

For the geometry of Fig. 14, we can treat the perfectly conducting sidewalls as a rectangular waveguide propagating in the z direction. The rectangular waveguide TE (transverse electric) and TM (transverse magnetic) modes in each dielectric region form a complete orthogonal basis for any source-free field. We sum all the TE and TM modes (over the composite summation index i) in each region so as to match tangential electric fields at the interface be-

tween the two regions ($z = h$)

$$\mathbf{E}_t^{(1)} = \sum_i V_i \frac{\sin(k_{iz}^{(1)} z)}{\sin(k_{iz}^{(1)} h)} \mathbf{e}_i \quad (7)$$

$$\mathbf{H}_t^{(1)} = \sum_i V_i Y_i^{(1)} \frac{\cos(k_{iz}^{(1)} z)}{\sin(k_{iz}^{(1)} h)} \mathbf{h}_i \quad (8)$$

$$\mathbf{E}_t^{(0)} = \sum_i V_i \frac{\sin(k_{iz}^{(0)}(c - z))}{\sin(k_{iz}^{(0)}(c - h))} \mathbf{e}_i \quad (9)$$

$$\mathbf{H}_t^{(0)} = \sum_i V_i Y_i^{(0)} \frac{\cos(k_{iz}^{(0)}(c - z))}{\sin(k_{iz}^{(0)}(c - h))} \mathbf{h}_i \quad (10)$$

where

$$\begin{aligned} Y_{i,TE}^{(p)} &= -jk_{iz}^{(p)} / \omega\mu_p \\ Y_{i,TM}^{(p)} &= -j\omega\epsilon_p / k_{iz}^{(p)} \\ k_{i,z}^{(p)} &= (k_p^2 - k_x^2 - k_y^2)^{1/2} \\ k_p &= \omega(\mu_p\epsilon_p)^{1/2} \end{aligned}$$

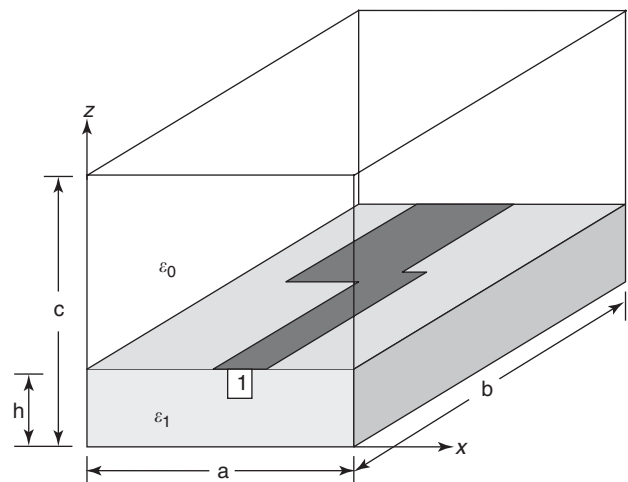


Figure 14. Shielded electromagnetic analysis views the conducting sidewalls of the shielding box as rectangular waveguide. The fields for each layer are written as a sum of waveguide modes.

$$\begin{aligned}
k_x &= m\pi/a \\
k_y &= n\pi/b \\
\mathbf{e}_{i,TE} &= N_1 g_1 \mathbf{u}_x - N_2 g_2 \mathbf{u}_y \\
\mathbf{e}_{i,TM} &= N_2 g_1 \mathbf{u}_x - N_1 g_2 \mathbf{u}_y \\
\mathbf{h}_I &= \mathbf{u}_z \times \mathbf{e}_i \\
g_1 &= \cos(k_x x) \sin(k_y y) \\
g_2 &= \sin(k_x x) \cos(k_y y) \\
N_1 &= (2/ab)^{1/2}, \quad m=0, \quad n \neq 0 \\
N_1 &= 0, \quad m \neq 0, \quad n=0 \\
N_1 &= 2(n/b)(ab/(n^2 a^2 + m^2 b^2))^{1/2}, \quad m \neq 0, \quad n \neq 0 \\
N_2 &= 0, \quad m=0, \quad n \neq 0 \\
N_2 &= (2/ab)^{1/2}, \quad m \neq 0, \quad n=0 \\
N_2 &= 2(m/a)(ab/(n^2 a^2 + m^2 b^2))^{1/2}, \quad m \neq 0, \quad n \neq 0
\end{aligned}$$

