

# Dynamic Voltage, Body Bias and Frequency Scaling for FD-SOI-Based Low-Power Edge Processors

Shrihari Gokulachandran\*, Navneet Jain<sup>†</sup>, Andreas Gerstlauer\*

\*Electrical and Computer Engineering, The University of Texas at Austin, <sup>†</sup>GlobalFoundries, USA

\*{shrihari.gerstl}@utexas.edu, <sup>†</sup>navneet.jain@globalfoundries.com

**Abstract**—Fully depleted silicon-on-insulator (FD-SOI) has emerged as a proven technology for energy-efficient edge deployments. FD-SOI offers a wide body biasing range, providing an additional knob for trading off energy and performance that has remained underexplored. Traditional approaches have applied body biasing only statically, at coarse granularity, or limited to corner tightening for adaptive PVT compensation. This overlooks the potential of body biasing for power-energy tuning on top of traditional dynamic voltage and frequency scaling (DVFS) in response to system or workload characteristics.

In this paper, we explore simultaneous dynamic voltage, body bias, and frequency scaling (DVBFS) to maximize power and energy efficiency on edge processors. Optimal operating points depend on the trade-off between leakage and dynamic energy, which is affected differently by supply voltage and bias while also varying with synthesis conditions. We introduce a methodology and automated framework to jointly determine optimal synthesis parameters and runtime DVBFS settings. We show that for workloads dominated by active execution, synthesizing for maximum performance together with DVBFS is near-optimal. By contrast, when workloads alternate between active and significant idle periods, using DVBFS with a netlist optimized for leakage delivers superior energy efficiency. We evaluate our approach on an embedded RISC-V processor running TinyML applications. Results demonstrate up to 29% energy savings from DVBFS compared to DVFS.

## I. INTRODUCTION

Edge computing and IoT devices often demand high performance with equally high power and energy efficiency. FD-SOI technology has emerged as a compelling alternative to traditional FinFET technology, offering comparable performance with a more cost-effective planar architecture and unique power-performance optimization capabilities [1]. In particular, body biasing in FD-SOI provides an additional dimension for power-performance tuning beyond traditional voltage and frequency scaling. However, existing approaches have not fully explored this potential. Biasing has traditionally been used for design-time optimization through static bias islands [2]–[4]. Several approaches have explored adaptive body biasing (ABB) to compensate for process, voltage, and temperature (PVT) variations [5]–[10], including combined with dynamic voltage and frequency scaling (DVFS) [11]–[14]. However, their focus has primarily been on performance boosting.

FD-SOI provides a wide body biasing range that opens new possibilities for workload-dependent dynamic power and energy optimization not considered in prior work. Supply-voltage and bias scaling affect performance, dynamic energy and leakage at different rates. This tradeoff depends on workload characteristics and synthesis conditions, which influence netlist parameters such as cell composition and logic quality.

Independent scaling of body bias and supply voltages thus creates the challenge of finding optimal operating points that minimize energy or power consumption at a given target frequency in response to workload variations. The optimal modulation of both supply voltage and body bias during runtime, influenced by synthesis parameters and workload characteristics, creates a complex design space and an open optimization problem that has not been studied so far.

In this paper, we propose a novel approach that combines design-time synthesis parameter selection with dynamic voltage, bias and frequency scaling (DVBFS) for workload-aware power and energy optimization in FD-SOI based designs. We present a comprehensive methodology and an automated framework for joint exploration of synthesis and runtime conditions. We apply this methodology to explore the DVBFS design space, in order to determine optimal supply voltage and body bias operating points, as a function of the operating frequency and synthesis parameters for a given design and workload. We outline strategies for determining optimal synthesis conditions to enable DVBFS, where results show that independent scaling of body bias and supply voltage can achieve better total energy efficiency than traditional power management techniques.

In summary, this paper makes the following contributions:

- A novel runtime power management strategy based on dynamic voltage, bias and frequency scaling (DVBFS), combined with a synthesis strategy to enable optimal DVBFS operating points.
- A systematic methodology and an automated design space exploration framework to determine optimal runtime operating conditions and design-time synthesis parameters, enabling power-energy optimization of a given RTL specification and workload.
- An evaluation of our approach using an embedded RISC-V processor running TinyML and biomedical edge applications, demonstrating that independent scaling of body bias and supply voltage can achieve up to 29% better total energy efficiency depending on workload characteristics.

The rest of this paper is organized as follows: Section II reviews related work, and Section III provides a background on bias scaling. Section IV describes our methodology and setup for joint synthesis and runtime exploration. Section V examines the impact of synthesis parameters, while Section VI evaluates runtime strategies across workload scenarios. Finally, Section VII presents experimental results and Section VIII concludes the paper with a summary.

## II. RELATED WORK

Prior work on body biasing in FD-SOI technology has explored both design-time and runtime strategies for body-biasing. At design time, industry practice typically relies on brute-forcing bias configurations for pre-defined bias domains using synthesis tools [2], [3]. Some research efforts [4], [15] explore algorithmic approaches for body-bias domain partitioning and static body-bias configurations. However, all of these methods do not account for dynamic biasing based on workload variability, focusing instead on independent system-level design-time partitioning and post-silicon optimization.

At runtime, several works [5], [7]–[9] employed adaptive body biasing (ABB) in FD-SOI technology to compensate for process, voltage, and temperature (PVT) variations. Beyond pure PVT compensation, [6], [10] also use ABB to modulate leakage power for reducing chip temperature. In all of these cases, however, biasing is only used to overcome performance degradations. Several works [11]–[14] explored joint body-biasing with DVFS, but they focused on implementation concerns without determining actual operating points.

In traditional CMOS technology, [16] combines ABB with DVS in deeply depleted channel CMOS to maintain an energy-optimal operating point for a given frequency. Martin et al. [17] proposed a combined DVFS and ABB approach for single-core processors. Yan et al. [18] extended this concept to heterogeneous embedded systems. These efforts relied on analytical models for optimization that ignore the much wider biasing range and critical effects, such as short-circuit energy, in FD-SOI. Furthermore, none of the existing works explored the dynamic behavior of different workload scenarios or the impact of synthesis parameters on optimal operating points and joint optimization of design-time and runtime parameters.

