Received 29 June 2020; revised 23 August 2020; accepted 13 September 2020.
Date of publication 21 September 2020; date of current version 1 February 2021.

Digital Object Identifier 10.1109/JXCDC.2020.3024798

Minimum-Energy Digital Computing With Steep Subthreshold Swing Tunnel FETs

DANIEL S. TRUESDELL (Member, IEEE), SHEIKH Z. AHMED (Student Member, IEEE), AVIK W. GHOSH (Senior Member, IEEE), and BENTON H. CALHOUN (Senior Member, IEEE)

Charles L. Brown Department of Electrical and Computer Engineering, University of Virginia, Charlottesville, VA 22904-474 USA

CORRESPONDING AUTHOR: D. S. TRUESDELL (dst4b@virginia.edu)

This work was supported in part by the Semiconductor Research Corporation under Award 2694.02 and in part by NSF under Award 1640053.

ABSTRACT Energy efficiency in digital circuits is limited by the subthreshold swing (SS), which defines how abruptly a transistor switches between its ON and OFF-states. The SS is particularly important for circuits targeting minimum-energy computation which operate in the subthreshold region between the ON and OFF-states of the transistor. The SS of MOSFET devices is fundamentally limited by thermionic emission, which has inspired a search for new devices whose SS can reach below the Boltzmann thermal limit. Tunnel field-effect transistors (TFETs) have emerged as a post-CMOS candidate with low (steep) SS and have been investigated using an evolving selection of geometries and materials that yield continuously improving device performance and circuit performance estimates. To unify previous works and guide future TFET iterations, this article provides a comprehensive theory on minimum-energy operation in the subthreshold region for steep-SS devices. We show that the optimal supply voltage for energy minimization and minimum obtainable energy are both proportional to the SS, and that a fundamental limit exists for the required I_{ON}/I_{OFF} to achieve operation at the minimum-energy point. We explore how device knobs affect the optimization space for minimum-energy operation, and analyze how common TFET nonidealities affect the potential for minimum-energy operation.

INDEX TERMS Energy efficiency, minimum energy, performance optimization, steep-slope devices, subthreshold swing (SS), tunnel field-effect transistor (TFET).

I. INTRODUCTION

DIGITAL CMOS circuit performance has been steadily approaching its fundamental limit as dictated by Boltzmann physics, which restricts the subthreshold swing (SS) of MOSFET devices to $k_BT \ln 10/q$ (60 mV/decade) due to thermionic emission. The SS describes how abruptly a device transitions between its ON and OFF-states which are defined as the operating regions above and below a certain threshold voltage, respectively. The SS defines the *I–V* characteristic of the sub-threshold region and is a key factor in limiting the ratio of current delivered by the device in its ON and OFF-states, known as the I_{ON}/I_{OFF} ratio. High-performance computing applications operate in the super-threshold region to maximize I_{ON}/I_{OFF}, and are constrained by power consumption and thermal dissipation. The SS sets a hard limit on the amount of power and thermal reduction that can be accomplished before the device fails to operate at a tolerable clock frequency, which has inspired a search for a post-CMOS device without a thermally limited SS [1]. The tunnel field-effect transistor (TFET) [2], [3] has emerged as a popular candidate device for achieving a sub-thermal SS, and has been analyzed from a variety of circuit perspectives including scaling, layout, reliability, digital, analog, RF, and memory design that collectively inform device and circuit design for performance-oriented TFET-based circuits. [4]–[12]

However, high-performance circuits are not the only victims of the SS limit. Energy-constrained applications, such as battery-powered and energy-harvesting circuits, operate in the sub-threshold region where the minimum-energy point

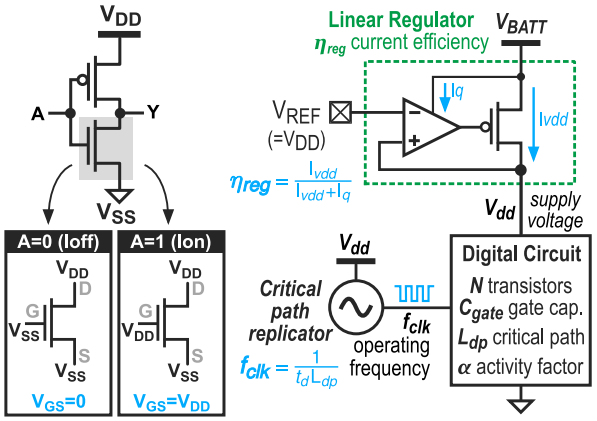

FIGURE 1. Digital circuit model using a linear voltage regulator for tunable supply voltage and critical path replicator for maximum clock speed.

occurs [13], so the SS directly affects the performance of minimum-energy circuits. Some works have highlighted and discussed minimum-energy operation for TFET circuits [6], [7], [12], [14], [15], but the exact results are dependent on specific device geometries and materials. Energy minimization has been explicitly proven for steep-SS TFETs using a theoretical approach with an emphasis on susceptibility to process variation [16], but operating frequency is not considered and the results do not account for losses in voltage regulation. Together, these works demonstrate the capability and potential impact of minimum-energy operation for TFETs, but a more comprehensive analysis is required for a foundational understanding of the device and circuit-level factors for achieving minimum-energy operation with steep-SS TFETs.

In this work, we implement a fully analytical digital circuit model based on foundational modeling techniques from both CMOS circuits and TFET devices ([13], [18]) that includes operating frequency and voltage regulation, and use it to evaluate energy minimization based on device parameters with emphasis on the SS. We demonstrate the impact of voltage regulation losses on the minimum-energy point and model the optimum supply voltage for minimizing energy based on the SS. We show how device and circuit parameters affect the limit on the minimum energy per cycle, and derive a theoretical limit on the device I_{ON}/I_{OFF} for achieving minimum-energy operation. We also solve for the optimal threshold voltage for minimizing energy at target operating frequency and discuss how device parameters influence the optimization space for a circuit. Finally, we compare these results with data from a quasi-analytical TFET model, showing how device nonidealities impact the potential for minimum-energy operation.

