Common Misconceptions about Saturation Voltage in Nanoscale MOS Transistors

Bertrand M. Grossman
Department of Physics and Technology
The City University of New York
Bronx Community College
Bronx, NY 10453
Bertrand.Grossman@bcc.cuny.edu

ABSTRACT

Many undergraduate programs in electronic engineering include a one semester course in semiconductor devices. In this course the student is introduced to lumped physical models that are derived from solid-state physics, statistical mechanics and electrostatics for the operating characteristics of transistors and other semiconductor devices. Among the many concepts taught is that the drain-to-source voltage that causes the current flow in an MOS transistor to saturate ($\partial I_{ds} / \partial V_{ds} \rightarrow 0$) is given by the simple mathematical expression $V_{dsat} = V_{gs} - V_T$. However, this simple expression is widely misapplied by both students and professionals to nanoscale MOS transistors used in modern high-performance microelectronic circuits. This work describes the three competing physical mechanisms that saturate drain-to-source current flow in an MOS transistor and gives a mathematical expression for V_{dsat} for each mechanism.

INTRODUCTION

The metal-oxide-semiconductor (MOS) transistor is a sandwich-like structure consisting of a semiconductor substrate isolated from a gate electrode by a thin insulating film. A cut-away view of an MOS transistor appears in Figure 1. Modern MOS transistors are typically fabricated from crystalline silicon (Si) substrates and silicon dioxide (SiO_2) insulating films although other semiconductor substrates, such as gallium arsenide [1] and indium phosphide [2], and other insulating films, such as silicon nitride [3] and aluminum oxide [4], may be used in specialized devices. The gate electrode is usually formed from heavily doped polycrystalline silicon (polysilicon), polysilicon in conjunction with a refractory metal (polycide), or a refractory metal alone. The insulating film is a thin dielectric that is either thermally grown or deposited onto the semiconductor substrate [5] and must be well matched to the crystalline structure of the underlying substrate. A mismatch between the insulator and substrate molecular structures introduces mechanical strain and other asperities [6] at the semiconductor-to-insulator interface that are detrimental to the performance of the transistor [7].

The basic physical features of an n-MOS transistor appear in Figure 1. The n-MOS transistor has two heavily doped n-type regions, the source and drain, at opposite ends of a p-type substrate. Current in an n-MOS transistor consists of a thin layer of electronic charge (≈ 10 Angstroms) that flows from source-to-drain on the semiconductor side of the

semiconductor-to-insulator interface. The thin layer of electronic charge is electrostatically induced by an electric field in the normal direction to the semiconductor surface (the "field effect") that is supported by the application of a gate-to-source voltage (V_{gs}). At the same time, charge carriers in the thin layer are driven toward the drain by an electric field in the drain-to-source direction that is supported by the application of a drain-to-source voltage (V_{ds}). Conversely, a p-MOS transistor has two heavily doped p-type regions at opposite ends of an n-type substrate and the conducting specie are positively charged holes.

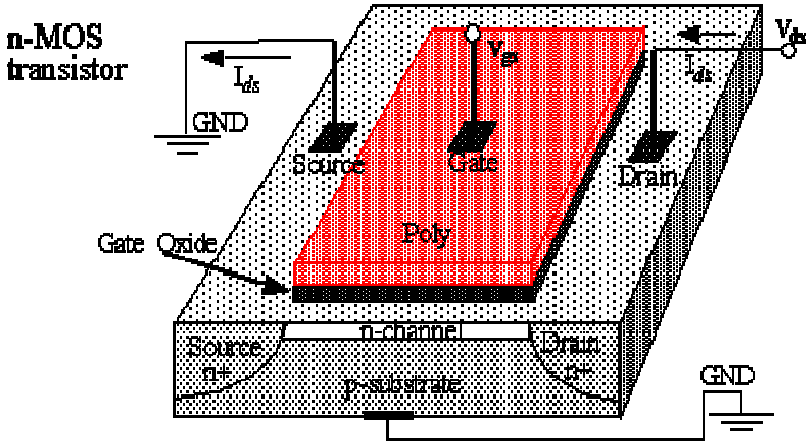

Figure 1**:** Cut-away view of MOS transistor

Drain-to-source current flow in an MOS transistor can be represented by an equivalent voltage dependent current source. The characteristic response of the current source is given by a monotonic, piecewise smooth and continuous function of the three independent voltages applied across the four terminals of the MOS transistor. The drain-to-source saturation voltage (V_{dsat}) is one of the critical points where the "pieces" of the function join.

