Bandwidth Extension in CMOS with Optimized On-Chip Inductors

Sunderarajan S. Mohan, Maria del Mar Hershenson, Stephen P. Boyd, and Thomas H. Lee

***Abstract*—We present a technique for enhancing the bandwidth of gigahertz broad-band circuitry by using optimized on-chip spiral inductors as shunt-peaking elements. The series resistance of the on-chip inductor is incorporated as part of the load resistance to permit a large inductance to be realized with minimum area and capacitance. Simple, accurate inductance expressions are used in a lumped circuit inductor model to allow the passive and active components in the circuit to be simultaneously optimized. A quick and efficient global optimization method, based on *geometric programming*, is discussed. The bandwidth extension technique is applied in the implementation of a 2.125-Gbaud preamplifier that employs a common-gate input stage followed by a cascoded common-source stage. On-chip shunt peaking is introduced at the dominant pole to improve the overall system performance, including a 40% increase in the transimpedance. This implementation achieves a 1.6-kΩ transimpedance and a 0.6-μA input-referred current noise, while operating with a photodiode capacitance of 0.6 pF. A fully differential topology ensures good substrate and supply noise immunity. The amplifier, implemented in a triple-metal, single-poly, 14-GHz $f_{T\max}$, 0.5-μm CMOS process, dissipates 225 mW, of which 110 mW is consumed by the 50-Ω output driver stage. The optimized on-chip inductors consume only 15% of the total area of 0.6 mm^2.**

***Index Terms*—CMOS analog integrated circuits, inductors, integrated circuit design, integrated circuit modeling.**

I. Introduction

THE explosive growth in the commercial wired telecommunications market has generated tremendous interest in low-cost implementations of radio-frequency (RF) broad-band receivers. The performance of such a receiver's front end is determined to a large extent by the preamplifier. Traditionally, this preamplifier has been fabricated in expensive GaAs and silicon bipolar technologies. However, the quest for low-cost solutions in the commercial market has spurred a desire to implement RFIC's in standard CMOS technology. An additional advantage of these CMOS processes is that they permit the integration of the analog and digital components, the holy grail for "system-on-chip" solutions. The performance of CMOS technologies is improving constantly and consistently, thanks to the scaling achieved by the highly competitive microprocessor market. In fact, submicrometer CMOS technologies now exhibit sufficient performance for radio-frequency applications in the 1–2-GHz range. This paper discusses how optimized on-chip inductors can be used to enhance the bandwidth of broad-band amplifiers and thereby push the performance limits of CMOS implementations. An attractive feature of this technique is that the bandwidth enhancement comes with no additional power dissipation.

This bandwidth enhancement is achieved by shunt peaking, a method first used in the 1940's to extend the bandwidth of television tubes. Section II describes the fundamentals of this approach. Section III focuses on how shunt-peaked amplifiers can be implemented in the integrated circuit environment. A well-accepted lumped circuit model for a spiral inductor is used along with recently developed inductance expressions to allow the inductor modeling to be performed in a standard circuit design environment such as SPICE. This approach circumvents the inconvenient, iterative interface between an inductor simulator and a circuit design tool. Most important, a new design methodology is described that yields a large inductance in a small die area.

The new method is implemented using a simple and efficient circuit design computer-aided design tool described in Section IV. This tool is based on geometric programming (GP), a special type of optimization problem for which very *efficient global* optimization methods have been developed. An attractive feature of this technique is that it enables the designer to optimize passive and active devices simultaneously. This feature allows a shunt-peaked amplifier with on-chip inductors to be optimized directly from specifications.

Sections V and VI illustrate how shunt peaking is used to improve the performance of a transimpedance preamplifier. A prototype preamplifier, intended for gigabit optical communication systems, is implemented in a 0.5-μm CMOS process. The use of on-chip shunt peaking permits a 40% increase in the transimpedance with no additional power dissipation. The optimized on-chip inductors only consume 15% of the total chip area.

Section VII summarizes the main contributions of this paper.

II. Shunt Peaking

Although inductors are commonly associated with narrow-band circuits, they are useful in broad-band circuits as well. In this section, we study how an inductor can enhance the bandwidth of a broadband amplifier.

We consider the simple common source amplifier illustrated in Fig. 1(a). For simplicity, we assume that the small signal frequency response of this amplifier is determined by a single

Manuscript received July 27, 1999; revised October 22, 1999.

S. S. Mohan and M. del Mar Hershenson were with the Electrical Engineering Department, Stanford University, Stanford, CA 94305-4070 USA. They are now with Barcelona Design, Inc., Mountain View, CA 94040 USA (e-mail: mohan@barcelonadesign.edu).

S. P. Boyd and T. H. Lee are with the Electrical Engineering Department, Stanford University, Stanford, CA 94305-4070 USA.

Publisher Item Identifier S 0018-9200(00)00561-8.

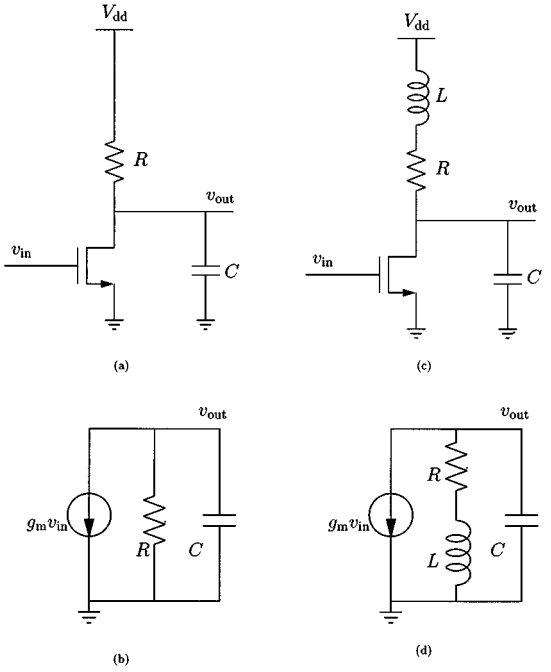


Fig. 1. Shunt peaking a common source amplifier. (a) Simple common source amplifier and (b) its equivalent small signal model. (c) Common source amplifier with shunt peaking and (d) its equivalent small signal model.

dominant pole, which is determined solely by the output load resistance R and the load capacitance C [Fig. 1(b)]

$$\frac{v_{\text{out}}}{v_{\text{in}}}(\omega) = \frac{g_m R}{1 + j\omega RC}. \tag{1}$$

The introduction of an inductance L in series with the load resistance alters the frequency response of the amplifier [Fig. 1(c)]. This technique, called shunt peaking, enhances the bandwidth of the amplifier by transforming the frequency response from that of a single pole to one with two poles and a zero [Fig. 1(d)]

$$\frac{v_{\text{out}}}{v_{\text{in}}}(\omega) = \frac{g_m(R + j\omega L)}{1 + j\omega RC - \omega^2 LC}. \tag{2}$$

The poles may or may not be complex (although, they are complex for practical cases of bandwidth extension). The zero is determined solely by the L/R time constant and is primarily responsible for the bandwidth enhancement.

