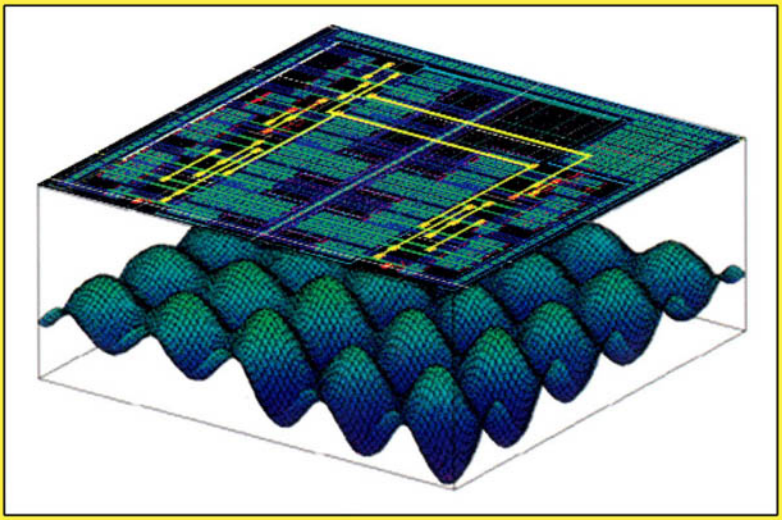
SUBSTRATE NOISE

Analysis and Optimization for IC Design

Edoardo Charbon
Ranjit Gharpurey
Paolo Miliozzi
Robert G. Meyer
Alberto Sangiovanni-Vincentelli



SUBSTRATE NOISE

Analysis and Optimization for IC Design

This page intentionally left blank

SUBSTRATE NOISE

Analysis and Optimization for IC Design

Edoardo Charbon
University of California/Berkeley

Ranjit Gharpurey
Texas Instruments/Dallas

Paolo Miliozzi
Conexant Systems/Newport Beach

Robert G. Meyer
University of California/Berkeley

Alberto Sangiovanni-Vincentelli
University of California/Berkeley

KLUWER ACADEMIC PUBLISHERS
NEW YORK, BOSTON, DORDRECHT, LONDON, MOSCOW

eBook ISBN: 0-306-48171-5
Print ISBN: 0-7923-7325-1

Created in the United States of America

Visit Kluwer Online at: http://kluweronline.com
and Kluwer's eBookstore at: http://ebooks.kluweronline.com

Contents

This page intentionally left blank

List of Figures

This page intentionally left blank

List of Tables

to Tokiko, Arianna, Avni,
Anupa, Usha, Ana, Chiara,
Teresa, Tonino, Liz, Marica,
Andrea, Maria, Marco

This page intentionally left blank

Acknowledgments

The authors would like to thank many of the former students at the University of California, Berkeley, who contributed to this book with discussions and comments. A special thank goes to Iason Vassiliou for designing some of the circuits presented here, Xiaojun Liu for implementing the switching behavioral model generator, Luca P. Carloni and Alberto Ferrari for discussions on substrate models, Amit Mehrotra and Stefano Zanella for help on phase noise and statistical modeling.

The authors are also grateful to Joel R. Phillips and Kenneth Kundert of Cadence Design Systems, Luis M. Silveira of INESC, Portugal, and Jacob White of the Massachusetts Institute of Technology for interactions on the numerical aspects of substrate noise extraction and substrate benchmarking. The authors would like to acknowledge Kannan Krishna and David Martin of Texas Instruments for insightful interactions and Raminderpal Singh of Cadence Design Systems for comments on the practical aspects of noise modeling.

This page intentionally left blank

Foreword

The understanding, prediction, and control of unforeseen interactions within integrated circuits has been a major impediment to the timely development of analog, mixed-signal, and high-performance digital integrated devices since the dawn of the industry in the 1960s. The most important of these have resulted from the effect of package lead impedance, thermal effects, and electrical interactions through the substrate. Since the physical origin of these was often not well understood, much less characterized and modeled, the interactions were often not evident until actual silicon devices were fabricated, requiring cycles of design changes to produce a manufacturable device.

This problem has become more severe as the complexity of mixed analog-digital integrated circuits has become ever greater, and the operating frequencies and bandwidths processed on chip have increased steadily. The semiconductor industry is moving towards an era in which analog and digital subsystems will reside together on large complex chips, resulting in far higher levels of integration than now. This can only occur if these unforeseen interactions can be more effectively predicted during the design process than is the case now, resulting in reasonable design cycles to achieve a manufacturable product.

This book addresses the critical problem of the modeling and characterization of interactions through the substrate, perhaps the most difficult of the classes of interactions to characterize and model. Solution of the problem involves finding modeling approaches that maintain the right compromise between computational complexity and degree of accuracy. Computational approaches that are well suited to the particulars of the problem, and finally developing design approaches that result in designs that are relatively insensitive to substrate interactions. The book provides an excellent summary of the state of the art in these areas.

Paul R. Gray

This page intentionally left blank

Chapter 1

INTRODUCTION

The semiconductor industry has grown at a yearly rate of 15% over the past decade. Currently, the industry expects to reach over $300 billion in revenues by 2001. Such a significant growth is constantly demanding greater feature miniaturization and device density, as well as higher production volumes and yields. To support the latest technological advances in semiconductor fabrication, the last two decades have seen a continuous restructuring of design practices. New paradigms have appeared, constantly pushing the design abstraction far higher and resulting in the enhancement of designers' productivity.

The design of systems on a single chip represents today's frontier in the semiconductor industry. There are several challenges associated with this new design paradigm, where several circuits of entirely different nature, operating modes, and functionality are integrated on the same substrate at short distances to one another. In this scenario, designing in the presence of substrate noise is proving to be one such challenge, perhaps one of the most problematic.

Substrate noise has had a constant and significant impact in the design of analog and mixed-signal integrated circuits. Due to its ubiquity on chip, substrate noise changes the performance of most components, thus resulting in partial or total loss of functionality in the entire system. Furthermore, some dynamic logic families are more susceptible to noise than static ones. Thus, restricting the applicability of substrate noise to mixed-signal systems is becoming unrealistic. Recently, with further advances in chip miniaturization and innovative circuit design, substrate noise has begun to plague fully digital circuits as well. Critical path delays may be impacted, other paths may become critical as a result of generalized delay increases. Localized delay degradation may cause clock skews and glitches.

The design of systems in the presence of substrate noise has been traditionally a manual process, where noise-aware optimization was generally not used.

To combat the effects of substrate noise, heavily over-designed structures are generally adopted, thus seriously limiting the advantages of innovative technologies. For this reason, the modeling of substrate noise is receiving renewed attention.

To date, research on substrate noise has focused on characterizing multilayer substrates and creating compact models for relatively simple substrate configurations. Noise sources are usually represented in terms of independent white or colored Gaussian wide-sense stationary or cyclo-stationary random processes. To help designers quickly assess the impact of substrate noise, researchers have built *ad hoc* noise models for specific circuits, such as voltage controlled oscillators and mixers. Very few attempts have been made to systematically analyze complex noisy systems, including noise generation, transmission, and reception.

This work is an attempt to answer these questions, addressing both design and physical implementation issues in light of the insights we gained over years of experience. At the same time we tried to form a set of guidelines intended to help make early decisions, while managing the layout problem.

1. SUBSTRATE NOISE PROBLEM

Substrate noise can be decomposed, at a macroscopic level, in intrinsic and switching noise. Intrinsic noise is a background spurious signal originated in the resistive component of the substrate due to thermal noise[1]. The strength of intrinsic noise is relatively low compared to that of switching noise. Thus, it is often ignored in most substrate noise analysis tools.

Switching noise originates in digital blocks where frequent state transitions, occurring in gates across the chip, cause current pulses to be absorbed from and transmitted to power and ground buses through direct feedthrough and the charge/discharge of loads. Such pulsing currents are partially injected into the substrate through impact ionization and capacitive coupling.

Accurately characterizing substrate noise is problematic for various reasons. Geometry and process variations may change the map of the noise distribution system and the manner in which it is delivered to sensitive circuits. Spurious currents injected into the substrate travel through the bulk reaching various depths and resurfacing to be collected by low-resistivity pickups. The paths followed by such currents are determined by the relative position of the injector, the pickup and the other contacts in the circuit, the substrate doping profile, and the backplate potential. Figure 1.1 shows substrate current flow lines for the case of (a) distant and (b) close injector/receptor systems in a typical low-resistivity substrate, as they appear in a two-dimensional device simulation.

Noise results from superposition of a large number of local and remote sources, each attenuated and delayed in a unique way. Modeling signal attenuation and delay individually may be extremely time consuming and would

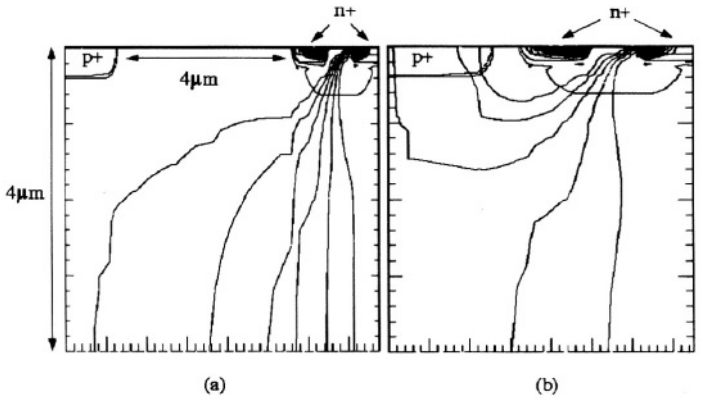


Figure 1.1. Current flow of substrate noise through low-resistivity substrate with different injector/receptor configurations, assuming grounded backplate contact: (a) distant; (b) close surface contact

require accurate *ad hoc* characterization of all the sources, which is in itself a hard problem. Switching noise coupled through the substrate is very destructive as it can be broadcast over great distances and picked up by sensitive circuits by way of capacitive coupling and body effect. Digital circuits are not immune from these effects, since the resulting threshold voltage modulation dynamically changes gate delays locally, thus impacting performance in ways that are difficult to predict.

Switching noise has an especially detrimental effect on dynamic logic, memories and embedded analog circuits. In these circuits the presence of sensitive structures and large noise injectors on the same chip makes it imperative for the designer to accurately and efficiently estimate the strength of substrate noise at various locations. Delay skews can be caused by various factors, among the most important, substrate noise, voltage drops across power supply interconnect (shown in Figure 1.2), and switching noise.

2. ANALYZING SUBSTRATE NOISE TRANSPORT

Accurate estimation of substrate noise requires several techniques to model switching noise injection, transport, and reception mechanisms, both at the microscopic and macroscopic level. Moreover, means are necessary to embed substrate noise in optimization loops and trend analysis tools. This is critical to help designers detect substrate related problems as early as possible, thus tailoring design practices in order to avoid lengthy redesign cycles.

The substrate analysis problem has been addressed by a number of authors since the 1970s, however, the advent of advanced miniaturization has funda-

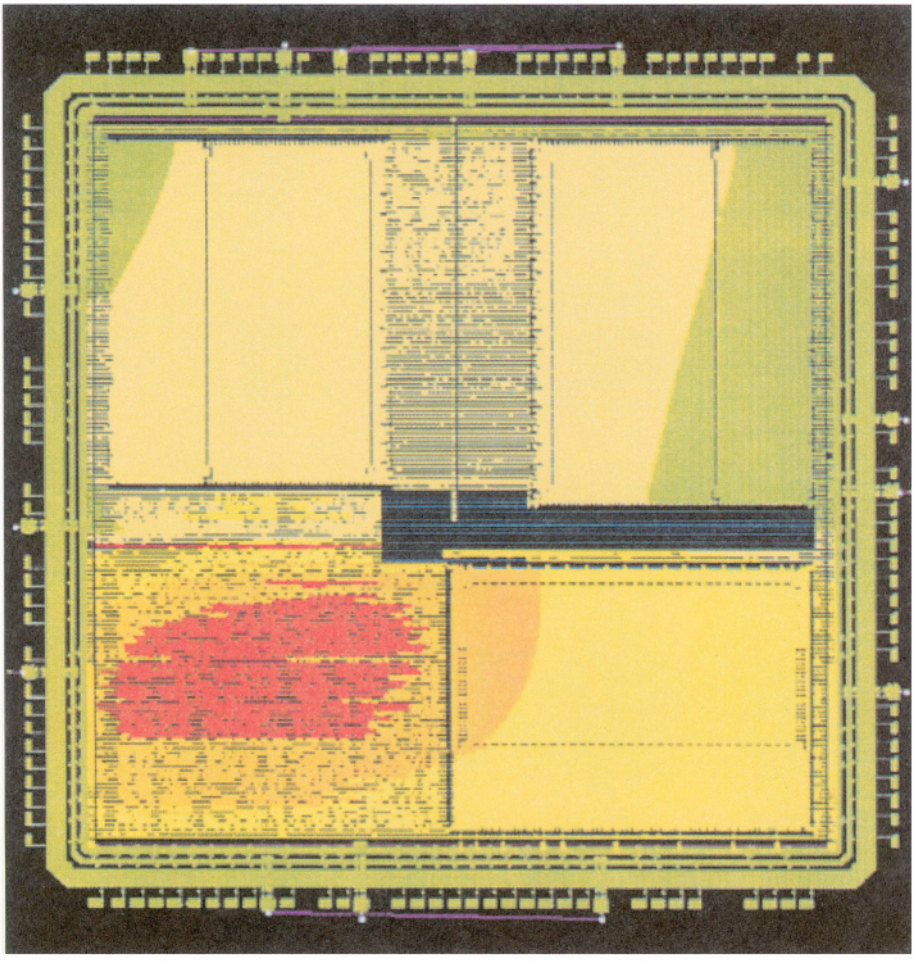


Figure 1.2. Map of delay skews due to non-ideal supply distribution. Darker tones represent greater, lighter ones smaller skew– Courtesy: Simplex Solutions

mentally revolutionized the field. Anisotropic substrates were first studied in detail in the 1970s [2]. In these approaches, the space was discretized into a resistive/capacitive mesh. DC/steady-state analysis was carried out by direct solution of the system of simultaneous thermal and electrical equations. Automated thermal characterizations of the substrate have also been proposed, for example in [3].

More recent substrate analysis approaches can be classified in two main types, the ones based on direct solution of the partial differential equation underlying the charge transport mechanisms and the ones solving an equivalent integral form of the equation.

The first approach involves the use of classical finite element [4] or difference [5] methods. The space is uniformly discretized into three-dimensional cubes each one of which is assigned a variable representing the potential. Charge transport mechanisms are modeled by a three-dimensional Laplace equation which in turn is solved by replacing the derivative of the potential in x-, y-, and z-direction by differences of the variables associated with adjacent boxes. Such methods are versatile and general in nature, since they can handle lateral and vertical resistivity variations and arbitrary substrate geometries.

However, to obtain accurate substrate characterization, a fine mesh is required, thus making storage and computational efforts often prohibitive. To overcome the formidable computational complexity of the problem, sparse non-uniform grids are often used. The grid size is made fine in areas close to substrate contacts and coarse in distant regions. The use of non-uniform or coarse grids involves speed-accuracy trade-offs, which are often difficult to evaluate *a priori.*

In the second approach the three-dimensional partial differential equation modeling charge transport is translated into an integral equation over the two dimensional surfaces that bound the problem domain, usually the substrate contacts and possibly material interfaces. This results in a reduction of the number of unknowns and, where they are applicable, integral equation techniques can provide superior performance, if efficient linear equation solvers are used.

Solving the equivalent integral form of the charge transport equation is especially useful in incremental substrate analyses. Hence, they are effective when incorporated in flows which include engineering change orders (ECOs) and optimization processes. For this reason such methods have been the focus of our work.

3. OPTIMIZATION IN THE PRESENCE OF SUBSTRATE NOISE

Early authors favored a bottom-up approach while designing systems operating in a noisy environment. Generally, large areas dominated by contacts were built in order to separate the critical sections of the chip [6]. Recently, top-down alternatives have been proposed in which substrate noise is taken into account at various stages in the design of high-performance circuits, from the initial architectural phase to the final physical design and verification phases. The aim is to drive the design towards solutions more resilient to substrate noise as early as possible. Experience has shown however that time complexity may become prohibitive when one wants to very accurately model substrate and/or analyze large systems.

This problem has been addressed by a number of authors, who proposed heuristics to speed up substrate analysis during physical assembly phases,

e.g. [7, 8]. The approaches solve the charge transport differential equation via a finite difference method using a coarse grid spanning the workspace. Asymptotic waveform evaluation (AWE) [9] is used to accelerate the solution of the resulting system of simultaneous algebraic equations. A potential problem with this approach is a strict requirement of alignment between grid and layout objects. Thus, unless specific tessellation [10] is used, iterative solvers based on progressive and often minimal modifications may not fully take advantage of the algorithms. In an optimization environment in which substrate must be considered, one has to take into account the *global* effects of small changes in the layout. Hence, approaches, consisting of solving coarse finite difference analyses, may reach such inaccuracy levels that the insights gained applying this method might not be beneficial but misleading, thus possibly resulting in sub-optimal solutions.

Alternative approaches based on the use of analytical models have been used in optimization problems. Such models are generally constituted by an impedance network spanning the set of all circuit contacts and the ground [11, 12]. However, it is not clear how accurate such models are when even minor modifications are performed during a progressive transformation of the layout.

A third approach uses sensitivities to evaluate the impact of layout adjustments over contact-to-contact impedances [13]. As an illustration, consider the phase lock loop (PLL) shown in Figure 1.3. The circuit consists of a phase frequency detector (PFD), a voltage controlled oscillator (VCO), three frequency dividers (D_1, D_2, D_3), a charge pump (CP), and a loop filter (LPF). The diagram shows the VCO's sensitivity to noise as a function of the directional displacement of the divide by n component (D_3).

Figure 1.4 shows a normalized diagram of the noise injected by the circuit components as it cumulates locally across the chip. As expected, the areas with the highest concentration of noise are those near the dividers, especially the divide by n component as it is operated at high frequency.

Sensitivities can also be used as a quality factor for the selection of the most cost-effective technology on the basis of a class of circuits one wants to fabricate with given specifications. Furthermore, one can characterize the *trend* of circuit performance when engineering changes are implemented on substrate geometry, technology parameters or design. Finally, the effects of technology migration/scaling can be carried out efficiently for a given chip without the need of performing a large number of complete substrate extractions. Sensitivities are typically used to build performance models accounting for discrete parasitics as well as substrate effects.

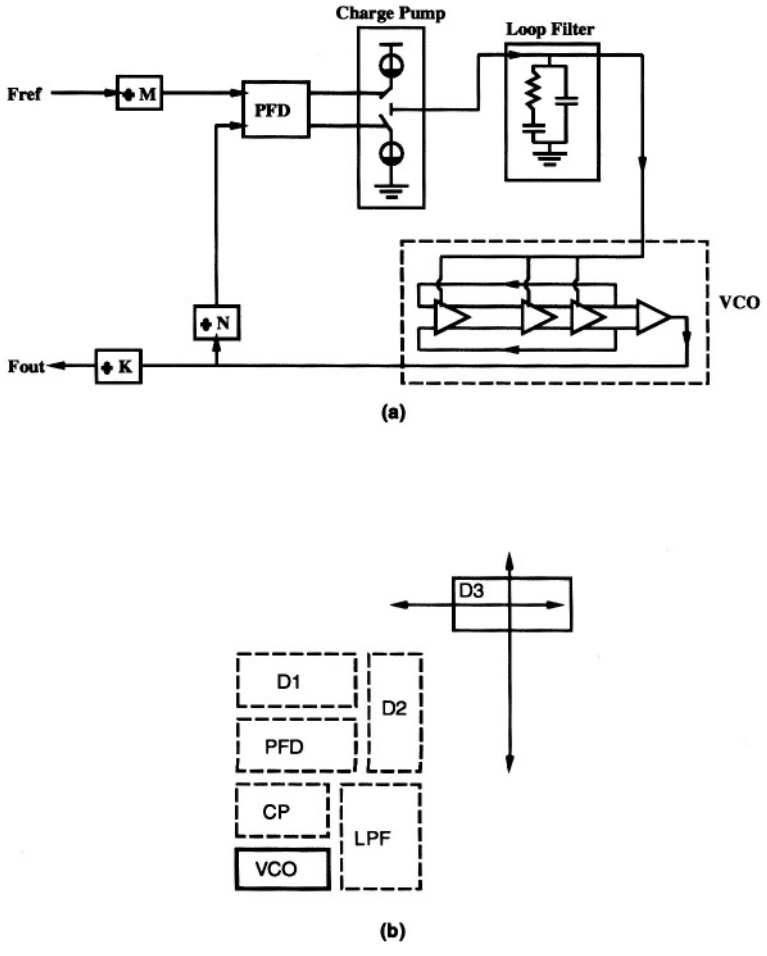


Figure 1.3. (a) PLL block diagram; (b) PLL floorplan with the rate of change of noise power at the VCO as a function of the direction in which divider D_3 is displaced (represented by arrows of different lengths)

4. DESIGN PRACTICES

Besides trends and sensitivities of a performance measure with respect to certain substrate parameters, the designer usually needs qualitative insight on a given architecture. Certain design decisions, especially early ones, require knowledge of possible interactions of the various components, which are generally still on the drawing board.

For this reason we have devoted significant space to describe several design practices which we feel have been an important factor in achieving the designed performance in our fabricated chips. The practices are also suitable to be

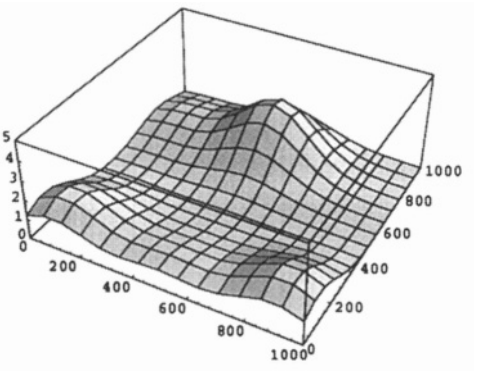


Figure 1.4. Normalized switching noise signal amplitude observed at various locations on chip

incorporated in semi-automated design flows, thus making them even more valuable.

Similar practices have been developed to select the best suitable process to a given design or, conversely, to choose the most suitable architecture for a given process. We believe that today's large number of options available in state-of-the-art processes can be best tailored to one's needs if one takes advantage of these rational tools and practices.

5. BOOK ORGANIZATION

This work gives a comprehensive view of the substrate noise problem from the perspective of system and component designers in digital and mixed-signal domains. Accelerated substrate analysis and noise-aware optimization techniques are presented in the context of the design of high-performance circuits.

Based on our measurements of fabricated mixed-signal designs and *ad hoc* test structures, the effects of substrate noise are quantified and a set of guidelines is described for the design of circuits and their physical implementation on silicon. Early architectural decisions are weighted in light of their implications to later implementation and performance issues.

Chapter 2 describes several noise sources common in semiconductor circuits and the physical phenomena underlying noise transmission. Chapter 3 reviews the known techniques for efficient noise transport characterization in large silicon substrates.

Chapter 4 outlines acceleration techniques for substrate extraction and simulation. Chapters 5 and 6 present methods currently in use for substrate-aware computer-assisted design and optimization.

Chapter 7 outlines the implications of certain design practices to noise performance of critical circuits. From these notions, we derive guidelines to create substrate-noise resistant circuits.

To conclude, Chapter 8 illustrates the noise transmission behavior of various types of commonly used substrates and proposes a set of guidelines for noise-aware physical design practices.

This page intentionally left blank

Chapter 2

NOISE COUPLING MECHANISMS

In this chapter the basic mechanisms behind the generation and transport of substrate noise are outlined in light of their effects on certain performance metrics. Noise can act on circuits in very different ways based on what is generating it, and how it is injected into and picked up from the substrate.

We attempt here to categorize a number of widely used substrate types based on their transmission properties. The most relevant noise injection and reception mechanisms are described in detail. In the reminder of the book we will often refer to such notions for various discussions on substrate noise effects.

1. SUBSTRATE NOISE TRANSMISSION

Figure 2.1 shows a typical cross-section of substrate epitaxial layer on which a CMOS inverter is integrated. Substrate noise is caused by switching currents, which are partially injected into the substrate at various depths and are picked up by near and distant devices.

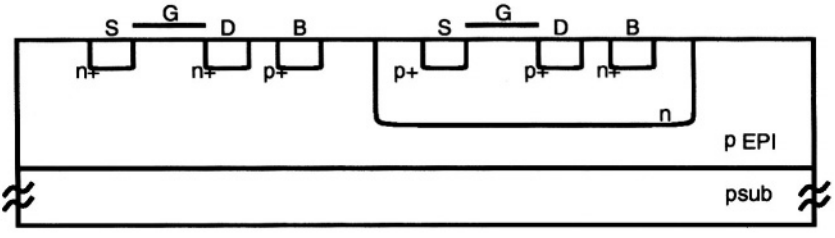


Figure 2.1. Substrate cross-section of a CMOS inverter

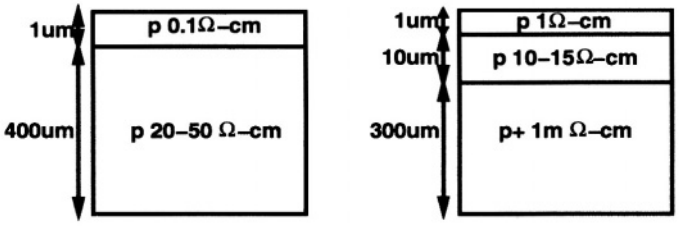


Figure 2.2. Typical substrate doping profiles: high resistivity (left); low resistivity (right)

In the whole spectrum of silicon substrates available today, one can recognize two main types: one referred to as *high-resistivity* and the other as *low-resistivity* substrate. Figure 2.2 shows examples of such types. In general, the first type is composed of a uniformly doped layer with a resistivity coefficient of $20-50\,\Omega\text{cm}$. The second type consists of a thick, high-resistivity epitaxial layer ($d \simeq 10\mu\text{m}$, $\rho \simeq 10-15\Omega\text{cm}$) and a low-resistivity bulk($\rho \simeq 1\text{m}\Omega\text{cm}$). Low-resistivity substrates are generally preferred for their good latch-up suppression properties [14]. On the other hand high-resistivity substrates are better suited to block substrate noise by using guard rings and physical circuit separation as shown in Chapter 8. At low and medium frequencies, typically less than 5GHz, all substrates show a resistive behavior.

Let us consider as an example the CMOS inverter from Figure 2.1, which we assume has been integrated into a high-resistivity substrate as in Figure 2.2(a). The plot in Figure 2.3 shows the input waveform and the resulting injected signals for both High-to-Low and Low-to-High transitions and various slew rates at the input (note the different time scale in the various waveforms). The plot was obtained from SPICE simulations using custom-fitted device models. Assuming that switching is synchronized with a clock signal, it can be shown that the power spectrum has energy components located in a wide spectrum, not necessarily centered around the clock frequency. A significant portion of this energy is usually concentrated around special frequency bands, e.g. at the inverse of the average gate delay. At DC or near DC frequencies, one also observes large spurious currents. This is due to the fact that impact ionization, for its very nature, only generates positive currents. Higher frequency components are due to glitches and fast switching phenomena occurring in large circuits [16, 17].

2. SUBSTRATE INJECTION MECHANISMS

Different types of active and passive devices in use in most IC technologies are shown in Figure 2.4. The cross-section of a bipolar npn transistor is shown in Figure 2.4(a). In these devices coupling to substrate is generally capacitive through collector-to-bulk reverse-biased junction capacitor C_{js}. The value of

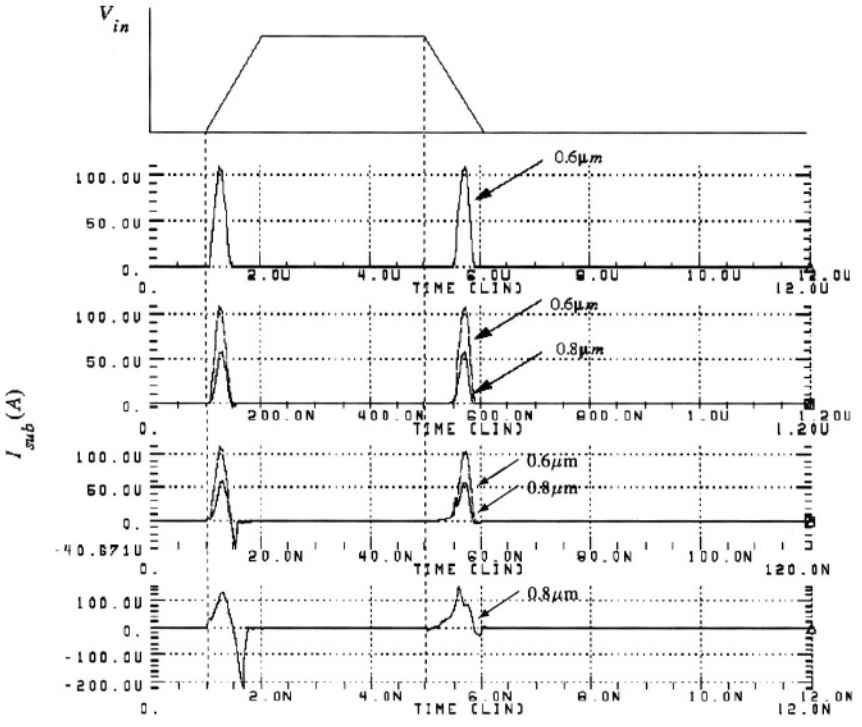


Figure 2.3. Noise spikes injected into substrate via impact ionization and capacitive coupling ($0.8\mu m$ and $0.6\mu m$ CMOS technologies). The waveforms shown were obtained with input waveforms of varying slew rates: note the different time scales of 2μs, 200ns, 20ns, and 2ns, while the input waveform is scaled accordingly

C_{js} for an abrupt junction can be estimated as

$$C_{js} = \sqrt{\frac{q\epsilon}{2(\psi_0 + V_{cs})}\left(\frac{N_C\, N_S}{N_C + N_S}\right)}, \tag{2.1}$$

where N_C, N_S are the collector and the substrate doping levels, ψ_0 the built-in junction potential, V_{cs} the collector-to-substrate bias voltage, q the electron charge and ϵ the substrate dielectric permittivity. For formulae of more complex doping profiles, see [18]. Injection also occurs through the parasitic pnp that forms when the main device approaches the saturation region of operation and its base becomes forward biased with respect to the collector. The base of the npn acts as the emitter of the parasitic pnp, the collector as its base and the substrate as its collector. However, the gain of the pnp will be necessarily small due to the high thickness of its base and the doping levels of its emitter.

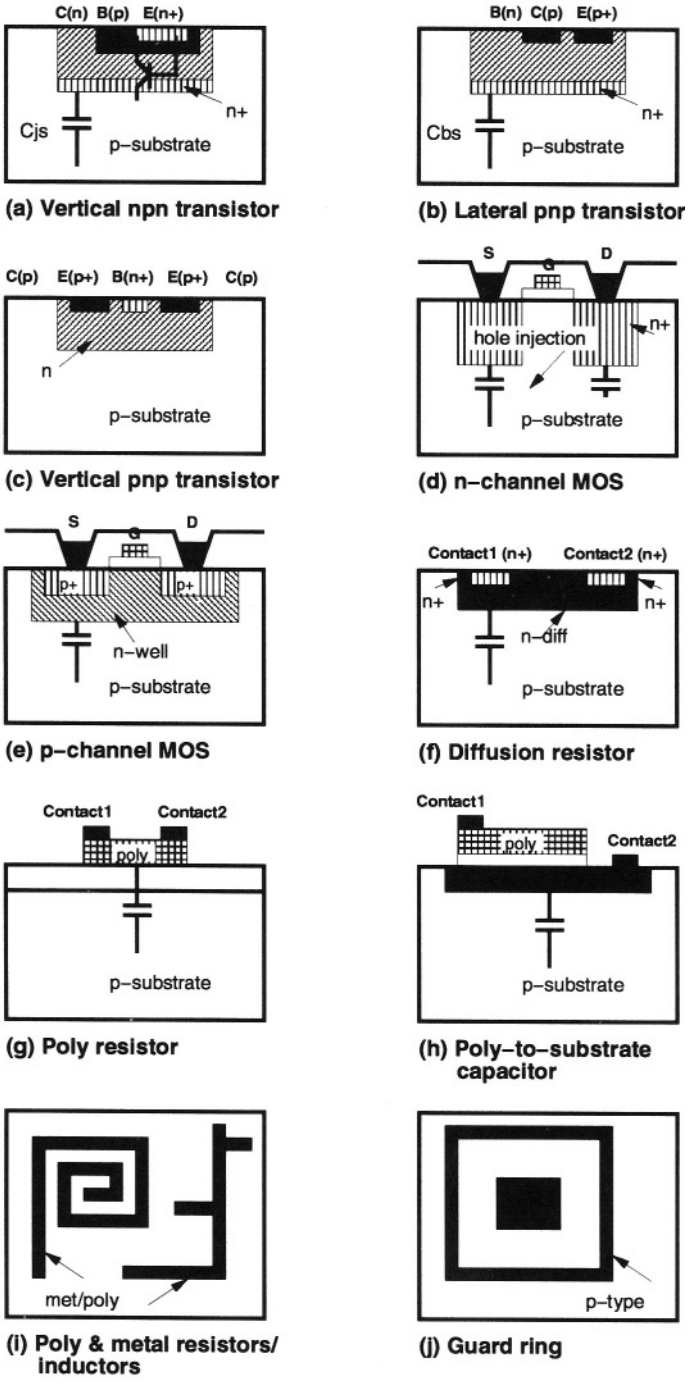


Figure 2.4. Injection and reception mechanisms for different types of devices

Lateral pnp transistors inject noise mainly through the base-to-substrate capacitance, as shown in Figure 2.4(b). On the contrary, in vertical pnp transistors the substrate is the collector node, shown in Figure 2.4(c). Thus, significantly higher currents can be injected in the substrate, unless low impedance draining is provided in immediate proximity of the device.

MOS transistors, shown in Figure 2.4(d),(e) for a n-well process, can interact with substrate in a number of ways: (1) capacitively, through the source(drain)-to-substrate junction; (2) resistively through *hot-electron injection* also known as *impact ionization.*

Impact ionization is caused by electron-hole pairs generated in the pinch-off region, when the electric field exceeds a given threshold. In the NMOS transistor case, while the electrons contribute to the channel current, the excess holes are collected in the region of substrate under the device and from there they are transported throughout the chip. Impact ionization currents are evaluated as

$$I_{impact} = \int_{E_s}^{E_m} I_d A e^{-B/E(x)} dx, \tag{2.2}$$

where E_s, E_m, $E(x)$ and I_d are source electric field, maximum electric field, local electric field and drain current, respectively. Constants A and B are material related coefficients. Formulae relating these parameters to measurable quantities and the derivation of (2.2) can be found in [14]. Since $E_m \gg E_s$, integral (2.2) can be approximated to

$$I_{impact} \simeq \frac{A}{B} l E_m I_d e^{-B/E_m} = C_1 (V_{ds} - V_{dsat}) I_d e^{-\frac{C_2}{(V_{ds} - V_{dsat})}}, \tag{2.3}$$

where l, V_{ds} and V_{dsat} are effective channel length, drain-source voltage and saturation voltage, respectively. C_1 and C_2 are material related coefficients [14]. Equation (2.3) is used by most MOSFET models to represent impact ionization currents [19].

Recent experimental evidence suggests that impact ionization is the dominant cause of substrate noise in NMOSFETs up to at least 100MHz [20]. Shorter device channels are likely to worsen the problem in the future, due to increased fields and smaller oxide thicknesses. For small-signal analysis, impact ionization can be modeled as a drain-to-body transconductance g_{db} given by

$$g_{db} = \frac{\partial I_{sub}}{\partial V_D} = \frac{C_2 \, I_{sub}}{(V_{ds} - V_{dsat})^2} \, . \tag{2.4}$$

The direct effect of g_{db}, which appears in parallel to r_0, is the reduction of the transistor output impedance.

Impact ionization currents in PMOS transistors are considerably small if compared with similarly sized NMOS transistors due to a lower hole ionization-coefficient. Substrate injection is further reduced by the fact that PMOS devices

in the process shown here are built in a locally AC-grounded well. The quality of the grounding is crucial, in fact if the well potential is allowed to vary with respect to the substrate potential, the entire well acts as a large injector, with a large reverse-biased well-to-substrate capacitance, thus worsening the effect.

Moreover, reverse-biased p-n junctions formed by all devices with substrate exhibit a steady DC leakage current. This current consists of carriers which are swept across the depletion barrier in the direction of the electric field. Electrons are injected into the n-region and holes into the p-region under the action of the field. Hence the substrate current induced by this mechanism is a majority-carrier drift current.

The passive components in typical processes are shown in Figure 2.4(f), (g), (h), (i). These components include resistors, capacitors, inductors and local diffusions. Resistors are in general implemented using either poly or diffusion. Poly resistors have a capacitance to substrate which is small if compared to that of diffused resistors. Assuming that one end of the resistance is connected to an AC ground, the current I injected into the substrate at low-frequencies, due to a voltage V_{in} applied at the other end of the resistor is given by

$$I = \sqrt{\frac{j\omega C}{R}} tanh\left(\frac{\sqrt{j\omega RC}\ell}{2}\right) V_{in}, \tag{2.5}$$

where C is the unit capacitance, R the unit resistance and ℓ the length of the resistor. Local diffusions in the substrate can be p- or n-type. N-type diffusions inject noise through a reverse bias capacitance. P-type diffusions are often used as substrate taps or guard rings, to tie down the substrate to a desired potential. If designed improperly, these diffusions can inject very high levels of noise into the substrate, as they act as wide ground-planes on the substrate and any voltage bounce on these diffusions is conveyed throughout their extent on the chip through a very low impedance path. Guidelines for the design of guard rings can be found in great detail in Chapter 8.

3. SUBSTRATE RECEPTION MECHANISMS

Capacitive sensing is the most common mechanism of noise reception in surface devices, as bipolar transistors, capacitors, resistors and interconnect lines. The junction with substrate in lateral pnp devices consists of the n-type base region. If the pnp device is used in a gain stage, then the base of the device must be carefully shielded, or connected to a low impedance node. Otherwise the substrate noise will be amplified by the gain of the circuit.

In addition to capacitive pickup through the source and drain depletion junctions, MOS devices also exhibit a more severe form of substrate interaction due to the *body effect*. In MOS devices threshold voltage V_t is a strong function of the substrate potential. For a uniform surface impurity concentration N_A,

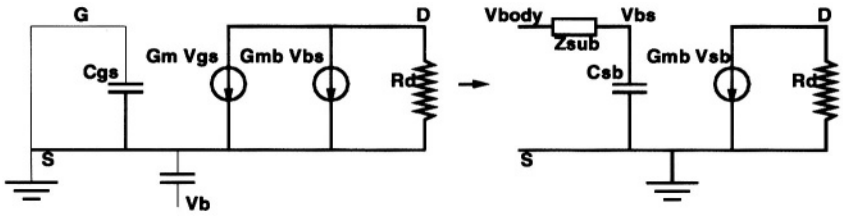


Figure 2.5. Body effect in MOSFETs

this dependence is given by [18]

$$V_t = V_{t0} + \frac{\sqrt{2q\epsilon N_A}}{C_{ox}} \left(\sqrt{2\phi_f + V_{sb}} - \sqrt{2\phi_f} \right) , \tag{2.6}$$

where ϵ is the substrate dielectric permittivity, N_A the substrate doping, C_{ox} the unit oxide capacitance, $2\phi_f$ the surface inversion potential and V_{sb} the source-to-body potential. The effect can be represented by a linearized model parameter g_{mb} in the small signal device model [1]. By shorting gate and source of transistor in Figure 2.5, a gain stage is created between substrate *S* and drain *D*. With suitable approximations [1] it can be shown that

$$\frac{g_{mb}}{g_m} = \frac{\sqrt{2q\epsilon N_A}}{2C_{ox}\sqrt{2\phi_f + V_{sb}}} , \tag{2.7}$$

where g_m is the small-signal transconductance of the device. Parameter g_m relates the drain current to the gate-to-source voltage. In typical processes the g_{mb}/g_m ratio varies from 0.1 to 0.3. The parasitic body-to-drain gain is thus only 14-20dB lower than the gate-to-drain gain. This fact makes MOSFET devices especially vulnerable to substrate noise reception at low to medium frequencies. On the contrary capacitive pickup, exhibited by most other devices, becomes significant only at frequencies above 1MHz.

4. DELAY EFFECT

So far we have outlined the effects of substrate noise coupling on mixed-signal circuit performance. Digital circuits are not immune from substrate noise. The noise is injected by logic gates during switching and glitch transients through impact ionization and capacitive coupling, and it is picked up by active devices via capacitive coupling and body effect. As a result the delay of the datapath may increase, thus possibly exceeding the predefined clock period. Such behavior is known as *delay effect.*

Gate delay $t_{50\%}$ is a function of several factors, including fanout, supply voltage, transistor geometry, input waveform, and charge excess caused by

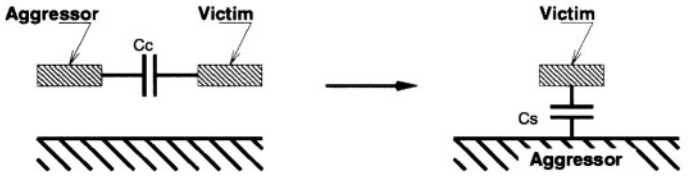


Figure 2.6. Analogy between interconnect crosstalk and substrate capacitive coupling

charge sharing effects. Ignoring the loading due to interconnect wiring, the gate delay is usually approximated by [21]

$$t_{50\%} = R_{EFF} \times C_G, \tag{2.8}$$

where $C_G = C_{OX} W L$ is the gate capacitance. R_{EFF}, the effective resistance, is the average transistor resistance during the output voltage swing and is proportional to R_{tr}, with

$$R_{tr} = \frac{L/W}{\mu C_{OX}(V_{DD} - V_T)}, \tag{2.9}$$

where L and W are the dimensions of the transistor, V_{DD} the supply voltage, C_{OX} the gate oxide capacitance, and V_T the threshold voltage. V_T is in turn proportional to the square root of the voltage V_{sb} applied between its substrate contact and the source. Hence,

$$t_{50\%} \propto \frac{1}{\sqrt{\phi_f + V_{sb}} - \sqrt{\phi_f}}. \tag{2.10}$$

Capacitive coupling also contributes to parasitic gate delay. The analysis of this effect is essentially identical to that of crosstalk between interconnect lines. In this case the aggressor is the substrate underneath the victim interconnect line or device. Figure 2.6 shows the coupling model equivalence, with coupling capacitances $C_c = C_s$. Resistor R_1 represents the impedance which holds the victim node at ground potential. Using standard analytic charge coupling models [21] one can estimate the charge noise present in the interconnect line due to substrate noise. Figure 2.7 shows the model proposed for a typical cross-coupling system. A close-form analytic model for the response voltage to charge injection $v_1(t)$ was derived in [21]. The voltage at the peak and the instant at which it occurs are respectively

$$t_p = \begin{cases} \tau_1 \\ \frac{\tau_2 \tau_1}{\tau_2 - \tau_1} ln(\frac{\tau_1}{\tau_2}) \end{cases} \quad v_1(t_p) = \begin{cases} \frac{C_c}{C_1 + C_c} e^{-1} & \text{if } \tau_1 = \tau_2 \\ \frac{C_c}{C_1 + C_c} (\frac{\tau_1}{\tau_2})^{\frac{\tau_1}{\tau_2 - \tau_1}} & \text{otherwise,} \end{cases} \tag{2.11}$$

where $\tau_1 = (R_1 || R_2)(C_1 + C_c)$ and the waveform of the aggressor node is assumed to be a decaying exponential step with time constant τ_2. From charge noise one can derive the extra delay present on a gate [21].

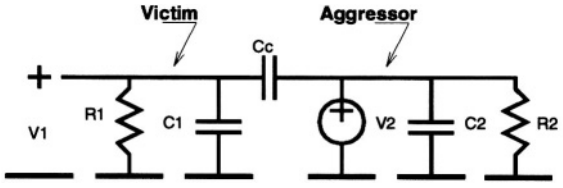


Figure 2.7. Standard cross-coupling model

Empirical delay models based on crosstalk have been proposed in the literature. One such model, relating the length of the parallel running wire l and the average spacing s to the extra delay $\triangle\tau$, was proposed in [22]. The model computes $\triangle\tau$ as

$$\triangle\tau = \frac{\alpha\, l^m}{s^n}, \tag{2.12}$$

where α is a fitting constant, while m and n were empirically observed to be near 2 and 1, respectively.

This page intentionally left blank

Chapter 3

ANALYSIS AND SIMULATION

Given a certain contact planimetry, the generation of the corresponding matrix of contact-to-contact impedances is not always the desired solution, due to the prohibitive computational complexity of calculating and storing it. Moreover, extracting the entire substrate impedance matrix is often unnecessary.

In this chapter we present a number of techniques for full or partial substrate extraction based on a rigorous analysis of the impedance matrix while it is formed. Macromodels for further simplification of the impedance matrix are also discussed in some detail. The material is presented in a consistent manner keeping in mind the acceleration and optimization techniques of the next chapters.

1. SUBSTRATE MACROMODELS

While it is possible to directly model the detailed transport and coupling processes in the substrate, usually it is desirable to obtain a compact representation of the interactions of circuit elements that couple through the substrate. A popular approach consists of creating equivalent circuits or simple analytical models, whose parameters are selected to fit either measured or simulated substrate conduction behaviors [11, 12, 23, 24].

Experiments are conducted on a small number of contacts, typically two or three. In such experiments, a known current is injected at one end and all the currents present in every other contact, including the backplate, are measured. The experiment is repeated for different geometries and the results are empirically fitted to the model based on a limited number of components.

An alternative approach is based on a numerical [7, 25, 26] or semi-analytical [27, 28, 29, 30] solution of the differential equations that describe substrate transport.

2. ELECTROMAGNETIC FORMULATION

The continuity equation

$$\nabla \bullet (\sigma \nabla \Phi(x,y,z,t)) + \frac{\partial}{\partial t}(\nabla \bullet (\epsilon \nabla \Phi(x,y,z,t))) = 0, \qquad (3.1)$$

where ϵ and σ are respectively the local dielectric permittivity and conductivity of the substrate, is the basic relation describing substrate transport. The equation characterizes the potential Φ at an arbitrary point in space (x, y, z) at time t. If the dielectric relaxation time, $\tau_e = \epsilon/\sigma$, is much smaller than any time-scale of interest, then the second term in Equation (3.1) may be neglected and the substrate treated as purely resistive. Except in very low-resistivity materials, this is typically a reasonable approximation up to around 2-5 GHz. As we mentioned, in general, silicon substrates are composed of differently doped layers of semiconducting material. Hence, areas of varying levels of conductivity are present in the vertical section of the chip, while lateral resistivity variations are due to device and well implants, as well as other integrated components.

To begin the discussion of electromagnetic analysis methods let us first consider regions of uniform material in the electrostatic approximation. In this case Equation (3.1) reduces to the Laplace equation

$$\nabla^2 \Phi = 0. \qquad (3.2)$$

The problem geometry and the substrate extraction process provide the boundary conditions necessary to solve Equation (3.2).

Contacts with the substrate are usually considered equipotential (or are subdivided until the equipotential assumption is valid). These regions will provide Dirichlet (known potential at given locations) boundary conditions for Equation (3.2). Other boundaries of the medium, such as edges, are treated by employing zero-normal-current (Neumann) conditions on the potential. At interfaces between regions of different conductivity, as shown in Figure 3.1, the current must be continuous, leading to the boundary condition

$$\sigma_+ \frac{\partial \Phi}{\partial n}|_+ = \sigma_- \frac{\partial \Phi}{\partial n}|_- \qquad (3.3)$$

where $\partial\Phi/\partial n$ refers to the derivative along vector $\hat{n}$ normal to the interface, and σ_+, σ_- refer to conductivity on opposite sides of the interface. To extract a column of the impedance matrix corresponding to a specific contact, the potential of that contact is set to 1 volt and the potential throughout the substrate is computed. The currents flowing into each of the other contacts, computed from integrating the normal derivative of the potential over each contact's surface, give the relevant mutual admittances.

