

C. W. Ho
D. A. Chance
C. H. Bajorek
R. E. Acosta

The Thin-Film Module as a High-Performance Semiconductor Package

This paper discusses a multichip module for future VLSI computer packages on which an array of silicon chips is directly attached and interconnected by high-density thin-film lossy transmission lines. Since the high-performance VLSI chips contain a large number of off-chip driver circuits which are allowed to switch simultaneously in operation, low-inductance on-module capacitors are found to be essential for stabilizing the on-module power supply. Novel on-module capacitor structures are therefore proposed, discussed, and evaluated. Material systems and processing techniques for both the thin-film interconnection lines and the capacitor structures are also briefly discussed in the paper. Development of novel defect detection and repair techniques has been identified as essential for fabricating the Thin-Film Module with practical yields.

Introduction

Very Large Scale Integration (VLSI) in semiconductor technology is expected to result in significant improvements in device cost, performance, reliability, and function, and concomitant improvements in systems based on such devices. However, many of the benefits of VLSI will be lost without significant improvements in device packaging techniques. This paper first summarizes key VLSI trends in chip technology and their implications for device packaging, the limitations of current packages, and how these have been overcome by IBM's latest multichip modules based on Multilayer Ceramic Technology (MLC). The main body of the paper discusses a packaging approach being investigated in this laboratory, the Thin-Film Module, which features thin-film transmission lines and on-module decoupling capacitors integrated into the body of the module, improvements which are likely to be required to package high-performance VLSI devices in the late 1980s.

VLSI trends in chip technology

Projections for semiconductor chips [1] suggest that the number of circuits per chip for field effect and bipolar transistor logic chips and memory bits per chip will continue to double approximately every two years during the 1980s.

As the number of circuits on a logic chip increases, the ratio of input/output connections to the number of circuits decreases. This improvement, however, is not sufficient to avoid the need for a substantial increase in the total number of input/output connections to a group of circuits on a chip. As image size decreases, power per circuit decreases, but the countervailing trend of increasing integration results in a significant increase in total chip power to levels as high as 20 W/chip in high-performance applications. Similarly, although current per circuit also decreases, the total current that needs to be delivered to a chip also significantly increases, and could approach the 10-ampere level in high-performance bipolar devices. The need to simultaneously switch an ever-growing number of circuits on a chip at shorter pulse rise times and higher total currents exacerbates problems related to maintaining adequate signal/noise ratios due to chip and package inductance.

Chip and package inductance can give rise to what is called simultaneous switching noise. This inductance is particularly critical in the power supply distribution to off-chip driver circuits typically used in high-performance digital systems (Fig. 1). As depicted in this figure, the

Copyright 1982 by International Business Machines Corporation. Copying is permitted without payment of royalty provided that (1) each reproduction is done without alteration and (2) the *Journal* reference and IBM copyright notice are included on the first page. The title and abstract may be used without further permission in computer-based and other information-service systems. Permission to *republish* other excerpts should be obtained from the Editor.

simultaneous switching of many such drivers can cause fluctuation of the voltage level at the individual driver transistors by an amount ΔV :

$$\Delta V = LN \frac{\Delta I}{\Delta t}, \quad (1)$$

where L is the effective chip and package inductance, N the number of drivers switched simultaneously, ΔI the current switched by each driver, and Δt the current rise time. Fluctuation of the power supply voltage level can cause delay in the rise time of interchip signals in active nets and, if excessive, can cause noise in quiet nets which can lead to false switching of receivers connected to the quiet nets. The latter effect can cause errors in computation.

Future VLSI drivers are expected to have ever-increasing switching speeds (shorter rise times) and increasing numbers of drivers switched simultaneously per unit area of chip or package, while maintaining approximately constant driver switching current. Referring to Eq. (1) and Fig. 1, containment of simultaneous switching noise to acceptable levels will therefore require significant reductions in chip and package inductance.

Implications for device packaging

The trends discussed in the previous section have direct implications for packaging of the devices. Future packages must be capable of providing corresponding increases in the number of input/output connections and in the wiring capacity to interconnect chips. Sophisticated cooling techniques and reduced package inductance will also be essential. Attainment of maximum performance for a given system design and circuit family calls for minimizing signal propagation delay by minimizing interchip distances and the dielectric constant of interchip and intermodule transmission lines. Improving the reliability and reducing the cost of future systems requires minimizing the number of physical interconnections between chips.

Continuing the present practice of containing the function of chip packages to that of a space transformer, packaging one chip to a module, exclusive use of additional package levels, cards and cables to provide intermodule (interchip) wiring, and reliance on peripheral input/output connections to the chip and module will not permit the necessary improvements in chip and package inductance and input/output connections and will result in a loss of many of the benefits of VLSI.

A very attractive packaging approach which can provide many of the improvements needed to capitalize on VLSI is IBM's recently announced multichip MLC-based module approach (Fig. 2) [2, 3]. This technology makes it possible to package as many as one hundred logic chips on a single

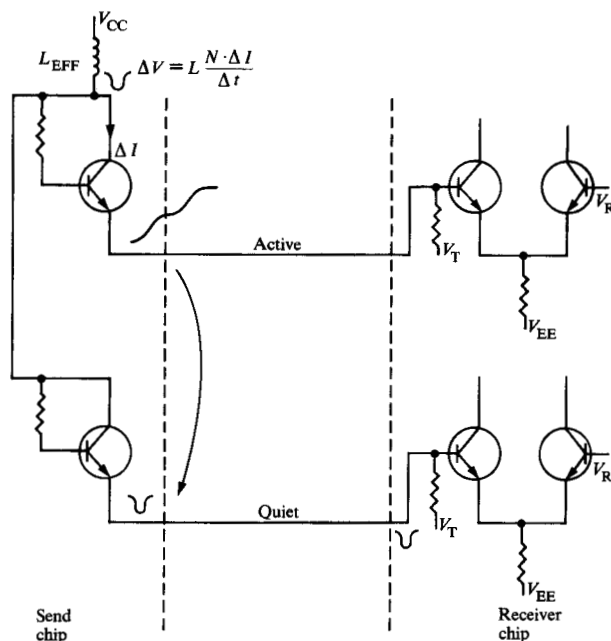


Figure 1 Simultaneous switching noise.