II. MINIMUM-ENERGY OPERATION FOR STEEP-SS DEVICES

Fig. 1 shows a digital circuit model that consists of N transistors each with a gate capacitance C_{gate}. The circuit has a critical path length of L_{dp} (normalized to the delay of an inverter), and an activity factor α. The critical path length describes the slowest (typically the longest) possible path of logic gates through which a signal might need to propagate in order for the circuit to complete a given operation. Thus, an L_{dp} of 20 indicates that the slowest logical path of a circuit is equivalent to the delay of 20 consecutive inverters. The activity factor of any gate corresponds to the probability (ranging from 0 to 1) that its input will change from low (logic 0) to high (logic 1) on a clock cycle. Gates that switch every clock cycle will have α of 1, while most CMOS logic ranges closer to $\alpha = 0.1 - 0.25$. In this model, the α-value represents the average activity factor of all gates in the entire circuit. A linear voltage regulator with current efficiency η_{reg} supplies the circuit with a supply voltage V_{dd}, which is regulated from a source voltage V_{batt}. The current consumption and maximum operating frequency of the circuit can be expressed using a combination of its static and dynamic currents, which can in turn be related to device currents in the ON and OFF-states. A simplified subthreshold current equation is given by

$$I_{ds} = I_0 10^{(V_{gs}-V_{th})/SS} \tag{1}$$

where I_0 is the current at threshold defined in this work to be 100 nA, SS is the SS, V_{th} is the threshold voltage, and drain-induced barrier lowering (DIBL) is neglected for simplicity. The specific selection of I_0 in this case is based on a population of TFET performance data in [17], where we assume a nanometer-scale device (10–100 nm). Then, the ON-current I_{ON} and OFF-current I_{OFF} for a given subthreshold supply voltage $V_{dd} < V_{th}$ can be expressed as

$$I_{ON} = I_0 10^{(V_{dd}-V_{th})/SS} \tag{2}$$

$$I_{OFF} = I_0 10^{-V_{th}/SS}. \tag{3}$$

For ideal circuit performance, n- and p-type devices should have symmetrical electrical characteristics (I_0, V_{th}, C_{gate}, etc.). However, to better consider the variations between n- and p-type devices, the average of their values can be used in these models. The total energy consumed from V_{batt} per clock cycle of the digital circuit is expressed as a sum of the dynamic energy $E_{cycle,dynamic}$ and static energy $E_{cycle,static}$

$$E_{cycle} = E_{cycle,dynamic} + E_{cycle,static} \tag{4}$$

$$E_{cycle,dynamic} = \frac{1}{\eta_{reg}} \alpha N C_{gate} V_{dd} V_{batt} \tag{5}$$

$$E_{cycle,static} = \frac{I_{OFF}}{\eta_{reg} f_{clk}} \frac{N}{2} \left(1 - \frac{\alpha}{L_{dp}}\right) V_{batt} \tag{6}$$

where f_{clk} is the clock operating frequency which is the inverse of the cycle time. A background on the derivation of (4)–(6) is given in [18]. The maximum clock frequency for any $V_{dd} < V_{th}$ can be modeled as

$$f_{clk}(V_{dd}) = \frac{1}{t_d L_{dp}} = \frac{I_{ON}}{K L_{dp} C_{gate} V_{dd}} = \frac{I_0 10^{(V_{dd}-V_{th})/SS}}{K L_{dp} C_{gate} V_{dd}} \tag{7}$$

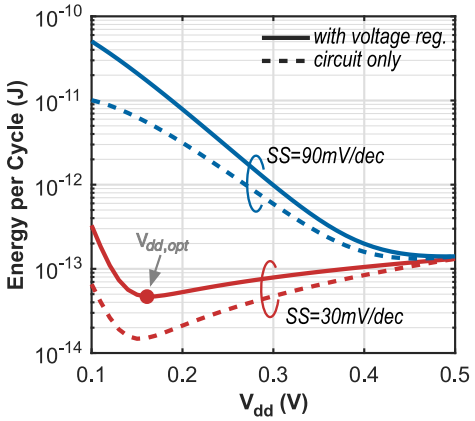

FIGURE 2. Energy per clock cycle versus supply voltage for SS = 30 mV/decade and SS = 90 mV/dec decade. Voltage regulator losses diminish the energy savings at low V_{dd} where the minimum-energy points occurs for steep-SS devices.

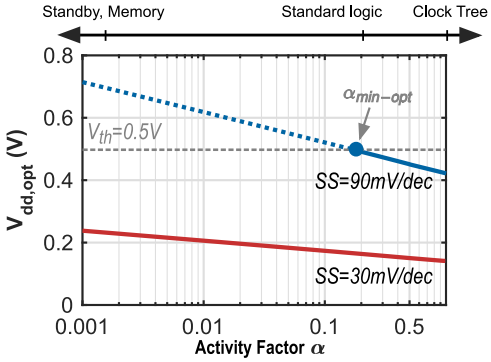

FIGURE 3. Optimum supply voltage $V_{dd,opt}$ for energy minimization versus activity factor α for SS of 30 and 90 mV/decade, given $V_{th} = 0.5$ V and assuming that f_{clk} is maximized under the V_{dd} limit according to (7). Depending on the topology and SS, $V_{dd,opt}$ for some α may exceed V_{th}, indicating that energy minimization is not possible.

where $t_d = KC_{gate}V_{dd}/I_{ON}$ represents the gate delay of a single inverter using the traditional CV/I approximation and K is a scaling factor used for calibration. Operation at the maximum frequency allowable by the critical path is a necessary condition for achieving minimum-energy operation due to a minimized integration time of leakage power without affecting dynamic energy [18]. Note that this result has an exponential dependence on V_{dd}, so small linear variations in C_{gate} across V_{dd} can be considered negligible to preserve modeling simplicity. Alternatively, (7) can be rearranged to find the minimum V_{DD} required to operate at a given f_{clk}

$$V_{dd}(f_{clk}) = \frac{-SS}{\ln 10} W_{-1}\left[\frac{-I_0 10^{-V_{th}/SS}\ln 10}{C_{gate} f_{clk} K L_{dp} SS}\right] \quad (8)$$

where W_{-1} is the negative (lower) branch of the lambert W function. Simplifying the energy per clock cycle in terms of the supply voltage, we combine (7) with (4)–(6)

$$E_{cycle}(V_{dd}) = \frac{1}{\eta_{reg}}\alpha N C_{gate} V_{dd} V_{batt} + \frac{1}{2\eta_{reg}}\left(L_{dp} - \alpha\right) C_{gate} K N 10^{-V_{dd}/SS} V_{dd} V_{batt}. \quad (9)$$

