HEWLETT PACKARD

Modeling and Analysis of High Speed VLSI Interconnects for Digital Applications

Al Barber and Ken Lee

HPL–92–03
January, 1992

packaging, connecting elements, high performance system design, high speed technology, VLSI design, interconnect, computer, SPICE

Increasing speed and density of VLSI technologies place increasing demands on packaging and interconnection structures to support them. We describe here a methodology for simulating the performance of electrical interconnects using SPICE and describe a library of models to support it. We then use this methodology to examine how the physical design choices of materials and dimensions affect the electrical results of crosstalk, ground bounce and reflections and, ultimately, noise margin. Real design examples are examined and issues relevant to sub-nanosecond signaling explored.

Copyright ©1992 Hewlett-Packard Company

1 Introduction

As both the speed and density of VLSI technologies and modern computer architectures continue to increase so do the demands on the packaging and interconnection structures to support them. Many believe that even today engineering compromises in packaging and interchip signaling place the greatest limitation on system performance. Designs using a combination of very high speed BiCMOS processes and interconnect hungry VLSI chips strain the packaging subsystem even further. The work described here has two purposes. The first is to provide a methodology for analyzing the bewildering array of potential combinations and answering the bottom line question: Will it work? The second is to investigate some technology choices available for high speed electrical signaling and packaging.

The emphasis here is on electrical signaling, and therefore package interconnections and, to a lesser extent, drivers and receivers. As overall measures of the quality of electrical signaling we chose two: noise margin and delay. Many design elements contributing to noise margin degradation were studied, including crosstalk, ground bounce, terminations, reflections and power supply inductance. Delay is influenced almost totally by device and circuit design of drivers and receivers and by the dielectric constant of the interconnect medium. With so many design elements influencing success, and with so many available technology choices affecting those design elements, it is important to have the modeling capability to predict the quality of any set of choices.

The sections that follow describe a methodology for simulating the performance of electrical interconnects using SPICE and describe the library of models needed to support it. This methodology is then used to independently examine the effects of design choices on crosstalk, ground bounce, terminations, reflections and power supply inductance. Then this methodology is applied to three examples of complete interconnection systems using currently available technology which we show will meet specificspecific technical requirements.

2 Methodology

Before settling on SPICE as our simulation tool we considered a number of other choices with the following goals in mind:

1. The tool must simulate nonlinear circuits (drivers and receivers).
2. It should have existing models for many of our components, including semiconductors and interconnects.
3. It should handle transmission lines well, including electromagnetic coupling, dispersion, and skin effect since our circuits are operating well into the microwave domain.
4. It should be computationally efficient; large numbers of drivers and receivers can interact on a single chip pair.

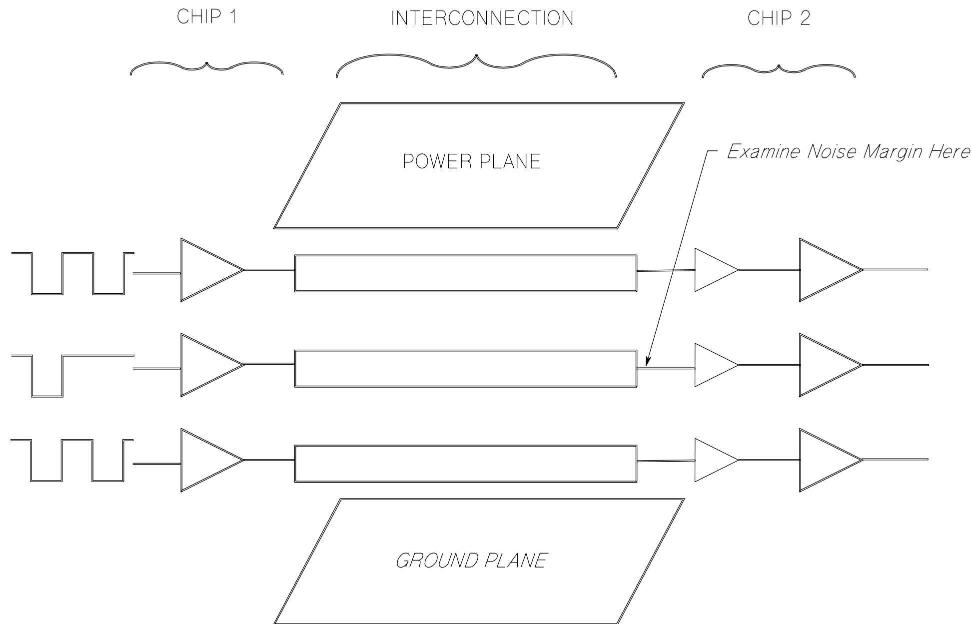

Figure 1. Drivers and receivers connected by transmission lines.

The best alternative to SPICE would be one of the microwave analysis packages, such as HP's MDS (Microwave Design System). They all satisfy goals 1, 3, and 4 very well and goal 2 well except for semiconductors. Their semiconductor models exist, but are incompatible with the SPICE models that are routinely and laboriously generated for our devices. SPICE, on the other hand, has all the semiconductor models well supported but fails to satisfy goals 3 and 4. We chose SPICE, judging it easier to overcome its shortcomings than to generate semiconductor models for the microwave simulators.

Section 3 of this report details how we taught SPICE about transmission lines and verified their accuracy. In addition it describes the sizeable library of components from which we assembled simulations.

Figure 1 shows, in diagram form, the extent of the simulations. The drivers and receivers are either CMOS or BiCMOS circuits assembled from the custom device models in SPICE. The interconnection portion may consist of multiple coupled transmission lines in multiple technologies with vias or solder bumps connecting them. Again, they are assembled from the library of components. We chose to simulate a maximum of three channels in order to keep the computation time reasonable. To do so with accuracy we can only examine the center of the three channels in order to include coupling from the adjacent channels. We assume that channels beyond the nearest neighbor do not influence the channel under examination. That this is a reasonable, though far from perfect, assumption is supported by the data shown in Figure 2, generated with the microwave analysis

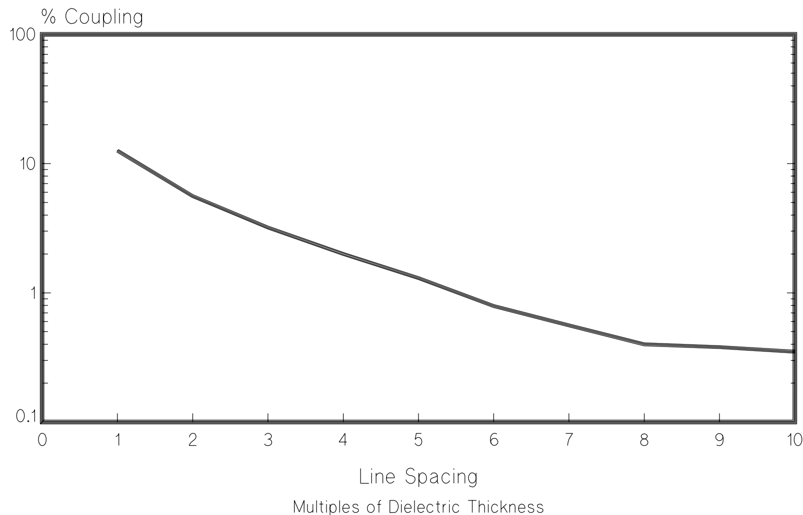

Figure 2. Voltage coupling vs. transmission line spacing.

software mentioned above. Coupled energy between lines falls off rapidly with increasing pitch. Typical tight spacings would be around two dielectric thicknesses for the nearest neighbor and five for the next nearest, or about 1% coupling from the ignored line. Also, we use 1:1 power and ground:signal ratio, which assures that no channel shares a power or ground connection with other than its nearest neighbor. These values are typical for designs operating at the speeds considered here. Lower ratios would primarily affect power and ground transients and are considered briefly in the chapter on ground bounce.

To enable comparisons of alternative design approaches, two figures of merit were chosen. They are delay ("How long does it take?") and noise margin ("Does the message get through?"). Delay is simply the elapsed time between the half-amplitude points at each end of the signaling link under consideration and is discussed in Section 8. Noise margin is more subject to interpretation and is described here.

For our purposes, noise margin is defined at the input to a receiver, as shown in Figure 3. Nominally, for simple inverters, the output of the receiver will switch between high and low states when its input voltage traverses the midpoint between V_{dd} and V_{ss}. More accurately, there is a narrow band of voltage around this midpoint in which the dynamic gain is very high and the output is effectively indeterminate. We take this band to be between the slope=1 points of the receiver transfer curve, V_{IL} and V_{IH}. It is about 200 mV wide for the CMOS receivers studied here and is shown cross-hatched in Figure 3. For simple CMOS inverters used as drivers, also nominally, the output will switch between V_{ss} and V_{dd}. For a noise free system those voltages will arrive uncorrupted at the receiver so that nominally $V_{OL}=V_{ss}$ and $V_{OH}=V_{dd}$. Noise margin is commonly defined [WEST85] as:

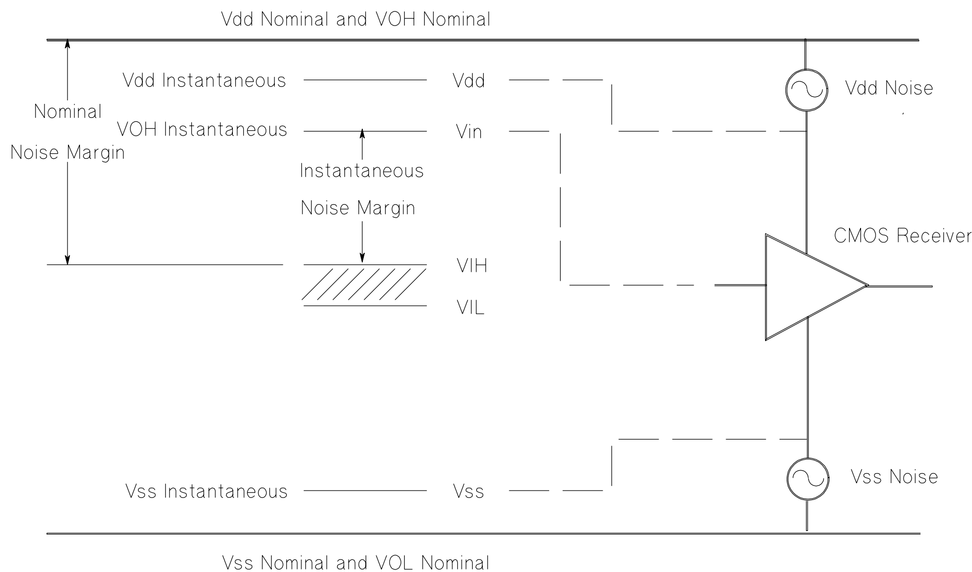

Figure 3. Noise margin calculation method.

Output High: $V_{OH} - V_{IH}$

Output Low: $V_{IL} - V_{OL}$

Here, we take this as the nominal noise margin and use nominal, that is noise-free, values for the V's, as shown in Figure 3. There are two broad categories of effects which can degrade noise margins below these nominal values. First are those that operate at d.c. Especially important are processing and temperature variations. These can be global as well as differences between driver and receiver, which are commonly on different chips. These were not studied by us, but [WEST85] discusses their effects. Second are those that operate at high frequencies and are the focus of this discussion. Crosstalk, reflections, and power and ground noise all corrupt the driver output signal before it arrives at the receiver. Also, power and ground noise cause the threshold of the receiver and thus V_{IL} and V_{IH} to move. In our analysis these high frequency effects are accounted for and an instantaneous noise margin is defined identically to the nominal noise margin above, except that instantaneous values for the voltages are used. Instantaneous noise margin is thus a waveform rather than a value. The minimum value of that noise margin is what interests us, so we define actual noise margin as:

Output High: $(V_{OH} - V_{IH})|_{min}$

Output Low: $(V_{IL} - V_{OL})|_{min}$

where instantaneous values are used for all the voltages and the minimum is taken over the time period of interest as indicated in the example below.

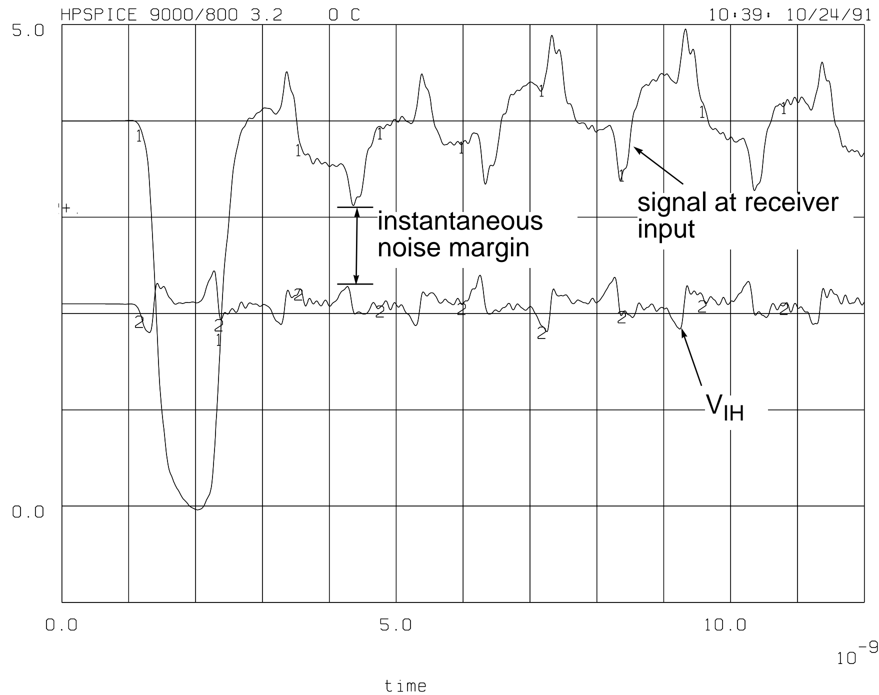

Figure 4. Instantaneous noise margin example.

Scripts have been written to extract these measures from SPICE simulations; examples of their use are shown in Figures 4 and 5. The signal at the receiver input in Figure 4 was generated by sending a single pulse down the transmission line from the driver. The noise after the pulse is received is caused by multiple reflections of the initial pulse, by coupling from adjacent active lines, and by power supply ringing. The trace labeled "V_{IH}", positioned 100 mV above the midpoint between V_{dd} and V_{ss}, is the top of the indeterminate zone and is noisy due to power supply and ground bounce. The instantaneous noise margin, as indicated, is the difference between the two. Figure 5 shows instantaneous noise margin derived from the same receiver input signal.

In general, noise margin is different for high and low signals because drivers and receivers have no symmetry in their operation with respect to voltages and currents. The rest of the interconnection system does have this symmetry, however. That is, for every effect of crosstalk, power or ground bounce, or reflection as it affects a high level signal, a complementary situation exists to produce the identical effect on a low level signal. Since this paper does not explore the subtleties of driver and receiver design, results will generally be shown only for a high or low signal, not both.

The value of this figure of merit is in providing a consistent, meaningful measure of performance as some design parameter is varied. It will be used throughout this

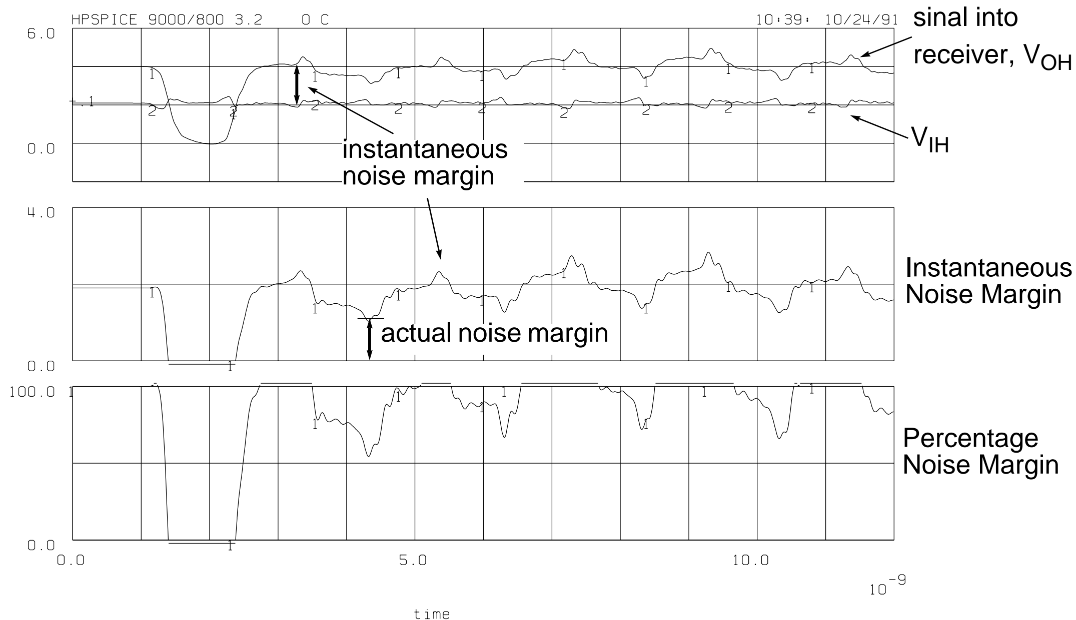

Figure 5. Instantaneous and percentage noise margin extracted from received signals.

paper in the manner shown in Figure 1. Three drivers are connected to three receivers through some interconnect structure. A single pulse is transmitted on the center line and continuous pulses are transmitted on the adjacent lines. The noise margin is examined at the receiver input and, through multiple simulations, the system parameter is varied. Then the value of noise margin can be plotted vs. the system parameter value.

3 Library of Component Models

In order to simulate the behavior of an interconnection system, the electrical model for each interconnection element has to be built. In this study, the interconnection elements between an output driver pad and input receiver pad were modeled with two types of interconnect elements. The first type is a lumped element model for discontinuities such as solder bumps and vias. The second type is the transmission line model for line structures such as striplines and microstrips which are used in chip carriers, MCM substrates, printed circuit boards, and impedance controlled connectors. In this section, these two types of interconnection models are presented and the accuracy of the transmission line model is verified. Also, the circuit models of the output drivers and input receivers used in simulations are presented.

Figure 6 shows a representative view of the studied interconnection. A signal travels from an output driver pad to a receiver pad through a solder bump, short transmission line on the carrier, via through the carrier, solder bump to the substrate,

transmission line on the substrate, and the same connection through the receiver carrier in the reverse order. The solder bumps and the vias through the substrate are modeled with their lumped equivalent circuits because their sizes(100-1000 μm) are small compared with the wavelengths of the frequency components of the signal, which are longer than 5 cm. On the other hand, the transmission lines are modeled with distributed models because their sizes are comparable or larger than the signal wavelengths.

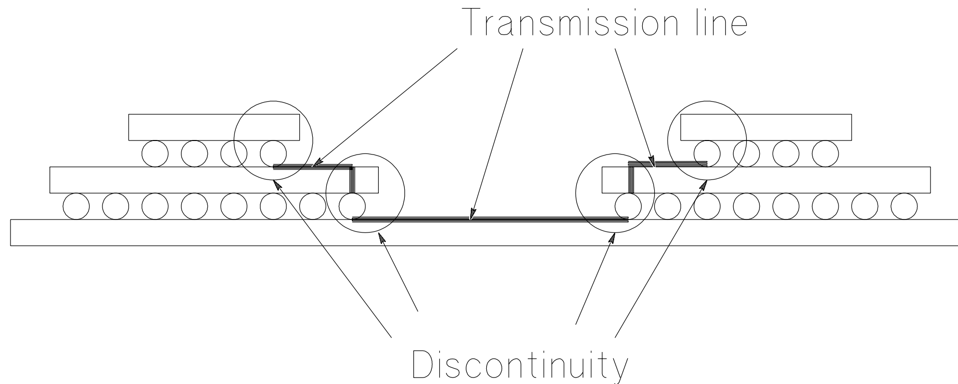

Figure 6. Interconnection models used in simulations.

The circuit model for the interconnect in Figure 6 is shown in Figure 7. The values of the electrical components were calculated using the electrical modeling programs: FCAP2[LEE88a], FCAP3[LEE88b], IND3[LEE89], and ICM[LIU90]. FCAP2 is a two-dimensional Poisson equation solver and is used to calculate the capacitance and inductance matrices of the coupled transmission lines. FCAP3 is a three-dimensional Poisson equation solver and is used to calculate the capacitance of the discontinuities. IND3 is a three-dimensional inductance program and is used to calculate the inductance of the discontinuities. These three programs can be used as stand-alone programs or can be called through ICM which has a library of commonly used interconnect geometries.

The transmission line sections of the model which are represented by boxes in Figure 7 consist of the cascaded connection of coupled RLC ladder circuits. The SPICE transmission line model cannot be used in this study because the model does not handle the coupling between the multiple transmission lines and the resistive loss of the lines which is significant in the thin film MCMs. The ladder circuit approach is accurate for transmission lines with a homogeneous dielectric (e.g. stripline with uniform dielectric) as long as the length of each section is much smaller than the wavelength of the signal. In this study, the length of each ladder section was limited to 1/10 of the highest frequency of interest. However, for transmission lines with an inhomogeneous dielectric medium (e.g. microstrip line and stripline with multiple dielectrics) the ladder circuit approach is acceptable only if the quasi-TEM approximation is valid.

