Engineering the Electron-Hole Bilayer Tunneling Field-Effect Transistor

Sapan Agarwal
James Teherani
Judy Hoyt
Dimitri Antoniadis
Eli Yablonovitch

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2013-249
http://www.eecs.berkeley.edu/Pubs/TechRpts/2013/EECS-2013-249.html

December 31, 2013

Acknowledgement

The work was supported by the Center for Energy Efficient Electronics Science (NSF Award 0939514)

Engineering the Electron-Hole Bilayer Tunneling Field-Effect Transistor

Sapan Agarwal, James T. Teherani, Judy L. Hoyt, Dimitri A. Antoniadis, and Eli Yablonovitch

***Abstract*—The electron-hole (EH) Bilayer Tunneling Field-Effect Transistor promises to eliminate heavy-doping band-tails enabling a smaller subthreshold swing voltage. Nevertheless, the electrostatics of a thin structure must be optimized for gate efficiency. We analyze the tradeoff between gate efficiency versus on-state conductance to find the optimal device design. Once the EH Bilayer is optimized for a given on-state conductance, Si, Ge, and InAs all have similar gate efficiency, around 40-50%. Unlike Si & Ge, only the InAs case allows a manageable work function difference for EH Bilayer Transistor operation.**

***Index Terms*—Electron hole Bilayer, quantization, semiconductor device modeling, tunneling, tunneling field effect transistor (TFET)**

I. INTRODUCTION

IN order to reduce the power consumption of modern electronics, the operating voltage needs to be significantly reduced. The electron-hole Bilayer Tunneling Field-Effect Transistor (EH Bilayer TFET) is a new device concept that has the potential for reduced voltage operation [1-4]. In general, TFETs may achieve a low operating voltage by overcoming the thermally limited subthreshold swing voltage of 60mV/decade, but results to date have been unsatisfying [5, 6]. The best subthreshold swings have been measured at a current density of around 1nA/μm, and performance degrades significantly at larger currents.

TFETs promise a small subthreshold swing voltage by abruptly turning on when the conduction band on the n-side aligns with the valence band on the p-side of a tunneling junction[7]. In actuality, the band edges are not perfectly sharp and there are states that extend into the band gap[8]. This is seen in the Urbach tail of optical absorption measurements [9, 10]. Below the band-edge energy, the absorption coefficient falls off exponentially due to a residual band-tail density of states. The same band-tail density of states will unfortunately smear the abrupt response and increase the subthreshold swing voltage of TFETs. In intrinsic GaAs the optical absorption falls off at a semilog slope of S_{Urbach}≡17meV/decade due to phonons [10]. If the GaAs is heavily doped to 10^{20}/cm^3, an impurity band forms and the absorption falls off more gradually, ~58meV/decade [11]. By eliminating doping in the tunneling junction, the EH Bilayer TFET avoids problems associated with these doping band-tails.

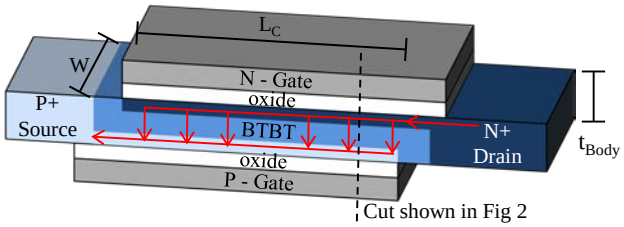

Fig. 1: (a) The electron-hole Bilayer TFET structure with the current path (red) and inversion layers (blue).

A EH Bilayer TFET consists of a p^+ source, an n^+ drain, and an undoped channel bound by offset top and bottom gates as shown in Fig. 1. The gates are oppositely biased to create an electron (hole) gas along the top (bottom) gate extending to the n^+ drain (p^+ source). The device turns on when sufficient potential is applied between the gates to align the energy levels, enabling vertical band-to-band tunneling (BTBT) across the channel. The band diagram along the tunneling path is shown in Fig. 2(a). The voltage difference between the gates can be accommodated by the work-function difference between the *n* and *p*-type gates. In addition to eliminating doping, the EH Bilayer TFET also has a higher on-state conductance as it provides a large overlap area to compensate for limited tunneling transmission. The double quantum confinement also assists the on-state conductance [12-14].

In this paper we focus on minimizing the subthreshold swing voltage, while maintaining a high on-state conductance[1]. We do this by optimizing the dc gate biases, the body thickness and the channel material. (Si, Ge, InAs and an InAs/AlGaSb heterostructure were considered). First we analyze the different factors that influence the subthreshold swing in Section II. We find that maximizing the gate efficiency (the ability of the gate to change the energy levels) has the largest impact on minimizing the subthreshold swing. Consequently, we analyze the tradeoff between gate efficiency versus on-state conductance, to find the optimal device design.

The work was supported by the Center for Energy Efficient Electronics Science (NSF Award 0939514).

S. Agarwal and E. Yablonovitch are with the University of California, Berkeley CA 94720 USA (email: sapan@berkeley.edu)

J. T. Teherani, J. L. Hoyt and D. A. Antoniadis are with the Microsystems Technology Laboratories, Massachusetts Institute of Technology, Cambridge, MA 02139 USA

[1] We consider conductance rather than current, as the speed of a low voltage device is limited by its RC time and not by its current density.

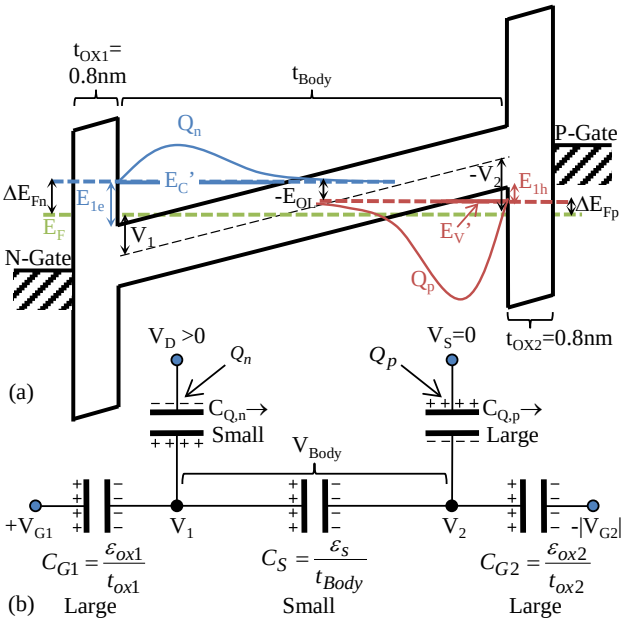

Fig. 2: (a) Band diagram with quantum well ground states shown. (b) Capacitive voltage divider model of the EH Bilayer device.