Fig. 2 shows the energy per cycle versus supply voltage for a circuit of $N = 5000$, $\alpha = 0.2$, $C_{gate} = 0.5$ fF/μm, and L_{dp}= 20. These values are arbitrarily chosen, but for reference a 32-bit ripple-carry adder can be characterized as $N \approx 1000$ and $L_{DP} \approx$ 10–20. V_{th} and V_{batt} are both set to 0.5 V. Different curves are shown for sub and super-thermal SSs of 30 and 90 mV/decade, respectively. With a higher SS, E_{cycle} is more easily dominated by the static energy contribution of leaking devices, and therefore experiences a declining energy efficiency (increasing energy per cycle) as the supply voltage is lowered. If low (steep) SS devices are used instead, the energy can be minimized by balancing the dynamic and static contributions by scaling the supply voltage further into the subthreshold region to an optimum value $V_{dd,opt}$, which can be calculated as

$$V_{dd,opt} = \frac{SS}{\ln 10}\left(1 - W_{-1}\left[\frac{2\alpha e}{K\left(\alpha - L_{dp}\right)}\right]\right). \quad (10)$$

The energy savings from operating at $V_{dd,opt}$ are reduced due to power losses in the linear regulator, which become more significant as V_{dd} is lowered, as shown in Fig. 2. For circuits using steep-SS devices with low $V_{dd,opt}$, it therefore becomes important to use a V_{batt} that is as close to the intended V_{dd} as possible. Note that $V_{dd,opt}$ is directly proportional to SS, and has no dependence on the regulator current efficiency or on operating parameters such as f_{clk} or V_{th}. Instead, it is dictated by the topological parameters α and L_{dp}. Fig. 3 shows the $V_{dd,opt}$ computed for a variety of α, ranging from highly inactive topologies such as memory ($\alpha \approx 0.005$) to typical logic designs ($\alpha \approx 0.1$) and highly active topologies such as clock trees ($\alpha = 1$). Steep-SS devices are less sensitive to the activity factor and therefore have a more consistent target $V_{dd,opt}$ for minimum-energy operation. For architectures with varying workload, frequency scaling, or standby modes that target dynamic energy minimization, this simplifies the voltage regulation range for keeping the circuit operating at the minimum-energy point. In some cases when static energy is particularly significant (when α is very low), the $V_{dd,opt}$ computed from (10) will exceed V_{th} which renders the subthreshold model invalid and corresponds to a case when no minimum-energy points exists in the subthreshold region. The minimum α that allows for energy minimization at a given V_{th} is calculated as the lowest α that yields a $V_{ddp,opt} < V_{th}$. Therefore, we equate (10) with V_{th} and solve for α

$$\alpha_{min-opt} = \frac{L_{dp}\left(SS - V_{th}\ln 10\right)}{SS\left(1 - \frac{2}{K}10^{V_{th}/SS}\right) - V_{th}\ln 10}. \quad (11)$$

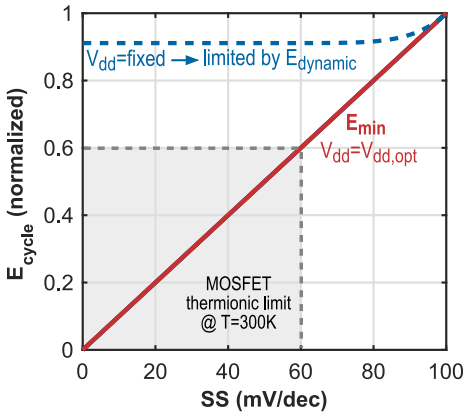

FIGURE 4. **Minimum energy per cycle E_{min} scales proportionally with SS regardless of f_{clk}, V_{th}, or I_0. E_{min} is restricted for thermionically-limited MOSFETs. Reducing SS without keeping V_{dd} set to $V_{dd,opt}$ will reduce E_{cycle} slightly but eventually becomes limited by dynamic energy.**

By setting the circuit supply voltage to $V_{\mathrm{dd,opt}}$, the minimum possible energy per cycle E_{min} can be calculated as

$$E_{\mathrm{min}} = \frac{\theta C_{\mathrm{gate}} N V_{\mathrm{batt}} \mathrm{SS}}{2\eta_{\mathrm{reg}} e \ln 10} \left(K \left(L_{\mathrm{dp}} - \alpha \right) e^{1-\theta} + 2\alpha e \right) \quad (12)$$

$$\theta = 1 - W_{-1}\left[\frac{2\alpha e}{K \left(\alpha - L_{\mathrm{dp}} \right)} \right]. \quad (13)$$

Equation (12) reveals that the value of E_{min} scales proportionally with SS as shown in Fig. 4 and is independent of any other device metrics such as I_0 or V_{th}, and is also independent of f_{clk}. Other dependencies are more intuitive, such as the proportional dependence of E_{min} on gate capacitance and the number of transistors in the circuit.

Another important insight is created by evaluating the $I_{\mathrm{ON}}/I_{\mathrm{OFF}}$ ratio at the minimum energy point, when $V_{\mathrm{dd}} = V_{\mathrm{dd,opt}}$. By using (10) with (2) and (3), we find that $I_{\mathrm{ON}}/I_{\mathrm{OFF}}$ required for energy minimization depends only on the circuit architecture

$$\left. \frac{I_{\mathrm{ON}}}{I_{\mathrm{OFF}}} \right|_{E_{\mathrm{min}}} = e^{\theta}. \quad (14)$$

This result is true regardless of the device parameters (V_{th}, I_0, SS) or target f_{clk}. Fig. 5 illustrates this fundamental limit by showing a device that is initially operating at E_{min} with $I_{\mathrm{ON}}/I_{\mathrm{OFF}} = e^{\theta}$ before the SS is decreased. The lower SS slightly reduces E_{cycle} since static energy is lower due to a reduced I_{OFF}, meaning that E_{cycle} is now dominated by dynamic energy. The reduced SS also proportionally reduces $V_{\mathrm{dd,opt}}$ and E_{min}. Since V_{dd} is higher than $V_{\mathrm{dd,opt}}$ now, $I_{\mathrm{ON}}/I_{\mathrm{OFF}} > e^{\theta}$ and E_{cycle} could be further reduced to E_{min} by setting V_{dd} to $V_{\mathrm{dd,opt}}$ to re-balance the dynamic and static energy. Frequency will be reduced from the lower V_{dd} if I_0 and V_{th} remain fixed, but this can be compensated by reducing V_{th} while holding I_0 constant so that the same I_{ON} now occurs at $V_{\mathrm{dd,opt}}$. For the circuit under consideration, e^{θ} is roughly 10^5–10^6 for α ranging from 0.01 to 0.5.