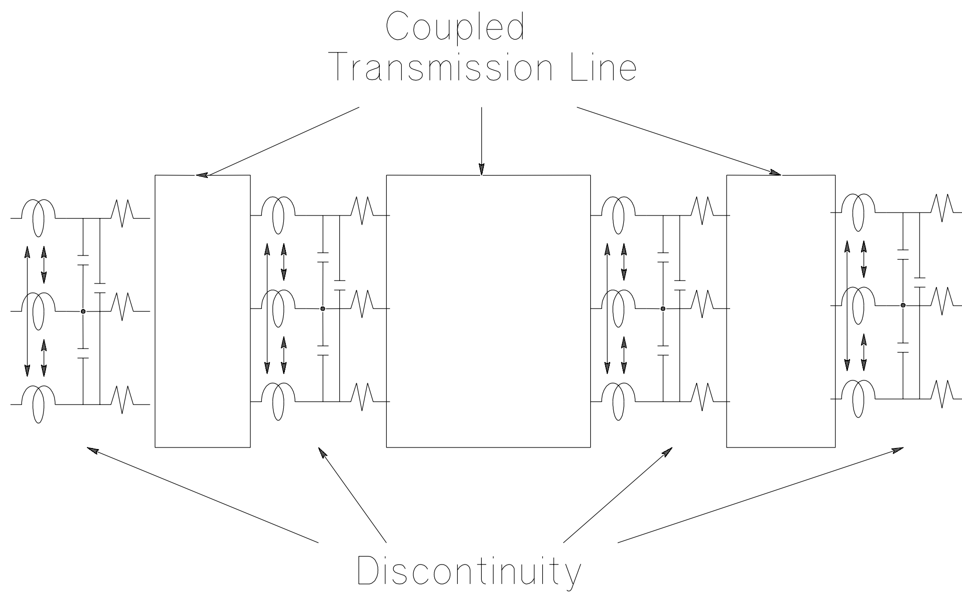

Figure 7. Circuit model for the interconnection in Figure 6.

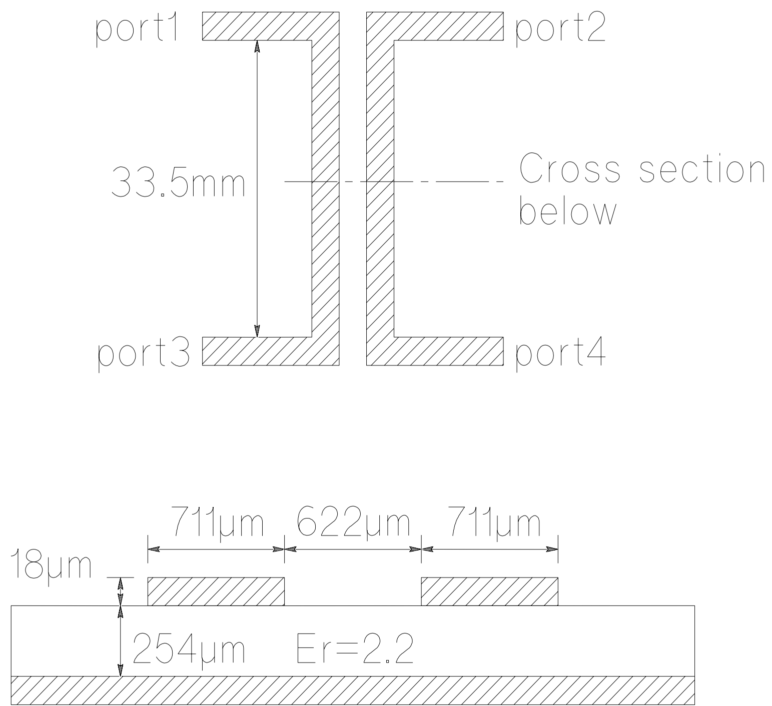

Figure 8. Duroid microstrip coupler.

In order to confirm the accuracy of the SPICE simulation with the ladder circuits, the SPICE simulation result was compared with the simulation result from MDS (Microwave Design System) and a measurement for a duroid microstrip coupler. MDS is an HP software product which is capable of solving the frequency response (scattering parameters) of microwave circuits. It has a library of microwave structures such as striplines and microstrips. The dimensions of the duroid coupler used in the measurement is shown in Figure 8. The signal propagation velocity in duroid is 20.2 cm/nsec. A spice model was built with 20-section ladder circuit and each section is 0.17 cm long. If the "1/10 of wavelength rule" is valid, this model should be valid up to about 10 GHz. Figure 9 is scattering parameter, S31, data from MDS and SPICE simulations. It shows for the directly transmitted signal SPICE has only 0.3 dB difference from the MDS result which is well calibrated with experimental data. Figure 10 is S41 data which is the forward crosstalk. The SPICE result is about 2 dB higher at 10 GHz. The most difference occurs with S21, the backward crosstalk. SPICE overestimates the backward crosstalk by about 4 dB. The measurement data with an HP8510 network analyzer is also shown and it matches well with the MDS result. The deviation from the MDS result between 7 GHz and 9 GHz is due to the parasitic elements of the measurement. It was found that if the length of each section is reduced to half, so is the difference between the SPICE and MDS results. However, the accuracy with a 20 section model is tolerable to this study since the 4 dB difference is only in the high frequency range and does not affect the time domain waveforms of signals too much. Figure 12 shows the time domain waveforms of the backward crosstalk which are simulated with 20-section and 40-section models of the duroid coupler. Even though the magnitude of high frequency ringing due to the discreteness of the models is smaller for the 40-section model, the overall waveforms are almost identical. Also, the crosstalk level itself is very low (below 20 dB) and the SPICE results always overestimate the crosstalk.

Using the SPICE ladder circuit model, the capacitive and inductive coupling in the multiple transmission lines can be included in the simulation and the resistive loss can be included with series resistance in the ladder circuit. One drawback with this approach is that if the transmission line becomes longer and/or many lines have to be considered simultaneously, the SPICE run time increases substantially. There are some dedicated programs to handle this problem. However, those programs usually do not have the active circuit elements which are essential to this study. Since SPICE is the only tool which has the device models for the HP's IC process, it was used throughout this study even though it is not the fastest tool for the transmission line simulation.

The circuit models for the substrates from different MCM vendors such as Kyocera and Alcoa were built depending on the design rules of the vendors. An electrical model was also built for the impedance-controlled connector from Beta-Phase Connector using a transmission line model and capacitive discontinuities at the connecting points. Most vendors have fast evolving sets of design rules as their technology advances. In this report, different combinations of these models are used for

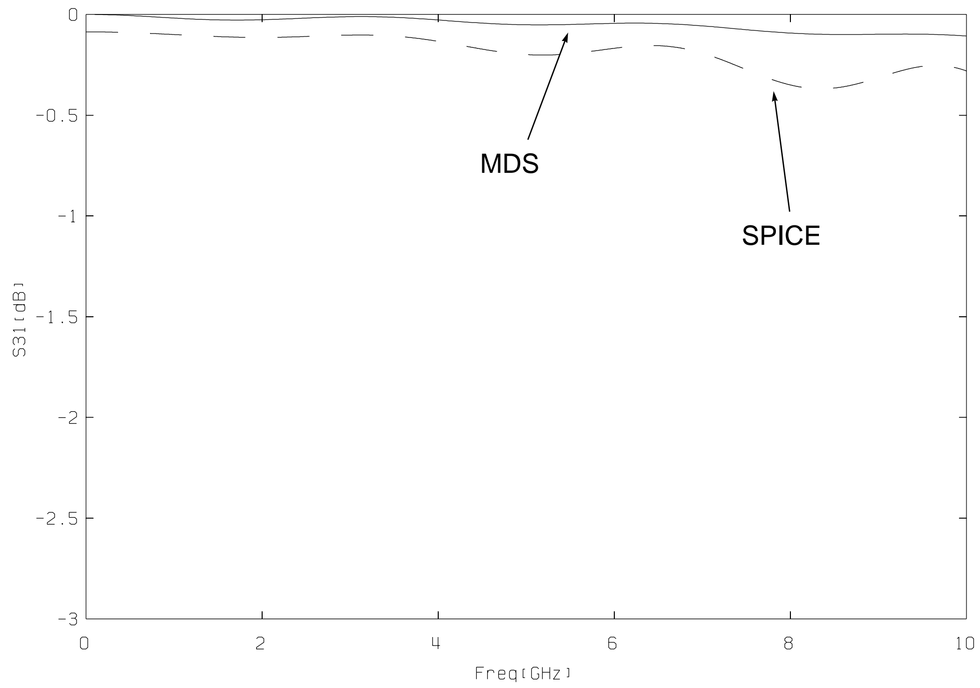

Figure 9. S31 parameter for the duroid coupler.

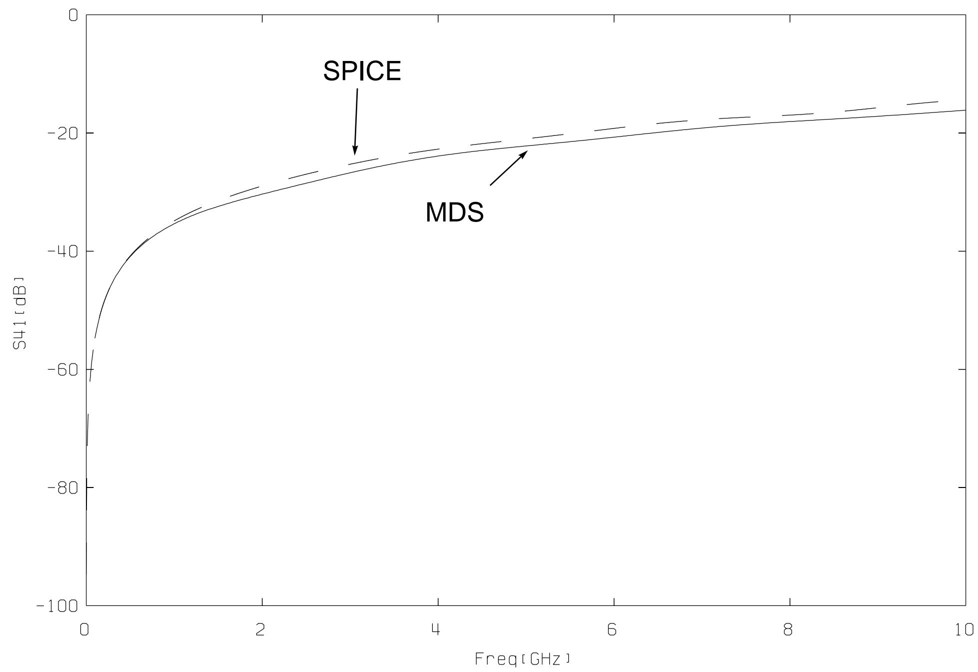

Figure 10. S41 parameter for the duroid coupler.

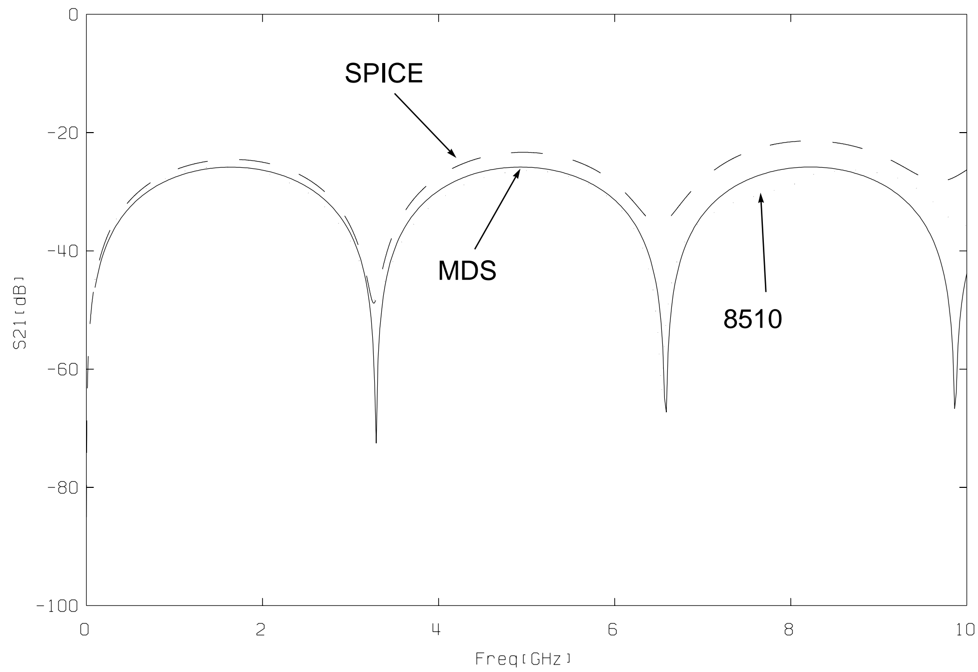

Figure 11. S21 parameter for the duroid coupler.

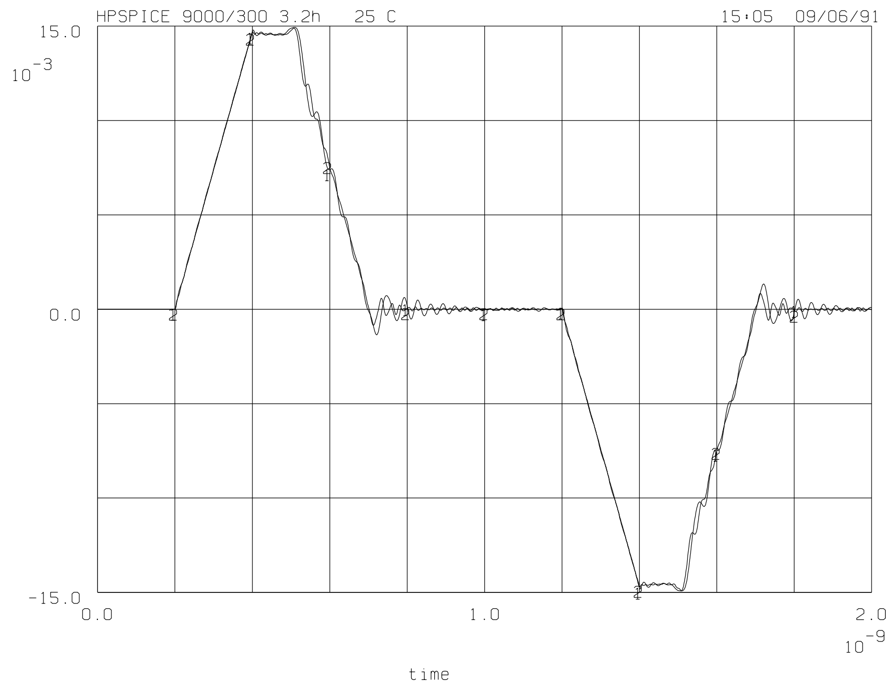

Figure 12. Time domain waveforms of backward crosstalk with 20-section(waveform with higher ringing) and 40-section models.

simulations. However, the relationship between the geometrical design rule and electrical performance (crosstalk, ground bounce, reflection, and attenuation) is generic across the different vendors.

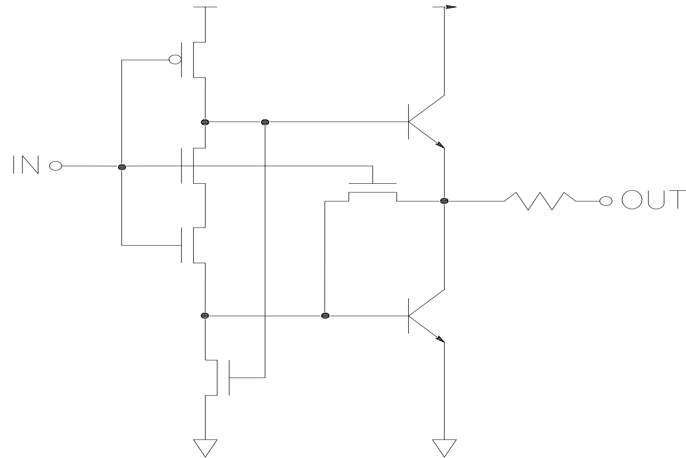

Figure 13. BiCMOS driver.

For the active elements, two types of driver circuits were used: CMOS and BiCMOS versions of pushpull drivers. The CMOS driver is a simple four-stage inverter chain and the BiCMOS driver is shown in Figure 13. The drivers are series terminated with a resistor (0-40 Ω) at the driver side. The matching receiver is a CMOS inverter for either driver. These circuits are implemented with the HP half-micron BiCMOS process.

The following is the list of interconnect and active circuit models used in simulations.

discontinuity models

sub.beta_up # Beta Phase connector - Connection from substrate to module

sub.beta_down # mirror image of beta_up

sub.cer_chip # ceramic to chip model using solder bumps

sub.chip_cer # chip to ceramic model using solder bumps

sub.cer_pcb # ceramic to PCB transition model

sub.pcb_cer # PCB to ceramic transition model

transmission line models

sub.carrier_cerZ135 # 135Ω fineline ceramic lines - micro strip

sub.carrier_polyZ60 # 60Ω polyimide lines - stripline

sub.kyoZ26 # 26Ω Kyocera polyimide - stripline

sub.kyoZ42 # 42Ω Kyocera polyimide - stripline

sub.kyoZ60 # 60Ω Kyocera polyimide - stripline

sub.pcbZ56 # 56Ω PCB lines - stripline

sub.pcbZ60 # 60Ω PCB lines - stripline

sub.subs_cerZ100 # 100 Ω ceramic lines - microstrip

sub.subs_cerZ70 # 70 Ω ceramic lines - microstrip

sub.subs_polyZ60 # 60 Ω polyimide lines - stripline

active circuit models
sub.cmos_drivers # four-stage inverter chain
sub.bicmos_drivers # BiCMOS driver
sub.receiver # CMOS inverter

4 Crosstalk

Crosstalk is the electromagnetic interference between signal lines. The noise margin of a victim line is decreased when the coupling noise from an adjacent line is injected to the victim line. Furthermore, a false signal can be received if the coupling noise is excessive. Crosstalk can be reduced by increasing the space between the signal lines or by placing a shield between the lines. For the signal lines in MCMs the first approach is usually used to limit the crosstalk even though the second approach is not impossible to use. This section mainly focuses on the first approach with some comments on the second at the end.

Since increasing the spacing between signal lines reduces the packing density of an MCM, a careful balance between the spacing and coupling noise has to be achieved. The coupling is determined by the geometry and materials of the MCM and as a result it is closely related to the characteristic impedance of the signal lines. The coupling and characteristic impedance were calculated for different MCM structures using the numerical tools mentioned in Section 2 and tradeoffs are studied in this section.

There are basically two types of transmission lines used in MCMs: stripline and microstrip. Figure 14 shows a typical stripline cross section and the dimensions and dielectric constant. Even though only two lines were included for simplicity of analysis, the analysis is valid for multiple lines. Especially for the three line case which is used through out this report, if the center line is the victim line and two neighboring lines are switching simultaneously, the crosstalk can be calculated by

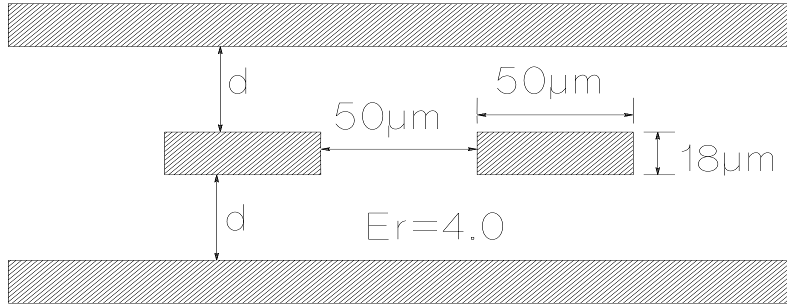

Figure 14. Stripline structure.

doubling the crosstalk from a single neighbor in the two line case. There are several design variables such as line width, line space, and dielectric thickness, and these variables make a complicated multi-dimensional electrical design space. In this analysis, we took different cross sections of the design space and examined the effects of the geometrical and material variables. First, the characteristic impedance of the coupled strip line in Figure 14 with respect to the dielectric thickness was calculated and plotted in Figure 15.