In Section III, we describe an analytical model for the EH Bilayer TFET. Finally, we discuss the results of the optimization in Section IV and compare the analytic model to a numerical simulation in *nextnano++*.

II. EH Bilayer Subthreshold Swing Voltage

An ideal TFET would rely upon a sharp band edge and would switch abruptly from zero-conductance to the desired on-conductance when the electron and hole eigenstate energies overlap. Unfortunately the band-edges are not perfectly sharp and thus there is a finite density of states (DOS) extending into the band gap, smearing out the desired abrupt response. Conventional TFET modeling does not account for the smeared band edge DOS. Consequently, we want to find the subthreshold swing voltage (SS) while accounting for the band edges. The subthreshold swing voltage is defined by:

$$SS \equiv \left(d\log(I)/dV_G\right)^{-1} \tag{1}$$

In order to evaluate SS, we need to include the band tails in the current model[15] [2]:

$$I \propto \int (f_C - f_V) \times \top \times D_C(E) \times D_V(E) \times \partial E \tag{2}$$

The difference in the Fermi occupation probabilities between the conduction and valence bands is $(f_C\text{-}f_V)$ and the transmission probability of a tunneling electron is $\top$. $D_C(E)$ and $D_V(E)$ are the conduction and valence band DOS. The product $D_C(E)\times D_V(E)$ is effectively a joint density of states.

[2] The band tail states will not have a well-defined E-k relationship and are likely to be localized. Consequently, conservation of transverse momentum will not hold when tunneling to band tail states. In this case, the current will be proportional to both the initial and final density of states. When the bands are overlapping conservation of momentum should be accounted for, resulting in a single density of states as is done in [12-13]

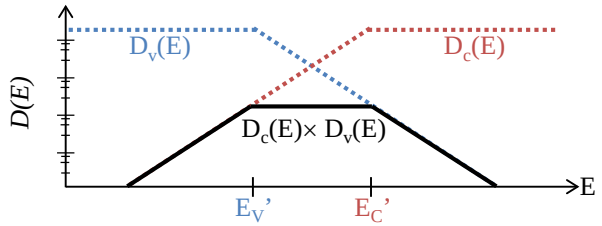

Fig. 3: The conduction and valence band density of states, $D_C(E)$ and $D_V(E)$, are shown. Below the band edges the density of states falls off exponentially. The product, $D_C(E)\times D_V(E)$ is also shown.

This model is valid when tunneling to band tail states where the electron and hole eigenstates, E'_C and E'_V respectively, are not aligned as shown in Fig. 2(a). $D_C(E)$ is given by:

$$D_C(E) = \begin{cases} D_C', & E \geq E_C' \\ D_{C0} \times e^{-(E_C' - E)/qV_0}, & E < E_C' \end{cases} \tag{3}$$

Above the band edge the DOS is given the by ideal DOS, D'_C, and is a constant with respect to energy in 2d. Below the band edge, we assume that the DOS falls off exponentially with a semilog slope of V_0 and constant pre-factor D_{C0}. An exponential falloff is typical of band edges as seen in the optical absorption edge [11]. Similarly, the valence band DOS will be given by:

$$D_V(E) = \begin{cases} D_V', & E \leq E_V' \\ D_{V0} \times e^{-(E - E_V')/qV_0}, & E > E_V' \end{cases} \tag{4}$$

Here D'_V is the ideal 2d hole DOS and D_{V0} is a constant pre-factor for the band tail DOS. For simplicity, we take the exponential slope, V_0, to be the same for conduction and valence band edges.

The combined DOS is given by $D_C(E)\times D_V(E)$. Ideally, no current would flow, but due to the band tails, an overlapping density of states exists as shown in Fig. 3. This gives:

$$D_C(E) \times D_V(E) = e^{\frac{E_{OL}}{qV0}} \times \begin{cases} D_C' \times D_{V0} \times e^{-(E - E_C')/qV0}, & E \geq E_C' \\ D_{C0} \times D_{V0}, & E_V' < E < E_C' \\ D_{C0} \times D_V' \times e^{-(E_V' - E)/qV0}, & E \leq E_V' \end{cases} \tag{5}$$

This is for the case where $E'_C > E'_V$. E_{OL} is the overlap energy between the electron and hole eigenstates shown in Fig. 2(a) such that: $E_{OL} = (E'_V - E'_C) < 0$. Since the combined DOS has a maximum plateau in the bandgap region between E'_C and E'_V, we can approximate the current integral as:

$$I \propto \left(\int_{E_V'}^{E_C'} (f_C - f_V) \times \top \times \partial E \right) \times e^{E_{OL}/qV_0} \tag{6a}$$

$$I \propto I_0 \times e^{E_{OL}/qV_0} \tag{6b}$$

where the tunneling pre-factor is:

$$I_0 \equiv \int_{E_V'}^{E_C'} (f_C - f_V) \times \top \times \partial E \tag{7}$$

Thus we have arrived at a simplified model for the tunneling current when band tails are present. Now we can compute the subthreshold swing voltage by plugging (6) into (1):

$$SS = \left(\frac{dV_{Body}}{dV_G} \times \frac{d\log(I_0)}{dV_{Body}} + \frac{dV_{Body}}{dV_G} \frac{dE_{OL}}{dV_{Body}} \times \frac{d\log\left(e^{E_{OL}/qV_0}\right)}{dE_{OL}} \right)^{-1} \quad (8)$$

In the first term we took the derivative with respect to the voltage across the semiconductor bilayer, V_{Body}, since tunneling transmission probability, T, depends sensitively on V_{Body}. In the second term we took the derivative with respect to E_{OL} as the band edge density of states depends on the band alignment. Finally, the subthreshold swing voltage in (8) can be expressed in the following form by replacing each term with the appropriate symbol to highlight the four contributing factors:

$$SS = \left(\eta_{el} \times \frac{1}{S_{tunnel}} + \eta_{el} \times \eta_{quant} \times \frac{1}{S_{DOS}} \right)^{-1} \quad (9a)$$

$$SS = \frac{1}{\eta_{el}} \times \left(\frac{1}{S_{tunnel}} + \frac{\eta_{quant}}{S_{DOS}} \right)^{-1} \quad (9b)$$

S_{DOS} is the semilog slope of the joint band edge density of states in mV/decade:

$$S_{DOS} = \frac{1}{q} \frac{dE_{OL}}{d\log\{e^{E_{OL}/qV_0}\}} = V_0 / \log(e) \quad (10)$$

S_{tunnel} is the semilog slope measuring how steeply the tunneling conductance pre-factor changes with respect to the voltage across the body, V_{Body}:

$$S_{tunnel} = dV_{Body} / d\log(I_0) \quad (11)$$

It is given in mV/decade and I_0 is given by Eq. (6). S_{tunnel} is the steepness that results from changing the thickness of the tunneling barrier with a changing bias as it is typically dominated by the voltage dependence of T.