For fixed values of V_{gs}, current flow in an MOS transistor saturates at the lowest non-negative value of drain-to-source voltage that either:

- pinches off charge carriers in the conducting channel at the drain junction;
- degrades the mobility of charge carriers in the conducting channel to the extent that drain-to-source current fails to increase with increments of V_{ds};
- or accelerates charge carriers in the conducting channel to their saturation velocities.

Two of these three competing physical mechanisms for saturating transistor current – mobility degradation and velocity saturation [7] – are rarely discussed in introductory courses on semiconductor devices or described in undergraduate textbooks. An incorrect value for V_{dsat} in an MOS transistor model could introduce a region of negative resistance

in the neighborhood of V_{dsat} that is nonphysical. In a computer simulation of a circuit, this region of negative resistance may cause the circuit simulator to converge to a false solution or, more likely, fail to converge as it endlessly hunts for a stable operating point. This work describes the three physical mechanisms that saturate drain-to-source current in an MOS transistor and gives a mathematical expression for V_{dsat} for each mechanism that is appropriate for classroom instruction or for manual calculation.

DISCUSSION

In the linear and triode regions of operation, where $V_{gs} \geq V_T$ and $0 \leq V_{ds} \leq V_{dsat}$, the drain-to-source current flow in an MOS transistor is given by

$$I_{ds} = C_{ox}\mu_{eff}\frac{W}{L}\left[(V_{gs} - V_T)V_{ds} - \frac{1}{2}V_{ds}^{\ 2}\right]. \tag{1}$$

This simple expression for I_{ds} is commonly known as the *Spice Level I* transistor model [9]. The notation for the variables and parameters in (1) appears in Table 1. Equation (1) makes use of the simplifying approximation that the areal density of fixed ionic charge in the depletion region beneath the conducting channel is invariant with position along the semiconductor surface [10]. (In this context, the areal density of fixed ionic charge is the net depleted charge per unit area of semiconductor surface.) More complicated *Level II* and *Level III* MOS transistor models include the effects of the variation of depletion charge density with position along the semiconductor surface [6]-[9].

Charge carriers (electrons and holes) are electronically scattered by the nuclei of the background atoms in the semiconductor as they traverse a path from the source to the drain along the semiconductor surface. These scattering events occur sufficiently often that the process for scattering charge carriers at the semiconductor surface can be lumped into an effective mobility parameter in (1) that is given by

$$\mu_{eff} = \left[\frac{1}{L}\int_0^L \frac{dy}{\mu_S}\right]^{-1}. \tag{2}$$

The functional form for μ_{eff} given by (2) is a type of weighted average of the mobility of charge carriers at the semiconductor surface (μ_S) that is assumed in the derivation of (1) [11] and is useful as a parameter in (1) only.

For classroom instruction or for the manual evaluation of I_{ds}, the effective mobility μ_{eff} in (1) is typically set equal to a constant which omits mobility degradation as a saturating mechanism. However, for practical applications, μ_{eff} is given as a function of terminal voltages V_{gs}, V_{ds} and V_{bs} because the value for μ_S is significantly degraded by the

electric field in the normal direction [12] that is supported by V_{gs} and by the electric field in the parallel direction that is supported by V_{ds} [13]-[15].

Table 1**:** Table of variables and parameters

Notation	
C_{ox}	areal capacitance across the gate insulator
$E_{\|}$	electric field in the parallel direction
I_{ds}	drain-to-source current
I_{dsat}	drain-to-source saturation current
K	slope of V_T versus $\sqrt{\varphi_{si} - V_{bs}}$ curve
L	channel length
Q_d	areal density of mobile charge at the drain
V_{bs}	substrate-to-source voltage
V_{ds}	drain-to-source voltage
V_{dsat}	drain-to-source saturation voltage
$V_{d\,max}$	value for V_{ds} that maximizes I_{ds}
V_{gs}	gate-to-source voltage
V_p	pinch-off voltage
V_T	threshold voltage
V_{vsat}	value for V_{ds} that accelerates carriers to v_{sat}
v_{sat}	saturation velocity of carriers at the surface
W	channel width
φ_{si}	surface potential in strong inversion at source
η	effective mobility parameter
μ_0	surface mobility for zero electric field
μ_{eff}	effective mobility parameter
μ_S	surface mobility
θ_0	effective mobility parameter