Since there are two signal lines, two modes of propagation exist. The even and odd mode impedances are calculated by

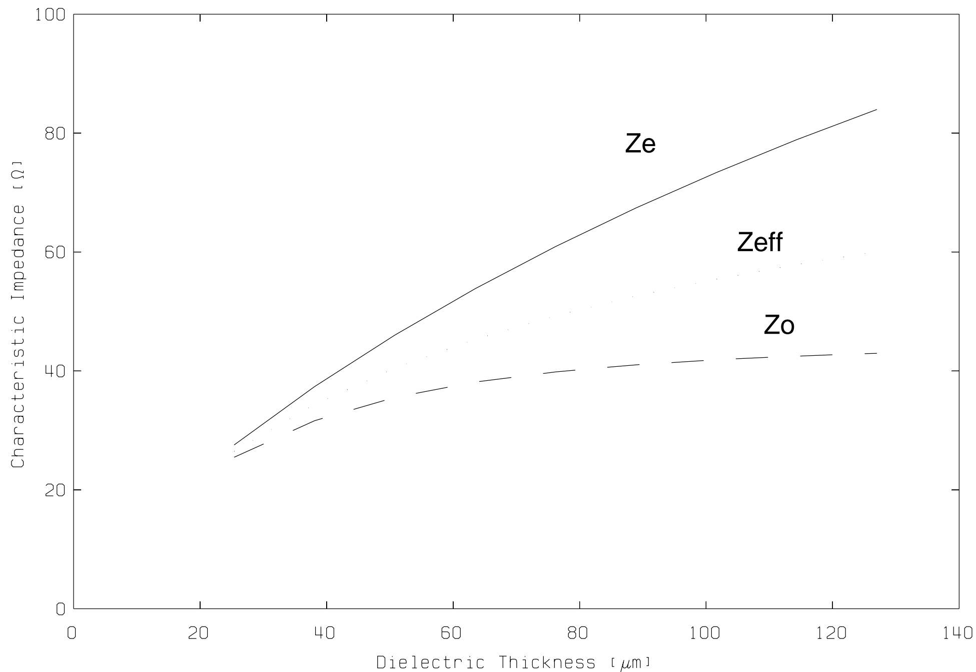

Figure 15. Stripline impedances v.s. dielectric thickness.

$$Ze = \sqrt{\frac{(L11 + L12)}{(C11 + C12)}}$$

$$Zo = \sqrt{\frac{(L11 - L12)}{(C11 - C12)}}$$

The capacitance and inductance matrices are calculated by FCAP2. For example, the values for the dielectric thickness of 50 μm are

C11 = 1.67e-16 (F/μm)

C12 = -2.14e-17 (F/μm)

L11 = 2.73e-13 (H/μm)

L12 = 3.50e-14 (H/μm)

C12 is a negative number because it is an off-diagonal element in the [C] matrix in the equation, [Q] = [C] [V]. Note that the ratios |C12|/C11 and L12/L11 are identical(0.128). This is because the dielectric medium is homogeneous for stripline. The even mode impedance is greater than odd mode impedance and the difference increases as the dielectric thickness increases. As shown later in this section, this difference is directly related to the crosstalk between the lines. The effective characteristic impedance is also shown in Figure 15 and is defined by

$$Zeff = \sqrt{ZeZo}$$

It is the averaged characteristic impedance of the even and odd mode signals and is a useful parameter to compare the impedance of different structures with multiple lines because it is not practical to compare the individual characteristic impedance for each mode. Coupled microstrip lines with the same line width and space as the stripline in Figure 14 are shown in Figure 16and the characteristic impedances are plotted in Figure 17. The characteristic impedance of microstrip is higher than that of stripline for the same dielectric thickness since it has only one ground plane, thus less capacitance.

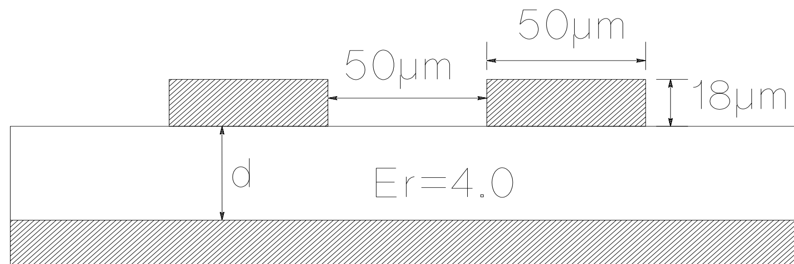

Figure 16. Microstrip structure.

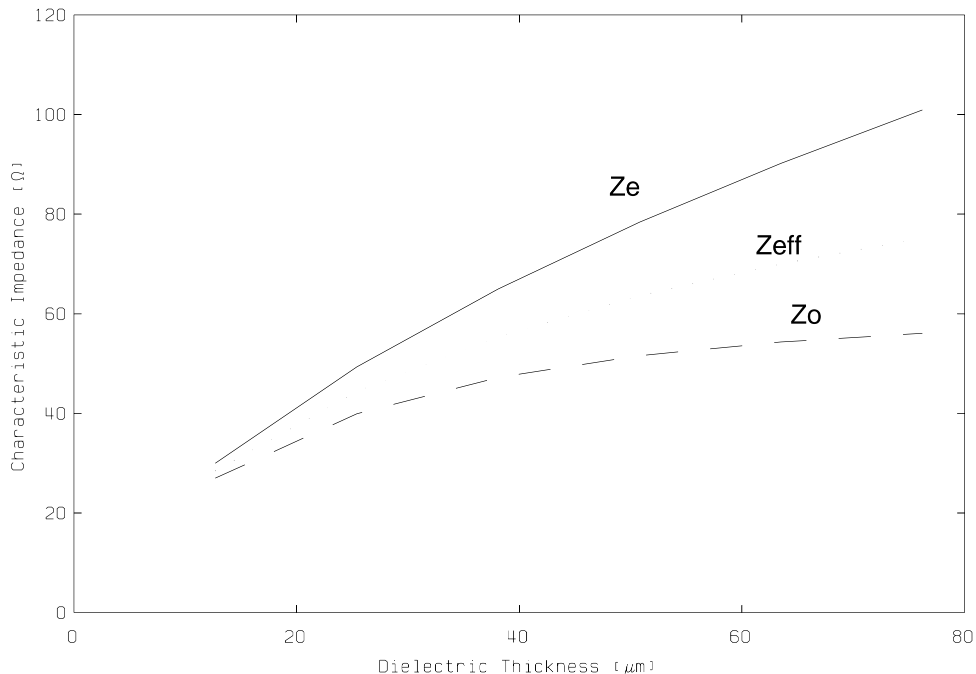

Figure 17. Microstrip impedances v.s. dielectric thickness.

In order to quantitatively compare the crosstalk of different transmission lines, we can use the crosstalk coefficient which is defined as

Crosstalk Coefficient = (Ze-Zo)/(Ze+Zo).

The crosstalk coefficients of the strip line and microstrip in Figure 14 and Figure 15 are plotted with respect to the effective characteristic impedance in Figure 18. It shows that for the lines with the impedance greater than 30 Ω, the stripline has higher crosstalk coefficient than the microstrip line. The actual crosstalk amounts at the near or far end of a victim line are different depending on the termination conditions, signal risetime, and line length. However, if the lines are terminated well with Zeff as shown in Figure 19 and the signal risetime is shorter than 2×(signal propagation time to the far end), the backward crosstalk measured at the near end (Vn1) is related to the crosstalk coefficient such that

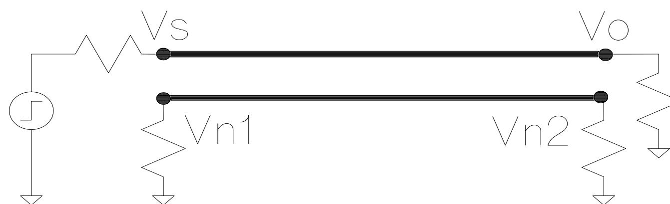

Figure 19. Termination condition for crosstalk simulation.

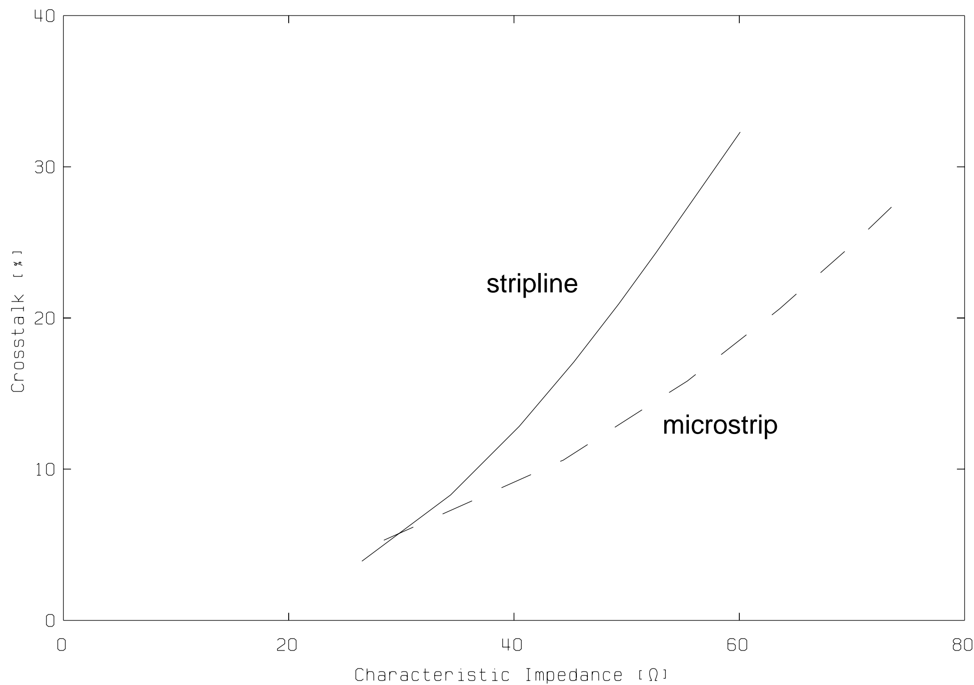

Figure 18. Crosstalk coefficient v.s. characteristic impedance of stripline and microstrip.

$$Vn1 = 0.5\times(\text{crosstalk coefficient})\times Vs.$$

Vn1 can also be calculated by[FELL65]

$$Vn1 = Vs \times (L12/L11 + |C12|/C11) / 4.$$

Note that the backward crosstalk is independent of signal risetime and line length.

The forward crosstalk measured at the far end (Vn2) cannot be calculated from the crosstalk coefficient. However, it can be calculated by

$$Vn2 = -[L12/(2\times Zeff) - (|C12|\times Zeff)/2]\times Vs\times L/Tr$$

where L is the line length and Tr is the signal risetime. The first term in the bracket is inductive coupling and the second term capacitive coupling. For a transmission line with a homogeneous dielectric medium such as stripline, these terms cancel each other resulting in Vn2 = 0. However, for microstrip line the inductive coupling is greater than the capacitive coupling. This is because the electric field is affected by the dielectric while the magnetic field is not. Figure 20 is the SPICE output for 8 cm long terminated microstrip line in Figure 16 with the dielectric thickness of 50 μm. The crosstalk calculated with the above equations are Vn1 = 103 mV and Vn2=-146mV which agree well with the SPICE results of 104mV and -141mV, respectively. Note that the duration of backward crosstalk is the summation of the rise time and two times the time of flight, while that of forward crosstalk is only the rise time.

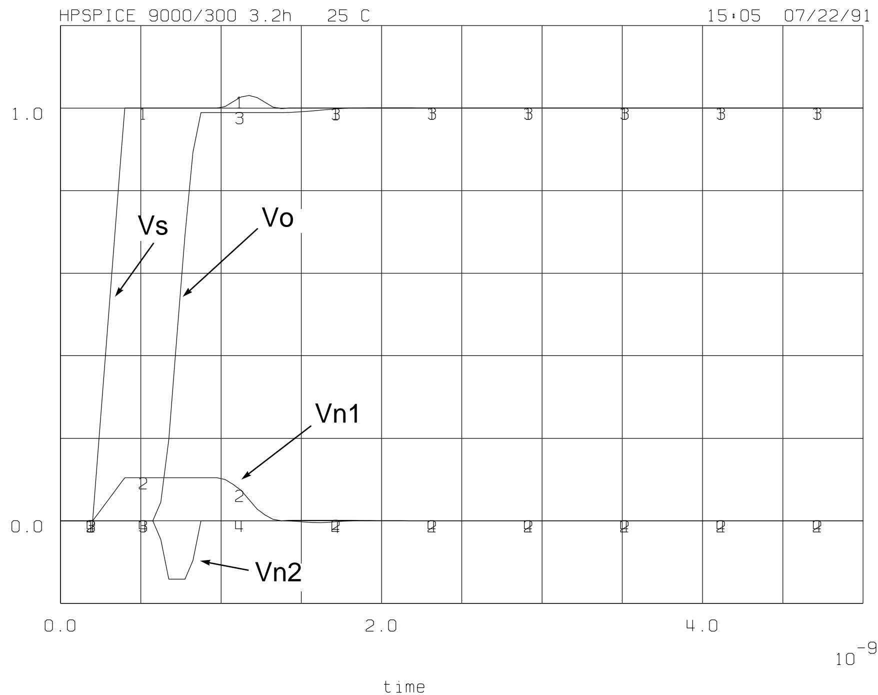

Figure 20. Forward and backward crosstalk in microstrip with the termination in Figure 19.

For the source termination scheme used in this study, the crosstalk waveform is more complicated due to the reflection at the far end. Figure 21 shows the termination conditions of two microstrip lines and Figure 22 is the SPICE output of the crosstalk simulation. When the voltage source rises from 0 V to 2 V with a risetime of 0.2 nsec, Vs rises 1V due to the voltage divider between source impedance and the characteristic impedance of the line. Since the far end is open the signal doubles back to 2V at the far end and the reflected wave goes back to the near end and is absorbed by the source impedance, raising Vs to 2V. On the victim line, the near end initially receives the backward crosstalk. The far end drops down due to the forward crosstalk when Vo rises, as in the terminated case. However, the reflected wave on the driving line, which now propagates toward the near end, generates backward crosstalk which

Figure 21. Series termination of coupled transmission lines.

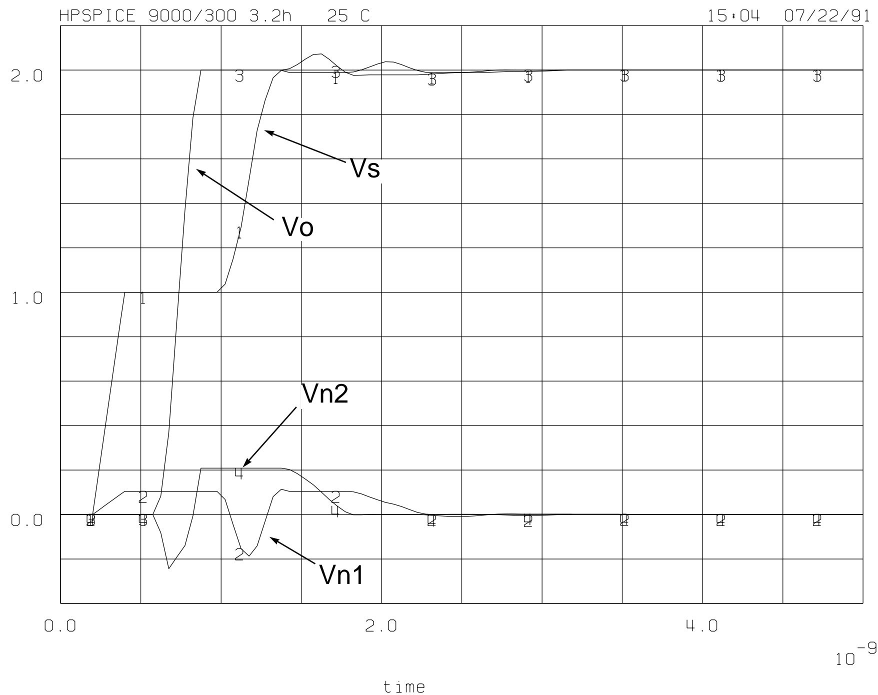

Figure 22. Forward and backward crosstalk in microstrip with the series termination in Figure 21.

goes to the far end of the victim line. This is the reason that Vn2 rises up immediately after the negative forward crosstalk hits the far end. This combined crosstalk itself is now reflected at the open far end of the victim line and propagates to the near end of the victim line. The Vn1 waveform between 1 nsec and 2.3 nsec in Figure 22 is due to this reflection.

The backward crosstalk components at the near and far end of the victim line are unavoidable for both microstrip and stripline. However, the forward crosstalk components which are the large negative spikes in Figure 22 can be eliminated by using stripline, where the electric and magnetic coupling cancel each other. On the other hand, for microstrip uncancelled coupling results in forward crosstalk, which also depends on the length of the line and the signal rise time. Therefore, when microstrip lines are used in an MCM, there are more design constraints compared to striplines, since the line length and rise time must be limited to a certain range in order to control the crosstalk.

The forward crosstalk in microstrip lines can be reduced by coating the lines with a dielectric material which has similar dielectric constant to the dielectric in the microstrip line. Basically, the coating makes the electromagnetic wave propagating in the dielectrics see a more homogeneous medium. As pointed out previously, if the medium is homogeneous the forward crosstalk does not exist. In order to examine the

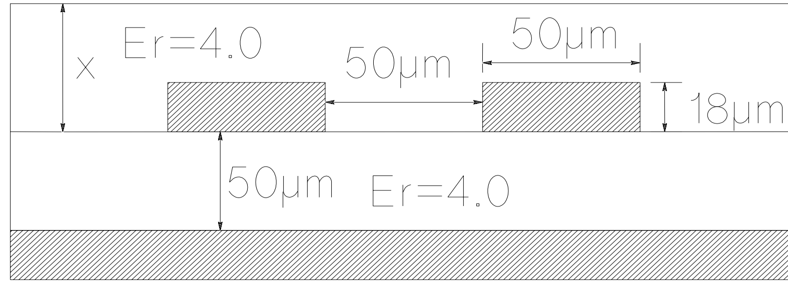

Figure 23. Microstrip lines with dielectric coating.

effectiveness of this approach a pair of microstrip lines were coated with the same dielectric material up to 75 μm as shown in Figure 23. The calculated characteristic impedances with respect to the overcoat thickness are shown in Figure 24. The thicker dielectric results in a higher capacitance, thus lower characteristic impedance. The effective characteristic impedance decreases from 59 Ω to 55 Ω. The crosstalk is shown in Figure 25 for the lines terminated at both near and far end. The line length used in the simulation is 8 cm and the signal rise time is 0.2 nsec. Without the dielectric coating the lines have 10.3% backward crosstalk and 14.5% forward crosstalk. As the coating thickness increases, the backward crosstalk increases because the lines are more strongly coupled capacitively through the dielectric

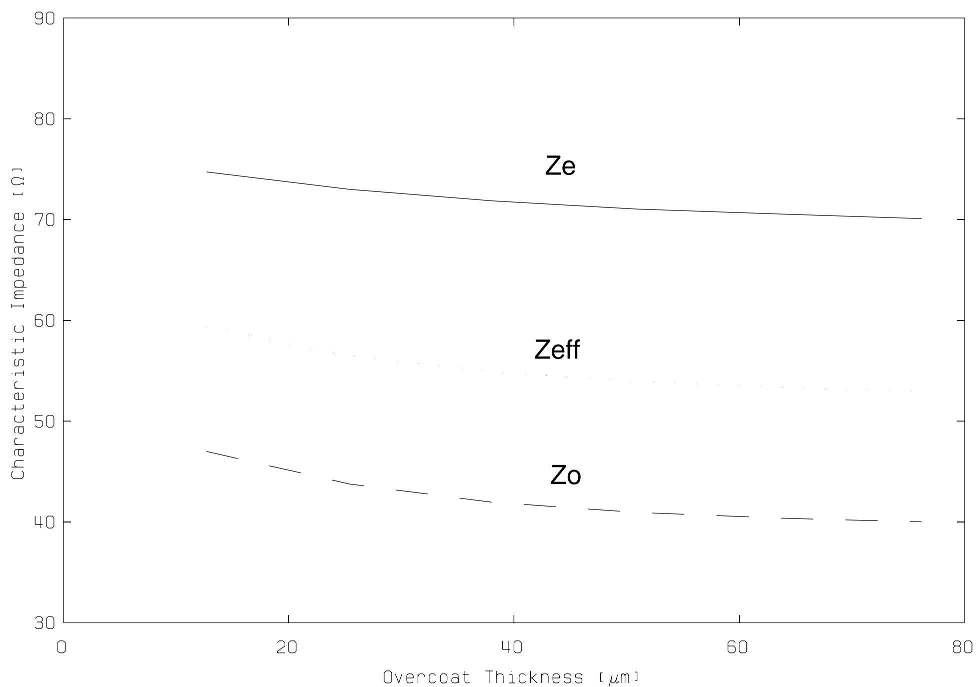

Figure 24. Microstrip impedances v.s. dielectric coating thickness.

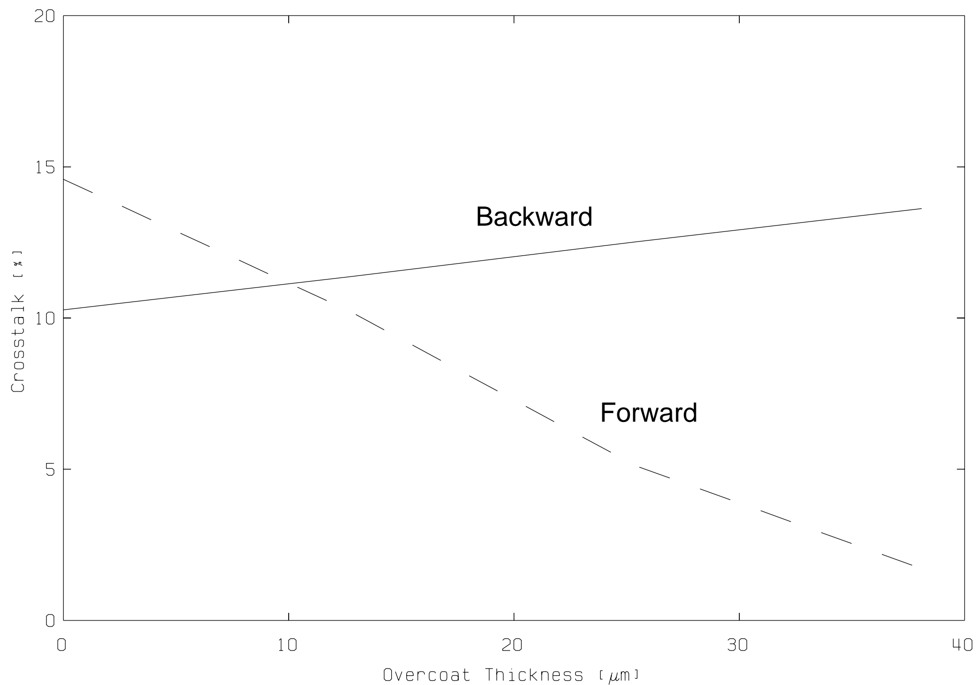

Figure 25. Backward and forward crosstalk in microstrip lines vs. the thickness of dielectric material coating.

material between the lines. However, the forward crosstalk decreases because the increased capacitive coupling now cancels the magnetic coupling more. As shown in Figure 25, 38 μm coating reduces the forward crosstalk down to less than 2%.Therefore even in microstrip lines, the forward crosstalk component can be reduced very effectively with a thin layer of dielectric material.