η_{quant} is the change in the band edge quantum level alignment with respect to the body voltage, V_{Body} due to level shifting. It is given by $dE_{OL}/d(qV_{Body})$. It can be significantly less than 1 because the shape of the triangular tunneling barrier is changing as V_{Body} changes, which causes the confinement energy to change [2].

The pre-factor efficiency, η_{el} is the electrostatic efficiency and can be found from the circuit model in Fig. 2(b). It is given by dV_{Body}/dV_G.

A small subthreshold swing voltage can be achieved by having either a small S_{tunnel} or a small S_{DOS}. Nonetheless, as discussed in [7], a small subthreshold swing voltage, cannot be achieved at high current densities by barrier thickness modulation, S_{tunnel}. High current densities require a high electric field, and any additional voltage will only result in a small change in the electric field and thus produce only a small change in the tunneling current. We verify this by computing S_{tunnel} in Section IV. Accordingly, at high current density, we find that S_{tunnel} is unfortunately >60mV/decade. Consequently, we must rely upon a sharp band edge density of states, S_{DOS}, to achieve a small subthreshold swing voltage.

By design, the EH Bilayer structure eliminates doping to improve the electronic S_{DOS}. To further improve the subthreshold swing, we focus on improving the gate efficiency. The overall gate efficiency, η_{gate}, is the change in the band alignment, E_{OL}, with respect to the gate bias, V_G and is:

$$\eta_{gate} \equiv \frac{1}{q} \frac{dE_{OL}}{dV_G} = \frac{1}{q} \frac{dE_{OL}}{dV_{Body}} \times \frac{dV_{Body}}{dV_G} = \eta_{quant} \times \eta_{el} \quad (12)$$

We need to optimize the gate efficiency that is reduced due to both electrostatics, η_{el}, and to quantum level shifts, η_{quant}.

III. EH Bilayer Modeling

To model the EH Bilayer TFET we consider the situation where the bias on the *n*-gate (V_{G1}) is changed while the bias on the *p*-gate (V_{G2}) is held constant. We calculate the carrier density by assuming a single Fermi level, E_F, as shown in Fig. 2(a). This is valid when there is a small source drain bias, corresponding to low voltage operation.

To be competitive with current CMOS transistors we assume an effective gate oxide thickness (EOT) of 0.8nm, and we consider a gate overlap region, L_C, shown in Fig. 1, of 10nm. For a given body thickness, channel material and Fermi level position, we first find the gate efficiency: $\eta_{gate} \equiv \Delta E_{OL}/(q\Delta V_{G1})$. Consequently, we need E_{OL} and V_{G1} in the on and off states. The device will turn on once the bands overlap and $E_{OL}=0$. In section III.A we find the gate biases, V_{G1} and V_{G2} required to achieve $E_{OL}=0$ and a given Fermi level position. In section III.B we determine E_{OL} and V_{G1} in the off state to find η_{gate}. After finding the gate efficiency, we find the tunneling conductance and the channel conductance in Section III.C and III.D respectively. In Section III.E we consider how the analysis would change for a heterojunction.

A. On-State Circuit Analysis

The first step of the analysis is to find the electron and hole quantum confinement energies, E_{1e} and E_{1h}, and the voltage across the body, V_{Body}, in the on-state. The overlap energy, E_{OL}, is given by [2]:

$$E_{OL} = qV_{Body} - (E_G + E_{1e} + E_{1h}) \quad (13)$$

This can be seen from Fig. 2(a). At zero overlap the voltage across the body is equal to the band gap, E_G, plus the confinement energies. The confinement energies are:

$$E_{1\alpha} \approx \left(\frac{9\pi}{8} \right)^{2/3} \times \left(\frac{(qV_{Body}/t_{Body})^2 \hbar^2}{2m^*_{\alpha,z}} \right)^{1/3} \quad (14)$$

where α represents either electrons (e) or holes (h) and t_{Body} is the thickness of the bilayer semiconductor body. The effective masses are tabulated in Table 1. We assumed an infinite triangular well model for the confinement energies.

In the on-state, the eigenstates are aligned such that $E_{OL}=0$ and so we can solve (13) for V_{Body} and then find E_{1e} and E_{1h}.

For a given Fermi level position, we can find the *n*-channel and *p*-channel potential, V_1 and V_2 as shown in Fig. 2. The potential is measured from the center of the band gap. From Fig. 2, we find:

TABLE I
MATERIAL PROPERTIES USED

	Si	Ge	InAs	GaSb	AlSb
E_g	1.12	0.66	0.354	-	-
$E_{g.eff}$	-	-	-	0.15[23]	0.26[23]
ε_S	11.7	16.2	15.15	15.7	12
γ_1	4.27[24]	13.4[25]	20.0[26]	13.4[26]	5.18[26]
γ_2	0.32[24]	4.25[25]	8.5[26]	4.7[26]	1.19[26]
γ_3	1.46[24]	5.69[25]	9.2[26]	6.0[26]	1.97[26]
$m^*_{e,t}$	0.38[27]	0.48[27]	.023[27]	-	-
$m^*_{e,z}$	0.92[27]	0.12[27]	.023[27]	-	-
m^*_{tunnel}	0.46[5]	0.058[5]	0.043	-	-

Eg,eff is the effective heterojunction band gap across the tunnel interface. The effective masses are calculated assuming a [100] wafer orientation. The transverse masses are density of states masses while the z direction mass is for confinement energy. The hole masses are computed from $m^*_{h,z}=1/(\gamma_1-\gamma_2)$ and $m^*_{h,z}=1/(\gamma_1+\gamma_2)$. For Si, $m^*_{e,z}=m_l$ and $m^*_{e,t}=2m_t$. For Ge $m^*_{e,z}=3m_l m_t/(m_t+2m_l)$ and $m^*_{e,t}=4m^*_{e,z}$. The tunneling mass is given by (26)&(27). All values for $Al_XGa_{1-X}Sb$ are linearly interpolated.

$$qV_1 = E_g/2 + E_{1e} - \Delta E_{Fn} \tag{15}$$

and

$$qV_2 = -E_g/2 - E_{1h} - \Delta E_{Fp} \tag{16}$$

At eigenstate alignment (the on-state of the device), the energy difference between the electron eigenstate and the Fermi level, ΔE_{Fn}, is equal to the energy difference between the hole eigenstate and the Fermi level, ΔE_{Fp}: $\Delta E_{Fn}=\Delta E_{Fp}\equiv\Delta E_F$.