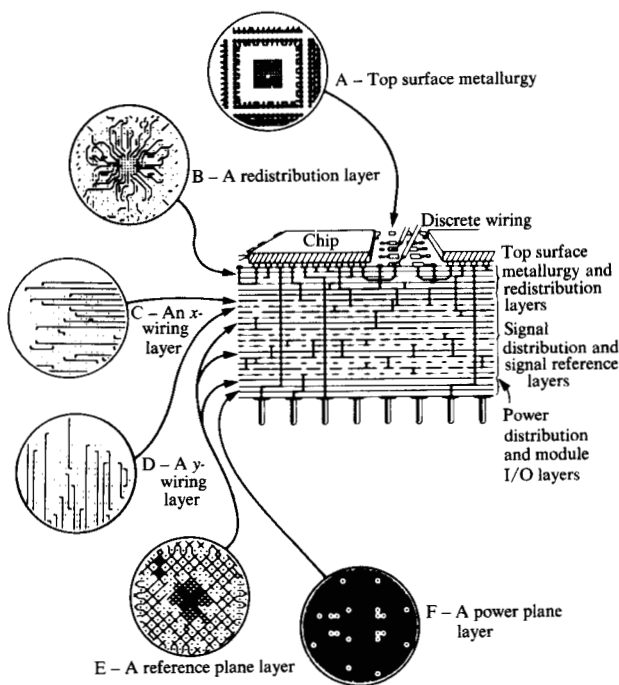


Figure 2 MLC construction detail.

ceramic substrate. At present levels of device integration, a single substrate can support as many as 60 000 logic circuits. This substrate no longer serves the simple role of a space expander; it also provides for permanent chip interconnec-

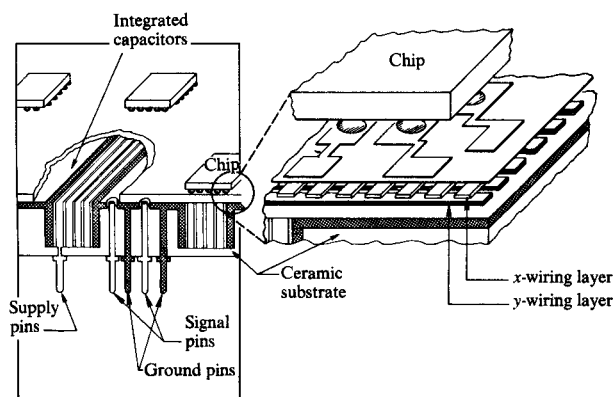


Figure 3 The Thin-Film Module.

Table 1 Impact of VLSI trends on package.

	Present	1985-1990
<i>Semiconductor</i>		
Lithography ground rules (μm)	2.0	1.0
Chip size (mm)	4.6	4.0
No. of circuits	600	2500-5000
No. of signal contacts	96	289-400
Voltage swing (V)	1.1	0.5-1.0
Rise time (ns)	1.1	0.2-0.4
<i>Package (normalized multipliers)</i>		
Package inductance		
Rise time effect	1.0	3.5
Delta I (switching current)	1.0	2.0
Noise intolerance	1.0	2.0
	1.0	14
Wiring density	1.0	3.0-4.0
Cooling	1.0	2.0-3.0

tions throughout the entire chip area which until now had to be made at the card level. It reduces the number of input/output connections at this level, the overall wire lengths (resulting in shorter time of flight between groups of circuits), and the power, while improving performance at the system level. The reduced number of interconnections between dissimilar packaging levels results in higher reliability and also makes it possible to improve system performance and cost by savings at the card, board, cabling, power supply, and frame levels. It also features means of performing engineering changes via discrete wires bonded to metal pads surrounding each chip site.

The extension of such multichip ceramic modules to accommodate more chips with larger numbers of input/output connections to the chip and to the module, and the

adaptation of this technology to semiconductor and magnetic bubble memory devices, are expected to make it the leading device package candidate of the 80s. However, further improvements will be required to serve the needs of VLSI devices expected beyond this time frame.

The level of improvement likely to be required in critical package functions is summarized in Table 1. The first part of Table 1 compares bipolar chip parameters of present devices with those expected toward the end of this decade. The second part compares the corresponding level of improvement that must be attained in packages to support such devices. Specifically, the rise times of driver circuits are likely to decrease by a factor of 3 to 4. The total number of drivers switching simultaneously is expected to increase by a factor of 4 as the current level per driver decreases to approximately one half of present devices, resulting in a factor of 2 increase in the total current switched simultaneously per chip. The reduced voltage level will decrease noise margins by a corresponding amount. These changes will require reducing future package inductance by more than an order of magnitude below that of present packages. The 4- to 10-fold increase in the number of circuits per chip is expected to require a 3- to 4-fold improvement in wiring capacity, and improvement in cooling capability by a factor of 2 to 3 beyond that of present packages.

The Thin-Film Module

One approach being investigated in this laboratory to achieve the aforementioned package improvements for high-speed digital systems is the Thin-Film Module (Fig. 3). It features the use of multilayer photolithographically defined interchip wiring and the use of decoupling capacitors integrated into the body of the substrate to control simultaneous switching noise. The combination of thin-film lines and integrated decoupling capacitors offers a very good potential to reduce the effective package inductance by more than an order of magnitude below that of present packages.

The following sections describe our efforts to date to understand the electrical characteristics of thin-film transmission lines and of power distribution networks which incorporate decoupling capacitors in the body of the substrate, as well as the materials and processes required to achieve the necessary structures.

• Thin-film lines

The majority of wires for chip interconnections in modules, cards, and boards used in modern digital systems are strip transmission lines. IBM's latest MLC multichip modules also use such strip lines. Figure 4(a) depicts the cross section of the strip lines in the MLC modules. The dimensions shown are also representative of strip lines in the cards and boards which carry such modules. The cross-sectional area

of strip lines used to date, even with use of relatively resistive metals such as Mo in the ceramic modules, has resulted in transmission lines with small resistive loss, thereby enabling interchip and intermodule signal propagation with negligible signal distortion. As discussed in the previous sections, future VLSI trends demand that future packages provide significant improvement in interchip wiring capacity. This objective can be attained via a combination of miniaturization of strip transmission lines to improve line density and use of additional wiring planes to improve total line capacity.

Figure 4(b) depicts the cross section of thin-film strip lines which could be fabricated with modification of state-of-the-art photolithographic processes used in chip fabrication. However, consideration of practical dielectric and metal material properties, dimensions, and processes used for such strip lines readily suggests that thin-film strip lines with characteristic impedance of 50Ω will have significant resistive loss. For example, a $50\text{-}\Omega$ strip line of $5 \times 9 \mu\text{m}$ of Cu will have a resistance per unit length of $4 \Omega/\text{cm}$. On the other hand, the potential for a 20-fold improvement in wiring density relative to the lossless case is very attractive *vis-à-vis* achieving significant wiring capacity while using as few as two wiring planes in future VLSI packages. The latter potential motivated us to attempt to understand the extent to which thin-film lossy strip lines could be used to provide the chip interconnection function in future packages.

