Impact Ionization and Hot-Electron Injection Derived Consistently from Boltzmann Transport

Paul Hasler, Andreas G. Andreou, Chris Diorio, Bradley A. Minch, and Carver A. Mead

Abstract— **We develop a quantitative model of the impact-ionization and hot-electron–injection processes in MOS devices from first principles. We begin by modeling hot-electron transport in the drain-to-channel depletion region using the spatially varying Boltzmann transport equation, and we analytically find a self consistent distribution function in a two step process. From the electron distribution function, we calculate the probabilities of impact ionization and hot-electron injection as functions of channel current, drain voltage, and floating-gate voltage. We compare our analytical model results to measurements in long-channel devices. The model simultaneously fits both the hot-electron-injection and impact-ionization data. These analytical results yield an energy-dependent impact-ionization collision rate that is consistent with numerically calculated collision rates reported in the literature.**

We develop a quantitative analytical model of the impact-ionization and hot-electron–injection processes in MOS devices that is derived consistently from a single spatially varying hot-electron distribution function. This approach not only provides a useful circuit model, but also complements and validates numerical results from Monte Carlo simulations.

We measure hot-electron–injection (gate) and impact-ionization (substrate) currents using an n-type MOSFET built with a high substrate doping ($1 \times 10^{17}\mathrm{cm}^{-3}$) operating with subthreshold currents. Figure 1 illustrates the cross section of the device. In subthreshold the channel current of a MOSFET is sufficiently small so that the mobile charge does not affect the surrounding electrostatics, resulting in a constant surface potential. Consequently, by operating the MOSFET in subthreshold, we obtain a high field region whose properties are independent of the channel current. This higher substrate doping is consistent with a 0.3μm channel length CMOS process; thus, these effects are directly applicable to modern processes.

For an electron to reach the floating gate, it must have energy greater than the oxide barrier height and must be directed towards the SiO_2 when the electron reaches that energy. The high electric fields in the drain-to-channel depletion region accelerate channel electrons to high energies (path 1). The high substrate doping increases the threshold voltage ($\approx$ 6V) and the drain-to-channel electric field, which generates high-energy electrons at subthreshold currents for positive gate-to-drain voltages; therefore, an electron surmounting the Si–SiO_2 barrier will be transported to the gate by the resulting oxide field (path 2).

As an electron gains energy due to the electric field in the z direction, the electron is confined by the electric field and the silicon–silicon-dioxide interface in the y direction.

The authors were associated with California Institute of Technology, Pasadena, CA 91125. E-mail: paul@ee.gatech.edu.

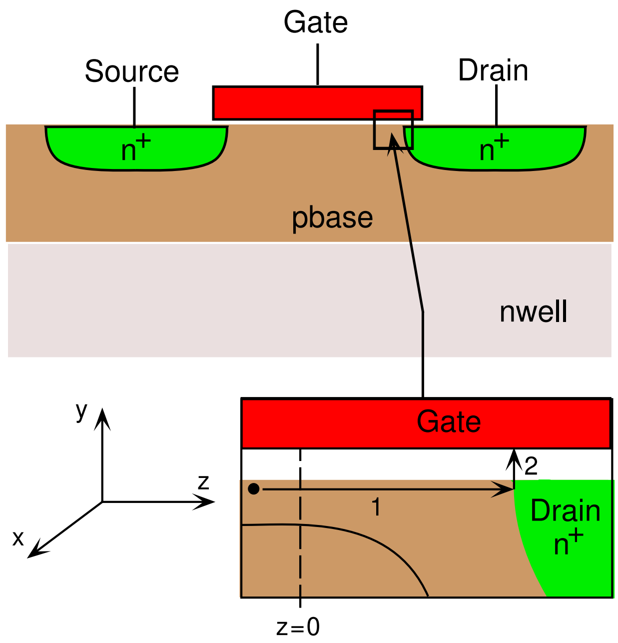

Fig. 1. Cross section of the MOSFET device we used to measure the hot- electron effects. It uses a highly doped ($1 \times 10^{17}\mathrm{cm}^{-3}$) substrate to achieve a high threshold voltage which allows hot-electron injection for bias current levels in subthreshold. The $n$well isolates the highly doped substrate region from the surrounding substrate, and allows measurement of substrate current. Holes resulting from impact ionization are measured at the $p$base contact. The hot-electron injection process is identical for the FET with or without the isolating nwell. Inset: the electron is accelerated through the drain depletion (path 1), and when it gains energy greater than the Si–SiO_2 barrier, the electron is injected over the Si–SiO_2 barrier to the floating-gate (path 2).