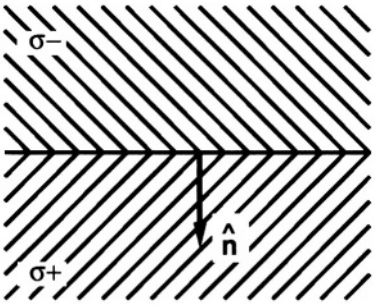


Figure 3.1. Interface between regions of different conductivity

When the dynamic part of Equation (3.1) is included, perhaps by defining a complex conductivity $\sigma' = \sigma + j\omega\epsilon$, the primary change to the procedure is that a more complicated model must be extracted. Either an equivalent circuit must be fit to data obtained by solving the differential equation at several points in the frequency domain[31, 32] or a model reduction procedure[33, 34, 35, 36, 37] must be performed. In either case, differential equations similar to the Laplace equation must still be solved and the numerical techniques are similar.

Methods for solving Equation (3.2) can be classified into two types, those based on a direct solution of the differential equation, and those that solve an equivalent integral form of the equation. Methods based on differential formulations have the advantage that they can easily analyze substrates with spatially-varying resistivities. When the material conductivities and permittivities are relatively homogeneous, then integral equation techniques are applicable[38, 39]. Generally, for integral methods to be competitive, the resistivity must vary only along one dimension, usually the vertical, and/or be piecewise-constant with only a few regions of different conductivity.

One way of deriving an integral equation is to start from a function $\Psi(\mathbf{r}, \mathbf{r}')$ that satisfies the underlying differential equation, e.g.

$$\nabla^2 \Psi(\mathbf{r}, \mathbf{r}') = \frac{1}{\sigma} \delta(\mathbf{r} - \mathbf{r}') \tag{3.4}$$

for the case of free standing charges. From Green's theorem[40] it follows that the potential $\Phi(\mathbf{r})$ at a point $\mathbf{r} = (x, y, z)$ can be expressed as

$$\Phi(\mathbf{r}) = \int_V \rho(\mathbf{r}') \Psi(\mathbf{r}, \mathbf{r}') d^3\mathbf{r}' + \sigma \oint_S (\Psi \frac{\partial \Phi}{\partial n} - \Phi \frac{\partial \Psi}{\partial n}) \, ds', \tag{3.5}$$

where $\partial/\partial n$ symbolizes the derivative with respect to $\hat{\mathbf{n}}$, the unit outward normal vector to surface S enclosing volume V and $\rho(\mathbf{r}')$ represents free sources in the volume (see Figure 3.2). Physically, $\Psi(\mathbf{r}, \mathbf{r}')$ represents the potential at $\mathbf{r}$ due to a point charge placed in $\mathbf{r}'$ and is called the Green's function.

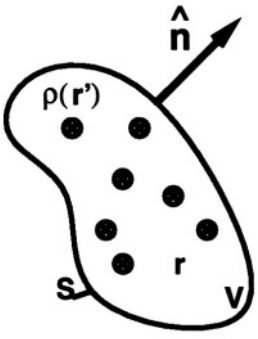


Figure 3.2. Physical meaning of Green's theorem

If the Green's function is known, Equation (3.5) allows one to determine the potential at any point in the volume V due to a known, arbitrarily-distributed charge density. If the distribution of the potential or normal derivatives are known over the region boundary, then Equation (3.5) can be used to write an equation that can be solved for the dual quantity. Note that since the currents are given by the product of the conductivity and the gradient of the potential,

$$\mathbf{J} = \sigma \nabla \Phi \tag{3.6}$$

solving for the normal derivatives of the field is equivalent to solving for the normal current densities.

3. BOUNDARY ELEMENT METHODS

In the electrostatic case, the problem of computing the resistance between a substrate contact and all the others can be translated into that of computing the charge at the contact when set at a potential of 1V, while the other contacts and the backplate contact are grounded. The reason for this is the following. Capacitance C_{ij} between contacts i and j is defined as the ratio of the charge on contact j to the potential of contact i, or $C_{ij} = Q_j/\Phi_i$. By Stokes' theorem,

$$C_{ij} = \frac{1}{\epsilon} \oint_S \mathbf{E} \bullet \hat{\mathbf{n}} ds, \tag{3.7}$$

where $\hat{\mathbf{n}}$ is the unit outward normal vector to the surface S which encompasses the contact. $\mathbf{E}$ is the electric field in the medium. Similarly, the resistance between contacts is defined as

$$R_{ij} = Y_{ij}^{-1} = \left[-\sigma \oint_S \mathbf{E} \bullet \hat{\mathbf{n}} ds \right]^{-1} = -\frac{1}{\sigma \epsilon} \frac{\Phi_i}{Q_j}, \tag{3.8}$$

where σ is the medium conductivity. Note that in both the resistive and the capacitive cases the potential satisfies the Laplace equation, thus the problems can be interchanged freely.

At frequencies up to 4-5GHz, substrate susceptance is typically much smaller than the conductance, hence it may be ignored and all substrate impedances may be considered real. Consider the problem of computing the resistance between contacts **1** and **2**, and toward ground in Figure 3.3. This represents a mixed-boundary problem, since zero potential in the chip's backplate contact is assumed (Dirichlet condition) and vanishing normal electric field on the other faces (Neumann condition). Under these conditions, Equation (3.5) simplifies to

$$\Phi(\mathbf{r}) = \int_V \rho(\mathbf{r}')G(\mathbf{r}, \mathbf{r}')d^3\mathbf{r}', \tag{3.9}$$

where V is the chip's volume region and $G = \Psi$ the Green's function. The potential of a contact is computed as the result of averaging all internal contact partitions. Hence, using (3.9) the potential of contact i can be derived as $\bar{\Phi}_i = \frac{1}{V_i}\int_{V_i}\int_{V_j}\rho_j G dv_j dv_i$, V_i and V_j being the volumes of contacts i and j and ρ_j the charge distribution on j. If a uniform charge distribution $\rho_j = Q_j/V_j$ is chosen over j, we obtain

$$\bar{\Phi}_i = \frac{Q_j}{V_j V_i}\int_{V_i}\int_{V_j} G dv_j dv_i. \tag{3.10}$$

The solution to Equation (3.9) for each contact pair yields the *coefficient of potential* matrix **P**. The relation between matrix **P** and vector $\bar{\mathbf{\Phi}}$, the average potential at each contact, and **Q**, the charge associated with all contacts, is described as

$$\bar{\mathbf{\Phi}} = \mathbf{PQ} \quad \text{and} \quad \mathbf{Q} = \mathbf{c}\bar{\mathbf{\Phi}}, \tag{3.11}$$

where $\mathbf{c} = \mathbf{P}^{-1}$ is called *coefficient of induction* matrix. For a contact i, the capacitance to ground C_i and all mutual capacitances C_{ij} are characterized as

$$C_i = \sum_{j=1}^{N} c_{ij}, \quad C_{ij} = -c_{ij}, \tag{3.12}$$

where N is the size of matrix **c**. Using Equations (3.7) and (3.8) in combination with relations (3.12), all mutual and ground resistances can be easily derived.

4. GREEN'S FUNCTION COMPUTATION

In the absence of any boundaries, that is in the free-space case, the function $\Psi(\mathbf{r}, \mathbf{r}')$ reduces to $1/(4\pi\sigma|\mathbf{r} - \mathbf{r}'|)$. In principle, the free-space Green's function may be used for substrate extraction calculations, however, the boundary

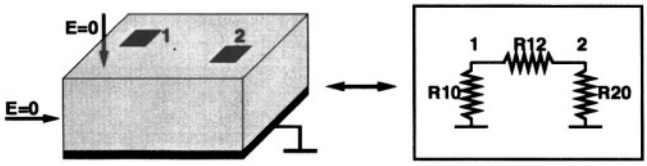


Figure 3.3. Substrate boundaries and contact resistance modeling

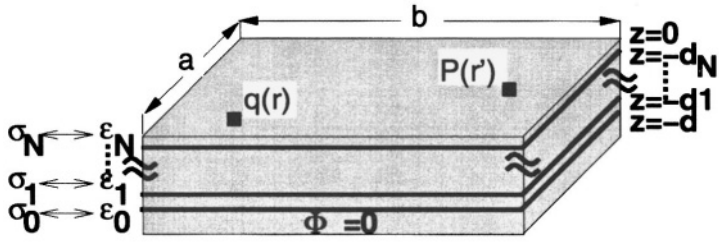


Figure 3.4. Multi-layer doping profiles

conditions at domain boundaries must be explicitly enforced, which implies discretizing the boundaries[41]. In the substrate analysis problem, where there are multiple substrate layers, each with a different conductivity, it is usually more convenient to derive a Green's function tailored to the layered-media boundary conditions. These Green's functions incorporate any effects due to vertically-varying conductivity and possibly finite extent of the substrate. For example, Equation (3.6) must be satisfied at each layer boundary where the conductivity changes, and Figure 3.3 shows a finite domain where the normal field must be zero at the top and edges.

Using the layered-media the Green's function simplifies the numerical procedure considerably, since the integral equation only needs be written over the multiply-connected surface defined by substrate contacts that are usually in the top layer of the material. However, the price paid for this simplification is that the Green's function can become complicated and expensive to compute. Possible methods for evaluating the Green's function include image-based techniques[42, 27, 28], separation-of-variables (SOV)[15, Chapter 3], and spectral domain analysis [43].

The full derivation of the Green's function for multi-layered problems can be found in [15, 29]. Here, we shall outline the basic steps to justify the sensitivity analysis and some optimization techniques proposed in this book. Figure 3.4 shows the multi-layered structure for which a Green's function must be computed. The figure shows for each layer i its conductivity σ_i and its permittivity ϵ_i associated with the equivalent electrostatic problem. Consider

the case in which the point-charge at $\mathbf{r} = (x, y, z = 0)$ and the observation point at $\mathbf{r}' = (x', y', z' = 0)$ are localized to a layer with dielectric permittivity ϵ_N. The Green's function corresponds to an infinite series of sinusoidal functions

$$\begin{aligned} G(r, r') = G_0|_{z=z'=0} + \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} f_{mn} C_{mn} \\ \times cos(\tfrac{m\pi x}{a}) cos(\tfrac{m\pi x'}{a}) cos(\tfrac{n\pi y}{b}) cos(\tfrac{n\pi y'}{b}), \end{aligned} \tag{3.13}$$

where $C_{mn} = 0$ for $m = n = 0$, $C_{mn} = 2$ for $m = 0$ or $n = 0$, but $m \neq n$, and $C_{mn} = 4$ for all $m, n > 0$. Parameters a, b and d are the dimensions of the substrate in x-, y- and z-direction (see Figure 3.4). Formulae for terms $G_0|_{z=z'=0}$ and f_{mn} can be found in [15] and [44]. From Equation (3.10), adapted for surface contacts, one can derive an expression for the average potential at contact i due to the charge on contact j.

$$\bar{\Phi}_i = \frac{Q_j}{S_j S_i} \int_{S_i} \int_{S_j} G(s_j, s_i) ds_j ds_i, \tag{3.14}$$

Consequently, the entry p_{ij} of matrix $\mathbf{P}$, computed as the ratio of $\bar{\Phi}_i$ and Q_j, becomes

$$p_{ij} = \frac{\bar{\Phi}_i}{Q_j} = \frac{1}{S_i S_j} \int_{S_i} \int_{S_j} G(s_j, s_i) ds_j ds_i, \tag{3.15}$$

where S_j and S_i are the surfaces of the contacts. Replacing (3.13) into (3.15) and integrating, one obtains an explicit formula for p_{ij}.

$$\begin{aligned} p_{ij} = G_0|_{z=z'=0} + \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} k_{mn} \\ \times \frac{\left[sin(m\pi \frac{a_2}{a}) - sin(m\pi \frac{a_1}{a})\right]\left[sin(m\pi \frac{a_4}{a}) - sin(m\pi \frac{a_3}{a})\right]}{(a_2 - a_1)(a_4 - a_3)} \\ \times \frac{\left[sin(n\pi \frac{b_2}{b}) - sin(m\pi \frac{b_1}{b})\right]\left[sin(m\pi \frac{b_4}{b}) - sin(m\pi \frac{b_3}{b})\right]}{(b_2 - b_1)(b_4 - b_3)}, \\ \text{with } k_{mn} = \frac{a^2 b^2 f_{mn} C_{mn}}{m^2 n^2 \pi^4}. \end{aligned} \tag{3.16}$$

Parameters (a_1, a_2) and (b_1, b_2) are the x- and y-coordinate of node i and (a_3, a_4) and (b_3, b_4) those of node j. Appropriately rewriting the second term of (3.16), after proper scaling, as a cosine series we obtain

$$\sum_{m=0}^{\infty} \sum_{n=0}^{\infty} k_{mn} \, cos(m\pi \frac{a_{1,2} \pm a_{3,4}}{a}) \, cos(n\pi \frac{b_{1,2} \pm b_{3,4}}{b}) \tag{3.17}$$

which is a compact representation of a sum of 64 terms forming all possible combinations of signs and indices. By replacing the ratios of contact coordinates and the substrate dimensions with ratios of integers $\tilde{p}, \tilde{q}$ and summing over finite limits $\tilde{P}$ and $\tilde{Q}$, term (3.17) becomes

$$K(\tilde{p}, \tilde{q}) = \sum_{m=0}^{\tilde{P}-1} \sum_{n=0}^{\tilde{Q}-1} k_{mn} cos(m\pi \frac{\tilde{p}}{\tilde{P}}) cos(n\pi \frac{\tilde{q}}{\tilde{Q}}), \tag{3.18}$$

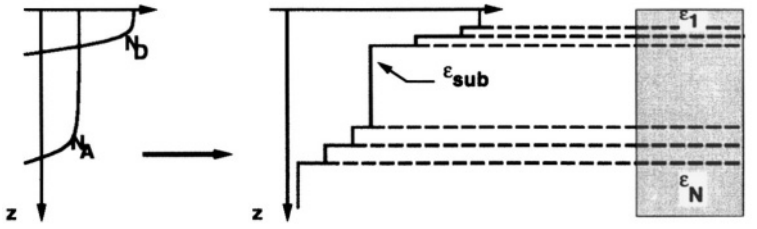


Figure 3.5. Discretization of non-abrupt doping profiles

a two-dimensional discrete cosine transform (DCT) of k_{mn}. Hence, the computation of p_{ij} ultimately requires only a simple DCT [15, 29]. Several techniques exist for efficient computation of the DCT, e.g. FFT based techniques only require a computation complexity $O(\tilde{P}\tilde{Q}\ log_2\tilde{P}\tilde{Q})$. Note that the value of k_{mn} is solely dependent on the properties of the substrate in z-direction. Hence, for a given substrate structure the DCT needs be derived **only once**. Any modification in the relative position of one or more nodes is captured completely by the Fourier transform, thus only matrix **P** needs be calculated and inverted. However, due to the relatively small size of **P**, typically 50-5,000, this process does not require a significant CPU time. Non-abrupt doping profiles can be analyzed at low CPU cost by simply discretizing in z-direction with a gradually changing value of permittivity as shown in Figure 3.5. The methodology denominated SUBRES[29, 13] incorporates all these techniques, including the acceleration methods which will be discussed in Chapter 4.

5. COMPUTATIONAL TECHNIQUES

Discretization of the equations governing the substrate can potentially generate large systems of linear equations, and sophisticated techniques are required for their efficient solution. Fortunately, the solution of Equation (3.2) is one of the most well-studied problems in applied mathematics literature, and a variety of methods are available to solve the linear equations that result from the typical discretizations.

Finite difference[5] or finite element[4] techniques are typically used to discretize Equation (3.2). For example, the simplest finite difference prescription involves converting Equation (3.2) into a set of algebraic equations by replacing the derivatives of Φ by the differences

$$\frac{\partial^2\Phi}{\partial^2 x} = \frac{\Phi_{i+1,j,k} - \Phi_{i-1,j,k} - 2\Phi_{i,j,k}}{\triangle_i^2}\ , \tag{3.19}$$

where similar substitutions are made for y- and z-direction, $\triangle_i$, $\triangle_j$, $\triangle_k$ being the spacing of the finite-difference grid at the grid point indexed by $\{i, j, k\}$. The resulting system of equations is usually very sparse since only a few

elements in each row are nonzero. For example, in the simple discretization of Equation (3.19), each row contains seven nonzero elements. The first option is to apply standard techniques for the solution of sparse linear systems[45], such as direct factorization methods based on Gaussian elimination. This approach can be prohibitively expensive in time and storage, unless relatively few mesh nodes exist in the vertical direction. This is due to the large degree of matrix fill that occurs during the factorization of a matrix that derives from a three-dimensional mesh. Usually an iterative matrix solution algorithm is needed.

Modern iterative algorithms are usually based on Krylov-subspace algorithms[46] such as the conjugate-gradient or GMRES methods. For the discretization of elliptic differential equations, preconditioning is required to achieve convergence in a reasonable number of iterations. Incomplete factorization[47, 46] preconditioners are popular, but preconditioners based on multigrid[48] or multiresolutional ideas[49] can be considerably more effective. In some cases rapid elliptic solvers (see [50] for references) can form the basis of good preconditioners[51].

In the engineering community, the numerical solution of electromagnetic integral equations is usually done via method-of-moment[52] or boundary-element techniques, though there has been some recent progress with Nystrom methods[53]. The simplest such scheme is to discretize the domain of the integral (in this case, the substrate contacts) into a number of polygonal sections called panels. Given Dirichlet boundary conditions on the panels, the unknowns are the injected currents, and on each panel the injected current is assumed to be constant. The potential of a panel is defined as the result of summing over the contribution from current injected by all the other panels in the domain and averaging the potential over the panel. Hence, using the simplified version of (3.5) the potential of contact i can be derived as

$$\bar{\Phi}_i = \sum_j \frac{I_j}{A_j A_i} \int_{A_i} \int_{A_j} G(\mathbf{x}, \mathbf{x}') da_j da_i = \sum_j Z_{ij} I_j \tag{3.20}$$

where the sum runs over all panels j, A_i and A_j are the areas of contacts i and j respectively, and I_j is the current injected from panel j, and the integral is over the panel surfaces. This procedure produces a matrix equation

$$\mathbf{Z}\mathbf{I} = \bar{\mathbf{\Phi}} \tag{3.21}$$

where the matrix $\mathbf{Z}$ is *dense*, that is, every entry is nonzero because a normal current injected from any panel induces a potential at *every* other panel in the substrate. A solution of this matrix equation is required for every column in the compact admittance or impedance matrix that is to be extracted.

In realistic problems, the matrix in Equation (3.21) can be quite large. Constructing and directly inverting the full $\mathbf{Z}$ matrix for the entire substrate contact

configuration can be prohibitively expensive, and so more efficient methods have been sought by many authors. Physically based heuristics involving approximations to the inverse of the $\mathbf{Z}$ matrix[15, 54, 26] can accelerate the matrix solution process as well as the following nonlinear simulation. Numerical stability and error control in these procedures can be difficult to quantify, however.

More rigorous analysis acceleration techniques typically exploit the analytic properties of the Green's function. For problems with bounded domains, the multilayer Green's function can be computed in $O(n\ log\ n)$ time using fast cosine transform (FCT) techniques. The FCT can be used to build a technology-dependent table that is used to accelerate the matrix construction procedure for one of the direct techniques. When combined with a matrix simplification procedure and low-rank update techniques[13], the overall procedure can be effective, particularly when embedded in the loop of an optimization procedure.

When direct techniques are no longer feasible, iterative matrix solution algorithms such as GMRES[55] must be used. The dominant computational cost in such an algorithm is the computation of a matrix-vector product with the matrix $\mathbf{Z}$. The speed of the FCT can be exploited to directly compute the matrix-vector products in such an iterative procedure [56] in nearly optimal time, if the contacts are uniformly subdivided and are fairly densely spaced. For complicated contact distributions, several algorithms have been developed that can compute a matrix-vector product in close to $O(n)$ time and memory by approximating the action of the matrix $\mathbf{Z}$.

Most of the matrix acceleration algorithms are based on the fact that the potential induced by an injected current has a spatially complex profile only near the injection source. Far away from the source it can be easily approximated. The approximations involved can always be controlled to within a user-specified tolerance. Several approaches have been developed in the context of layered-media and substrate analysis. FCT and FFT related techniques can be applied by using local corrections [57, 58] that remove any constraints on the relation between the FCT/FFT grid and the underlying discretization. The authors of [30] have developed an algorithm that interpolates the Green's function in a hierarchically spatially-decomposed manner, and then uses an SVD-like procedure to further compress the interpolation elements.

Recently, algorithms have been developed that combine the matrix approximation with an acceleration of the iterative matrix solution procedure itself. The multigrid method of [59] is based on constructing a hierarchical representation of the irregular problem domain. At each level of hierarchy, a coarser representation of the discretized problem is constructed by using a geometric moment-matching scheme to approximate the rough features of the finer geometry. The coarser grid problems can be solved relatively cheaply, so the solutions to the coarse grid problems are used to accelerate the iterative solution

of the linear systems on the finer levels. The convergence of the iterative solver is extremely rapid, requiring only a few iterations to converge to engineering tolerances. A similar multiresolution approach was described in [60], where a wavelet-like basis for the panel unknowns is constructed by matching moments of the multipole field expansions. The wavelet-like basis is used to perform rapid matrix-vector products and also provides a natural, and very effective, preconditioner.

This page intentionally left blank

Chapter 4

SUBSTRATE MODELING

In Chapter 3 we have seen how complex substrate configurations can be analyzed relatively fast. Substrate-aware optimization requires yet a much faster evaluation of substrate noise at a global level. This goal is often accomplished with the use of compact substrate models, which capture only the relevant aspects of noise injection. Similar models can be built for noise injected into supply lines in form of current or voltage ripple.

In this chapter various modeling techniques are outlined for both types of noise effects. The techniques generally present a trade-off between simplicity and accuracy, thus ultimately impacting the computational efficiency of noise evaluation.

1. SWITCHING NOISE AND NOISE SIGNATURES

A signal transition occurring in a typical logic gate causes a spike of current to be absorbed from the supply to charge a load. A similar spike traveling toward ground is generated when the load is discharged. A significant portion of transitional current is discharged to ground through direct feedthrough. Spurious currents can also be injected directly into the substrate through various mechanisms, as described in Chapter 2. The cumulative effect of spurious microcurrents absorbed/discharged by switching gates, is referred to as *switching noise*. Switching noise can quickly travel through interconnect coupling, power/ground buses and substrate, to be picked up by sensitive devices through capacitive coupling and body effect.

Any given circuit injects a unique current waveform into the substrate as a direct consequence of switching noise. Such waveform, which we will refer to as *substrate noise signature,* is dependent on the circuit implementation, technology and input vector set. Switching noise is also responsible for ripple currents at power and ground buses. Such currents, known cumulatively as

supply noise signature, cause various non-idealities in digital and analog blocks by modulating both supply and bias voltages. Approximate substrate and supply noise signatures can be used *au lieu* of extensive SPICE simulations of extracted substrate impedance matrices. Noise signatures can be useful during early architectural phases to select the appropriate building blocks and their specifications. They can also be used during high-level physical design, i.e. floorplanning and block placement, to determine the optimal topology of a layout and to size power/ground buses.

Early models of noise signatures were based on a single Gaussian white or pink noise current or voltage source connected to the substrate and the supply. The underlying assumption was that the global switching activity of the circuit is uniformly distributed over a large section of the spectrum. Substrate/supply noise signatures have also been modeled in the literature [7, 8] as a capacitively and/or resistively coupled current or voltage generator whose waveform was derived from the circuit's global clock. The accuracy of these models is often the limiting factor in circuit performance evaluation. Simple approximations for injected noise often capture only a relatively small portion of the entire noise energy spectrum. Thus, potentially detrimental noise components may be underestimated.

An alternative to white/pink or clock-synchronous noise modeling, consists of creating traditional multi-port substrate models with parameterization. The goal is that of obtaining simple attenuation formulae for all noise sources, which can be thus represented in full detail. To date, this is the single most used approach in the design community [23, 11, 12].

More recently, a new methodology called SUBWAVE has been proposed to accurately capture substrate noise signatures [16, 17]. SUBWAVE is based on a two-phase process. First, all the cells of a given library are characterized in terms of the current they inject into the substrate. Then, the switching activity of the circuit is combined with the library's substrate injection patterns to build the complete map of the total substrate noise signature.

2. USE OF NOISE SIGNATURES

During floorplanning, specific well-isolated areas can be allocated to noisy circuits. Minimum distance requirements can be computed based on the overall noise spectral energy produced by such circuits and the maximum levels of spurious energy tolerated by sensitive circuits [61]. When space is not available, specific guard rings can be designed to block those frequencies in the spectrum which could interfere with the operation of surrounding circuits. The design of guard rings and other blockage devices can be tuned to work optimally for problematic noise spectra. Guard rings and their effects on design are extensively discussed in Chapter 8.

Rapid characterization of injected noise can be used to test whether redesigned logic blocks are compatible with existing circuitry or if special measures - including further redesign - must be taken. Similarly, spectral characterization of substrate noise could be provided as part of intellectual property interface description, along with the block basic functionality, to reduce the risk of system failure due to unexpected second-order effects.

The efficient generation of substrate noise signature models can be used to drive logic synthesis in circuits which have a limitation in the amount of noise they can produce. A model of the performance degradation due to the effects of noise at specific frequencies can be embedded in the synthesis tool or used to assist a designer. Electromagnetic compatibility requirements for block and systems can be tested at or before actual integration, or *a posteriori* to verify existing problems and causes. Finally, substrate noise signatures can be used as fingerprints for fault analysis and diagnosis.

3. MULTI-PORT SUBSTRATE MODELS

Multi-port models can be obtained by full extraction of substrate impedance matrix in combination with sensitivity analysis. Computing the sensitivity of substrate coupling with respect to a number of technology parameters is useful for several reasons. First, it allows to evaluate the effects of slight imperfections in the fabrication process on the performance of a circuit and, ultimately, its yield. Second, it can be used for the selection of the best cost-effective technology on the basis of the class of circuits one wants to fabricate with given specifications. Finally, the technique can be used during substrate-aware optimization to help the decision process by providing a **trend** to the best possible improvement as suggested in [62]. Trend analysis is discussed in detail in Chapter 6.

In Chapter 3 we have shown how resistive substrate couplings can be efficiently computed using the Green's function and the DCT. In this section we develop the theory for the computation of substrate coupling sensitivities with respect to doping profiles and geometries. The relation between circuit performance K and technology, via substrate-related parasitics, is obtained using the following expression

$$\triangle K = \sum_{\ell} \frac{\partial K}{\partial T_\ell} \triangle T_\ell, \text{ with } \frac{\partial K}{\partial T_\ell} = \sum_{i,j} \frac{\partial K}{\partial Y_{ij}} \frac{\partial Y_{ij}}{\partial T_\ell}, \tag{4.1}$$

where (i, j) represents a contact pair, Y_{ij} the substrate conductive coupling between i and j, and T_ℓ a technology parameter. Hence, assuming $\frac{\partial K}{\partial Y_{ij}}$ exists[1], $\triangle K$ can be easily evaluated as a linear function of technology parameters T_ℓ, provided that term $\frac{\partial Y_{ij}}{\partial T_\ell}$ has been computed. This term is generally ignored due

to the extremely high complexity of its evaluation if traditional finite difference methods are employed.

Assume that the capacitive problem has been solved and that the equivalent resistive network has been computed from the coefficient of induction matrix **c**. Furthermore, let **c** be scaled in such a way that the node-to-node conductance Y_{ij} and the ground conductance Y_{ii} can be computed directly using

$$Y_{ii} = \sum_{j=1}^{N} c_{ij}, \quad Y_{ij} = -c_{ij}. \tag{4.2}$$

Let us define **Y** as a NxN matrix consisting of Y_{ii} on the diagonal and Y_{ij} everywhere else. Let us call $\partial\mathbf{Y}/\partial T_\ell$ the sensitivity of matrix **Y** with respect to technology parameter T_ℓ. The components of the sensitivity matrix are terms $\partial Y_{ii}/\partial T_\ell$ on the diagonal and $\partial Y_{ij}/\partial T_\ell$ everywhere else. The terms are computed using

$$\frac{\partial Y_{ii}}{\partial T_\ell} = \sum_{j=1}^{N} \frac{\partial c_{ij}}{\partial T_\ell}, \text{ and } \frac{\partial Y_{ij}}{\partial T_\ell} = -\frac{\partial c_{ij}}{\partial T_\ell}. \tag{4.3}$$

Recall that N is the size of matrix **c**. In order to derive $\partial c_{ij}/\partial T_\ell$, Equation (3.11) is differentiated on both hand-sides and solved with respect to $\partial\mathbf{Q}/\partial T_\ell$. Using the fact that $\partial\bar{\bar{\mathbf{\Phi}}}/\partial T_\ell$ vanishes, we obtain

$$\frac{\partial \mathbf{Q}}{\partial T_\ell} = -\mathbf{P}^{-1}\left(\frac{\partial \mathbf{P}}{\partial T_\ell}\mathbf{Q}\right). \tag{4.4}$$

Using the definition of c_{ij} we obtain

$$\frac{\partial c_{ij}}{\partial T_\ell} = \frac{1}{\mathbf{\Phi}_j}\frac{\partial Q_i}{\partial T_\ell}, \tag{4.5}$$

where $\partial Q_i/\partial T_\ell$ is computed using Equation (4.4). Now, only the derivative $\partial\mathbf{P}/\partial T_\ell$, i.e. $[\partial p_{ij}/\partial T_\ell]_{i,j=1,\ldots,N}$, remains to be computed. From Equation (3.16), assuming zero-depth contacts and $T_\ell \neq \epsilon_N$, d, a or b

$$\begin{aligned}\frac{\partial p_{ij}}{\partial T_\ell} = {} & \frac{\dot{\Gamma}_N\beta_N - \Gamma_N\dot{\beta}_N}{ab\epsilon_N{\beta_N}^2} + \sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\dot{k}_{mn} \\ & \times \frac{\left[sin(m\pi\frac{a_2}{a}) - sin(m\pi\frac{a_1}{a})\right]\left[sin(m\pi\frac{a_4}{a}) - sin(m\pi\frac{a_3}{a})\right]}{(a_2-a_1)(a_4-a_3)} \\ & \times \frac{\left[sin(n\pi\frac{b_2}{b}) - sin(m\pi\frac{b_1}{b})\right]\left[sin(m\pi\frac{b_4}{b}) - sin(m\pi\frac{b_3}{b})\right]}{(b_2-b_1)(b_4-b_3)},\end{aligned} \tag{4.6}$$

where $\dot{\Gamma}_N = \partial\Gamma_N/\partial T_\ell$, $\dot{\beta}_N = \partial\beta_N/\partial T_\ell$ and $\dot{k}_{mn} = \partial k_{mn}/\partial T_\ell$. Expressions for Γ_N, β_N, k_{mn} for all-depth contacts have been derived in [15]. The calculation of the derivatives can be found in Appendix B.

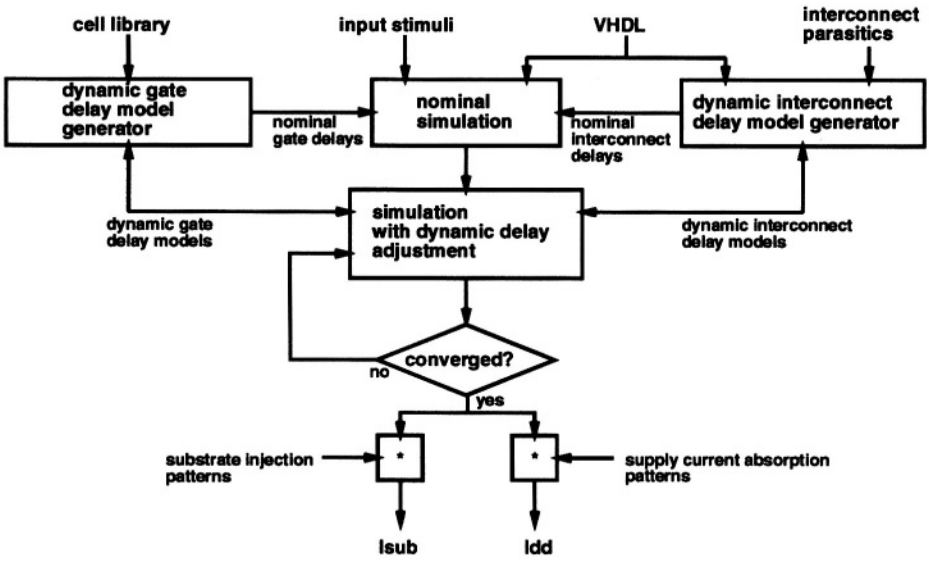


Figure 4.1. Flow of the noise signature model generator

The first term of (4.6) can be easily calculated from the formulae in the appendix, while the second term can be efficiently computed using the DCT by replacing k_{mn} with $\dot{k}_{mn}$ in Equation (3.18). The DCT can be computed for each location in the grid and repeated for all parameters $T_\ell, \ell = 1, \ldots, N_T$, where N_T is the number of technology parameters considered. Notice that this calculation need be performed **only once** for a given substrate structure.

To generate matrices $\partial \mathbf{c}/\partial T_\ell$ and $\partial \mathbf{P}/\partial T_\ell$, it is necessary to compute sensitivities $\partial p_{ij}/\partial T_\ell$ and $\partial c_{ij}/\partial T_\ell$ for all pairs of partition elements composing each contact. Every sensitivity measure requires additional *N*x*N* storage. As an example, assume $N_T = 10$, i.e. ten technology parameters T_ℓ are considered, moreover assume that a grid of 1024x1024 points is used. Then, the total storage needed by our approach is 41.9 MByte, which is relatively low considering that a $1\mu m$ resolution would be achieved on a 1**x**1mm chip size.

4. GENERATING NOISE SIGNATURES

A noise signature model generator was built, as part of the package SUBRES, using some of the techniques proposed in [16, 17]. The overall flow of the tool is described in Figure 4.1. The method consists of three phases. First, a dynamic delay model (DDM) is constructed to characterize the noise injection and delay of a variety of circuit components, such as gates, custom macrocells, and interconnect configurations. In the figure two DDM generators are shown, one for gates and macrocells and one for interconnect. Each model is parameterized

for supply ripple amplitude, input signal slope, and fanout loading. The effects of signal crosstalk on interconnect delays are also captured by the model since the interconnects are now modeled as circuit elements with possibly multiple fanin/fanouts and internal logic. Extending the model to account for more dramatic effects of signal crosstalk, such as the temporary flip-over of signal values, is straightforward.

Secondly, based on the initial estimates of the delays, an event-driven simulator computes the switching activity at every node of the circuit based on some user-defined input stimuli. The simulator, based on the principles described in [63], accepts VHDL as circuit description language and a *ad hoc* format for the DDM description. During the first iteration the gate and interconnect delays used by the simulator are the nominal ones as defined by the DDM. The result of this simulation is a *switching trace,* i.e. a collection of time-varying signals $s_n(t)$, one for each node n in the circuit, of the form

$$s_n(t) = \sum_{k=1}^{K_n} A_k \delta(t - t_k), \tag{4.7}$$

where K_n is the number of observed switching transitions for node n, t_k represents the kth transition time, and $A_k = \{+1, -1\}$ its direction (low-to-high or high-to-low). In any given circuit and for a given input vector, there exists N distinct switching traces, where N is the total number of nodes in the circuit. The switching trace is used to recompute the DDMs for all instantiated standard cells, which will then be utilized by the next iteration. The process continues until convergence is reached, i.e. no more DDM updates are necessary.

The third phase consists of computing the cumulative noise injection of the circuit, i.e. substrate noise signature I_{sub}, as a convolution of each node's activity with the substrate injection pattern of each instantiated component.

$$I_{sub} = \sum_{n=1}^{N} s_n(t) * i_{sub_n}(t) = \sum_{n=1}^{N} \sum_{k=1}^{K_n} A_k \delta(t - t_k) * i_{sub_n}(t), \tag{4.8}$$

where $i_{sub_n}(t)$ is the substrate injection pattern associated with node n and the operator "*" represents the convolution. As a byproduct, one can also compute the ripple current of each gate and, cumulatively, the supply noise signature I_{dd}. This signature is obtained as follows

$$I_{dd} = \sum_{n=1}^{N} s_n(t) * i_{dd_n}(t) = \sum_{n=1}^{N} \sum_{k=1}^{K_n} A_k \delta(t - t_k) * i_{dd_n}(t), \tag{4.9}$$

where $i_{dd_n}(t)$ is the supply current absorption pattern associated with node n. Both the substrate injection and supply current absorption patterns can

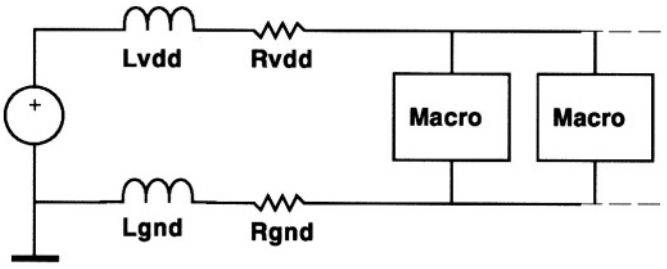


Figure 4.2. Power supply network

be computed explicitly through simulation [16], or by fitting measurement or simulated data onto analytical models.

The key factors that determine the efficiency and accuracy of the process are the quality of the DDMs and the convergence properties of the iterative refinement process. To achieve an efficiency close to that of gate-level logic simulation, the delay models must evaluate fast and be integrated properly into event-driven simulation. To deliver results with acceptable accuracy, the delay models must accurately characterize the effects of digital noise on gate/interconnect delays, and the iterative refinement process goes in the right direction.

The original implementation in SUBRES included the methodology described above applied to substrate noise injection. It was later extended to supply ripple without significantly alter the general implementation architecture.

4.1 CREATING DDMS

Switching noise produced by the cumulative switching activity of the circuit is propagated to all circuit components through the power supply network and the substrate. This behavior can be modeled in terms of ripple signal feeding every component through inductive and resistive leads, as shown in Figure 4.2. The current pulse resulting from a single gate switching ranges from $1\mu A$ to $100\mu A$ with a width of one half to five times the typical delay. In digital circuits, switching noise is the aggregate effect of many gate switching at around the same time. This aggregate effect changes the delay of a switching gate by producing a drop of the power supply V_{DD} and/or a raise of the ground level V_{GND}.

The DDM relates a gate, macrocell, or interconnect delay to ripple amplitude in V_{DD} and V_{GND} (a sinusoidal or continuous ripple is assumed), input rise time (linear ramps are assumed with discrete slopes), fanout (an integer multiple of the loading generated by an inverter is assumed). The test setup shown in Figure 4.3 is used to derive data for building the DDM for a particular gate,

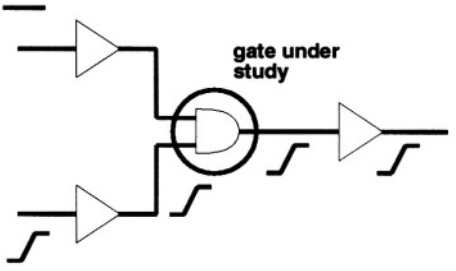


Figure 4.3. Test setup to evaluate a DDM

in this case an AND. The two intermediate signals in Figure 4.3 are used in measuring the delays. The most straightforward way to build such models is to use lookup tables. For every gate, tables are built for each driving input/output rise/fall combination, varying V_{DD} from a minimum to a maximum around its typical value. The inputs must also be devised based on the behavior (low-to-high/high-to-low) and speed (rise/fall slew reate). All possible input combinations must hence be considered.

The main disadvantage of using lookup tables is their space complexity. Upon examining the simulation data, it was found that one cannot use superposition to separate changes in V_{DD} and V_{GND} and reduce the space complexity of the tables. However, the change in gate delays is sufficiently smooth with respect to V_{DD} and V_{GND}, thus prompting use of 2-dimensional curve fitting to build delay equations for the DDM. The same observation was extended to variable input slopes and fanout. The equation below shows an example of a model for one particular input/output rise combination in the case of a NAND gate with simple fanout

$$\begin{aligned}\tau_{MDD}[sec] = {} & 6.25 \times 10^{-10} + 1.25 \times 10^{-11}x - 1.11 \times 10^{-11}y \\ & +1.99 \times 10^{-13}x^2 + 4.32 \times 10^{-14}y^2 - 7.50 \times 10^{-14}xy \\ & -3.20 \times 10^{-15}yx^2 + 4.00 \times 10^{-16}y^2x - 8.00 \times 10^{-18}y^2x^2,\end{aligned} \tag{4.10}$$

where $x = \triangle V_{DD}/0.05$, $y = \triangle V_{GND}/0.05$, and 0.05 is a normalization factor. A gate delay is not affected if the changes in V_{DD} and V_{GND} do not occur within a certain sensitive window around the time at which the gate switches. For each gate, the opening of this sensitive window is also determined as part of the DDM.

The substrate injection and supply current absorption patterns are obtained from SPICE simulations using custom-fitted device models. The obtained current patterns $i(t)$ are then fitted onto a second order filter impulse response [64]

$$i(t) = Ae^{-t/\tau}cos(\omega t - \phi) \tag{4.11}$$

A	ω	τ	ϕ
100 μA	1.3 sec^{-1}	6.3 nsec	0

Table 4.1. Fitting parameters for an analytical model of a typical absorption pattern

whose parameters are amplitude A, resonance frequency ω, damping factor τ and phase ϕ. As an example, consider the CMOS inverter from Figure 2.1, which has been integrated into a high-resistivity substrate similar to that of Figure 2.2(a). By fitting the waveform in the plot of Figure 2.3, we obtain the parameter values listed in Table 4.1.

4.2 ITERATIVE REFINEMENT PROCESS

After the initial event-driven simulation, which uses nominal delays for gates and interconnect, a series of event-driven simulations are performed, where delays are adjusted at each iteration based on DDMs. The process of integrating the DDM into event-driven simulation requires changing gate delays and extracting a signal trace during simulation.

To properly handle these tasks a basic simulator was built and used in both the nominal and iterative process of Figure 4.1. The iterative refinement process consists of the following sub-processes, performed in each node:

1. store trace from previous iteration
2. if the switching instant is in gate sensitivity window, then compute $\triangle V_{DD}$ and $\triangle V_{GND}$
3. based on previous trace, $\triangle V_{DD}$, and $\triangle V_{GND}$ update DDM
4. compute new trace
5. if stored and new traces satisfy convergence criteria, then stop, else go to next iteration

The convergence criterion requires that, during the refinement, the signal traces obtained from the last two event-driven simulations are *matched.* Two successive traces are matched when the node is switching in the same direction, at approximately the same time, and the same number of times. In other words, for the ith iteration the following equations must hold simultaneously

$$\begin{cases} K_n^{(i)} = K_n^{(i+1)} \\ A_k^{(i)} = A_k^{(i+1)}, \ \forall k = 1, \ldots, K_n^{(i)} \\ |t_k^{(i)} - t_k^{(i+1)}| < \triangle t^{(i+1)}. \end{cases} \tag{4.12}$$

The global sub-criterion mandates that $\triangle t^{(i)}$ be less than a maximum tolerance *TOL* for all values of i. In the implementation there also exists a local criterion which requires that a change in delay be relatively near the maximum of all delay changes for the gate at all iterations, or

$$\tau^{(i+1)} < \max\{\frac{\max_{j=1,\ldots,i} \tau^{(j)} - \tau^{\text{nominal}}}{\text{REL\_TOL}}, \frac{\tau^{\text{nominal}}}{\text{ABS\_TOL}}\}, \tag{4.13}$$

where $\tau^{nominal}$ is the nominal gate delay and $\tau^{(i)}$ is the delay computed at iteration i. `REL_TOL` and `ABS_TOL` are user-defined constants. The disvantage of using the local criteria is that error may accumulate when a signal propagates through several gates in the circuit. For all the test results given in the next section, the global criteria was adopted.

In this implementation, parameters R_{VDD}, R_{GND}, L_{VDD}, and L_{GND}, used in modeling the power supply network, were set using the estimates proposed in [65]. Nominally $R_{VDD} + R_{GND}$ was set to 0.25Ω. When the power supply network is extended to a particular macro block inside the circuit, the resistances were increased. Considering that in our experiments we usually had a better ground connection, the values of 1Ω and 0.5Ω were chosen in our tests. L_{VDD} and L_{GND} were set to zero due to their negligible effects.

Parameters `REL_TOL` and `ABS_TOL` were chosen to be 200 and 2000, respectively. These values are justified by the fact that the smallest gate delay is less than 4000 time units used in event-driven simulation, the convergence criteria set by these values are strict in the sense that they nearly reach the limit posed by the time resolution used for event-driven simulation. The value chosen for *TOL* was 3, which corresponds to 0.3ps. In the test, the convergence criteria set by this value effectively requires that upon convergence, the last two signal traces are identical.

5. CASE STUDY

We tested our minimalistic event-driven simulator against Verilog-XL with the `accu_path_delay` option, on the combinational logic circuits from ISCAS85. The resulting signal traces were identical.

Three combinational logic circuits were selected from ISCAS85 for the test: C432 with the smallest gate count, C499 with a medium gate count, and C7552 with the largest gate count. The test was conducted for a total of four input stimuli (pattern #1 through #4). The results are given in Table 4.2 together with several statistics of the circuits under test.

In our tests, convergence was reached in less than 5 iterations for all the convergent cases. The two non-convergent cases reveal two different types of non-convergent behavior. In the first case, C432 under pattern #2, the iteration oscillates between two matching signal traces, but the switching times in the two

Benchmark	# gates	# inputs	# outputs	# level	pattern	# iterations
C432	163	36	7	22	#1	3
C432	163	36	7	22	#2	-
C432	163	36	7	22	#3	3
C432	163	36	7	22	#4	4
C499	600	41	32	25	#1	4
C499	600	41	32	25	#2	4
C7552	2988	207	108	49	#1	4
C7552	2988	207	108	49	#2	-

Table 4.2. Convergence of iterative refinement for different input patterns

traces do not meet the global convergence criteria. In the second case, C7552 under pattern #2, the iteration oscillates between two non-matching signal traces. A major factor affecting the convergence of the iterative refinement process is the intensity of the switching activities in a digital circuit. Large number of simultaneous gate switchings will cause significant changes in gate delays between successive iterations, resulting in significant differences in successive signal traces. To further investigate the convergence properties of the iterative refinement process and improve its convergence, statistics such as the number of simultaneous switching gates and the distribution of gate switchings in time are needed.

An interesting analogy can be drawn between the waveform relaxation method and our iterative refinement process. In waveform relaxation method, the target of iterative refinement is a set of analog waveforms; in our iterative refinement process, the target is the signal trace in a digital circuit under a particular input stimuli.

Figure 4.4 shows the supply noise signature I_{dd} in benchmark C499 when a fast random process is used as stimulus. Note the components at significantly higher frequencies present in the signature. Figure 4.5, on the contrary, depicts the resulting signature for a relatively slow input stimulus applied to benchmark C432. In this case the hazardous signal is generated with a considerable delay, thus its effect may be tame. The impact of the stimuli and their statistical properties is clearly illustrated by these plots. It is therefore imperative that modeling reflect the potential stimulus or collection of stimuli which is likely to be presented to the actual circuit during operation.

