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
Copyright © 2013, by the author(s).
All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

Acknowledgement

The work was supported by the Center for Energy Efficient Electronics Science (NSF Award 0939514)
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 $S_{\text{Urbach}}$=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.

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.

\(^1\) 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.
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:

\[ SS = \left( \frac{d \log(I)}{dV_G} \right)^{-1} \]  

(1)

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

\[ I \propto \left( f_C - f_V \right) \times \int \times D_C(E) \times D_V(E) \times \hat{c}E \]  

(2)

The difference in the Fermi occupation probabilities between the conduction and valence bands is \((f_C-f_V)\) and the transmission probability of a tunneling electron is \(\hat{c}E\). \(D_C(E)\) 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]

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:

\[ D_C(E) = \begin{cases} \frac{D'_C}{D_{C0} \times e^{(E'_C-E)/qV_0}}, & E \geq E'_C \\ \frac{D'_V}{D_{V0} \times e^{-(E-E'_V)/qV_0}}, & E < E'_V \end{cases} \]  

(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:

\[ D_V(E) = \begin{cases} \frac{D'_V}{D_{V0} \times e^{-(E-E'_V)/qV_0}}, & E \leq E'_V \\ \frac{D'_C}{D_{C0} \times e^{(E'_C-E)/qV_0}}, & E > E'_C \end{cases} \]  

(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:

\[ D_C(E) \times D_V(E) = \begin{cases} \frac{E_{OL}}{e^{qV_0}} \times \frac{D'_C \times D_{V0} \times e^{-v(E'_C-E)/qV_0} \times E \geq E'_C}{D_{C0} \times D_{V0},} & E_V < E < E'_C \\ \frac{D'_V \times D_{C0} \times e^{-(E'_V-E)/qV_0} \times E \leq E'_V}{D_{C0} \times D_{V0},} \end{cases} \]  

(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:

\[ I \propto \left( \int \frac{E'_C}{E'_V} (f_C - f_V) \times \hat{c}E \times e^{E_{OL}/qV_0} \right) \times \hat{c}E \]  

(6a)

\[ I \propto I_0 \times e^{E_{OL}/qV_0} \]  

(6b)

where the tunneling pre-factor is:

\[ I_0 = \int \frac{E'_C}{E'_V} (f_C - f_V) \times \hat{c}E \]  

(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):

---

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

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

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

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

### III. EH BILAYER MODELING

To model the EH Bilayer TFET we consider the situation where the bias on the n-gate ($V_{\text{GI}}$) is changed while the bias on the p-gate ($V_{\text{GZ}}$) 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: $\eta_{\text{gate}}=\Delta E_{\text{OL}}/(qAV_{\text{GI}})$. Consequently, we need $E_{\text{OL}}$ and $V_{\text{GI}}$ in the on and off states. The device will turn on once the bands overlap and $E_{\text{OL}}=0$. In section III.A we find the gate biases, $V_{\text{GI}}$ and $V_{\text{GZ}}$ required to achieve $E_{\text{OL}}=0$ and a given Fermi level position. In section III.B we determine $E_{\text{OL}}$ and $V_{\text{GI}}$ in the off state to find $\eta_{\text{gate}}$. 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, $E_{\text{le}}$ and $E_{\text{lh}}$, and the voltage across the body, $V_{\text{Body}}$, in the on-state. The overlap energy, $E_{\text{OL}}$, is given by [2]:

$$E_{\text{OL}} = qV_{\text{Body}} - (E_G + E_{\text{le}} + E_{\text{lh}}) \quad (13)$$

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:

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

where $\alpha$ represents either electrons (e) or holes (h) and $t_{\text{Body}}$ is the thickness of the bilayer semiconductor body. The effective triangular 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 $E_{\text{OL}}=0$ and so we can solve (13) for $V_{\text{Body}}$ and then find $E_{\text{le}}$ and $E_{\text{lh}}$.

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

For Si, the accumulation charge to capture the effect of the quantum capacitance. We also assumed that all the accumulation charge is located at the oxide interface. 

The eigenstates and the Fermi level, 

where 

The hole masses are computed from 

and 

The tunneling mass is given by: 

All values for AlGaAs are linearly interpolated. 

TABLE I

<table>
<thead>
<tr>
<th>MATERIAL PROPERTIES USED</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>$E_g$</td>
</tr>
<tr>
<td>Si</td>
</tr>
<tr>
<td>Ge</td>
</tr>
<tr>
<td>InAs</td>
</tr>
<tr>
<td>GaSb</td>
</tr>
<tr>
<td>AlSb</td>
</tr>
</tbody>
</table>

$E_{g,o}$ 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 

and 

and 

The tunneling mass is given by: 

(26) and (27). All values for AlGaAs are linearly interpolated.