An obvious solution to reduce the coupling between transmission lines is to increase the space between the lines. A quantitative analysis was done with a coupled stripline and is shown in Figure 26. If the design goal of the characteristic impedance is 60 Ω, the proper dielectric thickness of 127 μm can be found in Figure 15. Using the set of dimensions as a starting point, the space between the lines is increased from 50 μm to 230 μm and the corresponding characteristic impedances are shown in Figure 26. Note that Zeff increases and converges to 64 Ω. Also Ze and Zo converge to the same value. As shown before, the less the difference between Ze and Zo, the less is the crosstalk. The crosstalk coefficient is plotted with respect to the line space in Figure 27. From this curve, the line space can be determined for a given crosstalk design goal. For example if the crosstalk coefficient of 10% is given, a line space of 152 μm should be used. However, Figure 26 shows that the characteristic impedance of the lines with 152 μm spacing is 64 Ω which is higher than the design goal of 60 Ω. In order to meet the goal, a new set of curves is needed for dielectric thickness smaller than 127 μm.

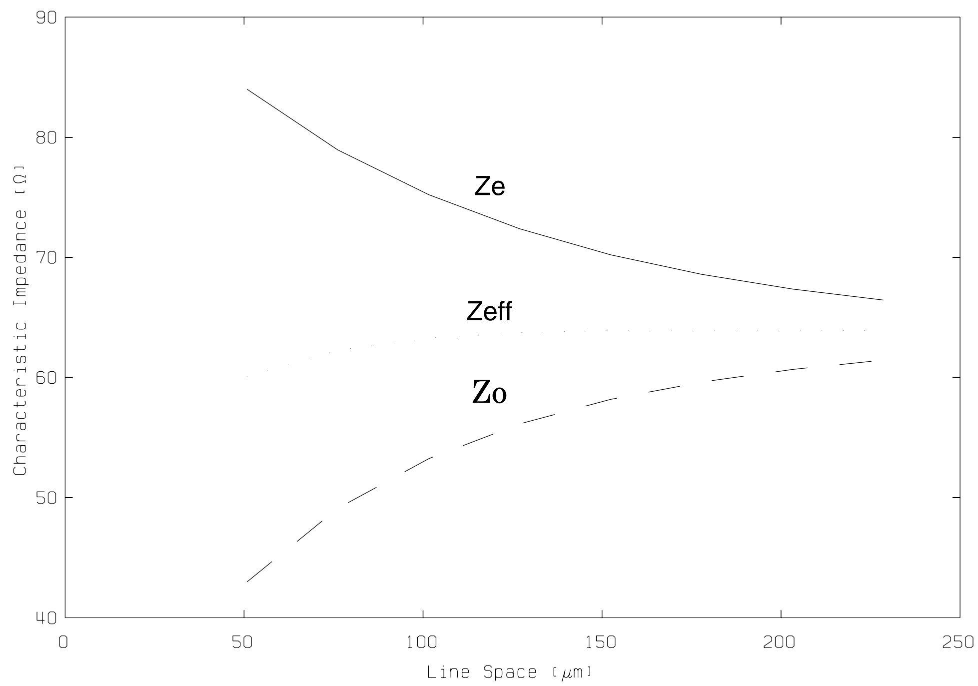

Figure 26. Stripline impedances v.s. line spacing.

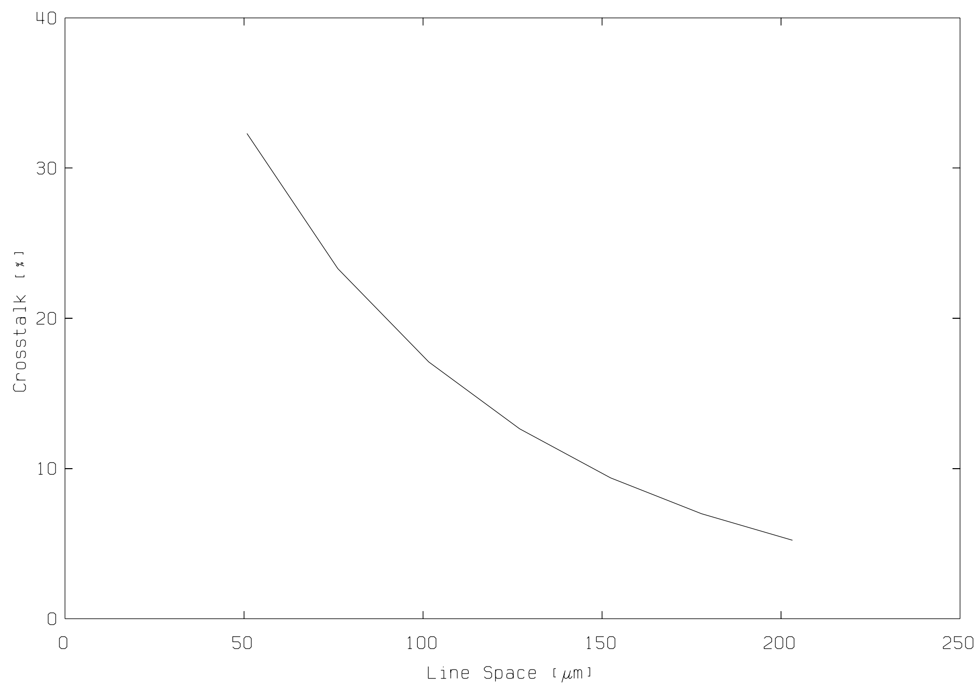

Figure 27. Crosstalk coefficient v.s. line spacing of striplines.

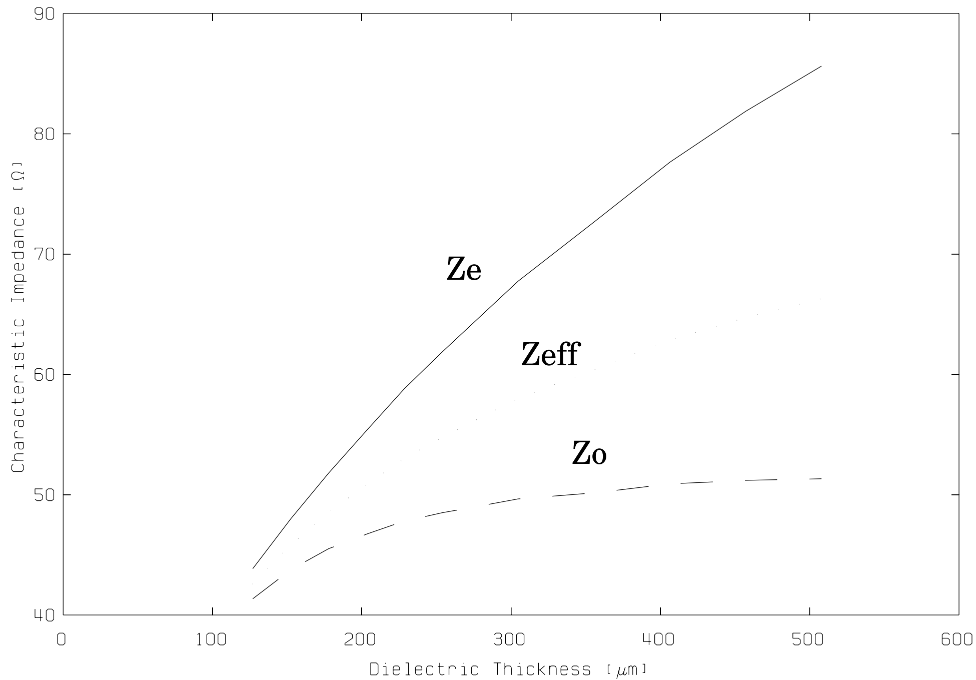

Figure 28. Stripline impedances v.s. dielectric thickness (ε_r = 9).

Another important design parameter is the dielectric constant of the MCM substrate since it has a big impact on crosstalk and characteristic impedance. For example, if a material with the relative dielectric constant of 9.0 is used instead of 4.0 which was used in the previous analysis with 50 μm line width and 50 μm spacing, an entirely new set of geometrical dimensions should be used to obtain the same electrical design goal(60 Ω, 10% crosstalk). Figure 28 shows characteristic impedance v.s. dielectric thickness and 356 μm thick dielectric yields 60 Ω. For the thickness of 356 μm, the characteristic impedance with respect to the line spacing is plotted in Figure 29 and the corresponding crosstalk coefficient is plotted in Figure 30. From Figure 30 the spacing of 406 μm can be chosen for 10% crosstalk. The resulting structure which has a pitch of 457 μm is much larger than 203 μm pitch when the dielectric constant of 4.0 is used. This is another benefit of a low dielectric material besides the inherent benefit of the higher signal propagation speed.

As studied in this section, the design process to achieve the proper characteristic impedance and crosstalk is a multi-parameter optimization process. A designer has to have a good understanding of the relationship of the parameters. It can be a time-consuming process because the electrical parameters have to be calculated through the numerical programs. Work is being done in HP Labs and ICBD R&D to accelerate the process using a database program with interpolation capability[CHAN91].

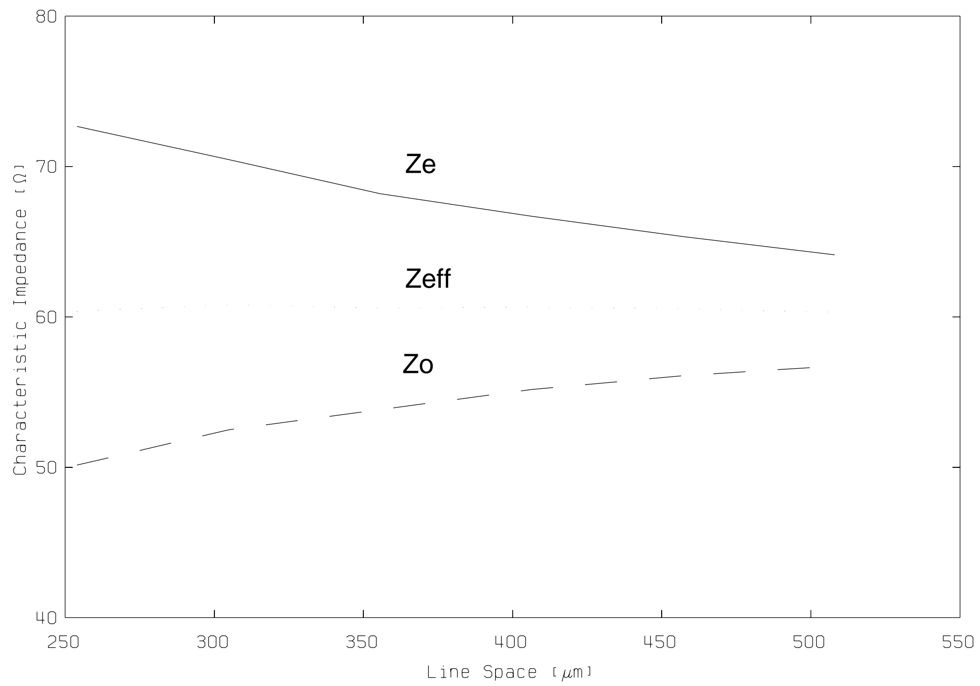

Figure 29. Stripline impedances v.s. line spacing (ε_r = 9).

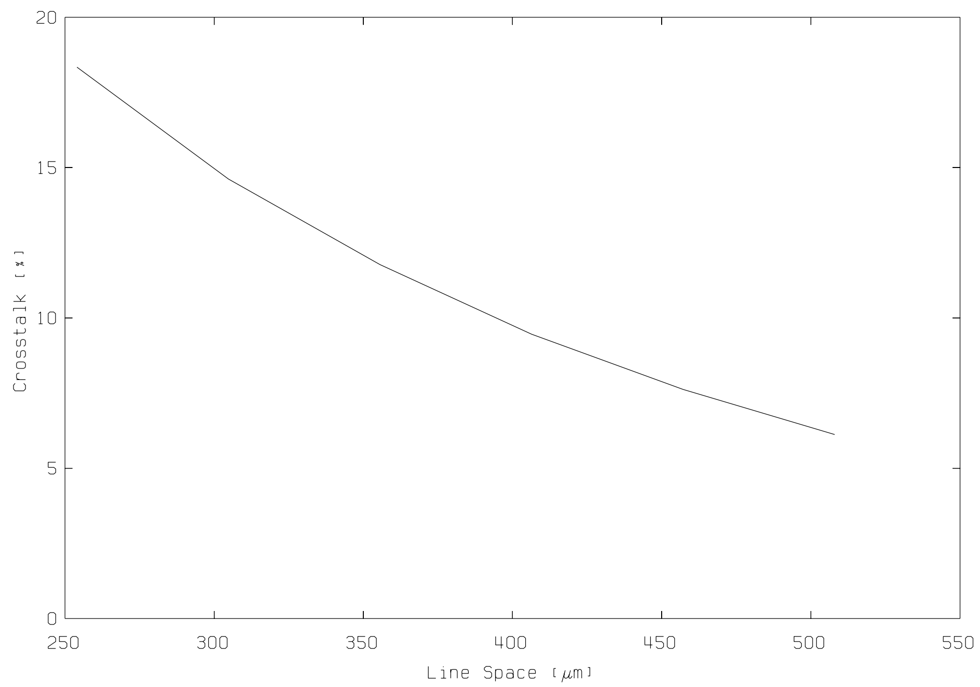

Figure 30. Crosstalk coefficient v.s. line spacing of striplines (ε_r = 9).

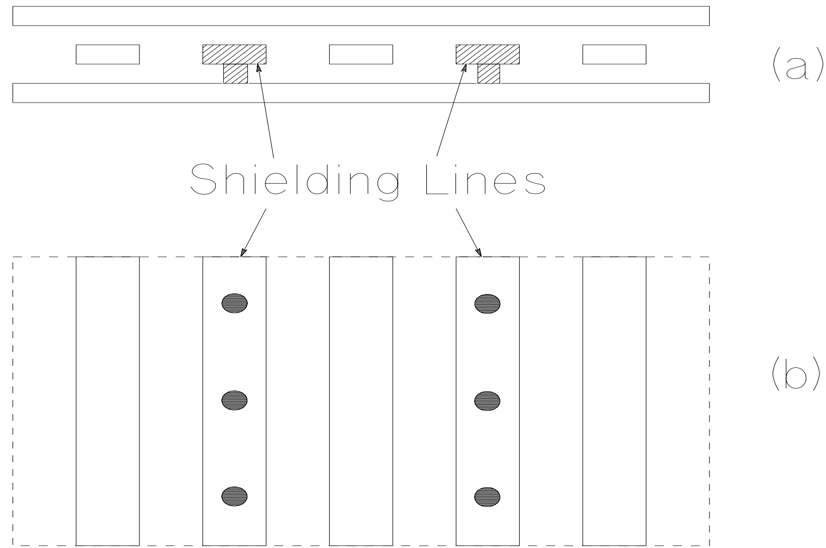

Figure 31. Shielding of transmission lines. (a) cross section. (b) top view.

As mentioned at the beginning of this section, crosstalk can be reduced using shielding lines between the signal lines as shown in Figure 31(a). Obviously, the packing density of signal lines is reduced to half because every other line is taken up for shielding. Another problem with this approach is that the shielding lines have to be connected to the planes with vias as shown in Figure 31(b) periodically in order to maintain a constant potential. Otherwise, the shielding lines do not properly shield the neighboring signal lines. Moreover they can actually help the signal lines couple to each other because they behave like floating conductors. The period of the via connections has to be smaller than 1/10 the wavelength of the highest frequency component in order to make the shielding lines effective. Since making via connection on a minimum width line is usually very difficult in terms of physical process, shielding the signal lines in an MCM is an expensive option. In some cases, the elimination of the shield lines allowing more space between signal lines (three times more than the minimum space if the minimum line width and space are the same) results in a reasonably small crosstalk.

5 Ground Bounce

Ground bounce during the simultaneous switching of output drivers is one of the major noise sources in VLSI circuits. The problem can be described with Figure 32. The power and ground of an IC is supplied through several levels of packages. Ideally, the power and ground voltages should be constant regardless of switching currents of the output drivers. However, due to the parasitic inductance of the IC package leads actual power and ground voltages on the chip are different from those of the power supply. The power supply in the right side of Figure 32 is assumed to be ideal. When

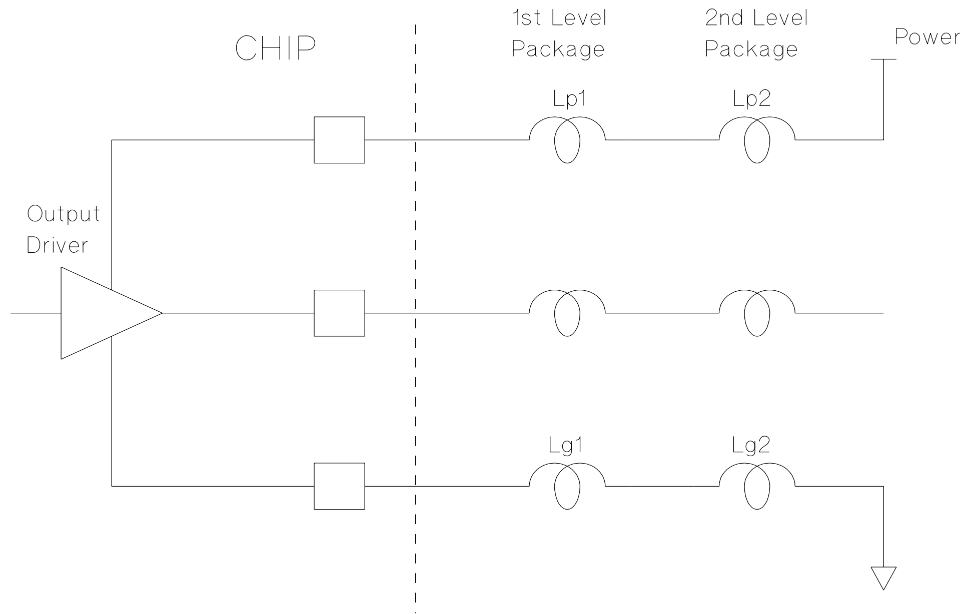

Figure 32. Output driver with parasitic inductances in power and ground connections.

the output driver is switching from high to low, the current coming into the driver has to be drained through the ground line of the package. This current surge is significant in high speed drivers and creates the inductive induced voltage LdI/dt on the line. Therefore, the voltage at the ground pad on the chip momentarily bounces from the ideal ground. This problem becomes more significant if the ground lines are shared by other drivers and those drivers are switching together. The resultant induced voltage is NLdI/dt, where N is the number of drivers switching simultaneously. If this noise is excessive, it can cause a false signal. The simultaneous switching noise also occurs on the power lines when the output drivers are switching from low to high. The power line noise is not discussed separately here because it is exactly the same in nature as the ground line noise.

There are three kinds of solutions to the ground bounce problem apparent from the equation, Switching Noise = NLdI/dt. First, we can reduce N, the number of drivers sharing the same ground line. This solution requires more ground lines. However, the number of connections available for grounds in IC packages is quite limited in many cases. Second, reducing the inductance can directly lower the noise. There have been a lot of efforts to reduce the size of IC packages resulting in lower parasitic inductance. However, the inductance cannot be totally eliminated unless the package itself is removed. Third, the switching noise can be reduced if dI/dt is made small. But this forces the entire system to operate at a slower speed, which is not desirable. This

situation can be avoided if differential drivers are used since dI/dt is small for each complementary output pair. However, the number of signal pins increases by a factor of two in the differential drivers.

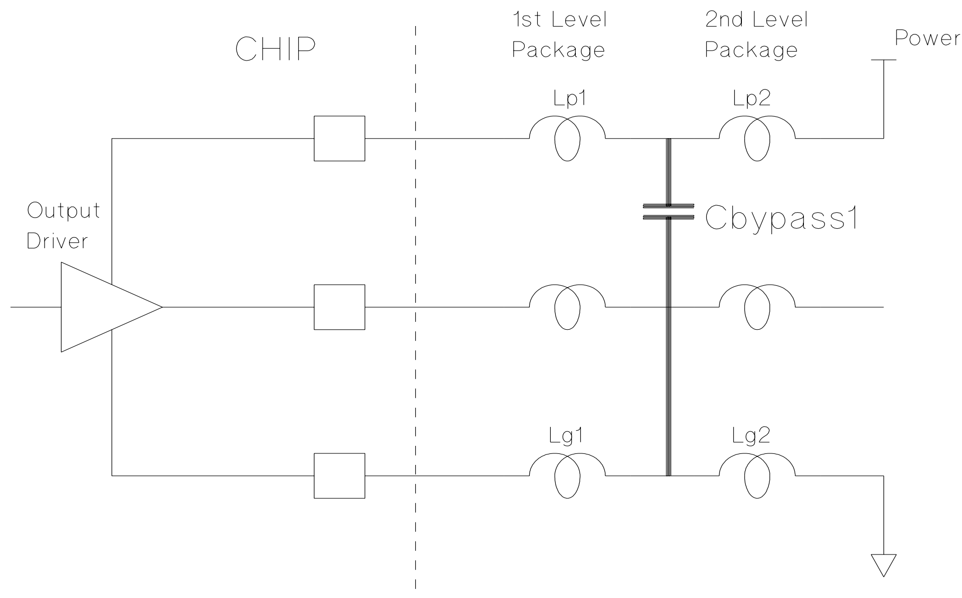

Figure 33. Output driver with a conventional bypass capacitance.