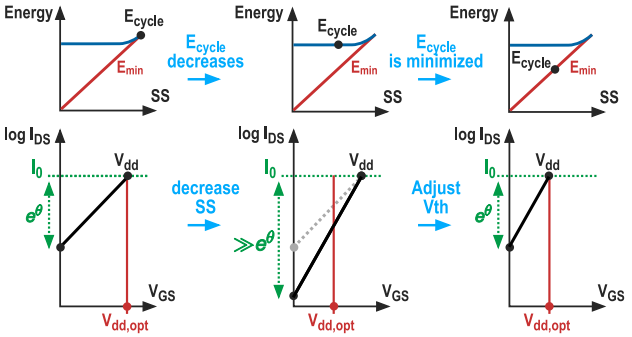

FIGURE 5. **Energy is minimized when $V_{\mathrm{dd}} = V_{\mathrm{dd,opt}}$, which occurs when $I_{\mathrm{ON}}/I_{\mathrm{OFF}}$ is equal to e^{θ}. Reducing SS and obtaining a higher $I_{\mathrm{ON}}/I_{\mathrm{OFF}}$ without adjusting V_{dd} will reduce energy but not minimize it.**

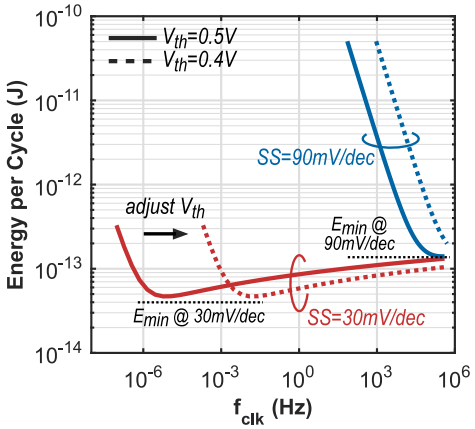

FIGURE 6. **Energy per clock cycle versus operating frequency f_{clk} for different SS, assuming the V_{dd} is minimized for operation at each f_{clk} by using (8). Changing the threshold voltage allows the minimum energy point to be shifted to any target f_{clk}.**

Much of the existing work on TFETs has focused not only on reducing SS but also sustaining it over as many decades of current as possible to achieve the highest $I_{\mathrm{ON}}/I_{\mathrm{OFF}}$ ratio [17]. Increasing I_{ON} improves operating speed according to (7) while reducing thermal dissipation by running at lower V_{dd}, but devices aimed at minimum-energy applications should emphasize SS reduction once the required $I_{\mathrm{ON}}/I_{\mathrm{OFF}}$ ratio has been met.

A. DESIGNING FOR MINIMUM-ENERGY AT A TARGET CLOCK FREQUENCY

Up to this point, our analysis has implicitly set f_{clk} to its maximum possible value across V_{dd} [in accordance with (7)] for the given circuit parameters while keeping I_0 and V_{th} fixed. If we instead represent the energy per clock cycle as a function of f_{clk} by using (8) with (4)–(6), we find the frequency f_{opt} at which E_{min} occurs, which is shown in Fig. 6.

f_{opt} can be directly calculated by substituting (10) into (7)

$$f_{opt} = \frac{10^{-V_{th}/SS} e^{\theta} I_0 \ln 10}{\theta C_{gate} K L_{dp} SS}. \quad (15)$$

For the steep-SS device, the minimum-energy frequency is exceptionally low (given $I_0 = 100$ nA and $V_{th} = 0.5$ V), making it unfit for modern processing applications. Although operation at higher frequencies is easily possible by increasing V_{dd} beyond $V_{dd,opt}$, doing so would preclude operation at E_{min}. Instead, the ideal solution is to keep V_{dd} fixed to $V_{dd,opt}$ and shift E_{min} to a new frequency by finding a new knob to adjust f_{opt} without affecting the value of E_{min}. According to (12)–(15), this can be done by changing V_{th} or I_0 since they have no impact on E_{min}. Decreasing V_{th} while maintaining the same I_0 and SS moves the minimum-energy point to a higher frequency, as shown by the dotted lines in Fig. 6. The same effect can be accomplished by increasing I_0 at a fixed V_{th}. Note that since $V_{dd,opt} < V_{th}$, no devices in the circuit will actually be operating at V_{th} and therefore the current I_0 at $V_{gs} = V_{th}$ may simply represent an extrapolation of I_{ds} beyond $V_{gs} = V_{dd,opt}$. Based on these concepts, minimum-energy operation can be achieved for steep-SS devices at a target f_{clk} by first determining $V_{dd,opt}$ and then choosing $V_{th,opt}$ in order to position the minimum-energy point at the target frequency. $V_{th,opt}$ can then be calculated by equating (8) with (10) and solving for f_{clk}

$$V_{th,opt} = SS \log\left[\frac{e^{\theta} I_0 \ln 10}{\theta C_{gate} f_{clk} K L_{dp} SS}\right]. \quad (16)$$

If the resulting $V_{th,opt}$ for a given f_{clk} is smaller than the calculated $V_{dd,opt}$, this again corresponds to a case when the energy cannot be minimized in the subthreshold region. That is, if any $V_{th,opt} < V_{dd,opt}$ were used with (11), the resulting $\alpha_{min\text{-}opt}$ would be larger than the original design value of α, indicating that the original α was too small to allow for energy minimization at $V_{th} = V_{th,opt}$. Rather than using (11) to constrain the optimization space, the domain of (16) can be more intuitively limited by finding the maximum target clock frequency $f_{max\text{-}opt}$ that yields a $V_{th,opt} < V_{dd,opt}$

$$f_{max\text{-}opt} = \frac{e^{1-\theta} I_0 \ln 10}{K e^{1-\theta} \left(C_{gate} L_{dp} SS - \alpha\right) + 2\alpha e}. \quad (17)$$

Fig. 7 shows $V_{th,opt}$ versus f_{clk} for different SS and α values. As the SS decreases, $V_{th,opt}$ also decreases and becomes less sensitive to both α and f_{clk}. The maximum frequency that can be reached at E_{min} ($f_{max\text{-}opt}$) also improves slightly as SS is reduced, but is most strongly dependent on I_0 as shown in (17).