The frequency response of this shunt peaked amplifier is characterized by the ratio of the L/R and RC time constants. This ratio is denoted by m so that $L = mR^2C$.

Fig. 5 illustrates the frequency response of the shunt-peaked amplifier for various values of m. The case with no shunt peaking ($m = 0$) is used as the reference so that its low-frequency gain and its $\omega_{3\,\text{dB}}$ (3-dB bandwidth) are equal to one ($RC = 1$ and $g_m R = 1$). The frequency response is plotted for the values of m listed in Table I [1].

As expected, the 3-dB bandwidth increases as m increases. The maximum bandwidth is obtained when $m = 0.71$ and yields an 85% improvement in bandwidth. However, as can be clearly seen in the magnitude plot, this comes at the cost of significant gain peaking. A maximally flat response may be obtained for $m = 0.41$ with a still impressive bandwidth improvement of 72%.

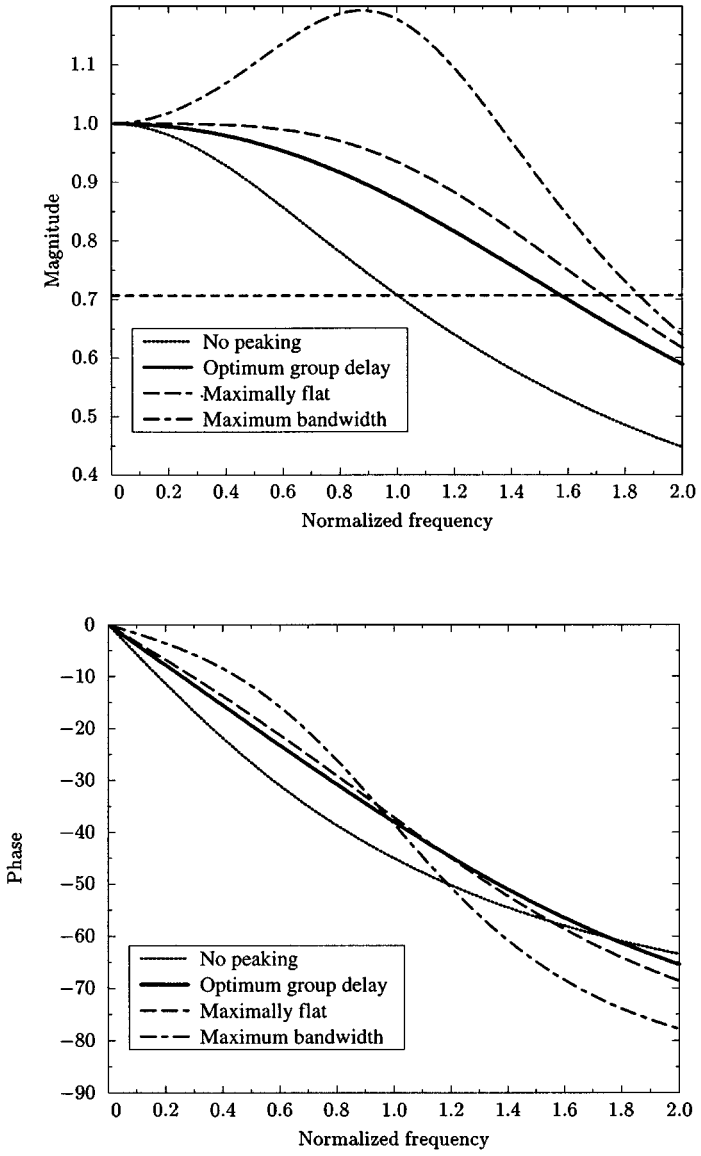


Fig. 2. Frequency response of shunt-peaked cases tabulated in Table 1

TABLE I
PERFORMANCE METRICS FOR SHUNT PEAKING

Factor (m)	Normalized $\omega_{3\text{dB}}$	Response
0	1.00	No shunt peaking
0.32	1.60	Optimum group delay
0.41	1.72	Maximally flat
0.71	1.85	Maximum bandwidth

Another interesting case occurs when $m = 0.32$. As seen in the phase plot, this best approximates a linear phase response up to the 3-dB bandwidth, which is 60% higher than the case without shunt peaking. This case, called the *optimum group delay case*, is desirable for optimizing pulse fidelity in broad-band systems that transmit digital signals and is used in the prototype preamplifier described in Sections V and VI.

III. On-Chip Shunt-Peaking

The nonidealities of on-chip inductors present several challenges for implementing monolithic gigahertz circuitry. In shunt-peaking applications, the biggest issue is the reduction in bandwidth improvement because of the additional parasitic capacitance introduced by the on-chip inductor. On-chip inductors are usually realized using bondwires or on-chip spirals. Although bond wires exhibit much higher quality factors (Q) than spiral inductors, their use is constrained by the limited range of realizable inductances and large production fluctuations. Furthermore, the bondpad capacitance associated with the bondwire can degrade performance. This capacitance is typically ≈70–200 fF, which is significant for gigahertz circuitry considering that the maximum realizable inductance is only ≈4 nH. Thus, although bondwires have been used as shunt-peaking elements, the net improvement in bandwidth is only ≈10–15% [2]. Moreover, differential implementations of shunt-peaked amplifiers experience a degradation in power-supply rejection ratio (PSRR) because of the inductance mismatch between the two bondwires.

On the other hand, spiral inductors exhibit good matching and are therefore suitable for differential architectures. Furthermore, they permit a large range of inductances to be realized. However, they possess smaller Q values and are more difficult to model.

A. Modeling of Spiral Inductors

The inductance as well as the parasitic elements of a spiral inductor are determined by both the parameters of the inductor and the process parameters. The geometry of a polygonal spiral is defined by the following lateral parameters: the outer diameter d_{out}, the conductor width w, the conductor spacing s, the number of turns n, and the number of sides in the polygon p [3]. While square spirals $(p = 4)$ are the most popular because of the ease of their layout, hexagonal $(p = 6)$, octagonal $(p = 8)$ spirals have also been commonly used [4].

The parasitic elements of an on-chip spiral entail important engineering tradeoffs. For example, while a large w and a large d_{out} are desirable for minimizing the series resistance of a spiral, a small w and a small d_{out} are required to minimize the spiral's area (as well as its parasitic capacitance). A small d_{out} is also desirable to minimize the resistive loss due to magnetic coupling to the substrate [5]. Thus, spirals need to be modeled properly to permit the designer to choose the optimal inductor for a given application.

While field solvers can model spirals accurately, they are inconvenient because they cannot be incorporated in a standard circuit design environment (such as SPICE). Thus, significant work has gone into modeling spiral inductors using lumped circuit models [6], [4], [7], [8]. In this paper, we use the simple and well-accepted lumped π model proposed in [6].

