Optimal Doping Profiles via Geometric Programming

Siddharth Joshi, Stephen Boyd, *Fellow, IEEE*, and Robert W. Dutton, *Fellow, IEEE*

***Abstract*—We first consider the problem of determining the doping profile that minimizes base transit time in a (homojunction) bipolar junction transistor. We show that this problem can be formulated as a geometric program, a special type of optimization problem that can be transformed to a convex optimization problem, and therefore solved (globally) very efficiently. We then consider several extensions to the basic problem, such as accounting for velocity saturation, and adding constraints on doping gradient, current gain, base resistance, and breakdown voltage. We show that a similar approach can be used to maximize the cutoff frequency, taking into account junction capacitances and forward transit time. Finally, we show that the method extends to the case of heterojunction bipolar junction transistors, in which the doping profile, as well as the profile of the secondary semiconductor, are to be jointly optimized.**

***Index Terms*—Base doping profile, base transit time minimization, cutoff frequency maximization, geometric programming, Ge-profile optimization, optimal doping profile.**

I. Introduction

The BASE transit time τ_B is an important parameter in determining the speed and the high frequency response of a bipolar junction transistor (BJT). The base transit time is a function of the doping profile; minimizing base transit time, by proper choice of doping profile, is a well studied problem [29]. Methods that have been proposed include iterative schemes [19], variational calculus [30], and optimal control [25], [26], which are compared in [30]. None of these methods, however, can guarantee (global) optimality of the resulting doping profiles.

In this paper we introduce a new method for finding the doping profile that minimizes the base transit time, subject to constraints. The method is based on formulating the problem as a *geometric program* (GP), a special type of mathematical optimization problem. Recently developed numerical methods for GPs can solve even large scale problems very efficiently, always guaranteeing that the global optimum is found [6], [7], [15]. In particular, the method described in this paper finds the *globally optimal doping profile* (for the models used).

The method described here can in addition handle a variety of practical constraints, such as limits on the doping concentration and its gradient, current gain, base resistance, and breakdown voltage. The method also extends to the problem of minimizing base transit time in heterojunction bipolar transistors (HBTs), as studied in [3], [4], [17], [18]. In the case of SiGe HBT, our approach applies to several problems: determining the base doping profile, given the Ge-profile; determining the Ge-profile, given the base doping profile; and the *joint optimization problem*, i.e., jointly optimizing both profiles, to minimize base transit time. This joint optimization problem has not been addressed in the literature.

We start with the model of base transit time in a homojunction BJT, and show how to formulate the basic optimal doping profile problem as a geometric program. In Section III we describe some model refinements, e.g., taking velocity saturation into account, as well as various additional constraints that can be added. In Section IV we consider the problem of finding the optimal base doping profile to maximize the cutoff frequency of the device, and show that it too can be (at least approximately) formulated as a GP. In Section V we consider minimization of base transit time in HBTs. We show the joint profile optimization problem, where we optimize both the Ge-profile and the base doping profile to minimize base transit time, can be transformed to a GP.

Our approach is to formulate these various doping profile optimization problems in GP form. We refer the reader to [6] for an introduction to geometric programming, some of the basic approaches used to formulate problems in GP form, a number of examples, and an extensive list of references. For background on geometric programming and its relation to convex optimization, we refer the reader to [7, Sect. 4.5]. Geometric programming has been used to solve a variety of circuit design problems, including sizing of analog circuits [10], [13], [21], [33], digital circuits [5], [27], mixed-signal circuits [8], [12], and RF circuits [14], [22] (for a more complete list, see, e.g., [6]).

Geometric programs can be transformed to convex optimization problems, and therefore solved globally and efficiently, for example by recently developed interior-point methods (see, e.g., [7]). Sparse GPs (i.e., problems in which each constraint depends on only a few variables) with tens of thousands of variables and hundreds of thousands of constraints can be solved in minutes, on a small personal computer; problems with thousands of variables can be solved in seconds. Moreover, the solution found by these methods is guaranteed to be the *global solution*. This is in contrast to most numerical optimization methods, which typically produce a locally optimal solution, which need not be the global solution.

II. Base Transit Time Optimization

A. Base Transit Time Model

The doping profile, denoted $N_A(x)$, is a positive function of a space variable x over the interval $0 \leq x \leq W_B$, where W_B

Manuscript received March 25, 2005; revised August 15, 2005. This work was supported by the Stanford Graduate Fellowship, by the MARCO MSD Center, by the National Science Foundation under Grant 0423905 and (through October 2005) Grant 0140700, by the Air Force Office of Scientific Research under Grant F49620-01-1-0365, by the MARCO Focus Center for Circuit and System Solutions under Contract 2003-CT-888, and by the MIT DARPA under Contract N00014-05-1-0700. The review of this paper was arranged by Editor J. N. Burghartz.

The authors are with the Department of Electrical Engineering, Stanford University, Stanford, CA USA (e-mail: sidj@stanford.edu; boyd@stanford.edu; dutton@gloworm.stanford.edu).

Digital Object Identifier 10.1109/TED.2005.859649

0018-9383/$20.00 © 2005 IEEE

is the base width. A model for the base transit time τ_B in a homojunction BJT is given by [16]

$$\tau_B = \int_0^{W_B} \frac{n_i^2(x)}{N_A(x)} \left(\int_x^{W_B} \frac{N_A(y)}{n_i^2(y) D_n(y)} dy \right) dx \quad (1)$$

where $n_i(x)$ is the intrinsic carrier concentration and $D_n(x)$ is the carrier diffusion coefficient. This model assumes low-level injection, and neglects velocity saturation and carrier recombination in the base region, which are the commonly made assumptions. (In Section III we will consider a refined model that includes velocity saturation.)

At a point x, the intrinsic carrier concentration $n_i(x)$ and carrier diffusion coefficient $D_n(x)$ are functions of the doping concentration $N_A(x)$. The intrinsic carrier concentration $n_i(x)$ depends on the effective bandgap reduction $\Delta E_{g,\mathrm{eff}}(x)$ as [17], [19], [32]

$$n_i(x)^2 = n_{i0}^2 \exp\left(\frac{\Delta E_{g,\mathrm{eff}}(x)}{kT} \right) \quad (2)$$

where n_{i0} is the intrinsic carrier concentration in undoped silicon, k is the Boltzmann constant, and T is the temperature (Kelvin). In a homojunction BJT the effective bandgap reduction is only due to doping, i.e., $\Delta E_{g,\mathrm{eff}}(x) = \Delta E_{g,\mathrm{dop}}(x)$, which is given by

$$\Delta E_{g,\mathrm{dop}}(x) = 0.018 \log\left(\frac{N_A(x)}{N_{\mathrm{ref}}} \right) \text{ eV} \quad (3)$$

where N_{ref} is a constant. (In this formula and throughout this paper, log denotes the natural logarithm.) Substituting (3) into (2) we obtain

$$n_i(x)^2 = n_{i0}^2 \left(\frac{N_A(x)}{N_{\mathrm{ref}}} \right)^{\gamma_2} \quad (4)$$

where $\gamma_2 = 0.018/kT$. The Boltzmann constant $k = 8.617 \times 10^{-5}$ eV/K; at $T = 300$ °K, $\gamma_2 = 0.69$.

The carrier diffusion coefficient $D_n(x)$ is a complicated monotonic function of $N_A(x)$, but over the region of interest is well approximated by

$$D_n(x) = D_{n0} \left(\frac{N_A(x)}{N_{\mathrm{ref}}} \right)^{-\gamma_1} \quad (5)$$

where D_{n0} and γ_1 are positive constants. This is a simple model that captures well, the dependence of the minority carrier mobility for electron, on the doping concentration (see [17], [19], [20], and [30]). More accurate model, obtained by fitting *generalized posynomial* to the mobility data (more specifically the inverse of mobility), can be used without any change in the methodology. (For generalized posynomial fitting see [6].)

Using (4) and (5) the expression for τ_B becomes

$$\tau_B = \frac{1}{N_{\mathrm{ref}}^{\gamma_1} D_{n0}} \int_0^{W_B} N_A(x)^{\gamma_2 - 1} \left(\int_x^{W_B} N_A(y)^{1+\gamma_1-\gamma_2} dy \right) dx. \quad (6)$$

B. Basic Doping Profile Problem

The basic doping profile problem is to find the doping profile $N_A(x)$, $0 \le x \le W_B$, that minimizes the base transit time τ_B, subject to limits on the doping concentration

$$\begin{array}{ll} \text{minimize} & \tau_B \\ \text{subject to} & N_{\min} \le N_A(x) \le N_{\max}, \quad 0 \le x \le W_B. \end{array} \quad (7)$$

The problem (7) is an infinite dimensional problem, since the optimization variable is the doping profile $N_A(x)$, a function of x. The doping concentration constraint is a semi-infinite constraint, i.e., it is an infinite family of constraints, one for each $x \in [0, W_B]$.

C. Discretized Doping Profile Problem

We first discretize the base region with M uniformly spaced points, $x_i = iW_B/M$, $i = 0, 1, \ldots, M-1$. The doping profile is sampled at these points: we define $v_i = N_A(x_i)$, $i = 0, 1, \ldots, M-1$. The base transit time τ_B can then be approximated by the sum

$$\hat{\tau}_B = \frac{1}{N_{\mathrm{ref}}^{\gamma_1} D_{n0}} \left(\frac{W_B}{M} \right)^2 \sum_{i=0}^{M-1} v_i^{\gamma_2 - 1} \sum_{j=i}^{M-1} v_j^{1+\gamma_1-\gamma_2}. \quad (8)$$

The discretization error can be shown to satisfy $|\tau_B - \hat{\tau}_B| \le \eta/M$, where η is a constant; see Appendix A. This shows that by choosing M large enough, the discretization error can be made arbitrarily small. In the remainder of this paper we will work with the discretized (approximate) expression for base transit time, $\hat{\tau}_B$. To keep the notation simple, however, we will drop the hat from $\hat{\tau}_B$, and simply use τ_B to denote the discretized base transit time. Thus, from now on, we take

$$\tau_B = C \sum_{i=0}^{M-1} v_i^{\gamma_2 - 1} \sum_{j=i}^{M-1} v_j^{1+\gamma_1-\gamma_2} \quad (9)$$

where $C = W_B^2 / M^2 N_{\mathrm{ref}}^{\gamma_1} D_{n0}$.

D. Geometric Programming Formulation

Substituting τ_B in (9) into the basic optimal doping profile problem (7), we obtain the problem

$$\begin{array}{ll} \text{minimize} & \tau_B \\ \text{subject to} & N_{\min} \le v_i \le N_{\max}, \quad i = 0, 1, \ldots, M-1 \end{array} \quad (10)$$

which is a finite dimensional, nonlinear, constrained optimization problem, with variables $v_0, \ldots, v_{M-1}$. It has, however, a very special form: since τ_B in (9) is a *posynomial function* of the variables $v_0, v_1, \ldots, v_{M-1}$, this problem is a *geometric program.* (See [6] or [7, Sect. 4.5] for more on posynomials and geometric programs.) As a result, the problem (10) can be solved very efficiently, for example by interior-point methods. Moreover, such methods guarantee that the globally optimal solution is found.

We will denote the optimal doping profile by $N_A^*(x)$, and the corresponding minimum base transit time τ_B^*.

E. Sparse GP Formulation

A basic interior-point method for the problem (10) has a complexity that grows like M^3, since the objective function is not sparse (indeed, it depends on all the variables). Such a method would become slow for problems with M more than a few thousand. Fortunately we can derive an equivalent formulation of the problem (10), which has more variables and constraints, but has the very desirable feature of *sparsity*, i.e., its objective and each of the constraint function depends only on a handful of variables. A GP solver that can exploit sparsity can solve a sparse GP, even if it has more variables and constraints, far faster than an equivalent nonsparse formulation. For the optimal doping problem, the complexity of such a method grows like M, and essentially removes any practical limit on what M can be. For values of M on the order of a few thousand, the reformulated method gives extremely fast solution times, on the order of a second or two.