B. DEVICE DESIGN AND TRADEOFFS FOR MINIMUM-ENERGY OPERATION

Combining the results from this section, we can summarize several key considerations and tradeoffs for designing devices targeting minimum-energy operation. The minimum achievable energy per clock cycle E_{min} is limited by the SS and can be obtained when the supply voltage V_{dd} equals $V_{dd,opt}$ which balances the static and dynamic energy. This condition is universally satisfied when the device can achieve an I_{ON}/I_{OFF} of e^{θ}. If $I_{ON}/I_{OFF} < e^{\theta}$, then E_{cycle} will be higher than E_{min} due to high static energy, and if $I_{ON}/I_{OFF} > e^{\theta}$, E_{cycle} will be higher than E_{min} due to high dynamic energy. The operating frequency at E_{min} (while $V_{dd} = V_{dd,opt}$) can be independently scaled (without affecting the value of E_{min} or the I_{ON}/I_{OFF} ratio) by adjusting I_0 or V_{th}. On the device side, I_0 and V_{th} can be tuned by using different TFET structures to obtain different electric field profiles in the ON-state, or by changing materials used to form the source and channel bands. The value of E_{min} can be reduced by decreasing the SS and does not depend on I_0 or V_{th}. Therefore, strictly from an energy minimization standpoint, SS should be minimized at all costs as long as an $I_{ON}/I_{OFF} = e^{\theta}$ can be maintained. In many TFET designs, a result of reducing SS is that I_{ON} and I_{OFF} are lowered, which only means that the operating frequency at E_{min} will be very slow as shown in Fig. 6. Still, this scenario is more energy efficient than running faster (higher I_0 but same I_{ON}/I_{OFF} ratio) at the expense of a higher SS. Improving I_0 should be viewed as an orthogonal optimization that improves the frequency range over which E_{min} may be obtained. Fig. 8 shows this orthogonal relationship that I_0 and SS have with $f_{max\text{-}opt}$ and E_{min}.

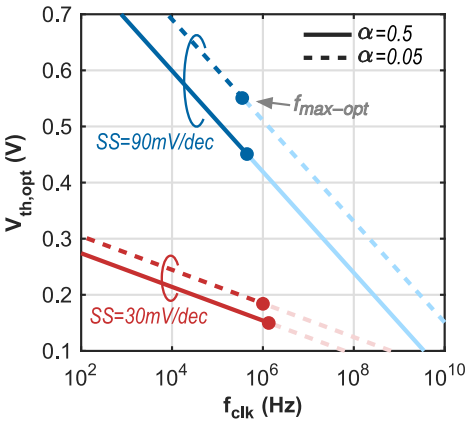

FIGURE 7. Optimum threshold voltage $V_{th,opt}$ to minimize energy consumption versus target operating frequency f_{clk}, shown for different SS. Supply voltage is fixed across f_{clk}, set by (10).

III. EFFECTS OF TFET NONIDEALITIES ON ENERGY MINIMIZATION

In order to study the effect of nonidealities on minimum-energy optimization, we created a quasi-analytical model for double gated III–V TFETs [18]. The model includes major nonidealities such as trap-assisted tunneling (TAT) and Auger generation. For the results discussed in this section, we simulate an n-type double gated InGaAs homojunction TFET.

We employ a two-band $k.p$ model to accurately model the real and imaginary bands of the materials. Our model is fitted to an environment dependent tight-binding model that is calibrated to state-of-the-art DFT band structure and wave functions [19]. This model allows us to accurately predict the

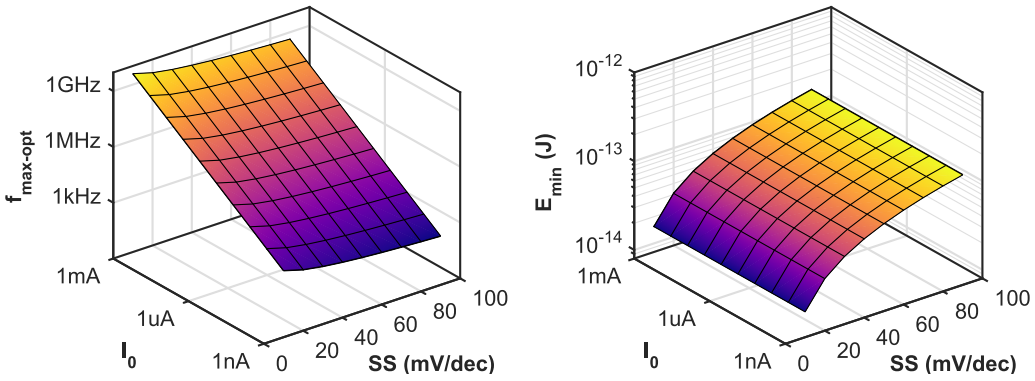

FIGURE 8. **Minimum possible energy per clock cycle E_{min} and the maximum clock frequency $f_{max\text{-}opt}$ that can be optimized to achieve E_{min}, shown as a function of I_0 and SS. $f_{max\text{-}opt}$ depends primarily on I_0, while E_{min} is affected only by SS.**

tunneling probability of electrons in the device. For obtaining the potential profile we solve a pseudo-2-D Poisson's equation. An analytical equation is used to model the channel surface potential [20]. The band-to-band tunneling (BTBT) current is computed using the modified Simmons equation, which is a popular chemistry equation used for calculating the Wentzel–Kramers–Brillouin (WKB) current through thin films [21]. We modify the equation for nonrectangular barriers and include effect of finite temperature. The modified equation is derived from the well-known Landauer formalism

$$I = \frac{q}{h}\int T(E)\left[f_S(E) - f_D(E)\right]dE \tag{18}$$

where $T(E)$ is the total transmission summed over all transverse states for a given energy, $f_{S/D}(E)$ represents the source/drain Fermi–Dirac functions that set the approximate energy window for tunneling electrons.