Although the parasitic resistors and capacitors in this model have simple physically intuitive expressions, the inductance value itself lacks a simple, accurate expression. The inductance has typically been calculated using the Greenhouse method [6], [9], [10]. Since this method operates by summing the self and mutual inductances of the segments of the spiral using the method of moments, the complexity of the calculation goes up as the square of the product or the number of sides and the number of turns. Thus, although the Greenhouse method offers sufficient accuracy and adequate speed, it cannot provide an inductor design directly from specifications and is cumbersome for initial design. We overcome this limitation by using the simple, analytical inductance expressions described in [11], which exhibit typical errors of ≈2–3% when compared to experimental data and field solvers. When combined with the lumped π model described in [6], these expressions allow the engineer to obtain design insight and explore tradeoffs quickly and easily. Thus, field solvers are now only needed to verify the final design.

B. Shunt Peaking with Spiral Inductors

In this section, we discuss a new design methodology for integrated shunt-peaked amplifiers that minimizes the adverse effects due to the nonidealities of an on-chip spiral.

Fig. 3 illustrates how the inductor Q can be made irrelevant by partitioning the total load resistance R between the inductor's series resistance (R_s) and the external resistance, which now takes on the value of $(R - R_s)$. Now that the series resistance of the spiral is part of the load, the inductor's turn width w and spacing s can be minimized to permit the desired inductance to be realized while minimizing the spiral area and capacitance. The minimum width w is now determined by current density considerations, while the minimum turn spacing s is set by lithography limitations.

The desired inductance is now a function of both the transistor's and inductor's parasitics as well as the load capacitance and external resistance. Thus the optimization of on-chip shunt peaking requires the simultaneous optimization of passive and active components. We facilitate this optimization problem by using the π lumped model and inductance expressions discussed earlier and by using a patterned ground shield (PGS) to eliminate the resistive and capacitive coupling to the substrate [12]. Although magnetic coupling to the substrate still exists (via eddy currents), it is not significant for frequencies up to 2 GHz [5], especially since the design methodology described in this paper allows the area of the spiral to be dramatically reduced.

In many inductor circuits, one terminal of the spiral is connected to incremental ground. In such cases, the spiral (with PGS) is well represented by the elements within the dashed box in Fig. 3 [12], [13]. Now, the inductor design problem boils down to choosing the geometrical parameters of the spiral such that the desired inductance is obtained while minimizing the capacitance C_{ind}. The next section discusses the optimization method used in this process.

IV. Optimization Via Geometric Programming

This section presents an efficient method for the optimal design and synthesis of RF CMOS inductor circuits. The method is based on *geometric programming*.

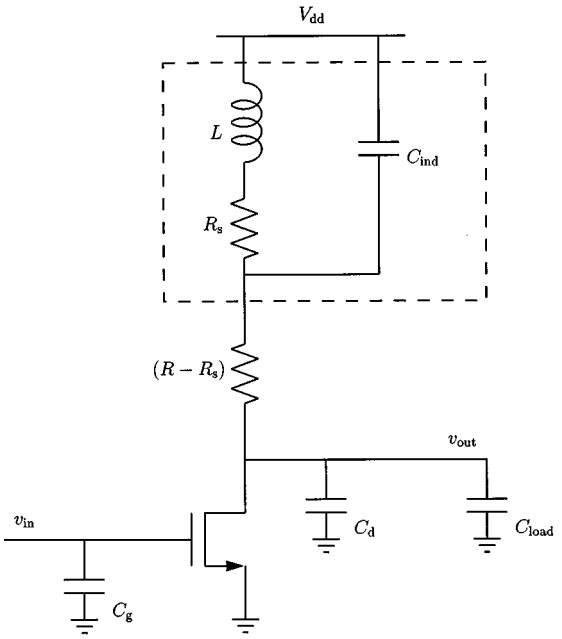


Fig. 3. Shunt peaking with optimized on-chip inductor.

A. Geometric Programming

Let g be a real-valued function of n real, positive variables $x_1, x_2, \ldots, x_n$. It is called a *monomial* function if it has the form

$$g(x_1, \ldots, x_n) = c x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}$$

where $c \geq 0$ and $\alpha_i \in \mathbf{R}$. A *posynomial* is a sum of *monomials*. Thus, for example, $2.3(x_1/x_2)^{1.5}$ is a monomial and $0.7 + 2x_1/x_3^2 + x_2^{0.3}$ is a posynomial. Posynomials are closed under sums, products, and nonnegative scaling.

A GP has the form

$$\begin{array}{lll} \text{minimize} & f_0(x) & \\ \text{subject to} & f_i(x) \leq 1, & i = 1, 2, \ldots, m \\ & g_i(x) = 1, & i = 1, 2, \ldots, p \\ & x_i > 0, & i = 1, 2, \ldots, n \end{array} \tag{3}$$

where f_i are posynomial functions and g_i are monomial functions. If f is a posynomial and g is a monomial, then the constraint $f(x) \leq g(x)$ can be expressed as $f(x)/g(x) \leq 1$ (since f/g is posynomial). From closure under nonnegativity, constraints of the form $f(x) \leq a$, where $a > 0$, can also be used. Similarly, if g_1 and g_2 are both monomial functions, the constraint $g_1(x) = g_2(x)$ can be expressed as $g_1(x)/g_2(x) = 1$ (since g_1/g_2 is monomial).

For our purposes, the most important feature of geometric programs is that they can be *globally* solved with great efficiency. GP solution algorithms also determine whether the problem is infeasible. Also, the starting point for the optimization algorithm does not have any effect on the final solution; indeed, an initial starting point or design is completely unnecessary. More information on geometric programmming can be found in [15].

B. Optimization of Inductor Circuits

Several circuit design problems may be posed as geometric programs [16], [17]. In particular, the design specifications of inductor circuits can be formulated in a way suitable for geometric programming [13]. For a polygonal spiral with a given number of sides (which in our case is four), the design variables that characterize the inductor are d_{out}, w, s, n, and d_{avg}. Note that this is a redundant set of variables as $d_{\text{avg}} = d_{\text{out}} - nw - (n-1)s$. Some of the design variables are discrete for practical designs. For example, n is restricted to be an integer multiple of 0.25 for a square spiral. During optimization, we ignore these rounding restrictions and consider the variables to be continuous. After optimization, the design variables are rounded to the closest grid point. We have not observed any significant error because of the rounding operation.

Each element in the inductor's lumped circuit model is a posynomial function of the design variables and multiplicative constants (which are determined by the frequency and the technology). For example, the inductance of the spiral is given by a monomial expression that has the form [3]

$$L_{\text{mon}} = \beta d_{\text{out}}^{\alpha_1} w^{\alpha_2} d_{\text{avg}}^{\alpha_3} n^{\alpha_4} s^{\alpha_5} \tag{4}$$

where the coefficients β and α_i are determined by the number of sides of the polygon. For a square, $\beta = 1.62 \cdot 10^{-3}$, $\alpha_1 = -1.21$, $\alpha_2 = -0.147$, $\alpha_3 = 2.40$, $\alpha_4 = 1.78$, and $\alpha_5 = -0.030$.