## III. BACKGROUND

In FD-SOI, Forward Body-Bias (FBB) reduces the threshold voltage. The threshold voltage  $V_{TN}$  and  $V_{TP}$  of NFET and PFET transistors shifts approximately linearly with their respective back-gate (well) biases  $V_{NW}$  and  $V_{PW}$  [19]–[21]:

$$V_{TX} = V_{TX,0} - \eta_X V_{XW}, \quad X \in \{N, P\}, \quad (1)$$

where  $V_{TX,0}$  is the front-gate threshold at zero back-gate bias and  $\eta_X$  is the body-factor. For an NFET in FBB,  $V_{NW} > 0$  and  $V_{TN,0} > 0$  so an increase in  $V_{NW}$  decreases  $V_{TN}$ . Similarly, for a PFET,  $V_{PW} < 0$  and  $V_{TP,0} < 0$ , i.e. an increase in  $|V_{PW}|$  pushes  $V_{TP}$  closer to 0, hence decreasing  $|V_{TP}|$ .

1) *Dynamic Energy*: Dynamic Energy ( $E_{dyn}$ ) =  $E_{sw} + E_{sc}$  consists of a switching component ( $E_{sw}$ ), associated with charging and discharging capacitances, and a short-circuit component ( $E_{sc}$ ), caused by direct-path current when both NFET and PFET are momentarily on during switching.

From [22]–[25], for a node with effective load capacitance  $C_{eff}$ , supply-voltage  $V_{DD}$ , and activity factor  $\alpha$ ,

$$E_{sw} = \alpha C_{eff} V_{DD}^2, \quad (2)$$

$$E_{sc} = \alpha \frac{\beta\tau}{12} [V_{DD} - V_{TN} - |V_{TP}|]^3, \quad (3)$$

where  $\beta$  is the process transconductance and  $\tau$  is the average input rise/fall time. Using (1) in (3), and grouping technology-specific terms into constants  $c_0$ – $c_3$  we get

$$E_{sc} = c_0\tau [c_1 + V_{DD} + c_2 V_{NW} + c_3 V_{PW}]^3. \quad (4)$$

2) *Leakage Power*: Leakage power ( $P_{leak}$ ) is given by

$$P_{leak} = V_{DD} I_{leak,avg}, \quad (5)$$

where  $I_{leak,avg}$  is the average leakage current. For this work, we approximate  $I_{leak,avg}$  as being dominated by subthreshold component  $I_{sub,avg}$ . Assuming a 50/50 input probability,  $I_{sub,avg}$  can be expressed in terms of NFET and PFET off-currents  $I_{sub,N}$  and  $I_{sub,P}$  as

$$I_{sub,avg} = \frac{1}{2} (I_{sub,N} + I_{sub,P}). \quad (6)$$

For simplicity, we use the drain-current equation in the sub-threshold region from [26] at the off-state ( $V_{GS} = 0$  and  $V_{DS} = V_{DD}$ ):

$$I_{sub,X} = I_0 \left[ 1 - \exp\left(\frac{-V_{DD}}{U_T}\right) \right] \left[ \exp\left(\frac{-V_{TX} - V'_{off}}{n_X U_T}\right) \right], \quad (7)$$

where  $V'_{off}$  is the offset voltage,  $n_X$  is the subthreshold slope and  $U_T$  is the thermal voltage. Using (1), (5)–(7),  $V_{DD} \gg U_T$  and grouping technology-specific terms into  $k_0$ – $k_5$  we obtain:

$$P_{leak} = V_{DD} (k_0 e^{k_1 + k_2 V_{NW}} + k_3 e^{k_4 + k_5 V_{PW}}). \quad (8)$$

3) *Delay*: The propagation delay ( $t_p$ ) of a CMOS stage can be written using an  $\alpha$ -power law [27]:

$$t_p = \frac{C_{eff} V_{DD}}{I_{d,sat}} \approx \frac{C_{eff} V_{DD}}{\beta (V_{DD} - V_{TX})^\alpha}, \quad (9)$$

where  $I_{d,sat}$  is the saturation drain current,  $\beta$  is the process transconductance and  $1 \leq \alpha \leq 2$  is the velocity saturation index. The maximum operating frequency is  $f_{max} \propto 1/t_p$ .

## IV. METHODOLOGY

Fig. 1 shows our workload-based analysis and optimization flow for joint exploration of synthesis and runtime parameters on edge CPU platforms. We cross-compile workloads to the target platform, convert the compiled binaries to hex files and load them onto the testbench for RTL simulation. We synthesize the design using the desired libraries and constraints, and perform gate-level simulation to capture glitching and accurate timing information for improved power analysis accuracy. Finally, on each netlist, we perform power analysis at different DVBFs operating points ( $V_{DD}$ , ( $V_{NW}$ ,  $V_{PW}$ )).

We emulate runtime scaling on each netlist, by performing power-energy analysis with different cell libraries representing different runtime operating conditions ( $V_{DD}$  and  $V_{XW}$  scaling), each at their maximum operating frequency  $f_{max}$ , determined by post-synthesis Static Timing Analysis (STA) on each netlist using the libraries used to describe the runtime scenario. We execute power simulations by propagating activities from either RTL simulation vectors or fully annotated gate-level vectors. From these power simulations, we obtain the cell internal power, switching power, cell leakage power, and total power used to determine the power and energy metrics under each scenario for the given workload.

We have developed a fully automated implementation of this flow using FuseSoC and Edalize to perform end-to-end studies. Our flow is highly configurable, parallelized and scalable

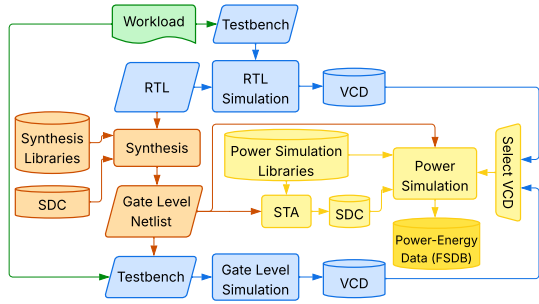


Fig. 1: Workload-based power-energy analysis methodology.