Pulse propagation, coupled noise, and dispersion of high-speed pulses in various configurations of lossy thin-film strip transmission lines have been modeled and experimentally investigated in our laboratory. The details of these studies are reported in [4]; the main conclusions are summarized below.

• Pulse propagation on lossy transmission lines

High-speed pulse propagation on uniform lossy transmission lines is characterized by the general transmission line equations [5-8]

$$\begin{aligned}\frac{\partial v}{\partial x} &= Ri + L \frac{\partial i}{\partial t}, \\ \frac{\partial i}{\partial x} &= C \frac{\partial v}{\partial t},\end{aligned}\quad (2)$$

where R , L , and C are resistance, inductance, and capacitance of the line per unit distance. No assumption on the magnitude of R is made; however, the conductance G through the dielectric material is known to be small for most practical dielectric materials and is therefore neglected here.

The solution for an infinitely long lossy line characterized by Eq. (2) in the frequency domain is

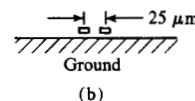
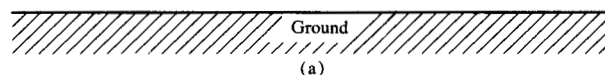
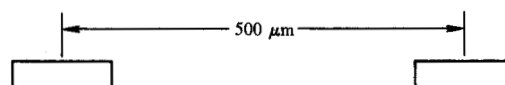


Figure 4 Strip transmission line dimensions for (a) lossless and (b) thin-film lines.

$$V(\ell, s) = e^{-\gamma(s)\ell} V_{in}(s), \quad (3)$$

where

$$\gamma(s) = \sqrt{(R + sL)sC} = s\sqrt{LC} \sqrt{1 + R/sL} \quad (4)$$

is the propagation constant, $V_{in}(s)$ is the Laplace transform of the input voltage, and ℓ is the distance from the source $V_{in}(s)$ to the point of interest on the line. It can be shown [9] that the inverse transformation of (3), which is the time domain solution of the line for a step input $V_{in}(s) = 1/s$, is

$$v(\ell, t) = e^{-R\ell/2Z_0} U(t - \ell\sqrt{LC}) + g(\ell, t) U(t - \ell\sqrt{LC}), \quad (5)$$

where the first term is an attenuated step function and the second term, represented by $g(\ell, t)$, is a rather complicated mathematical expression which is slow-rising like an RC circuit. The function $U(t)$ is a unit step function. The lossy line therefore behaves like an LC line and an RC line combined. A typical solution is plotted in Fig. 5. In this paper, we are primarily interested in the high-speed pulse propagation, and hence we concentrate on the fast-rising exponential term in Eq. (5).

When a pulse propagates on a lossy transmission line, it is attenuated exponentially, as described previously. When the attenuated pulse hits an open circuit, however, the current drops to zero. A reflection starts to propagate in the opposite direction, which also causes the voltage to double at the receiving end. If the line length, the line resistance, and its characteristic impedance Z_0 are within a certain range, the reflection can restore the lost pulse amplitude and the pulse rise time to a large extent, and yet not cause excessive ringing and distortion at the receiving end. Information can

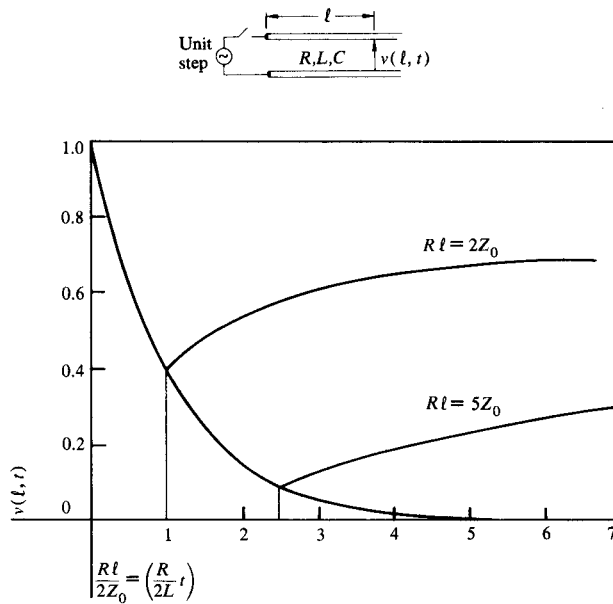


Figure 5 Solutions to general lossy transmission line equations:

$$\frac{\partial v}{\partial x} = L \frac{\partial i}{\partial t} + Ri,$$

$$\frac{\partial i}{\partial x} = C \frac{\partial v}{\partial t},$$

$$v(l, t) = \left[e^{-Rl/2Z_0} + \frac{Rl}{2Z_0} \int_{t-l\sqrt{LC}}^t \frac{e^{-Rt/2L}}{\sqrt{t^2 - (x\sqrt{LC})^2}} \times I_1 \left[\frac{R}{2L} \sqrt{t^2 - (x\sqrt{LC})^2} \right] dt \right] u(t - l\sqrt{LC}).$$

therefore be transmitted on the lossy line for receivers which have high input impedance compared to Z_0 attached to the open end of the line.

It is easily seen that if the product Rl is too large, the exponential term in the solution of Eq. (5) becomes insignificant and the line behaves practically as an RC line. However, if Rl is very small, such that it approaches the lossless case, assuming that the pulse rise time is still small compared with the total line delay $l\sqrt{LC}$, the line behaves like an LC line, and sustained reflections make meaningful transmission of information very difficult unless a clamp circuit is used to suppress the reflections. It can be shown that if we choose the total line resistance in the range

$$\frac{2Z_0}{3} \leq Rl \leq 2Z_0, \quad (6)$$

the pulse shape received at the open end is very similar to the transmitted pulse. For a uniform line, $Rl = 2Z_0$ establishes

the maximum line length $l_{\max} = 2Z_0/R$ and $Rl = 2Z_0/3$ establishes the minimum line length $l_{\min} = 2Z_0/3R$. In practical situations, it is often the case that the pulse rise time is not so small compared with the delay of the minimum length lines; i.e.,

$$t_{\text{rise}} \geq 2l_{\min} \sqrt{LC}. \quad (7)$$

In such cases, the reflections on a line whose length is shorter than $l_{\min}$ do not cause too much pulse distortion, since before the pulse can have a chance to double itself, the second reflection from the low-impedance voltage source starts to come in and pulls the pulse in the other direction. Hence the minimum line length condition is not critical, because the pulse shape is in general maintained at the receiving end. In those cases, it is important to see that the open-circuit lossy line can be used to transmit high-speed pulses with virtually the same speed as the lossless line. The attenuated pulse amplitude is restored due to the doubling effect, subject only to the condition that a maximum line length is not exceeded. For example, in a 50- Ω line system, with a line resistance per unit length of 4 Ω/cm , which can be readily achieved with $5 \times 9\text{-}\mu\text{m}$ copper lines, the maximum line length for pulse propagation with small distortion is 20 cm.