Many MOS transistor models used in circuit simulators include the degrading effects of the normal and parallel electric fields upon carrier mobility through the use of a widely accepted empirical function for μ_S [11],[16]. Substituting this empirical function for μ_S into (2), and invoking the *Level I* simplifying approximation that the density of fixed ionic charge in the depletion region is invariant with position along the semiconductor surface, gives the relation

$$\mu_{eff} = \frac{\mu_0}{1 + \theta_0 (V_{gs} - V_T + 2K\sqrt{\varphi_{si} - V_{bs}} - 0.5V_{ds}) + \eta V_{ds}} \tag{3}$$

where θ_0, η and K are parameters that can be extracted from measured transistor characteristics [16],[17]. (It is noteworthy that (3) omits the effects upon μ_{eff} introduced by the voltages dropped across the source and drain because of the small parasitic series resistances of these regions although more complicated closed-form expressions for μ_{eff} can be derived that include these effects [18].) In the model for effective mobility in (3), the value for μ_{eff} decreases as V_{bs} is made increasingly negative. Some models for μ_{eff} described in the engineering literature [19],[20] show the opposite tendency which gives the nonphysical result that μ_{eff} increases in value as the value for the normal electric field is made larger with negative increments of V_{bs}.

MECHANISMS FOR CURRENT SATURATION

As described in the introduction, drain-to-source current in an MOS transistor saturates because of one of three physical mechanisms: pinch-off, mobility degradation, or carrier velocity saturation. This section describes each of these physical mechanisms and gives simple mathematical expressions for V_{dsat} that are appropriate for classroom instruction or manual calculation.

The areal density of charge carriers in the conducting layer that is electrostatically induced by V_{gs} varies with position along the channel for values of V_{ds} that are greater than zero. A sufficiently high value for V_{ds} can cancel the inductive action of V_{gs} and cause the conducting layer to vanish near drain end of the conducting channel. If this occurs, the conducting channel is "pinched off" at the drain and the value for I_{ds} approaches a maximum value, or saturates, for values of V_{ds} that are greater than the pinch-off voltage [6]-[9]. For a *Level I* model, the areal density of mobile (electronic) charge per unit area of semiconductor surface at the drain end of the conducting channel is given by

$$Q_d = -C_{ox}(V_{gs} - V_T - V_{ds}) \tag{4}$$

for $V_{gs} \geq V_T$. The mathematical expression for Q_d in (4) is consistent with the level of approximation assumed in (1) and (3). The pinch-off voltage V_p is obtained by setting Q_d in (4) to zero and solving for V_{ds} which gives

$$V_p = V_{gs} - V_T \,. \tag{5}$$

The second current saturating mechanism to consider is mobility degradation which can saturate I_{ds} if μ_{eff} is functionally dependent upon V_{ds}. If the value for V_{ds} reduces parameter μ_{eff} in current equation (1) to the extent that I_{ds} fails to increase with increments of V_{ds}, then I_{ds} saturates because of mobility degradation. Current I_{ds} in (1) saturates because of mobility degradation at the value of V_{ds} that maximizes (1) which is

determined by substituting the right-hand-side of (3) for μ_{eff} into (1), differentiating with-respect-to V_{ds}, setting the resulting quadratic equation to zero for $V_{ds} = V_{d\,max}$, and solving for $V_{d\,max}$. Following this procedure gives

$$V_{d\,max} = \frac{B}{2A}\left[1 - \sqrt{1 - \frac{4AC}{B^2}}\right] \tag{6}$$

for parameters *A*, *B* and *C* given by

$$A = 0.5(0.5\theta_0 - \eta), \tag{7a}$$

$$B = 1 + \theta_0(V_{gs} - V_T + 2K\sqrt{\varphi_{si} - V_{bs}}) \quad \text{and} \tag{7b}$$

$$C = B(V_{gs} - V_T). \tag{7c}$$

Equation (6) gives non-negative Real values for $V_{d\,max}$ for the range of values of $4AC/B^2$ given by $-\infty < 4AC/B^2 \leq 1$. For $4AC/B^2 > 1$, the mechanism of mobility degradation cannot saturate I_{ds}. If the effect of mobility degradation upon I_{ds} is omitted by setting μ_{eff} equal to a constant ($\theta_0 = \eta = 0$), then

$$\lim_{\mu_{eff} \to \mu_0} V_{d\,max} = V_{gs} - V_T \tag{8}$$

and the value for $V_{d\,max}$ degenerates to pinch-off voltage V_p given by (5).