The modal admittances are those of the standing-wave modes, which differ from the more usual traveling-wave modes by the imaginary factor j . The amplitude of the i th mode (V_i) is determined as described below.

Note that the equations above, and this entire approach, maintain full validity for any degree of lossy or conductive substrates. The only change required is that the characteristic impedances and wavenumbers presented above become complex.

10.1. Calculating the Coupling between Subsections

The central problem is to determine the voltage on one subsection (the “field” subsection) caused by current on another subsection (the “source” subsection). It is this pairwise coupling that fills the moment matrix for every possible source–field pair of subsections.

In order to proceed, we must first assume a specific current distribution for the source subsection. We use the “rooftop” current distribution [15] (Fig. 15). In Fig. 15, the height indicates the current density over the rectangular area of the subsection. Several rooftops are overlapped to yield a piecewise linear approximation to the current in the direction of current flow (Fig. 16). Rooftops are placed side by side to yield a stepwise approximation to the actual current distribution in the direction transverse to current flow.

A given area must be subsectioned twice, once for x -directed current and a second time for y -directed current. Figure 17 shows how the x - and y -directed subsections overlap. Note that the centers of the x -directed and y -directed rooftops must be offset. This offset is required in order to allow current to flow from one rooftop to the next.

We now evaluate the V_i so that the discontinuity in tangential H field at $z=h$ equals the rooftop current at the

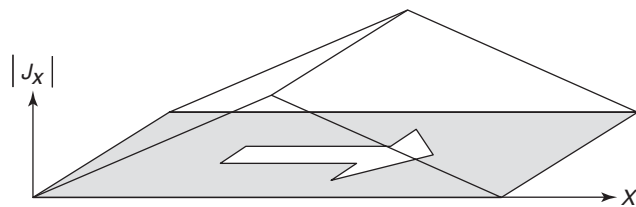


Figure 15. The vertical direction represents the current density on a rectangular subsection. Current is flowing in the x direction. This is called a “rooftop” basis function.

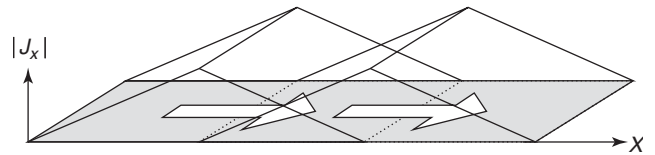


Figure 16. The total current from the sum of two overlapping rooftop basis functions yields a piecewise linear approximation to the current in the direction of the current flow.

source subsection. With \mathbf{J}_s representing the rooftop current distribution centered on the source subsection at (x_0, y_0) , we have

$$V_i = -Z_i \iint \mathbf{J}_s(x, y, x_0, y_0) \cdot \mathbf{e}_i(x, y) dx dy \quad (11)$$

$$Z_i^{-1} = Y_i^{(0)} \text{ctn}[k_{iz}^{(0)}(c-h)] - Y_i^{(1)} \text{ctn}[k_{iz}^{(1)}h] \quad (12)$$

Note that Z_i is just the impedance of the topcover and ground plane transformed by the intervening rectangular waveguide to the surface of the substrate (at $z=h$) and connected in parallel. Extension of this technique to multiple layers requires only the modification of Z_i .

When the full forms of the vector fields \mathbf{J}_s and \mathbf{e}_i are inserted above, the expression for V_i becomes complicated; however, it is simply the sine and cosine integrals. Analytic evaluation of the integrals is tedious but straightforward. The V_i are then used in Eqs. (7)–(10) to determine the tangential electric fields everywhere due to the subsection current distribution \mathbf{J}_s .