The preceding analysis thus shows that lossy thin-film transmission lines can, within certain limits, be used to propagate high-speed pulses with low distortion for distances of up to 20 cm, provided that such lines are terminated with an impedance much higher than the characteristic line impedance in combination with a clamp circuit to suppress reflections in short lines. Two important attributes of such thin-film lines in semiconductor packages are the attainment of high wiring density and the reduction of power dissipation due to elimination of the power dissipated in the terminating resistor of matched lossless transmission line designs.

A complete design of a transmission line structure also requires assessing coupled noise effects. The coupled noise characteristics of lossy transmission lines for both the near-end noise (NEN), which propagates toward the driver side of the quiet line, and the far-end noise (FEN), which propagates toward the receiver side of the quiet line, are to a first order identical to those of lossless designs, with the exception that loss exacerbates far-end noise since line resistivity increases the degree of inhomogeneity of lines above and beyond the inhomogeneities caused by crossover lines and by the vias which are needed to interconnect transmission line layers in lossy or lossless designs [4]. Coupled noise is also expected to depend strongly on specific line-to-line and line-to-ground-plane dimensions. We have evaluated several alternatives involving the use of one or more ground planes to meet the following requirements:

$$Z_0 = 50 \Omega,$$

$$l_{\max} = 20 \text{ cm},$$

$$\text{line pitch} = 25.4 \mu\text{m} \text{ (1000 lines per inch),}$$

$$\text{coupled noise} \leq 10\%, \quad (8)$$

where the line pitch is the center-to-center distance of two adjacent lines, and the coupled noise indicates the worst-case coupled voltage into a quiet line between two adjacent active lines switched simultaneously.

Figure 6 depicts the cross section of strip lines which meet the aforementioned requirements. All characteristic dimensions are shown in micrometers. The dielectric constant of the insulator is 3.5. The closed nature of this triplate design makes it ideally suited for use in a structure like the Thin-Film Module because it minimizes any electrical interaction between the fan-out layer shown in Fig. 3, needed to interconnect transmission lines with semiconductor chip terminals, and the x - y transmission lines. Connection between the fan-out layer and the lines and between lines requires vias of approximately $6.0 \mu\text{m}$ diameter. Representative signal responses and coupled noise for the indicated input voltage for various line lengths of this triplate design are shown in Fig. 7. The worst-case transient current amplitude is 11.2 mA for the input voltage amplitude of 600 mV . The propagation delay at the 50% signal level is 66.2 ps/cm and the worst-case coupled noise is 8.8% . The previously mentioned voltage oscillations at the high-impedance receiver can be seen in all cases, especially for the shortest line lengths. These reflections can be suppressed by means of a clamp network using Schottky barrier diodes at the receiver. The 20-cm line length case exhibits the beginning of the RC -type slowdown on the voltage incident at the receiver. Avoidance of this effect, which can cause undesirable delay in long lines, requires restricting the maximum line length to approximately 20 cm . This restriction is not expected to be serious since lines 20 cm long should enable wiring of large arrays of VLSI chips on multichip modules using thin-film transmission lines.

These results have been verified experimentally with a high-speed pulse sampling oscilloscope in conjunction with coaxial probes and thin-film line test samples specially designed to minimize discontinuities at the probe-to-sample interfaces [4]. The input pulse, the measured pulse, and a simulated pulse are shown in Fig. 8 for a $50\text{-}\Omega$ strip line with a conductor $12 \mu\text{m}$ wide and $5 \mu\text{m}$ thick. Figure 8 shows good agreement between the predicted and measured response.

Simultaneous switching noise

As mentioned previously, another key requirement in future VLSI packages is significant reduction of package inductance to minimize simultaneous switching noise. This noise

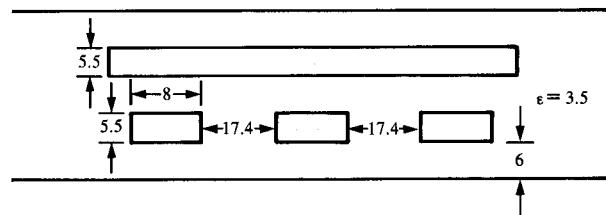


Figure 6 Triplate structure design for thin-film transmission lines. All dimensions are shown in micrometers. $\rho = 1.8 \mu\Omega\text{-cm}$, $R_0 = 4.1 \Omega/\text{cm}$, $L_{\max} = 2Z_0/R_0 = 20 \text{ cm}$, and $Z_0 = \sqrt{L_{22}/C_{22}} = 41 \Omega$.

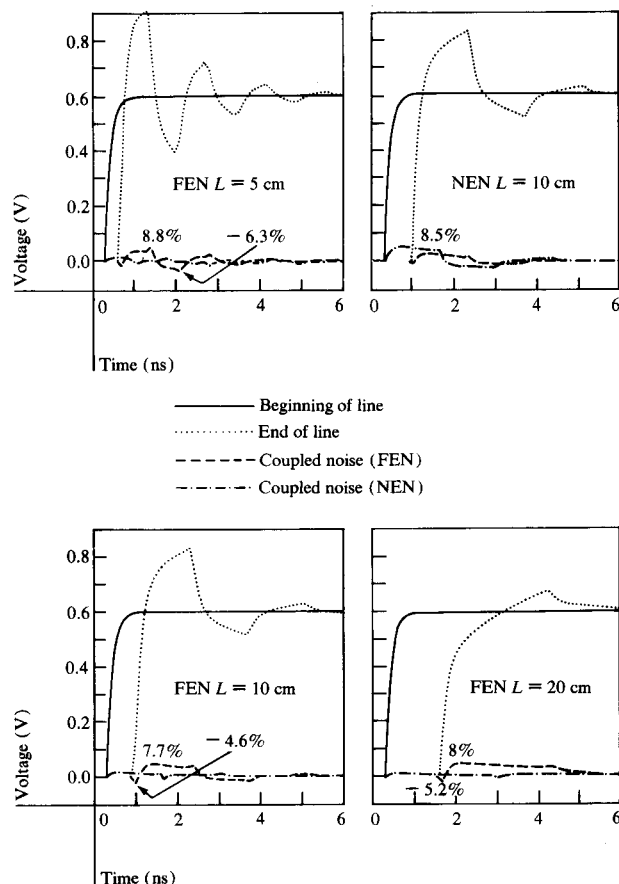


Figure 7 Signal and coupled noise responses for strip transmission lines of Fig. 6 for four line lengths.