TABLE I: Characterization study workloads.

Workload	Classification	Instruction Count	Cycles
Fibonacci Series	Active	254	18,872
NOP	Inactive	50	73
Memory stalls (testbench generated)	Inactive	1	20
Wait For Interrupt	Idle	1	NA

to enable workload-based power-energy exploration on any design, with configurations for selecting RTL or gate-level simulation vectors, power analysis type, library and frequency selection for synthesis, and runtime operating points.

#### A. Workload Classification

We study the impact of workload variations on optimal DVDFS synthesis and runtime conditions by isolating and categorizing distinct workload phases based on their activity characteristics. By examining these phases individually, we assess their power and energy behaviors, and then aggregate them to derive generalized results for complex workloads under diverse execution scenarios.

1) *Activity-based classification of workloads*: Based on switching activity, workload phases are classified into three representative classes: active, inactive, and idle phases. Active workloads exhibit high switching activity. By contrast, inactive workloads have low switching activity while still maintaining clocked operation. Finally, idle workloads involve sleep modes with clock-gating. For our study we pick representative workloads for each class summarized in Table I.

2) *Scenario-based execution of workloads*: Workload phases with different activity levels as described in Section IV-A1 are combined to form different execution scenarios. These execution scenarios are classified into three types: run to completion, run forever, and run periodically. In the run to completion scenario, the workload runs once without immediate re-invocation upon completion, involving a fixed amount of work (such as a fixed number of CPU instructions), with execution time depending on operating frequency; the goal here is to minimize average energy consumption, as seen in tasks like data encryption, arithmetic computations, or machine learning inference. In the run forever scenario, the workload runs continuously, as in real-time monitoring, data logging, or communication systems, where the focus shifts to minimizing average power. Lastly, in the run periodically scenario, workloads alternate between active and idle phases, such

TABLE II: Library parameters.

Corner	Typical
Temperature	25°C
Supply Voltage ( $V_{DD}$ )	0.5–0.8 V in steps of 0.05 V
Standard Cell Types	SLVT & LVT
Body Bias ( $V_{NW}, V_{PW}$ )	See Table III for configurations

TABLE III: Available body bias configurations.

$V_{PW}(\downarrow)$ , $V_{NW}(\rightarrow)$	0V	0.25V	0.5V	0.75V	1V	1.25V	1.5V
0V	✓	✓					
-0.25V	✓	✓	✓				
-0.5V		✓	✓	✓			
-0.75V			✓	✓	✓		
-1V				✓	✓	✓	
-1.25V					✓	✓	✓
-1.5V						✓	✓

as periodic data logging or sensing and control applications, aiming to minimize average power.

#### B. Experimental Setup

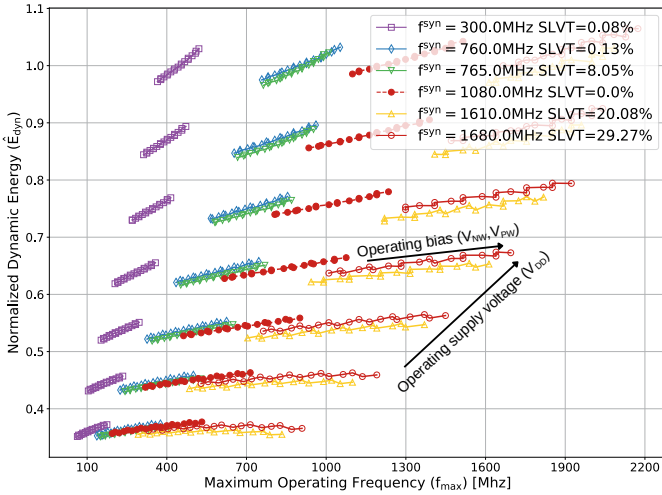
We utilize CV32E40P [28], a 4-stage pipelined embedded-class RISC-V CPU core implementing the RV32IM[F|Zfinx]C ISA alongside representative workloads shown in Table I. We compile the workloads with the standard RISC-V toolchain, perform RTL and gate-level simulations using Synopsys VCS, and RTL synthesis using Design Compiler. We use Synopsys Primetime and PrimePower for STA and power analysis. We use experimental libraries provided by GlobalFoundries described in Tables II and III for all our explorations.

### V. SYNTHESIS EXPLORATION

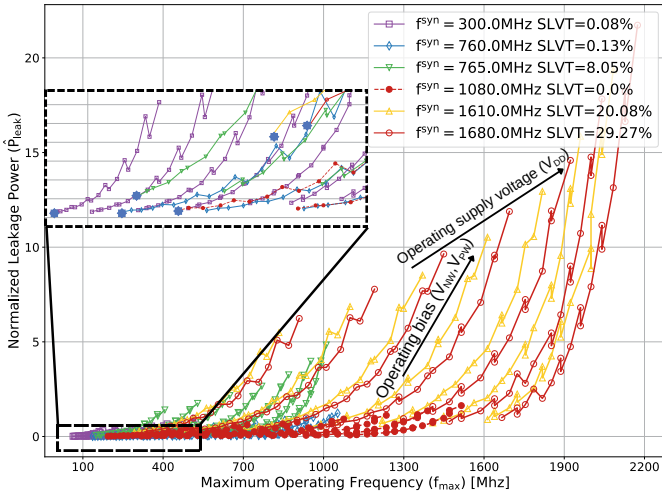
In this section, we first explore the impact of design-time parameters (synthesis frequency  $f^{syn}$ , supply voltage  $V_{DD}^{syn}$ , and body bias voltage  $V_{XW}^{syn}$  under different cell constraints) on  $E_{dyn}$  and  $P_{leak}$ . The study was conducted using the Fibonacci Series workload in a run-to-completion scenario.