The key to optimizing inductor circuits using geometric programming is to formulate the design specifications as monomial and poynomial functions of the design variables (of the inductors as well as the transistors) in a manner conforming to (3). Such formulations for a variety of inductor circuits (including shunt-peaked amplifiers) are presented in [13], [18], and [19]. In this paper, we outline the optimization methodology and focus on the circuit implementation details.

V. Design Example

This section illustrates how optimized on-chip spiral inductors can improve the performance of a preamplifier intended for the front end of a gigabit optical system. Fig. 4 shows the block diagram of a typical optical communication receiver. The key performance parameters of such a front end are bandwidth, sensitivity, stability, and dynamic range. The system's bandwidth and sensitivity are determined largely by the preamplifier ([20]–[22]). While a high bandwidth demands a small input resistance, good sensitivity requires the resistors in the signal path to be large in order to minimize thermal noise. Thus, the preamplifier is typically implemented using a transimpedance architecture, as it provides a large bandwidth by synthesizing a small input resistance using a much larger feedback resistor.

A. Transimpedance Limit

Fig. 5(a) illustrates the main elements of a transimpedance preamplifier. Assuming that the bandwidth of the amplifier is set by the input pole, we obtain

$$\omega_{3\,\text{dB}} = \frac{1}{R_{\text{in}} C_{\text{in}}} \tag{5}$$

where $\omega_{3\,\text{dB}}$ is the 3-dB bandwidth of the circuit, R_{in} is the input resistance, and C_{in} is the total input capacitance.

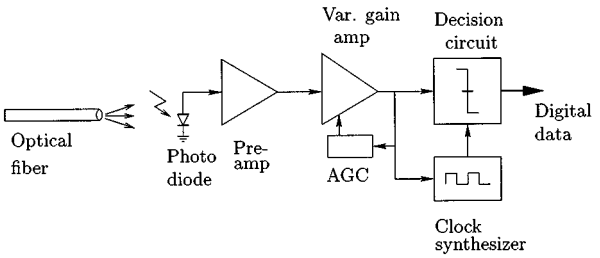


Fig. 4. System overview.

R_{in} is given by

$$R_{\text{in}} = \frac{R_f}{A+1} \approx \frac{R_f}{A} \tag{6}$$

where R_f is the feedback resistance and A is the open-loop gain of the amplifier. The approximate expression is valid when $A \gg 1$.

C_{in} is given by

$$C_{\text{in}} = C_g + C_D \tag{7}$$

where C_g is the input gate capacitance of the amplifier and C_D is the sum of the capacitances of the active area of the photo diode as well as associated parasitic capacitances (arising from bondpads, etc.). In GaAs implementations, C_D can be kept small by integrating the photo diode and the preamplifier on the same die. In such technologies, a C_D as small as 50 fF is common [23]. Silicon bipolar and CMOS implementations are not so fortunate: a C_D of ≈300–600 fF is typical.

The gain bandwidth product determines the maximum available gain for a given bandwidth. Denoting the transition frequency as ω_T, we relate the gain A to the 3-dB bandwidth $\omega_{3\,\text{dB}}$

$$A \approx \frac{\omega_T}{\omega_{3\,\text{dB}}}. \tag{8}$$

Substituting (6)–(8) into (5), we obtain a maximum achievable transimpedance $R_{f,\max}$

$$R_{f,\max} \approx \frac{\omega_T}{\omega_{3\,\text{dB}}^2 (C_g + C_D)}. \tag{9}$$

Noting that the transconductance g_m of the input stage is related to C_g by $g_m \approx \omega_T C_g$, and that for optimum sensitivity $C_g \approx C_D$, we conclude that the maximum achievable transimpedance is determined by the system bandwidth specification, the total input capacitance, and the process constant ω_T.

B. Circumventing the Transimpedance Limit

Fig. 5(b) illustrates a modified preamplifier architecture that circumvents the transimpedance limit. The transimpedance stage is decoupled from the photo diode by a common-gate stage, and the gain-bandwidth product of the transimpedance stage is enhanced by shunt peaking. Now, the sensitive feedback node of the transimpedance stage is more robust as its poles are not determined by any off-chip components. Furthermore, the common-gate stage permits the transistors of

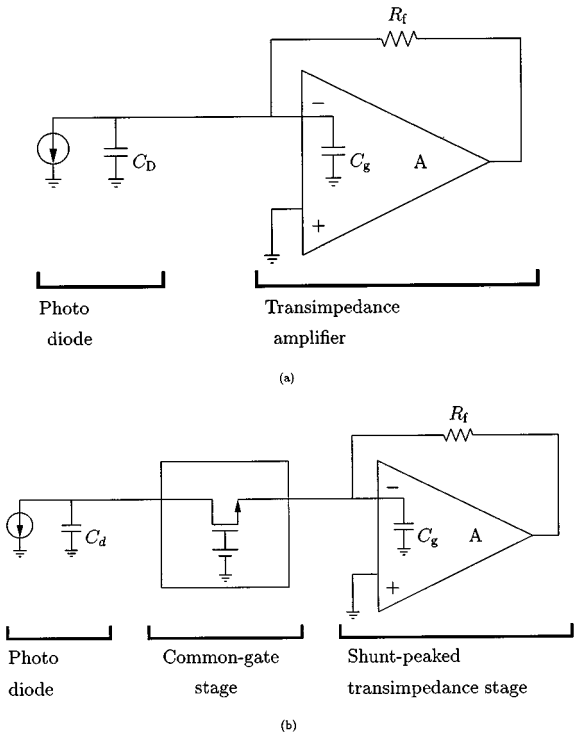


Fig. 5. (a) Conventional preamplifier architecture. (b) Modified architecture with a common-gate stage preceding the shunt-peaked transimpedance stage.

the transimpedance stage to be sized smaller, enabling a higher transimpedance to be achieved.

Note that the common-gate stage is not necessary to obtain the benefits of shunt peaking. If desired, one could connect the photo diode directly to a shunt-peaked transimpedance stage. Such an implementation is particularly attractive for applications that demand the best achievable sensitivity for a given power. However, such an implementation requires the parasitic impedance of the photo diode to be known so that the transimpedance stage can be sized for optimal performance. The introduction of the common-gate stage offers an additional degree of flexibility for the designer and permits stable operation over a wider range of photo-diode capacitances. This is valuable in cases (such as our prototype) where the capacitance of the photo-diode structure is not known in advance. The drawback of the common-gate source is the degradation in the high-frequency noise performance due to the source junction capacitance of the common-gate transistor, an issue that will be addressed more in Section V-E.

C. Shunt-Peaked Transimpedance Stage

Fig. 6 illustrates the shunt-peaked transimpedance stage. The cascode eliminates the bandwidth degradation due to the Miller capacitance of the common-source stage's gate-drain capacitance. This degradation is particularly significant in CMOS circuits, where the gate-drain capacitance can be as high as one-third of the gate-source capacitance. The cascode also enhances the overall gain by increasing the stage's output impedance.