Within the framework of the method of moments, the \mathbf{J}_s (one for each subsection) represent “basis” or “expansion” functions. The total current on a circuit is a sum of all the \mathbf{J}_s . To complete the method of moments, we must select “testing” functions. Here, we choose a Galerkin technique, in that the testing functions are the same as the basis functions.

Now, given a specific source subsection with current \mathbf{J}_s impressed on it, we calculate the voltage on a specific field subsection by multiplying the tangential electric field by a rooftop testing function centered on the field subsection and integrating over the area of the field subsection. This integration is of the same form as Eq. (11). It is once more a tedious but straightforward sine and cosine integration that is performed analytically.

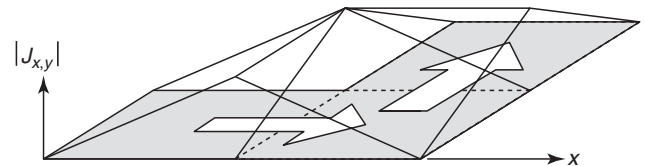


Figure 17. When a y -directed rooftop basis function overlaps half of an x -directed rooftop, current can flow around a corner. Note how the center of any x -directed rooftop must be offset from the center of any y -directed rooftop.

10.2. Solving the Moment Matrix

The process described above is repeated for every possible pair of source–field subsections. For N subsections, this fills an $N \times N$ impedance matrix, \underline{Z} . The amplitude of the current density on each subsection is stored in the $1 \times N$ vector \underline{J} and the total voltage on each subsection is stored in the $1 \times N$ vector \underline{V} . The resulting matrix equation is

$$\underline{V} = \underline{Z}\underline{J} \quad (13)$$

Typically most of the numerical effort is in inverting the \underline{Z} matrix. This operation is of order N^3 . Values of N up to $N=30,000$ can now be solved in about an hour on a 3-GHz Pentium. Prior to solving the matrix, we designate a few subsections as “port” subsections. Port subsections are subsections to which we plan to make outside circuit connections. After the matrix is solved, we have

$$\underline{J} = \underline{Y}\underline{V} \quad (14)$$

In order to meet the boundary condition of zero voltage on a conductor, we set the voltage on all nonport subsections to zero. Once we do that, we see that we don’t even need to solve for most of the \underline{Y} matrix. If solving the matrix with LU (lower/upper) decomposition, we still must do the full decomposition. However, nearly the entire back solve step is no longer needed.

With the nonport subsection voltages set to zero, we only need the portions of the \underline{Y} matrix that deal exclusively with port subsections. If there are two port subsections, the result is a 2×2 matrix. After converting current density \underline{J} to current, and possibly changing signs (when circuit theory positively directed current is of direction opposite that of EM analysis positively directed current), this is actually the \underline{Y} matrix of the circuit and is the solution to the problem.

If the current distribution is required, additional back solve effort is applied, yielding a larger portion of the \underline{Y} matrix. For example, if there are N total subsections and two of these subsections represent ports, we need either a $2 \times N$ or a $N \times 2$ portion of the \underline{Y} matrix. Then we use Eq. (14) to calculate \underline{J} for any possible port excitations.

10.3. Application of the Fast Fourier Transform

The summation of Eqs. (7)–(10) over i is actually a two-dimensional summation over rectangular waveguide mode numbers m and n for all TE and TM rectangular waveguide modes. Performing this summation repeatedly is very slow. It is performed more efficiently by making a simple trigonometric modification.

Simplifying to one dimension for illustrative purposes, one form of the required summation for a source subsection located at x_0 and a field subsection located at x_1 is

$$S = \sum_{m=0}^M C_m \cos \frac{m\pi x_0}{M} \cos \frac{m\pi x_1}{M} \quad (15)$$

For the complete summation, the summation index m goes to infinity. However, if x_0 and x_1 are restricted to integers

in the range from 0 to M and we take advantage of the periodicity of the trigonometric functions, the entire summation may be performed over the indicated summation range by appropriate modification of the C_m .