In case of a wide variety of stimuli possible, it is advisable to construct a parametric model based on the statistical properties of the signals. Alternatively, one can build a *super-model* based on a collection of models, each constructed with a particular collection of stimuli. All such models can then be

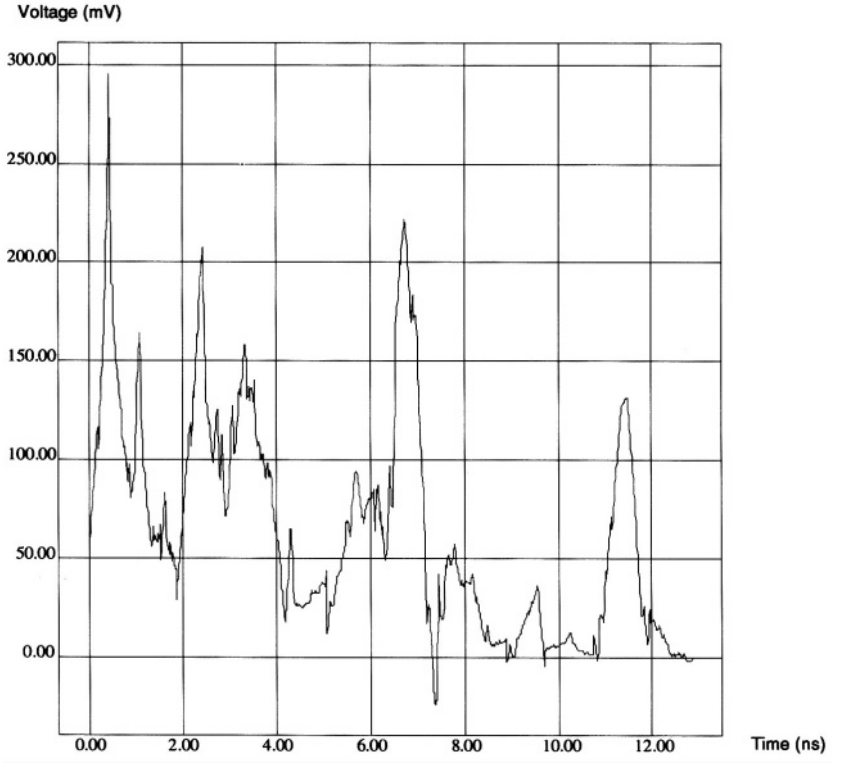


Figure 4.4. Time domain supply noise signature for C499

used during verification through an exhaustive analysis of the effects of each of them individually.

Notes

1 This term can be computed numerically in an efficient manner, during circuit simulation.

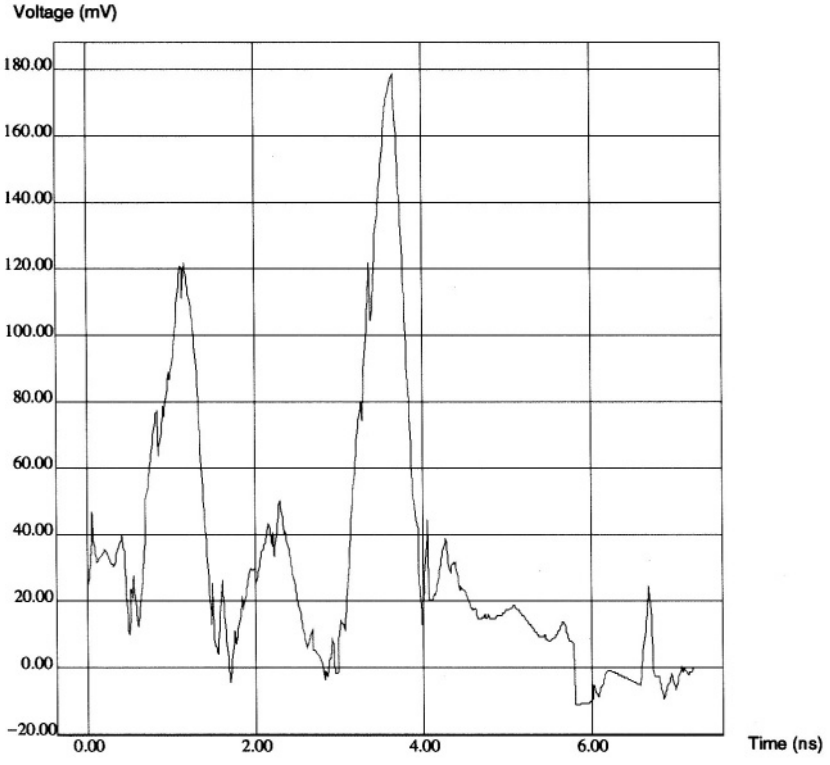


Figure 4.5. Time domain supply noise signature for C432

This page intentionally left blank

Chapter 5

CONSTRAINT GENERATION

Constraint generation has proven useful both during physical assembly and performance verification. This is mainly due to the existence of a constraint on a precise parasitic component can be used to guide a placer or a router to when a decision regarding a course of action is needed. Moreover, constraint violations can be easily detected and used to spot the causes of a system-level specification violation.

In this chapter we shall discuss the methods used in our work for the computation of constraints associated with substrate parasitics. Moreover, we shall show how our Green's function based analysis tools can be used for a fast and accurate evaluation of substrate parasitics and constraint enforcement during optimization.

1. LOCAL NOISE GENERATORS

Constraint generation in a strict sense requires that parasitics be entities associated with one or more physical structures of the layout being made. In the case of switching noise the physical location and transmissions paths through the substrate may not be known before the general floorplan is performed on the chip. In order to enable the constraint generation process, one needs to introduce artifacts for the characterization of the effects of substrate noise. To accomplish this task, we introduce the concept of *local noise generators*.

A local noise generator is defined as a model $G_n(t, \mathbf{\Pi})$ of all substrate noise present at node n, at time t. Vector $\mathbf{\Pi}$ represents all parameters relevant to characterize generator G_n at its nominal value. Due to the different nature of these parameters, $\mathbf{\Pi}$ can be split into basic components $\mathbf{\Pi} = \left[\mathbf{W}^T\ \mathbf{G}^T\ T\ V_0\right]^T$. $\mathbf{W}$ represents process-dependent and $\mathbf{G}$ layout-related parameters, T is the temperature and V_0 the local substrate poten-

tial. Variation vector $\triangle\mathbf{\Pi} = \left[\triangle\mathbf{W}^T \ \triangle\mathbf{G}^T \ \triangle T \ \triangle V_0\right]^T$ represents all variations of these parameters from nominal. Consider now an arbitrary performance vector $\mathbf{K}$, which is a compact representation of all performance measures characterizing a given circuit. For simplicity but without loss of generality, let us focus on one such performance measure K_i, which is the ith entry of $\mathbf{K}$. A typical performance measure could be unity gain bandwidth or offset in an amplifier or delay in a digital gate. Define sensitivity vector $\mathbf{S}_{\mathbf{\Pi}}^{K_i}$, expressed as

$$\mathbf{S}_{\mathbf{\Pi}}^{K_i} = \begin{bmatrix} \mathbf{S}_{\mathbf{W}}^{K_i} \\ \mathbf{S}_{\mathbf{G}}^{K_i} \\ \mathbf{S}_{T}^{K_i} \\ \mathbf{S}_{V_0}^{K_i} \end{bmatrix} = \begin{bmatrix} \left(\frac{\partial K_i}{\partial w_1} \ \frac{\partial K_i}{\partial w_2} \ \cdots \ \frac{\partial K_i}{\partial w_{|\mathbf{W}|}}\right)^T \\ \left(\frac{\partial K_i}{\partial g_1} \ \frac{\partial K_i}{\partial g_2} \ \cdots \ \frac{\partial K_i}{\partial g_{|\mathbf{G}|}}\right)^T \\ \frac{\partial K_i}{\partial T} \\ \frac{\partial K_i}{\partial V_0} \end{bmatrix}, \tag{5.1}$$

where $(w_1 \ \ldots \ w_{|\mathbf{W}|})^T$ and $(g_1 \ \ldots \ g_{|\mathbf{G}|})^T$ are the entries of vectors $\mathbf{W}$ and $\mathbf{G}$ respectively. The performance variation $\triangle K_i$ can be expressed by

$$\triangle K_i \ = \ (\mathbf{S}_{\mathbf{\Pi}}^{K_i})^T \cdot \triangle\mathbf{\Pi}. \tag{5.2}$$

The constraint generation process consists of finding a bound for each of the entries of $\triangle\mathbf{\Pi}$ so as to meet a constraint on the maximum degradation of the performance measure, represented by the following inequality

$$\triangle K_i \leq \overline{\triangle K_i}, \tag{5.3}$$

where $\overline{\triangle K_i}$ is such constraint. By substituting (5.2) into (5.3), one obtains

$$(\mathbf{S}_{\mathbf{\Pi}}^{K_i})^T \cdot \triangle\mathbf{\Pi} \leq \overline{\triangle K_i}, \tag{5.4}$$

which must be solved with respect to $\triangle\mathbf{\Pi}$. However, this problem has an infinite number of solutions. A possible solution is that of translating the above problem into the following constrained optimization problem

maximize	$\mathbf{Flexibility}(\mathbf{\Pi}^{(\mathbf{bound})})$
such that	$(\mathbf{S}_{\mathbf{\Pi}}^{K_i})^T \cdot \triangle\mathbf{\Pi} \leq \overline{\triangle K_i}$, $\mathbf{\Pi}^{(\mathbf{min})} \leq \mathbf{\Pi}^{(\mathbf{bound})} \leq \mathbf{\Pi}^{(\mathbf{max})}$,

where function `Flexibility(.)` quantifies how easily $\mathbf{\Pi}^{(\mathbf{bound})}$ can be implemented. $\mathbf{\Pi}^{(\mathbf{min})}$ and $\mathbf{\Pi}^{(\mathbf{max})}$ are physical bounds on the parameters.

Due to the mechanism of noise modeling obtained using local generators, constraints on noise parameters can be derived independently of a particular IC

process. Hence the constraint generation needs to be repeated only once for a given circuit. During layout synthesis, process-dependent substrate extraction methods are used to enforce bounds. From a theoretical point of view each generator could be supplied by a different signal waveform. However, since the size of the analog section of a mixed-signal circuit is small compared to the distance to the noise sources, it is assumed that all the substrate nodes are reached by an identical waveform with different phases. Suppose there exist M nodes each of them connected to a noise generator $G_m(t-\tau_m, \mathbf{\Pi}_m)$ *with* $m = 1, .., M$ where τ_m is the propagation delay of the waveform from one node to the other.

2. WORST-CASE SENSITIVITIES

Due to the potential non-linearity of some performance measures on parasitics, an additive linearization around a nominal value could inaccurately model the effects of substrate. For simplicity of notation but without any loss of generality, consider again performance measure K_i. The problem can be effectively addressed by deriving a worst-case sensitivity of K_i with respect to all parameters for which a linear behavior is observed. Let us split vector $\mathbf{\Pi}$ in two sub-vectors: $\mathbf{\Pi}'$, $\mathbf{\Pi}''$, which contain all the parasitics that show a linear and a non-linear behavior respectively. $\mathbf{\Pi}'$ is defined as the vector of all parameters such that

$$\left|(\mathbf{S}_{\mathbf{\Pi}'}^{K_i})^T \cdot \triangle\mathbf{\Pi}' - \triangle K_i\right| < \epsilon, \tag{5.5}$$

with $0 < \triangle\mathbf{\Pi}' < \delta$, for some $\epsilon, \delta > 0$.

The problem of finding a worst-case sensitivity $\overline{\mathbf{S}}_{\mathbf{\Pi}'}^{K_i}$ is equivalent to solving the maximization problem

maximize	$\mathbf{S}_{\mathbf{\Pi}'}^{K_i}$
such that	$\mathbf{\Pi}'' \in I$

where I is the feasibility interval of $\mathbf{\Pi}''$. Hence, the total linearized worst-case variation of K_i, due to node n, can be derived as

$$\triangle K_i|_n = (\overline{\mathbf{S}}_{\mathbf{\Pi}'_n}^{K_i})^T \cdot \triangle\mathbf{\Pi}'_n. \tag{5.6}$$

Using the same formalism of (5.2) we define the matrices:

$$\left[\overline{\mathbf{S}}_{\mathbf{\Pi}'}^{K_i}\right] = \begin{pmatrix} (\overline{\mathbf{S}}_{\mathbf{\Pi}'_0}^{K_i})^T \\ \vdots \\ (\overline{\mathbf{S}}_{\mathbf{\Pi}'_{N-1}}^{K_i})^T \end{pmatrix} \quad and \quad [\triangle\mathbf{\Pi}'] = \begin{pmatrix} (\triangle\mathbf{\Pi}'_0)^T \\ \vdots \\ (\triangle\mathbf{\Pi}'_{N-1})^T \end{pmatrix}. \tag{5.7}$$

The total degradation of performance measure K_i due to all the nodes can be expressed as

$$\triangle K_i = \mathbf{trace}\left(\left[\overline{\mathbf{S}}_{\mathbf{\Pi}'}^{K_i}\right] \cdot [\triangle \mathbf{\Pi}']\right). \tag{5.8}$$

The introduction of the worst-case sensitivity matrix $\left[\overline{\mathbf{S}}_{\mathbf{\Pi}'}^{K_i}\right]$ allows the reduction of the parameter space and the inclusion of non-linear behavior in a certain range of performance.

The local noise generator approach has several advantages. The effect of the substrate noise can be evaluated locally without taking into consideration the substrate configuration or the actual position of the devices which are injecting noise into it: the local noise generator can be seen as an antenna. Moreover, a standard sensitivity analysis can be used to analyze the effects of noise on performance. Furthermore, constraints on the various parameters of noise can be generated and accounted for during synthesis. Finally, once the substrate has been extracted the local substrate potential V_0 can be related to the noise generator substrate potential through an *isolation factor* α. From this value information on the placement of the analog part with respect to the digital part in the mixed-signal chip can be derived and eventually the necessity of guard rings can be highlighted.

3. CASE STUDY

In this study, a PLL is designed and fabricated using a top-down process which consists of three phases. In the first phase, behavioral models are created for all the blocks of the PLL. Based on these models, the PLL specifications are mapped onto a set of constraints on all the blocks, which are in turn used to design and size each block independently. As a byproduct of the block sizing, constraints on all physical parasitics are computed. These constraints are used in the second phase of the design, i.e. the physical design, where *ad hoc* tools enforce them. The final phase of the design is the parasitic extraction for a final verification of the specifications. Here, we will focus on the processes of constraint generation and enforcement.

3.1 PLL ARCHITECTURE

Let us consider the PLL shown in Figure 5.1. The architecture used as a basis for the PLL design consists of a crystal reference frequency generator (F_{ref}), a PFD, a charge pump, a second order RC type loop filter and a VCO. The PFD, charge-pump, and loop filter are similar to the ones used in [66]. The basic cell of the VCO has been derived from the one presented in [67]. Three frequency dividers are being used, represented by their dividing ratios n, m, k in Figure 5.1, to allow the synthesis of a number of frequencies. The output frequency is expressed in terms of the various dividing ratios as $F_{out} = F_{ref} \cdot n/mk$.

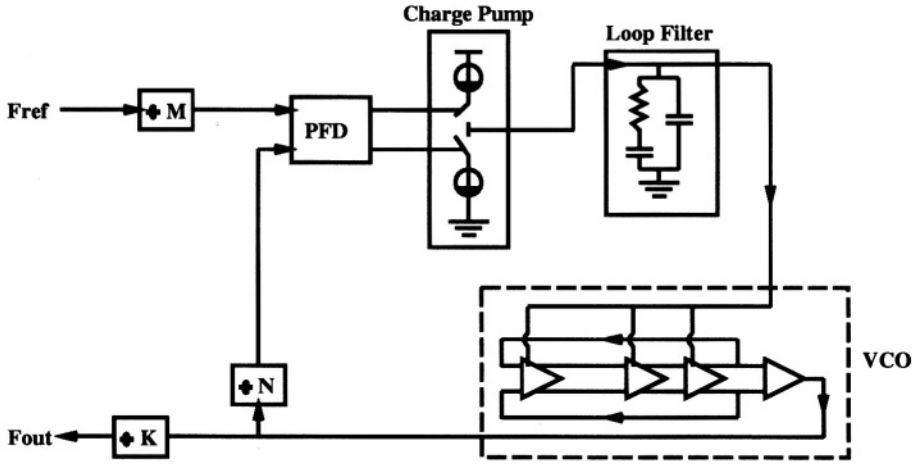


Figure 5.1. PLL Block Diagram

Each of the PLL blocks is characterized by a set of behavioral parameters. The PFD is characterized by a state transition table and a delay t_{pfd}, the dividers by a delay t_{ddiv}, and a divide ratio. In the PFD and the dividers extra delays induced by parasitics do not affect the PLL performance. The loop filter is determined by its component values R, C_1, C_2, the charge pump by its bias current I_{cp} and output resistance R_{out}. Let us focus now on the various components of the PLL to characterize the noise in the system.

The VCO is entirely determined by the frequency-to-voltage characteristic, which in turn can be significantly affected by process, temperature variations and layout parasitics. The PLL is optimized in such a way that performance degradation due to these non-idealities is kept within pre-determined tolerances. A further degradation due to layout parasitics can cause a system failure if not taken into account at the circuit optimization and layout level. Jitter performance of the system is mostly affected by the jitter of the VCO. Sources for VCO jitter are thermal noise and coupling of digital noise from the supply and the substrate. We will focus here on the effect of the substrate coupling to the peak-to-peak jitter performance. The VCO model used in this work is characterized by the following equation

$$F_{out} = F_0 + K_0 \cdot \triangle V, \tag{5.9}$$

where F_0 is the VCO central frequency of operation, K_0 the frequency-to-voltage gain, $\triangle V$ the deviation of the applied voltage in the control node from the nominal. Performance constraints for the PLL are:

1. stable frequency range of operation: $F_{min} \leq F_{out} \leq F_{max}$;

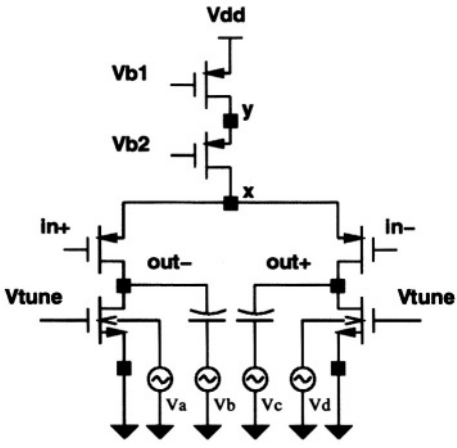


Figure 5.2. Example of noise injection from the substrate

2. rms timing jitter of the generated clock: $J_{rms} \leq \overline{J}_{rms}$;
3. peak-to-peak timing jitter: $J_{pp} \leq \overline{J}_{pp}$.

Using behavioral simulation, bounds are provided for the maximum allowed performance degradation [68].

Consider the VCO section of the circuit. Let $\mathbf{K} = [K_1, K_2]$ where K_1 is the oscillation period of the circuit and K_2 is the peak-to-peak jitter. To evaluate the sensitivity of $\mathbf{K}$ with respect to all parameters of interest, the following approximation has been used

$$\frac{\partial K_i}{\partial \pi_j} \simeq \frac{\triangle K_i}{\triangle \pi_j} \quad \text{for } \triangle \pi_j \rightarrow 0. \tag{5.10}$$

All layout parasitics affect the performance period while the substrate noise, in this architecture, affects mainly the performance jitter. The sensitivities of period K_1 with respect to all circuit parasitics are calculated using the method of finite differences by augmentation of the schematic and a perturbation scheme.

Local noise generators are placed at each node of the schematic so as to simulate all possible substrate coupling. Figure 5.2 shows four of the critical local noise generators. The generators, coupled through bulk contacts, simulate threshold voltage modulation, while the ones capacitively coupled represent direct signal coupling noise.

Let $G_n(t, \mathbf{\Pi})$ be the local noise generator associated with node n. Simulations showed that the worst case occurs when an impulsive function is used. Assume that we have M substrate nodes, then there exist M delayed generators $G_{n_i}(t - \tau_i, \mathbf{\Pi})$ *for* $i = 1, 2, \cdots, M$. Using the concept of worst-case sensitivities, one finds that there exists a vector $\tau = [\tau_1, \tau_2, \cdots, \tau_M]$ such that

Performance	Nominal	Max. Variation
K_0	$40MHz/V$	$4MHz/V$
F_0	$100MHz/V$	$10MHz$
J_{rms}	0	0.05%
J_{pp}	0	1%

Table 5.1. Performance constraints obtained by the behavioral optimization of the VCO

both $\triangle K_{1,2}$ are maximum. In the VCO basic cell the worst case corresponds to having all the generators switching synchronously with the cell itself. Under these circumstances, the sensitivity with respect to local substrate potential V_0 is almost constant to about 2/3 within the interval of interest (0-300mV). Considering the whole ring oscillator, the maximum degradation of the peak-to-peak jitter occurs when the delay between local noise generators is equal to the delay of the basic inverter cell, as shown in Figure 5.3.

The schematic of the loop filter of Figure 5.1 is depicted in Figure 5.4(a). Figure 5.4(b) shows the model with all the parasitic coupling to substrate which have been taken into consideration to evaluate the effect of the local noise generator. Using the same approach, the sensitivity of the peak-to-peak jitter with respect to the V_0 is evaluated, thus finding a similar linear dependency as before. The value of sensitivity results to be much less than the one of the VCO, hence no constraints need to be generated for the Loop Filter.

3.2 CONSTRAINT GENERATION

The robustness of the design can be significantly improved by use of sensitivity analysis applied to circuit performance. In what follows we show how sensitivities can be used in the optimization of VCO circuit level parameters.

Using behavioral simulation, constraints are computed on the performance of the VCO which ensure satisfaction of the PLL specifications. Table 5.1 lists the results. Note that performance deviations do not take into account process gradients.

The basic cell of the ring oscillator is shown in Figure 5.5. The optimization problem for the VCO can be written as

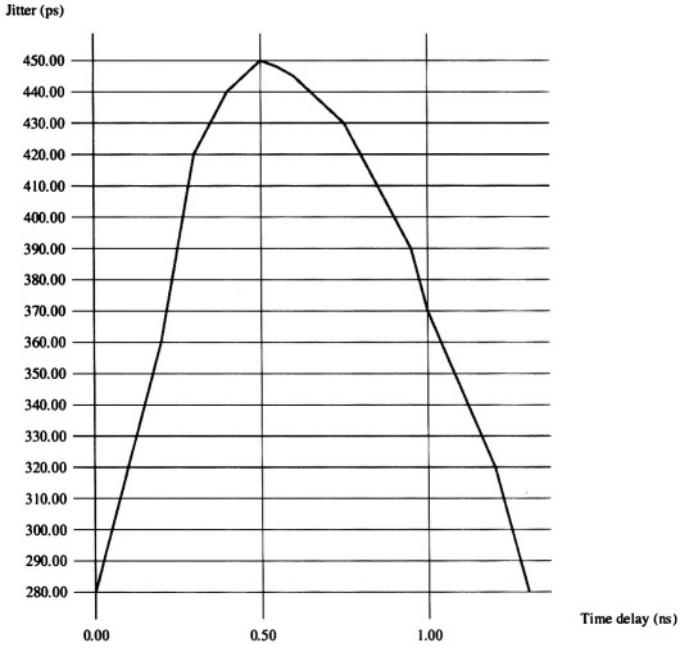


Figure 5.3. Peak-to-peak jitter as function of the time delay among the local noise generators associated with every cell. The maximum substrate voltage is kept constant to $V_0 = 300mV$

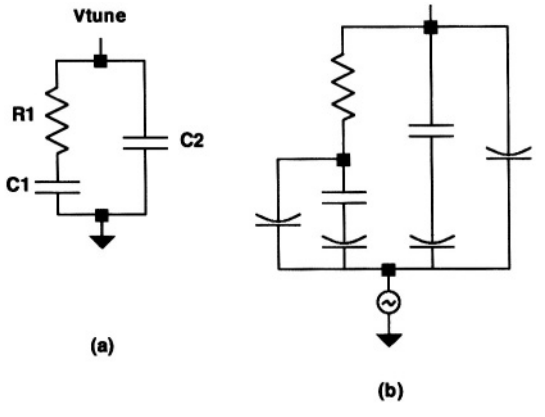


Figure 5.4. Loop filter used in the PLL: (a) schematic, (b) model for substrate coupling analysis.

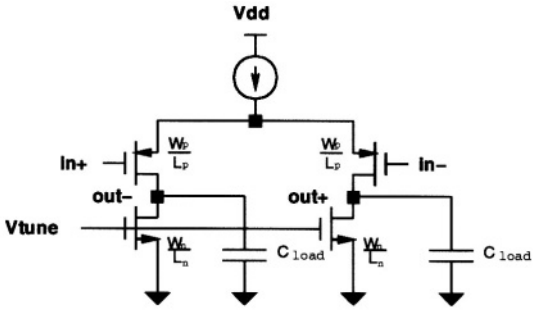


Figure 5.5. VCO basic cell

$$
\begin{array}{ll}
\textit{minimize} & \text{Total Power}(W_p, L_p, W_n, L_n) \\
\textit{such that} & F_{min_0} - \triangle F_{min_0} \leq F_{min} \leq F_{min_0} + \triangle F_{min_0} \\
& F_{max_0} - \triangle F_{max_0} \leq F_{max} \leq F_{max_0} + \triangle F_{max_0} \\
& J_{rms} \leq 0.05\% \\
& J_{pp} \leq 1\%
\end{array}
$$

where W_p, L_p, W_n, L_n are the transistor sizes of the basic delay cell shown in Figure 5.5. Function `Total Power(.)` estimates the power dissipation of the circuit. This optimization minimizes the current through the VCO and hence the size of the input transistors of the differential pair, which determines the cell delay, thus keeping the ratio I/C_{load} constant. Reducing transistor sizes has the drawback of increasing the sensitivity with respect to parasitics. In fact C_{load} is given by the gate capacitance of the next cell input transistor and by parasitics. Hence, if the layout parasitics are not well controlled, we may have a significant degradation of circuit performance. This can be avoided if the sensitivity information is used during the optimization process.

Using the notation introduced in this chapter, the new optimization problem becomes

$$
\begin{array}{ll}
\textit{minimize} & \text{Total Power}(W_p, L_p, W_n, L_n) \\
\textit{such that} & F_{min_0} - \triangle F_{min_0} \leq F_{min} \leq F_{min_0} + \triangle F_{min_0} \\
& F_{max_0} - \triangle F_{max_0} \leq F_{max} \leq F_{max_0} + \triangle F_{max_0} \\
& J_{rms} \leq 0.05\% \\
& J_{pp} \leq 1\% \\
& \left[\mathbf{S}_{\mathbf{\Pi}}^{K_i}\right]^T \cdot \triangle \mathbf{\Pi}^{(bound)} \leq \overline{\triangle K_i} \quad \forall i = 1, \cdots, N_k
\end{array}
$$

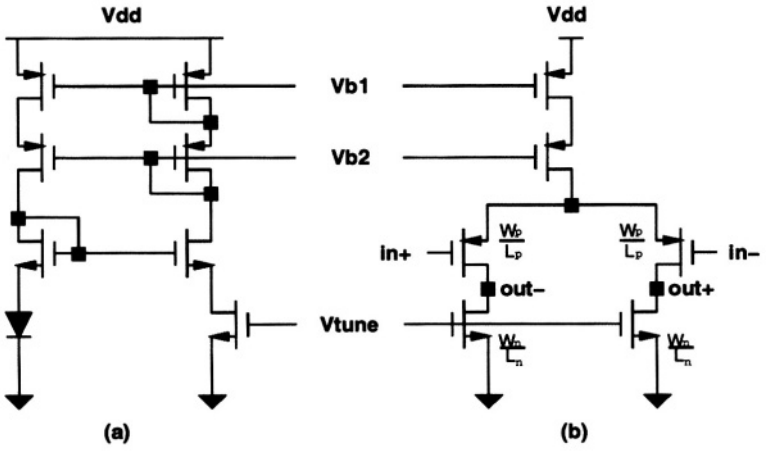


Figure 5.6. Basic VCO cells: (a) bias cell, (b) fully differential inverter

where $\triangle\Pi^{(bound)}$ represents the vector of the bounds on all critical parameters which, if enforced, ensure that performance specifications be satisfied.

3.3 CONSTRAINT ENFORCEMENT

The VCO is synthesized using module generator VCOGEN in CMOS technology [61]. The floorplan of the VCO was chosen to account for various considerations. First, critical capacitive and resistive parasitics could be easily minimized by abutting all delay elements of the ring oscillator. Second, this floorplan allows full scalability, both in power and frequency. Third, due to multiple folding of the ring oscillator's structure, technological mismatches can be contained. In addition, by keeping the cell's aspect-ratio low, the systematic component of the mismatch relevant to the circuit can be reduced to one dimension as suggested in [69]. The fully differential implementation of the VCO can acquire a better isolation to substrate noise coupling capacitively through interconnections with respect to a single ended one. Most of the coupled noise turns into common mode noise.

The delay element and the distributed bias of the VCO are depicted in Figure 5.6. The layout of the delay element was designed using a mirror symmetry. This configuration has two main benefits:

1. a minimization of rising and falling time mismatch;
2. the effect of thermal and substrate noise is minimized due to the balance of the two branches.

Let us consider the problem of enforcing a couple of constraints operating on the parasitic resistance and capacitance due to the same interconnection. These

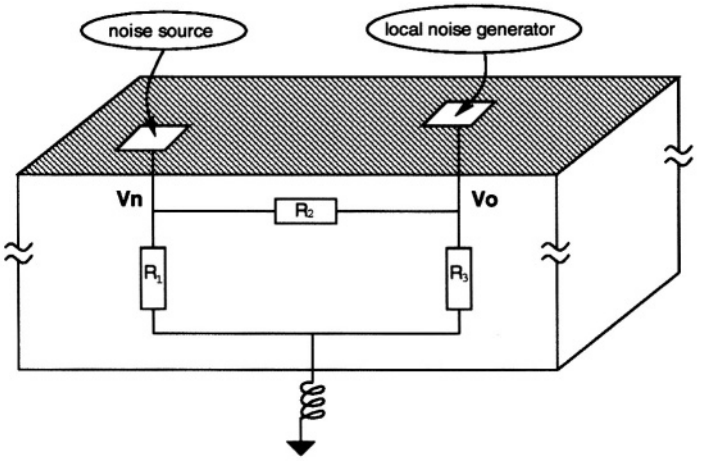


Figure 5.7. Substrate model implemented by SUBRES

two parasitics depend simultaneously on wire dimensions (Width and Length) $R = \rho \frac{L}{W}$ and $C = C_0\, W\, L$. The problem can be solved using the following algorithm

set $W = W_{min}$ **and** $L = L_{min} \implies C = C_0\, W_{min}\, L_{min}$;
do
evaluate $R = \rho \frac{L}{W}$;
if $R < R_{max}$ **then** exit the cycle;
else $W = W + \triangle W$;
while $C < C_{max}$;
if $C > C_{max}$ **or** $R > R_{max}$ **then** "infeasible" $\implies$ **stop**;
else "constraint enforced";

where $\triangle W$ is the minimum increment (λ) allowed by the process design rules.

The enforcement of substrate-related constraints is performed in the following way. First a technology-aware model of the substrate for the entire chip is built. Then, given that the signal injected by each noise source is known, the isolation factor α required to meet the derived constraint on the V_0 is calculated. Since SUBRES is based on the substrate model depicted in Figure 5.7, α is evaluated as the ratio between R_3 and R_2. Hence, the minimum required distance from the noise source can be derived and used for fully automated or hand layout design.

The relative position of the VCO with respect to the dividers, the main source of substrate noise, has been enforced using substrate resistance estimations performed by SUBRES.

Performance	Ideal Value	max degradation
oscillation period	6-20ns	10 %
peak-to-peak jitter	0	100ps

Table 5.2. Design specifications

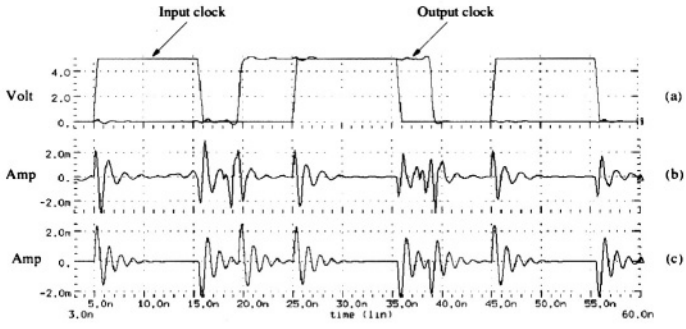


Figure 5.8. (a) Divider's input/output waveforms; (b) Noise injected in the substrate obtained by SPICE simulations; (c) Model derived for the equivalent current source

Table 5.2 illustrates the maximum performance degradations to be enforced. The layout of the PLL has been partitioned in eight clusters. A contact of size equal to the total injecting area (i.e. devices, interconnect, etc.) has been assigned to each cluster. Then the extraction of the substrate resistances connecting these contacts to the one associated to the VCO have been performed with SUBRES. Finally a SPICE simulation has been carried out using the behavioral model fitted from the waveforms of Figure 5.8(c).

The result of this simulation, shown in Figure 5.9, is the estimate of the noise present in the VCO substrate. It should be noted that the specific on the peak value is given in the worst case of having the switching noise reaching the VCO in the same exact moment of the transition of one of the delay elements. In this case, however, the divider is triggered by the VCO. Thus, the digital switching activity due to dividers always occurs with a certain delay with respect to the VCO trip point, hence giving a bigger safety margin.

Phase noise simulation imposes tight restrictions on the numerical simulation algorithms to be used. For accurate results, numerical noise created by the algorithm should be negligible when compared with the phase to be computed. In this particular example the numerical noise on the phase turned out to be about $3\ ps$. V_0 has been increased to have a value of numerical noise at most 10% of the jitter. The sensitivity analysis of interconnect in the

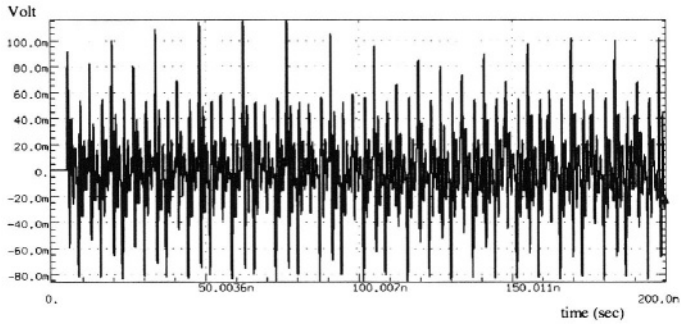


Figure 5.9. Substrate noise sensed at the contact associated with the VCO

Parasitic	Constraint	Extracted value
C_{out+}	12.34fF	11.60fF
C_{out-}	12.34fF	11.60fF
C_x	15.84fF	1.12fF
$V_0\|_{VCO}$	110mV	108mV

Table 5.3. Constraints obtained by the sensitivity analysis

Variable	size with parasitics	size without parasitics
W_p	36μm	25μm
L_p	1μm	1μm
W_n	2.6μm	2.6μm
L_n	4μm	4.6μm

Table 5.4. Transistor sizes obtained by circuit optimization.

VCO has been performed introducing 16 parasitics per cell. The constraints on critical parasitics are summarized in Table 5.3. Notice how the cell symmetry induces the same constraint on the parasitic capacitances of the two differential outputs. Table 5.4 shows the transistor sizes obtained from the optimization with and without taking into account parasitics analysis. It can be seen how in the first case the dimensions are bigger to reduce the sensitivity. The layout of the automatically generated VCO is shown in Figure 5.10. Finally, Table 5.5 shows the CPU times for each of the phases of the assembly.

Operation	CPU time (sec)
circuit optimization	1028 †
substrate sensitivity	2545 †
interconnect sensitivity	3256 †
constraint generation	115
layout generation	43

Table 5.5. Results obtained on a DECstation 5000/125 and on a DECstation alpha (†).

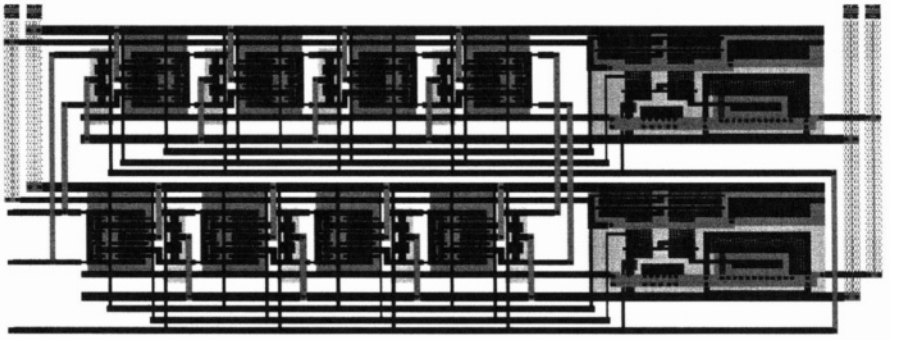


Figure 5.10. Layout of the VCO generated by VCOGEN

Chapter 6

OPTIMIZATION TECHNIQUES

Traditionally, the analysis of substrate noise has been performed after the completion of physical design as a verification step. Experience has shown the extreme time complexity required to accurately model substrate and to estimate performance degradations due to switching noise. In many design problems however, a *dynamic* substrate noise analysis would be preferable, since it could drive the design towards solutions more resilient to substrate noise.

Optimization phases typically performed in physical design include floorplanning, placement, routing and compaction. The parasitics usually considered in such optimization phases are generally localized to specific areas of the workspace. Hence, possibly global effects of parasitics may be ignored. The algorithms used in floorplanning and placement are based on incremental improvement techniques, consequently it is possible to derive compact and efficient ways of evaluating the degradation of performance due to parasitics while the optimization unfolds.

On the contrary, due to its "global" effects felt everywhere in the chip, substrate noise cannot be easily translated into a compact analytical model accounting for the entire substrate area. Hence, even if a small incremental modification is performed on the chip, the whole substrate analysis needs be reevaluated. Approaches found in the literature which make use of finite difference methods, do not have provisions to evaluate small incremental changes of the substrate geometry. Such changes often occur in complex optimization loops. The solution generally adopted is to solve the Laplace equation, accurately or approximately, over the entire workspace at each iteration [7, 8].

Hereafter, we describe several approaches based on integral equation techniques which make use of the locality of incremental changes. The key of such techniques is a fast computation of *variations* and *trends* of substrate transport given changes in its physical structure.

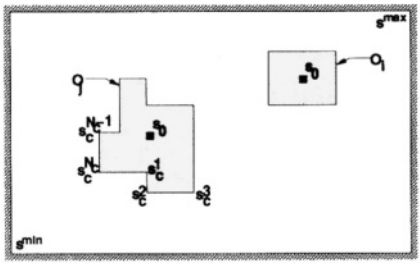


Figure 6.1. Object definition

1. SUBSTRATE-AWARE PLACEMENT

The placement problem is referred to as the task of finding a physical location for a number of layout objects in order to satisfy a number of *constraints* and to minimize a *cost*. Let O_j, $j = 1, \ldots, N_O$ be a layout object, $\mathbf{s_c^{(i)}}(O_j)$, $i = 1, \ldots, N_c(O_j)$ the vertices of its perimeter, and $\mathbf{s_0}(O_j)$ its center[1], as illustrated in Figure 6.1. Let C^A be the set of all absolute constraints on the center or perimeter of object O_j, typically of the form

$$\begin{gathered}\mathbf{s}^{(min)} \leq \mathbf{s_0}(O_j) \leq \mathbf{s}^{(max)}\ ,\ \ \forall j = 1, \ldots, N_O\ ,\ \text{or} \\ \mathbf{s}^{(min)} \leq \mathbf{s_c^{(i)}}(O_j) \leq \mathbf{s}^{(max)}\ ,\ \ \forall i = 1, \ldots, N_c(O_j)\ ,\ \ \forall j = 1, \ldots, N_O\ ,\end{gathered} \tag{6.1}$$

where $\mathbf{s}^{(min)} < \mathbf{s}^{(max)}$ are reference points. These constraints are required to fix the mobility of an object within boundaries determined by considerations on the entire chip or module under construction. Let C^R be the set of all relative constraints, i.e. the constraints relating pairs or groups of objects to each other, typically of the form

$$s_x(O_j) \gtrless s_x(O_i)\ ;\ \ s_y(O_j) \gtrless s_y(O_i)\ , \tag{6.2}$$

where $\mathbf{s}(O_j) = [s_x, s_y]^T$ is an arbitrary point (on the perimeter or in the center) relative to the O_j.

Assume that each object O_j can be represented in terms of a collection of simpler four-sided objects $\overline{O}_j$, called *primitives*. Each object $\overline{O}_j$ is completely specified by the following features: $\mathbf{s_0}$, r_0, ℓ and w. r_0 relates to the orientation of the object, while ℓ and w to its length and width, respectively. Furthermore, assume that a Manhattan style design is adopted for all our layouts. Then, r_0 can assume only the following self-explanatory values: `NO_ROTATE, ROTATE_90, ROTATE_180,ROTATE_270,MIRROR_X,MIRROR_Y, MIRROR_YX.`

For a given circuit, let us define the *placement configuration* S as the set of quadruples $\{(\mathbf{s_0}(\overline{O}_j), r_0(\overline{O}_j), \ell(\overline{O}_j), w(\overline{O}_j)) \mid \forall j = 1, \ldots, N_O\}$ which determine the location and orientation of each object in the layout. When S

satisfies the constraints of C^A and C^R, it is called *legal configuration.* Let us define the set of all configurations as $\{S\}$.

Finally, let us define a function $f(S)$ as the cost associated with configuration S. The placement problem in its most general formulation consists of finding a legal configuration S associated with the minimum cost f^*. The problem can be expressed in terms of the following optimization

minimize	f(S)
such that	C^A and C^R

In this formulation the placement problem is *N P*-hard, however over the years a number of heuristics have been developed to find sub-optimal solutions in lesser CPU time.

The best known optimization oriented heuristics are generally referred to as: *branch-and-bound search and partitioning-based, quadratic optimization-based,* and *iterative improvement techniques.* Iterative improvement techniques allow to best exploit progressive substrate extraction methods. The techniques proposed hereafter are designed for very fast estimation of *variations* and *trends* within computationally expensive algorithms.

1.1 INCREMENTAL SUBSTRATE EVALUATION

The first technique exploits the fact that small adjustments in the configuration of layout elements results in a small change in the coefficient of potential matrix $\mathbf{P}$. Let $\mathbf{P}'$ be the potential matrix associated with the new configuration. Note that in $\mathbf{P}'$ only row r and column c will differ from $\mathbf{P}$. Let $\delta\mathbf{P}_{r.}$ be the rth row and $\delta\mathbf{P}_{.c}$ the cth column of $\mathbf{P}'$, then $\mathbf{P}' = \mathbf{P} + \delta\mathbf{P}_{.c} + \delta\mathbf{P}_{r.}$. For simplicity consider only the modification due to $\delta\mathbf{P}_{r.}$. Using the Sherman-Morrison formula, $\mathbf{P}'^{-1}$ can be computed directly as

$$\mathbf{c}' = \mathbf{P}'^{-1} = \mathbf{c} + \delta\mathbf{c}, \text{ with } \delta\mathbf{c} = -\frac{\mathbf{c}_{.r}(\mathbf{c}\ \delta\mathbf{P}_{r.})}{1 + \delta\mathbf{P}_{r.}\ \mathbf{c}_{.r}}, \tag{6.3}$$

where $\mathbf{c}_{.r}$ is the rth column of $\mathbf{c}$. The computation of the entire resistive network is dominated by the Sherman-Morrison update, completed in $O(N^2)$ time for each contact partition being moved.

The second technique, known as *Gradient Based Method,* is based on the concept of *sensitivity to relocation.* Suppose that a contact or a collection of contacts z is to be relocated on the substrate surface from location $\mathbf{x}_0$ to $\mathbf{x}_k$ going through intermediate locations $\mathbf{x}_1, \ldots, \mathbf{x}_{k-1}$ (see Figure 6.2). One can

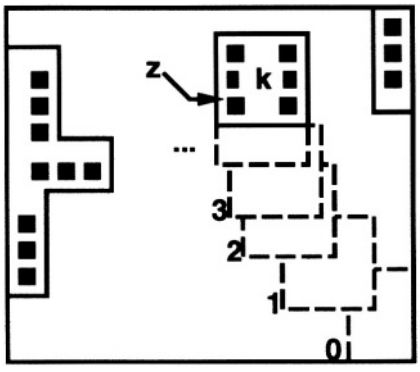


Figure 6.2. Sensitivity of resistive macro-model from transformation of a component and its contacts

easily show that

$$[\mathbf{c}]_k = [\mathbf{c}]_0 + \sum_{n=1}^{k} [\delta\mathbf{c}]_n \,, \tag{6.4}$$

where $[\mathbf{c}]_0$ is the coefficient of induction matrix associated with location $\mathbf{x}_0$, and $[\delta\mathbf{c}]_{n+1} = [\mathbf{c}]_{n+1} - [\mathbf{c}]_n$ is the $(n+1)$th update of $\mathbf{c}$. The updates $[\delta\mathbf{c}]_{n+1}$ can be computed using the Sherman-Morrison formula in $O(N^2)$ time.

To further speed-up the computation one can exploit the "gradient" information of resistive and conductive networks $\mathbf{R}$ and $\mathbf{Y}$, contained in $[\delta\mathbf{c}]_1$. Assume that a single contact z is relocated in direction $\mathbf{v}$ by an amount $|\mathbf{v}| \to 0$. Let us define the vector $\nabla_{\mathbf{v}}\mathbf{Y}$ to be

$$\nabla_{\mathbf{v}}\mathbf{Y} = [\mathbf{A}\ \mathbf{B}]^T, \tag{6.5}$$

where $\mathbf{A} = \partial\mathbf{Y}/\partial v_x$, $\mathbf{B} = \partial\mathbf{Y}/\partial v_y$, and $\mathbf{v} = [v_x v_y]^T$. The components of matrix $\mathbf{A}$ are defined as $A_{ij} = \partial\mathbf{Y}_{ij}/\partial\mathbf{v}_x$, those of $\mathbf{B}$ as $B_{ij} = \partial\mathbf{Y}_{ij}/\partial\mathbf{v}_y$. Recall that, since $N_c = N$, Y_{ij} is defined as the mutual conductance between contact partitions i and j for a given substrate configuration and that Y_{ii} is the ground conductance of i. The minimum step size in x- and y-direction corresponds to a unit of the grid of the DCT. Hence , matrix $\partial\mathbf{Y}/\partial\mathbf{v}_\mathbf{x}$ can be approximated by first computing differences $\delta p_{i,j\pm1}$ and $\delta p_{i\pm1,j}$ using Equation (6.6).

$$\delta p_{i,j\pm1} = p_{i,j\pm1} - p_{i,j}, \text{ and } \delta p_{i\pm1,j} = p_{i\pm1,j} - p_{i,j}. \tag{6.6}$$

Then, each component $\partial Y_{ij}/\partial v_x$ is calculated by replacing term c_{ij} with $\delta c_{i,j+1}$ in Equations (3.7), (3.8), (3.11) and (3.12). Notice that term $\delta c_{i,j+1}$ is derived directly from matrix $\mathbf{c}$ and $\delta p_{i,j+1}$ using the Sherman-Morrison formula. Moreover, the direct replacement of c_{ij} in the equations is legitimated

by the fact that all manipulations are linear. The same method is used to derive $\partial \mathbf{Y}/\partial v_y$. The time complexity of the operation is $O(N^2)$ since the Sherman-Morrison formula needs be repeated for all the contacts or partitions involved in the move.

Let us assume that $\partial \mathbf{Y}/\partial v_x$ and $\partial \mathbf{Y}/\partial v_y$ have been computed at the 0th step of our incremental algorithm. Call $[\partial \mathbf{Y}/\partial v_x]_0$ and $[\partial \mathbf{Y}/\partial v_y]_0$ these matrices. Assuming that the moving partition, contact or collection of contacts remains *close enough* to its position at step 0, then the conductance matrix at steps $1 \leq n \leq k$ can be approximated as

$$[\mathbf{Y}]_n \approx [\mathbf{Y}]_0 + \left[\frac{\partial \mathbf{Y}}{\partial v_x}\right]_0 \triangle v_x + \left[\frac{\partial \mathbf{Y}}{\partial v_y}\right]_0 \triangle v_y$$

$$= [\mathbf{Y}]_0 + \left[\nabla_{\mathbf{v}} \mathbf{Y}^{\mathbf{T}}\right]_0 \triangle \mathbf{v}, \tag{6.7}$$

where $\triangle \mathbf{v} = [\triangle v_x, \triangle v_y]^T$ is the vector representing the move of contact or partition z from step 0 to n.