Another common practice used for reducing the switching noise is to place a capacitance between power and ground lines as shown in Figure 33. This capacitance acts as a reservoir of charge which can be used when the output driver requires a sudden increase of current. Another way to look at this capacitance is that it bypasses the high frequency current needed by the driver. If this capacitance is fairly large it can reduce the ground noise due to the inductance Lg2. However, this still does not solve the noise problem due to the inductance Lg1, because the high frequency current still has to go through Lg1. This makes it apparent that the closer the bypass capacitance is to the chip, the more effective it is to reduce the ground bounce noise.

There have been many solutions proposed and implemented to place the bypass capacitance closer to the chip. For example, if a PGA is used to package an IC, the bypass capacitances are usually placed on the PGAs instead of PCBs. In some cases, the bypass capacitance is embedded in the IC package improving the performance. However, all these solutions have some residual inductance between the bypass capacitance and IC chip due to some interconnect lines such as bonding wires.

In this report, we describe a very effective way to place the bypass capacitance. The basic idea is to place the bypass capacitance right on the chip, which we call on-chip bypass capacitance(OBC) as shown in Figure 34. A sufficient amount of capacitance

can be obtained without modifying the existing IC process steps. The required area on the chip is minimal and the effectiveness of bypassing is excellent as shown below.

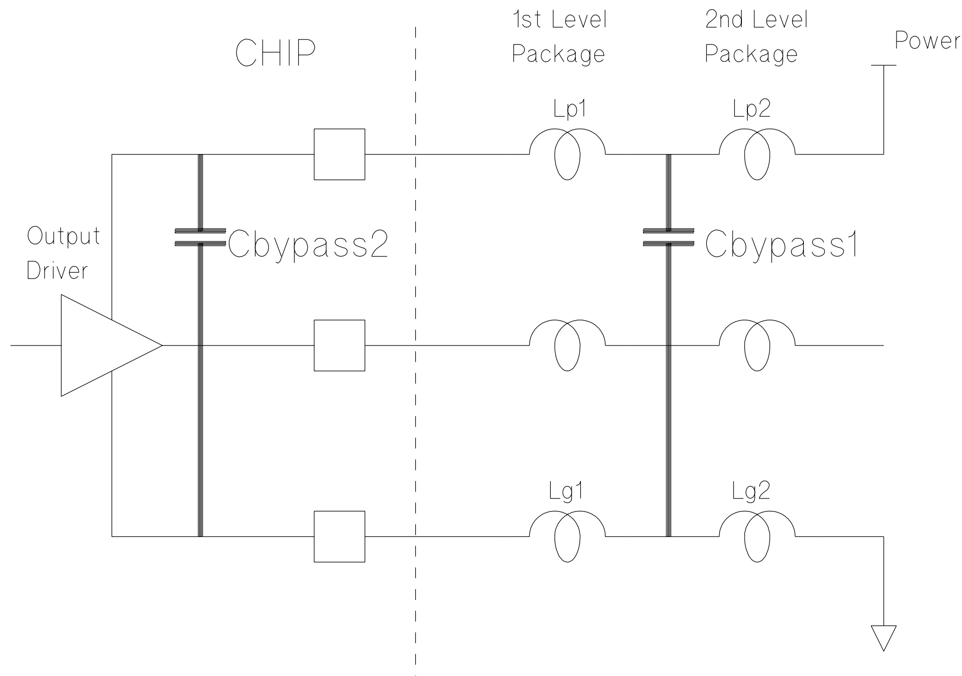

Figure 34. Output driver with an on-chip bypass capacitance.

OBC can be implemented in various ways depending on the process steps used for the IC chip. We demonstrated the effectiveness of OBC by implementing it on the HP BiCMOS14 process. The process has a 150 Å thick oxide gate for FET devices which provides 2.3 fF/μm^2 of available capacitance. If an IC chip is 1.4x1.4 cm and OBC is implemented as a 100 μm wide strip around the chip as shown in Figure 35, the total capacitance is 12.9 nF, or 23 pF for a bonding pad pitch if the pad pitch is 100 μm. The total area for the OBC is 5.6e-2 cm^2 and it is only 2.9% of total chip area.

Figure 36 shows the cross section of the OBC. The capacitance is formed with silicided polysilicon as the top plate and an n+ diffusion layer as the bottom plate. The polysilicon is the same layer as the gate of FETs and the n+ diffusion is the one for the collector of bipolar devices. The contact to the bottom plate is made by opening contact holes directly above the diffusion layer. By adjusting the number and location of the contact holes the series resistance of the diffusion layer can be reduced, which is essential for effective bypassing. The series resistance is on the order of 1 Ω for properly contacted OBCs.

The effect of the OBC is quite dramatic as shown in Figure 37. Figure 37 shows the net voltage between power and ground while the output drivers are switching when there is no OBC and power and ground inductances are 10 nH. It swings ±35% from

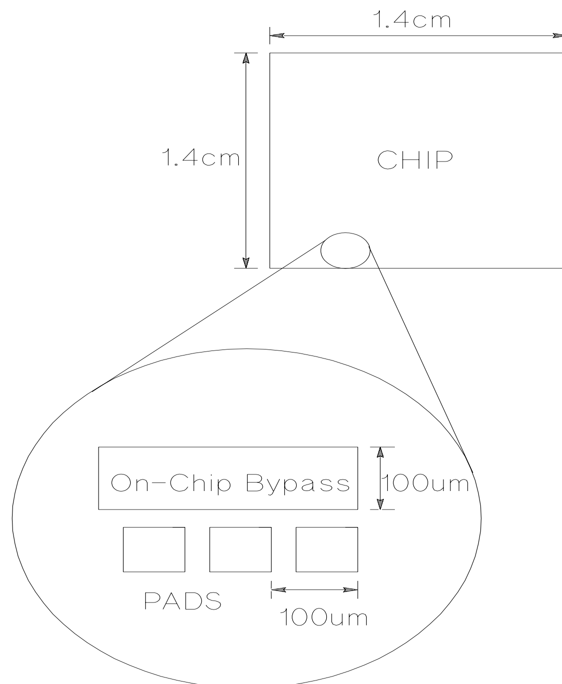

Figure 35. Dimensions of on-chip bypass capacitance.

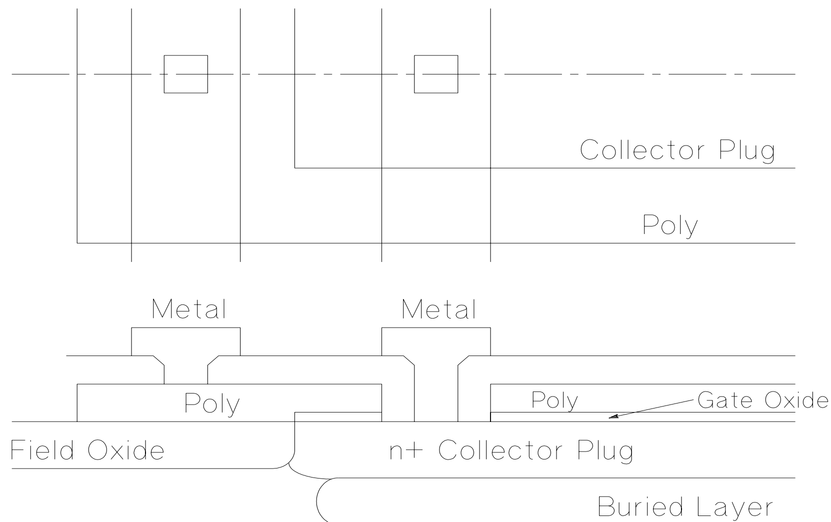

Figure 36. Layout and cross section of on-chip bypass capacitance.

its nominal value (4V), which is not acceptable in practical circuits. This noise is proportional to the inductance value. For a 2.54 mm long 25 μm diameter wirebond, which has 2.6nH of inductance, the noise is ± 9%. For the same circuit with 10 nH power/ground inductance, the net power-ground voltage swing is reduced to ± 2%

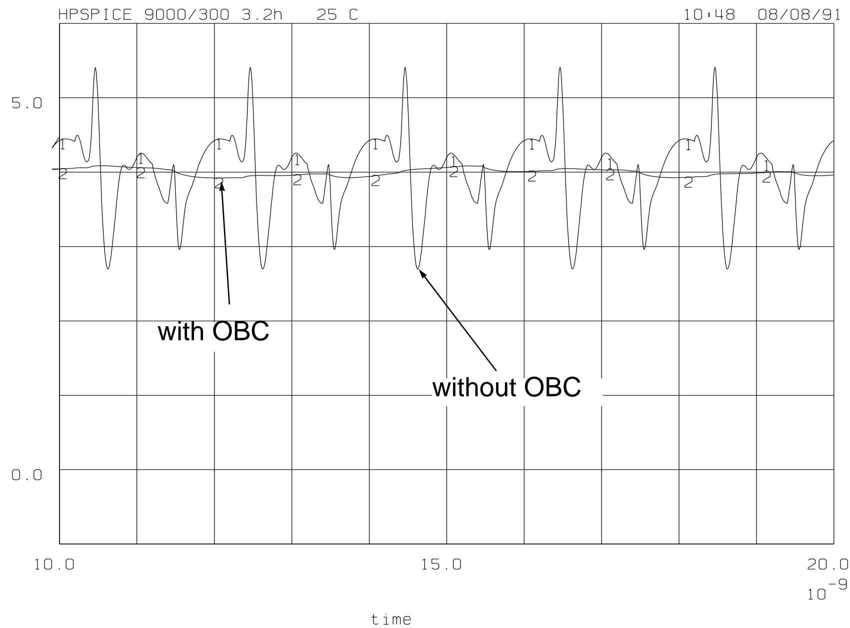

Figure 37. Net Vdd (Vdd-Gnd) on the pads with and without on-chip bypass capacitance when drivers are switching simultaneously.

when OBC is included as shown in Figure 37. The value of the OBC used was only 40 pF which is only two pad pitch worth of capacitance. This value is much smaller than those of conventional bypass capacitances. However, it is much more effective since it is implemented right adjacent to the driver.

In conclusion, the on-chip bypass capacitance is implemented using an existing IC process with minimum area penalty and it reduces the power and ground switching noise very effectively.

6 Terminations and Reflections

Failure to adequately terminate the transmission line between driver and receiver will result in reflections from any impedance discontinuity on the line. See [BAKO90] for a good discussion of transmission line fundamentals and [RAMO84] for a more thorough, detailed treatment. If there are multiple discontinuities there will be reflections from each which can combine at the receiver, causing intersymbol interference and potential bit errors. Even if the reflections are not sufficient alone to cause bit errors, they combine with other interference sources such as ground bounce and crosstalk, and the combination must be studied to understand the extent of noise margin degradation.

Lines can be terminated at the drivers, at the receivers, both, or neither. Even an unterminated network will eventually settle at a stable final value as multiple

reflections are attenuated by the inevitable loss in transmission lines. However in cases where timing requirements preclude such delays first incidence switching is required. First incidence switching implies that every receiver on a net switches to its final digital state upon the first arrival of the driver waveform and afterwards does not change state even momentarily with subsequent reflections. The maximum speed of operation of a net is reduced by non first incidence switching because the time interval between data bits must be greater than the settling time of the net to avoid intersymbol interference. Thus high speed systems are usually designed to be first incidence switching and this was an assumed requirement for this work. The following few examples describe the design parameters necessary to achieve first incidence switching as well as the penalties for failing to achieve it for a variety of network types.

Figure 38 shows a discretely loaded net with source termination. Receivers are modeled simply as capacitors at points 2 and 3 and lines are lossless for clarity in illustrating the principles. With source termination the propagating waveform from the driver is only half height initially due to the voltage divider action between the source termination resistor and the Zo of the transmission line. It doubles to full height only when it arrives at point 3, the open circuit termination. Consequently a receiver at point 2 will not see full amplitude voltage until the reflection arrives from point 3. This configuration cannot achieve first incidence switching; the minimum additional delay introduced beyond first incidence for a valid signal at point 2 is the time of one round trip to point 3. Additional reflections from (capacitive) receivers threaten to switch receiver states again. The negative spike in the point 3 waveform

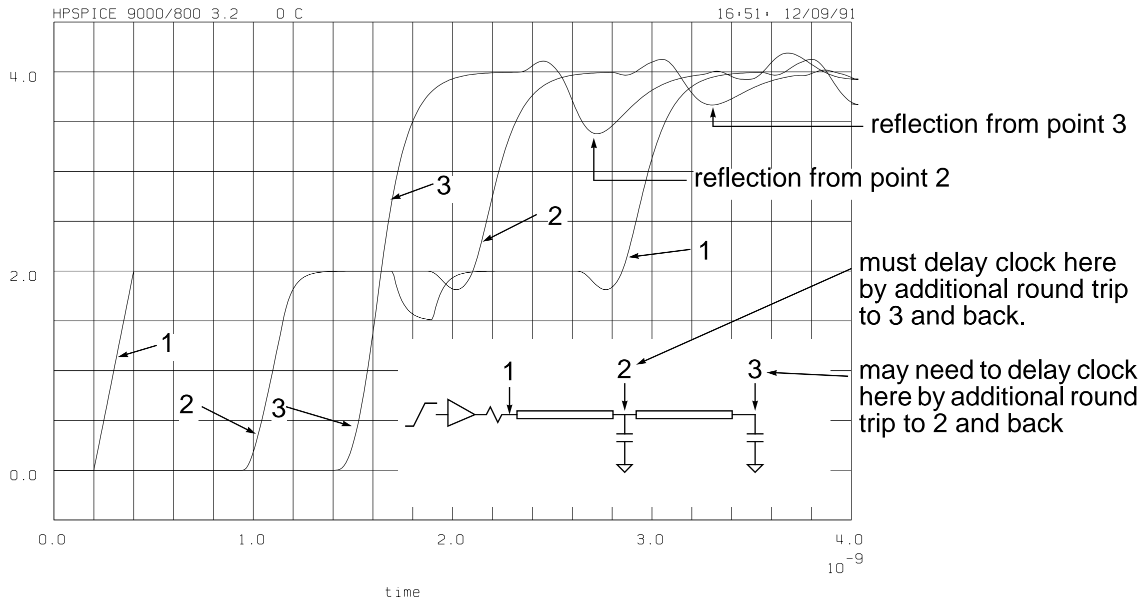

Figure 38. Discretely loaded net with source termination.

at the 2.8 nsec point in Figure 38 is caused by the initial reflection from point 3 being re-reflected from the capacitance at point 2. Similarly, the negative spike at the 3.5 nsec point in waveform #2 is caused by the previous spike reflected from the open circuit at point 3. Whether these spikes are sufficient to cause a receiver to switch state depends on factors such as line characteristic impedance, receiver and parasitic capacitance, driver waveshape, and aggravating circumstances such as ground bounce and crosstalk. Any proposed design should be analyzed in the manner described in Section 9. If the designer cannot be assured such reflections will not cause errors, the additional delay to await both receivers reaching assured final states is equal to the round trip propagation time between points 2 and 3.

Source end termination has the important advantage of low power consumption; in fact, static consumption is zero. As shown above, however, source termination cannot achieve first incidence switching with multiple discretely loaded receivers. This can be solved, at the expense of power consumption, with both end termination as shown in Figure 39. In this configuration the same voltage divider action is present as in the source end termination case described above except that, with no far end termination, the voltage doubling reflection does not occur and the final received voltage will be one half the open circuit driver output (2 volts in this example) and receiver threshold would be set at one fourth the open circuit driver output. The only threat to first incidence switching is the reflection from the terminating capacitance at point 3 incident on the receiver at point 2 at the 2.1 nsec point. This capacitive reflection will be similar in amplitude to the capacitive reflections described in the source termination case described above since it is caused by the same amplitude (half-height) waveform incident on the same value of capacitance. However, its effect on

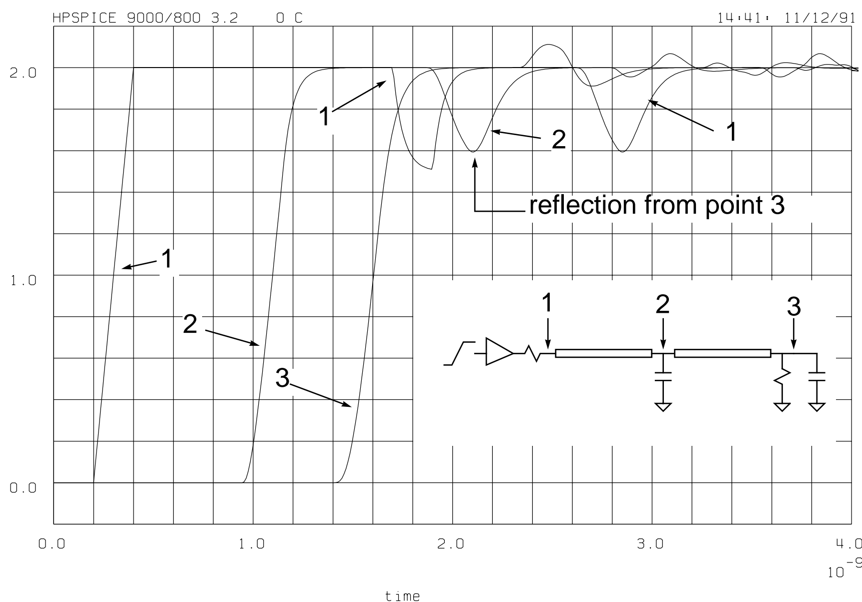

Figure 39. Discretely loaded net with source and load termination.

noise margin will be twice as great since the static noise margin for this doubly terminated case is half that for the source terminated case. Using far end termination only (that is, a voltage source as driver) will not relieve this effect either because, even though this will double the voltage delivered and the static noise margin, the voltage waveforms propagating on the line are now full height rather than half height and the capacitive reflections will be doubled as well. As before, a complete analysis would tell whether these reflections are tolerable, and the delay penalty if they are not is one round trip time between points 2 and 3.

A near end cluster is another way to drive multiple receivers with one driver, using separate lines for each receiver. Various circuit configurations can be used to connect the driver to the lines with the goal of coupling the maximum voltage waveform into the line while minimizing the inevitable interaction between the lines. Figure 40 shows one such configuration emphasizing the former at the expense of the latter. Since the lines are connected at their source end each will see a source impedance equal to the characteristic impedance of the other line in parallel with the impedance of the driver. Thus the reflections from the receiver end upon arriving back at the source end will be partly reflected (inverted) as well as partly transmitted down the other line. A transmitted pulse will then bounce repeatedly among the source and both receivers. With unequal line lengths the result is severe noise margin degradation as shown and no possibility of achieving first incidence switching.

A more optimum circuit configuration for a near end cluster potentially capable of first incidence switching is shown in Figure 41. Here the resistor values are chosen to provide a good impedance match at the source of each line so that reflections from the open circuit ends will not be re-reflected at the source end. This requires resistor

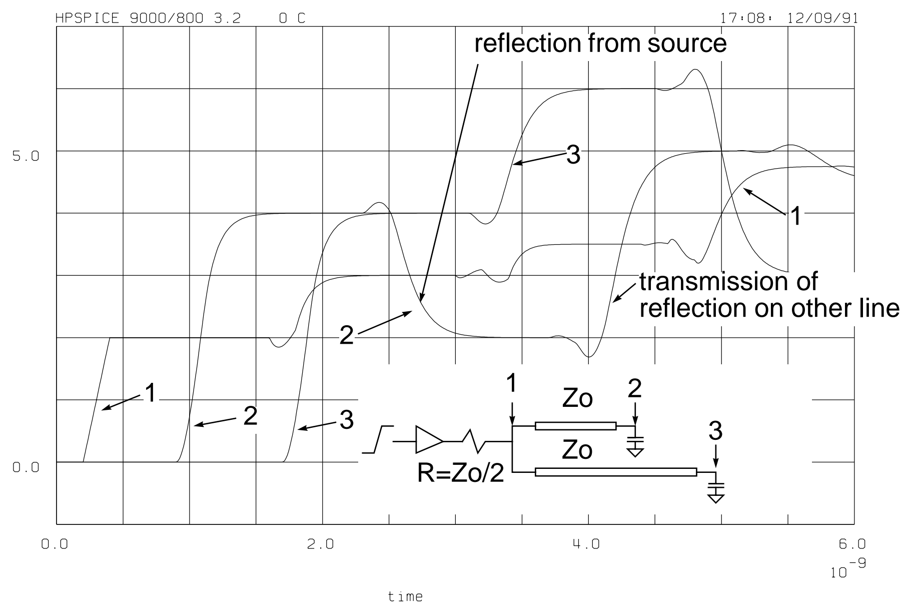

Figure 40. Near end cluster with source termination.