The key to developing a sparse formulation is to introduce some new variables to represent the sums in our expression for τ_B. We define

$$y_i = \sum_{j=i}^{M-1} v_j^{1+\gamma_1-\gamma_2}, \quad w_i = \sum_{j=i}^{M-1} v_j^{\gamma_2-1} y_j, \quad i=0,\ldots,M-1. \tag{11}$$

The base transit time can be expressed as $\tau_B = Cw_0$. The above equations can also be expressed as (backward) recursions

$$\begin{aligned} y_{M-1} &= v_{M-1}^{1+\gamma_1-\gamma_2}, \qquad w_{M-1} = v_{M-1}^{\gamma_2-1} y_{M-1} \\ y_i &= y_{i+1} + v_i^{1+\gamma_1-\gamma_2}, \quad i=0,\ldots,M-2 \\ w_i &= w_{i+1} + v_i^{\gamma_2-1} y_i, \quad i=0,\ldots,M-2. \end{aligned} \tag{12}$$

Using these expressions, we can formulate the optimization problem (10) as

$$\begin{array}{ll} \text{minimize} & \tau_B \\ \text{subject to} & N_{\min} \le v_i \le N_{\max}, \quad i=0,1,\ldots,M-1 \\ & y_{i+1} + v_i^{1+\gamma_1-\gamma_2} = y_i, \quad i=0,\ldots,M-2 \\ & w_{i+1} + v_i^{\gamma_2-1} y_i = w_i, \quad i=0,\ldots,M-2 \\ & y_{M-1} = v_{M-1}^{1+\gamma_1-\gamma_2} \\ & w_{M-1} = v_{M-1}^{\gamma_2-1} y_{M-1} \end{array} \tag{13}$$

where $v_0,\ldots,v_{M-1}, y_0,\ldots,y_{M-1}, w_0,\ldots,w_{M-1}$ are the variables, and $\tau_B = Cw_0$. This problem is *not* a GP, since it contains posynomial equality constraints. (The last two equality constraints are GP-compatible, since a GP can include *monomial* equality constraints.) Note that the problem is sparse, since the objective and each constraint function depends on only a few variables (at most four).

The next step is to relax the posynomial equality constraints to posynomial inequality constraints, which gives

$$\begin{array}{ll} \text{minimize} & \tau_B \\ \text{subject to} & N_{\min} \le v_i \le N_{\max}, \quad i=0,1,\ldots,M-1 \\ & y_{i+1} + v_i^{1+\gamma_1-\gamma_2} \le y_i, \quad i=0,\ldots,M-2 \\ & w_{i+1} + v_i^{\gamma_2-1} y_i \le w_i, \quad i=0,\ldots,M-2 \\ & y_{M-1} = v_{M-1}^{1+\gamma_1-\gamma_2} \\ & w_{M-1} = v_{M-1}^{\gamma_2-1} y_{M-1} \end{array} \tag{14}$$

TABLE I
MODEL AND PROBLEM PARAMETER VALUES USED FOR THE NUMERICAL EXAMPLE

Parameter	Value
$N_{\rm ref}$	10^{17} cm^{-3}
$N_{\max}$	5×10^{18} cm^{-3}
$N_{\min}$	5×10^{16} cm^{-3}
D_{n0}	20.72 cm^2/s
n_{i0}	1.4×10^{10} cm^{-3}
γ_1	0.42
γ_2	0.69
W_B	100 nm

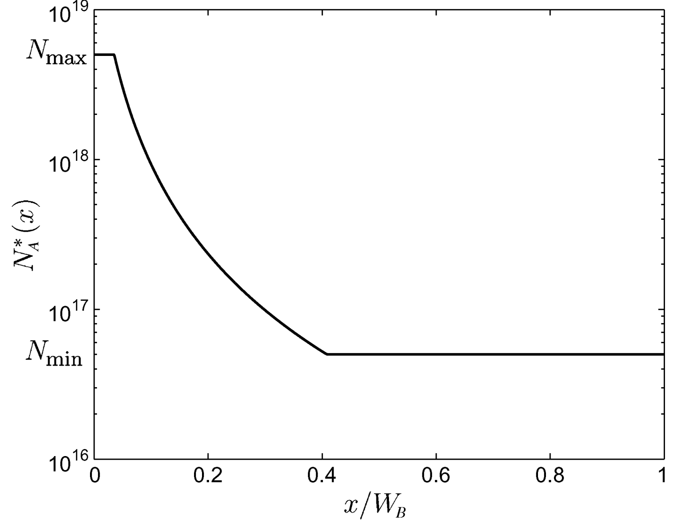

Fig. 1. Optimal doping profile $N_A^*(x)$.

which is a sparse geometric program. Since increasing y_i or w_i only increases the objective τ_B, we can conclude that at any optimal solution of the GP (14), the relaxed inequality constraints hold with equality. Thus, by solving the sparse GP (14), we solve the dense GP (10). Although the GP (14) has around three times the number of variables as the original GP (10), it is very sparse, and so can be solved extremely efficiently.

F. Example

We first consider a numerical example of the optimal doping profile problem (10). The model and problem parameters are taken from [30], and listed in Table I. We discretize the problem with $M = 1000$, and use the sparse GP formulation (14), which has 3000 variables and 4000 constraints. This GP is solved using the MOSEK GP solver [1], with a computation time on the order of 3 or 4 seconds, on a small personal computer. The resulting (globally) optimal doping profile is shown in Fig. 1. The (globally) optimal base transit time is $\tau_B^* = 1.52$ ps. To check that our discretization error is not significant, we also solved the problem with $M = 10\,000$ (which increases solution time to only 40 seconds), and find no change, to three significant figures, in the optimal base transit time or the optimal doping concentrations.

We compared our optimal doping profile with the one obtained by Kumar and Patri [19], who use an iterative method to minimize base transit time. The optimal doping profile (and base transit time) found by our method coincides with the one found by their iterative scheme. Since our method always finds the globally optimal solution, we conclude that (at least for this

one example) their iterative scheme did in fact produce the globally optimal doping profile.

For comparison, we give the base transit time achieved by a simple uniform doping profile. For such a doping profile we have $N_A(x) = N_A$, so the base transit time has the form

$$\tau_B = \frac{W_B^2}{2D_{n0}} \left(\frac{N_A}{N_{\rm ref}}\right)^{\gamma_1}. \quad (15)$$

Since $\gamma_1 > 0$, the minimum is achieved when $N_A = N_{\min}$. For this particular example, the base transit time obtained by a uniform doping profile is 1.80 ps, around 18% larger than the optimal value. (In other words, the optimal doping profile gives an 18% reduction in base transit time, compared to a simple uniform doping profile.)

G. Sensitivity Analysis

Before proceeding we comment on the sensitivity of the optimal doping profile to variations in model parameters and the doping profile, since optimization sometimes yields designs that are very sensitive to model variations. First we consider variations in the doping profile itself, i.e., we assume the true doping profile is $N_A(x_i) = N_A^*(x_i)(1+e_i)$, where $N_A^*(x_i)$ denotes the optimal doping profile, $N_A(x_i)$ denotes the actual doping profile, and e_i is a fractional error. As a simple model of variation, we assume that $|e_i| \le E$, where E is a maximum fractional deviation between of the optimal and the actual doping profiles. We can find a general bound on how much such variations increase τ_B by first imagining that the variations in v_i are different in the two sums appearing in the expression (9) for τ_B. (This assumption is conservative, i.e., it increases the effect of the doping variation, but it makes the analysis simple.) With this assumption we can substitute the extreme value $N_A(x_i) = N_A^*(x_i)(1+E)$ or $N_A(x_i) = N_A^*(x_i)(1-E)$ for each value of v_i, that maximizes τ_B. For the inner sum, i.e., for the terms $v_j^{1+\gamma_1-\gamma_2}$, we choose $e_j = E$ since $1+\gamma_1-\gamma_2 > 0$. Similarly for the outer sum, i.e., for the terms $v_i^{\gamma_2-1}$, we choose $e_i = -E$, since $\gamma_2 - 1 < 0$. This gives the (upper bound on) base transit time

$$\tau_B \le \overline{\tau}_B = (1-E)^{\gamma_2-1}(1+E)^{1+\gamma_1-\gamma_2}\tau_B^*.$$

Here $\overline{\tau}_B$ is an upper bound on the base transit time, for *any* profile $N_A(x)$ that is within a fraction E of $N_A^*(x)$.

For our example, if we have a maximum deviation of $E = 0.05$ (i.e., 5% variation in the doping), and γ_1 and γ_2 from Table I, the base transit time is no more than 1.60 ps, a 5% increase over the optimal value.

In a similar way we can study uncertainty in the parameters γ_1 and γ_2. For the optimal profile, we find that τ_B increases with decreasing γ_1 and increasing γ_2. It follows that if these two parameters lie within given intervals, the worst case (i.e., largest τ_B) occurs when γ_1 assumes its smallest value, and γ_2 assumes it largest value. For the example, we find that, for $\pm 10\%$ variations in γ_1 and γ_2, the largest base transit time is obtained when γ_1 is 0.9 times its nominal value and γ_2 is 1.1times its nominal value. The resulting worst case base transit time is $\tau_B = 1.62$ ps, a 7% increase over the optimal base transit time. Further, if the doping profile is optimized at the worst case values of γ_1 and γ_2, the optimal base transit time achieved is still the same, i.e., 1.62 ps. This shows that the optimal doping profile does not change appreciably with variation in γ_1 and γ_2.

In summary, we find that the optimal doping profile $N_A^*(x)$ found above does not suffer from excessive sensitivity to doping profile or model parameter values.

III. Extensions of Base Transit Time Optimization

In this section we consider various extensions to the basic doping profile optimization problem (10) [or the sparse formulation (14)], that still can be formulated as in GP form, and thus preserves our ability to solve the problem globally (and efficiently). First, we can change the model for τ_B, as long as it remains a posynomial (or generalized posynomial). In Section III-A, for example, we show how velocity saturation can be modeled, while keeping our expression for τ_B generalized posynomial. Second, we consider examples of additional GP compatible constraints that can be added to the doping profile problem. In Section III-B and Section III-C, we show how a constraint on the doping gradient and a contraint on the current gain of the device can be exactly transformed to a GP compatible format, respectively. In Section III-D and Section III-E, we show how a constraint on the intrinsic base sheet resistance and a constraint on the breakdown voltage can be approximately taken into account in the GP formulation. The base transit time optimization problem can be solved with any combination of these GP compatible constraints.

A. Velocity Saturation

A model for the base transit time in a homojunction BJT, taking velocity saturation at the base–collector into account, is given by [31]

$$\tau_B = \int_0^{W_B} \frac{n_i^2(x)}{N_A(x)} \left(\int_x^{W_B} \frac{N_A(y)}{n_i^2(y)D_n(y)}dy\right)dx + \frac{1}{v_{\rm sat}} \int_0^{W_B} \frac{n_i^2(x)N_A(W_B)}{N_A(x)n_i^2(W_B)}dx \quad (16)$$

where $v_{\rm sat}$ is the saturation velocity of electrons, which is a constant.

Using the simple discretization described above, the (discretized, approximate) base transit time can be expressed as

$$\tau_B = C\sum_{i=0}^{M-1} v_i^{\gamma_2-1} \sum_{j=i}^{M-1} v_j^{1+\gamma_1-\gamma_2} + \frac{W_B}{v_{\rm sat}M} v_{M-1}^{1-\gamma_2} \sum_{i=0}^{M-1} v_i^{\gamma_2-1}. \quad (17)$$

This function is also a posynomial of the variables $v_0, \ldots, v_{M-1}$, and so can be substituted into the basic base transit time minimization problem (10), which gives another GP.

The carriers can attain saturation velocity inside the base region. To take this velocity saturation into account we look at the following interpretation of the base transit time. The base transit time τ_B is the total time required by a carrier to travel across the base width. Let the apparent velocity of the particle,

i.e., not taking into account the velocity saturation, at point x in the base region be $v(x)$. Then the base transit time is given by

$$\tau_B = \int_0^{W_B} \frac{1}{v(x)} dx. \tag{18}$$

Velocity saturation means that the true velocity of the carriers cannot increase beyond $v_{\rm sat}$, and so the true velocity at a point x is $\min\{v(x), v_{\rm sat}\}$. Therefore, the base transit time is

$$\tau_{B,\rm vs} = \int_0^{W_B} \frac{1}{\min\{v(x), v_{\rm sat}\}} dx. \tag{19}$$

By comparing (1) and (18), the apparent velocity $v(x)$ is

$$v(x) = \left(\frac{n_i^2(x)}{N_A(x)} \left(\int_x^{W_B} \frac{N_A(y)}{n_i^2(y) D_n(y)} dy \right) \right)^{-1} \tag{20}$$

and base transit time considering velocity saturation in the base is

$$\tau_{B,\rm vs} = \int_0^{W_B} \max\left\{ \frac{1}{v_{\rm sat}}, \frac{n_i^2(x)}{N_A(x)} \left(\int_x^{W_B} \frac{N_A(y)}{n_i^2(y) D_n(y)} dy \right) \right\} dx. \tag{21}$$

We will now show that the problem of finding the base doping profile that minimizes $\tau_{B,\rm vs}$ as in (21), is a *generalized geometric program* (GGP) which can be transformed to a GP [6], and then solved efficiently. Using (2)–(6), and discretizing the base region we get

$$\tau_{B,\rm vs} = \sum_{i=0}^{M-1} \max\left\{ \frac{1}{v_{\rm sat}}, \frac{1}{N_{\rm ref}^{\gamma_1} D_{n0}} v_i^{\gamma_2 - 1} \times \left(\sum_{j=i}^{M-1} v_j^{1+\gamma_1-\gamma_2} \frac{W_B}{M} \right) \right\} \frac{W_B}{M}. \tag{22}$$

Note that this is a rather complex function of the variables v_i; for example, because of the maximum, it is not even differentiable (at some points, at least). But the base transit time $\tau_{B,\rm vs}$ is a sum of terms, each of which is the maximum of posynomials, and so is a *generalized posynomial* [6]. The problem of minimizing a generalized posynomial, subject to posynomial constraints, is a generalized geometric program; see [6, Sect. 5] for details. The problem can therefore be solved efficiently.