arises when many input/output terminals in one VLSI chip attempt to communicate with other chips simultaneously during a specific instant of a machine cycle, and a large amount of current has to be supplied instantly to the chip through its power supply terminals by the package. The current is divided and comes out of the chip as signals from its drivers, through signal I/O terminals to the transmission lines on the package. If the package is incapable of supplying the current quickly, extra delays or even malfunction may

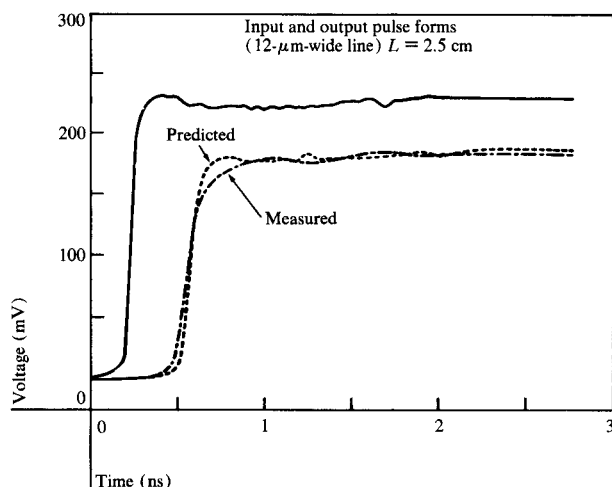


Figure 8 Comparison of theory and experiment for voltage response of 2.5-cm-long thin-film transmission line.

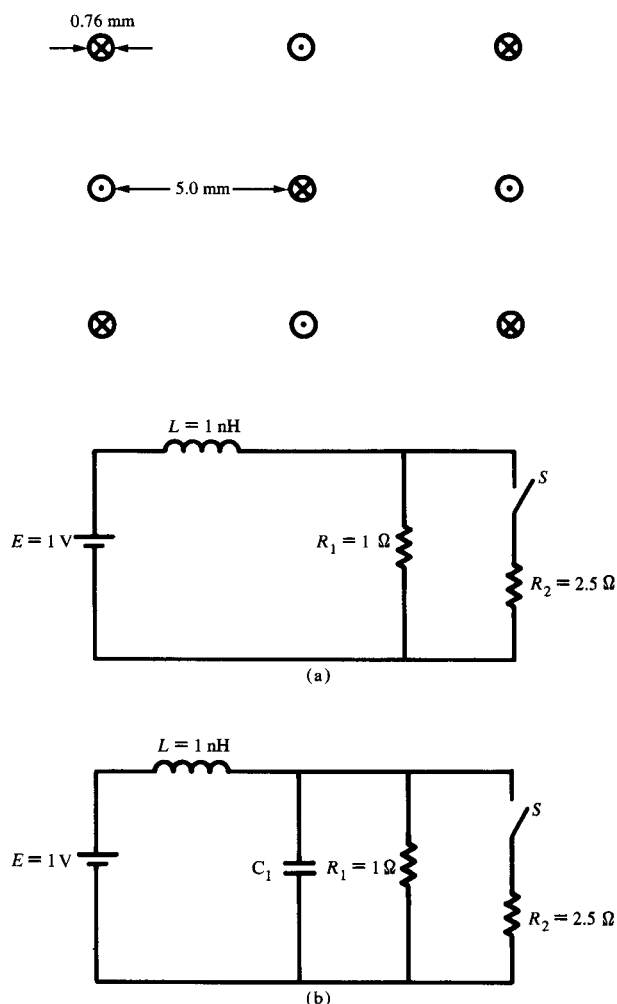


Figure 9 Package and chip power distribution equivalent circuits for modeling effect of (a) package inductance and (b) decoupling capacitor on simultaneous switching noise.

happen as a consequence. In this section a simultaneous switching example is used to illustrate the nature of the problem. The possible sources for supplying the driver currents and how an integrated-on-module capacitor design can be used as one possible solution to solve this important problem for the VLSI package are discussed.

Let us consider a set of on-module power supply pins, which bring power from the printed circuit board to the module, arranged in an array as shown in Fig. 9. If the pin-to-pin spacing and the length of the pin are 5.0 mm (200 mil), the total pin matrix inductance is calculated to be 1 nH. In the following discussions, 1 nH is used as the package inductance seen by the chip.

A typical bipolar chip draws current from the power supply for its internal circuits and external drivers. The power supply system of such a chip can be modeled with the equivalent circuit shown in Fig. 9(a), where R_1 represents the load due to the internal circuits, R_2 is the load of the drivers, and L is the package inductance. This example corresponds to a chip with internal circuit current of 1.0 A and with 20 drivers switching simultaneously into 50- Ω transmission lines which draw an additional 0.4 A at 1 V. When the switch S closes and the drivers start to draw current, the time constant of this simple R - L circuit is L/R_{eq} , where R_{eq} is the equivalent parallel resistance of R_1 and R_2 . Hence the time constant is $L/R_{eq} = 1.4$ ns. Since the circuit needs about 2.2 times the time constant to reach 90% of the final value, the corresponding time is 3.08 ns, which is too large compared with the 0.3-ns rise time of such drivers. The inductance within the power supply of the package is therefore excessive for the chip and there is a simultaneous switching noise problem.

The classical solution to this problem is to add a decoupling capacitor C to the chip as shown in Fig. 9(b), to supply the current needed by the chip and decouple it from the package inductance. Assuming that the power supply seen by the chip can tolerate only a 5% variation during simultaneous switching of driver circuits as modeled by resistor R_2 , the switching current in R_2 can only be supplied from two sources: the resistor R_1 and the capacitor C . A 5% drop of voltage at resistor R_1 of 1 Ω diverts 50 mA of its current to R_2 . Since R_2 needs a total current of $0.95 \text{ V} / 2.5 \Omega = 0.38 \text{ A}$ to reach 0.95 V, the remaining current of $0.38 \text{ A} - 0.05 \text{ A} = 0.33 \text{ A}$ has to be supplied by the capacitor C . The capacitor current is determined by $I = C(dV/dt)$ where, if $dV = 0.05 \text{ V}$, $dt = 1 \text{ ns}$ and $I = 0.33 \text{ A}$, $C = 6.6 \text{ nF}$. A key requirement is minimizing the inductance of the capacitor and chip to the capacitor interface, since an inherent assumption made in Fig. 9(b) is that the inductance between capacitor C and the resistors R_1 and R_2 is much smaller than L and is ignored.