1) *Sweep of  $f^{syn}$  constraint at fixed  $V_{DD}^{syn}$  and  $V_{XW}^{syn}$* : We begin by examining optimal  $f^{syn}$  across various  $V_{DD}$  and  $V_{XW}$ . Fig. 2a and 2b show the dynamic energy ( $\hat{E}_{dyn}$ ) and leakage power ( $\hat{P}_{leak}$ ) of netlists synthesized at different  $f^{syn}$  as a function of their  $f_{max}$ ,  $V_{DD}$  and  $V_{XW}$ , normalized to the  $E_{dyn}$  and  $P_{leak}$  of the netlist synthesized at maximum  $f^{syn}$  at its nominal ( $V_{XW} = 0V$ ,  $V_{DD} = 0.8V$ ) operating point. Each color/marker style represents a unique netlist, with lines indicating operation at specific  $V_{DD}$  and points along each line denoting operation at different  $V_{XW}$ , each with their supported  $f_{max}$  and associated  $\hat{E}_{dyn}/\hat{P}_{leak}$ . Solid lines with hollow markers represent netlists synthesized without cell composition constraints, while dashed lines with solid markers represent netlists containing only LVT (low threshold voltage) cells, where the legend indicates the actual SLVT (super-low threshold voltage) cell percentage.

From Eq. (9),  $f_{max}$  increases with  $V_{DD}$  and  $V_{XW}$  scaling, where the  $f_{max}$  range is determined by  $\alpha$  and  $\beta$  parameters. Netlists at higher  $f^{syn}$  with more SLVT cells generally achieve faster operation with a wider operating range. The proportion



(a) Dynamic energy vs. operating frequency.



(b) Leakage power vs. operating frequency.

Fig. 2: Dynamic energy and leakage power for netlists synthesized at different target frequencies.

of SLVT cells thereby grows with  $f^{syn}$ , surging sharply past a threshold frequency as determined by the synthesis tool.

For dynamic energy, as seen in Fig. 2a and following Eq. (1), at a fixed  $V_{DD}$ , increasing FBB ( $\uparrow V_{NW}, \downarrow V_{PW}$ ) lowers  $V_{TN}$  and  $|V_{TP}|$ , thereby enlarging the  $[V_{DD} - V_{TN} - |V_{TP}|]^3$  term in Eq. (4). Although the short-circuit duration  $\tau$  decreases, this reduction is outweighed by the stronger cubic dependence, such that  $E_{sc}$  and hence  $\hat{E}_{dyn}$  increase roughly linearly with bias. Conversely, at a constant bias ( $V_{NW}, V_{PW}$ ),  $E_{dyn} \propto V_{DD}^2$  in line with Eq. (2). Netlists synthesized at higher  $f^{syn}$  generally have somewhat higher  $\hat{E}_{dyn}$  compared to slower netlists. This follows an increase in the proportion of faster and smaller SLVT cells. At the same time, tightening frequency constraints also leads to more complex netlists with a larger total cell count and both higher  $\hat{E}_{dyn}$  and  $f_{max}$ . The higher  $f_{max}$  allows netlists with higher  $f^{syn}$  to operate at lower bias and thus lower  $\hat{E}_{dyn}$  to reach a given frequency.

As seen in Fig. 2b,  $\hat{P}_{leak}$  increases exponentially with rising

TABLE IV: Effect of  $V_{DD}^{syn}$  and  $V_{XW}^{syn}$  on cell composition, operating frequency ranges ( $f_{max}^{syn}, f_{min}^{syn}$ ), normalized dynamic energy at minimum and maximum operating points ( $\hat{E}_{dyn}^{min}, \hat{E}_{dyn}^{max}$ ) and normalized idle and maximum leakage power ( $\hat{P}_{leak}^{min}, \hat{P}_{leak}^{max}$ ).

$V_{DD}^{syn}$	$V_{NW}^{syn}, V_{PW}^{syn}$	SLVT	$f_{min}^{syn}$ [MHz]	$f_{max}^{syn}$ [MHz]	$\hat{E}_{dyn}^{min}$	$\hat{E}_{dyn}^{max}$	$\hat{P}_{leak}^{min}$	$\hat{P}_{leak}^{max}$
0.6V	0.0V, 0.0V	48.7%	505	2273	36.4%	108.5%	32.3%	2911.1%
0.6V	0.75V, -0.75V	42.4%	429	2222	36.3%	107.2%	36.3%	3235.5%
0.6V	1.5V, -1.5V	37.0%	395	2222	36.4%	108.2%	28.6%	2562.4%
0.7V	0.0V, 0.0V	39.6%	355	2222	36.3%	108.0%	31.6%	2825.6%
0.7V	0.75V, -0.75V	32.7%	345	2174	36.1%	106.5%	30.3%	2694.0%
0.7V	1.5V, -1.5V	32.4%	350	2174	36.1%	106.5%	27.5%	2445.1%
0.8V	0.0V, 0.0V	29.3%	331	2174	36.2%	106.5%	24.4%	2172.7%
0.8V	0.75V, -0.75V	28.8%	289	2128	35.8%	104.9%	25.5%	2272.4%
0.8V	1.5V, -1.5V	24.5%	295	2174	35.6%	103.6%	21.9%	1940.5%

$V_{XW}$  and linearly with  $V_{DD}$  in line with Eq. (8). The effect of  $V_{XW}$  is particularly pronounced in netlists with high leakage, i.e. those with a larger proportion of SLVT cells. As shown in Fig. 2b,  $\hat{P}_{leak}$  sharply increases beyond a  $f^{syn}$  of 760 MHz, where the proportion of SLVT cells increases significantly. This trend is also observed at idle operating points, with a substantial rise in idle  $\hat{P}_{leak}$  that generally follows SLVT percentages. Similar to  $\hat{E}_{dyn}$ , while netlists synthesized at higher  $f^{syn}$  have higher  $\hat{P}_{leak}$  for the same operating point, they also achieve higher  $f_{max}$ . This outweighs the impact of increased  $\hat{P}_{leak}$ , where shifting curves to the right results in lower  $\hat{P}_{leak}$  at any given  $f_{max}$ .