The resulting electron distribution in y and k_y is nearly independent of the electron distribution in the other coordinates; therefore, some electrons at $y = 0$ are directed toward the SiO_2, and these electrons will enter the $Si0_2$ if they have gained sufficient energy.

I. Electron Transport in the Drain to Channel Depletion Region

We begin by modeling hot-electron transport in the drain-to-channel depletion region using the spatially varying Boltzmann transport equation. We can simplify the general Boltzmann equation to a 1-D problem along the channel (z) axis [1]; Fig. 2 shows the conduction band as a function of position through the MOSFET's channel re-

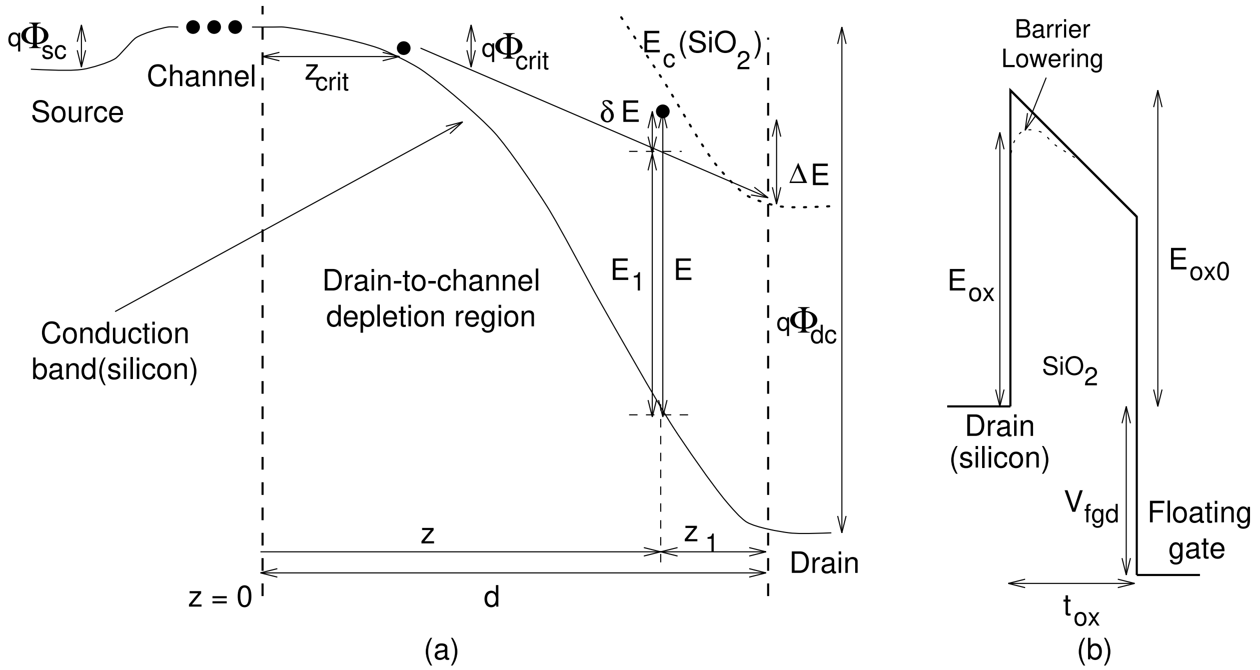

Fig. 2. Band diagram illustrating hot-electron injection in a MOSFET biased in subthreshold. The appropriate variables in the Boltzmann transport equation and its variable transformations are shown on the graphs. (a) Band diagram along the surface of the Si - SiO_2 barrier. This region is the lowest local potential in either material; therefore the electrons are most likely to travel along this path. This region corresponds to path 1 in the inset in Fig. 1. (b) Band Diagram of at the drain edge. This region corresponds to path 2 in the inset in Fig. 1.