As previously mentioned, the use of thin-film wiring structures on the surface of a ceramic substrate for interconnection allows the decoupling capacitor to be built into the ceramic substrate. The VLSI chips can be mounted on the thin-film surface on top of the substrate and are separated from the capacitor substrate only by the thickness of the thin-film layers. The chip-to-capacitor substrate distance is at most 0.1 mm. If an array of solder balls is used for power supply interconnection, the lead inductance from capacitor to chip may be reduced to several picohenrys. Two types of structures for the integrated capacitor substrate are shown and discussed below. Each structure has its advantages and disadvantages, but the basic tradeoff appears to be between simplicity of capacitor substrate fabrication and inductance of the interconnection from chip to capacitor.

● Integrated capacitor structures

The structure shown in Fig. 10 is a multilayer ceramic (MLC) substrate supporting thin-film lines and a VLSI chip. The integrated capacitor substrate is the typical MLC structure with horizontal ceramic layers interspersed with metal layers. Clusters of vias are provided under each chip as shown in Fig. 11. Vias for three levels of voltage are shown arranged in rows. Each row of vias for the different voltage levels is connected to appropriate capacitor planes. For vias passing through a capacitor plane, donut areas (ceramic areas) isolate each via from the capacitor plane. Also shown in Fig. 11 are signal vias and reference voltage vias, which are on a larger grid spacing than the power vias passing through the capacitor plane. The desired capacitance is provided by varying the number of layers, thickness, and dielectric constant of the ceramic substrate with the following limitation. Signals must also travel from chips through vias in the substrate to other modules and input/output devices. The signal flight time t_D is given by the relationship

$$t_D = \frac{\sqrt{\epsilon_r}}{c} \ell, \quad (9)$$

where ℓ is the substrate thickness, ϵ_r is the dielectric constant of substrate material, and c is the speed of light. If one limits the flight time for a high-speed module to less than 100 ps, for a substrate dielectric constant of 50 the substrate thickness should be less than 4.2 mm for this horizontal capacitor structure design.

An optimum inductance path from an integrated capacitor substrate to chips would be obtained if the chips were mounted directly on the edges of the capacitor plates. A close approximation to this optimum design is the structure with vertical capacitor plates shown in exploded perspective in Fig. 12. The structure consists of three elements, the capacitor inserts A, the main body of the substrate with signal and reference vias B, and the redistribution layers C.

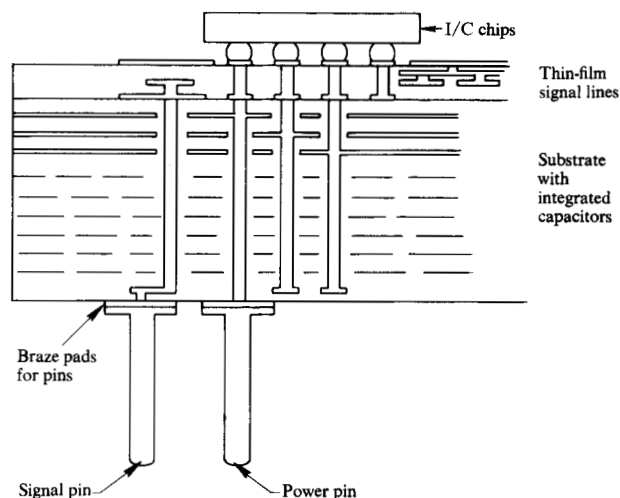


Figure 10 Cross section of substrate with integrated decoupling capacitance based on horizontally laminated power planes.

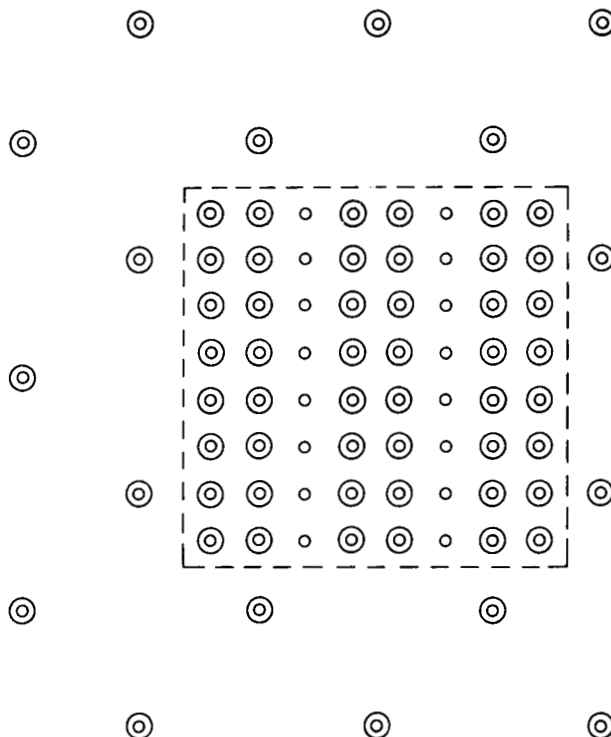


Figure 11 Arrangement of power supply and signal vias under each chip site of substrate shown in Fig. 10.

The inserts A are fabricated separately by laminating together a stack of ceramic and metallization layers of the appropriate pattern. In dicing the laminate to the desired size, tabs from the capacitor plates are exposed on the top for

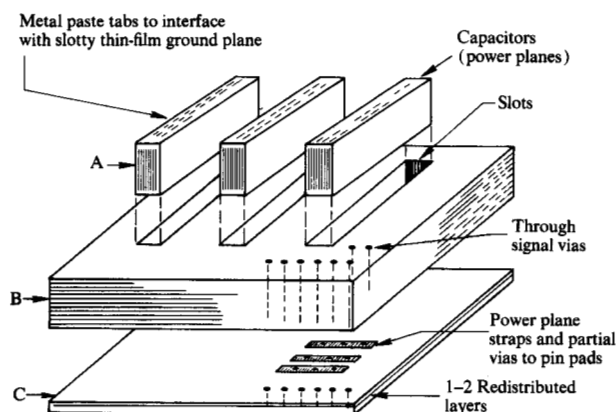


Figure 12 Assembly of substrate with integrated decoupling capacitance based on vertically laminated power plane.

connection to the thin-film structure, and on the bottom for connection to power plane straps. There is no via in these capacitor inserts. The main body of the structure, B, consists of ceramic layers with slots punched in the individual sheets which contain through vias for reference and signal. There is no metallization pattern on these layers. The redistribution layers C consist of at least two ceramic layers. The power plane straps are deposited on the top layer and, in addition, there are signal vias passing through all redistribution layers. On the lower layer or layers, the power and signal vias are redistributed by short lines of metallization to interconnect to pin pads on the bottom of the substrate, not shown.