2) *Sweep of  $V_{XW}^{syn}$  and  $V_{DD}^{syn}$  at optimal  $f^{syn}$* : We perform a combined sweep of  $V_{XW}^{syn}$  and  $V_{DD}^{syn}$  for netlists synthesized at maximum  $f^{syn}$  with and without cell composition constraints. Table IV highlights results for netlists synthesized without cell constraints at zero, middle, and maximum biases. We exclude  $V_{DD}^{syn}$  below 0.6V since we observed no further changes. Results show that higher  $V_{DD}^{syn}$  or  $V_{XW}^{syn}$  reduces the proportion of faster SLVT cells. As such, although all netlists maintain a similar  $\Delta f_{max}$ , those synthesized at lowest  $V_{DD}^{syn}$  and  $V_{XW}^{syn}$  achieve the highest  $f_{max}$ . Synthesizing at lower  $V_{DD}^{syn}$  also results in a decrease in  $\hat{E}_{dyn}$  and an increase in  $\hat{P}_{leak}$  following SLVT cell composition. However, no clear impact of  $V_{XW}^{syn}$  on  $\hat{E}_{dyn}$  or  $\hat{P}_{leak}$  is observed. Likewise, among netlists synthesized exclusively with LVT cells, no significant trends in  $\hat{E}_{dyn}$  or  $\hat{P}_{leak}$  are observed with changes in  $V_{XW}^{syn}$  or  $V_{DD}^{syn}$ .

3) *Synthesis Takeaways*: Synthesizing with tightest frequency constraints that cover a desired operating range, and at lowest  $V_{DD}^{syn}$  and  $V_{XW}^{syn}$  yields the fastest netlists with the lowest  $E_{dyn}$  and  $P_{leak}$  for any given operating  $f$ , but also with the highest  $P_{leak}$  in idle conditions. Imposing constraints on SLVT cell composition can modulate  $P_{leak}$  at the expense of increased  $E_{dyn}$ , i.e. creates a tradeoff that depends on workload characteristics and scenario, where restricting cells to LVT entirely also removes any influence of  $V_{DD}^{syn}$  or  $V_{XW}^{syn}$  on energy and power.

## VI. RUNTIME EXPLORATION

This section explores DVBFs across different workload scenarios, and compares against a static and a traditional DVFS approach that only scales  $V_{DD}$  with  $f$ .

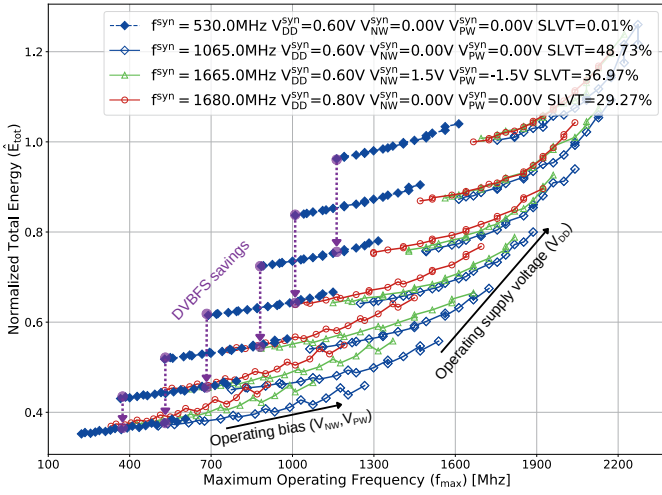


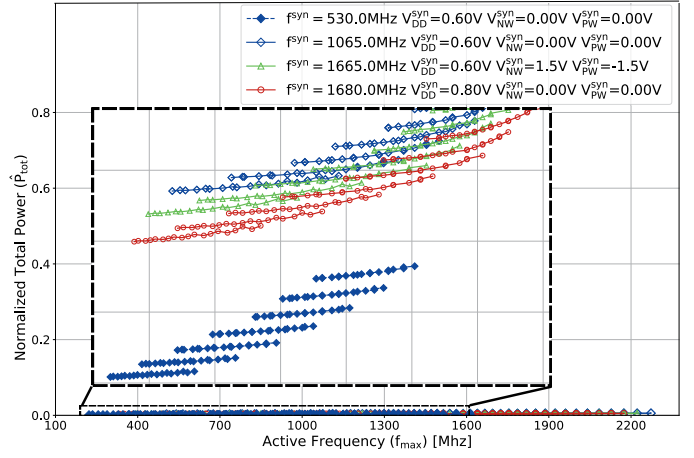
Fig. 3: Total energy vs. operating frequency for netlists synthesized at various design-time parameters, running Fibonacci series in run-to-completion scenario, for all runtime  $V_{DD}$  and  $V_{XW}$  conditions.

1) *Run to completion*: Fig. 3 illustrates the impact of runtime  $V_{DD}$  and  $V_{XW}$  scaling on the total energy  $E_{tot} = E_{dyn} + E_{leak} = E_{dyn} + P_{leak}N_{Cycles}/f_{max}$  of unconstrained and constrained netlists synthesized with optimal design-time parameters as well as on unconstrained netlists with one sub-optimal parameter, for the Fibonacci series workload in a run-to-completion scenario. Due to clock switching dominating and hence masking  $E_{dyn}$  reductions, energy trends for inactive workloads mirror those in Fig. 3 and are not reported here.