To transform this problem to a sparse GP, we define

$$\begin{aligned} y_i &= \sum_{j=i}^{M-1} v_j^{1+\gamma_1-\gamma_2}, \quad i = 0, \ldots, M-1 \\ u_i &= \max\left\{ \frac{1}{v_{\rm sat}}, K v_i^{\gamma_2-1} y_i \right\}, \quad i = 0, \ldots, M-1 \end{aligned} \tag{23}$$

where K is $W_B / M N_{\rm ref}^{\gamma_1} D_{n0}$. The base transit time can be expressed as

$$\tau_{B,\rm vs} = \frac{W_B}{M} \sum_{i=0}^{M-1} u_i. \tag{24}$$

The equations for y_is can be expressed as (backward) recursions, as in (12). Using (23) and (12), the optimization problem is

$$\begin{array}{ll} \text{minimize} & \tau_{B,\rm vs} \\ \text{subject to} & N_{\min} \le v_i \le N_{\max}, \quad i = 0, 1, \ldots, M-1 \\ & y_{i+1} + v_i^{1+\gamma_1-\gamma_2} = y_i, \quad i = 0, \ldots, M-2 \\ & y_{M-1} = v_{M-1}^{1+\gamma_1-\gamma_2} \\ & u_i = \max\left\{ v_{\rm sat}^{-1}, K v_i^{\gamma_2-1} y_i \right\}, \quad i = 0, \ldots, M-1 \end{array} \tag{25}$$

where $v_0, \ldots, v_{M-1}, y_0, \ldots, y_{M-1}, u_0, \ldots, u_{M-1}$ are the variables, and $\tau_{B,\rm vs}$ as in (24). This problem is not a GP, because it contains generalized posynomial equality constraints. To convert it to a GP, we relax these equality constraints to inequality constraints, and use the fact that

$$a \ge \max\{b_1, b_2\} \quad \text{if and only if} \quad a \ge b_1,\ a \ge b_2$$

which gives

$$\begin{array}{ll} \text{minimize} & \tau_{B,\rm vs} \\ \text{subject to} & N_{\min} \le v_i \le N_{\max}, \quad i = 0, 1, \ldots, M-1 \\ & y_{i+1} + v_i^{1+\gamma_1-\gamma_2} \le y_i, \quad i = 0, \ldots, M-2 \\ & y_{M-1} = v_{M-1}^{1+\gamma_1-\gamma_2} \\ & K v_i^{\gamma_2-1} y_i \le u_i, \quad i = 0, \ldots, M-1 \\ & v_{\rm sat}^{-1} \le u_i, \quad i = 0, \ldots, M-1 \end{array} \tag{26}$$

which is a GP. Since increasing y_i or u_i only increases the objective $\tau_{B,\rm vs}$, we can conclude that at any optimal solution of the GP (26), the relaxed inequality constraints hold with equality, and therefore problems (25) and (26) are equivalent. Note that for any given apparent velocity profile $v(x)$, $0 \le x \le W_B$, we have $\tau_B \le \tau_{B,\rm vs}$. Therefore, the optimal base transit time satisfies the inequality, $\tau_B^* \le \tau_{B,\rm vs}^*$.

To illustrate this, we again consider the numerical example of Section II-F. We take the saturation velocity to be $v_{\rm sat} = 10^7$ cm/s. We plot the apparent velocity profile $v(x)$ (20) for the optimal doping profile $N_A^*(x)$ shown in Fig. 1, and find that $v(x)$ is a increasing function of x, and reaches saturation at $x = 0.723 W_B$. We then solve the problem (26), which takes into account velocity saturation. The resulting optimal doping profile obtained is (very nearly) the same as the optimal doping profile obtained for the original problem (10), shown in Fig. 1. The optimal base transit time with velocity saturation in the base is found to be $\tau_{B,\rm vs}^* = 1.66$ ps, as compared to $\tau_B^* = 1.52$ ps, the optimal base transit time for problem (10). So in this example, modeling velocity saturation does not affect the optimal doping profile, but it does affect the predicted base transit time. Of course in general this need not be the case; the optimal doping profile for the model including velocity saturation need not be the same as the optimal doping profile when saturation is neglected.

B. Doping Gradient Constraint

Consider a maximum (percentage) gradient in doping concentration

$$|N_A'(x)| \le \alpha N_A(x)$$

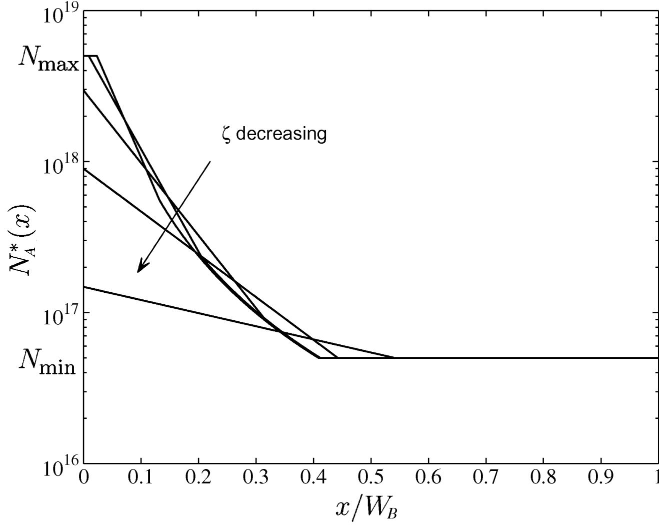

Fig. 2. Optimal doping profile $N_A^*(x)$ for different values of doping gradient limit ζ.

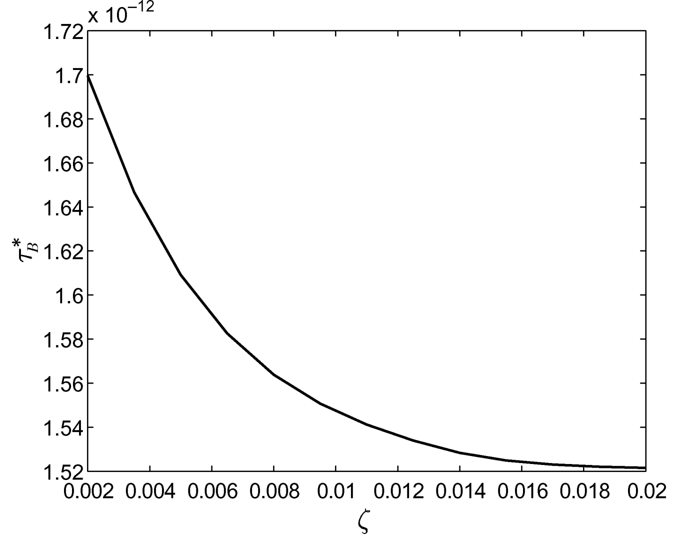

Fig. 3. Optimal base transit time versus doping gradient limit ζ.

where α specifies the maximum allowed gradient. After discretization of the base region this constraint can be expressed in terms of v_i as

$$\left(1-\frac{\alpha W_B}{M}\right) v_i \leq v_{i+1} \leq \left(1+\frac{\alpha W_B}{M}\right) v_i, \ i=0,1,\ldots,M-2. \quad (27)$$

These are $2(M-1)$ monomial inequality constraints, and so are GP compatible. We can add these constraints to the basic doping profile optimization problem, and obtain a GP.

To illustrate this, we consider our numerical example problem, with an added doping gradient constraint,

$$\begin{array}{ll} \text{minimize} & \tau_B \\ \text{subject to} & N_{\min} \leq v_i \leq N_{\max}, \quad i=0,\ldots,M-1 \\ & (1-\zeta)v_i \leq v_{i+1}, \quad i=0,\ldots,M-2 \\ & v_{i+1} \leq (1+\zeta)v_i, \quad i=0,\ldots,M-2 \end{array} \quad (28)$$

where $\zeta = \alpha W_B/M$. (The other problem and model parameters are given in Table I; we take $M = 1000$.) The optimal doping profiles obtained, for

$$\zeta = \infty, \quad 0.0155, \quad 0.0110, \quad 0.0065, \quad 0.0020$$

are plotted in Fig. 2. (The value $\zeta = \infty$ corresponds to the problem with no gradient constraint.) As ζ decreases the doping gradient constraint becomes tighter, and the value of minimum base transit time increases. Fig. 3 shows the increase in τ_B^* with decrease in ζ, or equivalently, the optimal tradeoff curve of maximum doping gradient versus base transit time.

C. Current Gain Constraint

The current gain β, i.e., the ratio of collector current to base current, is also affected by the base doping profile. In this section we show that a constraint requiring a minimum value of current gain, i.e., $\beta \geq \beta_{\min}$, is GP-compatible. The optimal doping profile problem, with minimum current gain constraint, can be expressed as

$$\begin{array}{ll} \text{minimize} & \tau_B \\ \text{subject to} & N_{\min} \leq N_A(x) \leq N_{\max}, \quad 0 \leq x \leq W_B \\ & \beta \geq \beta_{\min}. \end{array} \quad (29)$$

In the rest of this section we will show that the constraint $\beta \geq \beta_{\min}$ can be expressed as a posynomial inequality, so the problem (29) becomes a GP, after discretization, and therefore can be globally and efficiently solved.

The current gain β can be expressed as ratio of Gummel numbers

$$\beta = \frac{G_E}{G_B} \quad (30)$$

where G_E is the emitter Gummel number, and G_B is the base Gummel number [2], [32]. The emitter Gummel number G_E depends on the emitter doping profile and not on the base doping profile, and therefore can be treated as a positive constant for our purposes. The base Gummel number G_B depends on the base doping profile according to

$$G_B = \int_0^{W_B} \frac{N_A(x)n_{i0}^2}{D_n(x)n_i^2(x)} dx. \quad (31)$$

Using (2)–(5), this becomes

$$G_B = \frac{1}{N_{\text{ref}}^{\gamma_1-\gamma_2} D_{n0}} \int_0^{W_B} N_A(x)^{1+\gamma_1-\gamma_2} dx. \quad (32)$$

Discretizing the base region we get the approximation

$$G_B = \frac{W_B}{M N_{\text{ref}}^{\gamma_1-\gamma_2} D_{n0}} \sum_{i=0}^{M-1} v_i^{1+\gamma_1-\gamma_2} \quad (33)$$

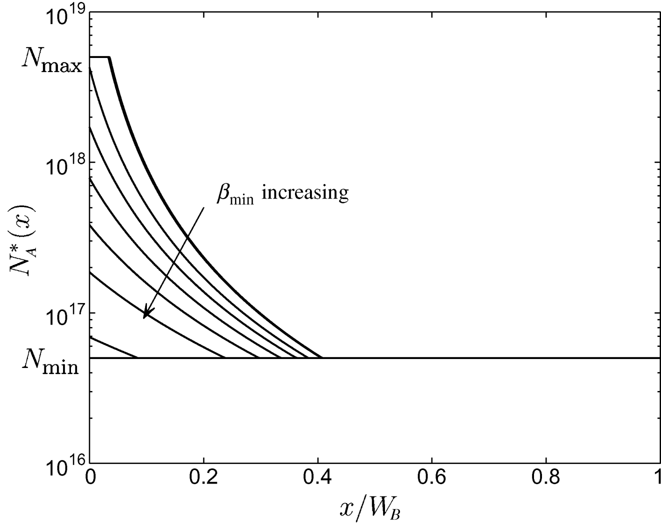

Fig. 4. Optimal doping profiles $N_A^*(x)$ for different values of minimum current gain $\beta_{\min}$.