This summation is starting to look like a Fourier cosine series, except for one problem. We have the product of two cosines, instead of one. This is easily remedied by rewriting the summation as the sum of two summations:

$$2S = \sum_{m=0}^M C_m \cos \frac{m\pi(x_0 + x_1)}{M} + \sum_{m=0}^M C_m \cos \frac{m\pi(x_0 - x_1)}{M} \quad (16)$$

Both summations are now a cosine series easily and rapidly summed by specialized FFT (fast Fourier transform) algorithms. In fact a single FFT cosine transform provides results for all possible values of $x_0 + x_1$ and $x_0 - x_1$ after taking into account the periodicity in M . **AU:3**

10.4. Improving Speed and Memory Requirements

For the actual two-dimensional summation (over all TE and TM waveguide mode numbers m and n), the 2D cosine transform is performed using a 2D FFT specialized for the cosine transform. Because the C_m depend on both the basis and testing functions, an additional 2D transform is required each time a different type of source or field subsection is used. For a single layer, typically three 2D FFTs are required, one each for x/x , y/y , and x/y coupling. Depending on the specific basis and testing functions, a sine transform may be required instead.

Multiple layers are accommodated by modification of Eq. (12). The modification is such that both extremely thin and extremely thick layers have no impact on accuracy or analysis time. This also changes the C_m in the summation above, so the FFTs must be repeated once more. This approach, when completely including all coupling between all layers, easily accommodates up to several hundred layers.

As mentioned above, all metal must be subsectioned twice, once for x -directed current, and a second time for y -directed current. Further, the center of all x -directed rooftops are offset with respect to the centers of the y -directed subsections. This offset is realized by restricting subsection center coordinates to either even or odd values. This means that the 2D FFTs need be performed for only half of the values of (x_0, y_0) and (x_1, y_1) . With further specialized modification, the size of the required FFTs is cut by half in both dimensions, resulting in a $4 \times$ speed increase.

Different basis and testing functions require only the modification of V_i in Eqs. (7)–(10). Other basis functions that have been implemented in shielded analysis include vias (for conducting current vertically through layers), and triangles [16]. Triangles are used to fill in the staircase that would otherwise exist when meshing with rectangular subsections. However, when comparing analysis results with and without the use of triangles, they are found to provide an advantage only when the transmission line is one or two rectangular subsections wide. If there is symmetry along one centerline axis in the circuit,

the FFTs need to include only half of the m or n modes, resulting in a further doubling of the FFT speed.

Unshielded analysis must also calculate coupling between subsections. In this case, it is worthwhile to store certain intermediate results that depend only on the dielectric geometry. In shielded analysis, the corresponding results are always calculated when needed. This is because the FFT-based calculation of these results is so fast that storing them for later use would actually slow down the analysis.

Matrix storage reduction is also possible. The most significant reduction is realized by noting that the \underline{Z} matrix is symmetric. It is relatively simple to organize LU decomposition so that storage is required for only half the matrix, cutting memory requirements by half.

Provided extremely low-frequency data are not required, all calculations can be performed in single precision, cutting storage requirements by half again. For both single and double precision, careful attention to avoiding loss of precision when dividing by the pivot element is critical. If there is potential loss of precision, then rows and columns must be pivoted. Rigorous testing for potential matrix solve problems requires running literally thousands of cases. Checking for success in only a few cases is certain to leave hidden problems.

Intel architecture computers utilize a specialized FPU (floating-point unit) that performs all calculations using 80 bits of precision. One way to minimize numerical precision problems is to leave the result of continuing calculations (e.g., dot products) and critical numbers (e.g., the pivot element) in the FPU, thus avoiding the precision loss caused by repeated truncation to single or double precision.