gion. Following a similar procedure to Baraff [2], we get

$$\frac{\partial f}{\partial z} + q\mathcal{E}\frac{\partial f}{\partial E} + q\mathcal{E}\frac{1-\zeta^2}{\zeta E}\frac{\partial f}{\partial \zeta} = \frac{m^*(c)}{\zeta c}S(f), \tag{1}$$

where $f(z,c,\zeta)$ is the distribution function, $\mathcal{E}(z)$ is the component of the electric field in the z direction, c is the magnitude of the average momentum vector, ζ is the cosine of the angle of momentum vector and the z axis, and $S(f)$ is the collision operator. $E = c^2/m^*(c)$ is the electron energy, where $m^*(c)$ is the effective mass of the electron that depends upon the silicon band structure.

Starting from Conwell's optical-phonon collision operator [3], we derive the following approximate optical-phonon collision operator for $E \gg E_R$ [1]:

$$S_{op}(f) \approx \left(\frac{c(E)}{m^*(E)}\right)\left(\frac{E_R^2}{2\lambda}\frac{\partial^2 f}{\partial E^2} + \frac{E_R}{\lambda}\frac{\partial f}{\partial E}\right), \tag{2}$$

where E_R is the energy of an optical phonon (E_R= 63meV in Si). A similar expansion and simplification has been done for polar optical phonons [4]. The mean free length for phonon collisions (λ) is known to be approximately constant for high energies. We can remove the bandstructure effects in (eq:boltz2a) by developing our collision models only in terms of a mean free length and terms of $\frac{c(E)}{m^*(E)}$. [1] Phonons have momentum, and the total momentum involved for a phonon absorption or emission must be conserved. To precisely model this effect, one would need to know the distribution function of momentum for the phonons in the drain–to–channel depletion region. Elsewhere we show that the scattering of the momentum distribution has a small effect on our zeroth order expressions [1].

Most proposed impact-ionization collision operators can be formulated in general as

$$S(f)_{ion} = -\frac{f}{\tau_{ion}(E)} = -\frac{c(E)}{m^*(E)}\frac{1}{L(E)}f \tag{3}$$

where τ_{ion} is the mean free time for an impact ionization collision, and L(E) is the mean free path, which is a function of the electron energy. We propose the following model for the energy dependence for the impact-ionization mean-free length

$$L(E) = (0.181\text{Å})\exp\left(\sqrt{\frac{119\text{eV}}{E - .95\text{eV}}}\right), \tag{4}$$

which is based on our experimental measurements of the impact-ionization mean free length, and corresponds to previous numerical calculations [5], [6], [7]. Figure 3 shows our functional form with these three numerically calculated models. We have assumed a constant velocity of 8.1×10^6 cm/s in converting from $L(E)$ to impact-ionization scattering rate, since our measured data is directly related to $L(E)$. This functional form is a curve fit to experimental data of $L(E)$ derived from our experimental measurements of hot-electron-injection and impact-ionization currents in Section III.

[1] Canceling out the effects of the bandstructure may limit the predictive power of this model. This insight by Karl Hess is appreciated.