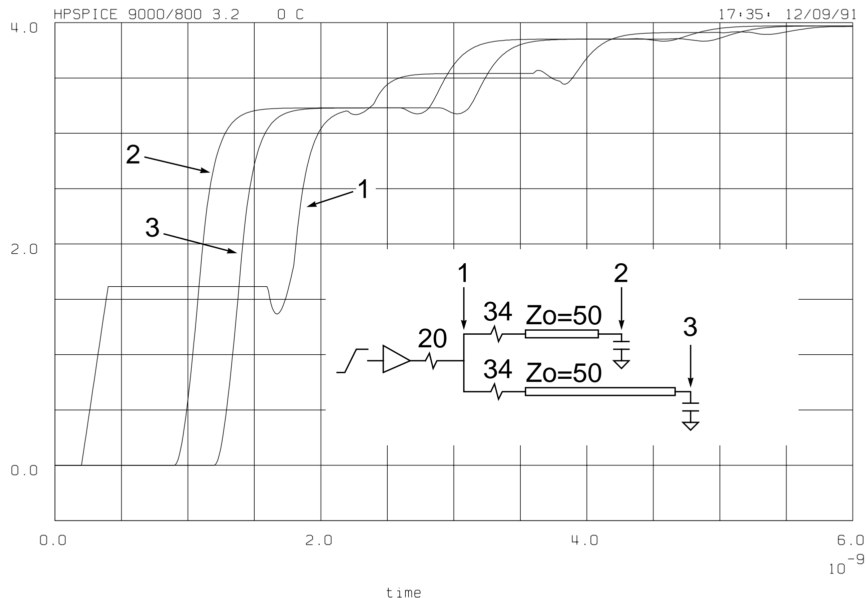

Figure 41. Near end cluster with multiple source terminations.

values that will underdamp the transmission line; that is, the voltage divider action on the driver pulse will initially inject a voltage of less than half height and only with multiple reflections on the transmission line will the far end voltage rise in an underdamped fashion to its final value as indicated in Figure 41. In this case, however, since there are no reflections at the source end, pulses propagating toward the source are transmitted to the other line. Thus each line is charged to its final value by reflections from the termination on the other line, which has the effect of slowing the charging time constant since reflections must travel the length of both lines to charge one of them up, not just the length of the one being charged. The threat to first incidence switching in this configuration is the degradation of noise margin while the lines are charging, the time between 1.5 and 3.5 nsec for the example of Figure 41, where nearly one half of the static noise margin has been lost. Once again only a full analysis will reveal whether this is acceptable. If it is not, the additional delay to await circuit settling is equal to at least one round trip down both transmission lines.

The applications studied here involve high speed switching with transition times as fast as 150 psec. Point to point interconnections, with no distributed taps or clustered nets, are necessary at these speeds because of all the issues discussed above. For receiver end termination a shunt resistor is the simplest solution; however, dc power consumption is often unacceptably high. Terminating the shunt resistor to ground with a capacitor, rather than directly, will eliminate the dc power consumption problem at the cost of the added complexity of the capacitor, whose value is usually too large for VLSI integration. Termination at the driver end is most easily

accomplished with a series resistor and a low impedance driver. Power consumption is very low and the full open circuit output voltage of the driver is delivered to the receiver. Thus a series termination at the driver end is very attractive for its low power consumption and is used for all the simulations in this paper.

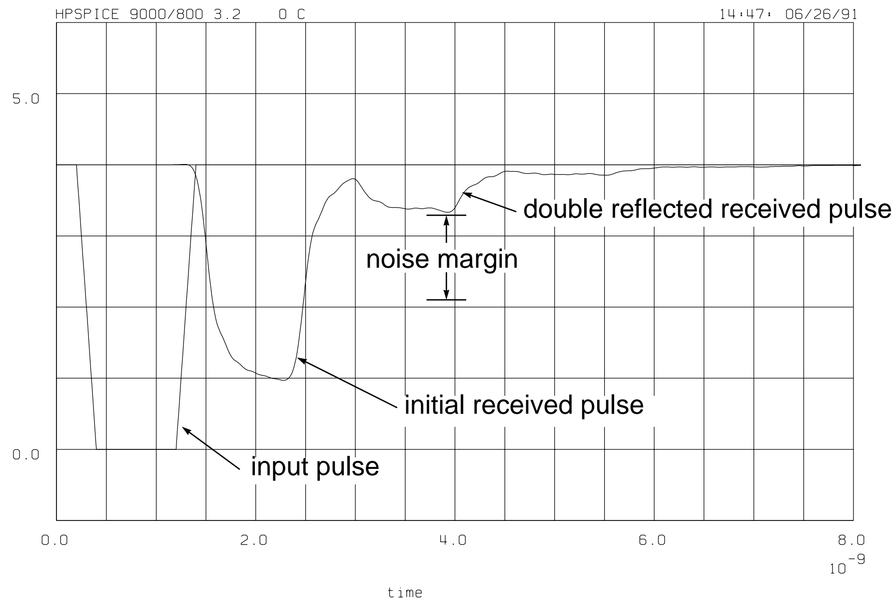

Figure 42. Received pulse for source termination resistance too high.

If the termination resistor value is not optimum, due to processing variations for example, the reflections will be partial and noise margin degraded by an amount depending on the amount the value deviates from optimum. Figure 42 shows a received pulse sent by a driver with a termination resistor value too high. Two effects are present degrading noise margin. First, the initial received pulse does not swing all the way to ground due to the voltage divider action initially between the source resistance and the transmission line characteristic impedance. If the transmitted pulse were wide enough to allow multiple reflections the received pulse would eventually settle at ground. Second, the double reflected pulse deflects the nominally quiescent received signal toward the receiver's transition voltage causing reduced noise margin at that point. Figure 43 shows a corresponding situation with a termination value less than the matched value. Here, the exact opposite effects are observed due to the favorable voltage divider effect and the inversion of the double reflected pulse. The quadruple reflected pulse, however, is again inverted and appears with a polarity and position to degrade noise margin.

Figure 44 shows the effect on noise margin for a complete range of source termination values. The simulation uses the circuit of Section 2 Figure 1 with no interfering signals on the adjacent lines in order to isolate these reflection effects from ground bounce and crosstalk. There is a value of source resistance which terminates the transmission line with a matched load, eliminating the reflections which degrade

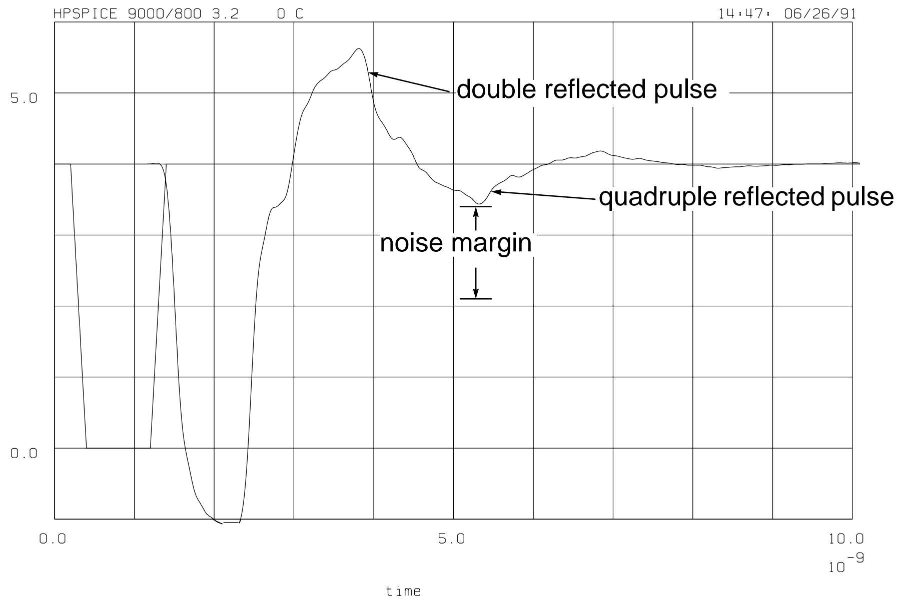

Figure 43. Received pulse for source termination resistance too low.

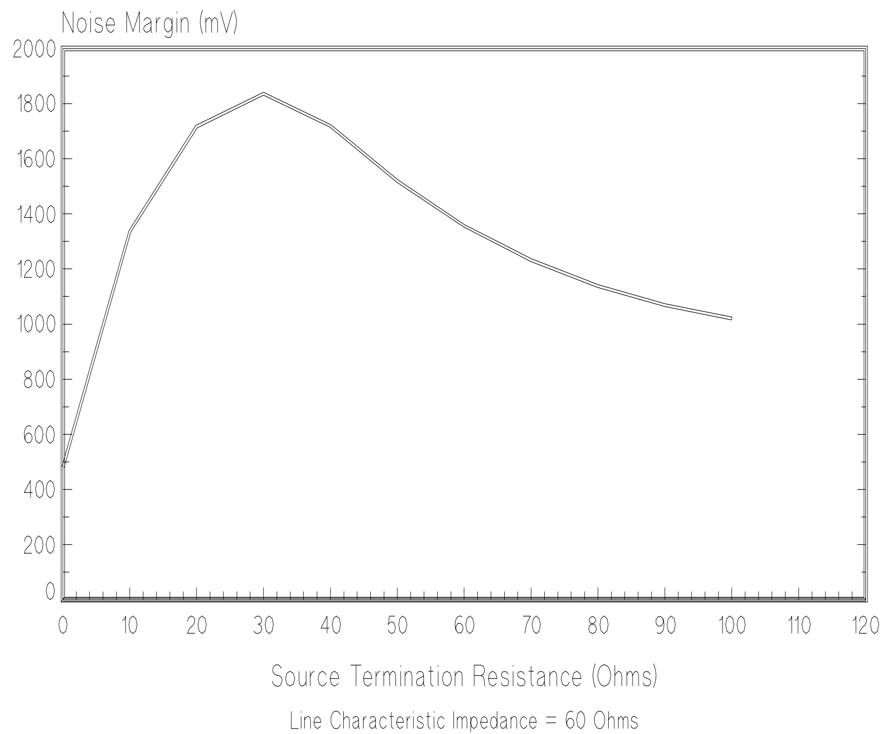

Figure 44. Noise margin degradation vs. source termination resistance.

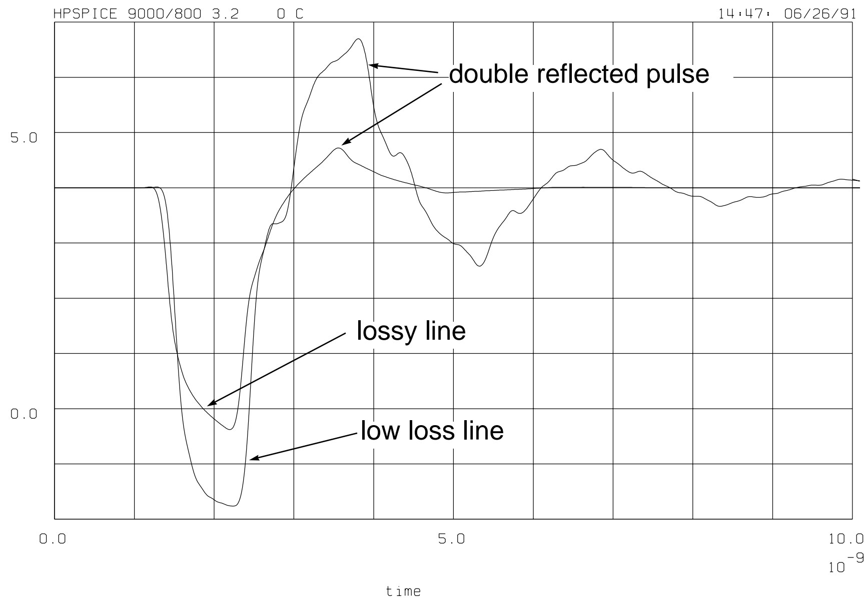

Figure 45. Received pulse with lossy line for source termination resistance too low.

noise margin. This occurs when the combination of the termination resistor and the effective source resistance of the driver equal the characteristic impedance of the transmission line, 60 ohms in this case. The drivers used here are simple CMOS inverters with an ON resistance of about 20 ohms.

The high frequency loss of the transmission line can dramatically change the effect of mismatches. The transmission line used in Figures 42, 43 and 44 has low loss, less than 1 dB at the frequencies and lengths used here. Examples include copper conductors on FR4 printed circuit boards, molybdenum or tungsten on multilayer ceramic, and some copper on polyimide structures. Extremely small geometry lines can have sufficient distributed resistance to yield a loss of several dB at lengths of 10 cm. Figure 45 compares the low loss line of Figure 43 with a line 2.5 µm thick and 20 µm wide on a polyimide dielectric 15 µm thick, yielding a characteristic impedance of 50 ohms and distributed resistance of 6 ohms per cm. The terminating resistor value is zero in both cases. The large double reflected pulse present in the low loss case is nearly completely attenuated by its three trips down the line in the lossy case. Further, the low resistor value yields a larger initial received pulse, now needed because of the higher loss of the line. Figure 46 compares the effect of resistor value on noise margin for both the low loss and high loss cases. The designer using a lossy transmission line must either use different termination values for different length lines or carefully choose a value that compromises reflections for short lines and initial received pulse attenuation for long lines.

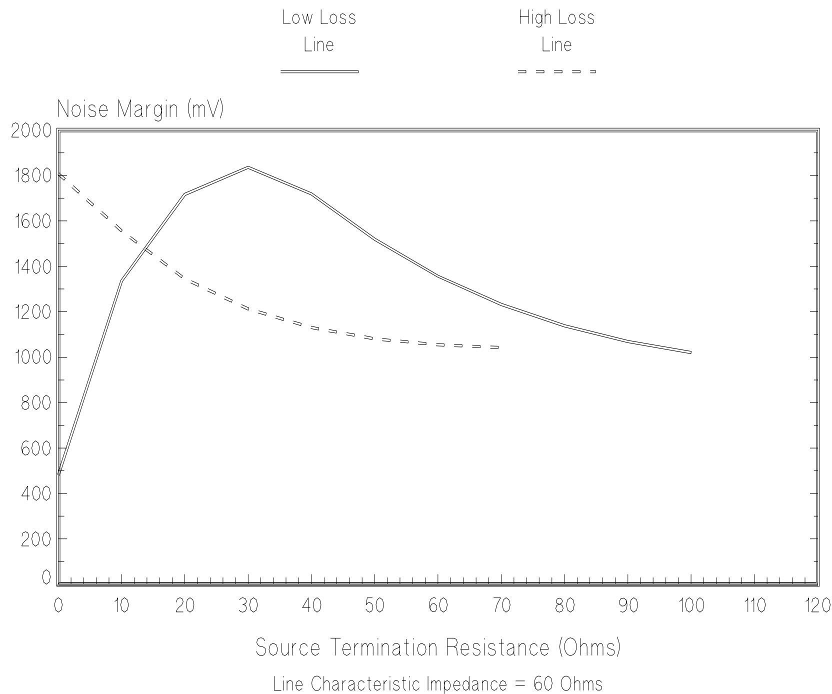

Figure 46. Noise margin degradation vs. source termination resistance.

Another interesting reflection phenomenon occurs when bipolar drivers are used instead of FETs, illustrated in Figure 47. The FET drivers exhibit a linear resistive behavior when turned on, regardless of the output terminal voltage or conducted current. In other words, the output FET dimensions are large enough that the FET will remain in its linear region even when absorbing 80 mA of current induced by a pulse incident on its output. It thus presents a good termination to the transmission line under all conditions. The emitter follower outputs of the bipolar drivers, on the other hand, are not conducting when quiescent and driving an open circuit. Consequently a reflected pulse incident on a quiescent driver will see an open circuit unless its amplitude and polarity are correct to cause the output transistor to conduct. For example, if an NPN emitter follower output is high, once the transmission line is charged no emitter current will flow and base-emitter voltage will be zero. When a pulse is incident on the driver in this state no emitter current will flow, and the impedance seen by the pulse will be high, until the emitter is pulled at least 600 mV below the base. That 600 mV of the incident pulse will be reflected rather than absorbed in the driver. If the transmission line loss is low, a pulse can bounce many times between such a driver and a receiver, only slowly decaying due to line attenuation as shown in Figure 47.

These examples have all assumed no interfering signals on adjacent lines for clarity. Crosstalk from adjacent lines will be primarily in the backward direction; however, a

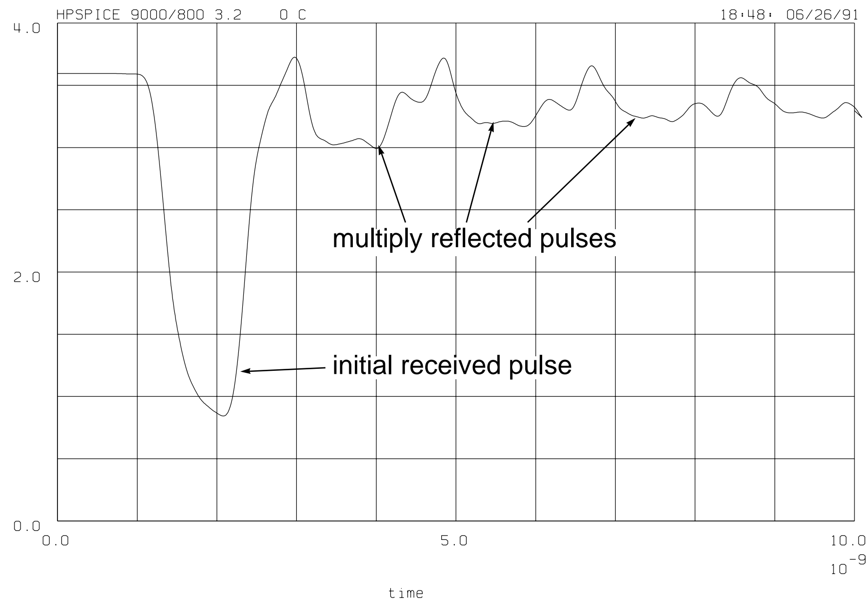

Figure 47. Multiple reflections induced by non-linear driver behavior.

poor source termination will reflect it in the forward direction, degrading noise margin. The combined effects of multiple interfering lines, ground bounce and reflections are studied further in Section 10, "Effect of all contributions combined."

7 Common Power Supply Inductance

The MCM studied in this report is to be placed on a printed circuit board which supplies power and ground connections. This power supply is not an ideal power supply since the local connection and supply plane shown in Figure 48 has some impedance. At the time of the simulations done for this report, there was not a good physical model to represent the power supply impedance, especially for the power supply planes. Therefore, an assumption was made that the power supply impedance is an inductance of up to 10 nH. This resulted in a low frequency resonance between the inductor and the on-chip bypass capacitance. The low frequency resonance is shown in Figure 49, where net Vdd (Vdd-Gnd) is plotted when all the drivers are switching at 500 MHz. The high frequency noise in Figure 49 is due to the ground bounce which was discussed in section 5. The noise margin defined in section 2 is plotted in Figure 50. The instantaneous noise margin drops down to 1.4 V from its nominal value of 1.9 V due to the low frequency resonance. This can actually happen if a power supply has a lumped inductance of 10 nH. However, if planes are used for a power supply, the impedance of the power supply is quite different from a lumped inductor. A physical model for the power supply planes is being created and will be published[LEE91]. The essence of the model is presented here.

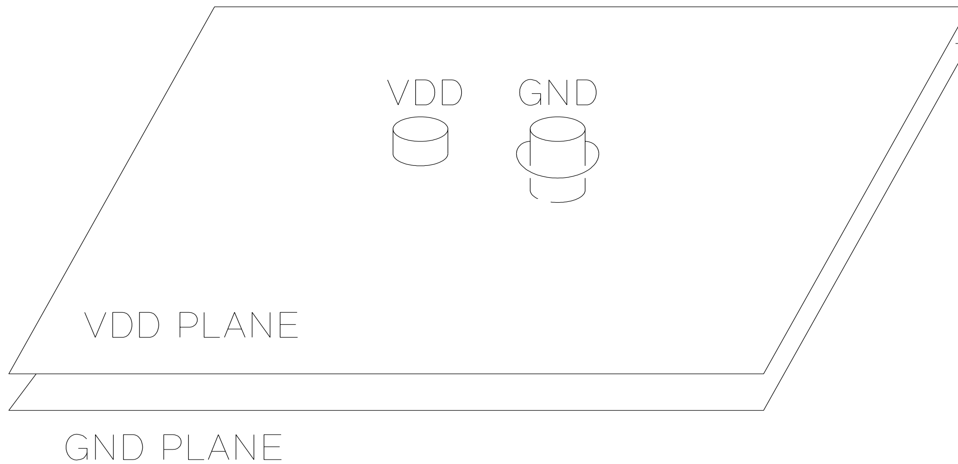

Figure 48. Power supply through the two planes of a printed circuit board.