11. DEEMBEDDING ELECTROMAGNETIC DATA

Whether derived from electromagnetic analysis or from measurement, high-frequency data must be properly “deembedded” if accurate results are to be obtained. In the early days of high-frequency measurement, slotted-line techniques were used. With these approaches, a standing wave is measured on a length of transmission line (e.g., rectangular waveguide). Then, with knowledge of the transmission-line characteristic impedance, the standing-wave measurements can be converted into S -parameter data.

Deembedding approaches analogous to slotted-line measurements are in use today in planar EM analyses, primarily in unshielded environment analyses. While easily implemented, such slotted-line techniques are compromised by the necessity to independently determine the transmission line characteristic impedance. This is usually done by a separate 2D cross-sectional EM analysis. Two additional sources of error are thus introduced: (1) error inherent in the cross-sectional EM analysis and (2) error due to the determination of the characteristic impedance.

This second source of error is due to the fact that the characteristic impedance must be determined from linear functionals of the cross-sectional fields. For example, voltage is determined by selecting a path from the line to

ground and integrating the electric field over that path. While unique for lossless homogeneous media, results can range over 20% or more for typical inhomogeneous microstrip media, depending on which circuit parameters (power, current, or voltage) are calculated and which path integrals are taken. This uncertainty in the “correct” value of characteristic impedance translates directly into error in the final result.

11.1. Modern Deembedding Algorithms

For modern high-frequency measurements, a device under test (DUT) is placed in a test fixture. The test fixture is typically microstrip for connections to the DUT, and either coax or ground–signal–ground (GSG) probepads for connection to the measurement equipment. In order to characterize and remove the effect of the test fixture, a series of known standards are inserted at the location of the DUT and measured. Typical standards might include a short (circuit), open (circuit), load, and through. However, both slotted-line and many modern deembedding approaches suffer from the fact that characteristic impedance must be independently determined. This is because these approaches cannot directly measure voltage at any point in the circuit.

In contrast, shielded EM analyses can determine both the voltage and current directly and unambiguously at ports. The shielded analyses use infinitesimal gap voltage sources at the edge of the substrate; a side view is shown in Fig. 18. Because the gap source is immediately adjacent to a perfect ground reference (the sidewall), the voltage is unique and precisely determined. With both port current and voltage known uniquely and unambiguously, the port input impedance is known as well.

11.2. Deembedding the Port Discontinuity

The problem with box sidewall ports is that the port input impedance also includes fringing fields excited by the port. The port fringing fields can be viewed as a port discontinuity including both series inductance and shunt capacitance. When loss is present, resistance and conductance may additionally be part of the port discontinuity. Now, the deembedding problem becomes characterization of the port discontinuity.

For the most straightforward deembedding approach, the port discontinuity is specialized to pure shunt admit-

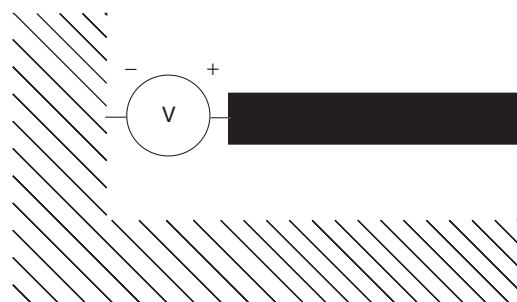


Figure 18. The shielded electromagnetic analysis uses an infinitesimal gap voltage source to excite the circuit, shown here from the side.

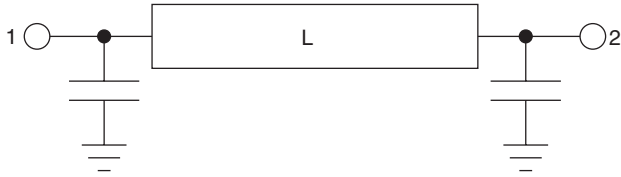


Figure 19. The L length through line shown schematically includes the port discontinuities, shown here as pure shunt capacitance. The $2L$ length through has the same port discontinuity and twice the length of the transmission line.

tance. This is realized in the EM analysis by forcing uniform voltage along the length of the gap. This disallows transverse current at the port, effectively short-circuiting out any series port inductance. That this assumption is correct is tested as part of the deembedding process, as described below.