Given a Fermi level position, ΔE_F, we can find the electron charge, Q_n, and hole charge, Q_p, in the channel:

$$|Q_n| = q\times N_{C,2D}\times\ln\left(1+\exp(-\Delta E_{Fn}/k_BT)\right) \tag{17}$$

$$|Q_p| = q\times N_{V,2D}\times\ln\left(1+\exp(\Delta E_{Fp}/k_BT)\right) \tag{18}$$

where

$$N_{C,2D}=\frac{m^*_{e,t}}{\pi\hbar^2}k_BT \quad\text{and}\quad N_{V,2D}=\frac{m^*_{h,t}}{\pi\hbar^2}k_BT \tag{19}$$

The effective masses are given in Table 1. Next we can use the capacitive voltage-divider model in Fig. 2(b) to solve for the corresponding gate voltages:

$$V_{G1} = V_1 + \left(|Q_n| + V_{Body}C_S\right)/C_{G1} \tag{20}$$

$$V_{G2} = V_2 - \left(|Q_p| + V_{Body}C_S\right)/C_{G2} \tag{21}$$

where C_S is the EH Bilayer body capacitance. C_{G1} and C_{G2} are the *n*-gate and *p*-gate oxide capacitances. We use the surface accumulation charge to capture the effect of the quantum capacitance. We also assumed that all the accumulation charge is located at the oxide interface. V_{G1} and V_{G2} provide the dc bias, or work function difference, needed to align the eigenstates and achieve a desired Fermi level position.

B. Off-State Circuit Analysis

In order to find the gate efficiency, we start with the off-state and then compute $\eta_{gate}\equiv\Delta E_{OL}/(q\Delta V_{G1})$. We define the gate efficiency this way because the quantum capacitance is non-linear and this definition contains the average gate efficiency. First we need to determine how much the overlap energy, ΔE_{OL}, needs to change in order to turn the tunneling off. A rough estimate is sufficient, since varying ΔE_{OL} from 50 to 200 meV only changes η_{gate} by a few percent. Consequently, we take ΔE_{OL}= -100 meV. If the band edge density of states S_{DOS} is 20 mV/decade (corresponding to the optically measured steepness), ΔE_{OL}= -100 meV will give 5 decades of on/off ratio.

To find V_{G1} in the off state, $V_{G1,OFF}$, we need to start at the opposite gate V_{G2} and work our way backwards through the capacitive voltage-divider model in Fig. 2(b). Since we are keeping the back-gate voltage, V_{G2}, fixed, we use the same value of V_{G2} that was found in the on-state from the capacitive voltage-divider model (21). Next, we need to find the body voltage (V_{Body}) across the semiconductor and the confinement energies (E_{1e} and E_{1h}) in the off state. We do this by solving the overlap energy definition (13) for E_{OL}=-100 meV. Now we can find V_2 by solving the capacitive voltage divider model (21) self-consistently for V_2. In (21), the charge density, $|Q_p|$, is a function of V_2 through ΔE_{Fp} from (16). Once we have V_2, we know $V_1=V_2+V_{Body}$. Then we can solve the capacitive voltage-divider model (20) for V_{G1} using (15) to define ΔE_{Fn} for the charge density. This gives us V_{G1} in the off-state, $V_{G1,OFF}$. As we already found V_{G1} in the on-state, $V_{G1,ON}$, at end of Section III.A, we can finally compute the gate efficiency as $\Delta E_{OL}/q(V_{G1,ON}-V_{G1,OFF})$.

C. Tunneling Conductance

As we are tunneling between two quantum wells we need to use the 2d-2d tunneling current formula [12, 13]:

$$G_{tunnel} = \frac{qm^*_{JDOS}L_CW}{\pi^2\hbar^3}\times\left(0.435E_{1e}\right)\times\left(0.435E_{1h}\right) \times\frac{q}{4k_bT}\times\mathcal{T}(F)\times\frac{1}{\cosh(\Delta E'_F/2k_BT)^2} \tag{22}$$

Where the tunnel transmission probability is:

$$\mathcal{T}(F) = \exp\left(\frac{-\pi(m^*_{tunnel})^{1/2}E_G^{3/2}}{2\sqrt{2}\hbar qF}\right) \tag{23}$$

and the electric field across the semiconductor layer is:

$$F = V_{Body}/T_{Body} \tag{24}$$

The length of the overlap region is L_C and the width is W as illustrated in Fig 1. The confinement energies, E_{1e} and E_{1h}, are given by (14). The Fermi level position relative to the closest eigenstate is given by $\Delta E'_F$. If E_F is below E'_C and E'_V then $\Delta E'_F=\Delta E_{Fp}$, given by (15). If E_F is above E'_C and E'_V then $\Delta E'_F=\Delta E_{Fn}$, given by (16). If E_F is in between E'_C and E'_V then $\Delta E'_F$ should be set to zero.

The tunneling probability is based on a two band WKB tunneling model and is given in [16]. Some care is needed in choosing the appropriate masses. The joint density of states mass is given by:

$$m^*_{JDOS} = 2\left(1/m^*_{e,t}+1/m^*_{h,t}\right)^{-1} \tag{25}$$

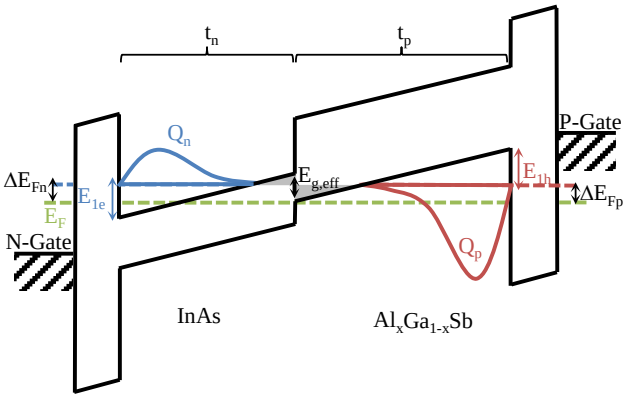

Fig. 4: Band diagram incorporating a heterojunction in the EH Bilayer.