One of the key issues in III–V TFETs is the presence of traps at the oxide-channel interface as well other hetero interfaces. These traps creates a leakage path for electrons in the OFF-state, leading to increased leakage current and SS. We consider TAT as a Fowler–Nordheim type tunneling process through a tilted barrier around the trap. The TAT current per unit width can be written in the following compact form: [18], [22]:

$$I_{\text{TAT}} = \frac{q}{2} v_{rcmb} n_i \Gamma d \left[1 - e^{-qV_{\text{DS}}/k_B T}\right] \tag{19}$$

where Γ is the electric field enhancement due to TAT and thermionic emission processes, d is the width of the trap active region, and n_i represents the intrinsic carrier concentration. The recombination velocity $v_{\text{rcmb}} = \sigma v_{\text{th}} N_t$, where N_t is the surface trap density per unit area at the midgap energy. Only midgap trap states are considered in our model as they have the highest occupation probability for electrons experiencing TAT.

The large electric field near the source/channel junction can trigger an additional current leakage mechanism called Auger generation even in immaculate interfaces with low trap densities. Auger generation involves scattering of electrons through Coulombic interactions. In this process, a "hot" electron promotes a valence band electron to the conduction band by colliding with it. This process limits the minimum SS that can be achieved in the TFET in the absence of traps [23]. The Auger generation rate is computed using Fermi's Golden Rule [18], [23]

$$G = \frac{1}{A}\frac{2\pi}{\hbar}\sum_{1,1',2,2'} P\left(1, 1', 2, 2'\right) \times |M|^2 \delta\left((E_1 - E_{1'} + E_2 - E_{2'}\right) \tag{20}$$

where P is the occupation probability and M is the matrix element that couples the initial states with the final states.

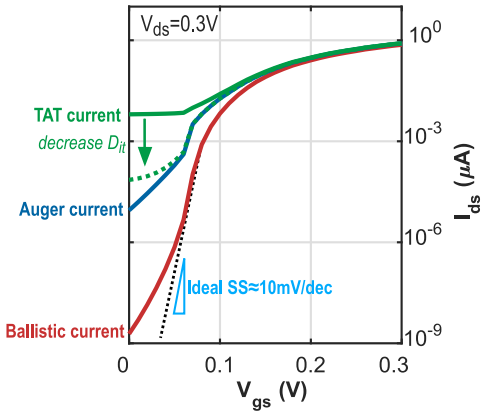

FIGURE 9. **I_{ds}–V_{gs} plot of homojunction InGaAs TFET showing deterioration of SS due to Auger and TAT currents.**

Fig. 9 shows the TFET drain–source current versus gate voltage computed using (18). Ballistic current closely matches an ideal SS of 10 mV/decade. Auger and TAT currents contribute to increased leakage current at $V_{\text{gs}} = 0$ V, which effectively increases the SS. The interface trap density is nominally set to $D_{\text{it}} = 5 \times 10^{12}$ m^{-2} eV^{-1}, which dominates the total current at low V_{gs}. Decreasing the trap density by 10 000× reduces TAT to the point that Auger current instead dominates the total current.

We can evaluate the impact of these device currents on the energy per cycle by using computed values for I_{ON} and I_{OFF} in place of (2) and (3), and substituting them directly into the energy per cycle model in (4)–(6). The results, shown in Fig. 10, demonstrate how the addition of nonidealities

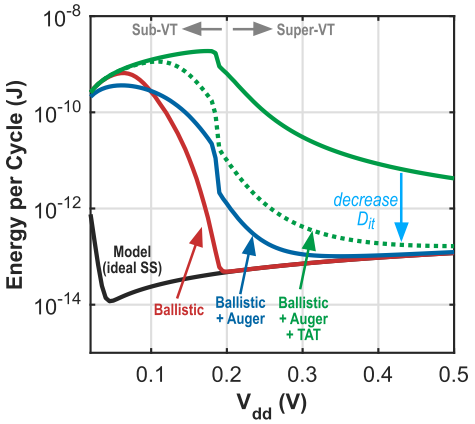

FIGURE 10. Energy per clock cycle versus supply voltage, shown for the ideal device model derived in Section II as well as the quasi-analytical TFET model that considers ballistic, Auger, and TAT currents. Nonidealities increase the energy per cycle and $V_{dd,opt}$.

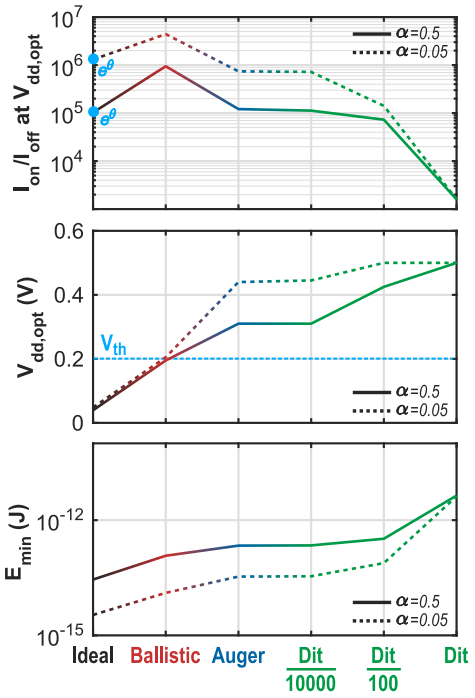

FIGURE 11. Addition of nonidealities in the TFET I-V characteristic increases the minimum obtainable energy per cycle E_{min} and optimum supply voltage $V_{dd,opt}$ for achieving E_{min}. When nonidealities are considered, circuits are overwhelmingly dominated by leakage energy, so the minimum-energy point occurs in the super-threshold region.

increases the energy per cycle of the circuit. Ballistic current mirrors an ideal SS down to low V_{gs} before it tapers away with rising leakage current that increases the static energy of the circuit. This effect appears as increased E_{cycle} at low V_{dd} that deviates from the ideal minimum-energy point and shifts $V_{dd,opt}$ higher. Adding Auger and TAT current to the TFET model increases the SS even more, so E_{cycle} and $V_{dd,opt}$ both rise further. Note that these results are in line with the theoretical models given by (10) and (12) despite $V_{dd,opt}$ exceeding V_{th} of the device.