The entire substrate is fabricated by stacking the redistribution layers C and the layers with punched slots B, then inserting presized laminates A in the slots. A final lamination is then performed, joining together the various elements in the "green" state. Cutting the module to size and sintering to burn off the organic binder to fuse the ceramic parts and metal together complete the fabrication process.

An integrated capacitor substrate may carry from 10 to 100 chips. If each chip requires a minimum capacitance of 20 to 40 nF, a total capacitance of 2 to 4 μ F in a substrate up to 10 $\times$ 10 cm in size is needed. The sheet thicknesses and numbers of layers for this capacitance and substrate size for several dielectric materials are given in Table 2. For the horizontal and vertical structures the numbers of layers range from 20 to 150 for materials with dielectric constants in the 50 to 100 range.

Depending on the number of layers and the dielectric constant desired, different dielectric and metal sets can be used. For example, alumina with a dielectric constant of 9 can be used with molybdenum or tungsten metallurgy

because of the high sintering temperature needed. On the other hand, glass ceramic material can be developed to obtain a dielectric constant between 10 and 100. Its sintering temperature is lower, and hence metals such as AgPd and Ni can be used.

Detailed analysis of the effective inductances of several horizontal and vertical integrated substrate designs indicates that the vertical design (Fig. 12) has two to three times lower inductance than the horizontal design [10]. However, both types of structures are expected to be able to support large arrays of chips switching simultaneously as many as 72 drivers per chip with better than 10% stability in the power supply at the chip terminals.

We are presently investigating materials and processes needed to implement such integrated capacitor substrates and we have partially succeeded in building satisfactory experimental structures. These results are beyond the scope of this paper and will be presented later.

Thin-film fabrication

Effort has also been devoted in this laboratory to developing material and processes for the thin-film structure. This section summarizes the essence of our preferred approaches and highlights the need to develop defect detection and repair techniques to achieve acceptable yield for a Thin-Film Module.

The wiring structure of the Thin-Film Module must possess the following characteristics: a minimum of five metal layers, two ground planes sandwiching two orthogonal wiring layers with dimensions exemplified in Fig. 6, and a topmost layer, the fan-out layer shown in Fig. 3, to connect the wiring layers to the semiconductor chip contacts. This fan-out layer can also be used to provide engineering change capability in a manner similar to that practiced on IBM's latest MLC multichip modules. Vias must be provided to vertically connect the various layers to the substrate vias and to make connections within the thin-film structure. The fan-out to chip contact and thin-film via to substrate via interfaces may require special metallurgical barriers to provide a reliable connection to the semiconductor chips and the substrate. Transmission line impedance control dictates the need to achieve better than 10% leveling at each film layer. A high degree of leveling and planarity is also required to achieve good image tolerance with photolithographic processes over large substrate areas and to minimize complications with subsequent assembly of devices with several hundred contacts per chip. The planarity requirements are comparable to those imposed on semiconductor device wafers and will most likely require planarizing of ceramic substrates prior to thin-film deposition. Minimizing resistive loss calls for wiring layers and vias with aspect ratios (height

Table 2 Geometrical requirements for integrated capacitors.

Substrate structure	Area of plates (cm × cm)	Dielectric thickness (μm)	Dielectric constant	No. of layers	Total capacitance (μF)
Horizontal	10 × 10	75	50	60	3.5
"	"	25	10	100	3.5
"	"	25	50	20	3.5
Vertical	0.5 × 10	25	10	150	0.26
"	(× 10 stacks)	"	"	"	(2.6)
"	"	25	50	30	0.26
"	"	"	"	"	(2.6)

to width ratio) and cross sections significantly larger than those typically required in semiconductor devices. Minimizing propagation delay calls for insulator films with low dielectric constant and low loss up to 1 GHz. The composite structure must be capable of withstanding repeated exposure to temperatures of several hundred degrees centigrade during processing and component attachment. This combination of properties cannot be achieved with state-of-the-art processes and will require significant materials and process innovations.

Cost, reliability, and resistivity considerations suggest that Cu is likely to be the most practical material for the metal layers. Polymer materials such as polyimide are likely to be best for the insulator. Such materials are known to have low dielectric constant in cured form ($\epsilon \leq 3.5$) and possess leveling and planarizing characteristics superior to those of inorganic films since they can be deposited as viscous liquid films which can flow during the curing process.

Both Cu and polymer films can be patterned by a variety of techniques. Subtractive techniques using liquid etchants are not likely to be adequate because these lead to undercutting under resist stencils, an undesirable effect *vis-à-vis* maximizing metal line and via aspect ratios and cross sections. This effect can be suppressed by use of dry etching techniques using reactive plasmas or by use of additive electroplating, electroless plating, or evaporation followed by a lift-off through a photoresist stencil. We have therefore emphasized using the latter techniques with partial success in key areas. The details of this work will also be presented later.

In manufacturing thin-film structures for both chips and packages, the resultant yield is usually gated by the size and the number of defects introduced into the structures by less than ideal manufacturing environments and processes. There are in general three major kinds of defects encountered in fabricating multilevel thin-film structures: intralevel metal line opens and shorts, interlevel metal line shorts, and

defective interlevel vias. Within a given level, the line opens and shorts are due to missing or excess of a portion of the metal lines. These can be the result of a mask defect, incomplete or excessive removal of photoresist in the additive approach, and particulate contamination. The defects are also process-dependent. For example, in electroplating the process starts at the bottom surface of the photoresist stencil. It is therefore important that the surface be clean and that there be no wetting problem or air bubbles between the electrolyte and the metal surface. In the lift-off process the metal is deposited through a photoresist stencil in a line-of-sight projection from the metal source in the evaporator. It is therefore important that there be no dust particles on the substrate surface that may block metal deposition.