The dominant pole in the amplifier occurs at the drain of the cascode transistor. The bandwidth of the amplifier is improved by applying shunt peaking at this node. The inductors, resistors,

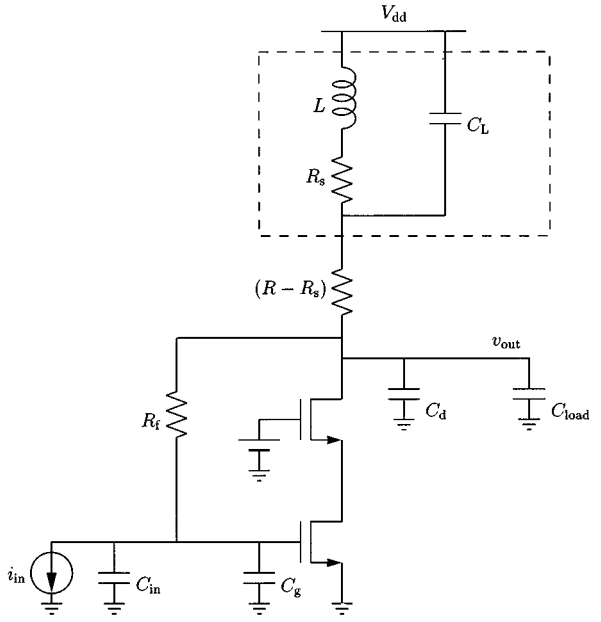


Fig. 6. Shunt-peaked transimpedance stage.

and transistors are sized for optimum group delay over the signal bandwidth. The design methodology can be summarized as follows:

1) Design and optimize transimpedance stage for desired signal bandwidth without shunt peaking.
2) Design optimal spiral inductor:
   a) use transistor current and interconnect current density specification to determine inductor's turn width w;
   b) determine minimum turn spacing s from process specifications;
   c) choose number of turns n and outer diameter d_{out} to realize optimum L while minimizing parasitic capacitance and area.
3) Increase the transimpedance resistance R_f and the total load resistance R.

In a manual design process, a few iterations may be required over these steps to arrive at the optimal design. This is because the presence of feedback, along with the presence of inductor and transistor parasitics, results in a more complex frequency response compared to the single-zero, double-pole system described earlier. Nevertheless, the simple treatment discussed earlier serves as an excellent starting point and permits the engineer to rapidly converge upon the final design.

The availability of a lumped inductor model (with analytical expressions for all elements including the inductance) and the use of geometric programming allows this entire design and optimization process to be automated so that no iteration is needed on the part of the designer. The geometric programming approach enables all three of the above steps to be accomplished simultaneously (with a few simplifying approximations), thereby yielding a globally optimal solution very quickly.

In our case, this optimization method allowed a 20-nH inductor to be realized with an outer diameter of only 180 μm. This inductor had 11.75 turns, a width of 3.2 μm, and a spacing of 2.1 μm, and was implemented on the third (top) metal layer with thickness 2.1 μm. The shunt peaking yielded a 40% increase in the transimpedance of this stage (for a fixed signal bandwidth) with no additional power dissipation. Alternatively, the shunt peaking could have been used to increase the signal bandwidth for a fixed transimpedance.

D. Differential Architecture

Compared to differential architectures, single-ended architectures consume less power, take up less die area, and exhibit better noise performance. However, at high frequencies, they are susceptible to supply noise and are plagued by stability problems stemming from parasitic feedback paths. By providing good common-mode rejection, differential architectures circumvent these disadvantages, and are therefore preferred in systems where the integration of the analog and digital functions is the ultimate goal. In keeping with this premise, the architecture described here is fully differential and provides complementary outputs, which is a necessity given that high-speed digital and clocking circuitry operate in differential mode.

Fig. 7 shows the schematic of the complete prototype preamplifier. The common-gate (CG) stage is followed by the common-source (CS) transimpedance stage, whose output goes to a source follower that buffers the output driver. The output driver is only needed for testing purposes and is not needed nor desired in a system where the analog and digital components of a receiver are integrated.

The chip consumes a total of 225 mW, of which the 50-Ω output driver consumes 110 mW. For optimum sensitivity, the total power consumption of the common-gate and common-source stages is roughly proportional to the photo-diode capacitance. This preamplifier has been designed to operate with an external capacitance as large as 600 fF. The need to support such a large capacitance arises because the photo diodes are external to the chip with correspondingly large bondpad capacitances. Recent research has explored flip-chip bonding techniques for reducing the capacitance loading of the front end to less than 100 fF. Such a low input capacitance would permit a higher input impedance and therefore allow smaller devices to be used throughout the preamplifier, resulting in a substantial power saving, while retaining the same bandwidth and improving sensitivity. Alternatively, the reduced capacitance would allow the design of preamplifiers with increased bandwidth supporting faster baud rates.

E. Noise Considerations

The sensitivity of the preamplifier is usually expressed as the equivalent integrated input-referred current noise density. Significant work has gone into deriving the minimum noise conditions for conventional optical preamplifiers [20], [24]. Some studies have also investigated how inductors can increase the sensitivity of optical preamplifiers implemented in GaAs [21].

The noise performance of the CG input stage followed by the CS transimpedance stage has been studied in GaAs HBT and BiCMOS processes [25]. Although a simulation involving a single-ended CMOS version was reported, it ignored the effects of the source and drain junction capacitances and did not

consider the impact of short-channel effects on small signal behavior and noise [26].

Junction capacitances in submicrometer CMOS processes are comparable to the gate capacitances and therefore significantly influence both noise behavior and bandwidth. A rigorous analysis that includes the impact of the junction capacitances and short-channel behavior yields two conditions for a noise optimum. First, the saturation-mode gate capacitance of the common-source stage must equal the saturation-mode drain capacitance of the common-gate stage so that $C_{\mathrm{gs,CS}} + C_{\mathrm{gd,CS}} = C_{\mathrm{gd,CG}} + C_{\mathrm{db,CG}}$. Second, the saturation-mode input capacitance of the common-gate stage (which is the gate-source capacitance $C_{\mathrm{gs,CG}}$ plus the source-substrate capacitance $C_{\mathrm{sb,CG}}$) must equal βC_{ext}, where $\beta \approx 0.8 - 1$. β is a function of both devices' ω_T, their coefficients of channel thermal noise (γ), and their ratios of junction capacitance to gate capacitance, all of which are bias dependent. For a typical CMOS device in saturation, $(C_{\mathrm{gs}} + C_{\mathrm{sb}})$ is around three to four times as big as $(C_{\mathrm{gd}} + C_{\mathrm{db}})$, and therefore the common-source stage can now be sized smaller, allowing a corresponding increase in the feedback resistance and a dramatic decrease in power consumption, while retaining the same device f_T. The buffer stage that follows the common-source stage can also be sized smaller, as can the width of all the interconnects, resulting in a smaller die area. The transconductance of the common-gate stage only needs to be large enough to ensure that the input pole is nondominant, enabling the power consumption of the first stage to be small.

