

ニアスレッシュヨルド電圧動作に適した単一電源で動作する高歩留まりオンチップメモリの設計

塩見 準¹, 石原 亨¹, 小野寺 秀俊¹

¹ 京都大学大学院情報学研究科.

{shiomi-jun, ishihara, onodera}@vlsi.kuee.kyoto-u.ac.jp

概要

集積回路中のオンチップメモリは集積回路全体の消費エネルギーの中で大きな割合を占めており、低消費エネルギーを実現するオンチップメモリが求められている。従って、電源電圧を低くすることがオンチップメモリにも求められている。本稿では、ニアスレッシュヨルド領域で動作するエネルギー効率の高い高歩留まりオンチップメモリの設計・評価実験を行う。商用 28 nm プロセスでの比較結果、設計したメモリは従来の SRAM 構造より読み出し操作で約 40% 省エネルギーであることが示された。また、商用プロセスのモンテカルロシミュレーションを行った結果、設計した回路の 5σ 遅延が SRAM 構造より小さいことがわかった。

High Yield On-Chip Memory Design for Single Supply Near-Threshold Computing

Jun Shiomi¹, Tohru Ishihara¹, Hidetoshi Onodera¹

¹ Graduate School of Informatics, Kyoto University.

Abstract

On-chip memory is one of the most energy consuming components in today's LSI. Aggressive voltage scaling is thus applied to the memory to obtain a quadratic reduction of dynamic energy consumption, which drastically degrades the memory yields. This paper discusses a design methodology for a high yield on-chip memory which uses a single near-threshold supply voltage. This memory has an energy-efficient readout structure in near-threshold operation. Circuit simulation using a commercial 28 nm process technology shows that the energy consumed in our memory for readout operation is about 40% more energy efficient than that of a conventional SRAM readout structure. Circuit simulation using a foundry provided Monte Carlo simulation package also shows that the 5σ worst case read-access time of our originally designed memory is smaller than that of a conventional SRAM circuit.

1 序論

近年、IoT (Internet of Things) などの応用分野では、高性能で消費エネルギーの小さいオンチップメモリが求められている。集積回路の動的エネルギーが電源電圧の 2 乗に比例することから、電源電圧を下げることは低消費エネルギーを実現する有効な方法である。伝統的に電源電圧をしきい値電圧まで下げる手法 (サブスレッシュヨルド動作) が用いられてきた。しかし、サブスレッシュヨルド領域ではトランジスタの特性ばらつきによる回路特性の悪化が顕著になり、回路の歩留まり確保が大きな課題となっている。

オンチップメモリとして代表的に用いられる SRAM (Static Random Access Memory) は回路の特性悪化に脆弱な構造であり、集積回路のロジック部分と比較して極低電圧動作が困難である。SRAM

を低電圧で動作させるために、様々な手法が取られている [1]。多くの手法では、複数の電源電圧を用いる構成がとられている。これは、設計コストの増大や、DC/DC コンバータの変圧によるエネルギー消費のオーバーヘッドが問題である。

メモリの低電圧動作時の問題点の別の解決方法として、スタンダードセルのみで構成されたメモリが提案されている [2, 3, 4]。この回路構成はデジタル回路のみで構成されているため、集積回路のロジック部分と同じ電源電圧までメモリの電源電圧を下げるができる。

本稿では、読み出し操作で高いエネルギー効率を実現するオンチップメモリの設計手法について述べる。設計を行うメモリ構造は、デジタル回路のマルチプレクサ (Multiplexer: MUX) とラッチで構成された回路である。極低電圧で動作するだけでなく、

読み出し操作の際に必要な最小限の回路のみ動作させることで高いエネルギー効率を実現する。一方、SRAM は大きな容量を持ったビット線を毎読み出しサイクル充放電し、エネルギー消費が大きい。本稿で設計するデジタル回路で構成されたメモリ構造と極低電圧動作するよう設計する SRAM の読み出し構造を比較し、設計するメモリ構造が高いエネルギー効率を達成していることを示す。