The tunneling mass can be computed from [16]:

$$m^*_{tunnel} = 2\left(1/m^*_{e,z} + 1/m^*_{h,z}\right)^{-1} \quad (26)$$

The transverse masses, $m^*_{e,t}$, and $m^*_{h,t}$, as well as the masses in the tunneling direction, $m^*_{e,z}$, and $m^*_{h,z}$, are given in Table I.

The WKB model and reduced mass work well in InAs where the carriers in a single conduction band, tunnel to a single valence band [17]. However, in silicon and germanium the band gap is indirect and there are many interacting bands and so the WKB model breaks down [17]. Consequently, we use an experimentally fitted tunneling effective mass derived in [5]. While [5] used a single band tunneling model, we used a two band tunneling model and need to adjust the mass:

$$m^*_{2Band} = \left((2\sqrt{2}/\pi)\times(4\sqrt{2}/3)\right)^2 m^*_{1Band} \quad (27)$$

This comes from comparing the tunneling equation in [5] with (23). A summary of all the material parameters used is given in Table 1.

We use the tunneling formula in (22) because it more accurately captures the benefits of quantum confinement in increasing the current, as discussed in the appendix. For simplicity, we assume that tunneling only occurs in the vertical direction, perpendicular to the gates, and neglect the two dimensional electrostatics and any lateral tunneling.

D. Channel Conductance

The last step is to calculate the channel conductance. The channel needs to have a minimum charge available to carry the current that has tunneled, or else the on-state conductance will be limited by the channel resistance instead of tunneling resistance. The device conductance will be given by the lower of the channel or tunneling conductance. At a high conductance near 1 mS/um, the channel conductance becomes the limiting conductance. Consequently, a smaller ΔE_F is required to increase the number of electrons.

The channel conductance is given by a ballistic model [18]:

$$G = Wqn_s v_T / (2k_B T/q) \quad (28)$$

where

$$v_T = \sqrt{2k_B T/\pi m^*_{e,t}} \text{ and } n_s = |Q_n|/q \quad (29)$$

E. Heterojunction Analysis

Using a heterojunction as shown in Fig. 4 can reduce the required dc bias or work function difference to achieve the desired band alignment. A heterojunction will also slightly improve the gate efficiency. We consider an InAs/AlGaSb heterojunction since the band alignment at the hetero-interface, or effective band gap, $E_{g,eff}$, can be widely tuned by changing the Al content.

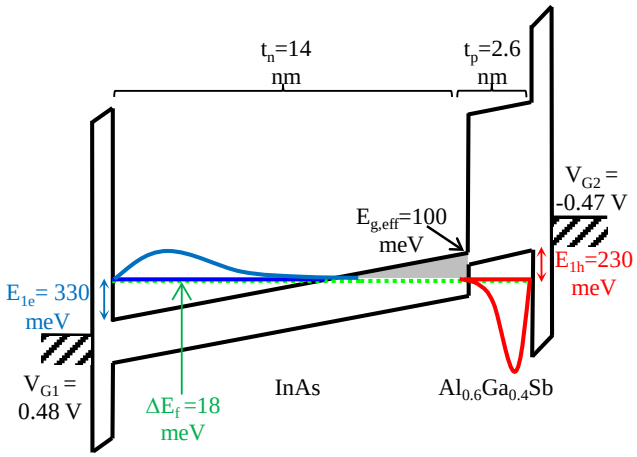

Fig. 5: The band diagram for the optimal heterojunction structure with an on state conductance of 1 mS/μm and maximum bias or Work Function difference<1eV is shown. The narrow p-well eigenstate energy changes little as the bias changes, helping the gate efficiency.

In order to account for the heterojunction, a few changes must be made. First we need to change the body capacitance:

$$C_S = \left(t_n/\varepsilon_{s,n} + t_p/\varepsilon_{s,p}\right)^{-1} \quad (30)$$

Here we have used n and p subscripts to refer to the device properties on the n and p sides, respectively. The electric field in each material is also different:

$$F_n = V_{Body}/\left(t_n + \varepsilon_{s,n}\times t_p/\varepsilon_{s,p}\right) \quad (31)$$

$$F_p = V_{Body}/\left(t_p + \varepsilon_{s,p}\times t_n/\varepsilon_{s,n}\right) \quad (32)$$

Next, we need to update the tunneling probability to account for the fact that we are tunneling through two triangular barriers. The triangular barrier heights on the n and p sides ($E_{B,n}$ and $E_{B,p}$) are given by:

$$E_{B,n} = qF_n \times t_n - E_{1e} \quad (33)$$

$$E_{B,p} = qF_p \times t_p - E_{1h} \quad (34)$$

The tunneling barriers are shaded in gray in Fig. 4. Looking at the p-side, the tunneling begins when the hole eigenstate energy enters the forbidden region. The height of the triangular tunneling barrier is given by (34). If a confinement energy, E_{1e} or E_{1h}, is large, the barrier height would be negative and so the tunneling begins in the other material. This situation can be seen for the hole energy in Fig. 5. Now we can model the tunneling probability with two single band tunneling approximations such that $T = T_n \times T_p$ and:

$$T_n = \exp\left\{-4(2\times m^*_{e,z})^{1/2} E^{3/2}_{B,n} / (3\hbar q F_n)\right\} \quad (35)$$

$$T_p = \exp\left\{-4(2\times m^*_{h,z})^{1/2} E^{3/2}_{B,p} / (\hbar q F_p)\right\} \quad (36)$$

Since there is an abrupt transition from the tunneling energy being close to the valence band, and then close to the conduction band, at the heterojunction, the tunneling process