The introduction of the common-gate stage introduces three new noise sources: the thermal noise of the source resistor, the thermal noise of the drain resistor, and the thermal channel noise of the common-gate transistor. Of these terms, careful design ensures that the resistors are made large enough so as not to significantly affect the noise performance. The thermal channel noises of the common-gate and common-source devices are reflected at the input by equivalent current noise spectral densities proportional to the square of the frequency. When integrated over frequency, these thermal channel noise terms dominate, a behavior typical of short-channel CMOS processes, where carrier velocity saturation conditions cause thermal channel noise to increase due to excess noise stemming from hot electron effects [27]. Balancing that degradation is the continuing reduction in gate length delivered by higher f_T CMOS devices, thereby improving noise performance. However, carrier velocity saturation causes the small signal transconductance (and f_T) to be smaller than that predicted by long-channel (square-law) approximations.

VI. Layout and Experimental Details

As shown in the die photo (Fig. 8), the chip area is dominated by the passive components, which is typical of RFIC's. However, the two inductors combine for less than 15% of the total area, thanks to the optimized shunt-peaking technique described in the earlier sections. A patterned ground shield is used beneath the inductors to reduce substrate coupling [12]. Differential symmetry and cross quad layout are used to ensure maximum matching, thereby reducing common-mode noise and systematic offset. On-chip capacitance of 16 pF is used to provide supply decoupling. Several substrate contacts, placed around the transistors, minimize source inductance. The floor plan keeps

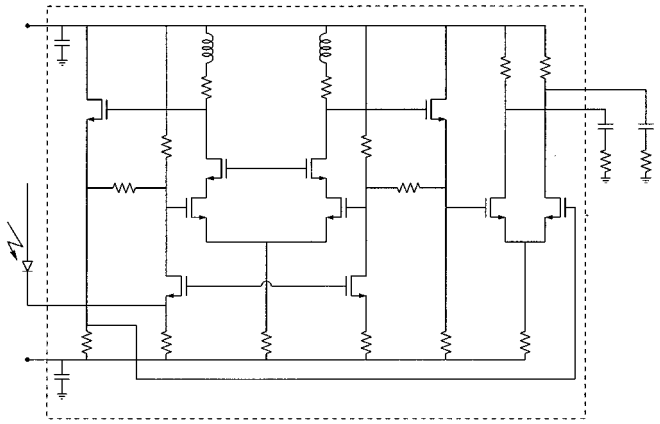


Fig. 7. Simplified circuit diagram.

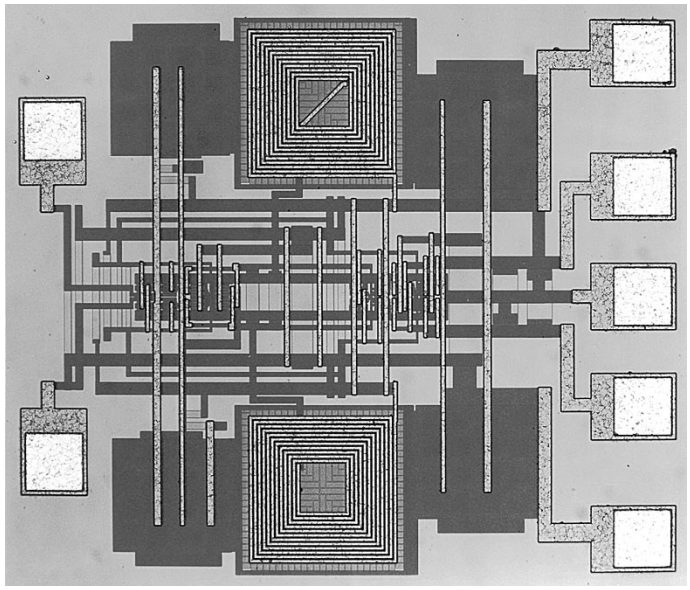


Fig. 8. Preamplifier die photo.

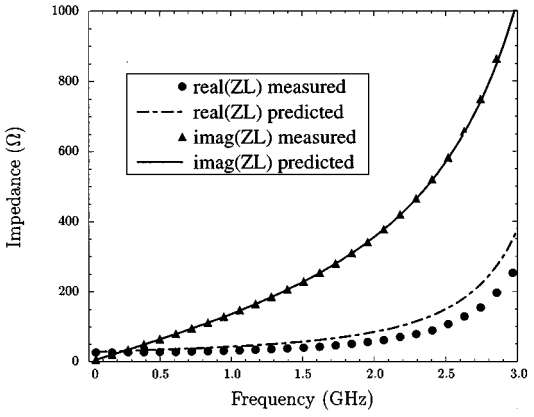


Fig. 9. Simulated and measured one-port impedance of the spiral inductor used for shunt peaking: $d_{\mathrm{out}} = 180\ \mu$m, $n = 11.75$ turns, $w = 3.2\ \mu$m, $s = 2.1\ \mu$m, and $t = 2.1\ \mu$m with $L = 20$ nH.

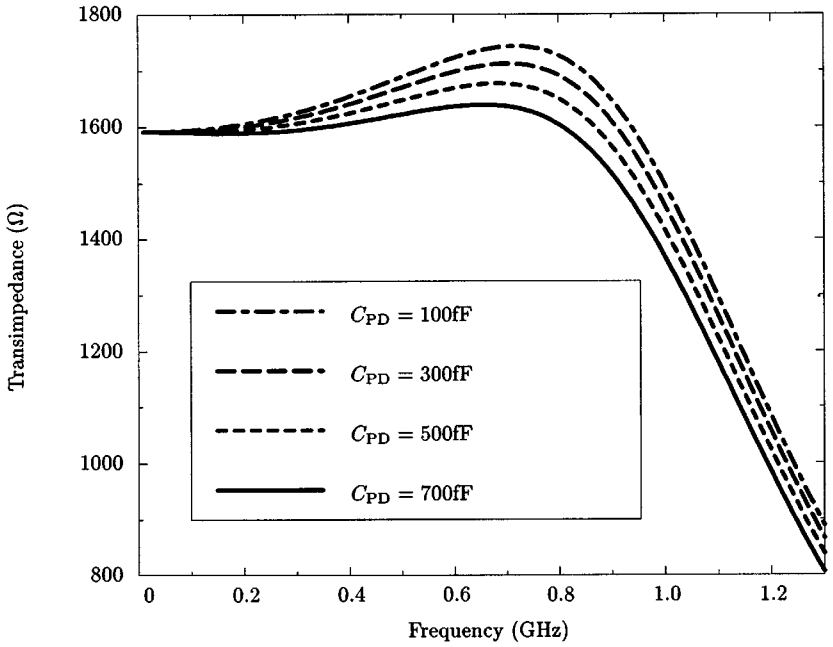


Fig. 10. Simulated transimpedance versus frequency.

the sensitive input bondpads as far away from the other pads as possible.

The S-parameters of the inductor were measured using coplanar ground-signal-ground (GSG) probes and an open calibration structure. The inductance and the one-port impedance (which is the relevant measure in our amplifier) were extracted from these measurements.