As $V_{dd,opt}$ rises with increasing nonidealities, energy minimization becomes more difficult. While energy in a ballistic-only device can be minimized in the sub-threshold region, the addition of Auger currents moves the minimum-energy point to the super-threshold region, and TAT currents eliminate the existence of a true minimum-energy point. Fig. 11 summarizes these trends and revisits the optimum I_{ON}/I_{OFF} ratio for minimum-energy operation. In the ideal model, $I_{ON}/I_{OFF} = e^{\theta}$ as expected. Energy for the ballistic model is minimized in the subthreshold region but since the SS increases and I_0 and V_{th} are not compensated, I_{ON}/I_{OFF} rises above e^{θ} (see Fig. 5). When Auger and TAT currents are present, the minimum-energy point lies outside the subthreshold region where our models and assumptions are no longer valid. Still, $V_{dd,opt}$ and E_{min} continue to increase into the super-threshold region which agrees with our model, while the I_{ON}/I_{OFF} ratio decreases due to increasing I_{OFF}.

IV. CONCLUSION

This work presents a theoretical model for evaluating minimum-energy operation in steep-SS TFET devices. We show that the minimum energy consumption and optimal supply voltage scale proportionally with SS, and that voltage regulation losses diminish energy savings as SS is reduced. Steep-SS devices can be optimized for minimum-energy operation as long as their I_{ON}/I_{OFF} ratio reaches a fundamental threshold limit that is dictated only by circuit architecture, and device knobs such as threshold voltage and ON-current can be used to achieve minimum-energy operation at a target clock frequency. Typical device tradeoffs between SS and ON-current have orthogonal control over the energy optimization space by dictating the minimum obtainable energy E_{min} and maximum clock frequency that can be reached under minimum-energy operation. TFET nonidealities such as Auger and TAT currents deteriorate the SS and increase the minimum energy point.

REFERENCES

[1] D. E. Nikonov and I. A. Young, "Benchmarking of beyond-CMOS exploratory devices for logic integrated circuits," *IEEE J. Explor. Solid-State Comput. Devices Circuits*, vol. 1, pp. 3–11, 2015, doi: 10.1109/JXCDC.2015.2418033.

[2] A. C. Seabaugh and Q. Zhang, "Low-voltage tunnel transistors for beyond CMOS logic," *Proc. IEEE*, vol. 98, no. 12, pp. 2095–2110, Dec. 2010, doi: 10.1109/JPROC.2010.2070470.

[3] A. Ionescu and H. Riel, "Tunnel field-effect transistors as energy-efficient electronic switches," *Nature*, vol. 479, pp. 329–337, Nov. 2011.

[4] H. Lu, P. Paletti, W. Li, P. Fay, T. Ytterdal, and A. Seabaugh, "Tunnel FET analog benchmarking and circuit design," *IEEE J. Explor. Solid-State Comput. Devices Circuits*, vol. 4, no. 1, pp. 19–25, Jun. 2018, doi: 10.1109/JXCDC.2018.2817541.

[5] J. Min and P. M. Asbeck, "Compact modeling of distributed effects in 2-D vertical tunnel FETs and their impact on DC and RF performances," *IEEE J. Explor. Solid-State Comput. Devices Circuits*, vol. 3, pp. 18–26, Dec. 2017, doi: 10.1109/JXCDC.2017.2670606.

[6] M. Alioto and D. Esseni, "Tunnel FETs for ultra-low voltage digital VLSI circuits: Part II–evaluation at circuit level and design perspectives," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 22, no. 12, pp. 2499–2512, Dec. 2014, doi: 10.1109/TVLSI.2013.2293153.

[7] S. Strangio, P. Palestri, M. Lanuzza, F. Crupi, D. Esseni, and L. Selmi, "Assessment of InAs/AlGaSb tunnel-FET virtual technology platform for low-power digital circuits," *IEEE Trans. Electron Devices*, vol. 63, no. 7, pp. 2749–2756, Jul. 2016, doi: 10.1109/TED.2016.2566614.

[8] U. E. Avci, D. H. Morris, and I. A. Young, "Tunnel field-effect transistors: Prospects and challenges," *IEEE J. Electron Devices Soc.*, vol. 3, no. 3, pp. 88–95, May 2015, doi: 10.1109/JEDS.2015.2390591.

[9] V. Saripalli, S. Datta, V. Narayanan, and J. P. Kulkarni, "Variation-tolerant ultra low-power heterojunction tunnel FET SRAM design," in *Proc. IEEE/ACM Int. Symp. Nanosc. Architectures*, San Diego, CA, USA, Jun. 2011, pp. 45–52.

[10] B. Sedighi, X. S. Hu, H. Liu, J. J. Nahas, and M. Niemier, "Analog circuit design using tunnel-FETs," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 62, no. 1, pp. 39–48, Jan. 2015, doi: 10.1109/TCSI.2014.2342371.

[11] Y. Lee *et al.*, "Low-power circuit analysis and design based on heterojunction tunneling transistors (HETTs)," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 21, no. 9, pp. 1632–1643, Sep. 2013, doi: 10.1109/TVLSI.2012.2213103.

[12] U. E. Avci *et al.*, "Energy efficiency comparison of nanowire heterojunction TFET and Si MOSFET at Lg=13nm, including P-TFET and variation considerations," in *IEDM Tech. Dig.*, Washington, DC, USA, Dec. 2013, p. 33, doi: 10.1109/IEDM.2013.6724744.

[13] B. H. Calhoun, A. Wang, and A. Chandrakasan, "Modeling and sizing for minimum energy operation in subthreshold circuits," *IEEE J. Solid-State Circuits*, vol. 40, no. 9, pp. 1778–1786, Sep. 2005.