The Green's function and its DCT are well-behaved functions everywhere in the workspace [15]. Hence, necessarily terms $\delta p_{i,j\pm1} < \infty$ and $\delta p_{i\pm1,j} < \infty$. No "high-frequency" components are present in the function, making it an ideal candidate for a highly accurate use of a gradient based method. In fact, in our experiments the method has shown a 1% accuracy when the move occurred in the vicinity (less than five steps away) of the position at step 0, while a 10% accuracy was reached when the move was up to one tenth of the chip size.

1.2 MODIFIED PLACEMENT ALGORITHM

A substrate-aware placement methodology has been implemented in a framework based on simulated annealing (SA) with analog constraints, called PUPPY-A[70, 71]. The annealing, fully characterized by search space, cost function, move-set and cooling schedule, is described in detail in [44, Chapter 4]. Improvements on the performance degradation due to substrate-induced switching noise can be achieved by placing noise injecting and noise sensitive modules at a certain distance or by creating special structures, such as low-resistivity guard-rings, around noise injectors as described in Chapter 8. The first provision is implemented in the placer using the conventional SA move-set. The second issue is generally solved by extending the search space, allowing the annealing to choose from a number of alternative implementations for a module, including one with a guard-ring implemented around it. In this work we restrict our attention to the first option, where our Green's function based substrate analysis method is used for the evaluation of the substrate at each annealing step.

In order for a placer to be effective in preventing violations to performance specifications, the following features must be implemented in the tool. First,

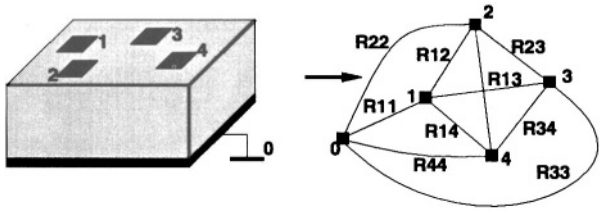


Figure 6.3. Mapping substrate onto fully connected graph $G_S(V, E)$

a model for each noise injecting module must exist. The model should characterize the *waveform* and the *spatial location* where the noise is injected as precisely as possible. Second, a compact model of substrate transport should be available and efficient substrate current evaluation should be possible, independently of the circuit configuration. Third, a model for substrate noise absorption and its effect on performance should be defined. All such models have been described in detail in previous chapters.

For each noise injecting module j a local noise generator is defined, which accurately reproduces substrate injected noise, taking into account both impact ionization and capacitive coupling through devices and interconnect lines. Using the notation introduced in Chapter 5 for the injection model associated with node j, let us define $G_j(t, \mathbf{\Pi}_j)$. The problem of evaluating the effects of substrate on performance is approached in the following way.

1. compute constraints for node of noise-sensitive modules
2. generate resistive network associated with substrate
3. quantify violations to constraints

In step 1 a set of bounds $\mathbf{\Pi}_j^{(bound)}$ is generated for a subset of *critical nodes* n_c using the constrained optimization techniques described in Chapter 5. Subset n_c is generated from the cumulative impact of all parasitic noise sources acting on each node as in [72].

In step 2 a given placement configuration is mapped onto a fully connected graph $G_S(V, E)$, whose vertices V are the substrate contacts and edges E are weighted by the conductance Y_{ij} or resistance R_{ij} between the corresponding vertices i and j of the $G_S(V, E)$. Figure 6.3 shows the mapping procedure. The techniques for the evaluation of the edges have been described in detail in this chapter. The calculation of all violations in step 3 to the given constraints is carried out by solving the circuit underlying $G_S(V, E)$ and evaluating the appropriate parameters at each critical node.

At each stage of the annealing, only steps 2 and 3 need be repeated, since step 1 is carried out only once for each chip. The efficiency of a Green's

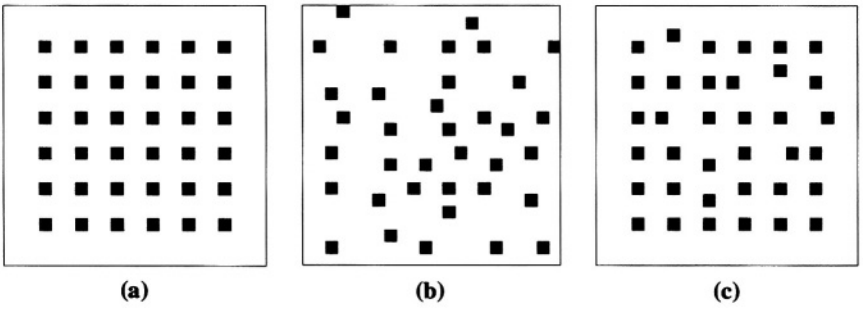


Figure 6.4. (a) Initial contact grid; (b) Reshuffling of contacts at high temperatures; (c) Resulting grid at lower temperatures

function based substrate simulator, though high, is still insufficient for such computationally intensive algorithms as SA, hence, appropriate heuristics must be developed. In SA, at high annealing temperatures, considerable reshuffling is allowed on the components of the layout. Hence, the locations of switching noise generators and receptors can be significantly modified. At lower temperatures on the contrary, modules move by lesser amounts in average. Hence, the edges of $G_S(V, E)$ change by a lesser extent and with lower frequency.

As an illustration consider a regular 36 contact grid shown in Figure 6.4(a). The plot of Figure 6.5 shows the average variation of the resistive components of the substrate network when high-temperature, Figure 6.4(b), and low-temperature, Figure 6.4(c), contact perturbations occur during the unfolding of SA. On the other hand, only when changes in the edges of $G_S(V, E)$ reflect a significant change in any performance measure K_i, the entire substrate network should be evaluated along with the estimate of performance degradation $\triangle K_i$. When a new temperature T_k is reached, the full graph $G_S(V, E)$ is solved, i.e. all the edges in V are evaluated exactly, using the Sherman-Morrison update to obtain the new matrix $\mathbf{P}^{-1}$. After a new move m_k and the associated translation $\triangle \mathbf{v} = [\triangle v_x \triangle v_y]^T$ is selected by the annealing algorithm, the sensitivity of the edges of $G_S(V,E)$ can be efficiently computed using the techniques outlined in this chapter. Suppose subset n_c of all critical receptors has been derived for the circuit, moreover let n_s be the subset of all noise injecting nodes. Let $[\mathbf{Y_c}]_{m_k}$ be the conductance matrix of all the nodes in n_c and in n_s and let $[\triangle \mathbf{Y_c}]_{m_k}$ be its update. By Equation (6.7), term $[\triangle \mathbf{Y_c}]_{m_k}$ is estimated as

$$[\triangle \mathbf{Y_c}]_{m_k} \approx \left[\nabla_{\mathbf{v}} \mathbf{Y}^{\mathbf{T}}\right]_0 \triangle \mathbf{v}, \tag{6.8}$$

where term $[\nabla_{\mathbf{v}} \mathbf{Y}^{\mathbf{T}}]_0$ is defined as in Equation (6.7) for matrix $\mathbf{Y_c}$. After updating $\mathbf{Y_c}$, the resistive network is solved and $\mathbf{\Pi}_j$ can be evaluated for all

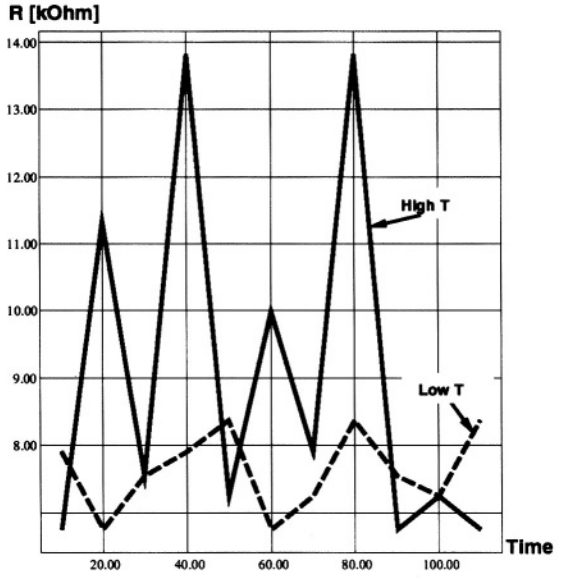


Figure 6.5. Resistive network reacting to high- and low-temperature contact reshuffling

```
foreach temperature T_k
    evaluate_substrate_network_exactly;   // Sherman-Morrison update
    repeat
        m_k = select_move;
        estimate_network_change(m_k);      // use gradient based method
        foreach node j ∈ n_c
            G_j(t, Π_j) = estimate_cumulative_noise;
            if Π_j ≥ Π_j^(bound)
                evaluate_substrate_network_exactly;
            go_to_next_node;
        evaluate_cost_function;
        accept_or_reject_move;
    until equilibrium_reached
```

Figure 6.6. Heuristic for the combined use of Sherman-Morrison and gradient based methods

critical nodes. By comparing $\mathbf{\Pi}_j$ with the bound $\mathbf{\Pi}_j^{(bound)}$ one can obtain the corresponding violation. If a violation to specifications has occurred, then a precise extraction step must be performed and the precise value for the violation

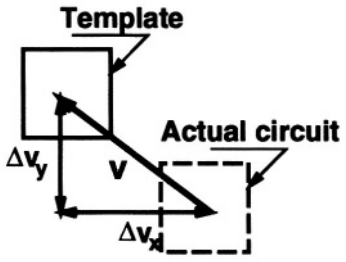


Figure 6.7. Computation of update matrix δc based on contact displacement relative to template

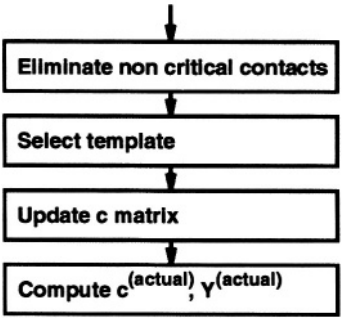


Figure 6.8. Block diagram of the template-based substrate extraction algorithm

is used to drive the cost of the annealing in a manner similar to [70]. Otherwise the contribution of substrate noise to node j in degrading performance K_i is considered negligible and the cost function will not take it into account. However, the cost relative to the remaining analog-specific constraints, as well as area and wiring length, will be computed. The placement algorithm is proved to converge to a global minimum under the same conditions of [73] and [74] when it is modified to account for noise substrate transport evaluation [44].

2. TEMPLATE-BASED EXTRACTION

In Chapters 3 and 4 techniques were presented to speed-up the extraction process and to simplify the schematic based on the knowledge of contact loading. In this section we discuss a method for further reduction of the extraction time of large circuits that share a set of recurring contact patterns.

Figure 6.8 illustrates the technique through a block diagram. First, a set of templates with N_c or more contacts, for which an extracted schematic exists, is compared to the sample layout. Among the available ones, a template is selected and its pre-computed coefficient of induction matrix $\mathbf{c}^{(\mathbf{template})}$

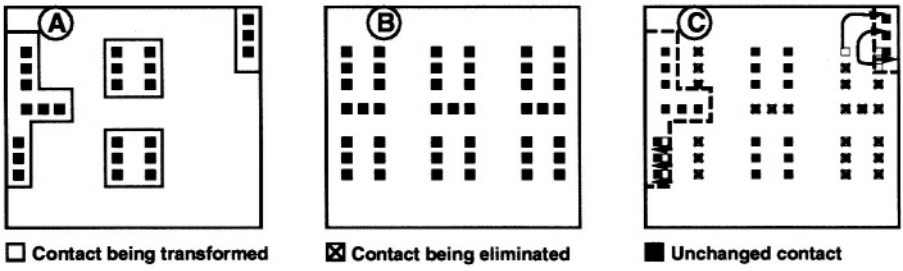


Figure 6.9. Speed-up mechanism for the extraction of large substrates

is used to compute $\mathbf{c}^{(\mathbf{actual})}$, the matrix associated with the actual circuit. Each progressive update matrix $\delta\mathbf{c}$ is computed based on the displacement $\mathbf{v} = [\triangle v_x \triangle v_y]^T$ of each contact non overlapping exactly with a corresponding contact in the template, as shown in Figure 6.7. Finally, the partial conductance matrix $\mathbf{Y}^{(\mathbf{actual})}$ is computed directly from $\mathbf{c}^{(\mathbf{actual})}$ as

$$\mathbf{Y}^{(\mathbf{actual})} = \mathbf{X}^{\mathbf{T}}\mathbf{c}^{(\mathbf{actual})}\,\mathbf{X}. \tag{6.9}$$

Figure 6.9(A) shows an example of physical layout being extracted. The template selected for this circuit is shown in Figure 6.9(B). The procedure of eliminating and aligning some of the contacts of the template onto the actual circuit is shown in Figure 6.9(C). In order to derive bounds on the time complexity of the procedure, consider the following cases. First, assume the worst-case scenario, i.e. no contact exists which overlaps exactly with a contact in the template. In this case, N updates are needed for complete substrate evaluation, the resulting complexity is therefore $O(N^3)$. This case is equivalent to a full inversion of matrix $\mathbf{P}$, hence no improvement is achieved over the non-simplified substrate extraction. Second, consider the case in which the sample and the template are identical. In this case no computation is needed, hence the extraction complexity is zero. The second scenario, or one as near as possible to it, is most desirable.

Since the complexity of computing an update of matrix $\mathbf{c}$ is independent of the transformation involved, an effective criterion for selecting the template is one aimed at maximizing N_o, the number of contacts exactly overlapping a contact in the actual circuit layout. Consequently, assuming that $N_o < N$ contacts differ in location from corresponding contacts of a template, the complexity of the procedure could be a fraction of that needed to invert $\mathbf{P}$.

In real circuits however, a large number of contacts rarely overlaps to those on the template. To cope with this problem, we propose a criterion based on performance sensitivities for the template selection and the minimization of updates needed for full extraction given pre-defined accuracy constraints. The modified template-based substrate extraction algorithm is described in

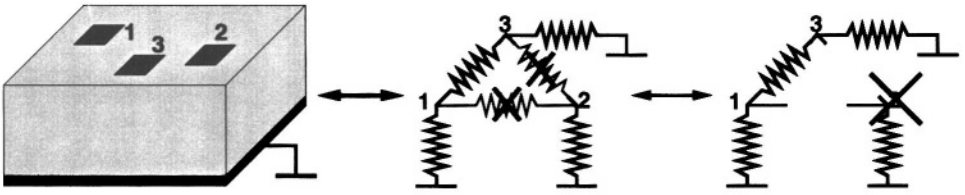


Figure 6.10. Elimination of all non-critical conductances and contacts

Figure 6.11. For simplicity but without loss of generality, let us consider only one performance function K. Assume that the matrix $S_{Y_{ij}}$ of the sensitivities of K with respect to all partial conductances Y_{ij} has been computed or estimated. Moreover, assume that estimates exist for the maximum values of all substrate conductances[2]. Using a fraction of the specified maximum degradation of K as threshold, all conductances, whose cumulative effect on performance is lower than the threshold, are eliminated from the schematic. All nodes connected to one or zero conductances are also eliminated as illustrated in Figure 6.10. The resulting substrate configuration must be then compared with a set of templates and the best template must be selected. This problem is solved using optimization. A byproduct of the selection procedure is the set D of all contacts that need be extracted in all details. The displacements of the contacts in D, relative to the selected template, are identified and updates needed for the computation of $\mathbf{c}^{(\mathbf{actual})}$ are computed using the Sherman-Morrison formula. Partial conductance matrix $\mathbf{Y}^{(\mathbf{actual})}$ is finally derived directly from $\mathbf{c}^{(\mathbf{actual})}$ using Equation (6.9).

Hereafter, the template selection procedure is illustrated. Let us consider matrix update $\delta\mathbf{c}|_i$ representing the move of contact i from its location in the template to that of the actual circuit. The coefficient of induction matrix $\mathbf{c}^{(\mathbf{actual})}$ associated with the actual circuit is computed as

$$\mathbf{c}^{(\mathbf{actual})} = \mathbf{c}^{(\mathbf{template})} + \sum_{i \in D} \delta\mathbf{c}|_i, \tag{6.10}$$

where D is the set of all the contacts whose locations in the template and in the actual circuit are non-identical and hence need be extracted in full detail. Combining Equations (6.10) and (6.9) one obtains

$$\mathbf{Y}^{(\mathbf{actual})} = \mathbf{Y}^{(\mathbf{template})} + \sum_{i \in D} \mathbf{X}^{\mathbf{T}} \delta\mathbf{c}\Big|_i \mathbf{X}, \tag{6.11}$$

where $\mathbf{Y}^{(\mathbf{template})} = \mathbf{X}^{\mathbf{T}}\, \mathbf{c}^{(\mathbf{template})}\, \mathbf{X}$ is the pre-computed partial conductance matrix of the template. Let us define the error matrix, i.e. the update

needed to translate $\mathbf{Y^{(template)}}$ into $\mathbf{Y^{(actual)}}$, as

$$\boldsymbol{\epsilon} = \mathbf{Y^{(actual)}} - \mathbf{Y^{(template)}} = \sum_{i \in D} \boldsymbol{\epsilon}|_i, \tag{6.12}$$

where $\boldsymbol{\epsilon}|_i = \mathbf{X^T}\, \delta\mathbf{c}\Big|_i \mathbf{X}$ is the error matrix due to the displacement of contact i in the actual circuit relatively to the template[3]. Assume one could calculate $\boldsymbol{\epsilon}|_i$, $\forall i \in D$ *a priori*. Using the sensitivity[4] of performance K with respect to matrix $\mathbf{Y}$, performance degradation $\triangle K$ due to the displacement of contacts in the actual circuit relatively to the template can be calculated as

$$\triangle K = \mathbf{e}^T \left(\frac{\partial K}{\partial Y_{ij}} \odot \boldsymbol{\epsilon}\right) \mathbf{e}, \tag{6.13}$$

where $\mathbf{e}$ is a Nx1 unity vector such that $\mathbf{e} = [1, \ldots, 1]^T$. The $\odot$ operator is defined as following: $\mathbf{A} = \mathbf{B} \odot \mathbf{C}$ such that $a_{ij} = b_{ij}\, c_{ij}$. Combining (6.12) and (6.13), one obtains

$$\triangle K = \sum_{i \in D} \mathbf{e}^T \left(\frac{\partial K}{\partial Y_{ij}} \odot \boldsymbol{\epsilon}|_i\right) \mathbf{e}. \tag{6.14}$$

Let us define *weighted extraction inaccuracy* A_K of an extracted schematic with respect to performance K as the relative amount by which K varies if some or all parasitics are inexactly estimated. The weighted extraction inaccuracy is expressed as

$$A_K = \frac{|\triangle K| + \epsilon_p + \epsilon_r}{K_v}, \tag{6.15}$$

where ϵ_p and ϵ_r are the errors due to inaccurate parasitic and performance models, respectively, and K_v is the nominal performance value. Moreover, (6.15) reduces to $A_K \approx \frac{|\triangle K|}{K_v}$ if $\epsilon_p + \epsilon_r \ll |\triangle K|$. Suppose now that a constraint on the weighted accuracy $\overline{A}_K$ has been set

$$A_K \leq \overline{A}_K. \tag{6.16}$$

Then, Equations (6.15) and (6.16) can be used as a criterion for selecting the appropriate template

minimize	D $\forall$ templates	(6.17)
such that	$A_K \leq \overline{A}_K$.	

Problem (6.17) is *guaranteed* to have a solution, since a template with at least N_c contacts, all of them not overlapping with the actual circuit's contacts,

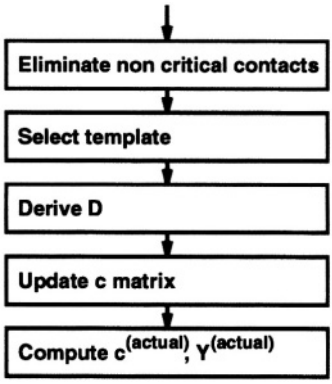


Figure 6.11. **Block diagram of the modified template-based substrate extraction algorithm**

exists by construction. Hence, arbitrarily small values of A_K can be achieved by simply extending D to include all the contacts i, $1 \leq i \leq N_c$. Problem (6.17) is solved by exhaustively calculating the minimum set D needed for each template for a given inaccuracy A_K. The procedure of calculating A_K and D has a time complexity of $O(N^2)$, while the overhead of computing $S_{Y_{ij}}$ is generally not accounted for since the evaluation is performed beforehand during circuit synthesis. Hence, a circuit with N_c contacts and a specification on A_K (6.16) can be extracted in $O([N_T + |D|]N^2)$ time, where N_T is the number of template circuits and $|D|$ the size of set D.

The final issue to be addressed is the efficient calculation of estimate $\epsilon|_i$, which can be computed exactly from update $\delta\mathbf{c}|_i$ using mapping $\mathbf{X}$ of Equation (6.11). However, a more efficient computation of $\epsilon|_i$, can be obtained using the approximation of (6.7). Consider all the contacts $i \in D$, assume that the locations of i in the template and in the actual circuit are *close enough.* Then, a two-dimensional Taylor expansion for $\delta\mathbf{c}|_i$ can be constructed as

$$\delta\mathbf{c}|_i \approx \left.\frac{\partial\mathbf{c}}{\partial v_x}\right|_i \triangle v_x + \left.\frac{\partial\mathbf{c}}{\partial v_y}\right|_i \triangle v_y = \nabla_{\mathbf{v}}\mathbf{c}|_i \triangle\mathbf{v}|_i, \tag{6.18}$$

where vector $\triangle\mathbf{v}|_i = [\triangle v_x \triangle v_y]^T$ represents the displacement needed to bring i from the template location to the location in the actual circuit. Term $\nabla_{\mathbf{v}}\mathbf{c}|_i = [\left.\frac{\partial\mathbf{c}}{\partial v_x}\right|_i, \left.\frac{\partial\mathbf{c}}{\partial v_y}\right|_i]^T$ is calculated using the Sherman-Morrison formula as in (6.7) and is valid for *small* displacements of contact i. Assume now that there exists a contact j in the vicinity of i which is displaced by $\triangle\mathbf{v}|_j$, where $|\triangle\mathbf{v}|_j|$ is also *small.* Assuming that the surrounding objects' relative distances from i and j are similar, one can estimate the cumulative effects of the displacement of the contacts as

$$\delta\mathbf{c}|_{i,j} \approx \nabla_{\mathbf{v}}\mathbf{c}|_i \triangle\mathbf{v}|_i + \nabla_{\mathbf{v}}\mathbf{c}|_i \triangle\mathbf{v}|_j, \tag{6.19}$$

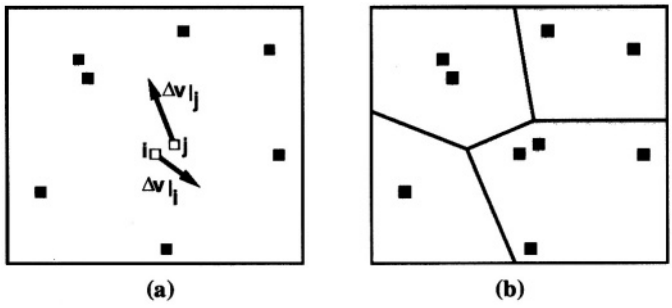


Figure 6.12. (a) Displacement of contacts i and j in a single landscape. (b) Partitioning of substrate to minimize the number of different contacts for which $\nabla_{\mathbf{v}}\mathbf{c}$ need be computed explicitly

where vectors $\triangle\mathbf{v}|_i$ and $\triangle\mathbf{v}|_j$ relate to the displacements of i and j as shown in Figure 6.12(a). Ideally, one would like to be able to compute $\mathbf{c}|_i$ using Equation (6.19) for each contact $i = 1, \ldots, N_c$. However, far contacts "see" a completely different landscape, which causes term $\delta\mathbf{c}|_i$ to change by moving within the workspace. To improve the accuracy of (6.19), one could partition of the workspace in order to minimize the number of contacts for which a new $\nabla_{\mathbf{v}}\mathbf{c}$ needs be computed. Figure 6.12(b) shows such a partitioning. Notice that only one contact per partition, the *pole,* is used for the computation of $\nabla_{\mathbf{v}}\mathbf{c}$.

The problem of minimizing the number of partitions of Figure 6.12(b) can be time-consuming, since it requires the estimation of each contact displacement to select the best candidates for the partitions and its poles. The complexity of this partitioning would nullify the efforts for an efficient substrate extraction. In addition, the needed parasitic estimate accuracy ϵ_p in Equation (6.15) is not high. Hence, in our experiments a single contact was used to estimate $\delta\mathbf{c}|_i, \forall i$ with an error of 50% or less. Moreover, this error could be modeled as term ϵ_p in Equation (6.15) and hence accounted for while determining D.

3. CASE STUDY

The circuit used in our experiments is a 140MHz monitor display controller (RAMDAC) including three D/A converters, a PLL frequency synthesizer, and digital control logic. The circuit was integrated in a Mosis HP 1μm CMOS technology. A low-resistivity substrate with a doping profile similar to the one depicted in Figure 2.2(b) was used. The converters were generated using dedicated silicon compilers [75]. The PLL needed particular care due to its extremely high sensitivity to thermal noise and spurious signals originated within the chip.

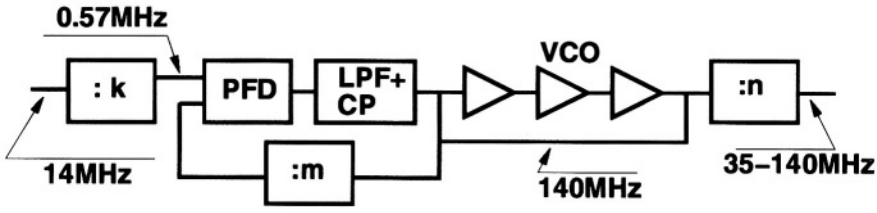


Figure 6.13. PLL schematic

Measure	PLL input freq.	n	VCO freq.	Specs
Stability	0.56MHz	100	56MHz	Yes
		250	140MHz	Yes
Jitter $\triangle T/T$	0.56MHz	250	140MHz	≤ 0.007
Ph. Margin	-	-	-	$\geq 45^{o}$

Table 6.1. PLL specifications

The PLL architecture, shown in Figure 6.13, was derived from [66], Device sizing was performed using a modified version of the supporting hyperplane algorithm and SPICE for circuit evaluation [76]. The circuit consists of a digital section, i.e. three divide-by-n modules and a phase-frequency detector (PFD), and a number of analog components, i.e. an analog low-pass filter (LPF) and a charge pump (CP). The interface between analog and digital sections is represented by the voltage-controlled oscillator (VCO), which generates a digital output at a frequency proportional to the input voltage. Typical frequencies of operation are shown in the various branches of the circuit in Figure 6.13.

The specifications for the PLL are summarized in Table 6.1. The jitter $\frac{\triangle T}{T}$ is defined as the ratio between variation from nominal of the oscillation period $\triangle T$ and period T. Due to the time-variance of $\frac{\triangle T}{T}$, it is generally measured in terms of its peak-to-peak or RMS value.

3.1 PHYSICAL DESIGN

The jitter performance of the PLL is entirely dependent on the jitter produced by the VCO. Using this fact, a sensitivity based model of the PLL could be constructed relating the PLL jitter performance to the level of the noise voltage peak-to-peak present at some 85 critical locations in the VCO. All critical substrate noise receptors were identified in the delay elements and in the two bias circuits using SPICE simulations accounting for both impact ion-

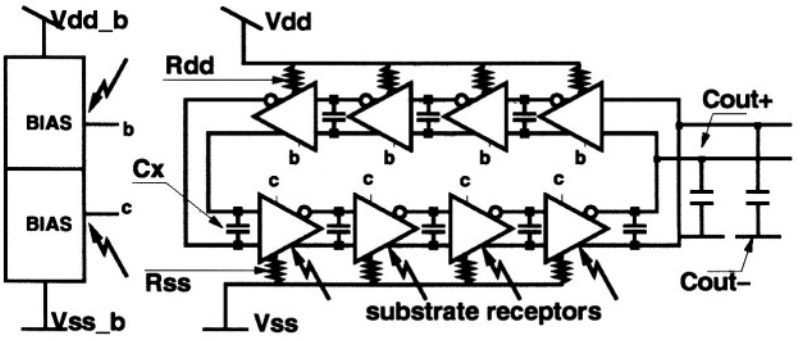


Figure 6.14. Interconnect parasitics and substrate noise receptors

ization and capacitive coupling. Interconnect parasitics and IR drops were also identified (see Figure 6.14). Sensitivities with respect to all parasitics (RC for interconnect and V_0 for substrate receptors) were computed. Then, constraint generator PARCAR [72, 61] was used to derive a minimal set of constraints on the maximum admissible noise voltage in each one of the receptors and on the maximum RC values for the interconnect parasitics in the VCO. The CPU time needed for the sensitivity analysis and constraint calculation was in total 2545 seconds, the results are shown in Table 5.3. Interconnect parasitic constraints were exploited by a constraint-based module generator VCOGEN to synthesize the VCO. TIMBERWOLF was used for the internal divider. The module generation step required a total of 163 seconds on a DEC AlphaServer 2100 5/250.

The next step was the placement of the component blocks of the PLL and of the other circuits in the RAMDAC. The placement was carried out using PUPPY-A. In the circuit there exist three major switching noise injectors, corresponding to the dividers. In order to accurately verify if the constraints on the maximum admissible noise voltage were violated, an accurate model was constructed of the injectors using the tool SUBWAVE [16,17]. SUBWAVE generates simplified substrate noise models, accounting for currents injected via capacitive coupling and impact ionization from active device areas and supply lines as described in Chapter 4.

Assuming that the substrate shows a purely resistive behavior, the calculation of the peak-to-peak voltage at each node of the surface can be carried out by performing a simple DC analysis on the positive and negative peak values of the current of the injector. The placement was performed using the heuristics summarized in Figure 6.6. The constraints on the maximum admissible noise voltage at each node of the VCO were used in the cost function of the annealing in a manner identical to [70]. Figure 6.15 shows the estimated values of switching noise voltage at each location in the chip at different temperatures

during the annealing. Figure 6.15(c) shows the substrate noise distribution at the end of the SA run. As expected, the algorithm successfully minimized the noise present in the substrate underlaying the PLL (compare layout in Figure 6.17). The plot of Figure 6.16 shows the impact of estimation algorithms on the relative error in substrate noise measured at the receptors during the annealing. All relative errors are obtained by comparison with an exact method, i.e. the Sherman-Morrison update. Curves (a) through (c) and (d) show how the constraint violation is driven towards zero depending on whether or not the proposed substrate injection control is used. Figure 6.17 shows the final placement performed using PUPPY-A. As expected divider n was placed at a large distance from the sensitive components of the PLL, namely the CP, VCO

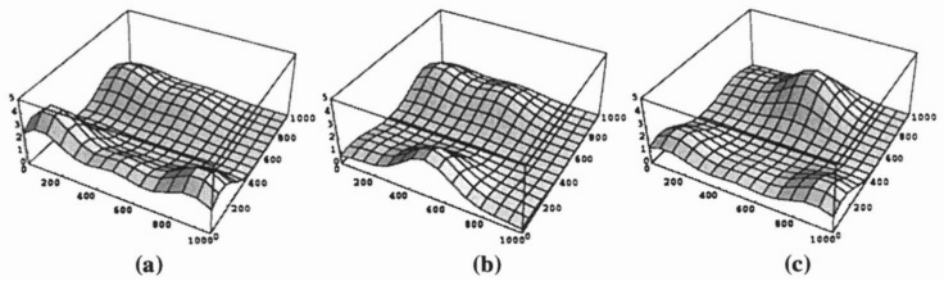


Figure 6.15. Estimated switching noise signal amplitude resulting from divider injection during SA. The signal was normalized with respect to the lowest constraint over the entire 1000x1000 μm chip. Noise injection at (a) high; (b) medium and (c) low SA temperature

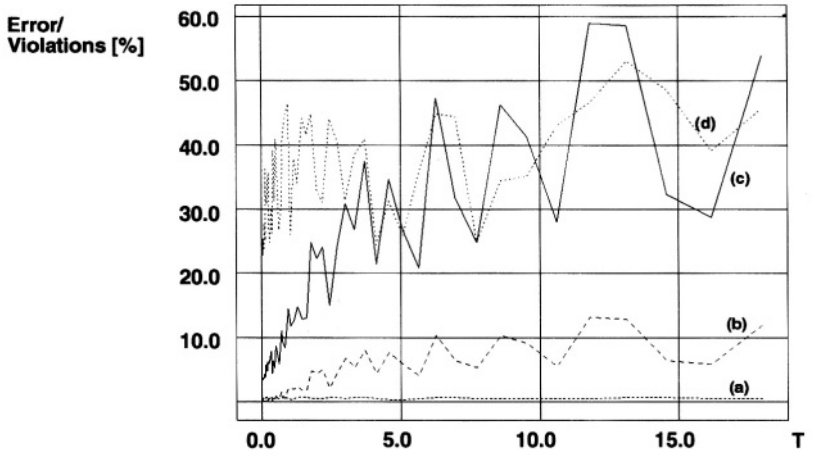


Figure 6.16. Error in substrate injection estimation using: (a) combined heuristic; (b) gradient based method only. All substrate violations using: (c) combined heuristic; (d) no substrate control

Mode	CPU (sec)	Area (λ^2)	Est. Jitter
Manual	-	5637 x 6481	-
Parasit.	406.74	6765 x 6528	0.1
Subst.+Parasit.	885.20	7322 x 7716	0.005

Table 6.2. Placement statistics obtained on a DEC AlphaServer 2100 5/250

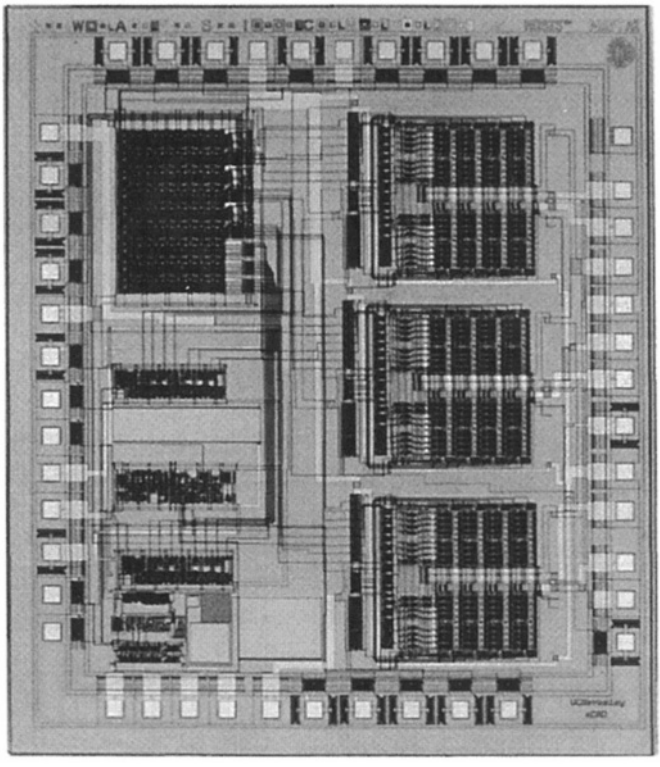


Figure 6.17. Microphotograph of the PLL integrated in the RAMDAC– Courtesy: Iason Vassiliou

and LPF. On the contrary, The sensitivity of these components with respect to the switching noise produced by divider k is small, hence it could be placed accordingly. For divider m the placer had to perform a trade-off between the strength of the switching noise received by it and the parasitics introduced when large interconnect capacitances are introduced. Using the same performance model employed in the constraint derivation, the jitter performance predicted in the PLL is summarized in Table 6.2.

Component	Number of receptors	Number of injectors
divider	-	152
PFD	-	23
VCO	85	-
LPF	5	-
CP	46	-
Total	136	175

Table 6.3. Noise injector and receptor statistics in the components of the PLL

3.2 TREND ANALYSIS AND TECHNOLOGY SCALING

All the potential sources of switching noise in the PLL are localized in the dividers, while the receptors are in the VCO, CP and LPF. Injection occurs by impact ionization through the active areas of NMOS devices (in a N-well processes) and by capacitive coupling through junctions and interconnect. Receptors are in the active areas of sensitive devices and supply lines. Table 6.3 lists the main sources and receptors of noise in the various components of the design.

Suppose one is interested in finding the change of jitter performance if a new lightly doped substrate is to be used instead of the low-resistivity substrate for which the circuit was designed. In this case performance K is the expression $\frac{\Delta T}{T}$. Since sensitivity $\partial \frac{\Delta T}{T} / \partial Y_{ij}$ is known, expression $\partial Y_{ij} / \partial T_\ell \ \forall i, j$ remains to be calculated. For simplicity, consider a single point-to-point substrate conductance, representing the resistive current path between a divider and one of the VCO's delay cells. Call such conductance Y_{13}. Note that in this case T_ℓ is a particular doping level associated with the layer of interest. The plot in Figure 6.18(a) shows the values of the sensitivities of entry $R_{13} = 1/Y_{13}$ at various nominal doping levels $(t_1, \ldots, t_3)$. Substrate impedance R_{13} was chosen as an illustration due to the high sensitivity of jitter with respect to it. The impedance is in fact responsible for approximately 20% of the noise generated in divider n and picked up by the VCO.

Consider now the dependence of impedance R_{13} as a function of another technology-specific parameter, namely the contact layer depth c. See plot in Figure 6.18(b). Lines $(t_1, \ldots, t_4)$ in Figure 6.18(b) represent the sensitivities of R_{13} at several values of c as computed using the formulae in Appendix B. Let us now consider the effects of changes in the doping profiles in Figure 3.4. Assume that the number of layers stays constant but the epitaxy expands towards the ground-plane while the underlaying layer shrinks. The plot in Figure 6.18(c) shows the sensitivities (t_1, t_2) of R_{13} as a dependence of the thickness of the

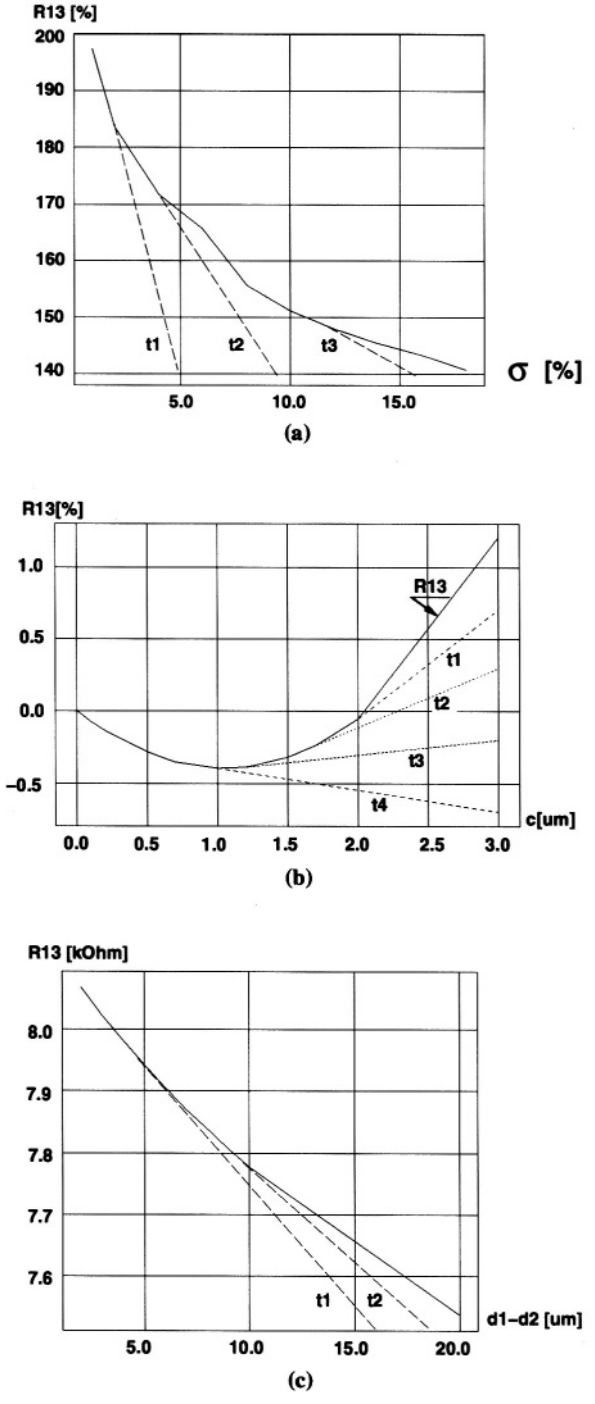


Figure 6.18. R_{13}: sensitivity with respect to (a) epitaxial doping levels, (b) contact depth, (c) epitaxial depths

Experiment	CPU times (sec)	Jitter trend
Epitaxial doping	3038.88	1.34
Contact depth	2858.83	0.95
Profile change	4005.46	0.55

Table 6.4. CPU times on a DEC AlphaServer 2100 5/250 for the trend analysis for the proposed experiments on the PLL with 311 noise sources / receptors. The CPU times include DCT, parameter and sensitivity computation. For the calculation of 311 contacts the inversion of matrix $\mathbf{P}$ was performed in 1525.0 seconds. The size of $\mathbf{P}$ was 1244x1244

epitaxial layer. Table 6.4 reports all CPU times for the sensitivities computed in the experiments and the estimated trend of jitter performance degradation calculated using (4.1).

3.3 ACCELERATED EXTRACTION AND TECHNOLOGY SELECTION

Consider a uniform 10x10 contact grid configuration. Table 6.5 lists the values of matrix Y using full and sensitivity based extraction for two configurations. All CPU times are referred to a DEC AlphaServer 2100 5/250 and relate to all computations except for the Green's function, which is performed once for a given substrate structure. The error is reported for all configurations. Note that a large circuit with 2,500 contacts could not be handled unless an extraction acceleration scheme was used.

Consider the uniform 10x10 contact grid again. Table 6.6 lists the mean and variance of the entries of matrix $\mathbf{R}$ as a function of depth variance $\sigma^2(c)$, assuming $\mu(c) = 1\mu m$. The execution times for the extraction of the mean and variance of $\mathbf{R}$ are also reported. For the uniform 10x10 grid example, suppose that all six substrate resistances R_{ij} and R_{i0} were critical and that constraints on each resistance were set as listed in Table 6.7. Technology $\mathcal{T}_2$ is more likely to meet the above specifications and hence it should be selected as best candidate.

Notes

1 The center of a polygon can be defined in a number of ways (center of mass, arbitrary edge, etc.). In this work polygons are always approximated with rectangular objects and the center is assumed to be the center of mass of the each object.

Coupling	method I	method II	error
R_{12}	8154.03Ω	8098.60Ω	0.67 %
R_{13}	1866.15Ω	1848.10Ω	0.96 %
R_{23}	3788.62Ω	3743.80Ω	1.18 %
R_{10}	893.85Ω	893.08Ω	0.08 %
R_{20}	460.40Ω	458.59Ω	0.39 %
R_{30}	690.77Ω	688.59Ω	0.31 %

uniform 10x10 contact grid

# contacts	method I	method II	max. error
100	73.8 sec	9.8 sec	3 %

uniform 10x10 contact grid

# contacts	method I	method II	max. error
2,500	not completed	57 min	5 %

industrial mixed-signal circuit

Table 6.5. Substrate extraction in presence of varying technology parameters using method I (full extraction) and method II (sensitivity based extraction)

Coupling	$\mu(R)$ @ $\mu(c) = 1\mu$m	$\sigma^2(R)$ @ $\sigma^2(c) = 0.25\mu\text{m}^2$	$\sigma^2(R)$ @ $\sigma^2(c) = 0.1\mu\text{m}^2$
R_{12}	8066.33Ω	$0.0270\Omega^2$	$0.0109\Omega^2$
R_{13}	1836.69Ω	$0.0040\Omega^2$	$0.0017\Omega^2$
R_{23}	3716.29Ω	$0.0324\Omega^2$	$0.1296\Omega^2$
R_{10}	892.11Ω	$2.5\times10^{-7}\Omega^2$	$10^{-7}\Omega^2$
R_{20}	456.35Ω	$10^{-4}\Omega^2$	$4\times10^{-5}\Omega^2$
R_{30}	685.69Ω	$2.25\times10^{-4}\Omega^2$	$9\times10^{-4}\Omega^2$

uniform 10x10 contact grid

# contacts	$\mu(R)$	$\sigma^2(R)$
100	73.8 sec	16.0 sec

uniform 10x10 contact grid

Table 6.6. Mean and variance of the entries of matrix $\mathbf{R}$ as a function of depth variance. All values are referred to a mean depth of 1μm. The execution times are reported for a uniform 10x10 contact grid

Coupling	$R_{ij}^{(bound)}$	$P_{\mathcal{T}_1}$ @ $\mu(c) = 0, \sigma^2(c) = 0.1\mu\text{m}^2$)	$P_{\mathcal{T}_2}$ @ $\mu(c) = 1, \sigma^2(c) = 0.25\mu\text{m}^2$)
R_{12}	8066.5Ω	0.849569	0.948270
R_{13}	1836.8Ω	0.959005	0.996184
R_{23}	3716.4Ω	0.729437	0.620028
R_{10}	893.0Ω	1.0	1.0
R_{20}	457.0Ω	1.0	1.0
R_{30}	685.8Ω	1.0	0.999877
Total	-	0.923001	0.927393

Table 6.7. Selection of most suitable technology based on the probability of satisfying all constraints on substrate coupling resistances

2 Rough estimates of the maximum/minimum value of substrate conductances can be easily computed from a simple set-up of two contacts located at chip edges or in close proximity

3 Assume all the other contacts are **not** displaced.

4 The sensitivity of K with respect to matrix $\mathbf{Y}$ is a $N_c \mathbf{x} N_c$ matrix, whose terms in the i**th** row and j**th** column are given by the expression $\partial K / \partial Y_{ij}$.

This page intentionally left blank

Chapter 7

IMPACT OF SUBSTRATE ON PERFORMANCE

In this chapter, we consider the effects of substrate impedance on circuit and device performance. We will classify these effects into three primary types. As we have discussed previously, the silicon substrate is a distributed impedance. Devices act as embedded localized signal injectors and sensors in this medium. The substrate provides a parasitic coupling path for signals within the same circuit. Thus it acts as a feedback mechanism, which may have grave effects on circuit performance. In other cases, especially in highly integrated mixed signal applications, signals from one circuit can be coupled into another and degrade circuit performance due to interference effects. In many cases, the coupled signals are not correlated with the signals at the internal circuit nodes, and hence appear as tonal or wide band noise in the circuit of interest, which is another effect of substrate coupling. The third effect arises from the dominant real lossy part of the substrate impedance. Losses in the substrate degrade the performance of passive components such as inductors. The resistance of the substrate also has thermal noise associated with it. In some applications, where a very low noise floor is desirable, the thermal noise of the substrate can degrade circuit performance by increasing the noise floor.

The above effects of substrate coupling are often unintended and undesirable. It is not possible to mitigate these effects completely. However prior knowledge of these effects can help the circuit designer avoid obvious sources of error. The application of sophisticated extraction and simulation methodologies presented earlier will help in discovering some more insidious effects. We will demonstrate the above classes of substrate effects with the aid of practical examples in the following sections. Note that the considerations made in this chapter are valid in general for current standard low-resistivity and high-resistivity substrates. Future trends in technology development however do not suggest fundamental changes of the conclusions resulting from this study.

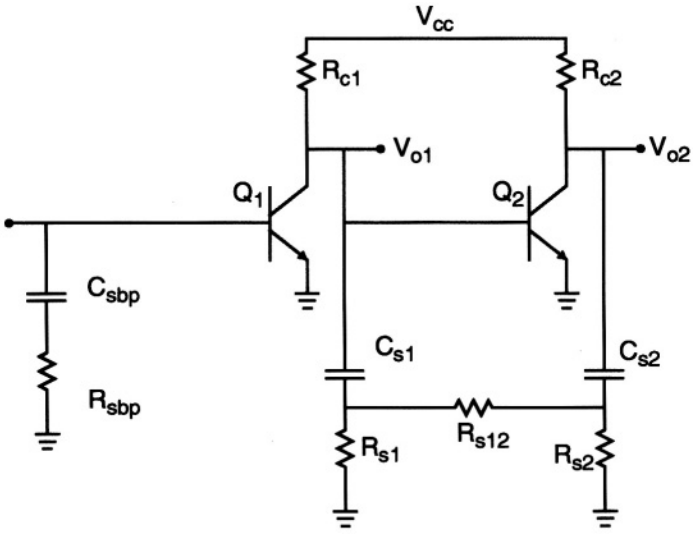


Figure 7.1. Accurate representation of substrate feedback in an amplifier

1. FEEDBACK

In order to examine the feedback path which exists through the substrate, we consider a two-stage bipolar transistor amplifier shown in Figure 7.1. A partial circuit diagram of a cascade of common-emitter amplifiers is shown in the figure. C_{sbp} is the capacitance between the bond-pad and the substrate and R_{sbp} is the substrate resistance from the bond-pad to the ground. C_{sbp} is typically in the range of 0.5 pF to 1 pF while R_{sbp} can vary from several ohms to several $\mathbf{k\Omega}$, depending on the type of the substrate used. $C_{s1(2)}$ and $R_{s1(2)}$ are the collector-to-substrate capacitances and the substrate resistances respectively for $Q_{1(2)}$. R_{s12} models the interaction of Q_1 and Q_2 through the substrate.