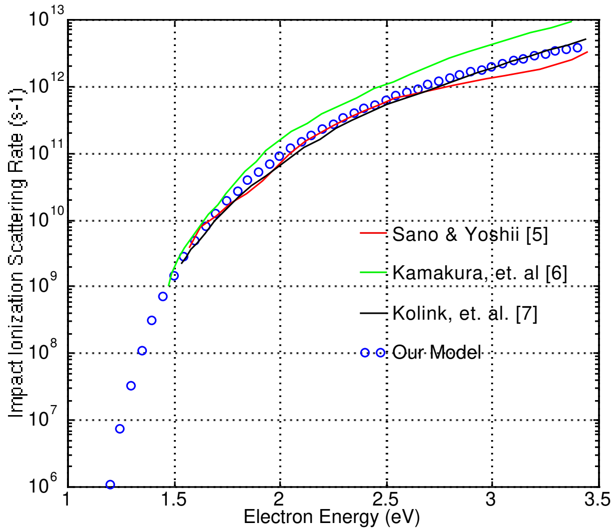

Fig. 3. Plot of previous calculations of impact-ionization rate versus electron energy in silicon, and our derived impact-ionization rate from our measured impact-ionization and hot-electron injection data. We have assumed a constant velocity, since our model measures the impact-ionization mean-free length. Our measured data is directly related to $L(E)$ and not impact-ionization scattering rate.

II. Solution of the Transport Equation

We analytically solve the resulting Boltzmann transport equation,

$$\frac{\partial f}{\partial z} + \left(q\mathcal{E} - \frac{E_R}{\lambda \zeta} \right) \frac{\partial f}{\partial E} + q\mathcal{E} \frac{1-\zeta^2}{\zeta E} \frac{\partial f}{\partial \zeta} = \frac{E_R^2}{2\lambda\zeta} \frac{\partial^2 f}{\partial E^2} - \frac{f}{L(E)}, \quad (5)$$

for a self-consistent distribution function using a two-step process. Elsewhere, we show that the transport along $\zeta = 1$ for hot-electron injection and impact ionization closely approximates the exact solution [1]; for clarity, we will only consider the $\zeta = 1$ case here. In the first step, we solve for the average hot-electron trajectory in energy and direction as a function of position through the depletion region. The average hot-electron trajectory is the flow line for the hyperbolic P.D.E. operator, and is related to the numerical method that Budd presented previously [9]. In this model, the average electron starts gaining energy at the position (z_{crit} in Fig. 2) where the phonon restoring force is equal to the energy increase due to the local electric field ($\mathcal{E}(z)$). This breakaway field—the minimum electric field at which the electron gains energy at the same rate as it loses energy to phonon collisions—is expressed as $\frac{E_R}{q\lambda}$, which for our parameters is 9.7V/μm. The average energy, $E_1(z)$, that the electron gains after reaching z_{crit} is

$$E_1(z) = qV(z) - qV(z_{crit}) - E_R \frac{z - z_{crit}}{\lambda}, \quad (6)$$

or the difference between the potential from z_{crit} to the position z in the drain-to-channel depletion region, and the number of phonon collisions in this region. We show the electron in Fig. 2 taking a linear path because of the functional form of (6).

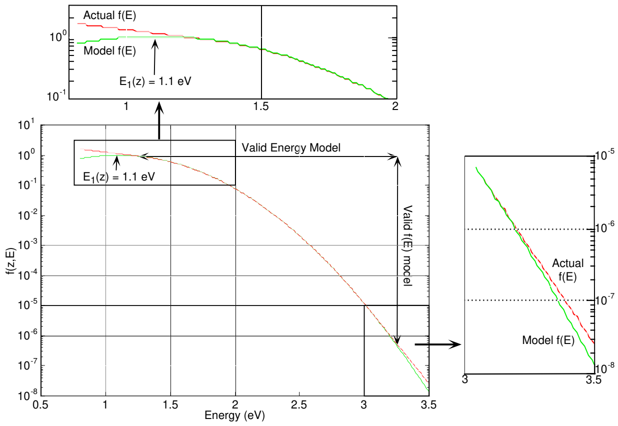

Fig. 4. Picture of the distribution function for an electron in the drain-to-channel beyond $z = z_{crit}$. This figure compares our approximate model to the solution using the exact conditions around z_{crit}. For energies at or below the average electron energy, the distribution function shows the cumulative effect of electrons leaving the conduction band after z_{crit}. For large positive energies, the distribution function does not change as fast as the Gaussian, but rather at a slope due to an effective temperature.