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

and

$$ qV_2 = -\frac{E_g}{2} + E_{1h} - \Delta E_{F_p} \tag{16} $$

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

Given a Fermi level position, $\Delta E_F$, we can find the electron charge, $Q_n$, and hole charge, $Q_p$, in the channel:

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

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

where

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

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:

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

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

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 $\eta_{gate} = \Delta E_{OL}/q\Delta V(G1)$. 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, $\Delta E_{OL}$, needs to change in order to turn the tunneling off. A rough estimate is sufficient, since varying $\Delta E_{OL}$ from 50 to 200 meV only changes $\eta_{gate}$ by a few percent. Consequently, we take $\Delta E_{OL} = -100$ meV. If the band edge density of states $S_{DOS}$ is 20 mV/decade (corresponding to the optically measured steepness), $\Delta E_{OL} = -100$ meV will give 5 decades of on/off ratio.

To find $V_{GI}$ in the off state, $V_{GIOFF}$, 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 (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, $\left|Q_p\right|$, is a function of $V_2$ through $\Delta E_{Fp}$ 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 $V_{GI}$ using (15) to define $\Delta E_{Fp}$ for the charge density. This gives us $V_{GI}$ in the off-state, $V_{GIOFF}$. As we already found $V_{GI}$ in the on-state, $V_{GION}$, at end of Section III.A, we can finally compute the gate efficiency as $\Delta E_{OL}/q(V_{GION}-V_{GIOFF})$.

C. Tunneling Conductance

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

$$ G_{tunnel} = \frac{q m_{J DOS}^* L_e W}{2 \pi \hbar^3} \times \left(0.435 E_{1e}\right) \times \left(0.435 E_{1h}\right) \times \frac{q}{4 k_B T} \times \frac{1}{\cosh(\Delta E_F/2 k_B T)^2} \tag{22} $$

Where the tunnel transmission probability is:

$$ T(F) = \exp \left(-\frac{\pi (m_{tunnel}^*)^{1/2} E_G^{3/2}}{2 \sqrt{2} h q F}\right) \tag{23} $$

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 $E_{1c}$ and $E_{1v}$ then $\Delta E_F = \Delta E_{Fp}$, given by (15). If $E_F$ is above $E_{1c}$ and $E_{1v}$ then $\Delta E_F = \Delta E_{Fp}$, given by (16). If $E_F$ is in between $E_{1c}$ and $E_{1v}$ 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:

$$ m_{J DOS}^* = 2\left(1/m_{e,t}^* + 1/m_{h,t}^*\right)^{-1} \tag{25} $$
The tunneling mass can be computed from [16]:

\[ m_{\text{tunnel}}^* = 2\left(\frac{1}{m_{e,z}^*} + \frac{1}{m_{h,z}^*}\right)^{-1} \tag{26} \]

The transverse masses, \( m_{e,z}^* \) and \( m_{h,z}^* \), as well as the masses in the tunneling direction, \( m_{e,t}^* \) and \( m_{h,t}^* \), 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 use a two-band tunneling model and need to adjust the mass:

\[ m_{2\text{Band}}^* = \left(\frac{2\sqrt{2}}{\pi} \times 4\frac{\sqrt{2}}{3}\right)^2 m_{\text{Band}}^* \tag{27} \]

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/μm, the channel conductance becomes the limiting conductance. Consequently, a smaller \( \Delta E_F \) is required to increase the number of electrons.

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

\[ G = \frac{W q n_s v_T}{2 k_B T / q} \quad \text{(28)} \]

where

\[ v_T = \sqrt{2 k_B T / m_{e,t}^*} \]

and \( n_s = |Q_n| / q \)

\[ \text{E. Heterojunction Analysis} \]

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, \( E_{g,\text{eff}} \), can be widely tuned by changing the Al content.

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

\[ C_S = \left(\frac{t_n}{\varepsilon_{s,n}} + \frac{t_p}{\varepsilon_{s,p}}\right)^{-1} \tag{30} \]

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:

\[ F_n = V_{\text{Body}}/\left(\varepsilon_{n,n} t_n / \varepsilon_{s,n}\right) \tag{31} \]

\[ F_p = V_{\text{Body}}/\left(\varepsilon_{p,p} t_p / \varepsilon_{s,n}\right) \tag{32} \]

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:

\[ E_{B,n} = q F_n t_n - E_{i,n} \tag{33} \]

\[ E_{B,p} = q F_p t_p - E_{i,b} \tag{34} \]

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, \( E_{i,e} \) or \( E_{i,h} \), is large, the barrier height would be so 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:

\[ T_n = \exp\left[-4(2 \times m_{e,z}^*)^{1/2} E_{B,n}^{3/2} / (3 h q F_n)\right] \tag{35} \]

\[ T_p = \exp\left[-4(2 \times m_{h,z}^*)^{1/2} E_{B,p}^{3/2} / (h q F_p)\right] \tag{36} \]

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...
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]:

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

$E_{tri}$ 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 $\hbar^2 \pi^2 / (2m_e^* a^2)$ or $\hbar^2 \pi^2 / (2m_h^* p^2)$. The trapezoidal quantum well shape improves the quantum confinement efficiency, $\eta_{square}$, 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 $[\eta_{gate} = dEOL/d(qV_{d})]$. 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 $\mu$S/$\mu$m to 1 mS/$\mu$m. 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, $C_{Q,p}$ and maximizing the hole quantum capacitance $C_{Q,n}$.