The final current saturating mechanism to consider is velocity saturation. A rigid object that is accelerated by a force field through a viscous medium attains a maximum or terminal velocity when the viscous drag on the object cancels the force that is accelerating the object forward [21]. An analogous situation occurs in the conducting channel of an MOS transistor. Charge carriers that are accelerated by a drain-to-source electric field ($E_{\|}$) along a semiconductor surface attain a maximum velocity when the various scattering events along the semiconductor surface cancel the force that is accelerating the carriers toward the drain. These scattering events limit the maximum velocity of charge carriers to the value given by $\mu_S E_{\|}$. However, the process for increasing the velocity of charge carriers in the channel by increasing the strength of $E_{\|}$ is not unbounded. Eventually, a point is reached where the additional energy supplied by increasing the strength of $E_{\|}$ is primarily emitted by the charge carriers as optical phonons [22] so that any differential change in the value for carrier velocity $\mu_S E_{\|}$ with $E_{\|}$ becomes vanishingly small. The saturation velocity v_{sat} is the maximum value that $\mu_S E_{\|}$ can attain. The value for v_{sat} for charge carriers in the conducting channel of an MOS transistor is critically dependent upon the way that the semiconductor surface was prepared during the

fabrication of the transistor. Modern fabrication methods can yield values for v_{sat} that are as high as 7.9 x10^6 cm/s for MOS transistors operating at room temperature [23] although the value for v_{sat} can be expected to vary over a wide range of values for different fabrication processes.

The drain-to-source current for charge carriers in the conducting layer that are accelerated to saturation velocity v_{sat} by drain-to-source voltage V_{vsat} is given by

$$I_{dsat} = W \mid Q_d \mid v_{sat} \tag{9}$$

which, upon substituting the *Level I* expression in (4) for Q_d , becomes

$$I_{dsat} = WC_{ox}(V_{gs} - V_T - V_{vsat})v_{sat} . \tag{10}$$

Setting the right-hand-side of (10) equal to the right-hand-side of *Level I* current equation (1) for $\mu_{eff} = \mu_0$ and $V_{ds} = V_{vsat}$, and solving the resulting quadratic equation for V_{vsat} gives [11]

$$V_{vsat} = V_p + \frac{v_{sat}L}{\mu_0} - \sqrt{V_p{}^2 + \left(\frac{v_{sat}L}{\mu_0}\right)^2} \tag{11}$$

where the quantity V_p is the pinch-off voltage given by (5).

It can be shown that (11) only returns positive values for V_{vsat} by representing the first two terms on the right-hand-side of (11) as the lengths of the sides of a right triangle and noting that the Pythagorean Theorem implies that the term in the square root represents the length of the triangle's hypotenuse. The geometry of the right triangle dictates that the sum of the lengths of the sides is always greater than the length of the hypotenuse which implies that the sum of the first two terms on the right-hand-side of (11) is always greater than the term in the square root which gives the result that $V_{vsat} > 0$. Geometric considerations also dictate that the difference between the sum of the lengths of the sides and the length of the hypotenuse of a right triangle is always less than the length of either side of the triangle. This last corollary implies that the maximum value for V_{vsat} is less than the value of either of the first two terms on the right-hand-side of (11) or, that V_{vsat} is bounded by $0 < V_{vsat} <$ Min [V_p , ($v_{sat}L / \mu_0$)] for the case that μ_{eff} is constant ($\mu_{eff} = \mu_0$).