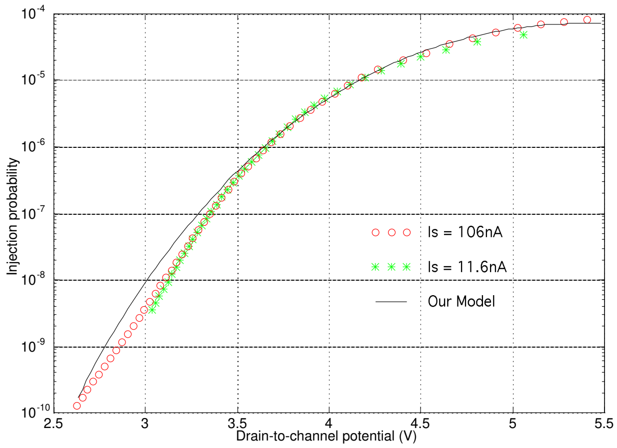

Fig. 5. Measurements of hot-electron–injection efficiency verses drain-to-channel voltage for two values of source current. The drain-to-channel voltage is computed from the source current and the drain-to-source voltage. For each sweep, we used a constant gate voltage to chose a particular channel current; the actual oxide barrier height changes slightly due to image force lowering, because the floating-gate–to–drain voltage is not constant.

In the second step, we solve for the electron distribution function around this average electron trajectory. In this coordinate system, phonon collisions diffuse the electron distribution spatially, and impact-ionization collisions remove high-energy electrons. To simplify the analysis, we assume that the electron leaving at z_{crit} dominates the behavior of hot-electron injection and impact-ionization for a wide range of drain voltages; the limitations of this approximation are illustrated in Fig. 4. Using a more complicated initial and boundary conditions, f(z,E) nearly follows an effective temperature solution for high electron energies, and f(z,E) is the convolution of several Gaussians at low energies. From this analysis, the solution for the distribution

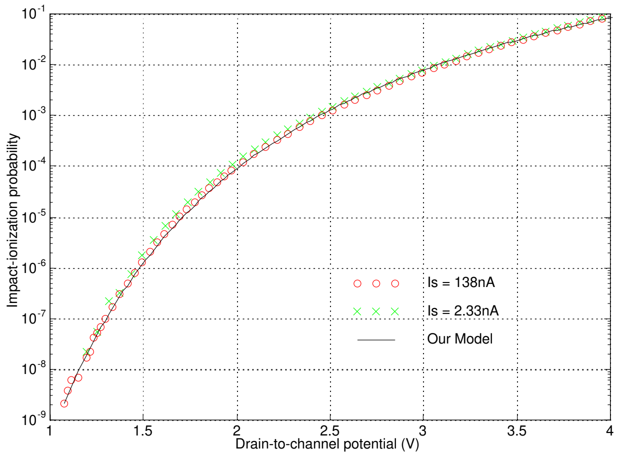

Fig. 6. Measurements of impact-ionization efficiency vs. drain to channel voltage for two source currents (gate voltages). We plot a curve fit to the analytic model in (11); the model closely agrees with the experimental data.

function, $f(z,E)$, is

$$f(z,E) = \exp\left(-\frac{\lambda}{z - z_{crit}}\left(\frac{E - E_1(z)}{2E_R}\right)^2\right) a(z,E), \quad (7)$$

where $a(z,E)$ models the electrons lost to impact ionization, and is approximated by

$$a(z,E) = \exp\left(-\frac{1}{(q\mathcal{E}(z)\lambda - E_R)}\int_{E=0}^{E}\frac{\lambda}{L(E)}dE\right). \quad (8)$$

This solution shows that the assumption of a constant electron temperature is not valid at energies at which impact ionization and hot-electron injection occur.