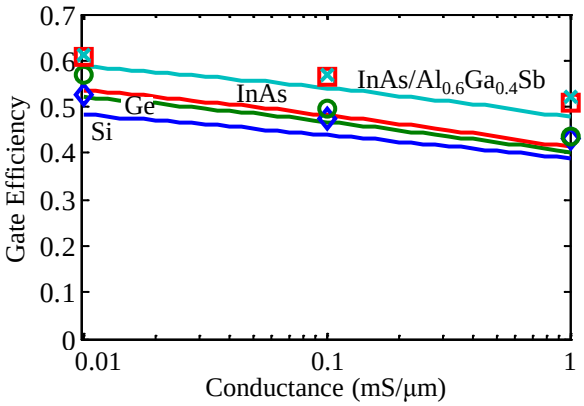

Fig. 6: The analytically computed gate efficiency for all three optimized homojunctions is similar. Using the heterojunction only slightly improves the gate efficiency. The device properties that give the optimal gate efficiency are summarized in Table II. A gate oxide thickness of 0.8 nm and a 10 nm channel length overlap, L_C, was assumed. The gate efficiency was averaged over a band misalignment of ΔE_{OL} = 100 meV. Numerically computing the gate efficiency in *nextnano++* gives similar results and is plotted using the following markers: Si – diamonds, Ge – circles, InAs – squares, $InAs/Al_{0.6}Ga_{0.4}Sb$ – crosses.

is divided into two discrete steps. As the tunneling primarily occurs within a single band on each side of the junction, a single band model is used.

The last change is that carriers are in trapezoidal quantum wells instead of triangular quantum wells. While this should be solved numerically and the finite barrier heights should be accounted for, we can get a qualitative understanding of what happens using the following approximation [19][3]:

$$E_1 \approx \sqrt{E_{tri}^2 + E_{square}^2} \tag{37}$$

E_{tri} is the energy in a triangular well given by (14). E_{square} is the standard quantum confinement energy in a square potential well given by $\hbar^2\pi^2/\left(2m_e^* t_n^2\right)$ or $\hbar^2\pi^2/\left(2m_h^* t_p^2\right)$. The trapezoidal quantum well shape improves the quantum confinement efficiency, η_{quant}, over a triangular well by reducing the change in the energy level, E_1, when the bias changes.

IV. Results and Discussion

In order to maximize the performance of the EH Bilayer TFET, the subthreshold swing voltage must be minimized while maintaining a high on-state conductance. Minimizing the subthreshold swing requires us to maximize the gate efficiency [$\eta_{gate} \equiv dE_{OL}/d(qV_{G1})$]. The easiest way to maximize the gate efficiency is to vary the body thickness, Fermi level position and channel material and determine the combination that gives the highest gate efficiency for a given on-state conductance. Consequently, we do this for an on-state conductance in the range from 10 μS/μm to 1 mS/μm. This optimization results in a low electron density and a high hole density. As seen from the capacitive voltage divider in Fig. 2(b), this corresponds to minimizing the electron quantum capacitance, $C_{Q,n}$ and maximizing the hole quantum capacitance $C_{Q,p}$.

In Fig. 6 we show the highest gate efficiency for a given conductance after the device thickness and dc biases are optimized relative to on-state conductance. We see that the gate efficiency is quite similar for all three homojunction channel materials, Si, Ge, InAs, at the optimal body thickness, and is around 40% for an on-state conductance of 1mS/μm.

In Table II, we summarize the efficiencies, optimal thickness, Fermi level position, and gate biases for the Si, Ge, and InAs devices.

Interestingly, it should be possible to achieve an on-state conductance of 1 mS/μm in Si if the body is sufficiently thin. Unfortunately, that requires an unrealistically high electric field around 5.8 MV/cm, a tunneling barrier thickness of 1.9 nm and a tunneling probability of 3.9×10^{-3}.

Although all three materials have similar on-state conductance and optimized gate efficiency, the dc bias required is drastically different between the materials. Table II shows the dc bias or Work Function difference on each gate required to align the energy eigenstates. At 1 mS/μm, the voltage across the two gates will be 5.6 V, 2.6 V and 1.45 V for Si, Ge and InAs respectively. It may be possible to achieve the dc bias required for InAs by different gate work functions, but it will be very difficult to achieve the 2 volts or more, that are required for Si and Ge. Furthermore, the unrealistically high electric field required in silicon would cause the gate dielectric to break down [2]. Thus InAs is the best candidate.

Next, we calculate S_{tunnel} the semilog slope of tunnel probability versus body voltage using its definition (11): $S_{tunnel} = dV_{Body}/d\log(I_0)$. This will determine if changing the barrier width has a significant impact on the subthreshold swing voltage. As we are taking $d\log(I_0)$, any quantity proportional to I_0 can also be used as the argument of the log. In particular, G_{tunnel}, defined by (22), is proportional to I_0. I_0 is

TABLE II
Optimized Device Properties

	G_{ON}=10μS/μm			G_{ON}=1mS/μm		
	Si	Ge	InAs	Si	Ge	InAs
η_{gate}	0.48	0.52	0.54	0.39	0.40	0.41
η_{quant}	0.68	0.69	0.64	0.63	0.62	0.55
t_{body} (nm)	5.9	14.9	24.2	4.3	9.7	14.6
ΔE_F (meV)	81	84	30	61	63	17
V_{g1} (V)	1.78	0.82	0.51	2.48	1.25	0.84
V_{g2} (V)	-2.41	-1.01	-0.42	-3.07	-1.43	-0.61
V_{body} (V)	2.15	1.21	0.73	2.52	1.50	1.00
S_{tunnel} (mV/decade)	225	161	141	246	173	227

The analytic gate efficiency for Si, Ge and InAs Bilayer TFETS at an on-state conductance of 1mS/μm and 10μS/μm is summarized. The different material parameters and biases at the on-state required to achieve the optimal gate efficiency are also summarized. S_{tunnel} is the semilog slope measuring how steeply the tunneling conductance changes due to a changing barrier thickness with respect to the band overlap energy, E_{OL}.

[3] Ref [19] also modifies the effective masses to get a better fit for the energy. We capture the 1st order effect of the trapezoidal well by using the original masses. The error introduced by this and the other approximations is quantified in the numerical NextNano simulations.

composed of two key terms, (f_1-f_2) and $\top$. G_{tunnel} contains $\top$ and accounts for (f_1-f_2) through the $\cosh(\Delta E_{Fp}/2k_BT)$ term as discussed in the appendix. Thus we have: $S_{tunnel} = \Delta V_{Body} / \Delta \log(G_{tunnel})$. We can evaluate this by computing G_{tunnel} (22), and V_{Body} (13), for E_{OL}=0 and E_{OL}=100 mV. At 1 mS/μm, S_{tunnel} for Si, Ge and InAs are disappointingly 246, 173 and 227 mV/decade, respectively. Since S_{tunnel} is worse than 60 mV/decade, a steep swing cannot be achieved by electrostatically modulating the tunneling barrier width alone. The EH Bilayer TFET requires a steep band edge density of states, S_{DOS}, for a steep subthreshold swing!