An intralevel defect may or may not be "fatal" (affecting the electrical performance of the chip or package) depending on its size. For example, a metal line 8 μm wide may very well be able to tolerate a dust particle of 3 μm in size. On the other hand, a dust particle equal to or more than 8 μm in size which settles right on the line would cause a line open, a fatal defect. Similar considerations apply to defects located between metal lines which may or may not cause a short depending on their size. Experience has indicated that in a controlled clean-room environment, the density of defects occurring on a substrate is a strong function of the size of the defects [11]:

$$f(x) = k \frac{1}{x^3}, \quad (10)$$

where x is the defect size, $f(x)$ is the probability density function of defects of size x , and k is a constant. Once a critical defect size x_0 is determined, beyond which the defect can become a fatal one, Eq. (10) can be integrated from $x = x_0$ to $x = \infty$ to calculate the total number of defects D_0 with a size equal to or larger than x_0 occurring in a given area:

$$D_0 = k_1 \frac{1}{x_0^2}. \quad (11)$$

If we now define the critical area A_c to be an area on the

substrate where a defect with a size x_0 or larger will cause a fatal defect, a Poisson distribution for a uniform distribution of defects over the whole substrate is given by

$$Y = e^{-A_c D_0} \quad (12)$$

Note that Eq. (11) shows that the yield is really determined by the total number of defects that occur on the total critical area in a multilevel thin-film device. Another observation can be made with Eq. (12) if we substitute Eq. (11) into (12), using the fact that for a given structure on a square substrate of size y , if all horizontal dimensions of structures of this substrate are scaled proportionately to y , the critical area A_c is proportional to y^2 . The expected yield can therefore be expressed as

$$Y = e^{-k_2(y/x_0)^2} \quad (13)$$

where k_2 is another constant. Equation (13) therefore states that the yield is only a function of the *ratio* of the substrate size *versus* the critical defect size. If we scale down the substrate horizontal dimensions and all its internal horizontal features linearly, the critical defect size also scales down with the substrate size. Therefore, the ratio of y and x_0 remains constant and so does the yield. This means that given the above assumptions on defect distribution, the yield for fabricating a given device does not change when all its horizontal dimensions are scaled up or down simultaneously. Hence, achieving high yield in fabricating a Thin-Film Module with a substrate size of 10×10 cm and a minimum feature of $10 \mu\text{m}$ would be comparable to achieving high yield in fabricating a chip with the same structure but with a chip size of 0.5 cm and a minimum feature of $0.5 \mu\text{m}$, which surely exceeds the capability of present integrated circuit processes practiced in stringent clean-room environments.

The foregoing analysis suggests that achieving practical yields in building a Thin-Film Module with the complexity described in this paper will require development of novel defect detection and repair techniques for each layer in the thin-film structure. Electron-beam microscopy is likely to be best for defect detection since it offers the potential for contactless testing of what are likely to be delicate electrical structures [12]. Jet or laser-enhanced plating and etching are examples of potential defect repair techniques being investigated in this laboratory [13]. In our judgment, achieving practical yield in thin-film structures over large substrate areas through the use of novel defect detection and repair techniques is the principal obstacle to fabrication of the Thin-Film Module.

Summary and conclusions

Key trends in future VLSI devices and their expected impact on future semiconductor packages for high-speed digital systems have been summarized. Two key requirements for future semiconductor packages are significantly higher wiring capacity and containment of simultaneous switching

noise to acceptable levels. A novel package concept, the Thin-Film Module, featuring thin-film strip transmission lines for interchip wiring and power supply decoupling capacitors integrated into the body of the module, has been proposed as an alternative to meet these two requirements. Thin-film transmission lines of dimensions required to provide the necessary wiring density and capacity will have significant resistive loss. However, analysis of high-speed pulse propagation and coupled noise has established that with certain restrictions lossy lines can indeed be used to wire future VLSI devices in multichip packages. Novel MLC approaches for integrated decoupling capacitor substrates and material and process considerations for the required multilayer thin-film wiring structures have also been discussed. Novel defect detection and repair techniques for thin-film structures will be essential for fabricating the Thin-Film Module with practical yields.

References

1. E. Bloch, "VLSI and Computers—Challenge and Promise," keynote address, IEEE Computer Society Conference, San Francisco, February 1980.
2. A. J. Blodgett, Jr., "A Multilayer Ceramic Multichip Module," *IEEE Trans. Components, Hybrids, Manuf. Technol.* CHMT-3, 634-637 (1980).
3. B. T. Clark and Y. M. Hill, "IBM Multichip, Multilayer Ceramic Modules for LSI Chips—Design for Performance and Density," *IEEE Trans. Components, Hybrids, Manuf. Technol.* CHMT-3, 89-93 (1980).
4. A. Deutsch and C. W. Ho, "Triplate Structure Design for Thin Film Lossy Underminated Transmission Lines," presented at the 1981 International Symposium on Circuits and Systems, Chicago, April 27-29, 1981.
5. E. Weber, *Linear Transient Analysis*, Vol. II, John Wiley & Sons, Inc., New York, 1956, p. 383.
6. N. Arvanitakis, IBM General Technology Division, Endicott, NY, private communication.
7. C. W. Ho, "Theory and Computer-aided Analysis of Lossless Transmission Lines," *IBM J. Res. Develop.* 17, 249-255 (1973).
8. C. W. Ho, "Thin Film Lossy Line Package," U.S. Patent No., 4,210,885, July 1, 1980.
9. W. C. Johnson, *Transmission Lines and Networks*, John Wiley & Sons, Inc., New York, 1950.
10. G. V. Kopcsay, IBM Thomas J. Watson Research Center, Yorktown Heights, NY, private communication.
11. C. H. Stapper, A. N. McLaren, and M. Dreckmann, "Yield Model for Productivity Optimization of VLSI Memory Chips with Redundancy and Partially Good Product," *IBM J. Res. Develop.* 24, 398-409 (1980).
12. T. P. Chang, F. J. Hohn, P. J. Cohane, D. P. Kern, and W. H. Bruenger, "Electron Beam Testing of Packaging Models for VLSI Chip Arrays," *Proceedings of the 16th Symposium on Electron Ion Photon Beam Technology*, May 1981.
13. J.-Cl. Puipe, R. E. Acosta, and R. J. von Gutfeld, "Investigation of Laser Enhanced Electroplating Mechanisms," *J. Electrochem. Soc.* 128, 25-39 (1981).

Received July 13, 1981

C. H. Bajorek is located at the IBM Research laboratory, 5600 Cottle Road, San Jose, California 95193 and the other authors are located at the IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598.