The deembedding procedure [17] first analyzes two standards, an L length through and a $2L$ length through. A schematic of the L length through including port discontinuities is shown in Fig. 19. Alternatively, only a single $2L$ length through with an internal port at L from each box wall port may be analyzed. The data for the L length through is then obtained by exciting the $2L$ length box wall ports in an odd mode to determine Y parameters of the L length through.

To characterize the port discontinuity, first convert the data for the L and $2L$ length throughs to $ABCD$ cascading parameters. Then invert the $ABCD$ parameters for the $2L$ length through and pre- and postmultiply the result by the $ABCD$ parameters for the L length through. This leaves the only a cascade of the port discontinuity with itself; this is called the *double-port discontinuity*.

If the $2L$ line includes a third internal port, as mentioned above, then data for a single-port discontinuity can be uniquely determined [18]. However, if the port discontinuity is specialized to a pure shunt admittance, then the single-port discontinuity can be determined only on the basis of the L and $2L$ length $ABCD$ parameters.

If the port discontinuity is a pure shunt admittance, then $A=D=1.0$ and $B=0.0$. This leaves the C (of the $ABCD$ parameters) equal to the double-port discontinuity shunt admittance. The single-port discontinuity is determined by simply dividing C by 2.

If these conditions on A , B , and D are not met, then either the port discontinuity is not a pure shunt admittance, or another failure mode (discussed later) has occurred. If the more general technique described in Ref. 18 is used, then this self-consistency check is not possible and deembedding failure will not be detected until it is realized that the analysis calculated incorrect results; however, a general port discontinuity can be characterized.

Once the port discontinuity has been characterized, the DUT data are deembedded by inverting the single-port discontinuity $ABCD$ matrix and pre- and postmultiplying the DUT $ABCD$ data.

This deembedding approach is valid for both lossless and lossy situations, including metal loss, dielectric loss, and—under the restrictions noted below—radiation loss.

11.3. Determination of Characteristic Impedance

This deembedding approach has absolutely no need for knowledge of the characteristic impedance of the transmission lines, as it deals directly only with the terminal (port) voltages and currents. However, additional information can be obtained by deembedding the $ABCD$ parameters (in the same way the DUT was deembedded) of the L length through. Then, by noting the $ABCD$ parameters for an ideal TEM through line of length L

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} \cos(\beta L) & jZ_0 \sin(\beta L) \\ j\frac{\sin(\beta L)}{Z_0} & \cos(\beta L) \end{bmatrix} \quad (17)$$

we can then determine [19] the “TEM equivalent” characteristic impedance Z_0 and wavenumber β of the L length line by equating (17) with the calculated $ABCD$ parameters of the L length through. We emphasize that this determination of the characteristic impedance is not required in any way for the actual deembedding process itself.

Once the deembedded $ABCD$ parameters of the L length through have been obtained, they can be used (by inverting them and then pre- or postmultiplying (as desired) the DUT $ABCD$ parameters) to shift the reference plane of the DUT data into the box. This is sometimes done to remove the port connecting transmission lines.

11.4. Deembedding Failure Mechanisms

This deembedding approach fails if the through standards allow propagation of more than one transmission-line mode. This is easily detected in the double-port discontinuity data as A , B , and D are different from the values stated above. In addition, the L length line must be long enough that the port fringing fields on each end do not interact with each other. A length of at least one substrate thickness, and preferably two, is usually all that is required. If this condition is not met, then once more, A , B , and D are different from the above-stated values.

An additional failure mechanism is sometimes exhibited when radiation is allowed. Shielded analyses approximate radiation by making the topcover resistive, usually 377Ω per square. The box conducting sidewalls form a rectangular waveguide. The topcover must be positioned far enough above the DUT that it is not involved in fringing fields (i.e., near field) of the DUT. Placing the topcover above the DUT by about the size of the DUT is usually sufficient.