サブスレッショルド領域はエネルギー効率が良いと考えられているが、本稿では電源電圧をしきい値電圧よりやや高い電圧（ニアスレッショルド電圧）でエネルギー効率が最大になることについても述べ、ニアスレッショルド領域で動作するオンチップメモリ設計方法について述べる。

本稿の章構成を示す。2 章で関連研究を示し、本研究の新規性について述べる。3 章で低電圧動作するオンチップメモリについて述べ、ニアスレッショルド領域でエネルギー効率が最大になることを述べる。4 章でニアスレッショルド領域で動作するオンチップメモリの設計について述べる。5 章で設計する回路の読み出しエネルギーの評価実験を行う。6 章で本稿の結論を述べる。

2 関連研究

オンチップメモリとして代表的に用いられている SRAM を低電圧で安定して動作させるよう長年努力が続けられてきた。低電圧動作では、1 ビットの情報を記憶するビットセルの安定性を確保することが大きな課題の 1 つとなっている。ビットセルに入力する制御回路の電源電圧を変更することで、ビットセルの安定性を補償することができる [1]。しかし、この手法は複数の電源電圧を必要とするため、設計コストの増大や、DC/DC コンバータを導入することによるコストの増大をもたらす。ビットセルを構成するトランジスタのゲートサイズを拡大することで、トランジスタの特性ばらつきをおさえることができる。従って、トランジスタのビットセルを大きくすることで、低電圧動作での安定性を実現することができる [5]。しかし、低電圧動作させるために必要な面積オーバーヘッドが大きくなり、消費電力の増大、面積効率の低下が問題となる。ビットセルを構成するトランジスタ数に冗長性を与えることでビットセルの安定性を確保することができる [6, 7, 8, 9]。しかし、いずれの手法を用いても、読み出しサイクル毎に、小さなビットセルが大きな容量を持つビット線の電荷を放電し、センスアンプがビット線の電位の変動を判定する構造は変化しない。

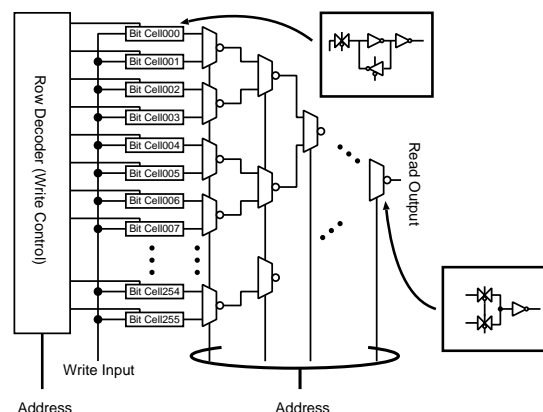


図 1: 文献 [2] で提案されているメモリ回路。毎読み出しサイクルですべての MUX を稼働するため、消費エネルギーが大きくなる。

サブスレッショルド領域で動作するオンチップメモリとして、スタンダードセルを用いたメモリ構造が古くから提案されている [2, 3, 4, 10]。プロセッサのロジック部分と同じスタンダードセルのみで構成されているため、メモリの電源電圧をロジック部分の電圧まで下げることができる。構造的に最適化されている SRAM と比較して、スタンダードセルを用いたメモリ構造は面積オーバーヘッドが大きいことが問題点であるが、最小限の設計コストでメモリを集積できることが利点である。文献 [2] では、ラッチと木構造に接続された 2 入力マルチプレクサで構成されたメモリ構造が提案されている。この構造は低電圧でも安定して動作を行うことができる。本稿では、消費エネルギーが小さいデジタル回路構成メモリを提案し、設計結果を報告する。提案するメモリは、文献 [2] で用いられた構造より 2 倍以上エネルギー効率が良い。