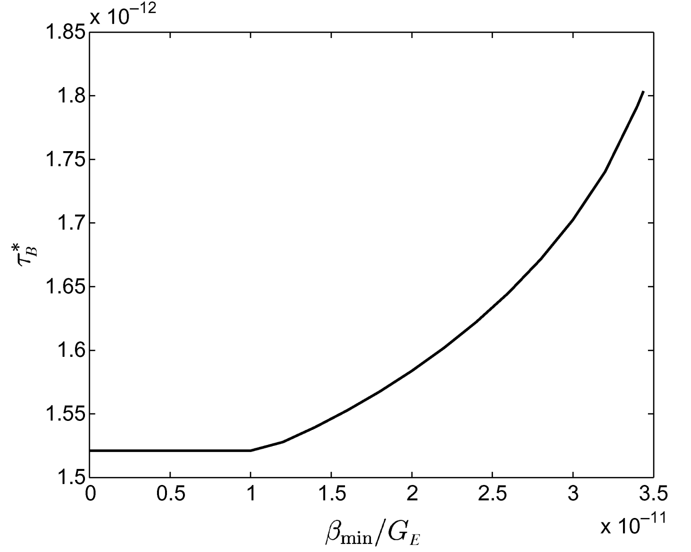

Fig. 5. Optimal tradeoff curve of base transit time versus minimum current gain $\beta_{\min}$.

which is a posynomial function of $v_i, \ldots, v_{M-1}$. The minimum current gain constraint $\beta \geq \beta_{\min}$ then becomes

$$\frac{1}{\beta} = \frac{G_B}{G_E} = \frac{W_B}{G_E M N_{\text{ref}}^{\gamma_1-\gamma_2} D_{n0}} \sum_{i=0}^{M-1} v_i^{1+\gamma_1-\gamma_2} \leq \frac{1}{\beta_{\min}} \tag{34}$$

which is a posynomial inequality constraint, and therefore GP compatible. We can therefore express the problem (29) (approximately) as the GP

$$\begin{array}{ll} \text{minimize} & \tau_B \\ \text{subject to} & N_{\min} \leq v_i \leq N_{\max}, \quad i = 0, 1, \ldots, M-1 \\ & \left(\dfrac{W_B \beta_{\min}}{G_E M N_{\text{ref}}^{\gamma_1-\gamma_2} D_{n0}}\right) \displaystyle\sum_{i=0}^{M-1} v_i^{1+\gamma_1-\gamma_2} \leq 1. \end{array} \tag{35}$$

By solving the problem (35) for various values of $\beta_{\min}$, we obtain the (globally) optimal tradeoff curve between τ_B and $\beta_{\min}$.

The minimum current gain constraint is particularly simple to express using the sparse GP formulation (14). From (11) we see that $\sum_{i=0}^{M-1} v_i^{1+\gamma_1-\gamma_2} = y_0$, so the minimum current gain constraint is just

$$\left(\frac{W_B \beta_{\min}}{G_E M N_{\text{ref}}^{\gamma_1-\gamma_2} D_{n0}}\right) y_0 \leq 1. \tag{36}$$

Adding this constraint to problem (14) gives a sparse GP formulation for the doping profile optimization problem with minimum current gain constraint (35).

To illustrate this, we consider the numerical example again, with an additional minimum current gain constraint. The resulting optimal doping profiles are shown in Fig. 4, for minimum current gains

$$\frac{\beta_{\min}}{G_E} \times 10^{11} = 0,\ 1.0,\ 1.4,\ 1.8,\ 2.2,\ 2.6,\ 3.0,\ 3.4,\ 3.43.$$

(The value $\beta_{\min} = 0$ corresponds to the unconstrained case, i.e., the optimal doping profile with no current gain constraint.)

The (globally) optimal tradeoff of base transit time versus current gain is shown in Fig. 5. The plot shows that when $\beta_{\min}$ is small the base transit time achieved is the same as in the unconstrained case, $\tau_B^* = 1.52$ ps. As the minimum current gain $\beta_{\min}$ is increased, the minimum base transit time increases. When the minimum current gain $\beta_{\min}$ is equal to the its maximum achievable value (which occurs with a constant doping profile), we have $\tau_B^* = 1.80$ ps.

D. Intrinsic Base Sheet Resistance Constraint

For a bipolar junction transistor it is desired to have low intrinsic base sheet resistance (denoted R_{Bi}). In this section we consider imposing a limit on the maximum intrinsic base sheet resistance, i.e., the constraint $R_{\text{Bi}} \leq R_{\text{Bi,max}}$. We will see that this constraint cannot be exactly incorporated in the GP framework, but a good approximation will result in a GP compatible constraint.

The intrinsic base sheet resistance is given by [32]

$$R_{\text{Bi}} = \left(q \int_0^{W_B} \mu_p(x) p_p(x) dx\right)^{-1} \tag{37}$$

where q is the magnitude of the charge of an electron, $\mu_p(x)$ is the mobility of holes in the p-type base, and $p_p(x)$ is the density of holes in the p-type base. The mobility is related to the carrier diffusion coefficient by

$$D_p(x) = \frac{kT}{q} \mu_p(x).$$

The carrier diffusion coefficient of the holes, which are the majority carriers in the p-type base, is well approximated by [3]

$$D_p(x) = D_{p0} \left(\frac{N_A(x)}{N_{\text{ref},2}}\right)^{-\gamma_3} \tag{38}$$

where D_{p0}, $N_{\text{ref},2}$ and γ_3 are positive constants with values

$$D_{p0} = 8.98\ \text{cm}^2/\text{s}, \quad N_{\text{ref},2} = 10^{17}\ \text{cm}^{-3}, \quad \gamma_3 = 0.3627. \tag{39}$$

The density of holes $p_p(x)$ is approximately equal to the base doping concentration $N_A(x)$, assuming complete ionization.

Substituting (38) into (37), we get the dependence of the intrinsic base sheet resistance on the base doping profile

$$R_{\rm Bi} = \frac{kT}{q^2 D_{p0} N_{\rm ref,2}^{\gamma_3}} \left(\int_0^{W_B} N_A(x)^{1-\gamma_3} dx \right)^{-1}. \quad (40)$$

Since $1 \geq \gamma_3$, we see that low doping concentration in the base leads to high base sheet resistance, which is not desirable.

We would like to solve the base transit time problem (7) (and (29)) with a maximum limit on $R_{\rm Bi}$, i.e., with an additional constraint $R_{\rm Bi} \leq R_{\rm Bi,max}$. This is equivalent to

$$\int_0^{W_B} N_A(x)^{1-\gamma_3} dx \geq \frac{kT}{q^2 D_{p0} N_{\rm ref,2}^{\gamma_3} R_{\rm Bi,max}}. \quad (41)$$

Discretizing the base region the inequality (41) becomes

$$\sum_{i=0}^{M-1} v_i^{1-\gamma_3} \geq \frac{kTM}{W_B q^2 D_{p0} N_{\rm ref,2}^{\gamma_3} R_{\rm Bi,max}}. \quad (42)$$

Unfortunately, this inequality is *not* a GP compatible constraint. Therefore, we seek GP compatible constraints that approximately limit the base resistance.

A crude method for controlling $R_{\rm Bi}$ is to change the minimum level of the doping concentration $N_{\rm min}$. We simply increase the minimum doping concentration $N_{\rm min}$ until the base resistance $R_{\rm Bi}$ is reduced to a satisfactory value.

We will now show a far better method to approximate the intrinsic base sheet resistance constraint (41) in a GP formulation. Using the arithmetic–geometric mean inequality we can produce a GP compatible approximation of the constraint (42). Since the arithmetic mean is greater than geometric mean, we have

$$\frac{1}{M} \sum_{i=0}^{M-1} v_i^{1-\gamma_3} \geq \left(\prod_{i=0}^{M-1} v_i^{1-\gamma_3} \right)^{\frac{1}{M}}.$$

Therefore, if the inequality

$$\left(\prod_{i=0}^{M-1} v_i^{1-\gamma_3} \right)^{\frac{1}{M}} \geq \frac{kT}{W_B q^2 D_{p0} N_{\rm ref,2}^{\gamma_3} R_{\rm Bi,max}} \quad (43)$$

holds, then the inequality (42) must also be satisfied. The inequality (43) is a GP compatible constraint, since the left-hand side is a monomial function [6]. The constraint (43) can be added to problem (10) [and (35)] to ensure that the intrinsic base sheet resistance of the device is below $R_{\rm Bi,max}$, i.e., $R_{\rm Bi} \leq R_{\rm Bi,max}$.

The approximation produces a conservative bound on the intrinsic base sheet resistance. The inequality (43) has the following interpretation. When the doping concentration is viewed on a logscale, the inequality (43) means that the average doping concentration (on a logscale) should be greater than a certain value, i.e.,

$$\frac{1}{M} \sum_{i=0}^{M-1} \log v_i \geq \frac{1}{1-\gamma_3} \log \left(\frac{kT}{W_B q^2 D_{p0} N_{\rm ref,2}^{\gamma_3} R_{\rm Bi,max}} \right).$$

To illustrate how the intrinsic base sheet resistance constraint changes the optimal doping profile we will use the numerical example of Section II-F, i.e., problem (10) with the additional constraint (43). The intrinsic base sheet resistance for the doping profile shown in Fig. 1 is found [using (40)] to be 10.3 kΩ. We solve the problem for two values of a maximum base resistance

$$R_{\rm Bi,max} = 5 \text{ k}\Omega, \quad 2.5 \text{ k}\Omega.$$

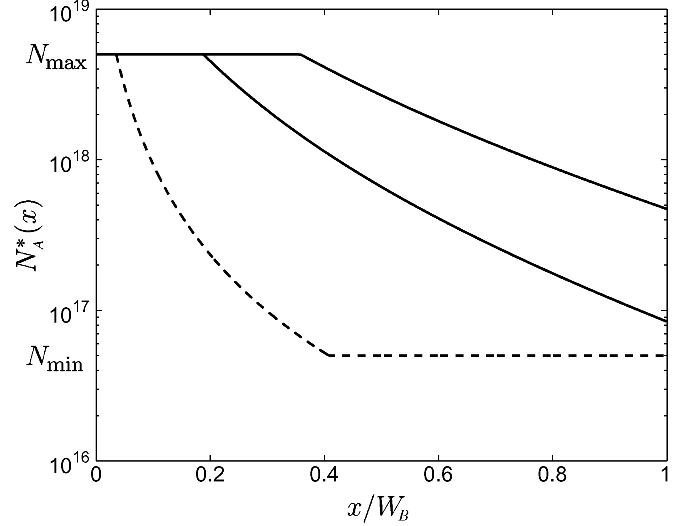

Fig. 6. Optimal doping profile $N_A^*(x)$ for different values of $R_{\rm Bi,max}$.

Fig. 6 shows the optimal doping profiles obtained. The dashed curve is the same doping profile as shown in Fig. 1. The solid curve with a greater average doping is the optimal doping profile for the case $R_{\rm Bi,max} = 2.5$ kΩ. Since the GP compatible constraint (43) is an approximation, it produces a conservative bound on the intrinsic base sheet resistance $R_{\rm Bi}$, and therefore we expect the value of $R_{\rm Bi}$ obtained by the optimal doping profiles to be lower than $R_{\rm Bi,max}$. The optimal doping profile for $R_{\rm Bi,max} = 5$ kΩ, has a $R_{\rm Bi} = 3.5$ kΩ (and $\tau_B = 3.2$ ps); and for $R_{\rm Bi,max} = 2.5$ kΩ, has a $R_{\rm Bi} = 2.2$ kΩ (and $\tau_B = 6.3$ ps).

E. Breakdown Voltage Constraint

In the design of bipolar junction transistors we are interested in achieving the collector-emitter breakdown voltage with the base open circuit $V_{\rm Br,CEO}$ above a certain threshold, i.e., include a constraint $V_{\rm Br,CEO} \geq V_{\rm Br,CEO,min}$ in the base transit time minimization problem. In this section we will show that this constraint cannot be exactly transformed to a GP compatible constraint, like the constraint on the intrinsic base sheet resistance in Section III-D. Using the same technique, however, we can form an approximation that is GP compatible.

The collector–emitter breakdown voltage with the base open circuit $V_{\rm Br,CEO}$ is related to the collector–base breakdown voltage with emitter open circuit $V_{\rm Br,CBO}$ by [32, Sect. 6.5]

$$V_{\rm Br,CEO} = \left(\frac{1}{\beta} \right)^{\frac{1}{m}} V_{\rm Br,CBO} \quad (44)$$

where m is a constant. The collector-base breakdown voltage with emitter open circuit $V_{\rm Br,CBO}$ is determined by the collector doping because the collector doping near the junction is less than the minimum base doping concentration, and the breakdown voltage of a junction is determined by doping on the lightly doped side. Thus, $V_{\rm Br,CBO}$ does not depend on the base doping concentration, and for a given collector doping $V_{\rm Br,CBO}$

is a known constant. The collector-emitter breakdown voltage $V_{\mathrm{Br,CEO}}$ depends on β and therefore on base doping profile.

A minimum breakdown voltage constraint on $V_{\mathrm{Br,CEO}}$, i.e., $V_{\mathrm{Br,CEO}} \geq V_{\mathrm{Br,CEO,min}}$ is reflected as a constraint on β as

$$\beta \leq \left(\frac{V_{\mathrm{Br,CBO}}}{V_{\mathrm{Br,CEO,min}}}\right)^{m}. \tag{45}$$

The breakdown voltage constraint forces β to be lower than a certain threshold (unlike the constraint on β in Section III-C).