As shown in Fig. 9, good agreement between the prediction of the lumped circuit inductor model and measured data is obtained for the equivalent one-port impedance of the spiral inductor used for shunt peaking. In particular, we note that the measured inductance of 20.5 nH matches the 20.3-nH value predicted by our simple inductance expressions to within a 1% error.

Fig. 10 shows the preamplifier's simulated transimpedance versus frequency for photo-diode capacitances varying from 100 to 700 fF. As can be seen, the 3-dB bandwidth is around 1.2 GHz and only weakly dependent on the photo-diode capacitance. Maximum gain peaking is 1 dB. These simulations are run with the output driving a 50-Ω resistance and 1-pF capacitance.

Fig. 11(a) and (b) displays the measured single-sided output eye diagrams for operation at 2.1 and 1.6 Gbaud, respectively. An open eye is obtained for single-sided output voltages extending from 4 to 500 mV. Table II summarizes the performance of the prototype chip.

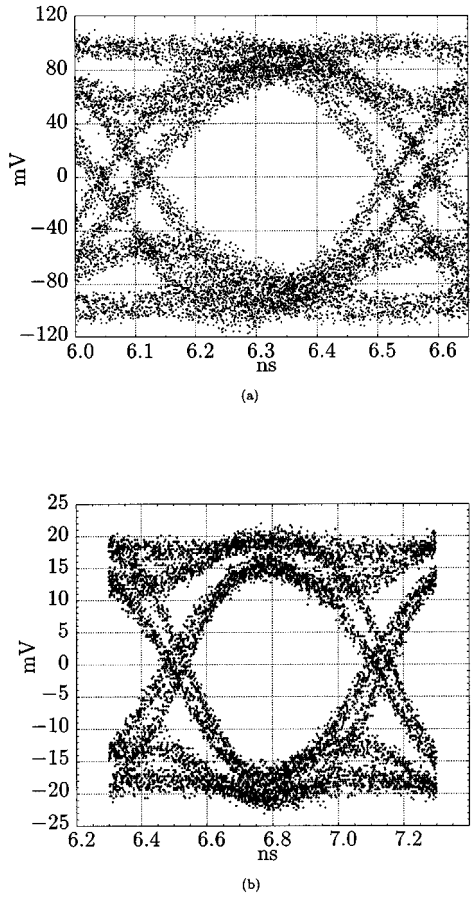


Fig. 11. Measured output eye diagrams at (a) 2.1 and (b) 1.6 Gbaud.

VII. Conclusion

This paper presented an area- and power-efficient technique for boosting the bandwidth of broad-band systems using optimized on-chip inductors as shunt-peaking elements. Simple, accurate inductance expressions were used in a lumped circuit inductor model to facilitate circuit design. The analytical expressions of this inductor model permitted inductor circuit problems to be posed as *geometric programs* so that globally optimal solutions could be obtained easily, with no iteration needed on the part of the designer.

TABLE II
PERFORMANCE SUMMARY

Transimpedance (small-signal)	1600Ω (differential)
	800Ω (single-ended)
Bandwidth (3dB)	1.2GHz
Max. photodiode capacitance	0.6pF
Max. input current	1.0mA
Simulated input noise current	0.6μA
Max. output voltage swing	1.0V_{pp} (differential)
(50Ω load at each output)	0.5V_{pp} (single-ended)
Power consumption	115mW (core)
	110mW (50Ω driver)
Die area	0.6mm^2
Technology	0.5μm CMOS

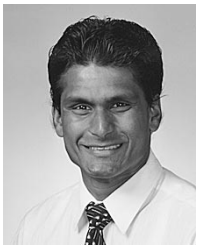
These bandwidth extension and circuit optimization techniques were applied in the implementation of a 2.125-Gbaud, 1.6-kΩ differential transimpedance preamplifier with an equivalent input current noise of 0.6 μA. The chip has a die area of 0.6 mm^2, of which less than 15% is consumed by the two inductors. Designed in a triple-metal, single-poly, 0.5-μm CMOS process, this chip was intended as a test vehicle to demonstrate how on-chip bandwidth extension techniques can push the limits of low-cost CMOS processes. To the best of the authors' knowledge, this chip is the first CMOS amplifier to use on-chip planar, spiral inductors for bandwidth enhancement.

ACKNOWLEDGMENT

The authors would like to thank Rockwell International for fabricating the chips. Particularly, they are obliged to Dr. C. Hull and Dr. P. Singh for invaluable discussions on layout and processing. They would also like to acknowledge Dr. A. Lu, Dr. T. Hofmeister, and Prof. L. Kazovksy for help with the optical measurements and Dr. C. P. Yue for helpful discussions on spiral inductors.