To arrive at a reasonable work function difference, we can employ a heterojunction in the EH Bilayer. Optimizing for only gate efficiency results in a very narrow quantum well and unreasonably large confinement energies and gate biases. Consequently, we limit the bias difference to 1 volt to limit the work function difference needed. This doesn't significantly hurt gate efficiency (<1%). We chose an aluminum concentration of 60% in the AlGaSb in order to give an effective band gap, $E_{g,eff}$, of 100 meV. Using pure GaSb or pure AlSb changes the optimized gate efficiency insignificantly (<1%). The optimized InAs/$Al_{0.6}Ga_{0.4}Sb$ structure for 1 mS/μm is shown in Fig. 5. We find that a 14 nm thick InAs layer and a 2.6 nm thick $Al_{0.6}Ga_{0.4}Sb$ layer gives the best gate efficiency. Overall, we can see that including the heterostructure only slightly increases the gate efficiency as shown in Fig 6.

To verify the accuracy of the analytic calculations we performed quantum simulations in *nextnano++* with a six band k·p model for the valence band and a single band model for the conduction band. We found the simulated gate efficiency for the body thickness and Fermi level positions given in Table II (the gate biases are adjusted to obtain the same Fermi level position). The simulated gate efficiency, η_{gate}, at a conductance of 1mS/μm for Si, Ge and InAs was 44%, 43% and 51% respectively. The confinement efficiency, $\eta_{quant}=dE_{OL}/d(qV_{Body})$, is 79%, 70%, and 66%, for Si, Ge and InAs respectively. The primary difference between the analytical results versus computer simulations is that the simulation does not assume an infinite triangular well as in (14). When simulating the heterojunction, we needed to reduce the thickness of the $Al_{0.6}Ga_{0.4}Sb$ layer to capture the benefit of the trapezoidal quantum well while maintaining the same tunneling barrier height and overall thickness. For a 14.6nm InAs thickness and a 2.0nm $Al_{0.6}Ga_{0.4}Sb$ thickness we simulated a gate efficiency, η_{gate}, of 52% and a confinement efficiency, η_{quant}, of 65%. The numerically computed gate efficiencies are also plotted in Fig. 6

V. Conclusion

We found that a 14.6nm thick InAs EH Bilayer represents the best tradeoff between gate efficiency (51%) and on-state conductance (1mS/μm), demanding a bias or gate work function difference of 1.45V for a homojunction EH Bilayer. We also found that a 9.7nm thick germanium EH Bilayer could achieve a gate efficiency of 43%, if a 2.6V gate work function difference could be engineered. Using an InAs/AlGaSb heterojunction in the EH Bilayer structure reduces the required work function difference to less than a volt, but does not significantly increase in the gate efficiency. Consequently, InAs seems to be the optimal channel material.

Appendix

The tunneling current can be modeled using the transfer Hamiltonian method as developed by Bardeen and Harrison [20, 21]. When applied to a bulk semiconductor, it yields the typical semi-classical WKB tunneling current. By starting with the transfer Hamiltonian formalism, the current can be extended to reduced dimensionalities such as tunneling between two quantum wells [12, 13]. For any arbitrary structure the tunneling current is given by [12, 21]:

$$J_{tunnel} = \frac{4\pi q}{\hbar} \sum_{k_i,k_f} \left|M_{fi}\right|^2 \delta(E_C - E_V)(f_1 - f_2) \tag{A.1}$$

The matrix element is given by [12]:

$$M_{fi} = \frac{-\hbar^2}{2m} i \int \left(\psi_f^* \nabla\psi_i - \psi_i \nabla\psi_f^*\right) \cdot d\vec{S} \tag{A.2}$$

where S is the tunneling interface area. In a square well the matrix element is given by [12]:

$$\left|M_{fi}\right|^2 = \frac{1}{\pi^2} E_{1e} \times E_{1h} \times \top . \tag{A.3}$$

E_{1e} and E_{1h} are the confinement energies and $\top$ is the tunneling probability. In a triangular well, the normalized exponential tail of the wavefunction is given by [22]:

$$\begin{aligned} \psi &\approx 1.426 \frac{1}{2\sqrt{\pi}} \times \frac{1}{\bar{Z}^{1/4}} \times \exp\left(-\frac{2}{3}\bar{Z}^{3/2}\right) \\ &= 1.426 \frac{1}{2\sqrt{\pi}} \times \frac{1}{L_0\sqrt{k}} \times \exp\left(-\int_0^Z k dZ\right) \end{aligned} \tag{A.4}$$

where

$$\bar{Z} = Z / L_0 \text{ and } L_0 = \left(\hbar^2 / (2mqF)\right)^{1/3} \tag{A.5}$$

Plugging (A.4) into (A.2) gives

$$M_{\text{triangular well}} = 0.435 \times M_{\text{square well}} \tag{A.6}$$

As the current is proportional to the square of the matrix element, the 2d-2d tunneling current in a triangular well is $(0.435)^2$ times lower than in a square well.

Furthermore, since the current is flowing over a small energy range near the threshold, the Fermi function difference f_1-f_2 can be Taylor expanded to give:

$$f_1 - f_2 \approx \frac{qV_{SD}}{4k_bT} \times \frac{1}{\cosh(\Delta E_F / 2k_BT)^2} \tag{A.7}$$

References

[1] L. Lattanzio, L. De Michielis, and A. M. Ionescu, "The Electron-Hole Bilayer Tunnel FET," *Solid-State Electronics*, vol. 74, pp. 85-90, Aug 2012.

[2] J. T. Teherani, S. Agarwal, E. Yablonovitch, J. L. Hoyt, and D. A. Antoniadis, "Impact of Quantization Energy and Gate Leakage in Bilayer Tunneling Transistors," *IEEE Electron Device Letters*, vol. 34, pp. 298-300, Feb 2013.

[3] J. P. Leburton, J. Kolodzey, and S. Briggs, "Bipolar tunneling field-effect transistor: A three-terminal negative differential resistance device for high-speed applications," *Applied Physics Letters*, vol. 52, pp. 1608-1610, 1988.