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 1 mS/$\mu$m.

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/$\mu$m 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 $3.9 \times 10^{-3}$.

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/$\mu$m, 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

---

1 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.

---

| TABLE II |
| Optimized Device Properties |
|---|---|---|---|---|
| $G_{on}=10\mu$S/$\mu$m | $G_{on}=1m$S/$\mu$m |
| Si | Ge | InAs | Si | Ge | InAs |
| $\eta_{gate}$ | 0.48 | 0.52 | 0.54 | 0.39 | 0.40 | 0.41 |
| $\eta_{quant}$ | 0.68 | 0.69 | 0.64 | 0.63 | 0.62 | 0.55 |
| $L_{body} (nm)$ | 5.9 | 14.9 | 24.2 | 4.3 | 9.7 | 14.6 |
| $\Delta E_f$ (meV) | 81 | 84 | 30 | 61 | 63 | 17 |
| $V_{gs}$ (V) | 1.78 | 0.82 | 0.51 | 2.48 | 1.25 | 0.84 |
| $V_{gd}$ (V) | -2.41 | -1.01 | -0.42 | -3.07 | -1.43 | -0.61 |
| $V_{body}$ (V) | 2.15 | 12.1 | 0.73 | 2.52 | 1.50 | 1.00 |
| $S_{tunnel}$ |
| (mV/decade) | 225 | 161 | 144 | 246 | 173 | 227 |

The analytic gate efficiency for Si, Ge and InAs Bilayer TFETS at an on-state conductance of 1 mS/$\mu$m and 10 $\mu$S/$\mu$m 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 tunnel conductance changes due to a changing barrier thickness with respect to the band overlap energy, $E_{OL}$. 

---

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 $\Delta E_{OL} = 100$ 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.
composed of two key terms, \((f_1-f_2)\) and \(\mathcal{T}\). \(G_{\text{tunnel}}\) contains \(\mathcal{T}\) and accounts for \((f_1-f_2)\) through the \(\cosh(\Delta E_{\text{F}_1}/2k_BT)\) term as discussed in the appendix. Thus we have:

\[
S_{\text{tunnel}} = \Delta V_{\text{Body}} / \log(G_{\text{tunnel}}).
\]

We can evaluate this by computing \(G_{\text{tunnel}}\) (22), and \(V_{\text{Body}}\) (13), for \(E_{\text{OX}}=0\) and \(E_{\text{OX}}=100\) mV. At 1 mS/\(\mu\)m, \(S_{\text{tunnel}}\) for Si, Ge and InAs are disappointingly 246, 173 and 227 mV/decade, respectively. Since \(S_{\text{tunnel}}\) 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, \(S_{\text{DOS}}\), 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, \(E_{\text{gap},\text{eff}}\), 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/\(\mu\)m 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 analytical 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, \(\eta_{\text{gate}}\), at a conductance of 1mS/\(\mu\)m for Si, Ge and InAs was 44%, 43% and 51% respectively. The confinement efficiency, \(\eta_{\text{conf}}=dE_{\text{OX}}/d(qV_{\text{Body}})\), 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, \(\eta_{\text{gate}}\), of 52% and a confinement efficiency, \(\eta_{\text{conf}}\), 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/\(\mu\)m), 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]:

\[
J_{\text{tunnel}} = \frac{4e}{\hbar} \sum_{k_1,k_f} \left| M_{f_1} \right|^2 \delta(E_C - E_V)(f_1 - f_2).
\]  

(A.1)

The matrix element is given by [12]:

\[
M_{f_1} = -\frac{\hbar^2}{2m} \int \left( \psi_{f_1}^* \nabla \psi_1 - \psi_1 \nabla \psi_{f_1}^* \right) d\mathbf{S}.
\]  

(A.2)

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

\[
\left| M_{f_1} \right|^2 = \frac{1}{\pi^2} E_{1e} \times E_{1h} \times \mathcal{T}.
\]  

(A.3)

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

\[
\psi \approx 1.426 \frac{1}{2\pi} \frac{1}{Z^{1/4}} \exp \left( -\frac{2}{3} Z^{3/2} \right)
\]

\[
= 1.426 \frac{1}{2\pi} \frac{1}{L_0 \sqrt{k}} \exp \left( -\int_0^Z k dZ \right)
\]  

(A.4)

where

\[
Z = Z / L_0 \quad \text{and} \quad L_0 = \left( \hbar^2 / (2mqF) \right)^{1/3}
\]  

(A.5)

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

\[
M_{\text{triangular well}} = 0.435 \times M_{\text{square well}}
\]  

(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:

\[
f_1 - f_2 \approx \frac{qV_{SD}}{4k_BT} \times \frac{1}{\cosh(\Delta E_{F}/2k_BT)^2}
\]  

(A.7)

REFERENCES


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.