III. Comparing Theory with Experiment

From the electron distribution function in (7), we can calculate the probabilities of impact ionization and hot-electron injection as functions of channel current and drain-to-channel voltage (Φ_{dc}). We use two free parameters, λ and E_{ox}, as well as our functional form for $L(E)$. The hot-electron–injection efficiency—the ratio of the injection current (I_{inj}) and the source current (I_s)—is approximately given by

$$\frac{I_{inj}}{I_s} = B_2 \exp\left(-\frac{\lambda}{d - z_{crit}}\left(\frac{E_{ox} - E_1(d)}{2E_R}\right)^2 - \frac{7.102}{\sqrt{E_1(d)}}\right), \quad (9)$$

where λ is equal to 6.5nm, $E_{ox} \approx 2.8$eV is the Si–SiO_2 barrier height at the drain, $B_2 = 4.55 \times 10^{-3}$, $d(\Phi_{dc})$ is the width of the drain-to-channel depletion region, $\sqrt{E_1(d)} = \sqrt{\Phi_{dc}} - \sqrt{\Phi_{crit}}$, and Φ_{crit} is the potential drop from $z = 0$ to $z = z_{crit}$. Our experimental data on the Early voltage versus Φ_{dc} show that the channel doping profile is approximately a step junction for a fixed gate voltage [1]. Figure 5 shows measured data of hot-electron–injection efficiencies as a function of drain-to-channel voltage for two channel currents; the hot-electron–injection efficiency is independent of source current. Figure 5 shows (9) fitted to the injection efficiency data. The curve fit shows close agreement to (9) except at large Φ_{dc} (> 5.0V), due to average-electron energy being near the energy of the silicon–silicon-dioxide barrier, and at small Φ_{dc}, probably due to the simplified modeling of the band-structure effects in the collision operators.

The impact-ionization efficiency—the ratio of the substrate current (I_{sub}) and the source current (I_s)—is

$$\frac{I_{sub}}{I_s} = \frac{\int_0^\infty (1 - a(d,E))\, e^{-\frac{\lambda}{d - z_{crit}}\left(\frac{E - E_1(d)}{2E_R}\right)^2} dE}{\int_0^\infty e^{-\frac{\lambda}{d - z_{crit}}\left(\frac{E - E_1(d)}{2E_R}\right)^2} dE.} \quad (10)$$

We get an approximate solution by substituting (4), (8) and expanding the function in the exponent around the function's maximum value in E. We show the general solution elsewhere [1]; the solution for substrate doping of $N_a = 10^{17}\text{cm}^{-3}$ is

$$\alpha = \frac{I_{sub}}{I_s} = e^{-\sqrt{\frac{179.5\text{V}}{\Phi_{dc}}}} \exp\left(\frac{\sqrt{28\text{V}}\left(1 - \frac{E_1(d)}{3.36\text{eV}}\right)^2}{\sqrt{\Phi_{dc}} - \sqrt{\Phi_{crit}}}\right) \quad (11)$$

Figure 6 shows experimental measurements of α versus drain-to-channel potential. The solid line is the curve fit of (11) to the experimental data; the fit closely agrees with the measured data. From measured values of α versus Φ_{dc}, our analytical model allows us to measure the energy-dependent impact-ionization collision rate from experimental data; (4) is a curve fit to these data.

Acknowledgements

We thank K. Hess for several helpful comments and suggestions, and L. Dupre for editing this manuscript.

References

[1] P. Hasler, "Impact Ionization and Hot-Electron Injection in MOSFETs Derived Consistently from Boltzmann Transport," in *Foundations of Learning in Analog VLSI*, Ph.D. Thesis, Computation and Neural Systems, California Institute of Technology, February 1997. *Also at www.pcmp.caltech.edu/anaprose/paul.*

[2] G.A. Baraff, "Maximum Anisotropy Approximation for Calculating Electron Distributions; Application to High Field Transport in Semiconductors," *Physical Review*, Vol. 133, 1964, pp. A26-A33.

[3] E. M. Conwell, *High Field Transport in Semiconductors*, Academic Press, New York, 1967.

[4] J. P. LeBurton and K. Hess, "Energy - diffusion equation for an electron gas interacting with polar optical phonons," *Physical Review B*, vol. 26, no. 10, pp 5623-5633.

[5] N. Sano, and A. Yoshii, "Impact ionization rate near thresholds in Si," *Journal of Applied Physics*, vol. 75, 1994, pp. 5102-5105.