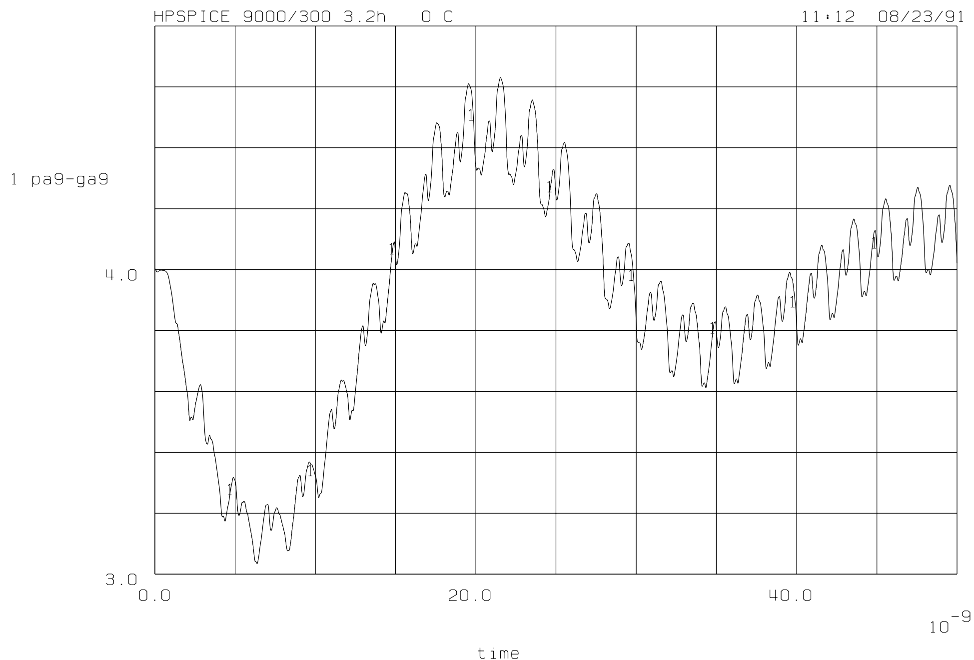

Figure 49. Net Vdd (Vdd-Gnd) with 10 nH common power supply inductance.

When a pulse of current is pulled out of the power supply planes, the pulse propagates radially toward the edge of the plane as shown in Figure 51. At the edge the pulse is reflected differently depending on the boundary conditions. Some of the energy travels through the power and ground connectors at the edge and becomes conductive emission noise if it is not properly bypassed[OTT88]. The pulse diminishes eventually

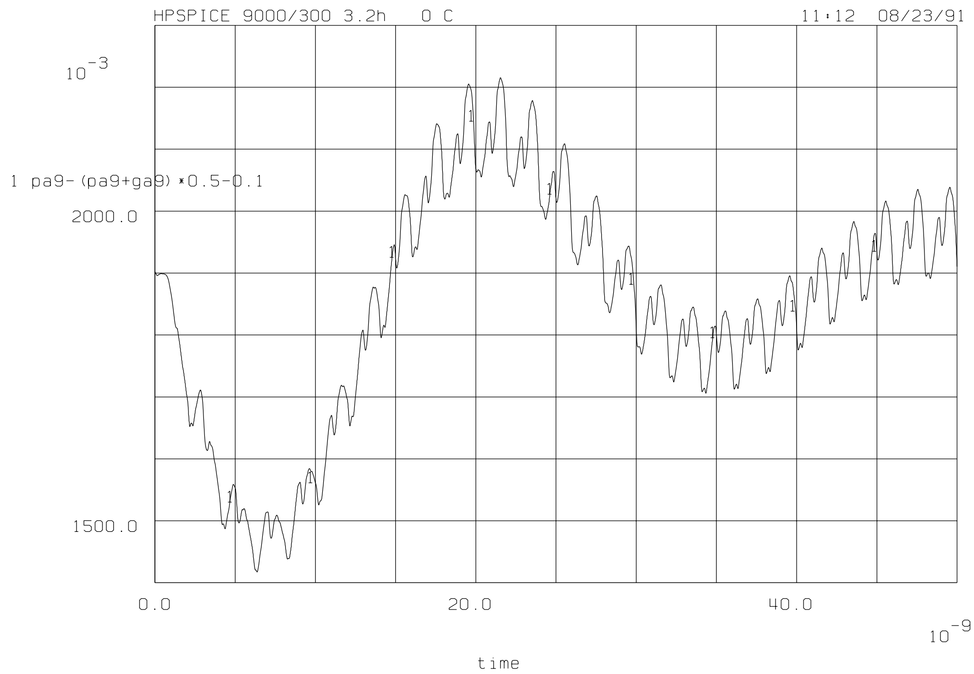

Figure 50. Noise margin with 10 nH common power supply inductance.

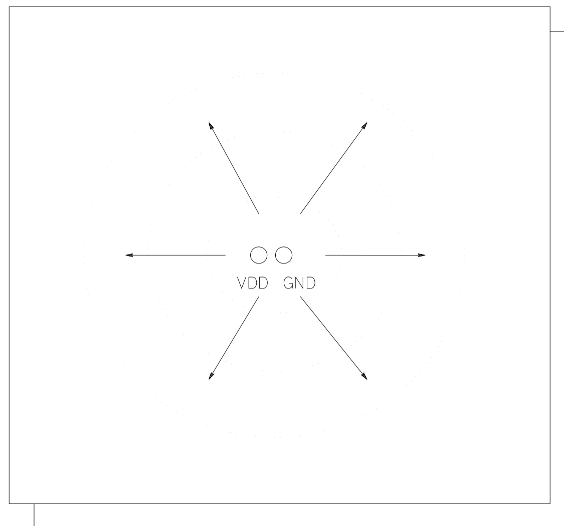

Figure 51. Noise propagation in power supply planes.

due to the loss in the conductor and dielectric and somewhat due to radiation at the edge. The initial pulse propagation near to the power/ground connection can be modeled as a radial transmission line if the power and ground connections are close

enough. Using this approach, the equivalent circuit model can be obtained for each power/ground connection. For a typical printed circuit board, with 114 μm thick FR4 dielectric and 356 μm diameter via connection, the equivalent circuit model is a simple parallel RL circuit with 6.4 Ω and 0.066 nH. This model has been verified with a high frequency measurement. Since this effective plane inductance is very small, the power supply impedance can be calculated only considering the inductance of the via connection. As mentioned before, the detailed derivation of the model and experimental verification will be presented in an HP technical report.

8 Delay

Noise margin and delay are the two measures chosen to evaluate the quality of a signalling system.Delay in a signalling path can be broken into three contributors: driver, receiver, and interconnect medium. A designer has many solution choices within each of these contributors. For this study at least two technologies were selected for each, chosen to be appropriate for a high speed VLSI central processing unit. A catalog of delay values was built for each technology and various combinations were tested for total delay in a hypothetical central processor with tight timing restrictions among three VLSI chips.

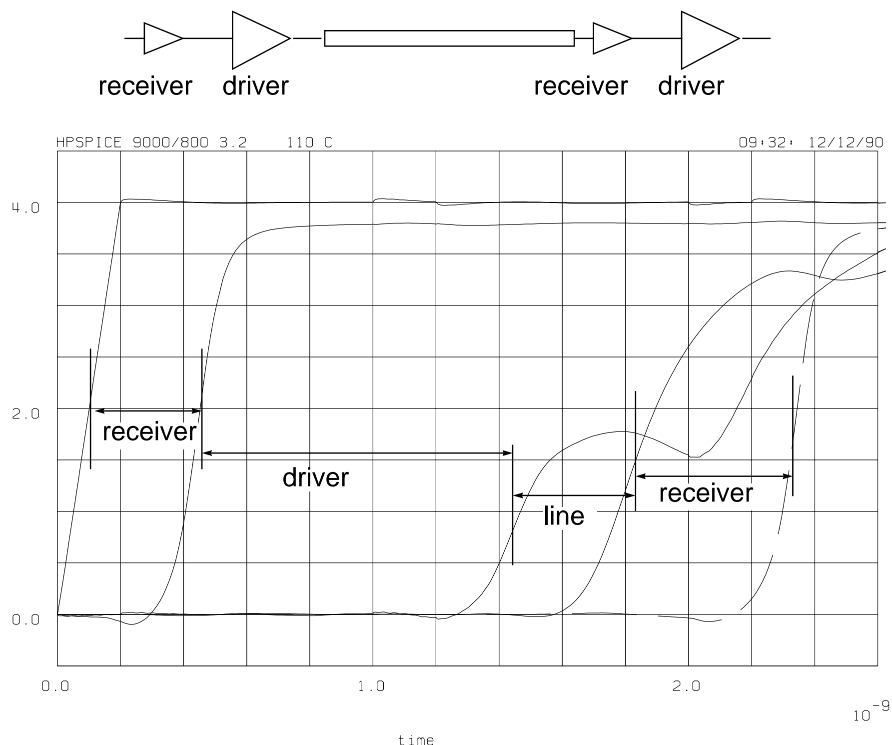

Figure 52. Delays extracted from SPICE results.

The technologies chosen were:

Drivers

1. CMOS driver constructed with a cascade of inverters with 200 micron n-channel width in the output device.
2. BiCMOS driver constructed with a CMOS pre-driver and emitter follower output transistors.

Each driver could be connected single ended or in pairs differentially.

Receivers

1. A single ended CMOS inverter.
2. A differential emitter coupled bipolar pair.

Each receiver could be connected to either of the drivers resulting in different delays due to the difference in output swing of the two drivers.

Interconnection

1. Thin film copper lines on polyimide dielectric.
2. Tungsten lines on co-fired alumina dielectric.
3. Copper lines on FR4 printed circuit board dielectric.

Many other technologies choices are possible. For example ECL or CML driver configurations, GaAs, Schmitt trigger receivers, stripline or offset stripline interconnections. The methodology of building models described here is easily adaptable to any of them. Reference [BAKO90] discusses driver and receiver circuit options and Reference [TUMM89] contains very good coverage of many transmission physical options.

The delay values were determined using SPICE models described in Section 3. The method is indicated in Figure 52. Components being measured are always embedded between other components rather than SPICE voltage sources or open circuits. The reason can be seen by examining the delay difference between the two receivers in Figure 52. The first exhibits lower delay due to the higher speed and zero impedance of the driving voltage source compared to the rounded waveform and finite impedance driving the second receiver. Likewise, a transmission line terminated with the capacitive load of a receiver with ESD protection will exhibit more delay than the same length transmission line driving a matched load, such as another transmission line. Data was taken for both worst case and best case semiconductor models as well as high and low temperature and power supply voltages. In general, one's interest is in the worst case situation but this extra data can aid evaluation of the benefits of better cooling, more benign ambient temperatures, and better power supply regulation.

The example module chosen to analyze is shown in Figure 53 with all chips shown mounted on carriers and the carriers in turn on a substrate. The transmission path analyzed is the round trip circuit from MCC to one SRAM to CPU back to MCC. The

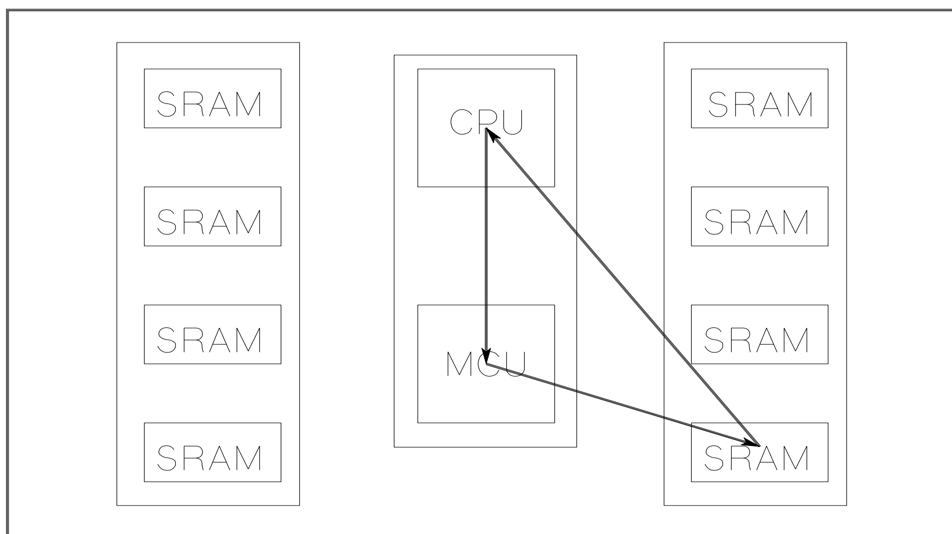

Figure 53. Example path for delay analysis.

library of delays is shown in Table 1. Three interconnect combinations were analyzed with both CMOS and BiCMOS drivers using differential signalling. In the first case copper/polyimide interconnections were used for carriers and substrate; in the second case co-fired alumina; and in the third copper/polyimide for carriers and co-fired alumina for substrate. The results are tabulated in Table 2 and summarized graphically in Figure 54.

Two conclusions can be drawn from this data. First, in recent years interconnection delay has come to dominate over electronics delay in typical workstation processors. This is because large PGA style packages with single chips connected by low density printed circuit boards resulted in relatively long interconnections, growing even longer as I/O counts and sizes of packages grew. At the same time improving IC technology steadily reduced the electronics delay component. In the future, as workstations move to dense, highly integrated multichip modules, the situation will be reversed again with electronics delay dominating over interconnect delay. In this example, even with the most aggressive driver design, electronics delay is more than double the interconnection delay. Thus, future systems will be able to reap the benefits of at least two more generations of IC speed improvement without having them erased by excessive interconnection delays.

The second, related point is the surprisingly small influence that dielectric constant has on overall delay. A pure alumina transmission line has 66% more delay than an equal length line in polyimide due simply to their dielectric constant difference (9.6 vs. 3.5). Overall delay is dominated by electronics delay as discussed above.

	Worst Case	Delta per C	Delta per V	Best Case
Drivers				
BiCMOS	310			170
CMOS	990	2.1	-230	400
Receivers (CMOS)				
Single Ended				
With BiCMOS Driver	550	1.2	-130	230
With CMOS Driver	490	1.0	-110	200
Differential				
With BiCMOS Driver	320	0.7	-58	170
With CMOS Driver	300	0.8	-51	150

Package Components

	Offset (pS)	Slope (pS/cm)
Cu-PI on Ceramic line (another line as load)	7	71
Cu-PI on Ceramic line (receiver input C=2pf load)	105	74
Substrate-Substrate transition with 30 mil via	9	0
PCB line (another line as load)	0	66
Co-fired Ceramic line (another line as load)	0	116
Co-fired Ceramic line (receiver input C=2pf load	105	121

Physical Distances (cm)

MCC Carrier Escape	1.08
SRAM Carrier Escape	0.95
CPU Carrier Escape	1.08
MCC-->SRAM	3.8
SRAM-->CPU	0.25
CPU-->MCU	0.25

Table 1. Library of delays for example. (all values in ps)

Polyimide		
CPU Paths (w/o electronics)		
MCC-->SRAM	536	536
SRAM-->CPU	284	284
CPU-->MCU	293	293
System Delays		
Total Cache SRAM Path (BiCMOS)	3003	2133
Total CacheSRAM Path (CMOS)	4983	2763

All ceramic		
CPU Paths (w/o electronics)		
MCC-->SRAM	786	786
SRAM-->CPU	375	375
CPU-->MCU	390	390
System Delays		
Total Cache SRAM Path (BiCMOS)	3441	2571
Total Cache SRAM Path (CMOS)	5421	3201

Polyimide carriers/ceramic substrates		
CPU Paths (w/0 electronics)		
MCC-->SRAM	700	700
SRAM-->CPU	288	288
CPU-->MCU	298	298
Systems Delays		
Total Cache SRAM Path (BiCMOS)	3176	2306
Total Cache SRAM Path (CMOS)	5156	2936

Table 2. Round trip system delays. (all values in ps)

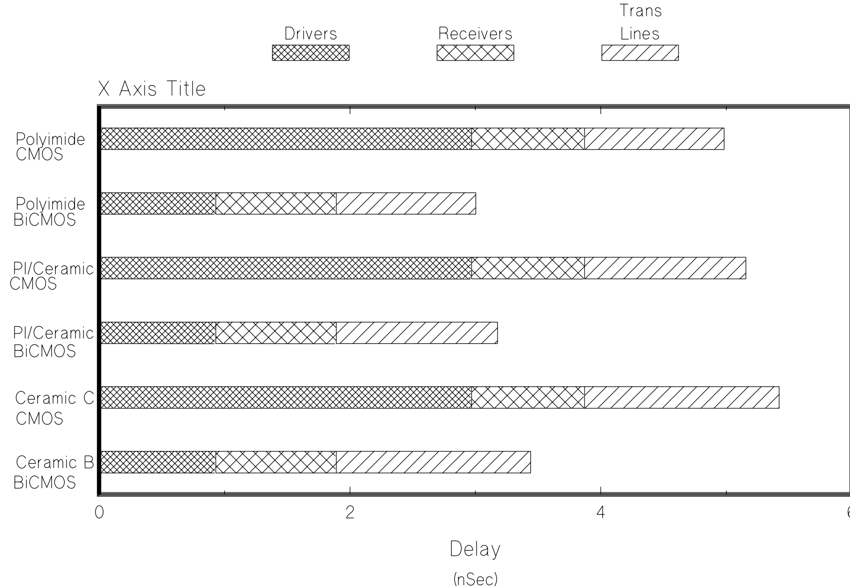

Figure 54. Round trip system delays.

Furthermore, even within the interconnection delay fully one third is time for the transmission line to charge the ESD and pad capacitance on the receiver end. Neither of these is affected by dielectric constant. The significance to the designer is that, with the short interconnections possible with multichip modules, one has more freedom in choice of dielectric materials. Considerations of thermal management (AlN has three orders of magnitude higher thermal conductivity than polyimide) or manufacturability (multi-layer alumina is lower in cost and more widely available than Cu-polyimide structures) may suggest a material with higher dielectric constant; the actual delay penalty to be paid may not be as high as the dielectric constant would suggest.

9 The Effect of All Contributions Combined, with Suggested Designs

As mentioned previously, the designer's choices of materials, processes, form, and dimensions in his interconnect structure each affect the electrical parameters of crosstalk, reflections, ground bounce and delay. In this section three interconnection designs will be analyzed in order to illustrate the methodology as well as to propose them as potential solutions to the high speed chip to chip interconnection problem. In the first two designs the chips are flip chip mounted on a carrier and the carrier in turn is flip chip mounted to a substrate. The transmission lines in the first of these two cases is low loss stripline in polyimide dielectric; in the second it is microstrip in a lossy, but less expensive, polyimide structure. The third design is similar to the first

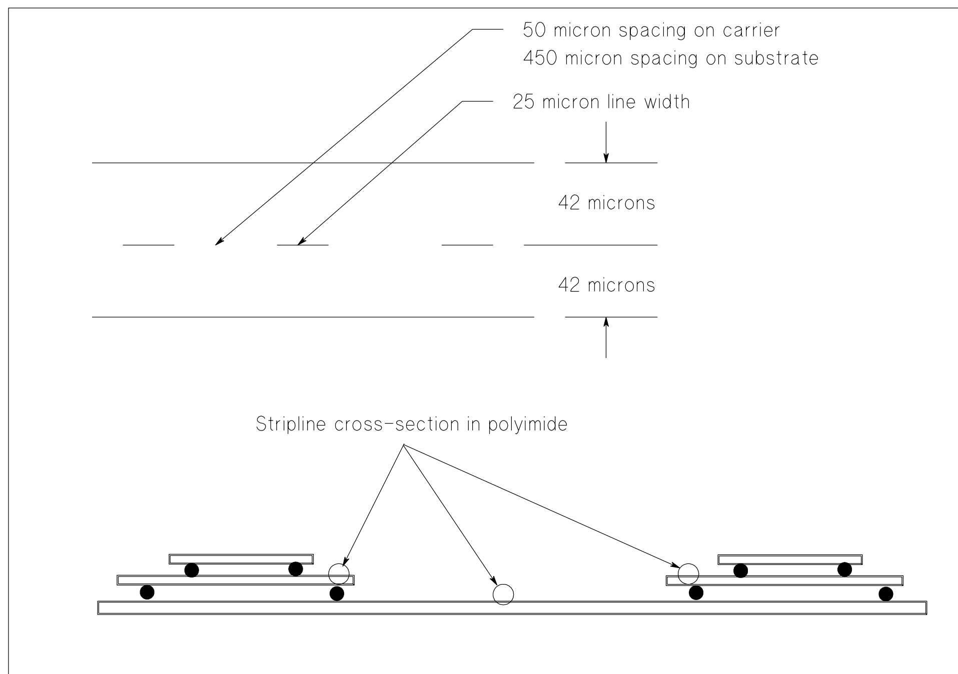

Figure 55. Design 1.

except the receiving chip on its carrier is right angle mounted to the substrate and electrically connected via a multi-layer flexible ribbon cable.

Design 1 is shown in Figure 55. The interconnecting transmission lines are of 6 micron thick copper embedded in polyimide dielectric with the dimensions shown. The line pitch is 3 mils on the carrier to escape from a dense VLSI chip and a wider 10 mils for lower crosstalk in the longer substrate runs. These are typical dimensional capabilities of at least two substrate vendors (NTK and Kyocera) and represent a relatively low loss, though expensive, structure. The simulation model includes, besides the transmission lines, all the parasitics of the solder bumps, vias, drivers and receivers. These models were described in Section 3.

The received pulse and noise margin are shown in Figure 56. Since interfering signals due to ground bounce, reflections, and crosstalk will all add in different phases at the receiver, it can be difficult to find the combination which provides the worst case noise margin. One effective way to do this is to vary the substrate transmission line length. Since crosstalk, ground bounce and reflections are each travelling different multiples of this length, their relative phases will vary and at some point they will add in the worst case phase. Figure 57 shows the results of multiple simulations each at different substrate line lengths (the carrier line length is kept constant at 1.1 cm each). Furthermore, the analysis is done with extreme values of source termination resistor, 25 and 40 Ω in this case. The results of Figure 56 are for the worst case of 25Ω source resistor and 4 cm line length.