[14] D. H. Morris, U. E. Avci, R. Rios, and I. A. Young, "Design of low voltage tunneling-FET logic circuits considering asymmetric conduction characteristics," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 4, no. 4, pp. 380–388, Dec. 2014, doi: 10.1109/JETCAS.2014.2361054.

[15] S. Strangio *et al.*, "Digital and analog TFET circuits: Design and benchmark," *Solid-State Electron.*, vol. 146, pp. 50–65, Aug. 2018.

[16] H. Fuketa *et al.*, "Design guidelines to achieve minimum energy operation for ultra low voltage tunneling FET logic circuits," *Jpn. J. Appl. Phys.*, vol. 54, no. 4S, Jan. 2015, Art. no. 04DC04.

[17] H. Lu and A. Seabaugh, "Tunnel field-effect transistors: State-of-the-art," *IEEE J. Electron Devices Soc.*, vol. 2, no. 4, pp. 44–49, Jul. 2014, doi: 10.1109/JEDS.2014.2326622.

[18] S. Z. Ahmed, Y. Tan, D. S. Truesdell, B. H. Calhoun, and A. W. Ghosh, "Modeling tunnel field effect transistors—From interface chemistry to non-idealities to circuit level performance," *J. Appl. Phys.*, vol. 124, no. 15, 2018, Art. no. 154503.

[19] Y. Tan, M. Povolotskyi, T. Kubis, T. B. Boykin, and G. Klimeck, "Transferable tight-binding model for strained group IV and III-V materials and heterostructures," *Phys. Rev. B, Condens. Matter*, vol. 94, no. 4, Jul. 2016, Art. no. 045311.

[20] C. Wu, R. Huang, Q. Huang, C. Wang, J. Wang, and Y. Wang, "An analytical surface potential model accounting for the dual-modulation effects in tunnel FETs," *IEEE Trans. Electron Devices*, vol. 61, no. 8, pp. 2690–2696, Aug. 2014, doi: 10.1109/TED.2014.2329372.

[21] J. G. Simmons, "Electric tunnel effect between dissimilar electrodes separated by a thin insulating film," *J. Appl. Phys.*, vol. 34, no. 9, pp. 2581–2590, Sep. 1963.

[22] R. N. Sajjad and D. Antoniadis, "A compact model for tunnel FET for all operation regimes including trap assisted tunneling," in *Proc. 74th Annu. Device Res. Conf. (DRC)*, Jun. 2016, pp. 1–2.

[23] J. T. Teherani, S. Agarwal, W. Chern, P. M. Solomon, E. Yablonovitch, and D. A. Antoniadis, "Auger generation as an intrinsic limit to tunneling field-effect transistor performance," *J. Appl. Phys.*, vol. 120, no. 8, Aug. 2016, Art. no. 084507.

DANIEL S. TRUESDELL (Member, IEEE) received the B.S. degree in electrical engineering from the University of Central Florida, Orlando, FL, USA, in 2016. He is currently pursuing the Ph.D. degree in electrical engineering with the University of Virginia, Charlottesville, VA, USA.

His current research interests include energy-efficient mixed-signal circuit design, on-chip oscillators and sensors, energy-harvesting and performance-scalable systems, and low-power digital circuit techniques.

SHEIKH Z. AHMED (Student Member, IEEE) received the B.Sc. degree from BRAC University, Dhaka, Bangladesh, in 2014. He is currently pursuing the Ph.D. degree with the University of Virginia, Charlottesville, VA, USA.

His research interests include the computational study of nanoelectronic and photonic devices, analytical modeling, quantum transport, and compact modeling.

AVIK W. GHOSH (Senior Member, IEEE) received the Ph.D. degree in condensed matter theory from Ohio State University, Columbus, OH, USA, and a Postdoctoral Fellowship in electrical engineering from Purdue University, West Lafayette, IN, USA.

He is a Professor with the Charles Brown Department of Electrical and Computing Engineering and the Department of Physics, University of Virginia, Charlottesville, VA, USA. He is the UVA Site-Director of the NSF-Industry University Cooperative Center on Multifunctional Integrated Systems Technology (MIST). He has authored over 125 refereed papers and a book (*Nanoelectronics – a Molecular View*, World Scientific, 2016) in the area of computational nano-materials and devices. He has given over 125 invited lectures worldwide.

Dr. Ghosh is a fellow of the Institute of Physics (IOP) and has received the IBM Faculty Award, the NSF CAREER Award, a 2006 best paper award from the Army Research Office, and UVA's All University Teaching Award.

BENTON H. CALHOUN (Senior Member, IEEE) received the B.S. degree in electrical engineering from the University of Virginia, Charlottesville, VA, USA, in 2000 and the M.S. and Ph.D. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2002 and 2006, respectively.

In January 2006, he joined the faculty at the Electrical and Computer Engineering Department, University of Virginia, where he is currently a Professor. His research has emphasized energy efficient and subthreshold circuit design for self-powered, batteryless wireless sensing systems. Starting from fundamental advances in subthreshold circuits, he has expanded his work to include complete self-powered nodes for Internet of Things (IoT) and body-worn applications. His research interests include self-powered wireless sensors for the internet of things, batteryless systems, body area sensor networks, low power digital circuit design, system-on-chip architecture and circuits for energy constrained applications, system driven embedded hardware/software design, wakeup receivers, energy harvesting-power management units, subthreshold digital circuits, subthreshold SRAM, energy efficient communication, power harvesting and delivery circuits, low power mixed signal design, and medical applications for low energy electronics. He is a coauthor of *Subthreshold Design for Ultra Low-Power Systems* (Springer, 2006), author of *Design Principles for Digital CMOS Integrated Circuit Design* (NTS Press, 2012), and has over 200 peer reviewed publications and 22 issued U.S. patents that contribute to the field of energy efficient circuits and systems for self-powered and energy constrained applications. He is a Campus Director and Technical Thrust Leader in the NSF Nanosystems Engineering Research Center (ERC) for Advanced Self-Powered Systems of Integrated Sensors and Technologies (ASSIST). He co-founded and is co-CTO at Everactive, Inc., which is selling self-powered, energy harvesting wireless sensing solutions in the industrial IoT market.