If the substrate model is not included in the circuit simulator, then the bottom plate of the collector-to-substrate capacitors in Figure 7.1 will be at ground potential. With the inclusion of the substrate model, a feedback path through the substrate between the collector and the base of transistor Q_2 can be observed from the figure. If R_{s1} and R_{s2} are large compared to the reactance of C_{s1} and C_{s2} at the frequencies of interest, and if R_{s12} is small[1], then the series combination of C_{s1} and C_{s2} appears as a Miller-multiplied capacitance at the collector of Q_1. This effect can be significant if Q_2 provides a large gain. For example, if C_{s1} and C_{s2} are 0.2 pF each and Q_2 has a gain of 10 then the Miller-multiplied term at the collector of Q_1 is 1 pF. This capacitance adds directly to the input capacitance (C_p) of Q_2 and causes a large fractional change in the bandwidth of the circuit. The reduction in the bandwidth of a

series-series feedback amplifier is discussed in [15], where inclusion of the substrate macromodel in the circuit simulator is shown to cause over a five percent bandwidth reduction from 1.9GHz to 1.8GHz.

There are several schemes to mitigate this effect. If a guard-ring is placed in close proximity to the collector of Q_1, then the bottom plate of the collector-to-substrate capacitance will be closely tied to the ground potential. R_{s1} and R_{s2} are effectively reduced by the guard ring, which reduces the feedback through C_{s1} and C_{s2}. In some process technologies, it may be possible to use a grounded backplate contact. This also has the effect of reducing R_{s1} and R_{s2}. A differential implementation of the amplifier reduces this problem as well. If we implement the two-stage amplifier of Figure 7.1 with differential amplifiers, then the bottom plates of the collector-to-substrate capacitances of the second differential pair are tightly coupled, if the devices of the differential pair are physically in close proximity. Thus, the signal at the bottom-plates does not vary to the first order, and feedback to the input of the second stage is minimized. This technique is expensive, since it involves doubling the power and device count in the amplifier. It is also possible to use feedback around the amplifier, to increase the bandwidth of the stage, but this is achieved at the expense of gain in the stage.

The substrate acts as a feedback path in MOS amplifiers as well and the Miller effect discussed above can be seen there too. A DC feedback path can also exist in MOS amplifiers due to body effect and the drain-to-body transconductance (g_{db}) caused by hot-electron induced holes. These effects can be studied in specific cases by using SUBRES.

In addition to the above effect, the substrate path effectively limits the maximum open-loop gain of amplifiers implemented on chip, since feedback through the substrate can lead to oscillation, if the Barkhausen criterion is satisfied.

The power gain of input and output matched amplifiers is also degraded by the presence of feedback through the substrate. This is easily observed in the gain of conditionally stable amplifiers, that have a maximum stable gain given by the ratio $|S_{21}/S_{12}|$. The electrical impact of feedback through the substrate is an increase in the magnitude of the reverse transmission parameter $|S_{12}|$. This directly impacts the maximum stable gain. Cascode amplifiers that achieve very high maximum stable gain by minimization of the reverse transmission (S_{12}) are especially sensitive to feedback through the substrate.

As a specific case we consider a conditionally stable common-emitter amplifier. This amplifier has a maximum stable gain given by $g_m/\omega C_\mu$, where g_m is the device transconductance and C_μ is the collector-to-base feedback capacitance. Substrate feedback is modeled as shown in Figure 7.2, and we assume for simplicity that the resistors R_{sbp} and R_{s1} are large enough that they can be ignored. If the amplifier is conditionally stable even after the substrate

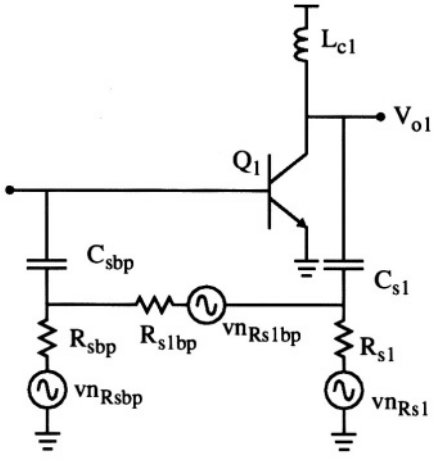


Figure 7.2. Substrate thermal noise sources

feedback is applied[2] the maximum stable gain is given by

$$G = \frac{g_m}{\omega[C_\mu + (C_{s1}||C_{sbp})/\{1 + \omega R_{s1bp}(C_{s1}||C_{sbp})\}]} \tag{7.1}$$

where $(C_{s1}||C_{sbp})$ is the series combination of C_{s1} and C_{sbp}. The decrease in power gain caused by substrate coupling is apparent from the above expression.

2. LOCALIZED POTENTIAL SHIFTS

DC effects in the substrate impact the bandwidth and gain of amplifiers. It can be shown easily that a change in the substrate potential, $\triangle V_{bias}$, changes the junction capacitance by the following value

$$\triangle C_j = -\frac{C_j}{n} \frac{\triangle V_{bias}/\psi_0}{\left(1 + \frac{V_{bias}}{\psi_0}\right)} \tag{7.2}$$

where ψ_0 is again the built-in potential of the p-n junction, V_{bias} is the externally applied potential across the p-n junction and C_j is the depletion capacitance of the junction for zero external bias across the substrate. Thus the poles of an amplifier transfer function, that are set by or depend on the capacitance to the substrate are changed by shifts in substrate potential. This alters the bandwidth of the circuit.

The presence of local substrate contacts sets the substrate potential quite effectively in its immediate vicinity, especially in high-resistivity substrates and

reduces modulation of the substrate junction capacitance. Hence surrounding sensitive nodes of an amplifier by substrate contacts reduces undesirable changes in bandwidth.

In MOS amplifiers, a change in the substrate bias, $\triangle V_{bias}$, changes the bias currents of MOSFETs by the following quantity,

$$\triangle I_d = g_{mb} \triangle V_{bias}. \tag{7.3}$$

The above change in bias will alter the gain of the amplifier. Using substrate contacts near the device is an effective means of setting the DC bias close to the device and reducing this effect as well.

3. THERMAL NOISE

Since the substrate is a lossy medium, it exhibits thermal noise associated with resistive losses. This can be a problem in some circuit applications that require low levels of thermal noise. An example is a low noise amplifier (LNA) integrated on a silicon substrate. Consider the substrate noise sources present in a bipolar transistor shown in Figure 7.2. Each resistor used to model coupling is a thermal noise generator. Noise coupled into the input of the amplifier from these noise sources has the most significant impact on the signal-to-noise ratio of the amplifier, since it is amplified by the transistor. It has been shown that this noise source can degrade the noise figure of the LNA considerably, unless adequate precautions are taken [77]. These include shielding the input bond-pad of the transistor with a low-impedance shield (Figure 7.3). The shield can be implemented with a buried n-well, connected to the supply, or in a lower-level metal layer. Typically when the reactive and resistive parts of the bond- pad impedance are similar in value, the thermal noise appearing at the output of the LNA is maximized [15]. Thus, it is also possible in principle, to use a very high-resistivity substrate to minimize the effect of substrate thermal noise. In Figure 7.2 if the resistors used to model the substrate are of sufficiently large magnitude, the effect of substrate thermal noise becomes insignificant. This requires the use of specialized substrates, and may not be a cost-effective option.

4. SUBSTRATE LOSSES

In this section we will study the effect of substrate losses on circuit performance in both passive and active circuits.

4.1 ON-CHIP PASSIVES

Losses in the substrate degrade the quality factors of on-chip passives such as inductors and capacitors. Inductors built on silicon substrates typically have much lower Q than off-chip inductors because of two reasons: higher resistance

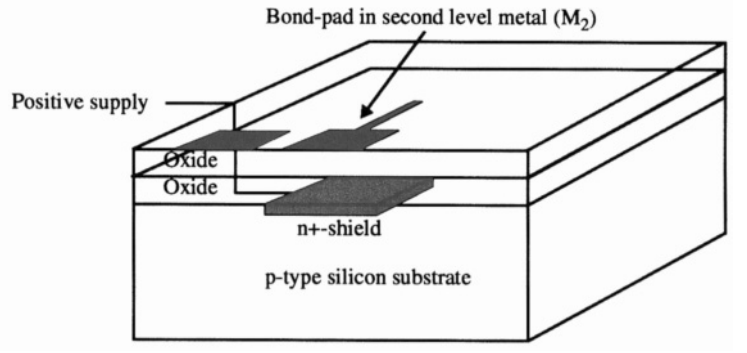


Figure 7.3. Bond-pad shield

of on-chip metallization and losses in the substrate. It has been shown that for frequencies near the self-resonance frequency of the inductor, substrate losses play a crucial part in lowering the Q of the inductor. Similar effects are seen in capacitors as well. Care must be taken in connecting capacitors to sensitive RF nodes. If the signal is applied to only one side of the capacitor, and the other side is connected to a reference node, then it is usually recommended to connect the top-plate to the active circuit node, and the bottom plate to the reference node.

Examples of this include on-chip LC tank circuits and high-frequency bypass applications. By connecting the capacitor in this manner, undesired current flow into the substrate is avoided, and the Q of the capacitor is maintained (Figure 7.4 (a)).

When used for AC coupling in single-ended circuits, the substrate resistance of the bottom-plate can be helpful, since it increases the net bottom-plate impedance. This reduces the attenuation caused by the bottom-plate parasitic. It is often the practice to extract accurate "brittle" (non-scalable) models of high-frequency active and passive components from experimental test-chips for library components. These are components of fixed electrical value. The extracted models are also useful for calibrating substrate simulators. The distributed nature of the substrate impedance brings forth special challenges in the model extraction and fitting procedure which is illustrated by the following potential pitfall. Consider two capacitors C_1 and C_2 used for on-chip AC coupling in a differential circuit (Figure 7.4 (b)). The capacitors are assumed to

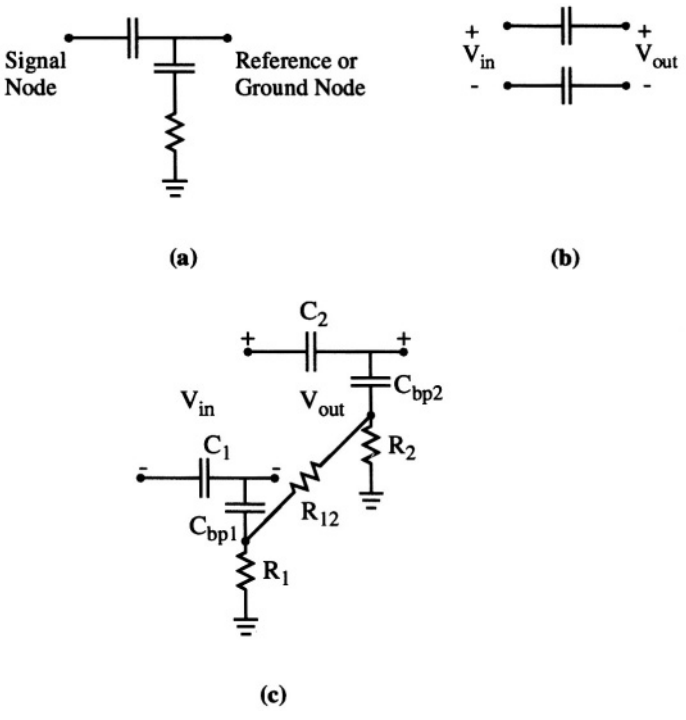


Figure 7.4. Capacitor modeling issues: (a) Minimizing substrate current flow; (b) Capacitors in a differential mode; (c) Exact model of differential capacitors

be laid out in close proximity. If the capacitors are modeled at high frequencies for single-ended applications, then the model of the bottom-plate impedance of each capacitor will typically consist of a capacitor and a series resistance. A brittle model for such a capacitor is shown in Figure 7.4 (a).

In high-resistivity substrates, the series resistance can be large compared to the bottom-plate reactance. Thus the bottom-plate impedance is constrained to a large value even at high-frequencies. However this single-ended model is a source of error if used for the differential application, because the bottom-plates are coupled quite closely by the element R_{12}, which is not reflected in the model of a single capacitor (Figure 7.4 (c)). Hence the capacitors C_{bp1} and

C_{bp2} attenuate the voltage at the input of the second stage. It is easy to see that, if the bottom plate capacitors are a certain ratio r of the primary capacitance, then the voltage at the output, V_{out}, is smaller than the input voltage V_{in}, by a factor $1/(1+r)$, if the input to the next stage has a much larger impedance than the bottom plate impedance. For $r = 0.2$, this translates into a 1.6dB signal loss, compared to the prediction from the brittle single-ended model. This is a situation where the use of a high-frequency model that includes substrate effects will lead to optimistic and erroneous results, since the model was developed for a different application.

This problem can arise in device models too, if a single-ended model is extracted for the impedance at the input and the output of the device. If the devices are used in a differential circuit, and placed in close proximity, the output of the device will see a smaller substrate impedance than it would in a single-ended application.

The modeling issue discussed above again points to the importance of using well-calibrated, substrate extraction tools. Single-ended models for device components may not evaluate coupling effects accurately. On the other hand, it is difficult to extract substrate models for multi-device structures at high frequencies, where a far higher simulation capability is necessary.

4.2 ACTIVE CIRCUITS

The impact of substrate losses can be significant, especially in tuned circuits, where an effort is made to keep parasitic resistive loading at the input and output nodes to a minimum. An example of this is an input and output matched LNA. A rudimentary case is shown in Figure 7.5.

With increasing losses at the input and the output of the LNA, the matching gains that can be achieved at these nodes are reduced. In a unilateral device, for example, the matching gain at the input is given by $1/(1-|S_{11}|^2)$ and at the output by $1/(1-|S_{22}|^2)$. For near-lossless input and output nodes, the total gain of the amplifier is enhanced considerably by impedance matching, since $|S_{11}|$ and $|S_{22}|$ are close to unity. However due to substrate losses, for example, losses in the bond-pad capacitance or the device junction-to-substrate capacitance, the magnitudes of these parameters are reduced [78]. This can be observed very simply, from the reflection coefficient of a series R-C network. Here the resistance *R,* represents the losses in the substrate, and the capacitance *C,* represents the capacitance of the junction to the substrate. Using the definition of reflection coefficient of a load, we have

$$|S_{11}| = \left|\frac{R + \frac{1}{j\omega C} - Z_0}{R + \frac{1}{j\omega C} + Z_0}\right| = \sqrt{\frac{(R-Z_0)^2 + 1/(\omega C)^2}{(R+Z_0)^2 + 1/(\omega C)^2}} \qquad (7.4)$$

For a negligibly small value of *R,* the reflection coefficient approaches unity. The same holds for a resistance that approaches infinity. In either case, the

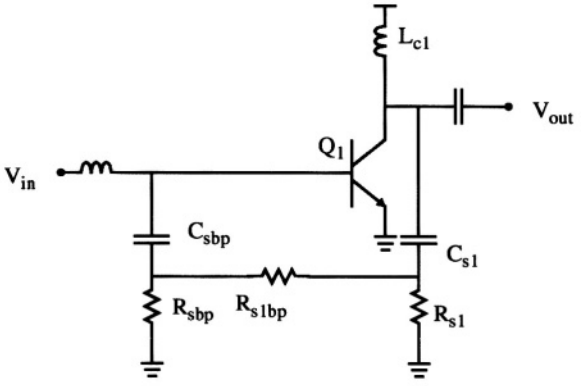


Figure 7.5. An input and output matched LNA

losses in the substrate approach zero, and it is possible in principle to achieve an infinite power gain in the matching network of the device. The reduction in matching gain can severely reduce gain in circuits with a very high output impedance like cascode amplifiers.

4.3 ON-CHIP COUPLING OF SPURIOUS SIGNALS

We now address another major impact of substrate coupling, namely the coupling of spurious signals into circuits. We had considered the impact of coupling on circuit performance in the context of feedback through the substrate. The difference here is that we consider signals, that are generated by other circuits on the same substrate, with weak or no correlation to the signals within our circuit of interest.

Most systems requiring a high dynamic range are implemented on boards, where each IC that performs a part of the signal-processing or any other function required in the system, is housed in a separate package. An example of this is a modern cellular phone handset. This is a particularly interesting example to consider, since the dynamic range requirement in these systems is large, often in excess of 100dB and several different signal-processing applications are required. Some of these applications include signal amplification, frequency conversion, frequency synthesis and analog-to-digital conversion. A typical transceiver board may consist of the following integrated circuits to implement these functions - VCOs, PLLs, a LNA and mixer front-end, an IC for intermediate frequency processing consisting of downconversion stages and analog-to-digital converters, a baseband digital IC, a power management

IC, an upconverter IC and a power amplifier (PA). It is theoretically possible to integrate all these circuits on one substrate. However a major problem with integration is interference between these blocks. Some of these possible interactions include:

1. Coupling of the PA output into the LNA input, especially in frequency-domain duplexed systems. The ratio between the power at the output of the PA to the sensitivity of the receiver can be greater than 100dB. In frequency-domain duplexed systems, the frequency bands of transmission and reception are different. Therefore isolation is achieved by means of narrow filters at the input of the receiver and the output of the PA. However, in an integrated implementation the substrate provides a strong coupling path that is not frequency selective. A large power level at the input of the LNA can reduce the gain of the LNA due to compression, and increase the noise floor by low-frequency noise upconversion.

2. In many implementations the digital clock frequencies can be several tens of MHz. Hence the switching noise generated by digital circuits has a spectrum that can extend to several GHz. For an ideal rectangular pulse train, the n-th harmonic is smaller than the fundamental by a factor of n, or $20\log(n)$dB. Thus, an ideal clock switching rail-to-rail in a 3V digital process, with a fundamental frequency of 10MHz, contains a harmonic close to a GHz, that is smaller by a factor of 40dB. In this ideal case, the fundamental has an amplitude of 1.91V and the above mentioned harmonic has an amplitude of 19.1mV. If the frequency of the received signal is similar to this harmonic, then this harmonic will appear in-band and look like noise. Filtering of high frequencies by junction capacitors and interconnect capacitors in the digital circuits does not help in reducing the level of the higher harmonics, since these capacitors are coupled to the substrate, and effectively convey high-frequency energy into the substrate. Device capacitors that are electrically isolated from the substrate, on the other hand, help in reducing the high-frequency signal content in the switching signals. These capacitors effectively couple the high-frequency signal energy into the circuit ground.

3. In addition to the harmonics of digital signals that couple in-band, low-frequency components of the digital switching noise can also be sensed and upconverted by circuits. As an example consider a single-ended LNA. In the presence of a "jammer" signal, which is an undesired large amplitude signal arising from a spurious external transmission, any low-frequency noise at the input port of the LNA will be upconverted to a frequency close to the desired radio-frequency. One mechanism responsible for this effect is the second-order harmonic distortion in the LNA. Assume that the

non-linearity of the amplifier can be expressed as a power series of the form

$$\nu_0 = \alpha_1 \nu_i + \alpha_2 \nu_i^2 + \ldots \tag{7.5}$$

If a jammer tone $\cos(\omega_j t)$ is applied to the input, and a low-frequency tone $\cos(\omega_{dig} t)$ is generated by the digital circuits, then the second order distortion term above will generate tones at $\omega_j + \omega_{dig}$ and $\omega_j - \omega_{dig}$. Depending on the frequencies of the jammer and that of the digital signal, either of these frequencies can fall in-band and appear like an effective increase in the noise floor at the output of the LNA. This is a severe problem in receiver circuits that exhibit high second-order distortion. This effect can be reduced by the use of differential circuits, that inherently have lower second-order (and even-order) distortion. This is similar to the desensitization process discussed in [79], where low-frequency circuit noise is shown to worsen the noise figure of an LNA.

4. Digital noise in the substrate can be transferred in band due to frequency translation and the generation of beat products in any non-linearity in the receiver. This problem exists in discrete designs as well, and the frequency planning of transceivers always incorporates information about spur generation and possible appearance of spurs in-band. In integrated applications it is important to understand the level of spurs coupled through the substrate, and incorporate this information into the frequency plan.

5. DC offsets can be generated in circuits that implement sampling or mixing functions. For example, in A/D converters, leakage of the clock waveform into the input of the sampler appears as a DC offset at the output. Similarly, leakage of the local oscillator of a mixer into its input can also lead to the appearance of DC offsets at the output of the mixer. This can be a problem in direct down-conversion receivers, since in this type of architecture, the down-converted signal is transferred to DC and can be much smaller than the DC offset caused by the substrate-coupled local oscillator signal.

6. The signal frequency of oscillators can be altered by the presence of large interferers near the frequency of oscillation. This phenomenon is known as frequency-pulling. It is also one of the primary determinants of isolation requirement in a transceiver. This is because of the very high power levels at the output of the Power Amplifier, that can modulate the frequency of the oscillator. Frequency-pulling depends on several factors - the quality factor of the resonator in the oscillator, the level of signal received at the oscillator input and the frequency separation between the signal frequency and the oscillator frequencies. In an integrated environment, due to increased coupling, the signal level from the PA into the oscillator is increased, which causes increased pulling.

7. Oscillatory loops can exist through the substrate, across high-gain stages implemented in the transceiver[3]. This problem is especially severe in architectures where very large gain is required at one frequency, for example in direct conversion receivers. In a classical heterodyne receiver, since the overall gain is achieved in a partitioned manner at several different frequencies, leakage from one part of the circuit to another will possibly be at a frequency which is attenuated by the circuit. Also, the isolation requirements at each stage are somewhat eased, since the gain implemented at a given intermediate frequency is modest. On the other hand, in direct conversion receivers, since almost the entire system gain (often in the range of eighty to a hundred dB), is implemented at one frequency, the isolation requirements in the stage are very stringent.

There can be several other issues brought forth by aggressive integration. Often these issues are difficult to anticipate beforehand, and require advanced simulation capability. Some of the above problems can be alleviated by circuit design or proper system design. For example, if a large digital circuit is switching at a certain rate, then information of the fundamental and the harmonics of the digital circuit should be incorporated into the frequency plan of the transceiver. In an analog-to-digital converter, the enabling of the output buffers can be staggered in time, compared to the sampling instant to reduce any coupling from the buffers into the sample-and-hold. Some problems, such as frequency pulling, are difficult to solve at the circuit design level, and one has to resort to the use of processing techniques to achieve the required isolation. We will discuss these issues at length later in this chapter.

4.4 A CIRCUIT APPLICATION FOR SUBSTRATE RESISTANCE

We have discussed several instances in the earlier section where the parasitics introduced by the IC substrate can cause performance degradation. One instance where the substrate resistance is actually used as a part of an application is in the design of Electrostatic Discharge (ESD) devices. These devices are used to absorb high-currents generated by electrostatic charge transfer on to circuit pins, from handling of circuits, without developing excessive potential on circuit pins, that can be destructive for the circuit. In a class of ESD circuits, an NMOS device is connected to the pin to be protected from ESD (Figure 7.6). As the voltage on the drain output rises, a current is injected into the substrate due to an avalanche carrier generation. The substrate current forward biases a parasitic substrate bipolar device which is used to absorb the ESD discharge. Details of this technique are found in [80]. Since the turn-on of ESD protection relies heavily on the value of the substrate resistance, modeling of this element is very critical in this application. In fact, while in the previous sections and

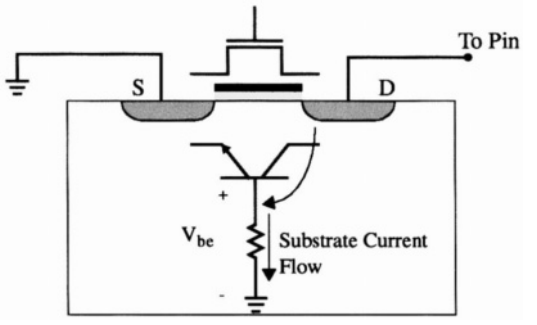


Figure 7.6. Electrostatic discharge through parasitic substrate npn

chapters we have looked at small-signal (linear) modeling of the substrate, it is important to include non-linear effects, such as conductivity modulation, for ESD design.

Notes

1 This can be the case, for example, in a high-resistivity substrate if the transistors are in close proximity, have large substrate capacitance, and the number and area of substrate contacts used to connect the substrate to ground are small.

2 If the device becomes unconditionally stable with the inclusion of the substrate feedback network, the gain will be less than the gain of Equation (7.1).

3 This is an effect considered in Section 1., since it is related to feedback through the substrate. However we mention it here for completeness.

This page intentionally left blank

Chapter 8

PHYSICAL DESIGN GUIDELINES

In this chapter several studies are presented to illustrate the behavior of substrate as a conductor as well as an insulator of noise. Based on these observations we propose a number of guidelines for noise-aware physical design. Let us consider again the substrate profiles depicted in Figure 8.1.

1. CHARACTERIZING CONDUCTION

Let us consider a single contact configuration. In Figure 8.2 the value of the impedance is plotted versus the width of the contact on a high-resistivity substrate. The experiment is repeated on a low-resistivity substrate. The results are shown in Figure 8.3. The resistance in the high-resistivity substrate shows a weak logarithmic dependence on the dimension of the contact, while the resistance in the low-resistivity substrate is a much stronger function of the contact dimension. In the case of the high-resistivity substrates, the presence of the low-resistivity epitaxial layer on the surface of the high-resistivity bulk region tends to increase the fringing fields. Thus the effective contact area appears to be much greater than the physical contact area. This explains the weak dependence of the resistance on the contact area. In the low-resistivity

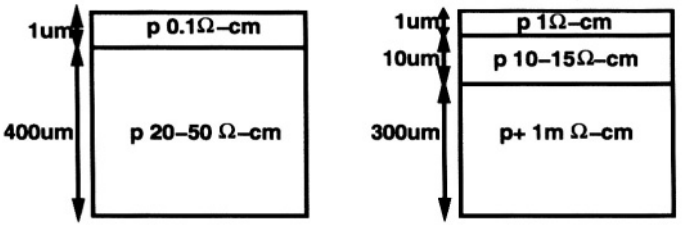


Figure 8.1. Typical substrate doping profiles: high resistivity (left); low resistivity (right)

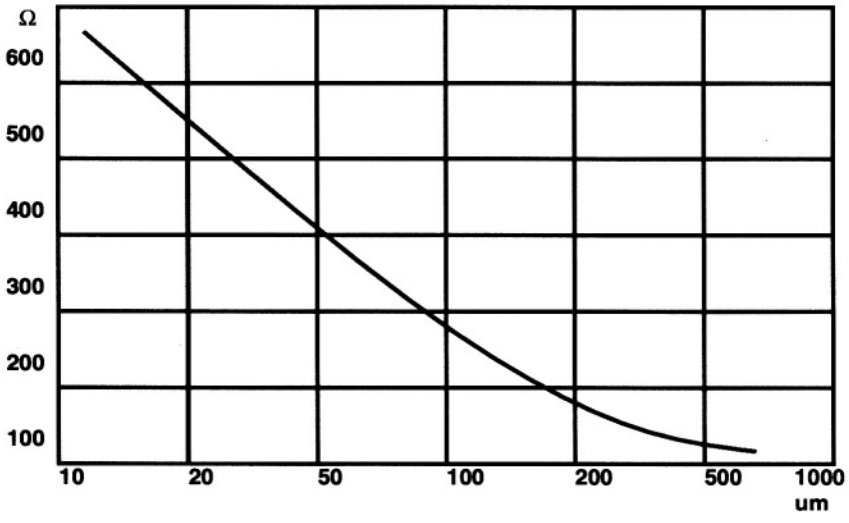


Figure 8.2. Impedance to ground for a single contact as a function of its width on a high-resistivity substrate

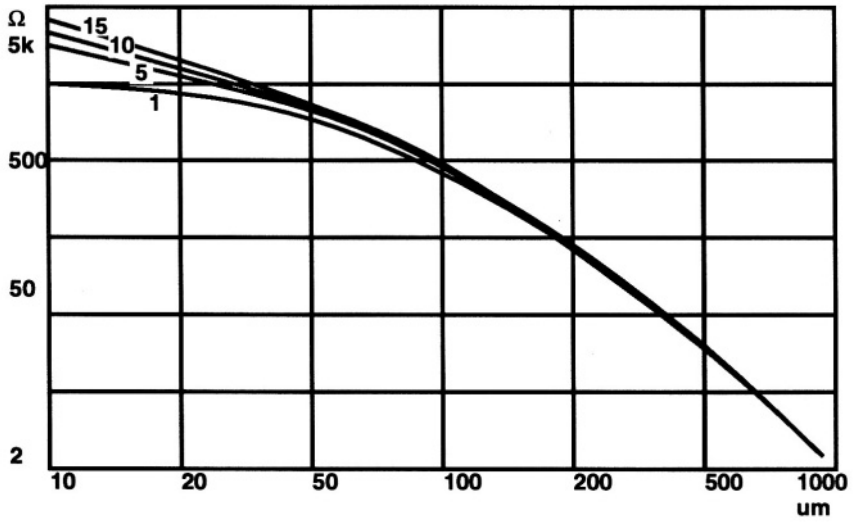


Figure 8.3. Impedance to ground for a single contact as a function of its width on a low-resistivity substrate with different values of epitaxial resistivity (1-15 Ωcm)

substrates, the presence of the thick low-resistivity bulk region very close to the surface reduces fringing fields. Thus a nearly inverse dependence of the contact resistance on the contact area is observed.

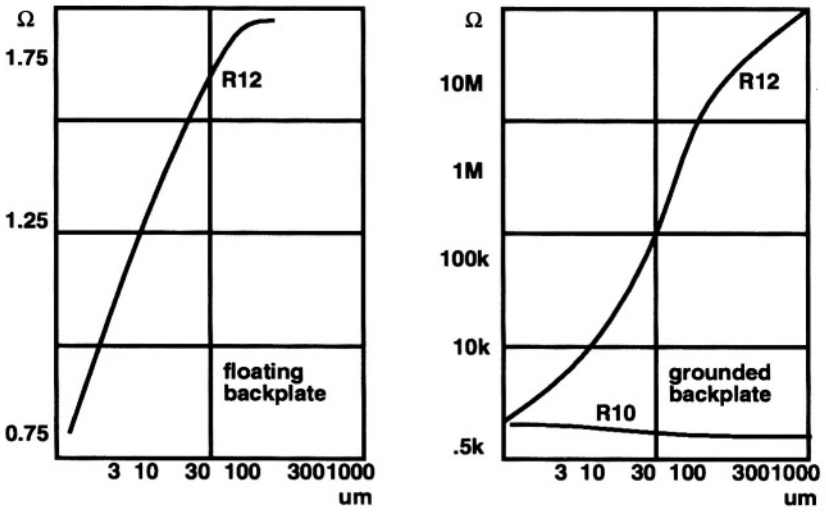


Figure 8.4. R_{10} and R_{12} simulations for a configuration with two $20\mu m$ square contacts as a function of their distance on a low-resistivity substrate

The dependence of the three-element model (Figure 3.3), in a two-contact problem, as a function of the distance between the contacts is shown in Figure 8.4, where the cross-coupling impedance R_{12} is simulated. The plots of Figure 8.4 show a low-resistivity substrate setup with floating and grounded backplate contacts.

Figure 8.5 refers to a high-resistivity substrate setup. The resistance between the two contacts increases monotonically as can be expected. The resistance to the backplate contact is not a monotonic function of separation. It is large when the two contacts are close, which implies that a significant portion of the current flow is in the surface region. As the distance between the contacts increases, most of the current flow is to the backplate, which causes a decrease in the resistance to the backplate. When the contacts are brought near the edge of the substrate, the effective resistivity of the substrate increases, as the resistivity beyond the edge is infinite. This leads to the increase in the backplate resistance for a large distance.

2. CHARACTERIZING ISOLATION

Isolation I between contacts is defined as the ratio of the voltage swing on the receiver contact V_{rcv} to the voltage swing on the injector contact V_{inj}. The isolation between two contacts depends on several factors besides the substrate model elements. These include the contact-to-substrate capacitances C_{sub}, the backplate contact impedance, the frequency of operation and the load impedance connected to the receiver contact R_{load}. The receiver contact

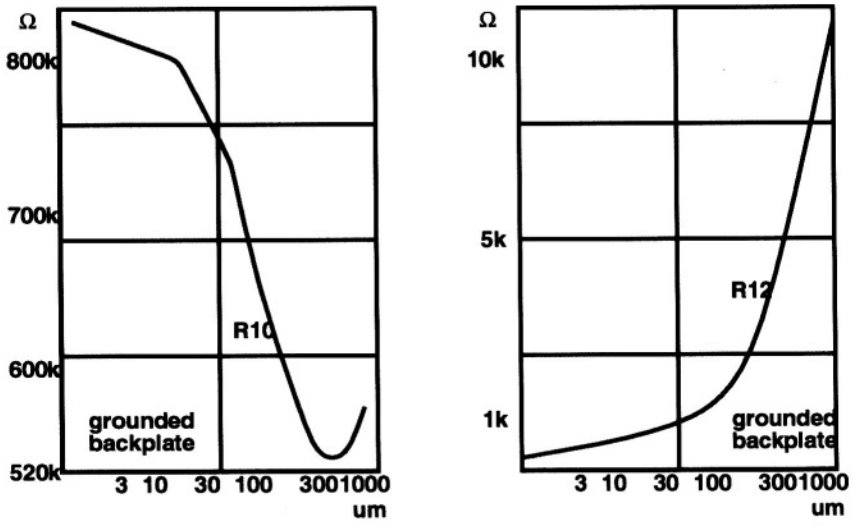


Figure 8.5. R_{10} and R_{12} simulations for a configuration with two $20\mu m$ square contacts as a function of their distance on a high-resistivity substrate

load may be, for example, the load resistance connected to the collector of a bipolar transistor. The collector of the device acts as a receiver due to the collector-to-substrate capacitance. The load has a significant influence on the magnitude of the isolation since the voltage on the receiver node depends on the value of the load. In what follows a nominal 50Ω load is used.

2.1 LOW-RESISTIVITY SUBSTRATES

Figure 8.6(a) shows the electrical setup used for the two-contact configuration of Figure 8.6(c). Both contacts are $50\ \mu$m in size, spaced by a variable factor δ. The central contact is fixed while the other contact is allowed to slide. The simulated isolation data are plotted in Figure 8.6(b). All the curves refer to a low-resistivity substrate with a profile similar to that of Figure 8. 1(b), with $\rho_{surface} = 1\Omega$cm, $\rho_{epitaxy} = 15\Omega$cm, and $\rho_{bulk} = 10$mΩcm.

In experiments (1) and (2) backplate contact is floating, in the other experiments a grounded backplate contact is used, while inductance L_{gnd} assumes values of 5nH in (3)-(4), 3nH in (5)-(6), 1nH in (7)-(8), and 0nH in (9)-(10). The frequency of operation is set to 1GHz in the odd- and 0.1GHz in the even-labeled experiments. Capacitance C_{sub} is assumed to be at 0.3pF in the plots of Figure 8.6(b) and 0.8pF in Figure 8.7(b).

In (1) and (2) the isolation is almost independent of the distance between the two contacts. This is due to the fact that the resistance R_{12} is much smaller than the reactance of C_{sub1} and C_{sub2} at the frequencies shown here. Thus

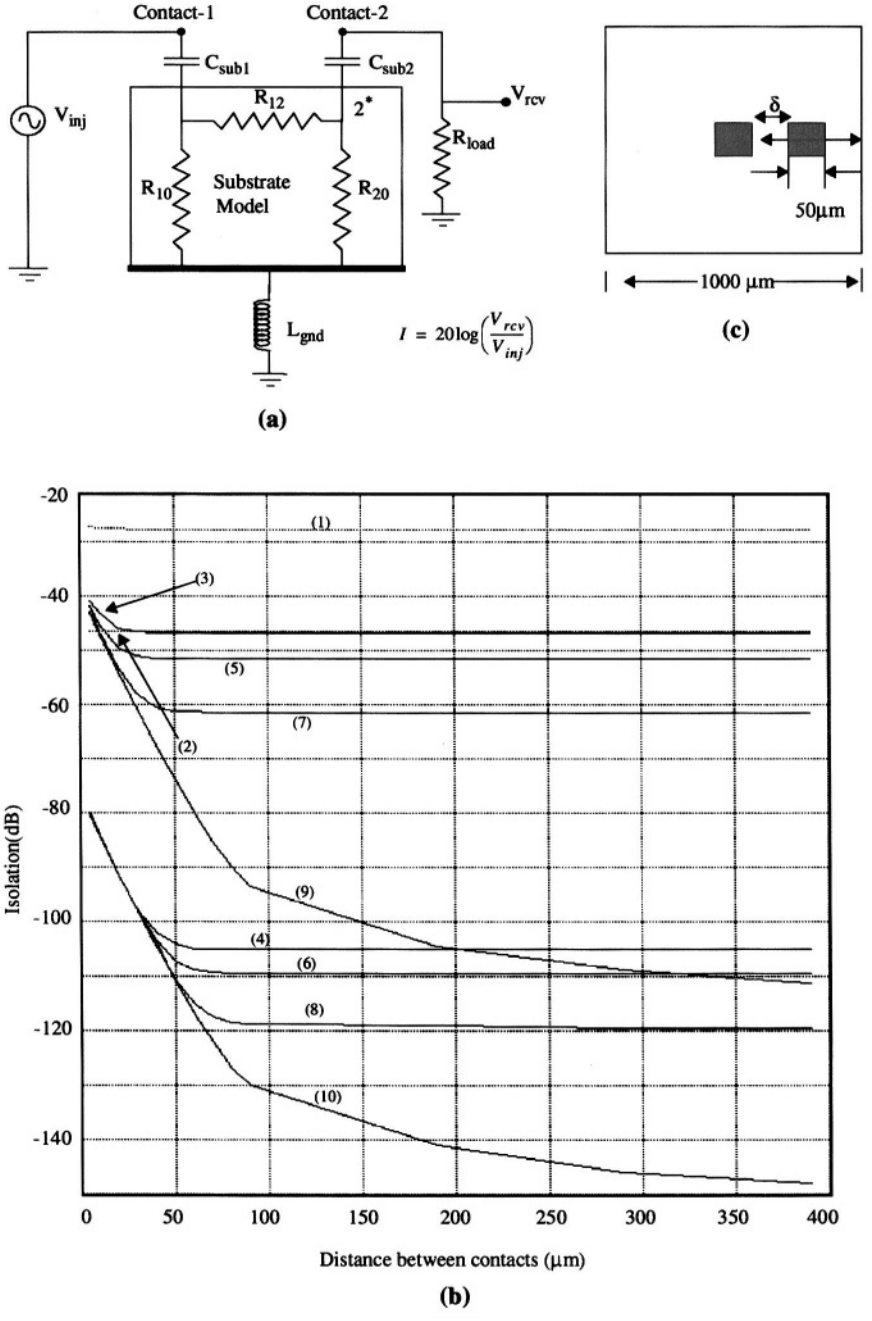


Figure 8.6. Isolation characterization (C_{sub} = 0.3 pF): (a) electrical setup; (b) simulations; (c) geometry

V_{rcv} is related to V_{inj} by the ratio of the resistor R_{12} and the reactance of the series combination of C_{sub1} and C_{sub2}. This is not a favorable environment for applications requiring high isolation since increasing the separation between the injector and receiver contacts is not helpful. In experiments (3)-(8), the isolation becomes independent of distance after a certain critical distance. Su et al. [11] found this critical distance to be four times the epitaxial thickness in their experiments. In fact, this distance is a function of the backplate impedance and varies from 2.5-5 times the epitaxial layer thickness (10μm) for L_{gnd} in the range of 1-5nH at 1GHz. For $L_{gnd} = 0$, the isolation continues to increase with increasing distance.

The isolation is very sensitive to the backplate impedance in these substrates. For a constant contact-to-contact distance, the variation in the substrate isolation is very large over the range of backplate inductance. Providing a small backplate impedance is critical in these substrates. If we assume that the reactance of C_{sub1} and C_{sub2} is small, the load resistance is small and R_{12} can be ignored. The impedance looking into the substrate between contacts 1 and 2 in Figure 8.7(a) is

$$Z'_{12}(j\omega) = R_{10} + R_{20} - j\frac{R_{10}R_{20}}{\omega L_{gnd}}. \tag{8.1}$$

The backplate inductance becomes a more severe problem for isolation between large contacts. This is the case because the resistance to the backplate falls almost inversely with contact area. Thus $Z'_{12}(j\omega)$ at high frequencies falls inversely with the area too, which implies that the high-frequency isolation worsens almost linearly with area. The reactance of C_{sub} also falls as the inverse of contact area. Thus the isolation will degrade considerably with area in these substrates.

The isolation values are very good for small ground-plane inductance values, as in experiments (7)-(10), especially at 100MHz. This is expected since the low-resistivity bulk restricts lateral current flow. Comparing the plots of Figures 8.6(b) and 8.7(b) shows that the isolation degrades as the capacitance C_{sub} increases.

2.2 HIGH-RESISTIVITY SUBSTRATES

The isolation between two surface contacts for a high-resistivity substrate is shown in Figure 8.8. All the curves refer to a high-resistivity substrate with a profile similar to that of Figure 8.1(a), with $\rho_{epitaxy} = 0.1\Omega$cm and $\rho_{bulk} = 20\Omega$cm.

In experiments (1) and (2) the backplate contact is floating, in the other experiments a grounded backplate contact is used, while inductance L_{gnd} assumes values of 5 nH in (3)-(4) and 3 nH in (5)-(6). The frequency of operation is set to 1GHz in the odd- and 0.1GHz in the even-labeled experiments. Ca-

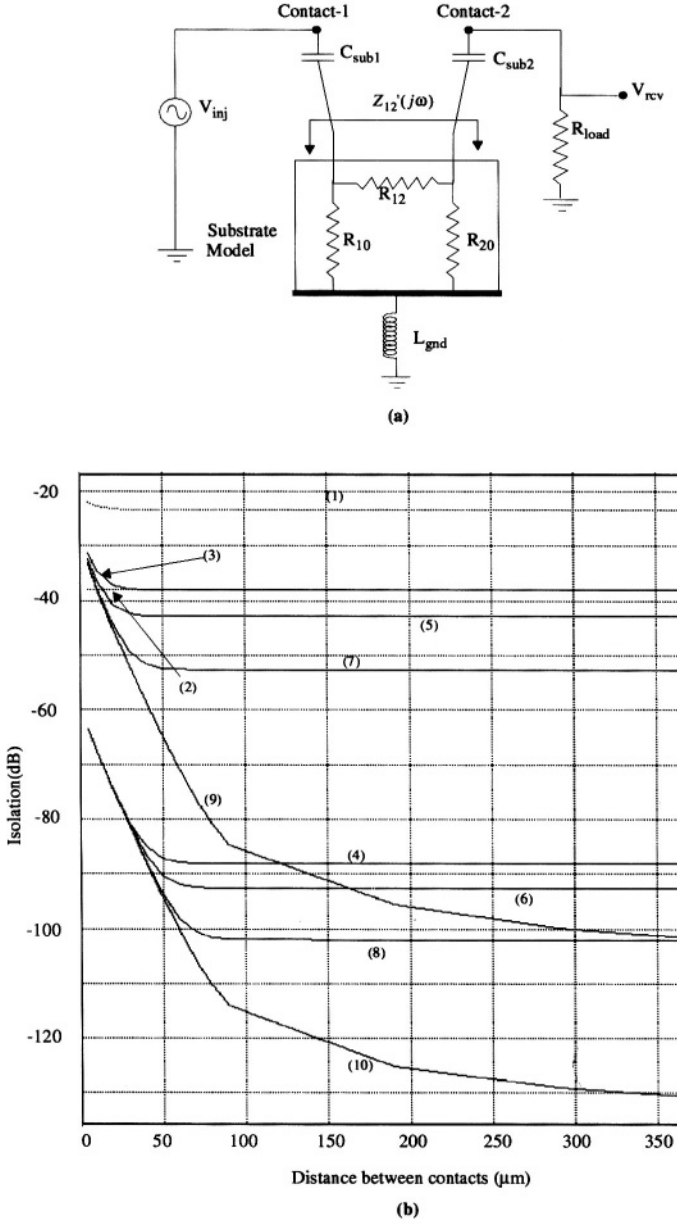


Figure 8.7. Isolation characterization (C_{sub} = 0.8 pF): (a) electrical setup; (b) simulations

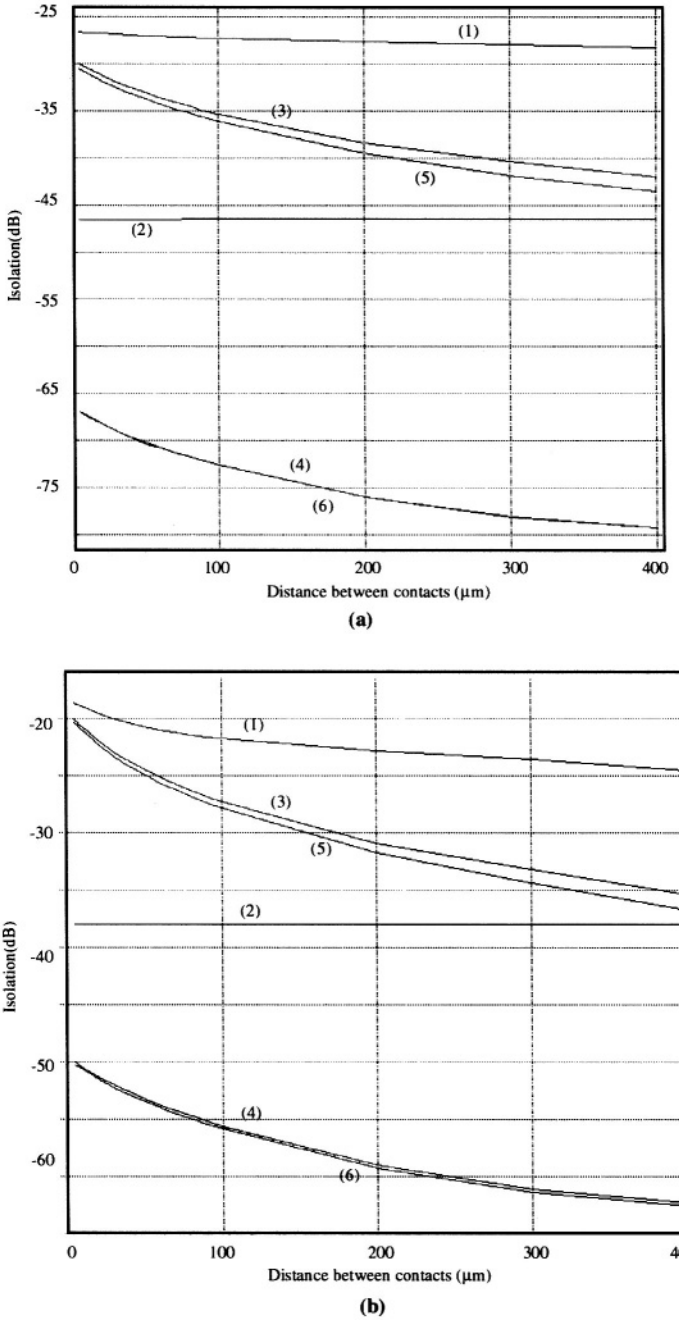


Figure 8.8. Isolation characterization for high-resistivity substrates: (a) C_{sub} = 0.3 pF; (b) C_{sub} = 0.8 pF

pacitance C_{sub} is assumed to be respectively at 0.3pF and 0.8pF in the plots of Figure 8.8(a) and (b). In experiments (1) and (2) the isolation is seen to be weakly dependent on the distance between the contacts at 1GHz. This is so because the value of R_{12} is comparable to the reactance of C_{sub} at this frequency. At 100MHz, the capacitive reactance is much larger than R_{12}. In high-resistive substrates the isolation is seen to be weakly dependent on the value of the backplate inductance, unlike that in the low-resistivity substrates. This is due to the significant surface conduction in these substrates. Experiments (4) and (6) almost coincide in Figures 8.8(a) and (b), which implies that almost all of the coupling takes place through the surface-layers. Since a significant part of the current flows at the surface, substrate taps can be expected to be more efficient in these substrates compared to the low-resistivity substrates.