Using equations (30)–(34) developed in Section III-C, after the discretizing the base region, the constraint (45) is a constraint on the doping profile

$$\frac{W_B}{G_E M N_{\mathrm{ref}}^{\gamma_1-\gamma_2} D_{n0}} \sum_{i=0}^{M-1} v_i^{1+\gamma_1-\gamma_2} \geq \left(\frac{V_{\mathrm{Br,CEO,min}}}{V_{\mathrm{Br,CBO}}}\right)^{m}. \tag{46}$$

This is *not* a GP compatible constraint. To produce a GP compatible approximation of the constraint (46), we use the arithmetic–geometric mean inequality (similar to Section III-D). Since the arithmetic mean is greater than geometric mean, if the inequality

$$\left(\prod_{i=0}^{M-1} v_i^{1+\gamma_1-\gamma_2}\right)^{\frac{1}{M}} \geq \left(\frac{V_{\mathrm{Br,CEO,min}}}{V_{\mathrm{Br,CBO}}}\right)^{m} \left(\frac{G_E N_{\mathrm{ref}}^{\gamma_1-\gamma_2} D_{n0}}{W_B}\right) \tag{47}$$

holds, then the inequality (46) must be satisfied. The inequality (47) is a GP compatible constraint. The inequality (47) gives a conservative bound that guarantees $V_{\mathrm{Br,CEO}} \geq V_{\mathrm{Br,CEO,min}}$.

IV. Cutoff Frequency Optimization

In this section we will consider the problem of determining the base doping profile that maximizes the cutoff frequency f_T of the bipolar junction transistor. The cutoff frequency is given by [2], [32]

$$\frac{1}{2\pi f_T} = \tau_F + \frac{C_{J,\mathrm{BE}} + C_{J,\mathrm{BC}}}{g_m} + R_C C_{J,\mathrm{BC}} \tag{48}$$

where τ_F is the forward transit time, $C_{J,\mathrm{BE}}$ is the base–emitter junction or depletion layer capacitance, $C_{J,\mathrm{BC}}$ is the base–collector junction or depletion layer capacitance, g_m is the transconductance, and, R_C is the collector resistance. We will now investigate how f_T depends on the base doping profile and then show the problem of optimizing cutoff frequency can be solved by the GP approach.

A. Transconductance

The transconductance g_m is given by

$$g_m = \frac{qI_C}{kT} \tag{49}$$

where I_C is the collector current. The collector current I_C is given by [2, Sect. 3.6]

$$I_C = \frac{qA_{\mathrm{BE}} n_{i0}^2}{G_B} \exp\left(\frac{qV_{\mathrm{BE}}}{kT}\right) \tag{50}$$

where q is the magnitude of charge of an electron, A_{BE} is the area of base–emitter junction, n_{i0} is the intrinsic carrier concentration in undoped silicon, and, V_{BE} is a given applied voltage across the base–emitter junction. Thus, I_C depends on the base doping profile through the base Gummel number G_B as given in (32). (We can put a maximum and minimum current constraint on I_C, which is same as having a maximum and minimum current gain constraint. These constraints can be accurately transformed to a GP compatible constraint as shown is Section III-E and Section III-C respectively.)

B. Junction Capacitances

The doping concentration in the collector region near the base–collector junction is assumed to be lower than the doping concentration in base. Therefore, the base–collector junction capacitance $C_{J,\mathrm{BC}}$, is determined by the collector doping profile. Thus, $C_{J,\mathrm{BC}}$ is independent of the base doping profile. The collector resistance R_C also depends on collector doping and thus independent of the base doping.

The base–emitter junction or depletion layer capacitance $C_{J,\mathrm{BE}}$ depends on the base doping profile, since the emitter doping is greater than the maximum base doping $N_{\max}$. To model the exact dependence of $C_{J,\mathrm{BE}}$ on the base doping profile is quite complicated, therefore we make some simplifying assumptions. We assume that the extention of base–emitter depletion layer into the base is small, which is the case because base–emitter junction is forward biased when the device is operating in active region. We also assume that the base doping near the base–emitter junction is constant, which is inspired from the solution of problem (10) shown in Fig. 1. Note that we can also add the constraint that the base doping profile is constant up to 5% of the base width. Using the capacitance relation for a pn-junction in [23, Sect. 7.2], $C_{J,\mathrm{BE}}$ for a constant base doping is

$$C_{J,\mathrm{BE}} = A_{\mathrm{BE}} \left(\frac{q\varepsilon_{\mathrm{Si}} N_A(0)}{2(V_{\mathrm{bi}} - V_{\mathrm{BE}})}\right)^{\frac{1}{2}} \tag{51}$$

where $\varepsilon_{\mathrm{Si}}$ is the permittivity of silicon, A_{EB} is the area of the base–emitter junction, V_{bi} is the built-in potential, V_{EB} is a given applied voltage across the emitter–base junction. The quantity $V_{\mathrm{bi}} - V_{\mathrm{BE}}$ depends very weakly on the base doping, and we take it to be a constant. Thus, $C_{J,\mathrm{BE}}$ is proportional to $N_A(0)^{1/2}$.

Other shapes of the base doping profile near the base–emitter junction can be considered. In general suppose the doping profile is given by

$$N_A(x) = bx^m, \quad 0 \leq x \leq 0.05W_B$$

where b is a variable, and m is a constant that depends on the shape of the profile. For example, $m = 0$ corresponds to constant doping, $m = 1$ corresponds to linear graded doping. The capacitance is given by

$$C_{J,\mathrm{BE}} = A_{\mathrm{BE}} \left(\frac{q\varepsilon_{\mathrm{Si}} b}{2(V_{\mathrm{bi}} - V_{\mathrm{BE}})}\right)^{\frac{1}{(m+2)}}.$$

For a particular shape of the profile, i.e., given m, the capacitance $C_{J,\mathrm{BE}}$ is a monomial in the variable b and therefore posses no problems in optimizing the cutoff frequency by the GP approach.

C. Forward Transit Time

The forward transit time has the following components [2, Sect. 5.2], [32, Sect. 8.3.3]

$$\tau_F = \tau_E + \tau_B + \tau_{\rm EB} + \tau_{\rm BC} \tag{52}$$

where τ_E is the emitter delay time, τ_B is the base transit time, $\tau_{\rm EB}$ is the emitter–base depletion region transit time, $\tau_{\rm BC}$ is the base–collector depletion region transit time. The emitter delay time is given by [2, Sect. 3.6, Sect. 5.2]

$$\tau_E = \frac{W_E p_{{\rm Eq},E}}{2n_{i0}^2} G_B \tag{53}$$

where W_E is the width of the emitter region, $p_{{\rm Eq},E}$ is the equilibrium concentration of holes in the emitter, n_{i0} is the intrinsic carrier concentration in undoped silicon, G_B is the base Gummel number. The quantities W_E, $p_{{\rm Eq},E}$ and n_{i0} are independent of the base doping profile $N_A(x)$. The base Gummel number G_B depends on $N_A(x)$ as given by (32). The base–collector depletion region transit time $\tau_{\rm BC}$ is given by

$$\tau_{\rm BC} = \frac{W_{\rm BC}}{2v_{\rm sat}} \tag{54}$$

where $W_{\rm BC}$ is the base–collector depletion width and $v_{\rm sat}$ is the saturation velocity of electrons. The base–collector depletion width $W_{\rm BC}$ is determined by the collector doping concentration near the base–collector junction which we assume to be lower than the base doping concentration. Therefore, $\tau_{\rm BC}$ does not depend on the base doping profile. The emitter–base depletion region transit time $\tau_{\rm EB}$ is typically small compared to the other terms in (52) and can be neglected.

D. GP Formulation of Cutoff Frequency

Now we can see the dependence of the cutoff frequency on the base doping profile. Substituting (49)–(51) in (48) we get

$$\begin{aligned}\frac{1}{2\pi f_T} = {} & \tau_B + \left(\frac{W_E p_{{\rm Eq},E}}{2n_{i0}^2}\right) G_B \\ & + \left(\frac{kT}{n_{i0}^2}\left(\frac{\varepsilon_{\rm Si}}{2q^3(V_{\rm bi}-V_{\rm BE})}\right)^{\frac{1}{2}} \exp\left(-\frac{qV_{\rm BE}}{kT}\right)\right) \\ & \times N_A(0)^{\frac{1}{2}} G_B \\ & + \left(\frac{C_{J,{\rm BC}}kT}{q^2 A_{\rm BE} n_{i0}^2} \exp\left(-\frac{qV_{\rm BE}}{kT}\right)\right) G_B + \frac{W_{\rm BC}}{2v_{\rm sat}} \\ & + R_C C_{J,{\rm BC}}\end{aligned} \tag{55}$$

where τ_B and G_B depend on base doping profile $N_A(x)$ as given by (6) and (32), respectively.

Discretizing the base region and using (9) and (33), we obtain $1/2\pi f_T$ as a posynomial in the variables $v_0, \ldots, v_{M-1}$. Therefore, the problem of maximizing f_T, equivalently the problem

$$\begin{array}{ll}\text{minimize} & \dfrac{1}{2\pi f_T} \\ \text{subject to} & N_{\min} \le v_i \le N_{\max}, \quad i = 0, 1, \ldots, M-1\end{array} \tag{56}$$

is a GP. Any other GP compatible constraints, e.g., as discussed in the Section III, can be included in (56).

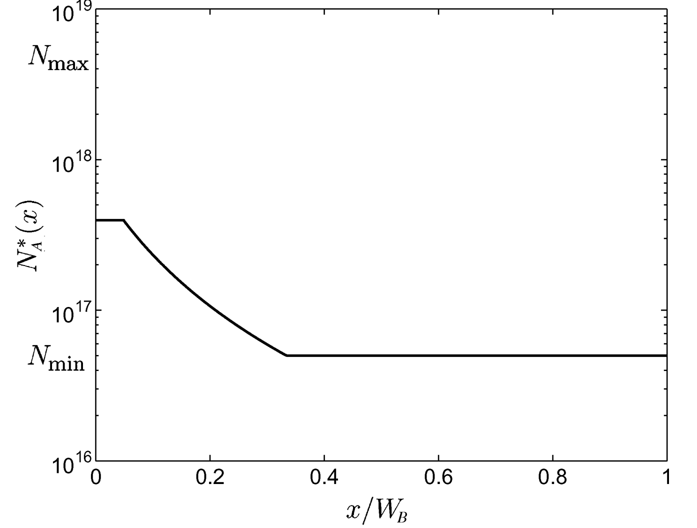

Fig. 7. Optimal doping profile $N_A^*(x)$.

E. Example

To get an idea of the shape of the optimal doping profile that maximizes the cutoff frequency, we consider a numerical example. Without explicitly giving all the parameter values in the expression (55) we will solve the following problem:

$$\begin{array}{ll}\text{minimize} & \tau_B + 4 \times 10^{-24} G_B + 8 \times 10^{-34} v_0^{\frac{1}{2}} G_B \\ \text{subject to} & N_{\min} \le v_i \le N_{\max}, \quad i = 0, 1, \ldots, M-1 \\ & v_i = v_{i+1}, \quad i = 0, \ldots, \lfloor 0.05M \rfloor.\end{array} \tag{57}$$

The parameter values in Table I are used. The optimal doping profile obtained is shown in Fig. 7, and the optimal objective value is 1.81 ps. The individual contributions to the objective are: the base transit time $\tau_B = 1.61$ ps, the sum of the forward transit time and the base–collector junction capacitance delay, $4 \times 10^{-24} G_B = 0.17$ ps and, the base–emitter junction capacitance delay $8 \times 10^{-34} v_0^{1/2} G_B = 0.03$ ps. If we take $W_{\rm BC}/2v_{\rm sat} + R_C C_{J,{\rm BC}} = 0.4$ ps, we obtain the cutoff frequency $f_T = 7.2$ GHz.

V. HBT Optimization

HBTs alters the structure of the semiconductor material in the base region by using another semiconductor. In the SiGe HBT, the base transit time is affected by the profile of both the doping and the Germanium content in the base. This leads to several possible optimization problems.

- *Optimal doping profile.* Determine the optimal doping profile, given the Ge-profile.
- *Optimal* Ge-*profile*. Determine the Ge-profile, given the doping profile.
- *Joint profile optimization*. Determine the optimal doping profile and the optimal Ge-profile.

We will show that all three problems can be expressed as GPs, and therefore solved globally with great efficiency. To be specific, we will consider a silicon germanium (SiGe) HBT throughout this section.