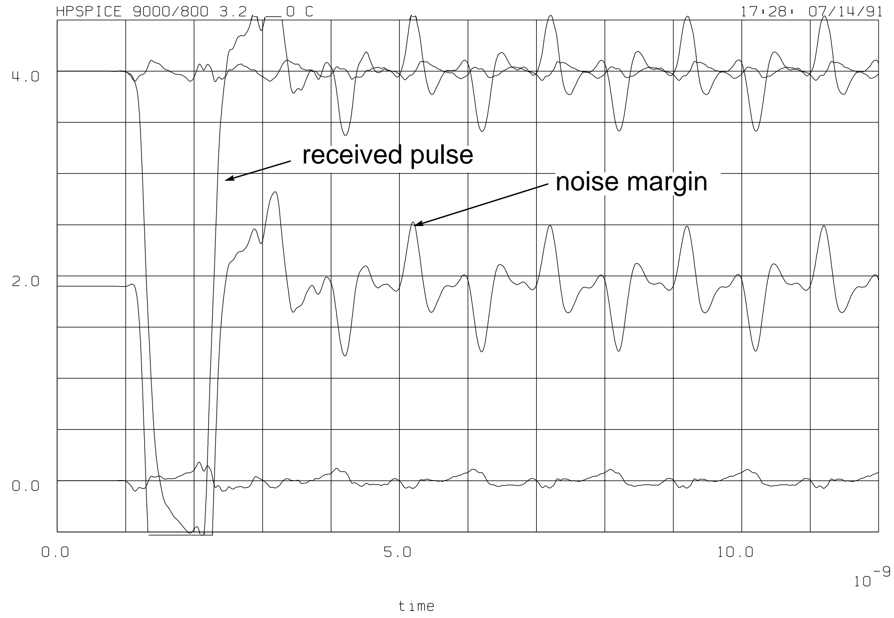

Figure 56. Design 1 received pulse and noise margin.

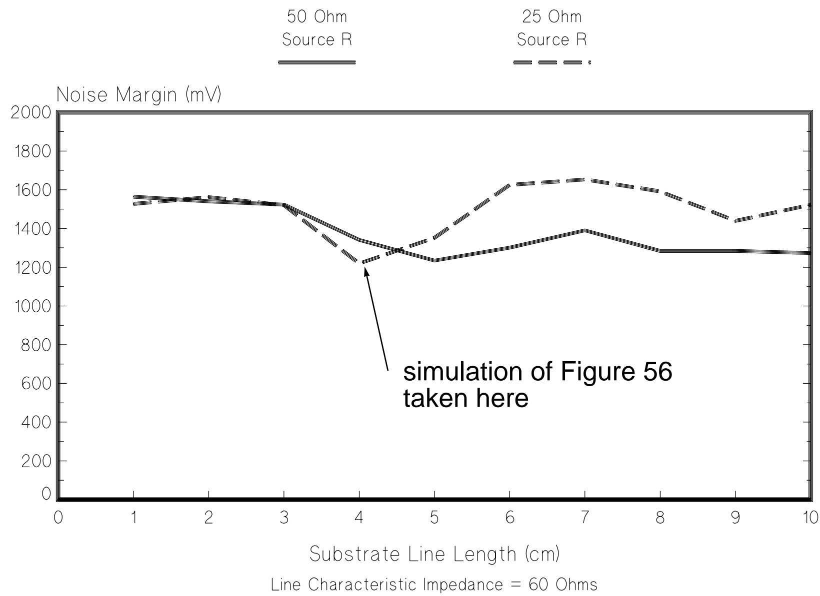

Figure 57. Design 1 noise margin degradation vs. substrate line length.

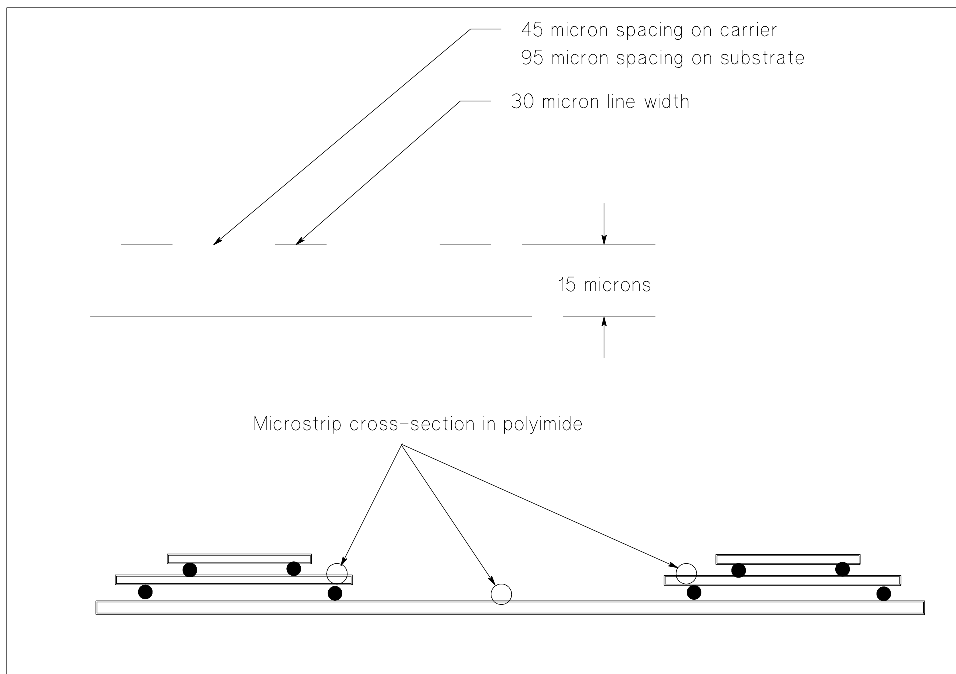

Figure 58. Design 2.

Design 2 is shown in Figure 58. It is identical to Design 1 except for the geometry of the transmission lines on the carriers and substrates. The conductor and dielectric thicknesses are less and 50 Ω microstrip is used instead of 60 Ω stripline. The effect of the thinner conductor (approximately 2 microns) is a high series resistance (approximately 6 Ω per centimeter) yielding a lossy line requiring greater care in design (more on this later). The thinner dielectric forces a narrower conductor for the same characteristic impedance; this in turn worsens the lossy line problem. However, the thinner dielectric allows closer spacing of conductors for the same crosstalk, enabling tighter pitch interconnections. The choice of microstrip instead of stripline is compelled by the thin dielectric. Stripline would require a much narrower conductor, too small to manufacture and of even higher loss. Fifty Ω is the highest characteristic impedance practically achievable in this structure. Geometries similar to this are realizable in products of several vendors, including Alcoa, IBM, and Pacific Microelectronics. If one can design around the shortcomings of its lossy lines, it will be substantially less expensive than Design 1.

The lossy line attenuates both the desired signal and the interfering crosstalk and reflections. Therefore, as described in Section 6, one would choose lower values of source termination resistor, 10 Ω in this case. Figure 59 shows the noise margin performance of this design for the worst case line length (5 cm) and a 7 Ω source termination resistor. Figure 60 shows the noise margin for the full range of line lengths and for 7 and 14 Ω source termination resistors.

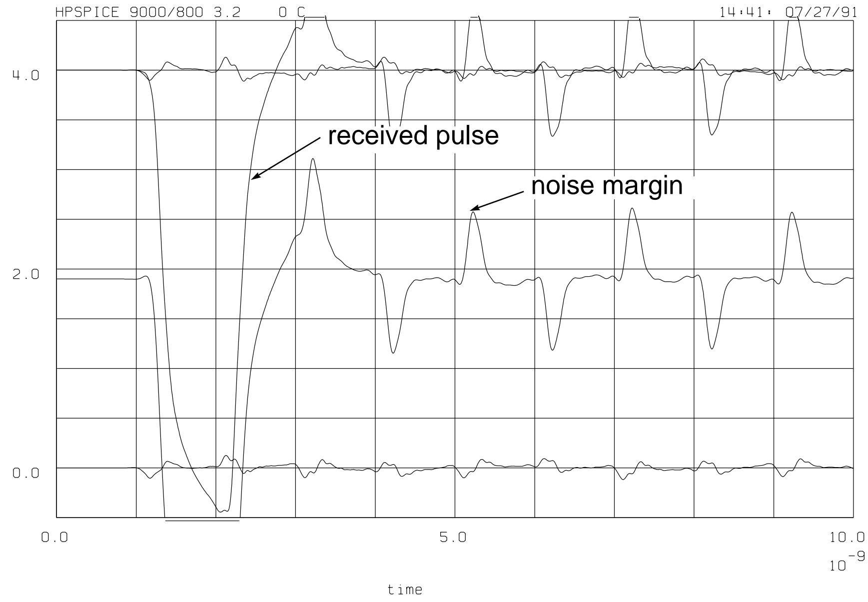

Figure 59. Design 2 received pulse and noise margin.

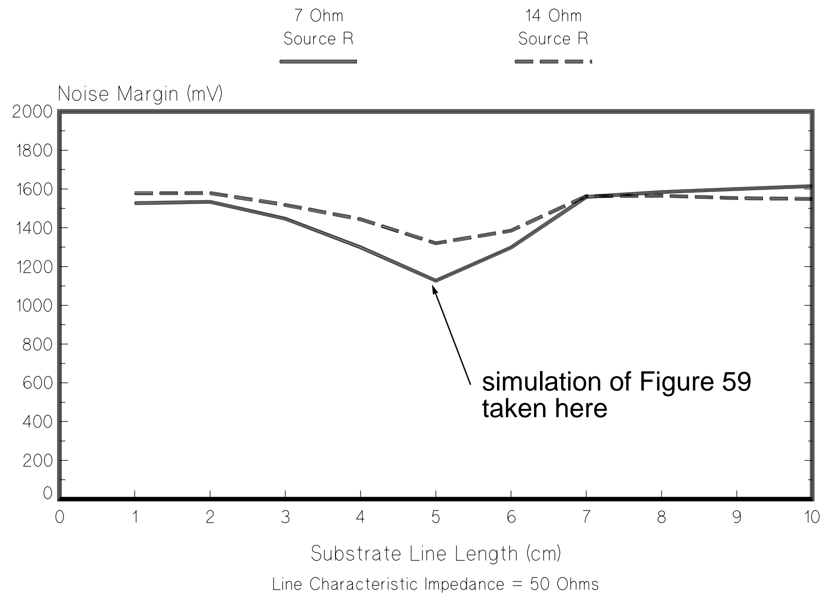

Figure 60. Design 2 noise margin degradation vs. substrate line length.

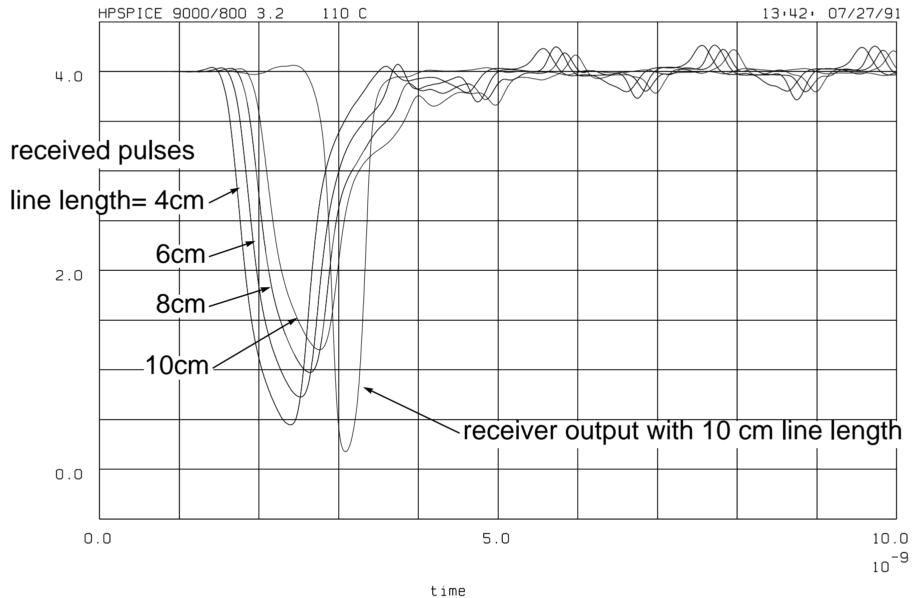

Figure 61. Figure 7: Received pulses for lossy transmission lines.

The simulations considered above were done with low temperature and “best case”, i.e. fast, device models for the reasons described in the introduction. However with a lossy line one must also consider the reliability of pulse transmission under worst case device conditions. A weak transmitted pulse could be attenuated enough in the transmission line to have difficulty triggering the receiver. This situation is shown in Figure 61 for a 14 Ω source termination resistor and four line lengths of 4, 6, 8 and 10 cm. Due to the temperature, device models, high value of source termination resistor, and line attenuation the received pulse does not swing all the way to ground even for a 4 cm line. Likewise the interfering signals from ground bounce, crosstalk and reflections are substantially smaller than those of Figure 59. While the receiver does successfully switch even for a 10 cm length line, its width is substantially narrowed. The conservative designer would limit use of this design to maximum line lengths of 6 cm.

Design 3 uses the same components as Design 1 except that the receiving end carrier is mounted vertically as shown in Figure 62. The interconnection between the substrate and vertically mounted carrier is with striplines in a multilayer kapton structure mechanically mounted with a BetaPhase apparatus. The electrical model for this transmission line was described in Section 3.

The noise margin performance for Design 3 is shown in Figure 63 and the performance for various line lengths and source termination resistor value is shown in Figure 64. Notice that the performance of this design is every bit as good as Design 1, yet this offers easy removal and replacement of circuits on the receiving end.

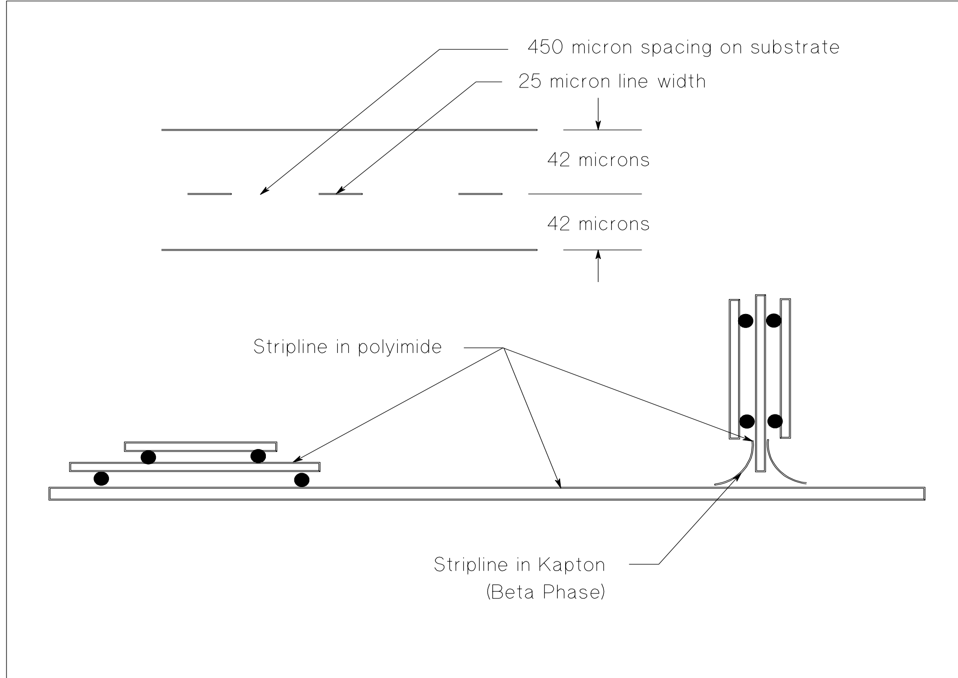

Figure 62. Figure 8: Design 3.

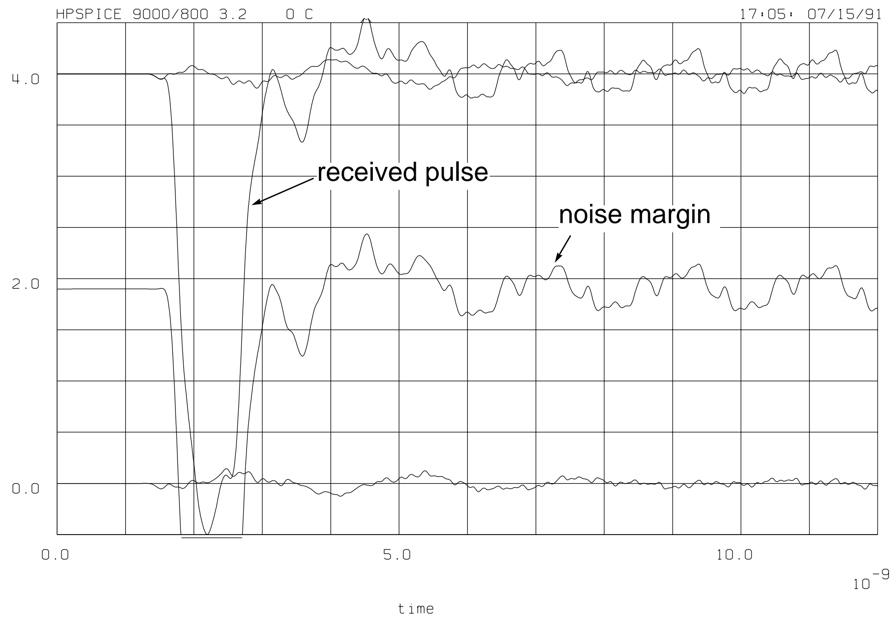

Figure 63. Design 3 received pulse and noise margin.

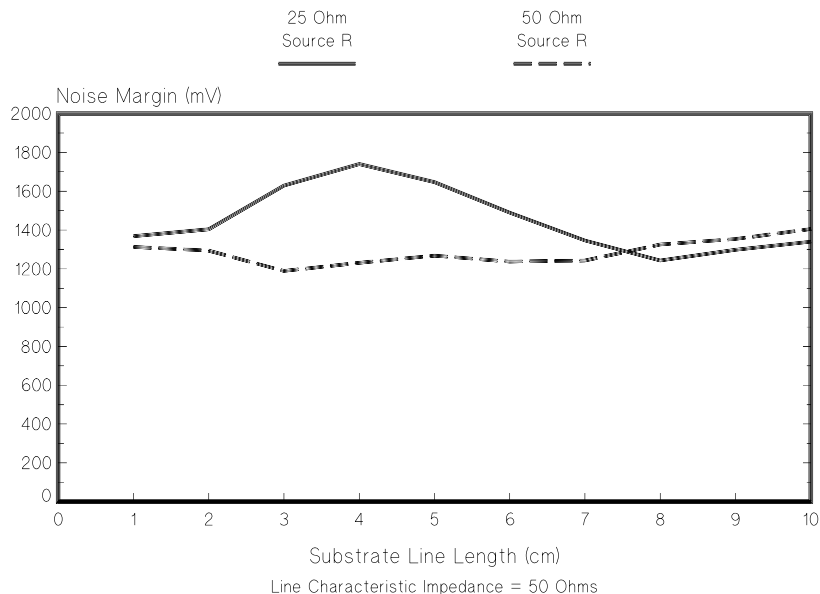

Figure 64. Figure 10: Design 3 noise margin degradation vs. substrate line

Other issues enter into transmission line choices, such as Electromagnetic Compatibility and manufacturability. See Reference [OTT88] for a good discussion of the former, Reference [TUMM89] for the latter.

References

[BAKO90] H. B. Bakoglu, “Circuits, Interconnections, and Packaging for VLSI,” Addison- Wesley, 1990.

[CHAN91] N. Chang, K. J. Chang, S. Y. Oh, and K. Lee, “Interconnect Performance Design Assistant,” to be published.

[FELL65] A. Feller, H. R. Kaupp, and J. J. Digiacomo, “Crosstalk and Reflections in High-Speed Digital Systems,” Proceedings - Fall Joint Computer Conference, pp. 511-525, 1965

[HEWL1203] “Modeling Passive High-Speed Digital Structures”, Hewlett Packard Application Note 1206.

[LAU89] Wai Yuen Lau, “Network Analysis Verifies Models in CAD Packages”, Microwaves & RF, November, 1989.

[LEE88a] K. Lee, “FCAP2 Manual”, HP Internal Document, Jul. 1988.

[LEE88b] K. Lee, “FCAP3 Manual”, HP Internal Document, Sep. 1988.

[LEE89] K. Lee, “Three-Dimensional Inductance Program,” Hewlett-Packard Design Technology Conference, May, 1989.

[LEE91] K. Lee, “An Electrical Model for Power and Ground Planes,” to be published.

[LIU90] R. Liu and K. Lee, “Interconnect Model Library - A Database Program for Electrical Models of Interconnects,” Hewlett-Packard Design Technology Conference, May, 1990.

[OTT88] Henry W. Ott, “Noise Reduction Techniques in Electronic Systems,”, John Wiley and Sons, 1988.

[RAMO84] Ramo, Winnery and Van Duzer, “Fields and Waves in Communications Electronics,” John Wiley and Sons, 1984.

[TUMM89] Rao Tummala and Eugene Rymaszewski, “Microelectronics Packaging Handbook,” Van Nostrand Reinhold, 1989.

[WEST85] Neil Weste and Kamran Eshraghian, “Principles of CMOS VLSI Design,” A Systems Perspective, Addison-Wesley, 1985.