The effective substrate impedance $Z'_{12}(j\omega)$ in these substrates is a much weaker function of contact area. Thus the isolation does not degrade as rapidly as in the previous case with area. This is especially true for larger areas, when the capacitive reactance of C_{sub} is small, and the isolation between the two contacts is dominated by $Z'_{12}(j\omega)$. Another consequence of the surface conduction is that the isolation continues to improve as the distance between contacts increases, even for finite values of the backplate impedance.

3. IMPROVING ISOLATION USING DIFFERENTIAL CIRCUITS

Differential circuits are often preferred over single-ended circuits in noisy environments. This is the case because the noise appears as a common-mode signal on the differential outputs. The differential noise signal is typically several orders of magnitude smaller than what would be observed in a single-ended implementation of the circuit. The use of differential circuits in applications requiring a high power-supply-rejection-ratio (PSRR) is a typical example. In this section, the use of differential circuits to reduce substrate noise is discussed. As in the previous section, the two types of substrates shown in Figure 8.1 are used to study isolation.

3.1 LOW-RESISTIVITY SUBSTRATES

The layout of the differential structure considered in the experiments is shown in Figure 8.9(c). The differential circuit consists of two $200 \times 10\mu m^2$ contacts placed at the center of the substrate surface. The surface injector is a single $50 \times 50\mu m^2$ contact and can slide in x- and y-direction. The equivalent circuit model is shown in Figure 8.9(a). The simulated isolation data are plotted in Figure 8.9(b). All the curves refer to a low-resistivity substrate with a profile similar to that of Figure 8.1(b), with $\rho_{surface} = 1\Omega\text{cm}$, $\rho_{epitaxy} = 15\Omega\text{cm}$, and $\rho_{bulk} = 10\text{m}\Omega\text{cm}$.

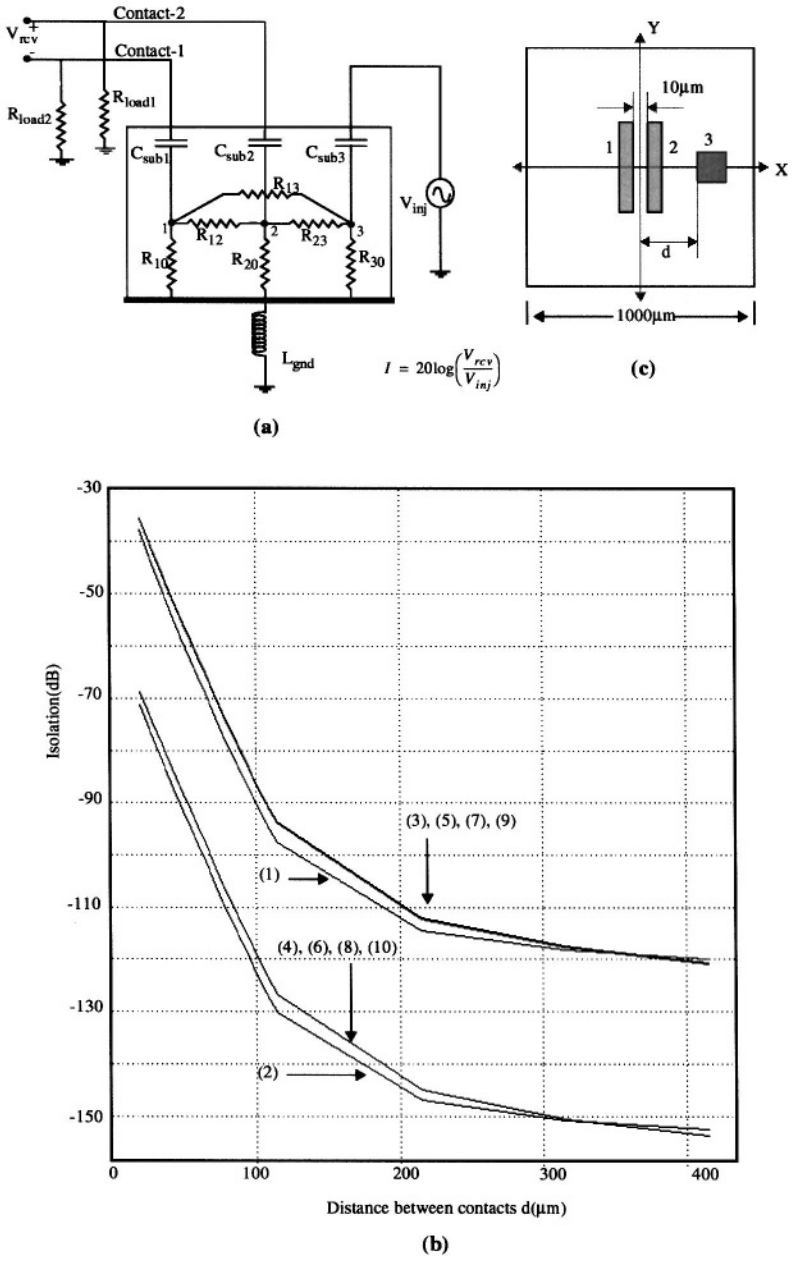


Figure 8.9. Isolation characterization for differential configurations: (a) electrical setup; (b) simulations; (c) geometry

In experiments (1) and (2) the backplate contact is floating, in the other experiments a grounded backplate contact is used, while inductance L_{gnd} assumes values of 5nH in (3)-(4), 3nH in (5)-(6), 1nH in (7)-(8), and 0nH in (9)-(10). The frequency of operation is set to 1GHz in the odd- and 0.1GHz in the even-labeled experiments.

If the injector contact is moved in the y-direction, then the circuit will be perfectly balanced with respect to the substrate, and the differential substrate-noise isolation will be infinite. The asymmetric axis-orientation used in this simulation results in the worst-case differential substrate-noise isolation. The differential receiver contacts in Figure 8.9(c) are assumed to be identical.

The isolation in Figure 8.9(b) falls rapidly with distance between the contacts and is insensitive to the backplate impedance. As the distance between the contacts increases, the direct coupling terms R_{23} and R_{13} increase rapidly in experiments (3)-(10). Thus the coupling takes place almost entirely through the low-resistivity bulk region. The presence of the low-resistivity bulk region close to the surface reduces the fringing fields. Another consequence of the reduction in the fringing effect is that the resistances to the backplate are almost independent of the location of the contact on the surface of the substrate. Two contacts with the same area will have nearly the same resistance to the backplate, regardless of their contact coordinates. Consequently, voltage excursions of the bulk region or the backplate are conveyed to the two contacts of the differential circuit with an equal amplitude. This leads to the excellent differential isolation seen in 8.9(b).

3.2 HIGH-RESISTIVITY SUBSTRATES

The simulation results for high-resistivity substrates are shown in Figure 8.10(a), where the same experiments as in low-resistivity substrate are used for characterization. The isolation improves significantly by using differential circuits. However, the improvement is not as much as in the case of low-resistivity substrates. This is due to the large surface component of the substrate currents. Contact-2 shields contact-1 from the injected noise. Thus, most of the injected noise appears on contact-2 which results in a relatively smaller improvement in isolation.

The simulations considered earlier assumed perfectly matched differential circuits. This situation is idealized, and practical differential circuits always suffer from component mismatch. The effect of component mismatch on differential-circuit substrate isolation is presented in Figure 8.10(b) for low-resistivity substrates. The layout of Figure 8.9(c) is considered but the substrate capacitors in the receiver pair are assumed to be mismatched by five percent. It can be observed from Figure 8.10(b) that a slight mismatch in the differential pair can significantly reduce the isolation. The degradation is especially severe for distant contacts. The isolation in Figure 8.9(b) is very large due to the

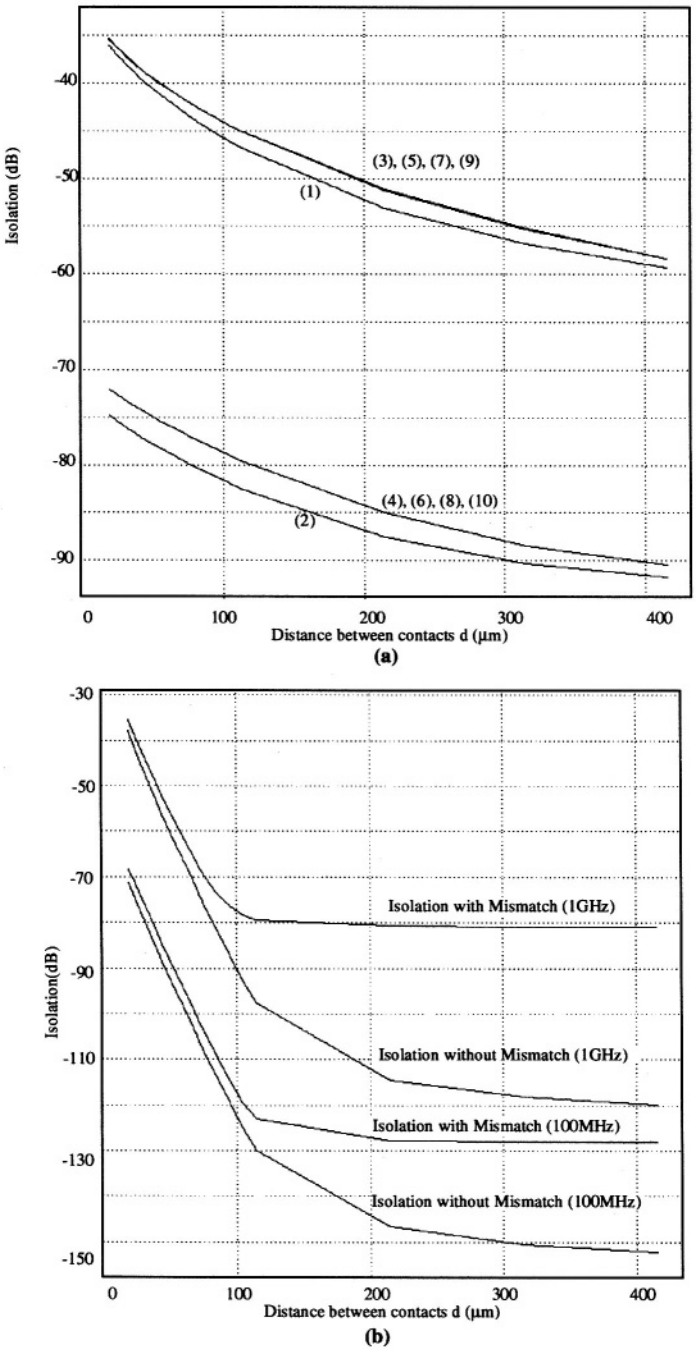


Figure 8.10. (a) Differential isolation on high-resistivity substrates with C_{sub} = 0.8 pF. (b) Effect of mismatch on differential isolation

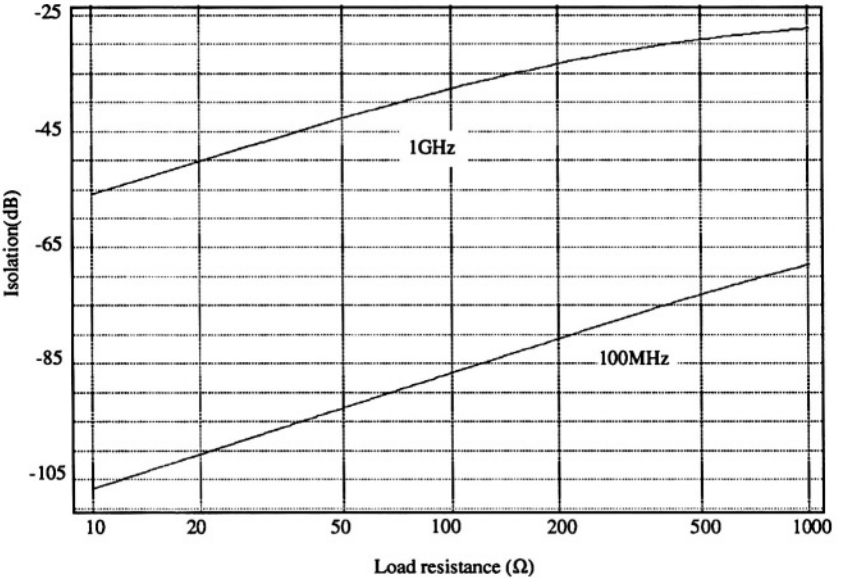


Figure 8.11. Effect of load impedance on isolation

idealized situation assumed there. Mismatches between circuit elements lead to common-mode substrate injection and reception, which make differential circuits exhibit single-ended behavior.

4. EFFECTS OF LOAD IMPEDANCE

In the previous examples the load impedance at the receiver has been assumed to be 50 Ω. The load impedance on-chip may be lower or higher. In this section, an earlier simulation is repeated with different values of the load resistance. The example shown here treats only the low-resistivity substrates, as the observations made here are valid for both types of substrates. Figure 8.11 shows the isolation between contacts 1 and 2 in Figure 8.6(a) for a fixed distance of 90 μm between the contacts. The L_{gnd} is fixed at 3 nH and C_{sub} at 0.8 pF.

Usually the load resistance value is chosen such that the capacitive reactance of the device-to-substrate capacitance is larger than the load in the frequency band of interest[1]. In other words, the pole frequency at the load due to the substrate capacitance is much higher than the highest frequency of interest. In such a case, it is reasonable to assume that the load impedance does not effect the voltage division ratio in the substrate macromodel. Referring to

Figure 8.6(a), we observe that the voltage at the internal substrate node (2*) is independent of the load impedance to the first order. The voltage at the load (V_{rcv}) can be calculated using the voltage division ratio of the load resistance and the substrate capacitance. If the above assumption regarding the load pole frequency is used, then V_{rcv} for different load values can be calculated by simply changing the voltage division ratios between nodes 2* and contact-2 in Figure 8.6(a).

This approximation is valid if the received substrate noise is in the same frequency band as the circuit itself. If the frequency of the substrate noise signal is higher than the receiver circuit's load-C_{sub} pole, then the value of the load changes the degree of coupling. This behavior is seen in Figure 8.11. The pole frequency with the highest load value of 1kΩ is approximately 200MHz. Thus at 100MHz the isolation scales linearly with the load resistance. This is not the case at 1GHz. The pole frequency equals 1GHz for a load resistance of 200Ω. Deviations from a linear increase at 1GHz are seen in Figure 8.11 for load resistances greater than 200Ω.

5. IMPACT OF GUARD RINGS

Guard rings and substrate taps are often used to reduce substrate coupling. In this section, the effectiveness of guard rings in different substrate types will be discussed. General guidelines for effective guard ring layout will be proposed. The layout of a typical guard ring is shown in Figure 8.12(a). The ring is a surface-region heavily doped with the majority-carrier dopant and is intended to form a Faraday shield around any sensitive devices which need to be protected from substrate noise. The ring could also be placed around the noise injectors. Figures 8.12(c) and (d) depict equivalent models of the substrate with and without the guard ring, respectively. The guard ring is effectively a current sink. By placing a current sink between the two contacts, the direct coupling model element R_{12} between contacts 1 and 2 is increased. Thus, with the ring connected to ground, the isolation between the two contacts increases. L_{gr} represents the inductance of the bond-wire and the pin used to ground the ring. L_{gnd} is the backplate contact impedance.

Contact-2 is effectively shielded from contact-1 by the ring, if the presence of the ring makes R_{12} large and the reactances of L_{gnd} and L_{gr} (the associated reactances X_{gnd} and X_{gr} respectively) are small at the frequency of interest. If X_{gnd} is large, then the current flow through the backplate can dominate and if X_{gr} is large, then the dominant current flow will be through the resistors R_{1gr} and R_{2gr}. Thus, care must be taken to keep these inductances small for the guard ring to be useful.

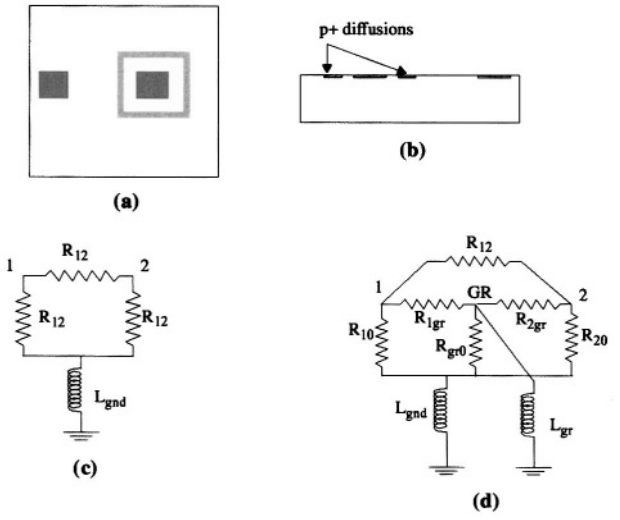


Figure 8.12. Guard rings: (a) layout; (b) cross-section; (c), (d) equivalent models

5.1 LOW-RESISTIVITY SUBSTRATES

The effect of placing a $10\mu m$ wide guard ring between two square contacts of area $2500\mu m^2$ is examined. The layout and cross-section are shown in Figure 8.12(a) and (b), respectively. The equivalent circuit of substrate conduction with and without the presence of the guard ring is shown in Figure 8.12(c) and (d).

The isolation between the two contacts as a function of distance is examined at two frequencies 100MHz and 1GHz. This simulation is similar to the simulations presented earlier in this chapter. C_{sub} has been chosen to be 0.8 pF. L_{gr} is stepped between 0 and 3 nH, while L_{gnd} is varied from 0 to 5 nH. Figure 8.13(a) shows the isolation for the two contacts, with the same set of parameters, with the guard ring included in the figure. On comparing the isolation in these two figures it can be seen that the improvement in isolation for the same separation of the contacts and the same value of L_{gnd} is in the range of 7-10dB. The isolation is a weak function of the guard ring inductance L_{gr}.

Large gains in substrate isolation are achieved only by lowering the backplate inductance. This behavior can be expected in low-resistivity substrates because current flow in these substrates is mostly through the bulk, while the guard ring is an effective current sink only for the surface component of the current. For an ideally grounded guard ring at 1GHz, the isolation seems to be improving

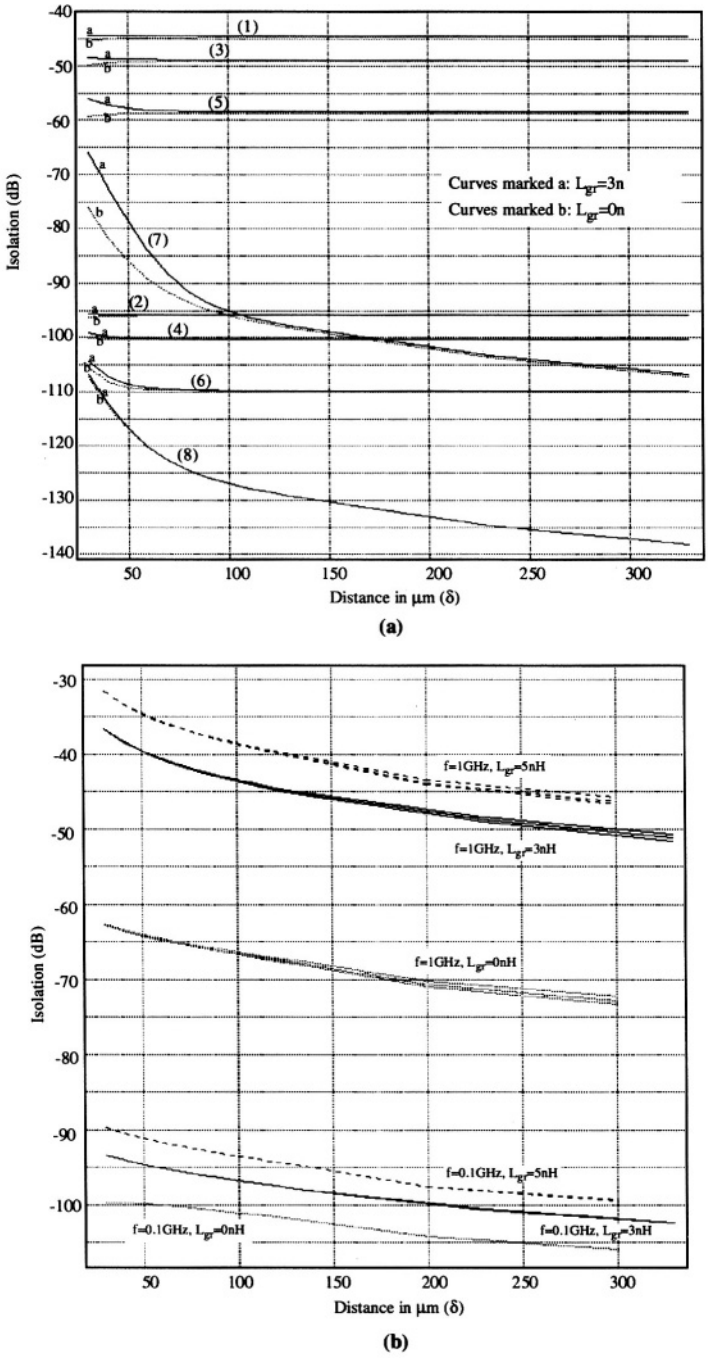


Figure 8.13. Effect of guard rings on isolation: (a) low-resistivity; (b) high-resistivity substrates

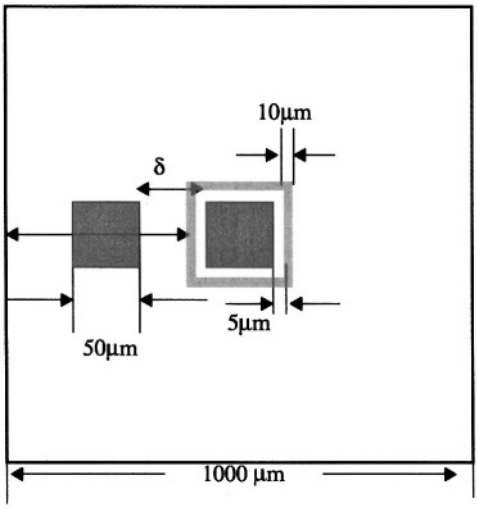


Figure 8.14. Guard rings in high-resistivity substrates

when the contacts come closer together for an L_{gnd} of 3 nH and 5 nH. This behavior is observed because the guard ring acts as a good ground for contacts close to it, especially since at high frequencies the backplate impedance is large and surface currents tend to increase slightly. In low-resistivity substrates, the most effective way of improving isolation is to provide a very good ground contact to the backplate. Surface isolation structures similar to guard rings are not very effective in these substrates.

5.2 HIGH-RESISTIVITY SUBSTRATES

A similar simulation was performed for the case of high-resistivity substrates as was done in the above subsection. The layout is shown in Figure 8.14. The results of the simulation are shown in Figure 8.13(b). If compared to the low-resistivity case, guard rings in high-resistivity substrates are very effective. Comparing Figure 8.13(b) with Figure 8.8(b), we observe that large improvement in isolation is obtained with the ring. It can also be seen that lowering the value of L_{gr} improves the isolation by large numbers. As was discussed earlier, a large fraction of the substrate current flows at the surface in these substrates. Hence the guard rings act as very good current sinks. The following study of guard rings will be restricted to this type of substrates only.

6. GUARD RINGS IN SINGLE-ENDED CIRCUITS

In this section the optimal sizing of guard rings in these substrates, in order to minimize the total noise appearing at the receiver node, is discussed. This

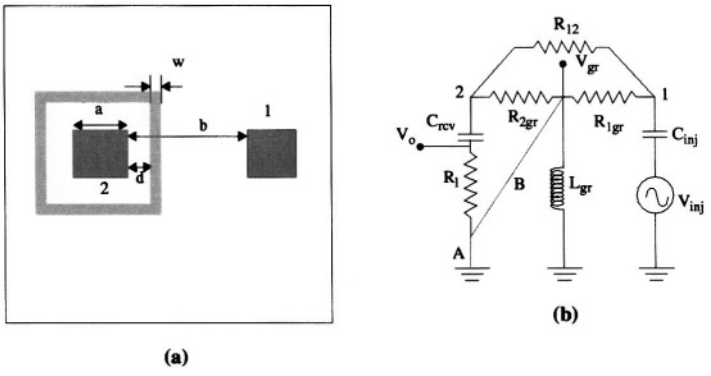


Figure 8.15. Guard ring at the receiver end: (a) layout; (b) equivalent circuit

section deals with single-ended circuits. Differential circuits are considered in the next section.

6.1 GROUNDING SCHEMES

A single guard ring is assumed to be laid out around the receiver contact as shown in Figure 8.15(a). The injector and the receiver are assumed to be capacitively coupled to the substrate. The substrate contacts are represented with 1 and 2, for the injector and the receiver end. The noise-injector is modeled by a sinusoidal AC source V_{inj}, and the noise received is depicted by V_o. We assume at first that the guard ring is connected to the ground through an independent bond-wire, modeled by an inductance L_{gr} as shown in Figure 8.16(a). The voltage V_o could be referred to the internal ground (V_o-V_{ig}) or to the global ground if the signal is measured externally. With regards to the isolation provided by the ring in this case, the reference against which V_o is measured is not important, at least in a behavioral sense. If the voltage V_o is referred to the global ground, then the impedance between the internal and global grounds, is merely an addition to R_l. On the other hand, if V_o is referred to the internal ground, then the signal across R_l will be directly proportional to the value of V_o referred to the global ground. Only the former case, when the signal is measured with reference to the global ground, is considered here.

When the ring is connected to the internal ground, as in Figure 8.16(b), it becomes necessary to identify the reference potential. The variation of isolation as a function of guard ring width and the bond-wire ground inductance L_{gr} is considered in Figures 8.17(a) and (b), and in Figure 8.18. The dimensions a, b, and d, shown in Figure 8.15(a), are kept constant and the width of the

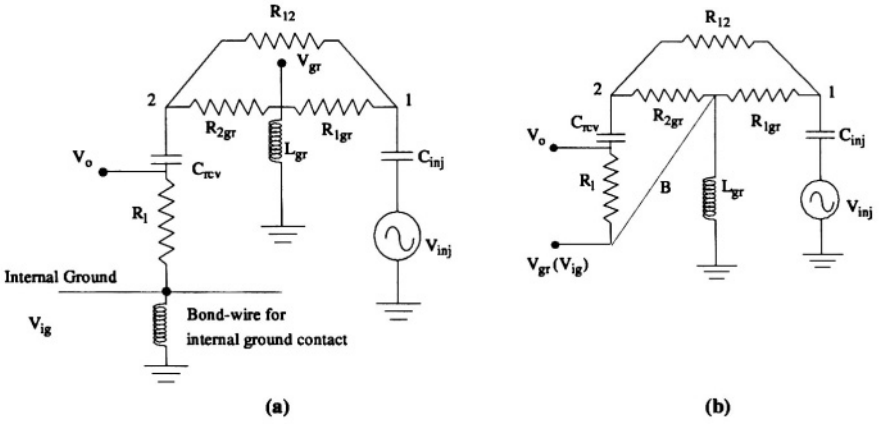


Figure 8.16. Guard ring configurations: (a) with an external ground connection; (b) with an internal ground connection

ring w is varied. The setup of Figure 8.16(a) is used to produce the results shown in Figure 8.17(a), while the setup of Figure 8.16(b) is used to produce the results shown in Figures 8.17(b) and 8.18. In Figure 8.17, the isolation is defined as the ratio of V_o and V_{inj} in dB. The isolation in Figure 8.18 is defined as $(V_o - V_{gr})/V_{inj}$ in dB. We observe from Figure 8.17, that increasing the thickness of the guard ring from 2 to 34μm, provides a small improvement in the isolation but only for the case $L_{gr} = 0$. For other values of L_{gr}, the isolation actually worsens. Thus, in both cases, it is advisable to use thin guard rings.

Further examination of Figure 8.17 reveals that the isolation in Figure 8.17(a) is significantly better, which leads us to the conclusion that if the signal V_o is to be referenced to an external ground, the guard ring must be connected to ground through an independent bond-wire. It can further be observed that the isolation in Figure 8.17(b), for large values of the bond-wire inductance, is worse than that without the guard ring. Thus, the guard ring must be connected to an internal ground, only if the signal is measured with respect to the internal ground. It can be seen from Figure 8.18 that the isolation provided by an internally grounded guard ring is excellent, if the signal is measured with reference to the internal ground. Further, the isolation is practically independent of the value of the bond-wire inductance. C_{rcv} and C_{inj} have been assumed to be 0.8 pF in all the simulations.

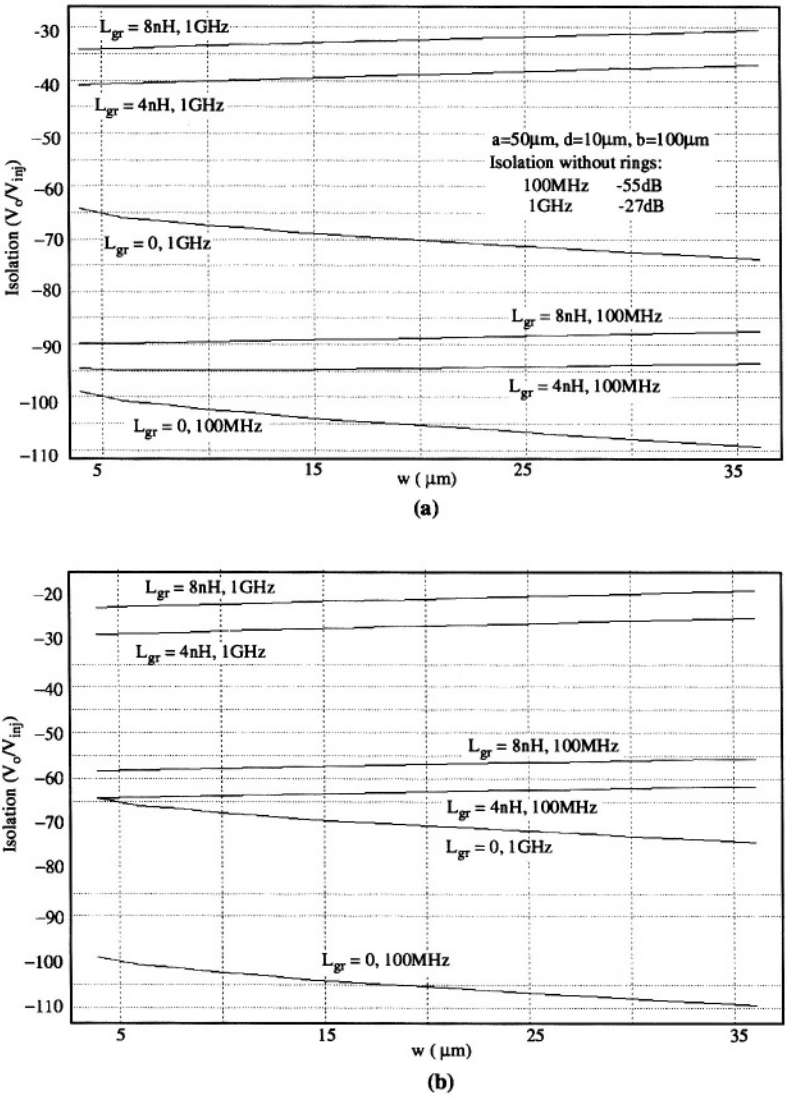


Figure 8.17. Guard ring simulations: (a) with an external ground connection; (b) with an internal ground connection

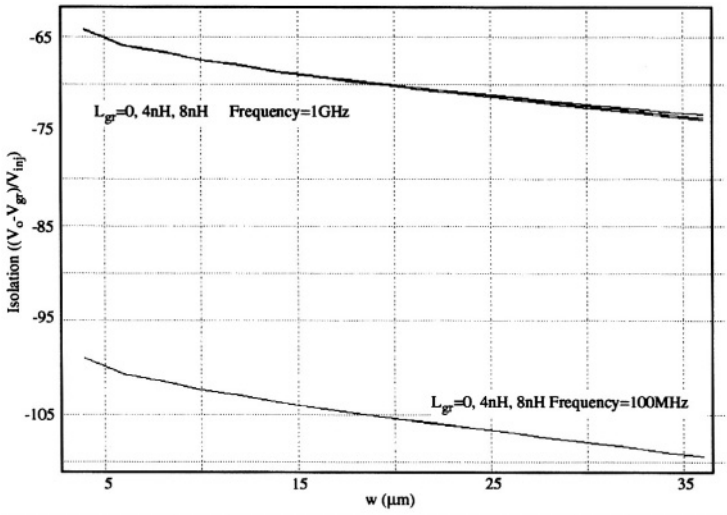


Figure 8.18. Guard ring simulations with internal ground connection

6.2 OPTIMIZATION OF GUARD RING WIDTHS

Guard rings are often connected to ground by the use of a separate ground-pin as in Figure 8.16(a). The associated bond-wire can pickup noise from adjoining bond-wires, and inject noise into the circuit. The optimization of the guard ring size with a noisy ground connection is treated in this section. In the case of Figure 8.16(b), the influence of noise on the guard ring bond-wire can be minimized by referring the receiver voltage to the internal ground.

For simplicity but without loss of generality in what follows, a substrate model with a floating backplate will be considered. In Figure 8.19(a), the lengths a and d are kept constant. The width of the ring is varied and the optimum size of the ring is determined. Contact-1 is the injector and contact-2 is the receiver. The injector voltage is shown as V_{inj}. The bond-wire used to ground the guard ring can also pickup noise from adjacent pins due to bond-wire mutual inductances. This noise is modeled by the generator V_{inj1}. The problem at hand is to reduce the influence of V_{inj} on V_2, while simultaneously keeping the effect of V_{inj1} low[2].

We assume, for ease of analysis, that the impedances Z_{inj} and Z_{rcv} be small. These impedances model the reactances of the device-to-substrate capacitors. Impedance Z_0 is the loading of the receiving node. Hence, node

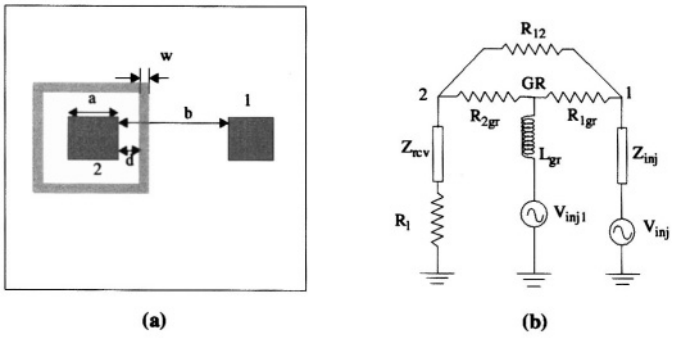


Figure 8.19. Guard ring: (a) layout; (b) substrate model

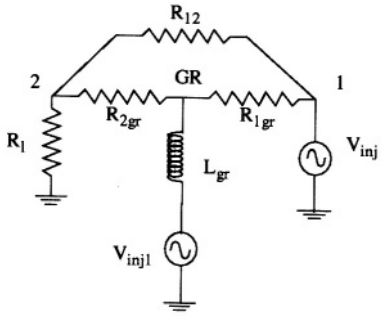


Figure 8.20. Modeling bond-wire noise in a guard ring

1 in Figure 8.19(b) is voltage driven and node 2 sees the resistance R_l. This approximation will be valid at very high frequencies. The isolation between nodes 1 and 2 is then determined by the reduced model shown in Figure 8.20. Under the approximations stated earlier and provided that R_l is small, the total output voltage at node 2 can be shown to be

$$V_2 = R_1 \left\{ \left(\frac{1}{R_{12}} + \frac{j\omega L}{R_{1gr}R_{2gr} + j\omega L(R_{1gr}+R_{2gr})} \right) V_{inj} + \left(\frac{R_{1gr}}{R_{1gr}R_{2gr} + j\omega L(R_{1gr}+R_{2gr})} \right) V_{inj1} \right\}. \tag{8.2}$$

Let us first consider the case when V_{inj1} is zero. The effect of increasing width w is that R_{12} increases while R_{1gr} and R_{2gr} decrease. At low frequencies the isolation between the contacts is determined by R_{12} and therefore for low-frequency operation a wide ring is beneficial. At high frequencies the second term in the bracket starts becoming significant. In fact, at very high frequencies, the isolation is governed by the sum $R_{1gr} + R_{2gr}$. If this sum is small at these frequencies, the isolation can actually be worse than that achievable without the ring. Further, as R_{1gr} and R_{2gr} are reduced, the frequency at which the ring begins to lose its advantages is also lowered for $\omega \approx (R_{1gr}R_{2gr})/L(R_{1gr} + R_{2gr})$. It must be mentioned, however, that this pole frequency is usually very large for typical values of L_{gr}. If V_{inj1} is nonzero, we observe a potentially more serious problem with a very wide guard ring. At low frequencies the noise term V_{inj1} appears at the output scaled by the ratio R_l/R_{2gr}. This ratio can be significant if R_{2gr} is too small[3].

An area-efficient way to increase R_{12} while keeping w small, is to make d small in Figure 8.19(a). Making d small also decreases R_{2gr}, which leads to a higher dependence of V_2 on V_{inj1}. If V_{inj1} is significant, then w and d must both be made large. Another option is to place the ring around the injector contact. It can be seen from (8.2) that the dependence of V_2 on V_{inj} is symmetric in R_{1gr} and R_{2gr}. Thus, increasing R_{2gr} and reducing R_{1gr} simultaneously has no effect on V_2 if V_{inj} is the only noise source. However, the effect of V_{inj1} on V_2 is significantly reduced by using this arrangement.

The alternate scheme of placing the ring around the noise-injectors may not always be preferable to the scheme shown in Figure 8.19(a). The principal cause of the appearance of the noise V_{inj1} is the mutual inductance which is present between the pins and bond-wires in a package. A usual design practice in mixed analog-digital circuits is to place the analog and digital portions of the circuit in different parts of the substrate. The pads are also laid out such that the analog and digital bond-wires are placed in different sections of the package. If layout constraints force the bond-wire used to ground the ring to be located in proximity to the digital bond-wires, then it is possible that the magnitude of V_{inj1} may increase significantly compared to the other case. The ideal scheme for isolating the sensitive circuits from the noisy circuits is to place the ring around the noisy circuits and lay out the bond-pads such that the guard ring bond-wire is placed along with the analog-section bond-wires.

The above discussion is not definitive but is meant to shed light on some of the trade-offs involved in the design of optimal guard rings. In the rest of the section results from simulations are presented. The noise appearing at contact-2 of Figure 8.19(b) for a 1 volt swing on contact-1 is shown in Figures 8.21 and 8.22. The dimensions of the contacts and the ring and the substrate data are presented in the figures. The width of the ring is swept from 4 μm to 32 μm in increments of 2μm and the output signal is measured. C_{sub}

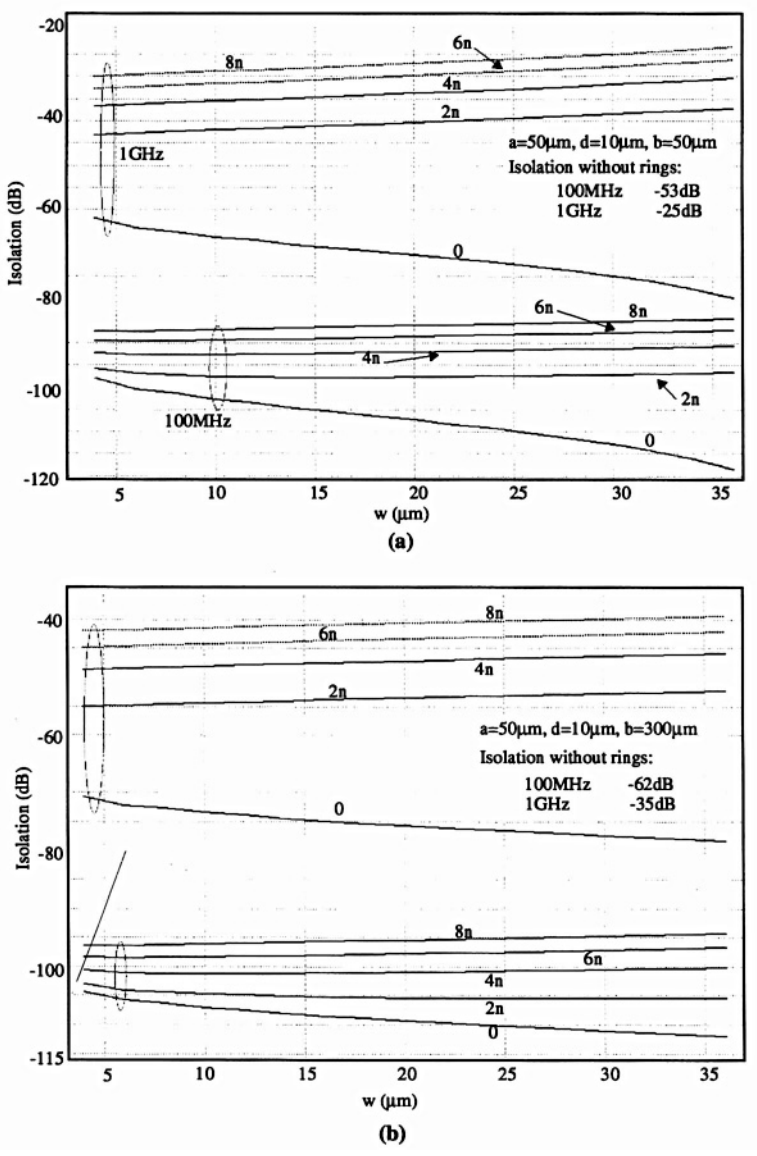


Figure 8.21. Guard ring isolation on a high-resistivity substrate ($\rho_{epitaxy} = 0.1\Omega$cm, $\rho_{bulk} = 20\Omega$cm, and $\rho_{epoxy} = 0$): (a) $d = 50\mu$m; (b) $d = 300\mu$m

is assumed to be 0.8pF in all the simulations. Figures 8.21(a) and (b) depict the substrate noise coupling in substrates without backplate contacts. 8.22(a) and (b) depict the case with ideally grounded backplates. In each figure the isolation at two frequencies, 100MHz and 1GHz is shown. The guard ring

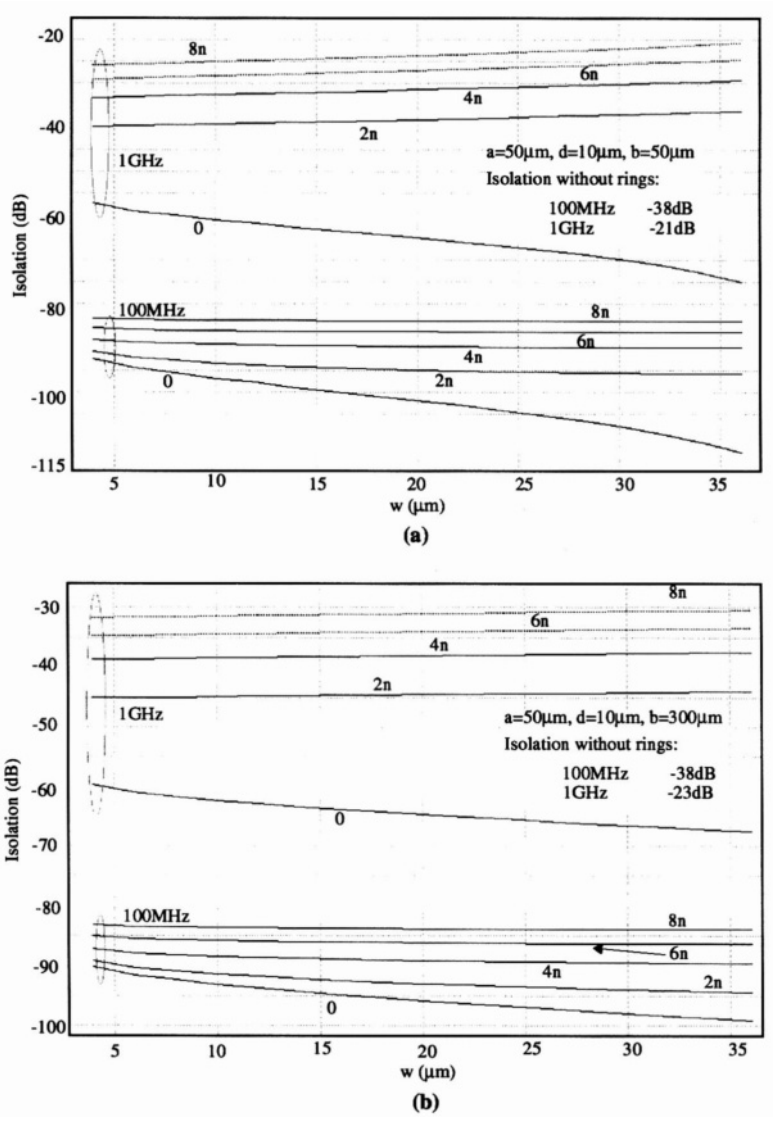


Figure 8.22. **Guard ring isolation on a high-resistivity substrate with floating backplate contact ($\rho_{epitaxy} = 0.1\Omega$cm, $\rho_{bulk} = 20\Omega$cm, and $\rho_{epoxy} = \infty$): (a) $d = 50\mu$m; (b) $d = 300\mu$m**

bond-wire inductance is stepped from 0 to 8nH in steps of 2nH. The isolation without the guard rings at the two frequencies is mentioned in each graph.

A common conclusion which can be drawn from each of the graphs is that at high frequencies **($\approx$ 1GHz),** thin guard rings provide better isolation than

wide rings. The evidence of the existence of an optimum guard ring width is found in the isolation curves at 100MHz in all the curves for different values of the inductance to ground. In Figure 8.21(a), for example, the optimal guard ring width for a 2nH ground path inductance is 15μm. The minimum is shallow. Thus, the best rule for sizing guard rings is to use minimum width rings and ensure a very good ground connection to the ring. The presence of the ideally grounded backplate improves the isolation by 5-10dB compared to the case without the backplate. This suggests that the primary path to ground is provided by the surface ground rings and not by the backplates in high-resistivity substrates.

It was mentioned earlier in this section that it is possible to place a guard ring around either the receiver or the injector contact. The dependence of the received noise on the noise terms V_{inj} and V_{inj1} is shown in Figure 8.23(a) and (b) respectively, with the ring placed around the injector and the receiver in each figure. The layout parameters and the substrate parameters used are shown in the figures. The bond-wire inductance is assumed to be 4nH. The upper curve in Figure 8.23(a) is the isolation from V_{inj} without the guard ring. It can be seen that the noise isolation provided by the ring, from the substrate injector is nearly the same, regardless of whether the ring is placed around the injector or the receiver. The isolation from V_{inj1} (Figure 8.23(b)) is better by approximately 20dB in the case when the ring is placed around the injector rather than around the receiver.

As a specific example, consider the case when the bond-wire adjacent to the guard ring bond-wire carries a current of 1mA at 1GHz and has a mutual inductance of 1nH with the guard ring bond-wire. V_{inj1} will thus be 6.3mV. If the ring is laid out around the receiver, then the voltage appearing on the receiver due to V_{inj1} will be 1.41mV. If the ring is laid around the injector, the noise at the receiver due to V_{inj1} will be 0.14mV. These voltages will scale linearly with the magnitude of the current in the adjacent bond-wire. For a 1V swing at 1GHz at the injector, the noise appearing at the receiver is 5.623mV. If the guard ring is placed around the injector, then the noise induced by V_{inj1} will exceed that caused by V_{inj} for current values exceeding 35mA. If the ring is placed around the receiver, the noise caused by V_{inj1} will dominate for current amplitudes greater than 3.5mA[4].