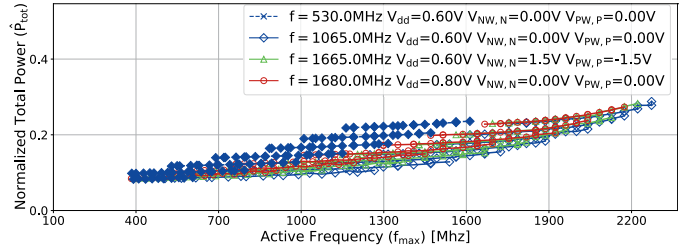
For a given netlist,  $E_{tot}$  increases with increasing  $V_{DD}$  and  $V_{XW}$  reflecting the effects of  $V_{DD}$  and  $V_{XW}$  scaling on  $E_{dyn}$  and  $E_{leak}$ . At higher  $V_{XW}$  and  $V_{DD}$ , leakage effects become more pronounced, leading to a steeper increase in  $E_{tot}$ . This is particularly pronounced for netlists with a higher proportion of SLVT cells. Without deadline constraints, the energy-minimal operating point is reached at the lowest operating frequency,  $V_{DD}$  and  $V_{XW}$ . Given a frequency (deadline) constraint, a DVBS policy that first applies FBB and only then raises  $V_{DD}$  yields lower  $E_{tot}$  than pure DVFS, because  $E_{sw} \propto V_{DD}^2$  dominates the comparatively smaller increases in  $E_{sc}$  and  $E_{leak}$  introduced by  $V_{XW}$ . Overall, since  $E_{dyn}$  dominates active operation of our design,  $E_{tot}$  of LVT-only netlists remains higher than that of unconstrained netlists. In particular, an unconstrained netlist synthesized at maximum  $f^{syn}$  with minimal  $V_{DD}^{syn}$  and  $V_{XW}^{syn}$  provides energy-minimal operating points. Similar trends were observed for  $E_{tot}$  in the run-forever scenario but omitted here for brevity.

2) *Periodic workloads*: Fig. 4 shows the total power ( $P_{tot}$ ) of various netlists for the Fibonacci workload running periodically with different periods. In this scenario, the workload alternates between an active phase and an idle phase. During the idle phase, the runtime  $V_{DD}$  and  $V_{XW}$  are set to their minimal values for optimal energy efficiency. We sweep runtime  $V_{DD}$  and  $V_{XW}$  in the active phase.

For longer periods with lower duty cycles (Fig. 4a), LVT-



(a) Period = 5 ms.



(b) Period = 0.05 ms.

Fig. 4: Total power vs. operating frequency of the optimal synthesized netlist running a periodic workload for all runtime  $V_{DD}$  and  $V_{XW}$  conditions.

only netlists synthesized at maximum  $f^{syn}$  with the lowest  $V_{DD}^{syn}$  and  $V_{XW}^{syn}$  are optimal, with DVBS operating points achieving on average 16.6% higher power efficiency than conventional DVFS. In contrast, for shorter periods with higher duty cycles (Fig. 4b), unconstrained netlists synthesized under the same conditions are optimal, where DVBS operating points deliver on average 22.5% higher power efficiency than DVFS. The power-minimal operating points occur at the lowest  $V_{DD}$  and  $V_{XW}$  that still satisfy the deadline constraint.

3) *Runtime takeaways*: At runtime, DVBS provides the power-optimal operating scheme. Active workloads and those with higher duty cycles, where  $E_{dyn}$  is critical favor faster netlists synthesized at lowest  $V_{DD}^{syn}$  and  $V_{XW}^{syn}$  without cell-composition constraints, while workloads with long idle periods and lower duty cycles, where leakage is significant benefit from constraining netlists to only use LVT cells.

## VII. EXPERIMENTS AND RESULTS

We evaluate our methodology using real-world applications from Biomedbench [29] and MLPerf Tiny [30] executed on the CV32E40P processor to highlight optimal strategies for both runtime and design-time configurations. We select four TinyML applications from Biomedbench for low-power wearables that exclusively use fixed-point operations, aligning with the configuration of the studied CPU core. Each Biomedbench workload consists of three phases—data acquisition, idle, and processing—where the acquisition and idle phases have fixed

TABLE V: Workload parameters.

(a) Biomedbench workloads (fixed idle period between runs)

Application	Cycles	Idle	Duty	Dominant Operations
Heartbeat Classifier	5.1M	15 s	Very Low	Branch-heavy (Fxp min/max search)
Seizure Detection—SVM	3.2M	60 s	Very Low	32-bit Fxp multiplications
Seizure Detection—CNN	160.0M	4 s	Medium	16-bit Fxp MAC
Cognitive Monitor	205.0M	56 s	Low	32-bit Fxp multiplications

(b) MLPerf Tiny workloads (fixed period/deadline)

Application	Cycles	Period	Duty	Dominant Operations
Visual Wake Words	124.6M	85 ms	High	8-bit Fxp MAC
Anomaly Detection	4.2M	7 ms	Medium	8-bit Fxp MAC
Keyword Spotting	92.3M	100 ms	High	8-bit Fxp MAC
Image Classification	339.2M	200 ms	High	8-bit Fxp MAC

durations, while the processing phase length scales with the active frequency. We impose an optional deadline on the processing phases based on its length at nominal operating conditions. We further incorporate four periodic MLPerf Tiny workloads characterized by fixed period lengths combining active and idle phases. Workload parameters and configurations are summarized in Table V. We perform power analysis across these workloads, applying our methodology to optimal netlists synthesized at maximum  $f^{syn}$  and minimum  $V_{DD}^{syn}$  and  $V_{XW}^{syn}$  in order to obtain total power metrics.

Fig. 5 summarizes results comparing normalized power of DVFS vs. DVBFS schemes across optimal LVT-only (striped) or unconstrained (solid) netlists for each application. Note that for Visual Wake Word and Image Classification, LVT-only netlists cannot satisfy the required deadlines and are therefore omitted. For Biomedbench applications, we compare against reference power when operating at lowest  $V_{DD}$  and  $V_{XW}$  without a processing deadline. Heartbeat Classifier, SVM Seizure Detection, and Cognitive Monitor have very low duty cycles and are therefore idle-dominated; the LVT-only netlist is most efficient and DVBFS provides modest additional savings of 1-11% over DVFS. By contrast, for CNN Seizure Detection, Anomaly Detection, Visual Wake Words, Keyword Spotting, and Image Classification, the higher duty cycle makes active energy a substantial portion of  $P_{tot}$ ; here, unconstrained netlists are optimal, and DVBFS achieves large gains of 23–29% by using  $V_{XW}$  first to satisfy deadlines before incurring the quadratic cost of additional  $V_{DD}$  scaling.