[4] C. Alper, L. Lattanzio, L. De Michielis, P. Palestri, L. Selmi, and A. M. Ionescu, "Quantum Mechanical Study of the Germanium Electron-Hole Bilayer Tunnel FET," *Electron Devices, IEEE Transactions on*, vol. 60, pp. 2754-2760, 2013.

[5] A. C. Seabaugh and Q. Zhang, "Low-Voltage Tunnel Transistors for Beyond CMOS Logic," *Proceedings of the IEEE*, vol. 98, pp. 2095-2110, Dec 2010.

[6] G. Dewey, B. Chu-Kung, J. Boardman, J. M. Fastenau, J. Kavalieros, R. Kotlyar, *et al.*, "Fabrication, Characterization, and Physics of III-V Heterojunction Tunneling Field Effect Transistors (H-TFET) for Steep Sub-Threshold Swing," *2011 IEEE International Electron Devices Meeting (IEDM 2011)*, p. 33.6, 2011 2011.

[7] J. Knoch, S. Mantl, and J. Appenzeller, "Impact of the Dimensionality on the Performance of Tunneling FETs: Bulk Versus One-Dimensional Devices," *Solid-State Electronics*, vol. 51, pp. 572-578, Apr 2007.

[8] M. A. Khayer and R. K. Lake, "Effects of Band-Tails on the Subthreshold Characteristics of Nanowire Band-to-Band Tunneling Transistors," *Journal of Applied Physics*, vol. 110, p. 074508, 2011.

[9] T. Tiedje, E. Yablonovitch, G. D. Cody, and B. G. Brooks, "Limiting Efficiency of Silicon Solar Cells," *IEEE Transactions on Electron Devices*, vol. 31, pp. 711-716, 1984.

[10] S. R. Johnson and T. Tiedje, "Temperature Dependence of the Urbach edge in GaAs," *Journal of Applied Physics*, vol. 78, pp. 5609-5613, 1995.

[11] J. I. Pankove, "Absorption Edge of Impure Gallium Arsenide," *Physical Review*, vol. 140, pp. A2059-A2065, 1965.

[12] S. Agarwal and E. Yablonovitch. (2011, September). Pronounced Effect of pn-Junction Dimensionality on Tunnel Switch Sharpness. *eprint arXiv:1109.0096*. Available: http://arxiv.org/abs/1109.0096

[13] S. Agarwal and E. Yablonovitch, "Using Dimensionality to Achieve a Sharp Tunneling FET (TFET) Turn-on," *69th Annual Device Research Conference (DRC)*, pp. 199-200, 2011.

[14] S. Agarwal and E. Yablonovitch, "Enhanced Tunneling Current in 1d-1d Edge Overlapped TFET's," *2012 70th Annual Device Research Conference (DRC)*, pp. 63-6464, 2012 2012.

[15] S. M. Sze and K. K. Ng, *Physics of semiconductor devices*: Wiley-Interscience, 2007.

[16] E. O. Kane, "Zener Tunneling in Semiconductors," *Journal of Physics and Chemistry of Solids*, vol. 12, pp. 181-188, 1959.

[17] M. Luisier and G. Klimeck, "Simulation of Nanowire Tunneling Transistors: From the Wentzel--Kramers--Brillouin Approximation to Full-Band Phonon-Assisted Tunneling," *Journal of Applied Physics*, vol. 107, p. 084507, 2010.

[18] M. Lundstrom and J. Guo, *Nanoscale Transistors: Device Physics, Modeling and Simulation*. New York: Springer 2006.

[19] S. Oh and H. S. P. Wong, "A Physics-Based Compact Model of III-V FETs for Digital Logic Applications: Current-Voltage and Capacitance-Voltage Characteristics," *IEEE Transactions on Electron Devices*, vol. 56, pp. 2917-2924, 2009.

[20] J. Bardeen, "Tunneling From a Many Particle Point of View," *Physical Review Letters*, vol. 6, pp. 57-59, 1961.

[21] W. A. Harrison, "Tunneling from an Independent-Particle Point of View," *Physical Review*, vol. 123, pp. 85-89, 1 July 1961.

[22] D. J. Griffiths, *Introduction to Quantum Mechanics*, 2nd ed. Upper Saddle River, NJ: Prentice Hall, Inc., 1994.

[23] H. Kroemer, "The 6.1 Å Family (InAs, GaSb, AlSb) and its Heterostructures: A Selective Review," *Physica E: Low-dimensional Systems and Nanostructures*, vol. 20, pp. 196-203, 2004.

[24] I. Balslev and P. Lawaetz, "On the Interpretation of the Observed Hole Mass Shift with Uniaxial Stress in Silicon," *Physics Letters*, vol. 19, pp. 6-7, 1965.

[25] P. Lawaetz, "Valence-Band Parameters in Cubic Semiconductors," *Physical Review B*, vol. 4, pp. 3460-3467, 1971.

[26] I. Vurgaftman, J. R. Meyer, and L. R. Ram-Mohan, "Band Parameters for III-V Compound Semiconductors and Their Alloys," *Journal of Applied Physics*, vol. 89, pp. 5815-5875, Jun 2001.

[27] S. L. Chuang, *Physics of Optoelectronic Devices*. New York: John Wiley & Sons, Inc, 1995.

Sapan Agarwal received the B.S. degree in electrical engineering from the University of Illinois, Urbana Champaign, in 2007 and the Ph.D. degree in from the University of California, Berkeley in 2012.

He is currently a Postdoctoral Fellow at UC Berkeley.

James T. Teherani received the B.S. in electrical and computer engineering from the University of Texas at Austin in 2008 and the M.S. from the Massachusetts Institute of Technology (MIT), Cambridge in 2010.

He is currently pursuing his PhD at MIT. His research interests include tunneling transistors, emerging materials, and high mobility devices.

Judy L. Hoyt received the B.S. in Physics and Applied Mathematics from UC Berkeley her Ph.D. in Applied Physics from Stanford University in 1987.

She is a Professor with the Department of Electrical Engineering and Computer Science at the Massachusetts Institute of Technology.

Dimitri A. Antoniadis received his B.S. from the National University of Athens in and his Ph.D. in Electrical Engineering from Stanford University 1976.

He is the Ray and Maria Stata Professor of Electrical Engineering at the Massachusetts Institute of Technology.

Eli Yablonovitch received his Ph.D. in applied physics from Harvard University in 1972.

He is a Professor of Electrical Engineering and Computer Sciences at UC Berkeley and the Director of the NSF Center for Energy Efficient Electronics Science.