CONCLUSION

Three physical mechanisms compete to saturate current flow in an MOS transistor. The physical mechanism that saturates current flow at the lowest drain-to-source voltage "wins" the competition and determines the value of V_{dsat} . This competitive process is mathematically expressed by the three-way minimum function given by

$$V_{dsat} = \text{Min}(V_p, V_{vsat}, V_{d\max}). \quad (12)$$

For classroom instruction or for the manual evaluation of I_{ds}, the complicating effect of mobility degradation upon transistor current flow is often omitted by setting μ_{eff} equal to a constant. The elimination of mobility degradation as a current saturating mechanism simplifies the derivation of a mathematical expression for V_{dsat} by reducing the three-way minimum function in (12) to the two-way minimum function given by $V_{dsat} = \text{Min}(V_p, V_{vsat})$. Combining this two-way minimum function with the condition deduced in the prior section that the upper bound on V_{vsat} is given by $\text{Min}[V_p, v_{sat}L/\mu_0)]$ further reduces the three-way minimum function in (12) to $V_{dsat} = V_{vsat}$ which, upon substituting the right-hand-side of (11) for V_{vsat}, gives the result

$$V_{dsat} = V_p + \frac{v_{sat}L}{\mu_0} - \sqrt{V_p{}^2 + \left(\frac{v_{sat}L}{\mu_0}\right)^2} \quad (13)$$

for the case that μ_{eff} is constant ($\mu_{eff} = \mu_0$).

It is instructive to investigate whether the velocity saturation model for V_{dsat} given by (13) is continuous with the pinch-off model for V_{dsat} given by (5) for limiting cases where only pinch-off can saturate I_{ds}. Physical sensibilities dictate that the mechanism of velocity saturation is eliminated for the limiting case where the velocity of carriers at the surface is unbounded ($v_{sat} \to \infty$) as well as for the limiting case where the channel length of the transistor is very long ($L \to \infty$) so that the strength of the electric field in the parallel direction (V_{ds}/L) is much smaller than the value for the parallel electric field (v_{sat}/μ_0) necessary for charge carriers to attain velocity v_{sat}. For the limiting case where $v_{sat} \to \infty$ and for the limiting case where $L \to \infty$, the values for the square root on the right-hand-side of (13) approach the value of $v_{sat}L/\mu_0$ which respectively give

$$\lim_{v_{sat}\to\infty} V_{dsat} = V_p \quad \text{and} \quad (14a)$$

$$\lim_{L\to\infty} V_{dsat} = V_p \quad (14b)$$

The limiting values for V_{dsat} given by (14) show that value for V_{dsat} given by (13) is continuous with the value for V_{dsat} given by the model for current saturation in (5) that is solely based upon pinch-off. In fact, for a long-channel MOS transistor for which $L >> (\mu_0 V_{dsat}/v_{sat})$, Equation (13) returns values for V_{dsat} that asymptotically approach the values for V_{dsat} given by (5) as an upper bound. However, modern MOS transistors used in high performance integrated circuits have channel lengths that are measured on the nanometer scale. For these short channel length devices, Equations (13) and (5) can

return markedly different values for V_{dsat} because the first two terms on the right-hand-side of (13) can have values that are comparable in magnitude. For a *Level I* MOS transistor model with μ_{eff} set to a constant, the correct value for V_{dsat} is given by (13) instead of the commonly used mathematical expression, $V_{dsat} = V_{gs} - V_T$, given by (5) that is valid only for the limiting case of saturation by pinch-off.

REFERENCES

[1] Ghandhi, Sorab, K. *VLSI Fabrication Principles: Silicon and Gallium Arsenide,* 2nd Edition, Wiley-Interscience, 1994, pp. 197-347.

[2] Wada, Osamu and Hideki Hasegawa. *InP-Based Materials and Devices: Physics and Technology,* Wiley-Interscience, 1999, pp. 71-339.

[3] Larsen, R., A. "A Silicon and Aluminum Dynamic Memory Technology." *IBM Journal of Research and Development,* Vol. 24, 1980, pp. 268-281.

[4] Guha, S., E. Cartier, N.A. Bojarczuk, J. Bruley, L. Gignac and J. Karasinski. "High-Quality Aluminum Oxide Gate Dielectric by Ultra-High-Vacuum Reactive Atomic-Beam Deposition." *Journal of Applied Physics,* Vol. 90, 2001, pp. 512-514.

[5] Plummer, James, D. *Silicon VLSI Technology: Fundamentals, Practice, and Modeling,* Prentice Hall, 2000.

[6] Richman, Paul. *MOS Field-Effect Transistors and Integrated Circuits,* John Wiley & Sons, 1973, pp. 158-160.

[7] Sze, Simon, M. *Physics of Semiconductor Devices*, 2nd Edition, Wiley-Interscience, 1981, pp. 431-486.