Both DVBFS and DVFS allow setting optimal operating points per workload and support active operation at different  $f_{max}$  depending on external constraints such as additional deadlines or overall system considerations. Table VI summarizes the power-optimal operating points for unconstrained and LVT-only netlists. Overall, a DVBFS approach provides more than 12% and up to 31% power gains over traditional DVFS approaches across a wide range of performance for both constrained and unconstrained netlists.

A dedicated back-bias generator (BBGEN) is required for applying forward or reverse body bias. The BBGEN from [31] occupies about 0.083mm<sup>2</sup> and draws 3.31μA when active

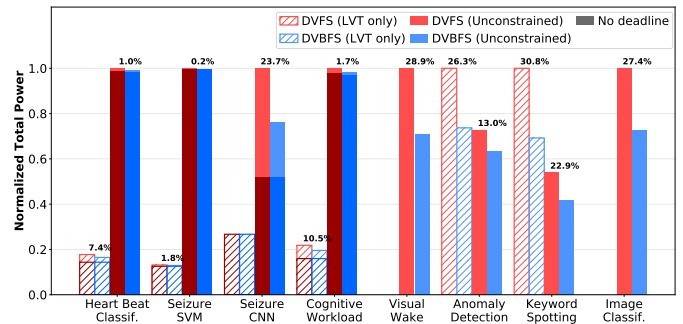


Fig. 5: DVFS vs. DVBFS power consumption.

TABLE VI: Optimal DVBFS vs. DVFS operating points.

$f_{max}$	DVFS $V_{DD}$	DVBFS		Power Gain
		$V_{DD}$	$V_{NW}, V_{PW}$	
<b>LVT-only Netlist</b> ( $V_{DD}^{syn} = 0.6V, V_{NW}^{syn} = V_{PW}^{syn} = 0V$ )				
366 MHz	0.55V	0.5V	(0.5V, -0.75V)	15.3%
529 MHz	0.6V	0.5V	(1.25V, -1.25V)	27.4%
690 MHz	0.65V	0.55V	(1.0V, -1.25V)	12.7%
885 MHz	0.7V	0.6V	(1.25V, -1.25V)	23.3%
1010 MHz	0.75V	0.65V	(1.0V, -1.25V)	13.0%
1163 MHz	0.8V	0.7V	(1.0V, -1.0V)	31.6%
<b>Unconstrained Netlist</b> ( $V_{DD}^{syn} = 0.6V, V_{NW}^{syn} = V_{PW}^{syn} = 0V$ )				
775 MHz	0.55V	0.5V	(0.75V, -0.5V)	12.9%
1064 MHz	0.6V	0.5V	(1.0V, -1.25V)	24.2%
1250 MHz	0.65V	0.5V	(1.5V, -1.5V)	25.8%
1492 MHz	0.7V	0.55V	(1.25V, -1.5V)	21.3%
1612 MHz	0.75V	0.6V	(1.25V, -1.25V)	28.9%
1754 MHz	0.8V	0.65V	(1.25V, -1.0V)	28.0%

and 87nA when idle at 0.8V, while a second-generation design shows further improvements [32]. These overheads remain small compared to a typical CV32E40P-based SoC, such as PULPissimo at 2.21mm<sup>2</sup> [33]. Furthermore, since our workloads operate on millisecond-level timescales, the 30μs startup and 10μs settling times reported in [31] have minimal impact on overall latency and energy.

## VIII. SUMMARY AND CONCLUSIONS

This work proposed a novel approach to power and energy optimization in FD-SOI based designs combining design-time synthesis parameter selection with dynamic, voltage, bias and frequency scaling (DVBFS). We developed a fully automated design space exploration framework to jointly determine optimal runtime operating points and design-time synthesis parameters. Using this framework, we provide guidance on design-time synthesis conditions to achieve optimal runtime operating points under different runtime scenarios, demonstrating that workload-aware joint optimization is necessary to unlock significant additional energy-saving opportunities from dynamic bias scaling. We evaluated our approach on an embedded RISC-V processor executing TinyML workloads, achieving up to 29% savings compared to DVFS. In future work, we plan to extend our approach to additional scenarios and platforms using improved exploration approaches.

## ACKNOWLEDGMENTS

This work was supported by a grant from GlobalFoundries.