3 低電圧動作するオンチップメモリ構造

3.1 完全デジタル型メモリ

スタンダードセルのみで構成されたオンチップメモリの構造として、MUX とラッチのみで構成されたメモリ構造が提案されている [2]。図 1 に回路図を示す。本稿ではこの回路を完全デジタル型メモリ (Fully Digital Memory: FDM) と呼ぶ。図 1 で示す回路は、読み出し回路を構成するマルチプレクサが毎読み出しサイクルですべて稼働する構成になっている。低消費エネルギーを実現する観点から、読み出し信号経路上のみのマルチプレクサを稼働する回路構造が要求される。図 2 に読み出し操作で信号

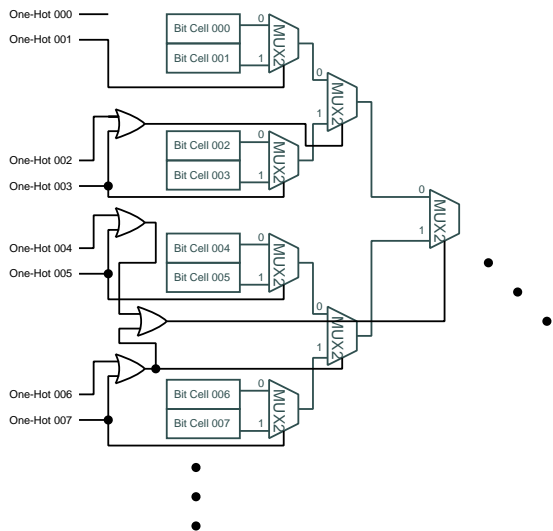


図 2: エネルギーが消費される回路部分を局所化することを目的に作り替えた回路構造. エネルギーが消費されるゲートは信号経路上に制限される.

経路上のゲートのみが稼働する構造を示す. 書き込み操作のために必要であるアドレスデコーダの出力信号を利用している. エネルギーが消費されるゲートを読み出し信号経路上のみのゲートに制限することができる. アドレスデコーダの出力信号を利用することにより, メモリ設計に使用するゲート数が多くなるが, 面積オーバーヘッドは生じない. 例えば 256 ワードのメモリを設計する場合, 図 2 で示すメモリは図 1 で示す回路より 231 個多くの NOR ゲートと 231 個多くの INV ゲートが必要になる. これは FDM 全体の中の約 7% の面積に対応する. 一方, 図 1 では, 多くの MUX を一度に動作させなければならないため, 大きなドライバが必要になり, ドライバの面積が FDM 全体の中の約 11% に対応する. 従って, 提案する回路構造は図 1 の回路と比較して面積オーバーヘッドなく消費エネルギーを小さくすることができる.

3.2 ニアスレッシュヨルド領域でのエネルギー効率

サブスレッシュヨルド領域では電源電圧が低いため, 容量の充放電に起因する動的電力は小さくなり, 動作周期あたりの消費エネルギーが非常に小さくなることが予想される. しかし, 指数関数的に遅延が増大するため, 漏れ電流による動作周期あたりの消費エネルギーも増大する. 従って, 電源電圧 V_{DD} を変化させた時, 回路の動作周期あたりの消費エネルギーはある点 V_{DDmin} で最小値をとる [11]. 一般的なデジタル回路での活性化率 0.1 を仮定すると,

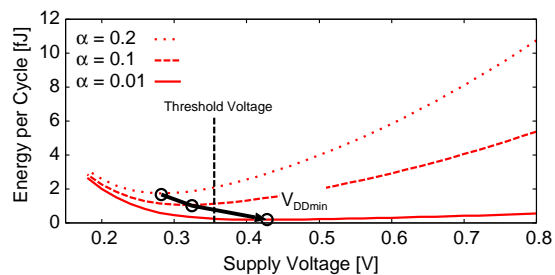


図 3: 53 段インバータリングオシレータの動作周期あたりの消費エネルギーと電源電圧の関係

V_{DDmin} はサブスレッシュヨルド領域に存在することが知られている. ここで, 活性化率とは, 回路の電位が 0 から 1 に上がる確率である [12].