[8] Baker, Jacob, R., Harry W. Li and David E. Boyce. *CMOS Circuit Design, Layout, and Simulation*, 2nd Edition, IEEE Press, 2004, pp. 201-417.

[9] Warner Jr., R., M. and B.L. Grung. *Transistors: Fundamentals for the Integrated-Circuit Engineer*, Wiley-Interscience, 1990, pp. 763-782.

[10] Sah, C., T. "Characteristics of the Metal-Oxide-Semiconductor Transistor." *IEEE Transactions on Electron Devices,* Vol. ED-11, 1964, pp. 324-345.

[11] Merckel, G., J. Borel and N.Z. Cupcea. "An Accurate Large-Signal MOS Transistor Model for use in Computer-Aided Design." *IEEE Transactions on Electron Devices,* Vol. ED-19, 1972, pp. 681-690.

[12] Schrieffer, J., R. "Effective Carrier Mobility in surface space-charge layers." *Physics Review,* Vol. 97, 1955, pp. 641-649.

[13] Sun, S., C. and J. D. Plummer. "Electron Mobility in Inversion and Accumulation Layers on Thermally Oxidized Silicon Surfaces." *IEEE Transactions on Electron Devices*, Vol. ED-27, 1980, pp. 1497-1508.

[14] Shichijo, H., "A Re-Examination of Practical Performance Limits of Scaled n-Channel and p-Channel MOS Devices for VLSI." *Solid-State Electronics*, Vol. 26, 1983, pp. 969-986.

[15] Sodini, C., G., P.-K. Ko and J.L. Moll. "The Effect of High Fields on MOS Device and Circuit Performance." *IEEE Transactions on Electron Devices*, Vol. ED-31, 1984, pp. 1386-1393.

[16] Sabnis, A., G. and J.T. Clemens. "Characterization of the Electron Mobility in the Inverted <100> Si Surface." 1979 International Electron Devices Meeting (Washington, D.C.). *1979 IEDM Technical Digest*, 1979, pp. 18-21.

[17] Sabnis, A., G. "Impact of Advances in Technology on the Properties of SiO_2/Si Interface." *Proceedings of the 1984 IEEE Reliability Physics Symposium*, 1984, pp. 156-160.

[18] Liu, William. *MOSFET Models for SPICE Simulation, Including BSIM3v3 and BSIM4*, Wiley-IEEE Press, 2001.

[19] Crawford, Robert, H. *MOSFET in Circuit Design: Metal-Oxide-Semiconductor Field-Effect Transistors for Discrete and Integrated Circuit Technology*, McGraw-Hill, 1967.

[20] Sheu, B., J., D.L. Scharfetter, P.-K. Ko and M.-C. Jeng, "BSIM: Berkley Short-Channel IGFET Model for MOS Transistors." *IEEE Journal of Solid-State Circuits*, Vol. SC-6, 1987, pp. 558-566.

[21] John, James, E., A. and William Haberman. *Introduction to Fluid Mechanics*, 3rd Edition, Prentice Hall, 1988, p. 516.

[22] Luryi, S. "Device Building Blocks." *High Speed Semiconductor Devices* (Editor: S. Sze), Wiley-Interscience, 1990, pp. 57-62.

[23] Chan, T.,-Y., S.-W. Lee and H. Gaw. "Experimental characterization and modeling of electron saturation velocity in MOSFETs inversion layer from 90 to 350 K." *IEEE Transactions on Electron Devices*, Vol. ED-11, 1990, pp. 466-468.

BIOGRAPHY

BERTRAND M. GROSSMAN is Assistant Professor of Physics and Technology at Bronx Community College of The City University of New York. He joined the faculty of Bronx Community College in 2003 after a long career as a research scientist and

electrical engineer at the IBM Thomas J. Watson Research Center in Yorktown Heights, NY. His professional career is in the design of VLSI circuits for high-performance microprocessors, semiconductor device design, and the mathematical modeling of semiconductors devices and transport processes. Some of the high performance circuits that he designed can be found in the IBM RS6000 line of computer servers, as well as in Apple computers and the Sony Playstation III game processor. Dr. Grossman earned Ph.D. and M.Phil. Degrees in electrical engineering, and M.S. and B.S. Degrees in applied physics, from Columbia University.