REFERENCES

[1] T. H. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*. Cambridge, U.K.: Cambridge Univ. Press, 1998.
[2] T. Yoon and B. Jalali, "Front-end cmos chipset for fiber-based gigabit ethernet," in *Symp. VLSI Circuits Dig. Tech. Papers*, June 1998, pp. 188–191.
[3] S. S. Mohan, M. Hershenson, S. P. Boyd, and T. H. Lee, "Simple accurate expressions for planar spiral inductances," *IEEE J. Solid-State Circuits*, vol. 34, no. 10, pp. 1419–1424, 1999.
[4] J. Crols, P. Kinget, J. Craninckx, and M. Steyeart, "An analytical model of planar inductors on lowly doped silicon substrates for analog design up to 3 GHz," in *Symp. VLSI Circuits Dig. Tech. Papers*, 1996, pp. 28–29.
[5] S. S. Mohan, "The design, modeling, and optimization of on-chip inductor and transformer circuits," Ph.D. dissertation, Stanford Univ., Stanford, CA, Dec. 1999.
[6] C. P. Yue, C. Ryu, J. Lau, T. H. Lee, and S. S. Wong, "A physical model for planar spiral inductors on silicon," in *Proc. IEEE IEDM'96*, 1996 , pp. 155–158.
[7] J. N. Burghartz, K. A. Jenkins, and M. Soyuer, "Multilevel-spiral inductors using VLSI interconnect technology," *IEEE Electron Device Lett.*, vol. 17, pp. 428–430, Sept. 1996.
[8] R. B. Merrill, T. W. Lee, H. You, R. Rasmussen, and L. A. Moberly, "Optimization of high Q integrated inductors for multi-level metal CMOS," in *Proc. IEEE IEDM'95*, 1995, pp. 38.7.1–38.7.4.
[9] K. B. Ashby, I. A. Koullias, W. C. Finley, J. J. Bastek, and S. Moinian, "High Q inductors for wireless applications in a complementary silicon bipolar process," *IEEE J. Solid-State Circuits*, vol. 31, pp. 4–9, Jan. 1996.
[10] H. M. Greenhouse, "Design of planar rectangular microelectronic inductors," *IEEE Trans. Parts, Hybrids, Packag.*, vol. PHP-10, pp. 101–109, June 1974.
[11] S. S. Mohan, "Modeling, design, and optimization of on-chip inductors and transformers," Ph.D. oral examination, Electrical Engineering Dept., Stanford Univ., Stanford, CA, June 1999.
[12] C. P. Yue and S. S. Wong, "On-chip spiral inductors with patterned ground shields for Si-based RF IC's," in *1997 Symp. VLSI Circuits Dig. Tech. Papers*, 1997, pp. 85–86.
[13] M. Hershenson, S. S. Mohan, S. P. Boyd, and T. H. Lee, "Optimization of inductor circuits via geometric programming," in *Proc. Design Automation Conf.*, session 54.3, June 1999, pp. 994–998.
[14] R. J. Duffin, E. L. Peterson, and C. Zener, *Geometric Programming—Theory and Applications*. New York: Wiley, 1967.
[15] S. P. Boyd and L. Vandenberghe, "Course reader for EE364: Introduction to convex optimization with engineering applications," Stanford Univ., Stanford, CA, 1997.
[16] S. Boyd and L. Vandenberghe, "Introduction to convex optimization with engineering applications (course notes),", Stanford Univ., Stanford, CA, 1997.
[17] M. Hershenson, S. Boyd, and T. H. Lee, "GPCAD: A tool for CMOS op-amp synthesis," in *Proc. IEEE/ACM Int. Conf. Computer Aided Design*, San Jose, CA, Nov. 1998, pp. 296–303.
[18] M. Hershenson, A. Hajimiri, S. S. Mohan, S. P. Boyd, and T. H. Lee, "Design and optimization of LC oscillators," in *Proc. IEEE/ACM Int. Conf. Computer Aided Design*, San Jose, CA, Nov. 1999.
[19] M. del Mar Hershenson, "CMOS analog circuit design via geometric programming," Ph.D. dissertation, Stanford Univ., Stanford, CA, Dec. 1999.
[20] R. G. Smith and S. D. Personik, "Receiver design for optical communication systems," in *Topics in Applied Physics: Semiconductor Devices for Optical Communication*, 2nd ed. Berlin, Germany: Springer-Verlag, 1982, vol. 39, pp. 89–160.
[21] N. Scheinberg, R. J. Bayruns, and T. M. Laverick, "Monolithic GaAs transimpedance amplifiers for fiber-optic receivers," *IEEE J. Solid-State Circuits*, vol. 26, pp. 1834–1839, Dec. 1991.
[22] S. S. Mohan and T. H. Lee, "A 2.125 Gbaud 1.6 kω transimpedance preamplifier in 0.5 μm CMOS," in *Custom Integrated Circuits Conf. Dig. Tech. Papers*, May 1999, pp. 513–516.
[23] Z. Lao , M. Berroth, V. Hurm, A. Thiede, R. Bosch, P. Hofmann, A. Hulsmann, C. Moglestue, and R. Kohler, "25 Gb/s AGC amplifier, 22 GHz transimpedance amplifier and 27.7 GHz limiting amplifier ics using AlGaAs/GaAs-hemts," in *ISSCC Dig. Tech. Papers*, vol. 40, 1997, pp. 356–357.
[24] A. A. Abidi, "On the choice of optimum FET size in wide-band transimpedance amplifiers," *J. Lightwave Technol.*, vol. 6, pp. 64–66, Jan. 1988.
[25] T. Vanisri and C. Toumazou, "Integrated high frequency low-noise current-mode optical transimpedance preamplifiers: Theory and practice," *IEEE J. Solid-State Circuits*, vol. 30, pp. 677–685, June 1995.
[26] C. Toumazou and S. M. Park, "Wideband low noise CMOS transimpedance amplifiers for gigahertz operation," *Electron. Lett.*, vol. 32, no. 13, pp. 1194–1196, June 1996.
[27] D. K. Shaeffer and T. H. Lee, "A 1.5 V, 1.5 GHz CMOS low-noise amplifier," *IEEE J. Solid-State Circuits*, vol. 32, pp. 745–759, May 1997.



Sunderarajan S. Mohan received the B.S., M.S., and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1992, 1993, and 1999, respectively. His Ph.D. work focused on the design of radio frequency integrated circuits in CMOS with a particular emphasis on the design and optimization of on-chip spiral inductors and transformer circuits.

During the summers of 1992 and 1993, he interned at Acuson Inc., Mountain View, CA, where he designed test systems and modeled the beam-forming characteristics of ultrasound transducers for medical applications. Since the summer of 1999, he has worked at Barcelona Design Inc., Mountain View, CA, where he designs RF and analog cells. He has served as a Teaching Assistant in many undergraduate and graduate lecture and laboratory classes.

Dr. Mohan received the Centennial Award for Teaching. He was supported by an IBM FMA fellowship from 1996 to 1998.

Maria del Mar Hershenson was born in Barcelona, Spain. She received the B.S.E.E. degree from the Universidad Pontificia de Comillas, Madrid, Spain, in 1995 and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1997 and 1999, respectively.

In 1994, she was an Intern at Linear Technology Corporation, Milpitas, CA, where she worked on low-power voltage regulators. Since the summer of 1999, she has been with Barcelona Design, Inc., Mountain View, CA, where she designs analog circuits. Her research interests are RF power amplifiers and convex optimization techniques applied to the automated design of analog integrated circuits.

Dr. Hershenson received an IBM fellowship in 1998.

Stephen P. Boyd received the A.B. degree in mathematics from Harvard University, Cambridge, MA, in 1980 and the Ph.D. degree in electrical engineering and computer science from the University of California, Berkeley, in 1985.

In 1985, he joined the Electrical Engineering Department at Stanford University, Stanford, CA, where he is now a Professor and Director of the Information Systems Laboratory. His interests include computer-aided control system design and convex programming applications in control, signal processing, and circuits.

Thomas H. Lee received the S.B., S.M., and Sc.D. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 1983, 1985, and 1990, respectively.

He joined Analog Devices in 1990, where he was primarily engaged in the design of high-speed clock recovery devices. In 1992, he joined Rambus, Inc., in Mountain View, CA, where he developed high-speed analog circuitry for 500-Mbyte/s CMOS DRAM's. He has also contributed to the development of PLL's in the StrongARM, Alpha, and K6/K7 microprocessors. Since 1994, he has been an Assistant Professor of Electrical Engineering at Stanford University, Stanford, CA, where his research interests are in low-power, high-speed analog circuits and systems, with a focus on gigahertz-speed wireless integrated circuits built in conventional silicon technologies, particularly CMOS. He has received six U.S. patents and is the author of *The Design of CMOS Radio-Frequency Integrated Circuits* (Cambridge, U.K.: Cambridge Univ. Press, 1998) and a co-author of two additional books on RF circuit design.

Dr. Lee has twice received the Best Paper award at the International Solid-State Circuits Conference (ISSCC), was co-author of a Best Student Paper at ISSCC, recently won a Packard Foundation Fellowship, and was named a Distinguished Lecturer of the IEEE Solid-State Circuits Society.