## REFERENCES

- [1] R. Carter *et al.*, “22nm FDSOI technology for emerging mobile, Internet-of-Things, and RF applications,” in *International Electron Devices Meeting (IEDM)*, Dec. 2016.
- [2] P. Rajput, R. Srinivasan, H. Narisetty, and R. Trichy, “Hierarchical design exploration to find automatic body bias configuration to achieve optimal PPA for GlobalFoundries 22 FD-SOI technology,” Synopsys Users Group Conference, 2017.
- [3] P. Rajput, “GlobalFoundries 22FDX body bias design exploration with Genus for optimal power/performance tradeoff,” Cadence Users Conference, 2017.
- [4] S. Garg and D. Marculescu, “System-level leakage variability mitigation for MPSoC platforms using body-bias islands,” *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 20, no. 12, 2012.
- [5] S. Höppner *et al.*, “How to Achieve World-Leading Energy Efficiency using 22FDX with Adaptive Body Biasing on an Arm Cortex-M4 IoT SoC,” in *European Solid-State Device Research Conference (ESSDERC)*, 2019.
- [6] S. Clerc *et al.*, “A 0.33V/-40°C process/temperature closed-loop compensation SoC embedding all-digital clock multiplier and DC-DC converter exploiting FDSOI 28nm back-gate biasing,” in *International Solid-State Circuits Conference (ISSCC)*, 2015.
- [7] D. Rossi *et al.*, “A self-aware architecture for PVT compensation and power nap in near threshold processors,” *IEEE Design and Test*, vol. 34, no. 6, 2017.
- [8] A. Quelen, G. Pillonnet, P. Flatresse, and E. Beigné, “A 2.5 $\mu$ W 0.0067mm<sup>2</sup> automatic back-biasing compensation unit achieving 50% leakage reduction in FDSOI 28nm over 0.35-to-1V VDD range,” in *International Solid-State Circuits Conference (ISSCC)*, 2018.
- [9] F. Schraut, H. Eisenreich, S. Höppner, and C. Mayr, “A fast lock-in ultra low-voltage ADPLL Clock generator with adaptive body biasing in 22nm FDSOI technology,” in *International Symposium on Circuits and Systems (ISCAS)*, 2019.
- [10] C. Oh *et al.*, “Thermal-aware body bias modulation for high performance mobile core,” in *International SoC Design Conference (ISOCC)*, 2012.
- [11] D. Puschini, J. Rodas, E. Beigne, M. Altieri, and S. Lesecq, “Body bias usage in utbb fdsoi designs: A parametric exploration approach,” *Solid-State Electronics*, vol. 117, 2016.
- [12] D. Jacquet *et al.*, “A 3 GHz dual core processor ARM Cortex™-A9 in 28 nm UTBB FD-SOI CMOS with ultra-wide voltage range and energy efficiency optimization,” *IEEE Journal of Solid-State Circuits*, vol. 49, no. 4, 2014.
- [13] M. Weiner, M. Blagojevic, S. Skotnikov, A. Burg, P. Flatresse, and B. Nikolic, “A scalable 1.5-to-6Gb/s 6.2-to-38.1mW LDPC decoder for 60GHz wireless networks in 28nm UTBB FDSOI,” in *International Solid-State Circuits Conference (ISSCC)*, 2014.
- [14] R. Wilson *et al.*, “A 460MHz at 397mV, 2.6GHz at 1.3V, 32b VLIW DSP, embedding FMAX tracking,” in *International Solid-State Circuits Conference (ISSCC)*, 2014.
- [15] J. M. Kühn, H. Amano, O. Bringmann, and W. Rosenstiel, “Leveraging FDSOI through body bias domain partitioning and bias search,” in *Design Automation Conference (DAC)*, 2016.
- [16] J. Lee *et al.*, “A Self-Tuning IoT Processor Using Leakage-Ratio Measurement for Energy-Optimal Operation,” *IEEE Journal of Solid-State Circuits*, vol. 55, no. 1, pp. 87–97, 2020.
- [17] S. Martin, K. Flautner, T. Mudge, and D. Blaauw, “Combined dynamic voltage scaling and adaptive body biasing for lower power microprocessors under dynamic workloads,” in *International Conference on Computer Aided Design (ICCAD)*, 2002.
- [18] L. Yan, J. Luo, and N. Jha, “Joint dynamic voltage scaling and adaptive body biasing for heterogeneous distributed real-time embedded systems,” *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 24, no. 7, Jul. 2005.
- [19] J. Colinge, “Thin-film SOI devices: A perspective,” *Microelectronic Engineering*, vol. 8, no. 3, pp. 127–147, 1988.
- [20] S. Cristoloveanu, E. Nowak, J. Barbot, L. Grenouillet, and I. Radu, “Enhanced threshold voltage tuning in FD-SOI MOSFET with ferroelectric buried oxide,” *Solid-State Electronics*, vol. 224, p. 109052, 2025.
- [21] A. Ohata, R. Ritzenthaler, O. Faynot, and S. Cristoloveanu, “Special size effects in advanced single-gate and multiple-gate SOI transistors,” *Journal of Telecommunications and Information Technology (JTIT)*, vol. 28, pp. 14–24, 2007.
- [22] J. M. Rabaey, A. Chandrakasan, and B. Nikolić, *Digital Integrated Circuits: A Design Perspective*, 2nd ed. Prentice Hall, 2003, ch. 5.
- [23] H. Veendrick, “Short-circuit dissipation of static CMOS circuitry and its impact on the design of buffer circuits,” *IEEE Journal of Solid-State Circuits*, vol. 19, no. 4, pp. 468–473, 1984.
- [24] N. Hedenstierna and K. Jeppson, “CMOS Circuit Speed and Buffer Optimization,” *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 6, no. 2, pp. 270–281, 1987.
- [25] K. Nose and T. Sakurai, “Analysis and future trend of short-circuit power,” *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 19, no. 9, pp. 1023–1030, 2000.
- [26] S. Salahuddin and C. Hu, *BSIM4 4.8.3 MOSFET Model User’s Manual*, 2025.
- [27] T. Sakurai and A. Newton, “Alpha-power law MOSFET model and its applications to CMOS inverter delay and other formulas,” *IEEE Journal of Solid-State Circuits*, vol. 25, no. 2, pp. 584–594, 1990.
- [28] “CV32E40P: A 32-bit RISC-V CPU Core,” [https://github.com/openhwgroup/cv32e40p/releases/tag/cv32e40p\\_v1.2.1](https://github.com/openhwgroup/cv32e40p/releases/tag/cv32e40p_v1.2.1), 2022, Accessed: Apr. 4, 2024.
- [29] D. Samakovlis *et al.*, “BiomedBench: A benchmark suite of TinyML biomedical applications for low-power wearables,” *IEEE Design and Test*, pp. 1–1, 2024.
- [30] C. Banbury *et al.*, “MLPerf Tiny Benchmark,” *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, 2021.
- [31] A. Siddiqi, N. Jain, and M. Rashed, “Back-bias generator for post-fabrication threshold voltage tuning applications in 22nm FD-SOI process,” in *International Symposium on Quality Electronic Design (ISQED)*, 2018.
- [32] N. Jain *et al.*, “Ultra Low Leakage IoT Design in 22FDSOI,” in *IEEE Electron Devices Technology and Manufacturing Conference (EDTM)*, 2024.
- [33] P. D. Schiavone, D. Rossi, A. Pullini, A. Di Mauro, F. Conti, and L. Benini, “Quentin: an Ultra-Low-Power PULPissimo SoC in 22nm FDX,” in *IEEE SOI-3D-Subthreshold Microelectronics Technology Unified Conference (S3S)*, 2018.