A. Base Transit Time Model

For a SiGe HBT the base transit time is given by

$$\tau_{B,\mathrm{SiGe}} = \int_0^{W_B} \frac{n_{i,\mathrm{SiGe}}^2(x)}{N_A(x)} \left(\int_x^{W_B} \frac{N_A(y)}{n_{i,\mathrm{SiGe}}^2(y) D_{n,\mathrm{SiGe}}(y)} dy \right) dx \tag{58}$$

where $n_{i,\mathrm{SiGe}}(x)$ is the intrinsic carrier concentration in SiGe, and $D_{n,\mathrm{SiGe}}(x)$ is the carrier diffusion coefficient of SiGe. At a point x, the intrinsic carrier concentration $n_{i,\mathrm{SiGe}}(x)$ depends on the effective bandgap reduction $\Delta E_{g,\mathrm{eff}}(x)$ as

$$n_{i,\mathrm{SiGe}}^2 = \gamma n_{i0}^2 \exp\left(\frac{\Delta E_{g,\mathrm{eff}}(x)}{kT}\right) \tag{59}$$

where γ is the ratio of the effective density of states in SiGe to the effective density of states in silicon, and is a constant. The effective bandgap reduction $\Delta E_{g,\mathrm{eff}}(x)$ has two components, one due to the doping profile $\Delta E_{g,\mathrm{dop}}(x)$ (as given earlier by (3)), and the other due to the Ge-content $\Delta E_{g,\mathrm{Ge}}(x)$

$$\Delta E_{g,\mathrm{eff}}(x) = \Delta E_{g,\mathrm{dop}}(x) + \Delta E_{g,\mathrm{Ge}}(x). \tag{60}$$

The component $\Delta E_{g,\mathrm{Ge}}(x)$ due to Ge-content depends linearly on $G(x)$, the Ge fraction in SiGe at the point x, i.e.

$$\frac{\Delta E_{g,\mathrm{Ge}}(x)}{kT} = \mu G(x) \tag{61}$$

and we require

$$0 \le G(x) \le G_{\max}, \quad 0 \le x \le W_B \tag{62}$$

where $G_{\max}$ is the maximum value of Ge fraction desired. Therefore, we have

$$n_{i,\mathrm{SiGe}}^2(x) = \gamma n_{i0}^2 \left(\frac{N_A(x)}{N_{\mathrm{ref}}}\right)^{\gamma_2} \exp\left(\mu G(x)\right) \tag{63}$$

where the Ge fraction $G(x)$ is a function of the space variable x. The carrier diffusion coefficient $D_{n,\mathrm{SiGe}}(x)$ also depends on Ge-content, and is given by

$$D_{n,\mathrm{SiGe}}(x) = \left(1 + K_{\mathrm{SiGe}} G(x)\right) D_n(x) \tag{64}$$

where K_{SiGe} is a constant and $D_n(x)$ is given by (5).

Using (63) and (64) we obtain an expression for base transit time

$$\tau_{B,\mathrm{SiGe}} = \frac{1}{N_{\mathrm{ref}}^{\gamma_1} D_{n0}} \int_0^{W_B} \exp\left(\mu G(x)\right) N_A(x)^{\gamma_2 - 1} \times \left(\int_x^{W_B} \frac{N_A(y)^{1+\gamma_1-\gamma_2}}{\exp\left(\mu G(y)\right)\left(1 + K_{\mathrm{SiGe}} G(y)\right)} dy \right) dx. \tag{65}$$

This expression depends on both the doping profile $N_A(x)$, and the Germanium profile $G(x)$; these are both functions of the space variable x, for $0 \le x \le W_B$.

B. Optimal Doping Profile

We first consider the problem of choosing the doping profile to minimize base transit time, when the Ge-profile is fixed. We will show that this problem results in a GP similar to the basic one considered above (7). Since the Ge-profile is known, the values of the functions $\exp(\mu G(x))$ and $\exp(\mu G(x))(1 + K_{\mathrm{SiGe}} G(x))$ are known for $0 \le x \le W_B$ (and are positive).

Discretizing the expression for τ_B, we get

$$\tau_{B,\mathrm{SiGe}} = C \sum_{i=0}^{M-1} \alpha_i v_i^{\gamma_2 - 1} \sum_{j=i}^{M-1} \beta_j v_j^{1+\gamma_1-\gamma_2} \tag{66}$$

where $C = W_B^2 / M^2 N_{\mathrm{ref}}^{\gamma_1} D_{n0}$ and α_i, β_i, v_i are the values of the respective functions at $x_i = iW_B/M$ for $i = 0, \ldots, M-1$, i.e.

$$\begin{aligned} \alpha_i &= \exp\left(\mu G(x_i)\right) \\ \beta_i &= \frac{1}{\left(\exp\left(\mu G(x_i)\right)\left(1 + K_{\mathrm{SiGe}} G(x_i)\right)\right)} \\ v_i &= N_A(x_i). \end{aligned} \tag{67}$$

Here, the values α_i, β_i, $i = 0, \ldots, M-1$ are fixed positive values; $v_0, \ldots, v_{M-1}$ are the variables. Thus, $\tau_{B,\mathrm{SiGe}}$ is a posynomial in $v_0, \ldots, v_{M-1}$, so the problem of choosing the doping profile, with a given Ge-profile, to minimize base transit time, i.e.

$$\begin{array}{ll} \text{minimize} & \tau_{B,\mathrm{SiGe}} \\ \text{subject to} & N_{\min} \le v_i \le N_{\max}, \quad i = 0, 1, \ldots, M-1 \end{array} \tag{68}$$

is a GP.

C. Optimal Ge-Profile

We now assume that the doping profile is fixed, and consider the problem of choosing the Ge-profile to minimize base transit time. To formulate this problem as a GP we need to make some simplifying assumptions. We assume the total Ge-content in the base is kept constant

$$\frac{1}{W_B} \int_0^{W_B} G(x) dx = G_{\mathrm{avg}} \tag{69}$$

and that the diffusion coefficient of SiGe is approximated by

$$D_{n,\mathrm{SiGe}}(x) = (1 + K_{\mathrm{SiGe}} G_{\mathrm{avg}}) D_n(x). \tag{70}$$

We note that $(1 + K_{\mathrm{SiGe}} G_{\mathrm{avg}})$ is a positive constant, since we assume G_{avg} is fixed.

Instead of working with $G(x)$ directly, we will work with its logarithm, $H(x) = \exp(G(x))$. The base transit time can be expressed in terms of H as

$$\tau_{B,\mathrm{SiGe}} = \frac{1}{N_{\mathrm{ref}}^{\gamma_1} D_{n0} (1 + K_{\mathrm{SiGe}} G_{\mathrm{avg}})} \int_0^{W_B} H(x)^{\mu} N_A(x)^{\gamma_2 - 1} \times \left(\int_x^{W_B} H(y)^{-\mu} N_A(y)^{1+\gamma_1-\gamma_2} dy \right) dx. \tag{71}$$

Now we discretize the base region, and define

$$z_i = H(x_i) = \exp\left(G(x_i)\right), \quad i = 0, 1, \ldots, M-1.$$

The base transit time is then (approximately)

$$\tau_{B,\mathrm{SiGe}} = \tilde{C} \sum_{i=0}^{M-1} z_i^{\mu} v_i^{\gamma_2 - 1} \sum_{j=i}^{M-1} z_j^{-\mu} v_j^{1+\gamma_1-\gamma_2} \tag{72}$$

where $\tilde{C} = W_B^2 / M^2 N_{\mathrm{ref}}^{\gamma_1} D_{n0} (1 + K_{\mathrm{SiGe}} G_{\mathrm{avg}})$ is a constant. Since the doping profile is given, v_i are positive constants; the variables here are $z_0, \ldots, z_{M-1}$, which describe the Ge-profile. Note that $\tau_{B,\mathrm{SiGe}}$ is a posynomial in $z_0, \ldots, z_{M-1}$.

We now translate the maximum and average Ge-profile constraints, given in (62) and (69), into GP-compatible constraints on $z_0, \ldots, z_{M-1}$. The maximum Ge-fraction constraint becomes

$$1 \le z_i \le \exp(G_{\max}), \quad i = 0, 1, \ldots, M-1 \tag{73}$$

which is a set of GP compatible constraints. The constraint on the average value of the Ge-profile, in discretized form, is

$$\frac{1}{M}\sum_{i=0}^{M-1} G(x_i) = G_{\text{avg}} \tag{74}$$

which can be expressed in terms of $z_0, \ldots, z_{M-1}$ as

$$\prod_{i=0}^{M-1} z_i = \exp(MG_{\text{avg}}). \tag{75}$$

This is a monomial equality constraint, and so is GP compatible.

To find the Ge-profile that minimizes base transit time, given the doping profile, we solve the GP

$$\begin{array}{ll} \text{minimize} & \tau_{B,\text{SiGe}} \\ \text{subject to} & \prod_{i=0}^{M-1} z_i = \exp(MG_{\text{avg}}) \\ & 1 \le z_i \le \exp(G_{\max}), \quad i = 0, 1, \ldots, M-1 \end{array} \tag{76}$$

with variables $z_0, \ldots, z_{M-1}$. The optimal Ge-profile is then given by

$$G(x_i) = \log(z_i), \quad i = 0, \ldots, M-1.$$

D. Joint Profile Optimization

We now consider the most general problem: finding the doping profile *and* the Ge-profile, simultaneously, that minimize base transit time. To the authors' knowledge, this problem has not been considered before.

We first observe that $\tau_{B,\text{SiGe}}$ in (72) is a posynomial of the variables $v_0, \ldots, v_{M-1}$ *and* $z_0, \ldots, z_{M-1}$. To form the joint optimization problem, we simply combine constraints from the doping problem (68) and the optimal Ge-profile problem (76), to obtain

$$\begin{array}{ll} \text{minimize} & \tau_{B,\text{SiGe}} \\ \text{subject to} & \prod_{i=0}^{M-1} z_i = \exp(MG_{\text{avg}}) \\ & 1 \le z_i \le \exp(G_{\max}), \quad i = 0, \ldots, M-1 \\ & N_{\min} \le v_i \le N_{\max}, \quad i = 0, \ldots, M-1. \end{array} \tag{77}$$

The problem is a GP with variables $v_0, \ldots, v_{M-1}$ and $z_0, \ldots, z_{M-1}$.

By solving this GP, we obtain the *jointly optimal* profiles, i.e., doping profile and the Ge-profile that jointly minimize the base transit time. This provides a one-step method, that finds the globally optimal solution. In particular, it is guaranteed to out-perform ad hoc methods for joint optimization, such as alternating between doping profile optimization (with fixed Ge-profile) and Ge-profile optimization (with fixed doping profile).

E. Sparse Joint Profile Optimization Formulation

We can obtain a sparse GP formulation of the joint doping profile problem (77), as an extension of the sparse GP formulation for homojunction BJT doping profile problem. The sparse GP formulation is

$$\begin{array}{ll} \text{minimize} & \tau_{B,\text{SiGe}} = \tilde{C}w_0 \\ \text{subject to} & 1 \le z_i \le \exp(G_{\max}), \quad i = 0, \ldots, M-1 \\ & N_{\min} \le v_i \le N_{\max}, \quad i = 0, \ldots, M-1 \\ & y_{i+1} + z_i^{-\mu} v_i^{1+\gamma_1-\gamma_2} \le y_i, \quad i = 0, \ldots, M-2 \\ & w_{i+1} + z_i^{\mu} v_i^{\gamma_2-1} y_i \le w_i, \quad i = 0, \ldots, M-2 \\ & z_{M-1}^{-\mu} v_{M-1}^{1+\gamma_1-\gamma_2} \le y_{M-1} \\ & z_{M-1}^{\mu} v_{M-1}^{\gamma_2-1} y_{M-1} \le w_{M-1} \\ & x_{M-1} = z_{M-1} \\ & x_i = z_i x_{i+1}, \quad i = 0, \ldots, M-2 \\ & x_0 = \exp(MG_{\text{avg}}) \end{array} \tag{78}$$

where the variables are $v_0, \ldots, v_{M-1}$, $z_0, \ldots, z_{M-1}$, $x_0, \ldots, x_{M-1}$, $y_0, \ldots, y_{M-1}$, and $w_0, \ldots, w_{M-1}$. Each constraint in this problem depends on at most five variables. This is a large, sparse GP and can be solved extremely efficiently.

F. Example

We give a numerical example to illustrate joint doping profile optimization. We use the values from Table I, along with

$$K_{\text{SiGe}} = 3 \quad \mu = 0.688 \text{ eV}/kT$$

which are taken from [17]. (At $T = 300$ °K, we have $\mu = 26.614$.) The maximum Ge-content is taken to be $G_{\max} = 0.25$, and we consider average Ge-content values

$$G_{\text{avg}} = 0,\ 0.02,\ 0.05,\ 0.08,\ 0.11,\ 0.14,\ 0.17,\ 0.2.$$

(The value $G_{\text{avg}} = 0$ corresponds to the BJT optimization problem solved in Section II-F.) The resulting sparse GP has 4000 variables, and 7000 constraints, and is solved (globally) using the MOSEK GP solver [1], in around 6 seconds, on a personal computer.