A radiation related deembedding failure mode is seen when the sidewalls of the box containing the L length through form a cutoff rectangular waveguide but the sidewalls of the $2L$ length through form a propagating waveguide. Thus, the L length through does not see radiation loss, but the $2L$ length through does. The error introduced is typically small because the radiation loss for a simple through line is typically small. However, careful inspection of the S parameters of an otherwise lossless passive structure can in this case actually show a small gain. If high-accuracy S parameters including radiation loss are

important, then one should take care to make sure that the box containing the L length through is large enough to allow propagation up to the topcover. This failure mode is not detected by inspection of A , B , and D as in other failure modes.

A final failure mechanism occurs if one of the standards excites a box resonance. This resonance appears in the deembedded data and also causes the values of A , B , and D for the port discontinuity to differ from the values stated above.

For typical high-frequency applications, the port discontinuity capacitance is a few tenths of a picofarad. Occasionally, such a port discontinuity is unimportant for a specific application and deembedding can be skipped.

11.5. Deembedding Multiple Coupled Ports

In the case of multiple coupled ports, A , B , and D of the $ABCD$ matrices above all become, themselves, matrices. For example, if there are two ports on one side of the box, both of which are to be deembedded, the L and $2L$ length throughs each become a pair of coupled lines and each element of the $ABCD$ matrix becomes a 2×2 matrix.

The approach is fully valid, no matter how many ports and no matter how tightly coupled they are. Analyses with several hundred ports, as tightly coupled as desired, are easily performed. The very high accuracy of this deembedding technique for tightly coupled ports is critical in the success of divide-and-conquer analysis strategies, as described above.

12. CONCLUSION

The field of applied high-frequency numerical electromagnetics began in the 1990s and has reached an advanced state of development. Today, numerical electromagnetics is considered a necessary part of nearly any high-frequency design. The designer can now complete numerous design iterations in days, or even in hours, that previously would have taken weeks or months. Once the design is complete, success on first fabrication is now regularly enjoyed.

Future advancements in this field are likely to be outside the realm of Maxwell's equations. Specifically, interoperability with design frameworks, faster computers, and a maturation of the generally accepted RF design cycle are all likely to see activity in the near future. Numerical electromagnetics promises to play an important role in the future of both commercial and military high-frequency design.

BIBLIOGRAPHY

1. K. Breitfelder and M. Geselowitz, IEEE Virtual Museum, Microwave Featured Exhibit, Institute of Electrical and Electronics Engineers, April 2003 (online) <http://www.ieee-virtual-museum.org/> (Feb. 4, 2004).
2. J. Estes, R. Lucero, A. Pavio, and L. Zhao, Optimization of multi-layer circuits using companion models and space mapping, *MTT Int. Microwave Symp. Digest*, Workshop WFA, June, 2003.