[6] Y.Kamakara, H. Mizuno, M. Yamaji, M. Morifuji, K. Taniguchi, C. Hamaguchi, T. Kunikiyo, and M. Takenaka, "Impact ionization model for full band Monte Carlo simulation," *Journal of Applied Physics*, vol. 75, 1994, pp. 3500-3506.

[7] J. Kolnik, Y. Wang, I.H. Oguzman, and K.F. Brennan, "Theoretical investigation of wave-vector-dependent analytical and numerical formulations of the interband impact-ionization transition rate for electrons in bulk silicon and GaAs," *Journal of Applied Physics*, Vol. 76, 1994, pp. 3542-3551.

[8] C.H. Lee, U. Ravaioli, K. Hess, C. Mead, and P. Hasler, "Simulation of a long term memory device with a full bandstructure

Monte Carlo Approach," *IEEE Electron Device Letters*, vol. 16, no. 8, 1995, pp. 360-362.
[9] H. Budd, "Path variable formulation of the hot carrier problem," *Physical Review*, vol. 158, no. 3, 1967, pp. 798-804.

Author Biographies

Paul Hasler received his B.S.E. and M.S. degrees in electrical engineering from Arizona State University in August 1991, and a Ph.D. in computation and neural systems from the California Institute of Technology in 1997. He is currently an Assistant Professor in electrical engineering at Georgia Institute of Technology. His research interests include using floating–gate MOS transistors to build adaptive systems in silicon, investigating the solid-state physics of floating-gate devices, and modeling high-field carrier transport in Si and SiO_2.

Andreas Andreou received his M.S.E. and Ph.D. degrees in electrical engineering and computer science from Johns Hopkins University in 1983 and 1986, respectively. From 1987 and 1989 he was a postdoctoral fellow and associate research scientist at Johns Hopkins University, where he became assistant professor in 1989, associate professor in 1993, and full professor in 1997. His research interests are in the areas of device physics, integrated circuits/systems and neural computation. During 1995-1996 he was a visiting associate professor of computation and neural systems at Caltech.

Chris Diorio received a B.A. in physics from Occidental College in 1983, an M.S. in electrical engineering from the California Institute of Technology in 1984, and a Ph.D. in electrical engineering from the California Institute of Technology in 1997. He is currently an Assistant Professor in computer science at the University of Washington. His interests include analog integrated circuit design, ultra-high-speed digital circuit design, and semiconductor physics. His current research involves using floating-gate MOS transistors for silicon learning applications. He is currently employed as a Staff Engineer at TRW Inc. in Redondo Beach, CA, and has worked as a Senior Staff Scientist at American Systems Corporation in Chantilly, VA, and as a Technical Consultant at The Analytic Sciences Corporation in Reston, VA.

Bradley A. Minch received a B.S. in electrical engineering with distinction from Cornell University in 1991, and a Ph.D. in computation and neural systems from the California Institute of Technology in 1997. He is currently an Assistant Professor in electrical engineering at Cornell University. His research interests include current-mode circuits and signal processing, the use of floating-gate MOS transistors to build adaptive systems in silicon, and silicon models of dendritic computation.

Carver A. Mead, Gordon and Betty Moore Professor of Engineering and Applied Science, has taught at the California Institute of Technology for more than 30 years. He has contributed in the fields of solid-state electronics and the management of complexity in the design of very large scale integrated circuits, and has been active in the development of innovative design methodologies for VLSI. He has written, with Lynn Conway, the standard text for VLSI design, *Introduction to VLSI Systems*. His recent work is concerned with modeling neuronal structures, such as the retina and the cochlea using analog VLSI systems. His newest book on this topic, *Analog VLSI and Neural Systems*, was published in 1989 by Addison-Wesley. Professor Mead is a member of the National Academy of Sciences, the National Academy of Engineering, the American Academy of Arts and Sciences, a foreign member of the Royal Swedish Academy of Engineering Sciences, a Fellow of the American Physical Society, a Fellow of the IEEE, and a Life Fellow of the Franklin Institute. He is also the recipient of a number of awards, including the Centennial Medal of the IEEE.