The *jointly optimal* doping and Ge-profiles are plotted as the solid curves in the two plots in Fig. 8. The optimal base transit time, obtained by the jointly optimal profiles, as a function of the average Ge-content, is shown as the solid curve in Fig. 9. When the Ge-content is zero, the optimal doping profile coincides with the profile shown in Fig. 1. The optimal base transit time attains the minimum when G_{avg} is 0.12. When $G_{\text{avg}} = 0.12$, the optimal base transit time is $\tau^*_{B,\text{SiGe}} = 0.28$ ps, a factor of 5.4 smaller than when no Ge-content is used.

G. Extensions

The extensions described in Section III, for the homojunction BJT case, are readily adapted to the case of HBTs. In particular, a model that accounts for velocity saturation is readily developed, gradient constraints (on doping and Ge-profile) can be imposed, and, a minimum current gain constraint can also be imposed. All of these can be transformed to GP compatible constraints. Constraints on the intrinsic base sheet resistance and breakdown voltage can be approximated to GP compatible constraints.

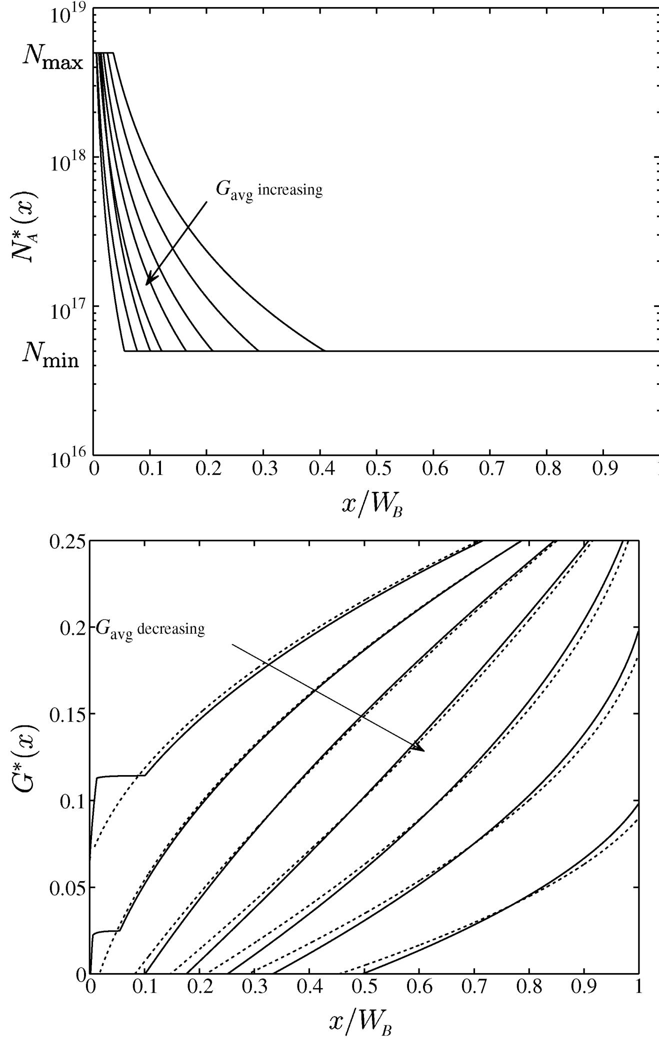

Fig. 8. Optimal doping profiles $N_A^*(x)$ and optimal Ge-profiles $G^*(x)$, with $G_{\max} = 0.25$, for different values of G_{avg}. (Top) Jointly optimal doping profiles. (Bottom) The solid curves show the jointly optimal Ge-profiles. The dotted curves show the optimal Ge-profiles for uniform doping profile $N_A(x) = N_{\min}$.

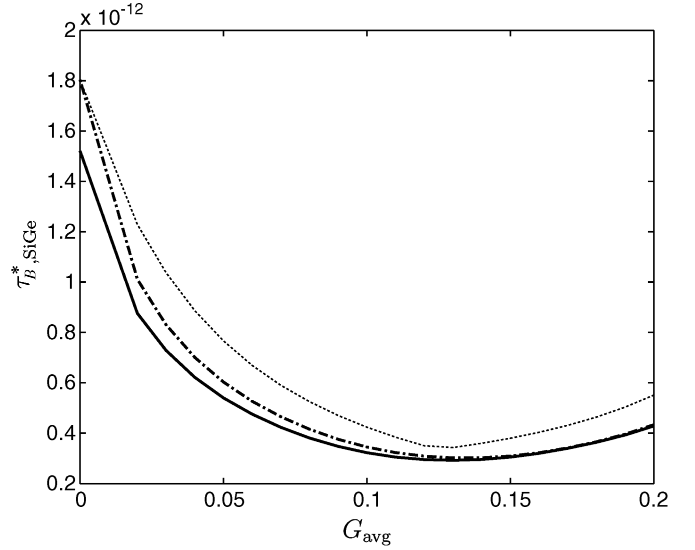

Fig. 9. Optimal base transit time $\tau^*_{B,\text{SiGe}}$ versus average Ge-content G_{avg}, with $G_{\max} = 0.25$. The solid curve shows the optimal base transit time for the jointly optimal profiles. The dotted curve shows the optimal base transit time for the uniform doping profile $N_A(x) = N_{\min}$ and optimal linear Ge-profile. The dot-dashed curve shows the optimal base transit time for uniform doping profile $N_A(x) = N_{\min}$ and optimal Ge-profiles.

H. Comparison With Conventional Suboptimal Profiles

In this section, we compare the optimal doping and Ge-profiles to some more conventional ones, such as a uniform doping profile and optimal linear Ge-profile, uniform doping profile and optimal Ge-profile (i.e., with just the maximum Ge-content constraint and the average Ge-content constraint).

Optimal Linear Ge-Profile: We assume the doping profile is uniform, and the Ge-profile is linear, i.e.

$$N_A(x) = N_A, \quad G(x) = g_0 + mx, \quad 0 \le x \le W_B.$$

Given G_{avg} and g_0, the slope m can be determined as $m = 2(G_{\text{avg}} - g_0)/W_B$. This profile is usually referred to as a *triangular profile* when $g_0 = 0$, and a *trapezoidal profile* when $g_0 > 0$. For this joint profile the base transit time is given by

$$\tau_{B,\text{SiGe}} = \frac{W_B^2}{D_{n0}(1+K_{\text{SiGe}}G_{\text{avg}})}\left(\frac{N_A}{N_{\text{ref}}}\right)^{\gamma_1}\left(\frac{\exp(-\rho)-1+\rho}{\rho^2}\right) \quad (79)$$

where $\rho = 2\mu(G_{\text{avg}} - g_0)$.

To minimize $\tau_{B,\text{SiGe}}$, (as before) we choose $N_A = N_{\min}$. The function $(\exp(-\rho) - 1 + \rho)/\rho^2$ is decreasing in ρ, so to minimize it we should choose g_0 as small as possible. Therefore, the linear Ge-profile is increasing in x. The maximum value of the Ge fraction is attained at $x = W_B$, which should be less than $G_{\max}$. This gives a constraint on how small g_0 can be, which is, $g_0 \ge 2G_{\text{avg}} - G_{\max}$. Thus, the optimal value of g_0 for the linear Ge-profile is

$$g_0 = \max\{2G_{\text{avg}} - G_{\max}, 0\}. \quad (80)$$

Given a value of $G_{\max}$ and G_{avg}, the optimal base transit time for this joint profile, which is a uniform doping profile $N_A = N_{\min}$, and linear Ge-profile, is given by (79) and (80). For $G_{\max} = 0.25$ and G_{avg} between 0 and 0.2 the optimal base transit time obtained by this joint profile is plotted as the dotted curve in Fig. 9. The difference in the optimal base transit times obtained by this joint profile and the jointly optimal profile is the smallest for values of G_{avg} around 0.12. This is because the optimal linear profile comes closest to the optimal Ge-profile (shown in Fig. 8) when G_{avg} is around 0.12, but is quite different for values of G_{avg} at the two extreme points.

We note that, as $G_{\text{avg}} \to 0$, $\rho \to 0$, we have $(\exp(-\rho) - 1 + \rho)/\rho^2 \to 1/2$. Thus, the base transit time $\tau_{B,\text{SiGe}}$ in (79) reduces to τ_B in (15), as expected. This is verified by Fig. 9, which shows the optimal base transit time obtained at $G_{\text{avg}} = 0$ is 1.80 ps, the same as obtained earlier by (15) for $N_A = N_{\min}$.

Optimal Ge-Profile: It is also interesting to compare the optimal base transit time obtained by varying the Ge-content and keeping the doping profile constant at $N_{\min}$, i.e., we want to find the performance loss by choosing a uniform doping profile $N_A(x) = N_{\min}$. Thus, we want to solve the optimization problem (76), with $\tau_{B,\text{SiGe}}$ given by (72) and $v_i = N_{\min}$, $i = 0, \ldots, M-1$. For a given $G_{\max}$ and G_{avg} the optimal base transit time of this joint profile should be between the optimal base transit times obtained by jointly optimal profile

and joint profile with uniform doping profile $N_A(x) = N_{\min}$, and optimal linear Ge-profile. For $G_{\max} = 0.25$, the optimal base transit time for this joint profile is shown in Fig. 9 by the dot-dashed curve. The optimal Ge-profiles are plotted as dotted line in Fig. 8.

When the average Ge-content is low, the base transit time is dominated by the doping profile; but when the average Ge-content is high, the base transit time is dominated by the Ge-profile. Therefore, the optimal base transit curve for the joint profile with constant doping profile $N_A(x) = N_{\min}$, and optimal Ge-profile, in Fig. 9, starts along the optimal base transit time curve for optimal linear Ge-profile, uniform doping profile ($N_A = N_{\min}$) and ends along the optimal base transit time for jointly optimal profiles.

VI. Conclusion

We have shown that a variety of base doping profile optimization problems can be formulated, after discretization, as GPs. These problems range from the simplest formulation, minimizing base transit time subject to bounds on doping concentration, to more complex formulations involving multiple dopants, and constraints on doping gradient and current gain.

Our method has several advantages. First, the GP approach yields efficient computational solutions, that scale linearly with the number of discretized regions, and in any case can be solved in seconds for relatively fine discretizations. The approach is extensible, in the sense that other GP compatible constraints can be added, without loss of efficiency. In the same spirit, mode accurate models of base transit time could be used, provided they are GP compatible.

The method also handles multiple dopants, not just in the modeling but in the optimization as well. We showed this in detail for SiGe profiles in Section V, but the same ideas apply to problems with even more dopants, e.g., $\mathrm{Si}_{1-x-y}\mathrm{Ge}_x\mathrm{C}_y$. For such problems the method gives the optimal solution without the iteration required in a method that cycles through the dopants, optimizing each one with the others fixed. Moreover, the GP approach guarantees that the globally optimal solution is found; this need not be the case for a method that cycles through optimizing over several dopants separately.

Another general advantage of the GP method is that it is guaranteed to always find the globally optimal solution. In particular, the GP method cannot be "trapped" in a locally optimal design. While we suspect that other methods, such as the iterative scheme proposed by Kumar and Patri [19], often find the global solution, it is not clear to us why this should always be the case. Since the global solution is always found by the GP method, it can be used to find the (absolute) limit of performance. This can be useful in practice even when a simpler profile will ultimately be implemented. By comparing the performance achieved by a simple profile (such as the ones found in Section V-H) to the optimal performance, we can assess performance lost by using a simple profile.

The main limitation of the GP approach lies in the models, which are restricted to have a specific analytical form, i.e., posynomial or generalized posynomial. While this form is fairly general, and includes a number of fairly accurate models that have been used in the literature, we do not expect analytical models like the ones described in this paper to have the ultimate accuracy of more complex TCAD models and tools. This observation suggests a method for combining the GP-based methods described in this paper with device optimization methods that rely on detailed device simulations. GP-based methods (which are fast, and guarantee global solutions, but rely on less accurate analytical models) are used to rapidly explore the performance trade-offs, and to get very good starting points. Then, simulation-based methods (which are slow, find only local solutions, but give high accuracy) are then used to do the final tuning (and verification) of the design.