3. J. W. Bandler, R. M. Biernacki, S. H. Chen, P. A. Grobelny, and R. H. Hemmers, Space mapping technique for electromagnetic optimization, *IEEE Trans. Microwave Theory Tech.* **42**:2536–2544 (Dec. 1994).
4. J. C. Rautio, An ultra-high precision benchmark for validation of planar electromagnetic analyses, *IEEE Trans. Microwave Theory Tech.* **42**(11):2046–2050 (Nov. 1994).
5. R. C. Booton, *Computational Methods for Electromagnetics and Microwaves*, Wiley, New York, 1992.
6. W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes*, Cambridge Univ. Press, Cambridge, UK, 1986.
7. E. H. Lenzing and J. C. Rautio, A model for discretization error in electromagnetic analysis of capacitors, *IEEE Trans. Microwave Theory Tech.* **46**(2):162–166 (Feb. 1998).
8. S. B. Cohn, Problems in strip transmission lines, *IRE Trans. MTT* **3**:119–126 (March 1955).
9. J. C. Rautio and V. Demir, Microstrip conductor loss models for electromagnetic analysis, *IEEE Trans. Microwave Theory Tech.* **51**(3):915–921 (March 2003).
10. M. Mattes and J. R. Mosig, A novel adaptive sampling algorithm based on the survival-of-the-fittest principle of genetic algorithms, *IEEE Trans. Microwave Theory Tech.* **52**(1):265–275 (Jan. 2004).
11. J. Rautio, A conformal mesh for efficient planar electromagnetic analysis, *IEEE Trans. Microwave Theory Tech.* **52**(1):257–264 (Jan. 2004).
12. R. F. Harrington, *Field Computation by Moment Methods*, Macmillan, New York, 1968; reprinted IEEE, 1993.
13. J. C. Rautio, *A Time-Harmonic Electromagnetic Analysis of Shielded Microstrip Circuits*, Ph.D. dissertation, Syracuse Univ., Syracuse, NY, 1986.
14. J. C. Rautio and R. F. Harrington, An electromagnetic time-harmonic analysis of shielded microstrip circuits, *IEEE Trans. Microwave Theory Tech.* **MTT-35**:726–730 (Aug. 1987).
15. A. W. Glisson and D. R. Wilton, Simple and efficient numerical methods for problems of electromagnetic radiation and scattering from surfaces, *IEEE Trans. Anten. Propag.* **AP-28**:593–603 (1980).
16. J. C. Rautio, Triangle cells in an electromagnetic analysis of arbitrary microstrip circuits, *MTT Int. Microwave Symp. Digest*, June 1990, pp. 701–704.
17. J. C. Rautio, A deembedding algorithm for electromagnetics, *Int. J. Microwave Millimeter-Wave Comput.-Aided Eng.* **1**(3):282–287 (July 1991).
18. L. Zhu and K. Wu, Short-open calibration technique for field theory-based parameter extraction of lumped elements of planar integrated circuits, *IEEE Trans. Microwave Theory Tech.* **50**(8):1861–1869 (Aug. 2002).
19. J. C. Rautio, A new definition of characteristic impedance, *MTT Int. Symp. Digest*, June 1991, pp. 761–764.

FURTHER READING

- D. G. Swanson, Jr. and W. J. R. Hofer, *Microwave Circuit Modeling Using Electromagnetic Field Simulation*, Artech House, Norwood, MA, 2003.
- J. S. Hong and M. J. Lancaster, *Microstrip Filters for RF/Microwave Applications*, Wiley, New York, 2001.
- C. W. Sayre, *Complete Wireless Design*, McGraw-Hill Telecom, New York, 2001.
- I. D. Robertson, *MMIC Design*, IEEE, New York, 1995.

Author Query Form



Title: Encyclopedia of RF and Microwave Engineering

Article/Number: Applied Numerical Electromagnetic Analysis for Planar High-Frequency Circuits/040

Dear Author,

During the preparation of your manuscript for typesetting some questions have arisen. These are listed below. Please check your typeset proof carefully and mark any corrections in the margin of the proof or compile them as a separate list. This form should then be returned with your marked proof/list of corrections to John Wiley.

This form should then be returned with your marked proof/list of corrections to Kris Parrish, Production Editor, Scientific, Technical, and Medical division, John Wiley & Sons, Inc. 111, River Street, Hoboken, NJ 07030-5774, Mail Stop 8-01, Tel: 201-748-8620, Fax: 201-748-8888, E-mail: kparrish@wiley.com

Queries and/or remarks

| | |
|------|---|
| AU:1 | This does not have to be noted – many authors cite the same figure multiple times; this is totally acceptable. Also please note that figures are numbered consecutively throughout article, not by section (in case this article is revised later). |
| AU:2 | Note to Editor is deleted on ms page 28. Please check |
| AU:3 | Are numerators correctly delineated (marked for buildup)? |
| AU:4 | (17) insertion is OK? |