7. GUARD RINGS IN DIFFERENTIAL CIRCUITS

It was mentioned earlier that the improvement in isolation in differential circuits in high-resistivity substrates was not as significant as in low-resistivity substrates. The reason for the excellent differential isolation in low-resistivity substrates was found to be that the heavily-doped bulk, which lies in close proximity to the surface, acts as a good ground-plane. As a result, noise signals coupled through the substrate appear as common mode signals at the differential

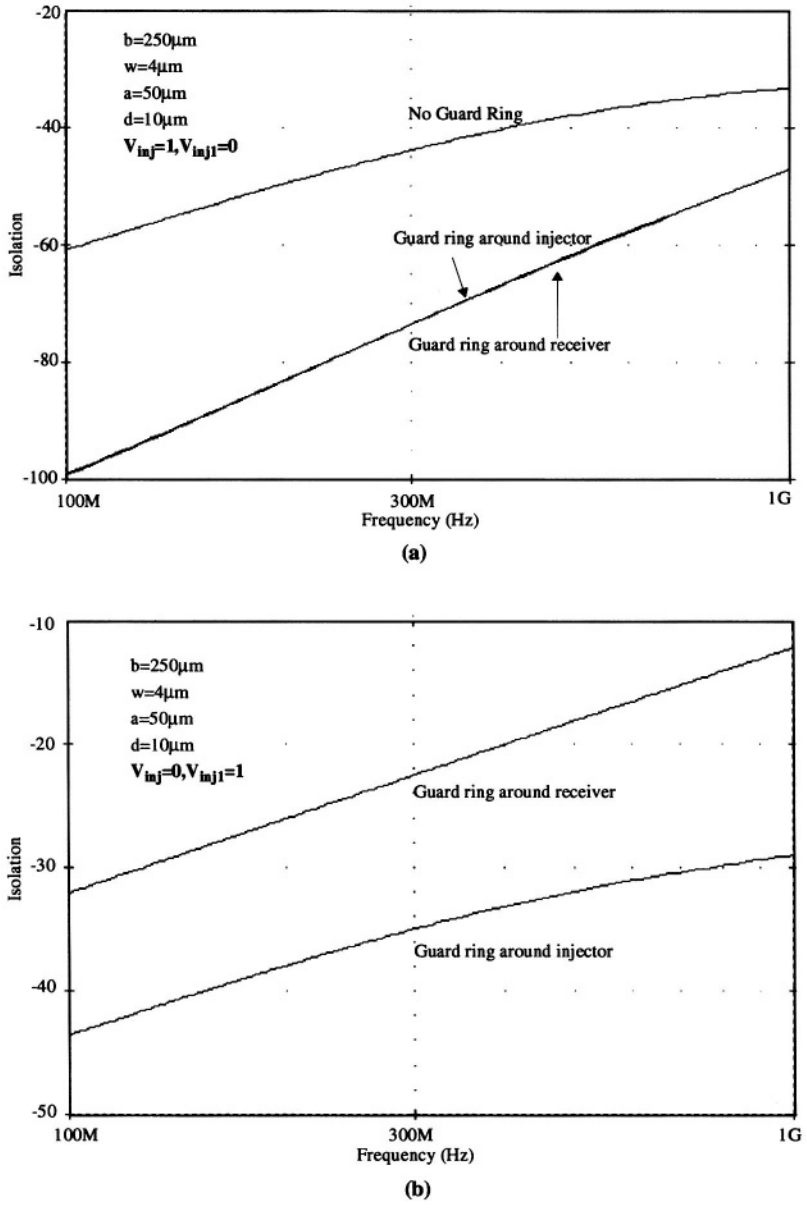


Figure 8.23. (a) Isolation with substrate injection only. (b) Isolation with guard ring bond-wire noise only

outputs. Guard rings in high-resistivity substrates also act as ground-planes in close proximity to the devices. Hence it can be expected that differential isolation in high-resistivity substrates will be improved in a similar fashion

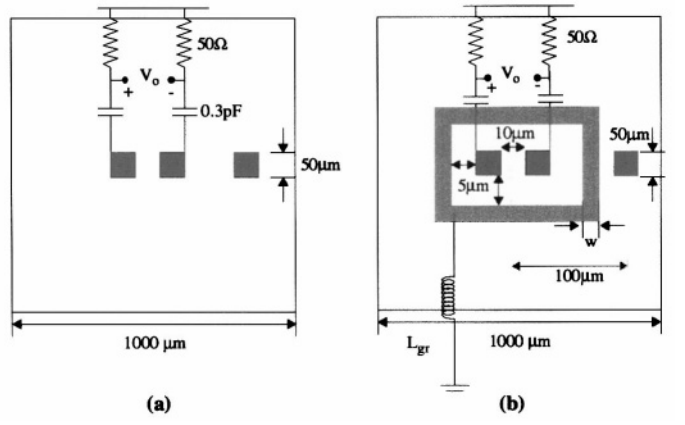


Figure 8.24. Differential signal simulation: (a) with guard ring; (b) without guard ring

by the use of guard rings. The effect of guard rings in differential circuits is studied in this section. Differential receivers with single-ended injectors are considered here.

Let us consider the differential substrate contacts with and without a surrounding guard ring shown in Figure 8.24. For the purpose of simulation, the guard ring shown in Figure 8.24(b) is connected to ground through the bond-wire inductance L_{gr}, which is stepped over three values, 0, 4 and 8nH. The width of the ring is stepped from 10μm to 48μm in steps of 2μm. The results of the simulation, that is the differential and common-mode isolation values, are shown in Figure 8.25. It can be observed from the figure that the differential-mode isolation improves slightly by the use of thick guard rings. Unlike the single-ended case in the previous section, the isolation is independent of the value of the bond-wire inductance. Thus, the isolation improves by a significant amount and is independent of the impedance to ground. Common-mode isolation, however, does worsen with increasing ring thickness for nonzero values of the bond-wire inductance.

Thus in differential circuits, as in the single-ended case, it is advisable to use thin guard rings. The relative independence of the differential-mode isolation on the value of L_{gr} is similar to that seen earlier in low-resistivity substrates. In both cases, the presence of an equipotential region in proximity to the circuits makes the noise a common-mode signal and is rejected. The presence of circuit mismatches will degrade isolation.

Circuits with differential injectors and receivers will show similar improvements in isolation in the presence of a guard ring around the receiver or the injector. These circuits have not been discussed separately in this section since

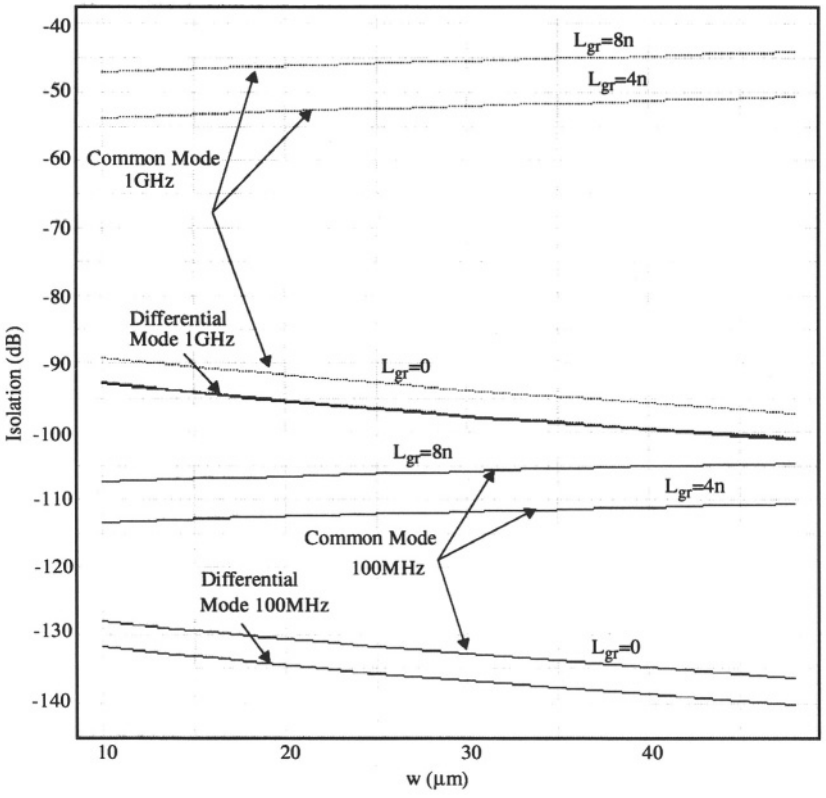


Figure 8.25. Isolation in a differential circuit with a guard ring

the mechanism which leads to the improvement in isolation is the same as that discussed above.

8. DUAL GUARD RINGS

The effect of one guard ring placed around the injector or the receiver has been discussed in the previous sections. It is also possible to place guard rings around both the injector and receiver contacts. The isolation between two single-ended contacts, with a guard ring placed around each contact is examined in this section. The layout is shown in Figure 8.26. The results of the simulation are shown in Table 8.1. The isolation at 100MHz to 1GHz, without guard rings, with a guard ring around the receiver contact only and with two guard rings, is shown in the table. The guard rings, shown in Figure 8.26, can be connected to ground separately or together. In the latter case, however, the two guard rings appear to act as one large ring, and all the conclusions of

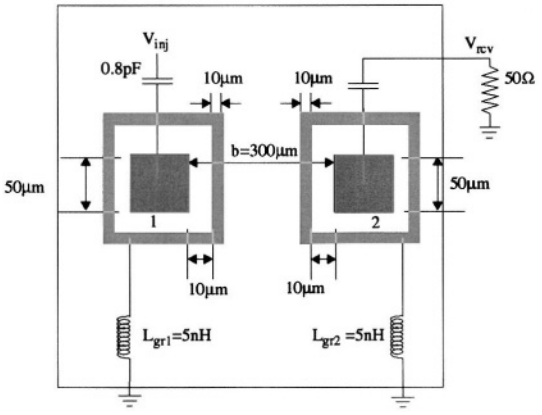


Figure 8.26. Dual guard ring layout

Setup	Frequency	Isolation
No guard ring	100MHz	-62dB
"	1GHz	-34dB
Guard ring around receiver only	100MHz	-99dB
"	1GHz	-45dB
Two guard rings	100MHz	-130dB
"	1GHz	-57dB

Table 8.1. The effect of guard rings on isolation

Section 6. in this chapter are applicable. Hence the guard rings are assumed to be grounded separately.

The guard rings are assumed to be connected to ground through independent bond-wires. The bond-wires are modeled by 5nH inductors. The injector and receiver contacts are squares of 50μm sides. The substrate capacitance is assumed to be 0.8pF. The improvement in isolation by placing two rings is seen to be significant, especially at low frequencies. Using two guard rings requires two package pins which may not always be possible. If, however, two pins are available, it is advisable to use two guard rings.

9. BURIED SUBSTRATE SHIELDS

It is possible to completely enclose a receiver or injector node for isolation. We refer to these isolation structures as buried shields. There are three primary types of structures that may be used. The first is a conductive or a Faraday

shield. This type of shield is an extension of the majority carrier guard ring and operates by providing an alternate low-impedance path to ground for substrate currents. The second uses dielectric isolation to isolate the nodes from the substrate. This scheme physically increases the impedance between the injector and the receiver by increasing the resistivity of the substrate that surrounds either of the two nodes. The third class uses a combination of both of the above types of isolation.

An example of the first class of structures is a p+-on-p type of shield as shown in Figure 8.27(a). A deep p+ implant is used below the devices to be shielded and substrate contacts are used to contact the implant. This structure is very effective in nearly ideally isolating the enclosed circuit with an ideal ground connection on the shield. Unfortunately, this structure is also very sensitive to the impedance of the ground path. A high ground path impedance causes the shield to float electrically. In a package, with sufficiently large bond-wire inductance, the effectiveness of this scheme can deteriorate considerably at high frequencies.

The second type of isolation scheme is implemented in Silicon-on-Insulator substrates (SOI). In these substrates bulk silicon is isolated from a thin active surface silicon layer, by a buried oxide layer. Devices are built on the thin surface silicon and isolated by pockets of trench oxide that can be built selectively to enclose them. This scheme, shown in Figure 8.27(b), provides very good isolation, but it adds to processing costs, since it requires the use of specialized silicon substrates. A key difference between this scheme and the other two is that SOI isolation typically uses an unpatterned buried oxide region, while the scheme discussed next is usually patterned, that is, the buried region exists only around the injector or the receiver.

The third type of isolation uses a buried minority-type enclosure around the device to be isolated. For example, the injector or the receiver nodes may be enclosed by a buried n+ well in p-substrates. The well is connected to the highest potential in the circuit as shown in Figure 8.27(c). Devices are built in the shallow p-region within the buried well. A depletion layer exists on both sides of the n-well. This provides dielectric isolation between the active devices and the bulk region. Since the n-type material is connected to a low-impedance supply pin, the well also forms a Faraday shield around the active device. The advantages of this scheme have been presented in detail in [81]. A disadvantage is that it is difficult to apply it around all types of devices. For example, while NMOS devices can be isolated by using this scheme, it is difficult to build a buried n+ region underneath the n-well in which PMOS devices are built. Similarly this scheme is difficult to implement with vertical npn devices, that use an n+ buried collector. This is because the n-well for PMOS devices and the buried collector of npn devices can extend

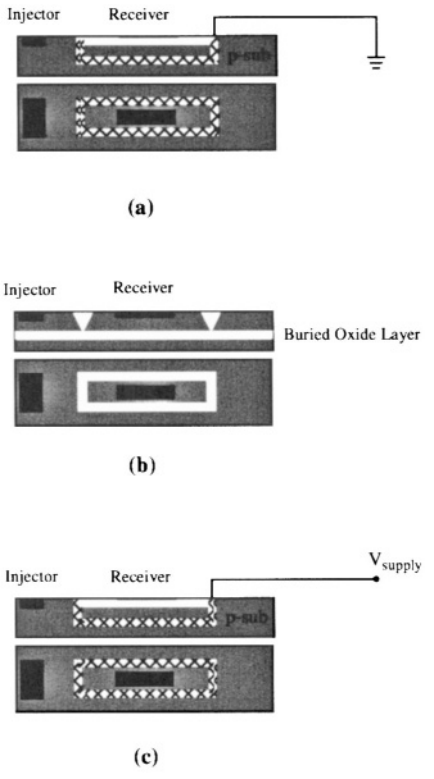


Figure 8.27. Cross-section and top view of various injector/receiver systems. Isolation with: (a) p+ buried layer; (b) buried oxide layer; (c) n+ buried layer

to a significant depth from the surface (up to 2μm in some processes). Adding a deeper implant is expensive and technologically challenging.

An approximate model for the isolation scheme of Figure 8.27(a) is shown in Figure 8.28. We assume that the p+ material has sufficiently low resistivity relative to the substrate material, so that the shield can be represented by a single node. If the resistivity of the shield is zero, coupling between the injector and the receiver can occur only through the shield node (Z_{12} is infinite in Figure 8.28). Thus, if the shield is connected to ground with an ideal connection, isolation between the two nodes is infinite. Nonzero resistivity of the shield material degrades the isolation since it effectively allows coupling between the injector and the receiver, even with a perfect grounding of the shield (finite Z_{12} in Figure 8.28). The ground-path impedance (Z_{shex}) also

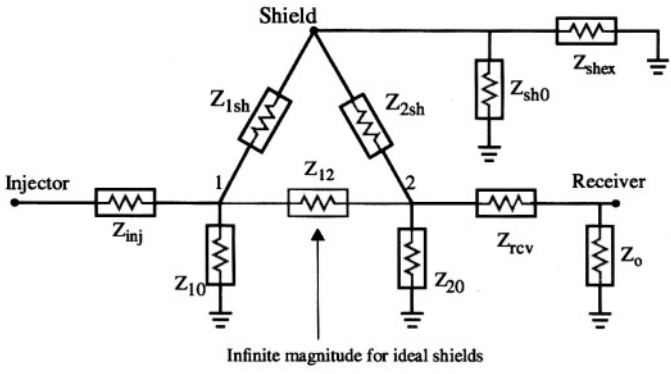


Figure 8.28. Physical model for p+-in-p shield

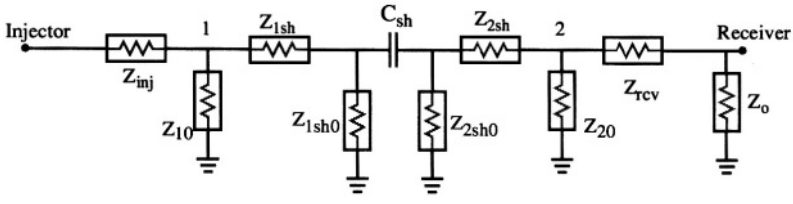


Figure 8.29. Physical model for oxide isolation

has a significant impact on isolation, which varies in inverse proportion to its magnitude. In Figures 8.28, 8.29, and 8.30 Z_{10}, Z_{20}, Z_{sh0}, represent impedances between taps, shields and ground. We assume that substrate taps are present inside and outside the shield. Elements Z_{1sh} and Z_{2sh} model connections between nodes 1, 2 and the shield.

The model for the dielectric isolation scheme is shown in Figure 8.29. This model is similar to the two-node model of Figure 3.3, except for the series capacitance in the path between nodes 1 and 2 (C_{sh}). Due to the capacitive nature of the impedance between the two nodes, this scheme provides excellent isolation at low frequencies, that progressively degrades with increasing frequency. Isolation can be improved by decreasing C_{sh}, which can be achieved by increasing the thickness of the isolation oxide. In the figure, Z_{1sh0} and Z_{2sh0} model the impedance to ground of the outer and inner surfaces of the oxide shield[5].

The model for the third scheme is shown in Figure 8.30. This scheme has

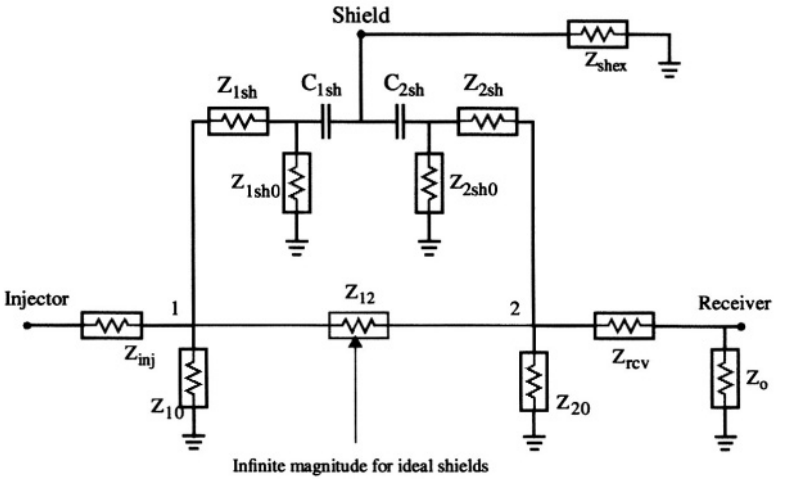


Figure 8.30. Physical model for p+-in-p shield

the advantage that isolation is achieved by two mechanisms: capacitively, by increasing the impedance between injector and receiver nodes as in Figure 8.29, and by providing an alternate low-impedance path to ground as shown in Figure 8.28. In Figure 8.30, C_{1sh} and C_{2sh} are junction capacitors formed between the two surfaces of the shield and the p-type silicon. These surfaces are represented by single nodes. Z_{1sh0} and Z_{2sh0} model the impedance to ground of the outer and inner surfaces. Z_{shex} is the impedance of supply connection, which is ideally zero. In this case too, we assume that the shield material has sufficiently low resistivity that it can be represented by a single node.

This scheme is advantageous at high-frequencies, because the isolation can be better than either of the above cases since the isolation is enhanced by two simultaneous effects. It should be noted that if C_{1sh} and C_{2sh} are removed from the model, then the model is identical to that shown in Figure 8.28. On the other hand, if Z_{shex} is set to infinity, i.e. the supply connection is eliminated, then the model is identical to that shown in Figure 8.29. This scheme is therefore much less sensitive to non-idealities in the ground-path impedance, compared to the p+-in-p shield. Similarly the sensitivity to the value of the series capacitance is smaller, compared to the scheme employing dielectric isolation. A comparison between the SOI and the n+ buried shield is presented in [81].

It is interesting to study the isolation for two limiting conditions of the conductivity of the n-type buried shield. For very high conductivity, the isolation

between the injector and the receiver is achieved by a T-type attenuator consisting of C_{1sh}, C_{2sh} and Z_{shex}. This scheme provides almost ideal isolation at low frequencies, degrading rapidly with frequency, and asymptotically approaching the isolation provided by C_{1sh} and C_{2sh} in the absence of Z_{shex}. In the other extreme case, when the conductivity of the shield is sufficiently small (low doping of the n-region), the resistance of the shield prevents any effective connection to ground. In this situation, isolation is achieved through a series combination of C_{1sh} and C_{2sh}. Lower doping of the n-region reduces the values of these capacitors. Thus, the isolation achieved at higher frequencies can be greater.

A designer may need to choose whether to place the shield around the injector or the receiver. Usually, with a dedicated ground connection, the shield is more effective if placed around the circuit with smaller area, and hence smaller C_{1sh} and C_{2sh}. This ensures a larger reactance of the capacitors relative to the impedance of the ground connection. In many mixed-signal ICs, the noise generating sections of the IC are digital circuits that can have a much larger area than the low-noise analog sections. In these cases, the shield is more effective if it is placed around the smaller analog sections.

Notes

1 The impedance of the load is the effective value of impedance seen looking into the device from the substrate. This definition of load includes the effective value of the load, including the effect of any feedback which may be present around the circuit.

2 Not placing the ring avoids the injection of V_{inj1} altogether. In some applications, where V_{inj1} is large, it is possible that adding the ring actually increases the noise.

3 It is assumed that R_l is always smaller than all the other resistors.

4 The noise without the guard ring can be shown to be 22mV. Therefore it is advantageous to use a guard ring in this example.

5 This representation is approximate due to the distributed nature of the problem.

This page intentionally left blank

Chapter 9

CONCLUSION

In today's push for extremely short design cycles and high system complexities, the clear trend is total integration. Electronic system integration, especially in deep submicron technologies, is becoming a great challenge for the emergence of previously unimportant parasitics and second order effects.

One such challenge is the understanding and proper modeling of the parasitic interactions on chip through its substrate. Modeling this interaction is a difficult task due to the heterogeneity in functionality and operating modes of each component, as well as the way in which substrate conducts spurious signals.

The substrate noise phenomenon is one of the major hurdles in the design of mixed analog-digital circuits today. Not surprisingly, the problem has become a significant issue in new high-speed digital and radio-frequency circuits already.

The study of substrate noise and the design of large systems-on-chip in the presence of noise has been the focus of this book. We have introduced and discussed several noise analysis techniques, as well as acceleration methods which have been used within optimization loops. Experimental verification of our models involving DC and radio-frequency circuits was provided and used to help designers select and optimally size protective structures in the presence of internal and external noise sources.

Guidelines for circuit-level design of highly critical devices have also been developed in view of the results obtained in our study. The guidelines have been developed always taking into account the effects of packaging, which often has an impact on the magnitude of substrate coupling. The guidelines were verified with current technologies and are intended to help designers make both architectural and implementative decisions during the design process.

This page intentionally left blank

Appendix A
Boundary Element Method Derivations

1. NONZERO DEPTH CONTACT CALCULATION

In order to consider contacts characterized by a nonzero depth c, equation (3.10) is re-written as follows [15]

$$p_{ij} = \frac{\bar{\Phi}_i}{Q_j} = \frac{1}{V_i V_j} \int_{-c_4}^{0} \int_{b_3}^{b_4} \int_{a_3}^{a_4} \int_{-c_2}^{0} \int_{b_1}^{b_2} \int_{a_1}^{a_2} G(x, y, z; x', y', z') dx\, dy\, dz\, dx' dy' dz' , \tag{A.1}$$

where the Green's function is computed as described in Section 4. and physical dimensions a, b, c of the contacts are illustrated in Figure A.1. After the

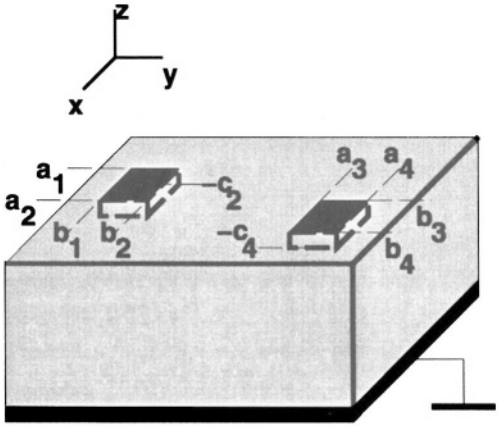


Figure A.1. Nonzero depth contacts and dimensions

appropriate manipulations the term p_{ij} is derived as

$$p_{ij} = \frac{1}{c_2 c_4 ab\epsilon_N \beta_N}\left(-\beta_N\left(\frac{c_s c_g^2}{2} + \frac{c_s^3}{6}\right) + c_2 c_4 \Gamma_N\right) + \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} k_{mn} \frac{\left[sin(m\pi\frac{a_2}{a}) - sin(m\pi\frac{a_1}{a})\right]\left[sin(m\pi\frac{a_4}{a}) - sin(m\pi\frac{a_3}{a})\right]}{(a_2 - a_1)(a_4 - a_3)} \times \frac{\left[sin(n\pi\frac{b_2}{b}) - sin(m\pi\frac{b_1}{b})\right]\left[sin(m\pi\frac{b_4}{b}) - sin(m\pi\frac{b_3}{b})\right]}{(b_2 - b_1)(b_4 - b_3)}, \quad \text{(A.2)}$$

where the term k_{mn} replaces the following expression

$$k_{mn} = C_{mn}\frac{a^2 b^2}{m^2 n^2 \pi^4}\left(f_{mn} - \frac{c_s c_g^2}{2ab\epsilon_N c_2 c_4}\right), \quad \text{with } c_g = max(c_2, c_4) \quad c_s = min(c_2, c_4). \quad \text{(A.3)}$$

Equation (A.2) becomes (3.16) and (A.3) turns into (3.16) when c_2 and c_4 are set to zero. ■

2. SCALING COEFFICIENT OF INDUCTION MATRIX

The solution of the capacitive problem can be used to solve the resistive problem as well. Using equations (3.7), (3.8) and (3.11) one can construct the coefficient of induction matrix $\mathbf{c}$ [15]. From $\mathbf{c}$, admittance matrix $\overline{\mathbf{Y}}$, relating each pairs of contacts i and j can be derived as following

$$\overline{\mathbf{Y}} = \overline{\mathbf{X}}^{\mathbf{T}}\mathbf{c}\,\overline{\mathbf{X}} - \mathbf{\Lambda}(\overline{\mathbf{X}}^{\mathbf{T}}\mathbf{c}\,\overline{\mathbf{X}}) + \mathbf{I}\left[(\overline{\mathbf{X}}^{\mathbf{T}}\mathbf{c}\,\overline{\mathbf{X}})\,\mathbf{e}\right], \quad \text{(A.4)}$$

where operator $\mathbf{\Lambda}(.)$ returns a diagonal matrix whose entries are identical the diagonal elements of the argument and $\mathbf{e}$ is a N_cx1 unity vector, defined as $\mathbf{e} = [\mathbf{1}, \ldots, \mathbf{1}]^{\mathbf{T}}$. Moreover, NxN_c matrix $\overline{\mathbf{X}}$ is defined as

$$\overline{\mathbf{X}} = [\mathbf{e_1}, \mathbf{e_2}, \ldots, \mathbf{e_{N_c}}], \quad \text{(A.5)}$$

where Nx1 vector $\mathbf{e_i}$ is

$$\mathbf{e_i} = [0, \ldots, 0, \underbrace{1, \ldots, 1}_{\text{indices assoc. w. contact } i}, 0, \ldots, 0]^T. \quad \text{(A.6)}$$

Let $\mathbf{b_i}$ be a N_cx1 vector with all entries equal to zero except those indexed i, then $\mathbf{\Lambda}(\overline{\mathbf{X}}^{\mathbf{T}}\mathbf{c}\,\overline{\mathbf{X}})$ can be re-written as

$$\mathbf{\Lambda}(\overline{\mathbf{X}}^{\mathbf{T}}\mathbf{c}\,\overline{\mathbf{X}}) = \mathbf{I}\left\{\sum_{i=1}^{N_c}\left[\mathbf{b_i}^{\mathbf{T}}(\overline{\mathbf{X}}^{\mathbf{T}}\mathbf{c}\,\overline{\mathbf{X}})\,\mathbf{b_i}\right]\mathbf{b_i}\right\}, \quad \text{(A.7)}$$

hence (A.4) simplifies to

$$\overline{\mathbf{Y}} = (\overline{\mathbf{X}}^{\mathbf{T}}\mathbf{c}\,\overline{\mathbf{X}}) + \mathbf{I}\left\{\left[(\overline{\mathbf{X}}^{\mathbf{T}}\mathbf{c}\,\overline{\mathbf{X}})\,\mathbf{e}\right] - \sum_{i=1}^{N_c}\left[\mathbf{b}_i^{\mathbf{T}}(\overline{\mathbf{X}}^{\mathbf{T}}\mathbf{c}\,\overline{\mathbf{X}})\,\mathbf{b}_i\right]\mathbf{b}_i\right\}. \quad \textbf{(A.8)}$$

By appropriate manipulations on **c,** (A.8) can be re-written in the form

$$\overline{\mathbf{Y}} = \overline{\mathbf{X}}^{\mathbf{T}}\,\tilde{\mathbf{c}}\,\overline{\mathbf{X}}\,. \quad \textbf{(A.9)}$$

For simplicity however, here the term $\tilde{\mathbf{c}}$ has been replaced with the **c** notation. ■

This page intentionally left blank

Appendix B
Sensitivity Analysis

The term $G_0|_{z=z'=0}$ is computed as

$$G_0|_{z=z'=0} = \frac{1}{ab\epsilon_N}\frac{\Gamma_N}{\beta_N} . \tag{B.1}$$

The terms Γ_N and β_N are computed recursively from equations (B.2).

$$\begin{bmatrix} \beta_k \\ \Gamma_k \end{bmatrix} = \begin{bmatrix} \frac{\epsilon_{k-1}}{\epsilon_k} & 0 \\ (\frac{\epsilon_{k-1}}{\epsilon_k} - 1)d_k & 1 \end{bmatrix} \begin{bmatrix} \beta_{k-1} \\ \Gamma_{k-1} \end{bmatrix} , \tag{B.2}$$

where the recursion begins with the values $\beta_0 = 1$ and $\Gamma_0 = d$. The term f_{mn} is computed as follows

$$\begin{aligned} f_{mn} &= \frac{1}{ab\gamma_{mn}\epsilon_N}\frac{\beta_N tanh(\gamma_{mn}d)+\Gamma_N}{\beta_N+\Gamma_N tanh(\gamma_{mn}d)} \\ \gamma_{mn} &= \sqrt{(\frac{m\pi}{a})^2 + (\frac{n\pi}{b})^2} . \end{aligned} \tag{B.3}$$

The terms β_N and Γ_N for $m \neq 0$ or $n \neq 0$ are computed recursively from equations (B.4).

$$\begin{bmatrix} \beta_k \\ \Gamma_k \end{bmatrix} = \begin{bmatrix} \frac{\epsilon_{k-1}}{\epsilon_k} cosh^2(\theta_k) - sinh^2(\theta_k) & (\frac{\epsilon_{k-1}}{\epsilon_k} - 1)\, cosh(\theta_k)sinh(\theta_k) \\ (1 - \frac{\epsilon_{k-1}}{\epsilon_k})\, cosh(\theta_k)sinh(\theta_k) & cosh^2(\theta_k) - \frac{\epsilon_{k-1}}{\epsilon_k} sinh^2(\theta_k) \end{bmatrix} \begin{bmatrix} \beta_{k-1} \\ \Gamma_{k-1} \end{bmatrix} , \tag{B.4}$$

where $1 \leq k \leq N$, $\theta_k = \gamma_{mn}(d - d_k)$, $\beta_0 = 1$, and $\Gamma_0 = 0$. ■

Assume $T_\ell = \epsilon_\ell$, then all Γ_k and β_k will not depend on ϵ_ℓ when $0 \leq k < \ell$, hence

$$\begin{bmatrix} \frac{\partial \beta_k}{\partial T_\ell} \\ \frac{\partial \Gamma_k}{\partial T_\ell} \end{bmatrix} = \mathbf{0} , \ \forall\, 0 \leq k < \ell .$$

Consider first the case in which $k = \ell$. Equation (B.4) becomes

$$\begin{bmatrix} \frac{\partial \beta_\ell}{\partial T_\ell} \\ \frac{\partial \Gamma_\ell}{\partial T_\ell} \end{bmatrix} = \frac{\epsilon_{\ell-1}}{\epsilon_\ell^2} \begin{bmatrix} -cosh^2(\theta_\ell) & -cosh(\theta_\ell)\, sinh(\theta_\ell) \\ cosh(\theta_\ell)\, sinh(\theta_\ell) & sinh^2(\theta_\ell) \end{bmatrix} \begin{bmatrix} \beta_{\ell-1} \\ \Gamma_{\ell-1} \end{bmatrix} \tag{B.5}$$

where $\Gamma_{\ell-1}$ and $\beta_{\ell-1}$ are already known, while $\theta_\ell = \gamma_{mn}(d - d_\ell)$ and $\gamma_{mn} = \sqrt{(\frac{m\pi}{a})^2 + (\frac{n\pi}{b})^2}$.

Secondly, consider the case in which $k = \ell + 1$. Equation (B.4) becomes

$$\begin{bmatrix} \frac{\partial \beta_{\ell+1}}{\partial T_\ell} \\ \frac{\partial \Gamma_{\ell+1}}{\partial T_\ell} \end{bmatrix} = \frac{1}{\epsilon_{\ell+1}} \begin{bmatrix} cosh^2(\theta_{\ell+1}) & cosh(\theta_{\ell+1})\, sinh(\theta_{\ell+1}) \\ -cosh(\theta_{\ell+1})\, sinh(\theta_{\ell+1}) & -sinh^2(\theta_{\ell+1}) \end{bmatrix} \begin{bmatrix} \beta_\ell \\ \Gamma_\ell \end{bmatrix}$$

$$+ \begin{bmatrix} \frac{\epsilon_\ell}{\epsilon_{\ell+1}} cosh^2(\theta_{\ell+1}) - sinh^2(\theta_{\ell+1}) & (\frac{\epsilon_\ell}{\epsilon_{\ell+1}} - 1)\, cosh(\theta_{\ell+1})\, sinh(\theta_{\ell+1}) \\ (1 - \frac{\epsilon_\ell}{\epsilon_{\ell+1}})\, cosh(\theta_{\ell+1})\, sinh(\theta_{\ell+1}) & cosh^2(\theta_{\ell+1}) - \frac{\epsilon_\ell}{\epsilon_{\ell+1}} sinh^2(\theta_{\ell+1}) \end{bmatrix} \begin{bmatrix} \frac{\partial \beta_\ell}{\partial T_\ell} \\ \frac{\partial \Gamma_\ell}{\partial T_\ell} \end{bmatrix}. \tag{B.6}$$

For $\ell + 1 < k \leq N$, $\partial \beta_k / \partial T_\ell$ and $\partial \Gamma_k / \partial T_\ell$ are computed as

$$\begin{bmatrix} \frac{\partial \beta_k}{\partial T_\ell} \\ \frac{\partial \Gamma_k}{\partial T_\ell} \end{bmatrix} = \begin{bmatrix} \frac{\epsilon_{k-1}}{\epsilon_k} cosh^2(\theta_k) - sinh^2(\theta_k) & (\frac{\epsilon_{k-1}}{\epsilon_k} - 1)\, cosh(\theta_k)\, sinh(\theta_k) \\ (1 - \frac{\epsilon_{k-1}}{\epsilon_k})\, cosh(\theta_k)\, sinh(\theta_k) & cosh^2(\theta_k) - \frac{\epsilon_{k-1}}{\epsilon_k} sinh^2(\theta_k) \end{bmatrix} \begin{bmatrix} \frac{\partial \beta_{k-1}}{\partial T_\ell} \\ \frac{\partial \Gamma_{k-1}}{\partial T_\ell} \end{bmatrix} \tag{B.7}$$

recursively, where $\partial \beta_{k-1} / \partial T_\ell$ and $\partial \Gamma_{k-1} / \partial T_\ell$ are obtained directly from equation (B.6). The recursion (B.7) ends when $\partial \beta_N / \partial T_\ell$ and $\partial \Gamma_N / \partial T_\ell$ are found. ■

Next, assume $T_\ell = d_\ell$, the layer thickness. Using a similar reasoning as before, consider first the case in which $k = \ell$. Equation (B.4) becomes

$$\begin{bmatrix} \frac{\partial \beta_\ell}{\partial T_\ell} \\ \frac{\partial \Gamma_\ell}{\partial T_\ell} \end{bmatrix} = -\gamma_{\mathrm{mn}} (\frac{\epsilon_{\ell-1}}{\epsilon_\ell} - 1) \begin{bmatrix} 2\, sinh(\theta_\ell)\, cosh(\theta_\ell) & cosh(2\theta_\ell) \\ -cosh(2\theta_\ell) & -2\, sinh(\theta_\ell)\, cosh(\theta_\ell) \end{bmatrix} \begin{bmatrix} \beta_{\ell-1} \\ \Gamma_{\ell-1} \end{bmatrix}, \tag{B.8}$$

where $\Gamma_{\ell-1}$ and $\beta_{\ell-1}$ are already known, while $\theta_\ell = \gamma_{mn}(d - d_\ell)$ and $\gamma_{mn} = \sqrt{(\frac{m\pi}{a})^2 + (\frac{n\pi}{b})^2}$.

Secondly, consider again the case in which $\ell + 1 \leq k \leq N$, $\partial \beta_k / \partial T_\ell$ and $\partial \Gamma_k / \partial T_\ell$ are computed as

$$\begin{bmatrix} \frac{\partial \beta_k}{\partial T_\ell} \\ \frac{\partial \Gamma_k}{\partial T_\ell} \end{bmatrix} = \begin{bmatrix} \frac{\epsilon_{k-1}}{\epsilon_k} cosh^2(\theta_k) - sinh^2(\theta_k) & (\frac{\epsilon_{k-1}}{\epsilon_k} - 1)\, cosh(\theta_k)\, sinh(\theta_k) \\ (1 - \frac{\epsilon_{k-1}}{\epsilon_k})\, cosh(\theta_k)\, sinh(\theta_k) & cosh^2(\theta_k) - \frac{\epsilon_{k-1}}{\epsilon_k} sinh^2(\theta_k) \end{bmatrix} \begin{bmatrix} \frac{\partial \beta_{k-1}}{\partial T_\ell} \\ \frac{\partial \Gamma_{k-1}}{\partial T_\ell} \end{bmatrix} \tag{B.9}$$

recursively, where $\partial\beta_{k-1}/\partial T_\ell$ and $\partial\Gamma_{k-1}/\partial T_\ell$ are obtained directly from equation (B.8). The recursion (B.9) ends when $\partial\beta_N/\partial T_\ell$ and $\partial\Gamma_N/\partial T_\ell$ are obtained. ■

Consider now the sensitivity of the term k_{mn} with respect to parameter T_ℓ. k_{mn} is defined in equations (3.16) and (B.3); after full expansion of its terms, it becomes

$$k_{mn} = \frac{ab\,C_{mn}}{m^2n^2\pi^4\gamma_{mn}\epsilon_N}\,\frac{\beta_N tanh(\gamma_{mn}d)+\Gamma_N}{\beta_N+\Gamma_N tanh(\gamma_{mn}d)}$$
$$\gamma_{mn} = \sqrt{(\tfrac{m\pi}{a})^2+(\tfrac{n\pi}{b})^2}\,.$$

Hence, assuming T_ℓ is either a doping level, which results in different ϵ_ℓ, or a layer thickness d_ℓ, the sensitivity of k_{mn} with respect to T_ℓ, $\forall\ 0 \le \ell < N$ is computed as

$$\begin{aligned}\frac{\partial k_{mn}}{\partial T_\ell} = {} & \frac{ab\,C_{mn}}{m^2n^2\pi^4\gamma_{mn}\epsilon_N} \times \frac{1}{[\beta_N+\Gamma_N tanh(\gamma_{mn}d)]^2} \\ & \times \Big([\dot\beta_N tanh(\gamma_{mn}d)+\dot\Gamma_N][\beta_N+\Gamma_N tanh(\gamma_{mn}d)] \\ & -[\beta_N tanh(\gamma_{mn}d)+\Gamma_N][\dot\beta_N+\dot\Gamma_N tanh(\gamma_{mn}d)]\Big)\,,\end{aligned} \tag{B.10}$$

where the terms $\dot\Gamma_N = \partial\Gamma_N/\partial T_\ell$ and $\dot\beta_N = \partial\beta_N/\partial T_\ell$ are computed from equations (B.7) and (B.9). Similarly, using (B.7), (B.9) and, slightly modified, (B.10), expressions can be easily derived for $T_\ell = d$ or ϵ_N. ■

Finally, consider the sensitivity of term p_{ij} with respect to contact depth c. Expressions for term p_{ij} in presence of zero depth are shown in equation (3.16). Formulae for nonzero depth can be found in [44]. Assume that all contacts have identical depth c, then sensitivity $\partial p_{ij}/\partial c$ is computed as follows

$$\begin{aligned}\frac{\partial p_{ij}}{\partial c} = {} & -\frac{2}{3}\frac{\beta_N}{ab\epsilon_N\beta_N} + \sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\frac{\partial k_{mn}}{\partial c} \\ & \times \frac{\left[sin(m\pi\frac{a_2}{a})-sin(m\pi\frac{a_1}{a})\right]\left[sin(m\pi\frac{a_4}{a})-sin(m\pi\frac{a_3}{a})\right]}{(a_2-a_1)(a_4-a_3)} \\ & \times \frac{\left[sin(n\pi\frac{b_2}{b})-sin(m\pi\frac{b_1}{b})\right]\left[sin(m\pi\frac{b_4}{b})-sin(m\pi\frac{b_3}{b})\right]}{(b_2-b_1)(b_4-b_3)}\,,\end{aligned} \tag{B.11}$$

where the term $\partial k_{mn}/\partial c$ is computed as

$$\frac{\partial k_{mn}}{\partial c} = \frac{\partial k_{mn}}{\partial c} = -C_{mn}\frac{a^2b^2}{m^2n^2\pi^4}\frac{1}{2ab\epsilon_N}, \tag{B.12}$$

where C_{mn} is defined in section 4.. Due to the linearity of the DCT, it is possible to compute the sensitivity of the coefficient of potential by simply calculating $\dot k_{mn}$ and by performing the DCT on it. Several DCTs related to a variety of different depth can be stored and used for the efficient calculation of the effects of technology on a particular circuit. ■

This page intentionally left blank

Appendix C
Convergence of Modified Placement Algorithms

1. MODIFICATION OF SEARCH SPACE

SA has been proven to converge to the global minimum for the class of problems to which placement belongs provided a set of conditions. In [73] conditions were set on the number of moves per annealing temperature, $t_k \rightarrow \infty$. A stronger result, proved, among others, by [74], guarantees convergence for $t_k = 1$, and $T_A(k) = a/ln\ k$, with a sufficiently large.

Theorem C.1: The conditions of [73] are sufficient for the convergence of a SA based placement algorithm where the search space $\{S\}$ is modified dynamically as discussed in section 6.1. during the unfolding of the annealing.
Proof : The annealing algorithm with the proposed modification is still a Markov Chain. Since all new states added to the search space are reachable, and the same properties apply to these states as to the others, the chain is regular and recurrent. Since, for fixed temperature, the transition rule to go from a state to the next is invariant, the chain is also homogeneous. Hence, the standard theory of homogeneous Markov Chains applies. As a result, at a given temperature all states of the chain converge to a stationary probability distribution (Theorems 3.1.4 and 3.2.1 in [73]). This implies that at each temperature a global minimum is reached, thus the convergence of the algorithm follows (Propositions 3.2.1, 3.2.2 and Theorem 3.2.2 in [73]). ■
Similar arguments can be used to prove convergence in the sense of [74], thus ensuring that the whole approach is robust.

2. SUBSTRATE-AWARE PLACEMENT

Theorem C.2: The conditions of [73] are sufficient for the convergence of a SA based placement algorithm with the modifications discussed in section 6.1.2.

Proof : Two equivalent resistive networks are used to represent the substrate in terms of its electrothermal behavior. Both networks are linear with no storage elements. Furthermore, due to finite dimensions of substrate geometries and to the inherent properties of the algorithm, each resistive component is bounded from above and below. Hence, all temperature and noise estimates are necessarily bounded. Consequently the properties of the Markov Chain underlying the annealing are not modified. Since all new states added to the search space are all reachable, and the same properties apply to these states as to the others, the chain is regular and recurrent. Since, for fixed annealing temperature, the transition rule to go from a state to the next is invariant, the chain is also homogeneous. Hence, the standard theory of homogeneous Markov Chains applies. As a result, at a given annealing temperature all states of the chain converge to a stationary probability distribution (Theorems 3.1.4 and 3.2.1 in [73]). This implies that at each annealing temperature a global minimum is reached, thus the convergence of the algorithm follows (Propositions 3.2.1, 3.2.2 and Theorem 3.2.2 in [73]). ■

Similar arguments can be used to prove convergence in the sense of [74], thus ensuring that the whole approach is robust. Convergence in the sense of [73] is also guaranteed when one or more cells are provided with a number of guard rings. In [82] convergence has been proven for SA in which swaps between cell implementations are allowed.

Appendix D
Measurement of Substrate Noise

Several schemes have been proposed for measurement of substrate noise. These schemes address the problem of substrate noise estimation from different perspectives. Some have been presented for accurate verification of results of simulation tools while others are used to study time-domain or frequency-domain effects of substrate noise. We will discuss the measurement methodologies of some of these schemes below and cover key results.

In this appendix we will outline some of these schemes depending on their application, namely DC measurements, small-signal measurements on two-ports, transient measurements and narrow-band, down-conversion measurements. The simplest schemes are best suited for calibration of simulators, which can then be used on more complex circuits. Transient and frequency-domain measurements on complicated circuits are not very useful for verifying simulators. On the other hand, they can be used to demonstrate electrical effects or physical phenomena in the substrate, or to determine trends versus certain physical parameters.

1. DC MEASUREMENTS

Experimental verification of the modeling techniques discussed in this book is presented here. DC measurements were performed on a chip consisting of several p+-to-p contacts. The experimental results are presented and correlated against the predictions of the simulator. A good match is observed between the simulated and experimental results. This experiment is used to verify the predictions of the simulator rather than explore the circuit effects of substrate coupling. An analysis of the sources of error is also performed in this appendix.

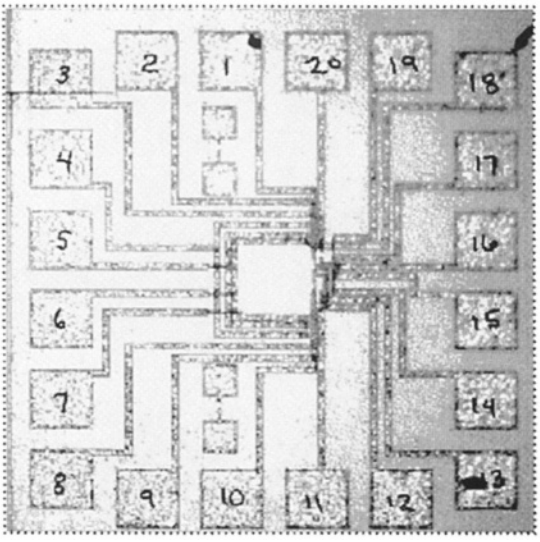


Figure D.1. Test chip microphotograph

1.1 TEST-CHIP DESCRIPTION

A die-photo of the test-chip is shown in Figure D.1. The cross-section of the test-chip substrate is assumed piecewise linear with a $1\mu m$ thick epitaxial layer of $0.1\Omega cm$, a $400\mu m$ thick p-substrate layer of $20\Omega cm$, and a backplate contact represented by a buried-p region of resistivity $0.1\Omega cm$. The test-chip consists of twenty ohmic substrate-contacts (p+-to-p). The minimum contact size is $2\mu m$ by $2\mu m$, while the largest contact size is that of the center square contact, which is $125\mu m$ to a side. There are four rings surrounding the center contact. The innermost ring is a large p+-contact while the outer rings consist of several individual p+ substrate-taps connected in parallel on the surface by metal interconnects. The contact-to-contact distances vary from $2\mu m$ to $100\mu m$. The die measured $950\mu m$ by $875\mu m$. A gold ohmic-contact was made to the back surface of the die and was connected to ground during the testing.