Appendix

A. Bound on Discretization Error

In this appendix we give a simple bound on the discretization error $|\tau_B - \hat{\tau}_B|$. We assume the doping profile $N_A(x)$ is Lipschitz continuous on $[0, W_B]$, i.e., satisfies

$$|N_A(y) - N_A(x)| \le l|y - x|, \quad \forall x, y \in [0, W_B] \qquad (81)$$

for some l, called the Lipschitz constant. We also assume that $N_A(x)$ satisfies

$$0 < N_{\min} \le N_A(x) \le N_{\max}, \quad \text{for } 0 \le x \le W_B.$$

We will show that the function $N_A(x)^\alpha$ is also Lipschitz on $[0, W_B]$, for any $\alpha \in \mathbf{R}$. Start with

$$N_A(y)^\alpha - N_A(x)^\alpha = \int_x^y dN_A(x)^\alpha = \int_x^y \alpha N_A(x)^{\alpha-1} N_A'(x) dx. \qquad (82)$$

As $N_A(x)$ is Lipschitz, $N_A'(x)$ exists (almost everywhere) and satisfies $|N_A'(x)| \le l$. The function $N_A(x)^{\alpha-1}$ is also bounded and thus the integral on the right in (82) is well defined on $[0, W_B]$. Let U_α be an upper bound on $N_A(x)^{\alpha-1}$, i.e., take $U_\alpha = N_{\max}^{1-\alpha}$ if $\alpha > 1$, and $U_\alpha = N_{\min}^{\alpha-1}$ if $\alpha < 1$. Then

$$\begin{aligned} |N_A(y)^\alpha - N_A(x)^\alpha| &= \left| \int_x^y \alpha N_A(x)^{\alpha-1} N_A'(x) dx \right| \\ &\le |\alpha| U_\alpha l \left| \int_x^y dx \right| \\ &= |\alpha| U_\alpha l |y - x| \end{aligned} \qquad .(83)$$

Thus $N_A(x)^\alpha$ is Lipschitz on $[0, W_B]$, with Lipschitz constant $l_\alpha = |\alpha| U_\alpha l$.

Now we show that $N_A(x)^\alpha N_A(y)^\beta$ is Lipschitz continuous on $[0, W_B] \times [0, W_B]$, for any α, $\beta \in \mathbf{R}$. This is the integrand which we deal with in expression (6) of τ_B. Consider

$$\begin{aligned} &\left| N_A(x)^\alpha N_A(y)^\beta - N_A(u)^\alpha N_A(v)^\beta \right| \\ &= \left| N_A(x)^\alpha \left(N_A(y)^\beta - N_A(v)^\beta \right) \right. \\ &\quad \left. + N_A(v)^\beta \left(N_A(x)^\alpha - N_A(u)^\alpha \right) \right| \\ &\le N_A(x)^\alpha l_\beta |y - v| + N_A(v)^\beta l_\alpha |x - u| \\ &\le V_\alpha l_\beta |y - v| + V_\beta l_\alpha |x - u| \\ &\le L_{(\alpha,\beta)} \left(|x - u| + |y - v| \right) \end{aligned} \qquad (84)$$

where $V_\alpha = N_{\max}^\alpha$ if $\alpha > 0$, and $V_\alpha = N_{\min}^\alpha$ if $\alpha < 0$, and $L_{(\alpha,\beta)} = 2\max\{V_\alpha l_\beta, V_\beta l_\alpha\}$. Thus, $N_A(x)^\alpha N_A(y)^\beta$ is Lipschitz.

Consider τ_B in (6) and $\hat{\tau}_B$ in (8). Let $\alpha = \gamma_2 - 1$, and $\beta = 1 + \gamma_1 - \gamma_2$. Then the difference between τ_B and $\hat{\tau}_B$ can be expressed as

$$\tau_B - \hat{\tau}_B = \frac{1}{N_{\text{ref}}^{\gamma_1} D_{n0}} \left(\int_0^{W_B} \int_x^{W_B} N_A(x)^\alpha N_A(y)^\beta dy dx - \sum_{i=0}^{M-1} \sum_{j=i}^{M-1} N_A(x_i)^\alpha N_A(x_j)^\beta \left(\frac{W_B}{M}\right)^2 \right). \quad (85)$$

Using the Lipschitz condition the change in the value of the function $N_A(x)^\alpha N_A(y)^\beta$ over a square of size $(W_B/M) \times (W_B/M)$ is less than or equal to $2L_{(\alpha,\beta)} W_B/M$; and there are $M(M+1)/2$ such squares. It follows that

$$|\tau_B - \hat{\tau}_B| \le \frac{2L_{(\alpha,\beta)} W_B}{M} \left(\frac{W_B}{M}\right)^2 \frac{M(M+1)}{2} \le \frac{2L_{(\alpha,\beta)} W_B^3}{M}. \quad (86)$$

This bound on the error shows that as M becomes large the error goes to zero, i.e., $\lim_{M\to\infty} |\tau_B - \hat{\tau}_B| = 0$. More sophisticated bounds on the error can be obtained but the simple bound (86) serves our purpose.

ACKNOWLEDGMENT

The authors would like to thank J. Plummer and K. Suzuki for very helpful comments and suggestions, G. Niu for providing very helpful insight, and suggestions related to SiGe devices and directions for further studies, two anonymous reviewers for very helpful suggestions on, in particular, the sections on velocity saturation, base sheet resistance, breakdown voltage and cutoff frequency optimization, which considerably improved this paper.

REFERENCES

[1] Mosek ApS, Optimization Software, E. Andersen and K. Andersen.*www.mosek.com/* [Online]

[2] P. Ashburn, *SiGe Heterojunction Bipolar Transistors*. New York: Wiley, 2003.

[3] A. Biswas and P. Basu, "An analytical approach to the modeling of intrinsic base sheet resistance in SiGe HBT and optimal profile design considerations for its minimization," in *Institute of Physics Publishing, Semiconductor Science and Technology*, Nov. 2002, pp. 1249–1254.

[4] ——, "Modeling of base transit time in $Si/Si_{1-y-z}Ge_yC_z/Si$ HBTs and composition profile design issue for its minimization," in *Institute of Physics Publishing, Semiconductor Science and Technology*, Aug. 2003, pp. 907–913.

[5] S. Boyd, S.-J. Kim, D. Patil, and M. Horowitz, "Digital circuit sizing via geometric programming," *Oper. Res.*, vol. 53, no. 6, 2005.

[6] S. Boyd, S.-J. Kim, L. Vandenberghe, and A. Hassibi, "A tutorial on geometric programming," Opt. Eng., 2005, to be published.

[7] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[8] D. Colleran, C. Portmann, A. Hassibi, C. Crusius, S. Mohan, S. Boyd, T. Lee, and M. Hershenson, "Optimization of phase-locked loop circuits via geometric programming," in *Proc. Custom Integrated Circuits Conf.*, Sep. 2003, pp. 326–328.

[9] J. D. Cressler and G. Niu, *Silicon-Germanium Heterojunction Bipolar Transistors*. Norwell, MA: Artech House, Dec. 2003.

[10] J. Dawson, S. Boyd, M. Hershenson, and T. Lee, "Optimal allocation of local feedback in multistage amplifiers via geometric programming," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 48, no. 1, pp. 1–11, Jan. 2001.

[11] D. Harame, J. H. Comfort, J. D. Cressler, E. F. Crabbe, J. Y.-C. Sun, B. S. Meyerson, and T. Tice, "Si/SiGe epitaxial-base transistors. I. Materials, physics, and circuits," *IEEE Trans. Electron Devices*, vol. 43, no. 3, pp. 455–468, Mar. 1995.

[12] A. Hassibi and M. Hershenson, "Automated optimal design of switched-capacitor filters," in *Proc. Design, Autmation Test Eur. Conf. Exhibition*, Paris, France, Mar. 2002, p. 1111.

[13] M. Hershenson, S. Boyd, and T. Lee, "Optimal design of a CMOS op-amp via geometric programming," *IEEE Trans. Computer-Aided Des. Integr. Circuits Syst.*, vol. 20, no. 1, pp. 1–21, Mar. 2001.

[14] M. Hershenson, A. Hajimiri, S. Mohan, S. Boyd, and T. Lee, "Design and optimization of *LC* oscillators," in *Proc. IEEE/ACM Int. Conf. Computer-Aided Design*, 1999, pp. 65–69.

[15] K. Kortanek, X. Xu, and Y. Ye, "An infeasible interior-point algorithm for solving primal and dual geometric progams," *Math. Program.*, vol. 76, no. 1, pp. 155–181, Jan. 1996.

[16] H. Kroemer, "Two integral relations pertaining to the electron transport through a bipolar transistor with a nonuniform energy gap in the base region," *Solid State Electron.*, vol. 28, no. 11, pp. 1101–1103, Feb. 1985.

[17] M. Kumar and V. Patri, "Profile design considerations for minimizing base transit time in SiGe HBTs.," *IEEE Trans. Electron Devices*, vol. 45, no. 8, pp. 1725–1731, Aug. 1998.

[18] ——, "Novel Ge-profile design for high-speed SiGe HBTs: Modeling and analysis," *Int. Electr. Eng. Proc. –Circuits Devices Systems*, vol. 146, no. 5, pp. 291–296, Oct. 1999.

[19] ——, "On the iterative schemes to obtain base doping profiles for reducing base transit time in bipolar junction transistor," *IEEE Trans. Electron Devices*, vol. 48, no. 6, pp. 1222–1224, Jun. 2001.

[20] T.-C. Lu and J. B. Kuo, "A closed-form analytic forward transit time model considering specific models for bandgap-narrowing effects and concentration-dependent diffusion coefficients for BJT devices operating at 77K," *IEEE Trans. Electron Devices*, vol. 40, no. 4, pp. 766–772, Apr. 1993.

[21] P. Mandal and V. Visvanathan, "CMOS op-amp sizing using a geometric programming formulation," *IEEE Trans. Computer-Aided Des. Integr. Circuits Syst.*, vol. 20, no. 1, pp. 22–38, Jan. 2001.

[22] S. Mohan, M. Hershenson, S. Boyd, and T. Lee, "Simple accurate expressions for planar spiral inductances," *IEEE J. Solid-State Circuits*, vol. 34, no. 10, pp. 1419–1424, Oct. 1999.

[23] R. F. Pierret, *Semiconductor Device Fundamentals*. Reading, PA: Addison-Wesley, Mar. 1996.

[24] D. M. Richey, J. D. Cressler, and A. J. Joseph, "Scaling issues and Ge profile optimization in advanced UHV/CVD SiGe HBTs," *IEEE Trans. Electron Devices*, vol. 44, no. 3, pp. 431–440, Mar. 1997.

[25] P. Rinaldi and H. Schättler, "Minimization of the base transit time in semiconductor devices using optimal control," in *Proc. Int. Conf. Dynamical Systems Differential Equations*, May 2002, pp. 1–10.

[26] ——, "An optimal control problem with state space constraints arising in the design of bipolar transistors," in *Proc. IEEE Conf. Decision Control*, Dec. 2004.

[27] S. Sapatnekar, *Timing*. New York: Kluwer Academic, 2004.

[28] R. Singh, D. Harame, and M. M. Oprysko, *Silicon Germanium: Technology, Modeling, and Design*. New York: Wiley, Oct. 2003.

[29] K. Suzuki, "Optimal base profile design for minimum base transit time," *IEEE Trans. Electron Devices*, vol. 38, no. 9, pp. 2128–2133, Sep. 1991.

[30] ——, "Optimum base–doping profile for minimum base transit time considering velocity saturation at base–collector junction and dependence of mobility and bandgap narrowing on doping concentration," *IEEE Trans. Electron Devices*, vol. 48, no. 9, pp. 2102–2107, Sep. 2001.

[31] K. Suzuki and N. Nakayama, "Base transit time of shallow-base bipolar transistor considering velocity saturation at base–collector junction," *IEEE Trans. Electron Devices*, vol. 39, no. 3, pp. 623–628, Mar. 1992.

[32] Y. Taur and T. Ning, *Fundamentals of Modern VLSI Devices*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[33] J. Vanderhaegen and R. Brodersen, "Automated design of operational transconductance amplifiers using reversed geometric programming," in *Proc. IEEE/ACM Design Automation Conf.*, San Diego, CA, Jun. 2004, pp. 133–138.

Siddharth Joshi received the B. Tech. (with Honors) degree in electrical engineering from the Indian Institute of Technology, Kharagpur, India, and the M.S. degree in electrical engineering from Stanford University, Stanford, CA, in 2002 and 2004, respectively. He is currently pursuing the Ph.D. degree in electrical engineering at Stanford University.

His current interests include application of convex optimization to various engineering applications.

Mr. Siddharth is a recepient of the Stanford Graduate Fellowship.

Stephen Boyd (S'82–M'85–SM'97–F'99) received the A.B. degree in mathematics from Harvard University, Cambridge, MA, in 1980, and the Ph.D. in electrical engineering and computer science from the University of California, Berkeley, in 1985.

He is the Samsung Professor of Engineering and Professor of Electrical Engineering in the Information Systems Laboratory at Stanford University, Stanford University. His current research focus is on convex optimization applications in control, signal processing, and circuit design.

Robert W. Dutton (M'70–SM'80–F'84) received the B.S., M.S., and Ph.D. degrees from the University of California, Berkeley, in 1966, 1967, and 1970, respectively.

He is Professor of Electrical Engineering, Stanford University, Stanford, CA, and Director of the Integrated Circuits Laboratory. He has held summer staff positions at Fairchild, Bell Telephone Laboratories, Hewlett-Packard, IBM Research, and Matsushita, respectively. His research interests focus on integrated circuit process, device, and circuit technologies–especially the use of computer-aided design (CAD) in device scaling and for RF applications.