1.2 MEASUREMENT PROCEDURE

Two experiments were performed on the test-chip. In the first experiment, a DC bias of 0.1V was applied to one contact. The voltage appearing at another contact was measured while all other contacts were set to ground. Two values of DC bias were chosen for this purpose (0.1V and 1V) in order to ensure that the junction nonlinearity or high-field effects in the bulk did not interfere with the measurement. It was found that some of the contacts exhibited weak nonlinearity at 1V and therefore data was extracted at 0.1V bias only. The

linearity around zero bias was verified using an I-V curve tracer. For any pair of contacts i and j, the voltage measured at contact j (V_j) with 0.1V applied at contact i is given by the following equation

$$V_j = \frac{0.1R_{j0}}{R_{ij} + R_{j0}}, \tag{D.1}$$

where R_{ij} is the resistance between contacts i and j and R_{j0} is the resistance between contact j and ground.

In a second experiment, the admittance matrix was determined. One contact was kept at a fixed DC bias of 0.1 V and the current flowing through the other contacts was measured. This process was repeated for all the contacts on two separate test-chips.

1.3 SIMULATION PROCEDURE

The two experiments discussed above were simulated using SUBRES and SPICE. A minimum feature size of approximately $1\mu m$ and junction depths of $0.5\mu m$ were used in the simulations.

1.4 DISCUSSION OF RESULTS

Experimental and simulated results are compared in the plots shown in Figure D.2 and D.3. The data from the voltage ratio experiment are shown in Figure D.2. Voltage ratios greater than 0.5 imply a strong coupling between two contacts. It can be seen from the figure that the match between measured and experimental data is less than 15% for voltage ratios greater than 0.5. The match is still good for voltage ratios larger than 0.1. Below a ratio of 0.1, however, the matching between experimental and simulated values degrades considerably.

Figure D.3 shows the current flowing through the contacts with the center square contact at 0.1V. The current varies from 2mA to 40mA which implies that the smallest contact-to-contact resistance is 25Ω. The match between experimental and simulated data is very good in this case. One of the sources of error in these experiments is the approximate substrate profile used in the simulations. The exact values of the bulk and the epitaxial layer resistivities can vary from one test chip to another.

The substrate multi layer setup is an approximation to the physical case of a gradually varying profile. The approximate nature of the resistivity profile can be identified as a possible source of error from the data points for voltage ratios less than 0.1 in Figure D.2. The simulated values are observed to be consistently lower than the experimental data. This implies that the potential falls off faster in the simulation than in the physical substrate. The potential roll-off is directly dependent on the substrate profile.

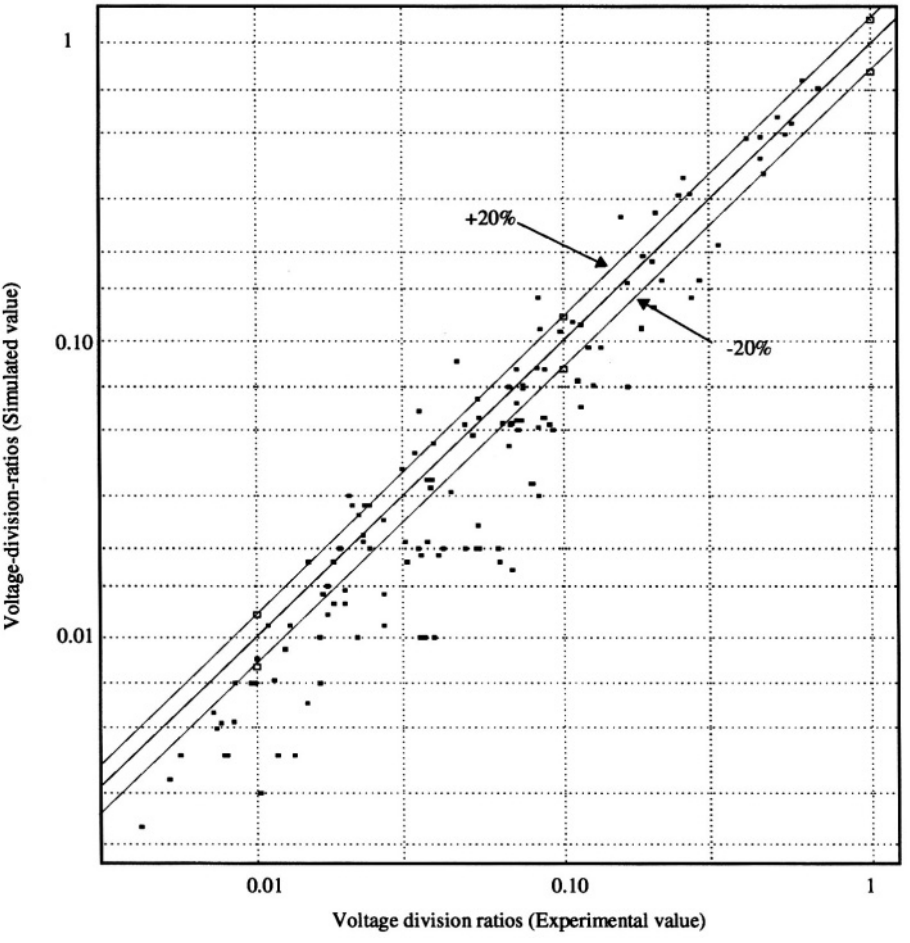


Figure D.2. Simulated vs. experimental results

Furthermore, while the exact match between the experimental and the simulated data degrades at low voltage ratios, the two sets of data show a good correlation, which indicates the presence of a systematic error. An approximate substrate profile will cause a systematic mismatch between the data. The experimental data is noisy at small voltage ratios. This could be measurement related noise. This issue needs to be investigated further. Parasitics not included in the simulator, for example the contact resistances, are yet another source of error. This fact also explains why the disagreement grows with the reduction of contact-to-contact resistances.

The good match obtained between the measured and simulated values implies that the simulator is accurate. Greater precision can be achieved by

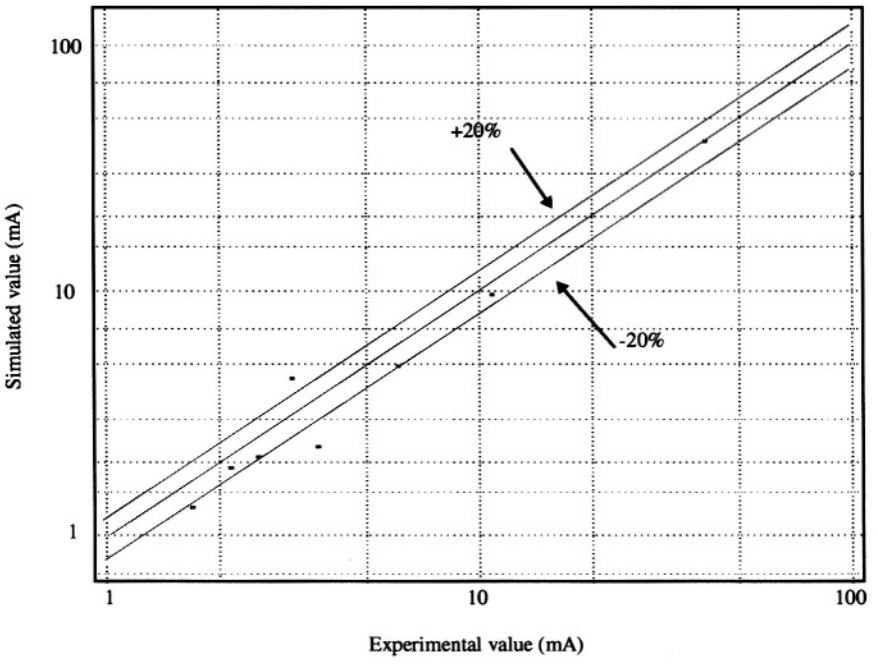


Figure D.3. Simulated vs. experimental results

the use of more accurate substrate profiles. The simulation technique can be extended to several vertical layers with ease, without major computational overhead, which is one of the attractive features of the technique. The test chip used for experimental verification provided a severe test for the capabilities of the simulator. This was due to the wide variation in the contact size and the contact-to-contact distance in addition to the small dimensions of the minimum-size contacts and the small distance between the closest contacts.

2. SMALL-SIGNAL HIGH-FREQUENCY ON-WAFER MEASUREMENTS

As we discussed earlier, substrate coupling can be studied at a fundamental level as a two-port problem, with one port acting as a noise injector and the other as a noise sensor. Simple two-port structures can be used to evaluate the strength of coupling. A typical measurement set-up and substrate contact layout is shown in Figure D.4. The ports can be implemented to resemble injectors and sensors that occur in integrated circuits, for example, the collector-to-substrate capacitance of bipolar devices, the bottom plate parasitic of capacitors or ohmic contacts with the substrate. It is possible to make measurements up to

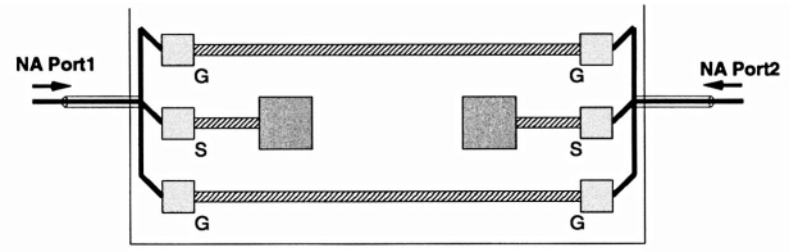


Figure D.4. Two port setup for high-frequency measurement of substrate coupling using a network analyzer

very high frequencies, using appropriate probes and measurement equipment. Measurements of this type up to 40GHz are presented in [31, 83].

High-frequency scattering parameters of two-port structures are measured and converted to Y-parameters using linear transformations. Measured Y-parameters are compared with parameters computed from a finite-element quasi-static simulator for verification of the simulator. An important result demonstrated in this study is that the quasi-static approximation used to model coupling through the substrate is valid up to several GHz. Time-varying magnetic fields can be ignored, and the substrate can be treated as a lossy dielectric, characterized by its resistivity and dielectric constant. This helps to simplify simulation algorithms and reduces simulation time considerably.

It is possible to develop a set of experiments where the parameters of the injector and sensor ports, such as the distance between them and the dimensions of the ports, can be varied over a range. By comparing the outputs of a simulator with the measured results over a wide range of parameter sweeps, the simulator can be calibrated and also validated for calculation of substrate networks for larger problems.

In [83] the authors have used n+ collectors of bipolar devices as injectors and sensors. Two shapes have been employed. In one set of measurements, one of the contacts is a circular disc, while the other node is a concentric ring that surrounds it. The distance between the edges of the contacts is 24μm. Other measurements are made with square contacts spaced at a distance of 240μm. The advantage of using concentric cylindrical contacts is that substrate coupling is primarily a function of the distance between contacts. For square or rectangular contacts, coupling takes place through spreading paths on the surface and therefore is a strong function of surface boundary conditions such as other devices that may be present in close vicinity, or the edge of the silicon die. For concentric cylindrical contacts, measurements have been presented with grounded and floating concentric p+ type guard rings placed between the contacts. These rings are placed at distances of 0.25μm and 7μm from the inner substrate contact. The authors discuss several interesting results

from their study. For example, the importance of modeling parasitic wiring inductance at very high frequencies (> 10GHz) is demonstrated for the guard ring that is laid out at $0.25\mu m$ spacing from the center contact. A rapid degradation of isolation at frequencies greater than 10GHz due to the parasitic inductance is reported.

This result sheds light on the importance of parasitic inductance in the path of the guard ring ground. Another interesting result is that the guard ring at the narrower spacing provided an improvement in isolation of approximately 22dB at a frequency of about 10GHz. This is a substantial improvement in isolation. It should be noted, however, that this improvement occurs for a guard ring that is placed very close to the injector, and has a very small ground-path impedance. Therefore, practically realizable values of improvement in isolation can be significantly smaller. The floor of measurements of this nature is set by the strength of spurious coupling paths, for example direct coupling between the probes used for measurement. It is important to determine these values for the measurement setup *a priori.* Careful calibration is required to minimize the effect of direct coupling. Post-calibration direct coupling through spurious paths can limit isolation measurements to values as large as -40 to -45 dB at several GHz. Similar two-port measurements are described in [81] up to a frequency of nearly 10GHz. A MOSFET of size $50x50\mu m^2$ and a p+ substrate contact of the same size are used as the injector-sensor pair. Isolation is measured between these devices without guard rings and with guard rings placed around both the devices.

An improvement in isolation of about 40dB[1] is demonstrated by the use of guard rings. Comparison of isolation is also provided for similar structures that use buried n+ wells and Silicon-on-Insulator (SOI) substrates for isolation. Buried n+ wells are shown to provide better isolation than SOI.

A drawback of this technique is that since the measurements are made on-wafer, it is difficult to study the impact of package parasitics on substrate coupling. For example, while it is possible to quantify the improvement in substrate coupling due to ideally grounded guard rings, it is difficult to measure the decrease in their effectiveness in the presence of ground-path inductance, which is a significant concern in packaged ICs.

Despite the above shortcoming, calibrated measurements of this level of complexity are perhaps the best suited for verifying the accuracy of simulators. More complicated measurements performed on packaged parts often include too many variables, such as coupling on the board or through the package. This makes it difficult to study coupling effects that occur exclusively through the substrate.

In addition to high-frequency measurements, that are useful for understanding the underlying electromagnetic nature of coupling, DC measurements can be performed on structures that are more complex than the two-port structures

discussed above, to verify simulation techniques in detail. One such structure is presented in [29]. Ohmic substrate contacts are used. The layout is such that the contacts have a large variation in size, varying from squares of $2\mu m$ to approximately $100\mu m$ to a side. Rectangular ohmic contacts with large aspect ratios are also implemented. The distance between contacts is varied over two orders of magnitude. Large variations in the electrical field can exist in the substrate due to the varied shapes and distances between contacts. Node-to-node resistance measurements thus allow for extensive verification of the simulation tool presented in [29]. Measured and simulated Y-matrices are compared for the pattern of contacts.

3. TIME-DOMAIN MEASUREMENTS

Fast on-chip switching activity can generate transient noise in the substrate. A scheme is presented in [11] to measure and characterize this type of noise. An integrated ring-oscillator is used to generate switching noise. Inverters, connected to the output of the ring oscillator and capacitively coupled to the substrate, are used to emulate digital noise injection into the substrate. Common-source MOS devices are used to sense substrate noise. Large substrate contacts are used to establish a DC bias in the substrate. Twelve inverters are used as injectors and ten MOS devices ($200\mu m \times 2\mu m$), configured as noise-sensing current sources, are distributed across the test structure. The relative effect of guard rings on isolation is studied by surrounding some of the current sources with guard rings. The guard rings are either grounded through a dedicated ground, or connected to the ground pad used to bias the substrate contacts. The outputs of the MOS devices are measured as the root-mean-squared (RMS) noise voltage induced on an external load. The impact of several physical parameters and isolation techniques on substrate noise is discussed by means of experiments and simulations. These include separation between the injector and the sensor nodes, the use of guard rings and package parasitic inductance. Experimental measurements have been presented for low-resistivity substrates.

Several interesting results are presented in this study. The isolation between the receiver and the injector is shown to be independent of the distance between the two nodes for distances greater than a multiple of the thickness of the epitaxial region (four times the epitaxial thickness for the case studied in [11]). It is found that majority-diffusion type guard rings with dedicated grounds, placed near the sensor node (about $6\mu m$ away) improve the isolation by approximately 20 percent. This is a small improvement, which is characteristic of low-resistivity substrates.

Rings that are far away (about $22\mu m$) do not have much impact on isolation at all. When guard rings are connected to the node used to bias the substrate, the sensors with the rings around them actually show an increase in the noise level. As explained by the authors, this is due to the noise on the package

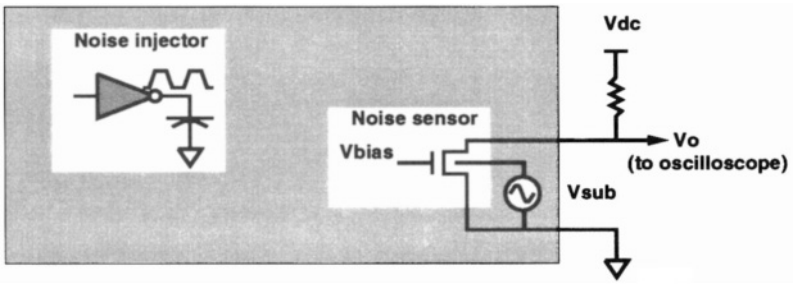


Figure D.5. **Setup for measurement of transient substrate noise**

inductance of the substrate contact, which is more effectively coupled to the sensor in the presence of a guard ring. The authors also demonstrate that minority-diffusion type rings in these substrates do not have much impact on isolation, since substrate current flow is mostly through the p+ bulk layer.

The authors do not attempt to numerically match measured and simulated values of substrate noise, because of approximations made in the simulator. Instead, they match trends obtained by varying several layout-dependent physical parameters, such as distances between noise injectors and sensors. A scheme for measuring transient noise using a statistical approach is presented in [84]. The authors use a high-gain sampling comparator to measure substrate noise. The comparator alternately samples an integrated voltage reference and a voltage source that is varied in incremental steps. A digital noise generator consisting of a bank of inverters is used to inject substrate noise. The integrated voltage reference is effectively the noise sensor, since its voltage level varies due to substrate noise. The clocking rate of the comparator and the inverters is different so that comparator sampling and noise injection is not synchronized. The swept input to the comparator is held at a fixed level for several sampling clock cycles. In the absence of substrate noise, the output of the comparator switches when the swept input equals or exceeds the reference input. If substrate noise is added to the reference, then the output switches when the input level equals or exceeds the modified noisy reference. Since the sampling clock and the clock of the digital noise generator are not synchronized, the level of noise on the reference can vary from sample to sample. At each level of the swept input, a switching probability is associated with the output of the comparator. The total noise power injected into the substrate and a noise signal amplitude are defined using the switching probabilities calculated at the various input voltage levels. Details regarding the statistical analysis techniques used to extract this information can be found in [84].

The technique presented in [84] uses a very good approximation of a large mixed-signal circuit to determine statistical noise measures with on-chip sam-

pling. The statistical noise measured using this technique can provide estimates of the noise floor caused by switching noise in the front-end of sampling data-converters. However, since simplified measures are used as noise estimators, information regarding the frequency content of the injected noise is not included. Spectral domain information of injected noise is useful in some sampling circuits, especially if the inputs to the circuit are narrow-band.

4. FREQUENCY DOMAIN MEASUREMENTS

In several radio-frequency and mixed-signal applications, substrate noise is of a narrow-band or tonal nature. In some situations, the injected noise may cover a relatively wide band width, but the sensor is sensitive to noise in a limited bandwidth. A scheme to measure narrow-band noise injected by an on-chip VCO is presented in [85].

In any measurement scheme, the path taken by the coupled signal is difficult to ascertain. For example, if a sensor is used to detect fast switching transients in the substrate, it is difficult to determine how much noise is coupled to the sensor through the substrate, through pin-to-pin mutual (inductive) coupling in the package or through trace-to-trace coupling on the board. A similar problem exists if we attempt to measure noise injected by narrow-band sources. The approach presented in [85] is to implement an on-chip heterodyne down-converter, that translates signals detected by a sensor to a different frequency and then brings them off-chip. In this way contamination of the substrate coupled signal by direct coupling through the package or through the board is avoided. Package and substrate-coupled signals are especially difficult to distinguish if their coupling strengths are similar. Detection of the two coupling paths by varying the operating frequency is also rarely feasible, since the coupling strengths of both mechanisms vary in a similar manner. Coupling typically increases at higher frequencies.

An on-chip differential VCO operating at 1.5GHz is used as a noise injector. This type of noise injector is important in wireless receivers where the VCO is integrated. The signal injected into the substrate by the VCO is especially a problem in direct-downconversion receivers, where the received signal band overlaps the VCO frequency. The same measurement methodology can conceivably be applied to other narrow-band noise injectors. On-chip sensors are implemented using bottom-plate capacitors of PMOS n-wells. The substrate signal is sensed by the capacitors and down-converted by integrated mixers to a different intermediate frequency and measured externally. The influence of signals coupled through the package and the board on the measured noise signal is reduced by this mechanism, since signals coupled through these media are primarily at the frequency of the VCO, while the substrate-coupled signal is measured at a different frequency. Package and substrate-coupled signals are very difficult to distinguish if their coupling strengths are similar. Detection

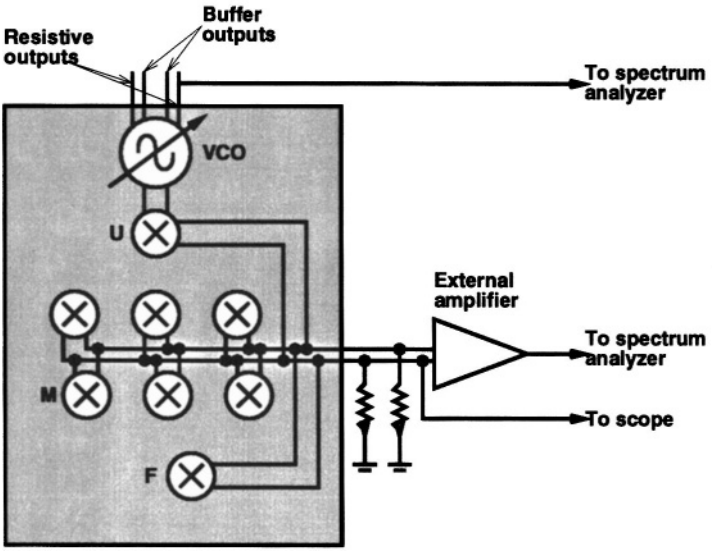


Figure D.6. Measurement set-up

of the two coupling paths by varying the operating frequency is also difficult, since the coupling strengths of both mechanisms vary in a similar manner. Coupling typically increases at higher frequencies.

In the measurement scheme eight down-conversion mixers with differential inputs and differential current-mode outputs are used. An abstract view of this measurement approach is presented in Figure D.6. One of the differential inputs of each mixer is connected to a corresponding noise sensing capacitor. The outputs of the mixers are connected together. The current outputs are converted to a voltage using an external transimpedance amplifier of sufficient bandwidth. Only one mixer is enabled at a time for making measurements. Two mixers are used solely for calibration of the noise floor and the upper reference of the measurement respectively. The upper limit is obtained from a mixer that is connected directly to the outputs of the VCO ("U" in Figure D.6).

The output voltage across the VCO can be a large signal, that may drive this mixer into a non-linear region of operation. Therefore the outputs of the VCO are attenuated using an external control, and the output of "U" is measured for various levels of VCO output. The peak down-converted signal for the unattenuated VCO signal is determined by extrapolating the linear portion of the mixer output vs. VCO signal curve, to avoid measurement artifacts arising from compression of this mixer. The outputs of the six mixers, marked "M" in the figure, are expressed relative to the measurement performed on mixer "U" in dBc. This is a measure of the attenuation of the VCO signal through the substrate. The outputs of these mixers are measured prior to the external

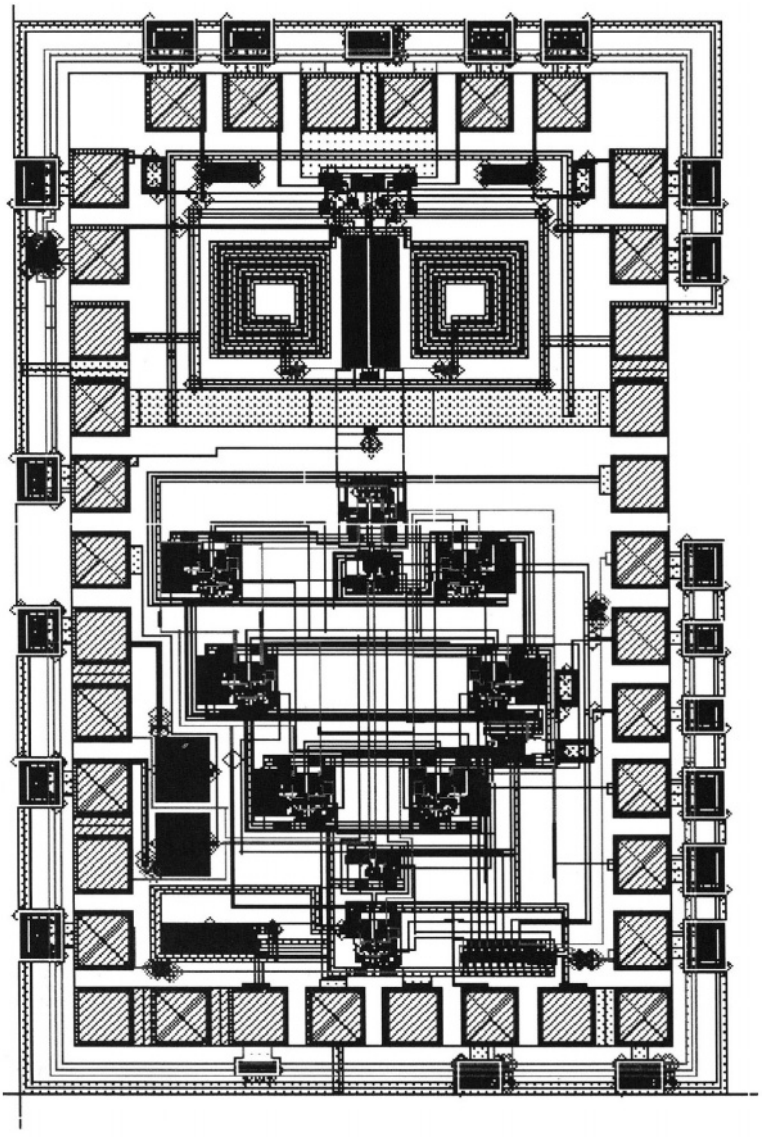


Figure D.7. Fabricated downconverter

amplifier on a digitizing oscilloscope and on the spectrum analyzer. This is done to calibrate the linearity of the external amplifier. The mixers are switched in sequence, one at a time. Measurements are made with the VCO operating unattenuated at its peak amplitude, since this would be the condition in a real integrated receiver.

The layout of the test chip is shown in Figure D.7. Mixers "M" are physically

placed at various distances from the VCO to determine the effect of distance on isolation. One of the mixers is surrounded by a guard ring that is connected to the substrate tie of the mixer. The mixer substrate tie is separated from the mixer ground. The measurement is repeated over different frequencies. The mixer for measuring noise floor ("F" in the figure) is not connected to a sensor, and therefore measures any ambient noise that may couple in through other sources, for example the junction capacitors of devices in the mixer or the bias circuits.

It is important for this measurement to be sufficiently lower than the outputs of mixers "M", in order for the results to be meaningful. The measured outputs from mixers "M" varied from -59 to -47dB below the peak VCO signal. This variation was observed for measurements made with different VCO terminations, including shorted and open outputs and outputs terminated in **50Ω**. The largest coupling was observed for open VCO outputs, while the smallest coupling was recorded for resistively terminated outputs. For a given termination, the outputs of all mixers were within a spread of 3dB. Since a low-resistivity substrate was used, this result is consistent, as it signifies that coupling is not a strong function of distance. A surprising result was that the mixer surrounded by the substrate guard ring had a consistently larger output than the other outputs, for all terminations and frequencies. This result is similar to that observed in [11] with the guard ring connected to the substrate tie of the IC. The substrate node was shared by all the mixers and the ESD cells, and therefore had a very large area, and a very low impedance to the ground of the VCO. The guard ring around the mixer merely coupled the voltage on the substrate node relative to the mixer ground, more effectively into the capacitive sensor. This resulted in a larger down-converted signal.

While the work described in [85] was primarily the description of a measurement technique, three relevant results were derived. The typical strength of coupled signals for a differential on-chip VCO topology, with integrated inductors and varactors was determined. The lack of dependence of substrate coupling on distance in low-resistivity substrates was verified and the importance of using the correct ground node for guard ring connections was highlighted. There are some possible problems with this approach, which include sensing of signals through parasitics, such as interconnect-to-substrate capacitance, and uncertainty in measurement due to matching of mixer gains. Some other potential sources of error are discussed in [85].

Notes

1 Exact measures of isolation are difficult to compare, since published results often use different technologies for implementing test structures. It is important to know the exact cross-section of the substrate to compare results. This information is very often not available. It is easier to comment on relative improvements in isolation for different types of isolation schemes in a given technology.

References

[1] P. R. Gray and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*, J. Wiley & Sons, New York, 3rd Edition, 1993.

[2] K. Fukahori and P. R. Gray, "Computer Simulation of Integrated Circuits in the Presence of Electrothermal Interactions", *IEEE Journal of Solid State Circuits*, vol. SC-11, pp. 834–846, December 1976.

[3] E. J. Pinto and B. B. Mikic, "Temperature Prediction on Substrates and Integrated Circuit Chips", *Heat Transfer Engineering*, vol. 9, n. 3, pp. 54–65, 1988.

[4] G. Strang and G. J. Fix, *An Analysis of the Finite Element Method*, Prentice-Hall, Englewood Cliffs,NJ, 1973.

[5] W. F. Ames, *Numerical methods for partial differential equations*, Academic Press, San Diego, 1992.

[6] J. C. Rudell et al., "Recent Developments in High Integration Multi-Standard CMOS Tranceivers for Personal Communication Systems", in *International Symposium on Low Power Electronics*, pp. 149–154, August 1998.

[7] B. R. Stanisic, N. K. Verghese, D. J. Allstot, R. A. Rutenbar and L. R. Carley, "Addressing Substrate Coupling in Mixed-Mode ICs: Simulation and Power Distribution Synthesis", *IEEE Journal of Solid State Circuits*, vol. SC-29, n. 3, pp. 226–237, March 1994.

[8] S. Mitra, R. A. Rutenbar, L. R. Carley and D. J. Allstot, "Substrate-Aware Mixed-Signal Macro-Cell Placement in WRIGHT", in *Proc. IEEE Custom Integrated Circuit Conference*, pp. 529–532, May 1994.

[9] L. T. Pillage and R. A. Rohrer, "Asymptotic Waveform Evaluation for Timing Analysis", *IEEE Trans. on Computer Aided Design*, vol. CAD-9, n. 4, pp. 352–366, April 1990.

[10] I. L. Wemple and A. T. Yang, "Mixed-Signal Switching Noise Analysis Using Voronoi-Tessellated Substrate Macromodels", in *Proc. IEEE/ACM Design Automation Conference*, pp. 439–444, June 1995.

[11] D. K. Su, M. Loinaz, S. Masui and B. Wooley, "Experimental Results and Modeling Techniques for Substrate Noise in Mixed-Signal Integrated Circuits", *IEEE Journal of Solid State Circuits*, vol. SC-28, n. 4, pp. 420–430, April 1993.

[12] K. Joardar, "A Simple Approach to Modeling Cross-Talk in Integrated Circuits", *IEEE Journal of Solid State Circuits*, vol. SC-29, n. 10, pp. 1212–1219, October 1994.

[13] E. Charbon, R. Gharpurey, R. G. Meyer and A. L. Sangiovanni-Vincentelli, "Analysis and Optimization of Substrate Noise in VLSI ICs", *IEEE Trans. on Computer Aided Design*, vol. CAD-18, n. 2, pp. 172–190, February 1999.

[14] C. Hu, *VLSI Electronics: Microstructure Science*, volume 18, Academic Press, New York, 1981.

[15] R. Gharpurey, *Modeling and Analysis of Substrate Coupling in ICs*, PhD thesis, University of California at Berkeley, May 1995.

[16] P. Miliozzi, L. Carloni, E. Charbon and A. L. Sangiovanni-Vincentelli, "SUBWAVE: a Methodology for Modeling Digital Substrate Noise Injection in Mixed-Signal ICs", in *Proc. IEEE Custom Integrated Circuit Conference*, pp. 385–388, May 1996.

[17] E. Charbon, P. Miliozzi, L. Carloni, A. Ferrari and A. L. Sangiovanni-Vincentelli, "Modeling Digital Substrate Noise Injection in Mixed-Signal ICs", *Trans. on Computer Aided Design*, vol. CAD-18, n. 3, pp. 301–310, March 1999.

[18] S. M. Sze, *Physics of Semiconductor Devices*, Wiley, New York, 1981.

[19] Y. Cheng, M. Chan, K. Hui, M. Jeng, Z. Liu, J. Huang, K. Chen, J. Chen, R. Tu, P. K. Ko and C. Hu, *BSIM3v3 Manual*, University of California at Berkeley, Berkeley, CA, 1996.

[20] R. B. Merril, W. M. Young and K. Brehmer, "Effect of Substrate Material on Crosstalk in Mixed Analog/Digital ICs", in *Proc. IEEE International Electron Devices Meeting*, pp. 433–436, December 1994.

[21] D. A. Kirkpatrick, *The Implications of Deep Sub-micron Technology on the Design of High Performance Digital VLSI Systems*, PhD thesis, University of California at Berkeley, December 1997.

[22] P. Parakh and R. B. Brown, "Crosstalk Constrained Global Route Embedding", in *Proc. ACM International Symposium on Physical Design*, pp. 201–206, April 1999.

[23] T. A. Johnson, R. W. Knepper, V. Marcello and W. Wang, "Chip Substrate Resistance Modeling Technique for Integrated Circuit Design", *IEEE Trans. on Computer Aided Design*, vol. CAD-3, pp. 126–134, 1984.

[24] N. P. van der Mejis and T. Smedes, "Accurate Interconnect Modeling: Towards Multi-Million Transistor Chips as Microwave Circuits", in *Proc. IEEE International Conference on Computer Aided Design*, pp. 244–251, November 1996.

[25] F. J. R. Clement, E. Zysman M. Kayal and M. Declercq, "LAYIN: Toward a Global Solution for Parasitic Coupling Modeling and Visualization", in *Proc. IEEE Custom Integrated Circuit Conference*, pp. 537–540, May 1994.

[26] N. K. Verghese, T. Schmerbeck and D. J. Allstot, *Simulation Techniques and Solutions for Mixed-Signal Coupling in ICs*, Kluwer Academic Publ., Boston, MA, 1995.

[27] T. Smedes, N. P. van der Mejis and A. J. Genderen, "Boundary Element Methods for 3D Capacitance and Substrate Resistance Calculations in Inhomogeneous Media in a VLSI Layout Verification Package", *Advances in Engineering Software*, vol. 20, n. 1, pp. 19–27, 1994.

[28] T. Smedes, N. P. van der Meijs and A. J. van Genderen, "Extraction of Circuit Models for Substrate Cross-Talk", in *Proc. IEEE International Conference on Computer Aided Design*, pp. 199–206, November 1995.

[29] R. Gharpurey and R. G. Meyer, "Modeling and Analysis of Substrate Coupling in ICs", *IEEE Journal of Solid State Circuits*, vol. SC-31, n. 3, pp. 344–353, March 1996.

[30] S. Kapur and D. E. Long, "IES^3: Efficient Electrostatic and Electromagnetic Simulation", *IEEE Computational Science and Engineering*, vol. 5, n. 4, pp. 60–67, October-December 1998.

[31] M. Pfost and H. M. Rein, "Modeling and Measurement of Substrate Coupling in Si-Bipolar ICs up to 40GHz", *IEEE Journal of Solid State Circuits*, vol. SC-33, n. 4, pp. 582–591, 1998.

[32] C. P. Coelho, J. R. Phillips and L. M. Silveira, "Robust Rational Function Approximation Algorithm for Model Generation", in *Proc. IEEE/ACM Design Automation Conference*, pp. 207–212, June 1999.

[33] X. Huang, V. R. Raghavan and R. A. Rohrer, "AWEsim: A Program for the Efficient Analysis of Linearized Circuits", in *Proc. IEEE International Conference on Computer Aided Design*, pp. 534–537, November 1990.

[34] P. Feldmann and R. W. Freund, "Efficient Linear Circuit Analysis by Pade Approximation via the Lanczos Process", *IEEE Trans. on Computer Aided Design*, vol. CAD-14, n. 5, pp. 639–649, May 1995.

[35] L. M. Silveira, M. Kamon, I. Elfadel and J. White, "A Coordinate-Transformed Arnoldi Algorithm for Generating Guaranteed Stable Reduced-Order Models of RLC Circuits", in *Proc. IEEE International Conference on Computer Aided Design*, pp. 288–294, November 1996.

[36] A. Odabasioglu, M. Celik and L. Pileggi, "PRIMA: Passive Reduced-Order Interconnect Macromodeling Algorithm", *IEEE Trans. on Computer Aided Design*, vol. CAD-17, n. 8, pp. 645–54, August 1998.

[37] J. R. Phillips, E. Chiprout and D. D. Ling, "Efficient Full-Wave Electromagnetic Analysis via Model-Order Reduction of Fast Integral Transforms", in *Proc. IEEE/ACM Design Automation Conference*, pp. 377–382, June 1996.

[38] D. Colton and R. Kress, *Integral Equations Methods in Scattering Theory*, Krieger, Malabar, Florida, 1992.

[39] K. E. Atkinson, "A Survey of Boundary Integral Equation Methods for the Numerical Solution of Laplace's Equation in Three Dimensions", in M. A. Goldberg, editor, *Numerical Solution of Integral Equations*, pp. 1–34, Plenum Press, New York, 1990.

[40] J.D.Jackson, *Classical Electrodynamics*, Wiley, 1975.

[41] S. M. Rao, T. K. Sarkar and R. F. Harrington, "The Electrostatic Field of Conducting Bodies in Multiple Dielectric Media", *IEEE Trans. on Microwave Theory and Techniques*, vol. MTT-32, n. 11, pp. 1441–1448, November 1984.

[42] W. T. Weeks, "Calculation of Coefficients of Capacitance of Multiconductor Transmission Lines in the Presence of a Dielectric Interface", *IEEE Trans. on Microwave Theory and Techniques*, vol. MTT-18, n. 1, pp. 35–43, January 1970.

[43] R. Crampagne, M. Ahmadpanah and J. L. Guiraud, "A Simple Method for Determining the Green's Function for a Large Class of MIC Lines Having Multilayered Dielectric Structures", *IEEE Trans. on Microwave Theory and Techniques*, vol. MTT-26, pp. 82–87, 1978.

[44] E. Charbon, *Constraint-Driven Analysis and Synthesis of High-Performance Analog IC Layout*, PhD thesis, University of California at Berkeley, December 1995.

[45] I. S. Duff, A. M. Erisman and J. K. Reid, *Direct Methods for Sparse Matrices*, Clarendon Press, Oxford, 1986.

[46] Y. Saad, *Iterative Methods for Sparse Linear Systems*, PWS Publishing, Boston, MA, 1996.

[47] J. A. Meijerink and H. A. van der Vorst, "An Iterative Solution Method for Linear Systems of which the Coefficient Matrix is a Symmetric M-Matrix", *Mathematics of Computation*, vol. 31, pp. 148–162, January 1977.

[48] W. Hackbusch, *Multigrid Methods and Applications*, volume 4 of *Computational Mathematics*, Springer, Berlin, 1985.

[49] J. H. Bramble, J. E. Pasciak and J. Xu, "Parallel multilevel preconditioners", *Mathematics of Computation*, vol. 55, pp. 1–22, 1990.

[50] R. W. Hockney and J. W. Eastwood, *Computer Simulation Using Particles*, Adam Hilger, New York, 1988.

[51] P. Concus, G. H. Golub and D. P. O'Leary, "A Generalized Conjugate Gradient Method for the Numerical Solution of Elliptic PDE", in J. R. Bunch and D. J. Rose, editors, *Sparse Matrix Computations*, pp. 309–332, Academic Press, New York, 1976.

[52] R. F. Harrington, *Field Computation by Moment Methods*, MacMillan, 1968.

[53] S. Kapur and D. E. Long, "High-order Nystrom Schemes for Efficient 3-D Capacitance Extraction", in *Proc. IEEE International Conference on Computer Aided Design*, pp. 178–185, November 1998.

[54] A. J. van Genderen, N. P. van der Mejis and T. Smedes, "Fast Computation of Substrate Resistances in Large Circuits", in *Proc. European Design and Test Conference*, pp. 560–565, March 1996.

[55] Y. Saad and M. Schultz, "GMRES: A Generalized Minimal Residual Algorithm for Solving Nonsymmetric Linear Systems", *SIAM Journal on Scientific and Statistical Computing*, vol. 7, n. 3, pp. 856–869, July 1986.

[56] J. P. Costa, M. Chou and L. M. Silveira, "Efficient Techniques for Accurate Modeling and Simulation of Substrate coupling in mixed-signal ICs", in *Design, Automation and Test in Europe*, pp. 892–898, February 1998.

[57] J. R. Phillips and J. K. White, "A Precorrected-FFT Method for Electrostatic Analysis of Complicated 3D Structures", *IEEE Trans. on Computer Aided Design*, vol. CAD-16, n. 10, pp. 1059–1072, 1997.

[58] J. P. Costa, M. Chou and L. M. Silveira, "Precorrected-DCT Techniques for Modeling and Simulation of Substrate coupling in mixed-signal IC's",

in *Proc. IEEE International Symposium on Circuits and Systems*, volume 6, pp. 358–362, May 1998.

[59] M. Chou and J. White, "Multilevel integral equation methods for the extraction of substrate coupling parameters in mixed-signal IC's", in *Proc. IEEE/ACM Design Automation Conference*, pp. 20–25, June 1998.

[60] J. Tausch and J. White, "A Multiscale Method for Fast Capacitance Extraction", in *Proc. IEEE/ACM Design Automation Conference*, pp. 537–542, June 1999.

[61] P. Miliozzi, I. Vassiliou, E. Charbon, E. Malavasi and A. L. Sangiovanni-Vincentelli, "Use of Sensitivities and Generalized Substrate Models in Mixed-Signal IC Design", in *Proc. IEEE/ACM Design Automation Conference*, pp. 227–232, June 1996.

[62] W. Nye, D. C. Riley, A. L. Sangiovanni-Vincentelli and A. L. Tits, "DELIGHT.SPICE: An Optimization-Based System for the Design of Integrated Circuits", *IEEE Trans. on Computer Aided Design*, vol. CAD-7, n. 4, pp. 501–519, April 1988.

[63] L. Lavagno and A. Sangiovanni-Vincentelli, *Algorithms for Synthesis and Testing of Asynchronous Circuits*, Kluwer, 1993.

[64] X. Liu, "Digital Noise Analysis in Gate-Level Logic Simulation", Unpublished report, University of California, Berkeley, December 1997.

[65] P. Larsson and C. Svensson, "Noise in Digital CMOS Circuits", *IEEE Journal of Solid State Circuits*, vol. SC-29, n. 6, pp. 655–661, June 1996.

[66] I. A. Young, J. K. Greason and K. L. Wong, "A PLL Clock Generator with 5 to 110 MHz of Lock Range for Microprocessors", *IEEE Journal of Solid State Circuits*, vol. SC-27, n. 11, pp. 1599–1607, November 1992.

[67] D. Reynolds, "A 320 MHz CMOS Triple 8b DAC with On-Chip PLL and Hardware Cursor", in *Proc. IEEE International Solid-State Circuits Conference*, pp. 50–51, February 1994.

[68] A. Demir, E. Liu, A. L. Sangiovanni-Vincentelli and I. Vassiliou, "Behavioral Simulation Techniques for Phase/Delay-Locked Systems", in *Proc. IEEE Custom Integrated Circuit Conference*, pp. 453–456, May 1994.

[69] E. Felt, A. Narayan and A. L. Sangiovanni-Vincentelli, "Measurements and Modeling of MOS Transistor Current Mismatch in Analog ICs", in

Proc. IEEE International Conference on Computer Aided Design, pp. 272–277, November 1994.

[70] E. Charbon, E. Malavasi, U. Choudhury, A. Casotto and A. L. Sangiovanni-Vincentelli, "A Constraint-Driven Placement Methodology for Analog Integrated Circuits", in *Proc. IEEE Custom Integrated Circuit Conference*, pp. 2821–2824, May 1992.

[71] E. Charbon, E. Malavasi, D. Pandini and A. L. Sangiovanni-Vincentelli, "Simultaneous Placement and Module Optimization of Analog IC's", in *Proc. IEEE/ACM Design Automation Conference*, pp. 31–35, June 1994.

[72] U. Choudhury and A. L. Sangiovanni-Vincentelli, "Constraint Generation for Routing Analog Circuits", in *Proc. IEEE/ACM Design Automation Conference*, pp. 561–566, June 1990.

[73] F. Romeo, *Simulated Annealing: Theory and Applications to Layout Problems*, PhD thesis, University of California at Berkeley, March 1989.

[74] B. Hajek, "Cooling Schedules for Optimal Annealing", *Mathematics of Operations Research*, vol. 13, n. 2, pp. 311–329, May 1988.

[75] R. Neff, P. Gray and A. L. Sangiovanni-Vincentelli, "A Module Generator for High Speed CMOS Current Output Digital/Analog Converters", in *Proc. IEEE Custom Integrated Circuit Conference*, pp. 481–484, May 1995.

[76] I. Vassiliou, H. Chang, A. Demir, E. Charbon, P. Miliozzi and A. L. Sangiovanni-Vincentelli, "A Video Driver System Designed Using a Top-Down, Constraint-Driven Methodology", in *Proc. IEEE International Conference on Computer Aided Design*, pp. 463–468, November 1996.

[77] N. Camilleri, J. Kirchgessner, J. Costa, D. Ngo and D. Lovelace, "Bonding Pad Models for Silicon VLSI Technologies and Their Effects on the Noise Figure of RF NPNs", in *Proc. IEEE Microwave and Millimeter-Wave Monolithic Circuits Symposium*, pp. 225–228, 1994.

[78] W. Liu, R. Gharpurey, M. C. Chang, U. Erdogan, R. Aggarwal and J. P. Mattia, "RF MOSFET Modeling Accounting for Distributed Substrate and Channel Resistance with Emphasis on the BSIM3v3 Spice Model", in *Proc. IEEE International Electron Devices Meeting*, pp. 309–312, 1997.

[79] R. G. Meyer and A. K. Wong, "Blocking and Desensitization in RF Amplifiers", *IEEE Journal of Solid State Circuits*, vol. SC-30, n. 8, pp. 944–946, 1995.

[80] A. Amarasekara, S. Ramaswamy, M. C. Chang and C. Duvvury, "Modeling MOS Snapback and Parasitic Bipolar Action for Circuit-Level ESD and High Current Simulations", in *Proc. IEEE International Reliability Physics Symposium*, pp. 318–327, 1996.

[81] K. Joardar, "Signal Isolation in BiCMOS Mixed Mode Integrated Circuits", in *Proc. Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 178–181, October 1995.

[82] E. Charbon, E. Malavasi, D. Pandini and A. L. Sangiovanni-Vincentelli, "Imposing Tight Specifications on Analog IC's through Simultaneous Placement and Module Optimization", in *Proc. IEEE Custom Integrated Circuit Conference*, pp. 525–528, May 1994.

[83] M. Pfost, H. M. Rein and T. Holzwarth, "Modeling Substrate Effects in the Design of High-Speed Si-Bipolar ICs", *IEEE Journal of Solid State Circuits*, vol. SC-31, n. 10, pp. 1493–1501, October 1995.

[84] K. Makie-Fukuda, T. Kikuchi, T. Matsuura and M. Hotta, "Measurement of Digital Noise in Mixed-Signal Integrated Circuits", *IEEE Journal of Solid State Circuits*, vol. SC-30, n. 2, pp. 87–92, February 1995.

[85] R. Gharpurey, "A Methodology for Measurement and Characterization of Substrate Noise in High-Frequency Circuits", in *Proc. IEEE Custom Integrated Circuit Conference*, pp. 487–491, May 1999.

Index