しかし, 回路の活性化率が低くなると, V_{DDmin} はニアスレッシュヨルド領域に存在する. 図 3 は 28 nm プロセスにおいて, 53 段のインバータリングオシレータを動作させたときの動作周期あたりの消費エネルギーを示している. 横軸は電源電圧 V_{DD} , 縦軸は動作周期あたりの消費エネルギーである. このプロセスのしきい値電圧は 0.35 V 程度である. 活性化率が高い 0.2 や 0.1 のときは V_{DDmin} はサブスレッシュヨルド領域に存在するが, 活性化率が 0.01 のとき, 消費エネルギー最小点はニアスレッシュヨルド領域に位置することがわかる. FDM のような, 稼働する回路が局所化された回路構成では, 活性化率が小さく, V_{DDmin} がニアスレッシュヨルド領域に存在すると予想される.

4 ニアスレッシュヨルド領域で動作するオンチップメモリの設計

本章では, ニアスレッシュヨルド領域で FDM が動作するよう設計を行う.

動作周期あたりのエネルギーが最小化されると考えられるニアスレッシュヨルド領域で, メモリが動作するよう設計を行う. 回路の特性ばらつきとして, トランジスタの大きさに応じてトランジスタのしきい値電圧のばらつき量に変化する Pelgrom モデルを用いる [13]. 1 KB のメモリが 3σ の歩留まり (約 1.5×10^{-3} の故障率) で動作するよう各ビットセルに対して 5σ の歩留まり (約 2.9×10^{-7} の故障率) で正しく動作できるよう設計を行う. ただし, 5σ の歩留まりを見積もるために必要な計算時間は非常に大きいので, 以下に示す方法で 5σ の歩留まりを満たしていることを確認する.

まず, ビットセルの安定性を考慮した設計を行う. 図 4(d) が FDM のビットセルである. このビットセルが正しく動作しない原因を分類する. 図 4(a) は,

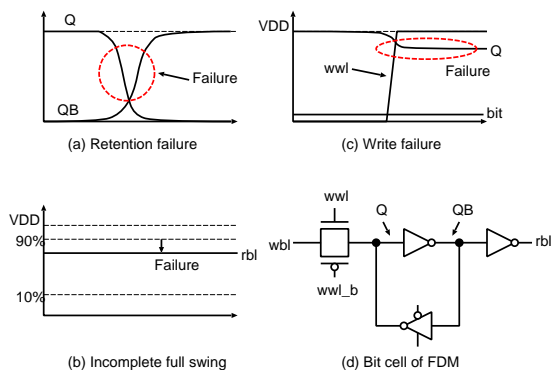


図 4: FDM のビットセルの動作不能条件の分類

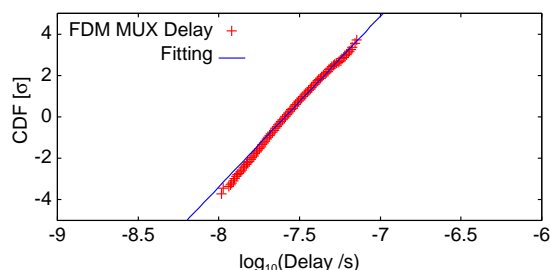


図 5: FDM のマルチプレクサ部分の読み出し遅延

ビットセルが値を維持できないことが原因、図 4(b) は、ビットセルに値を正しく書き込めないことが原因、図 4(c) は、ビットセル出力部分のインバータの出力がフルスイングしない（本稿では 90% を条件とする）ことが原因である。本稿では、電源電圧を確率変数とし、各原因による失敗確率が正規分布に従っていると仮定し、フィッティングを行い、 5σ の歩留まりを予測した。消費エネルギー、面積の観点でビットセルを構成するトランジスタはできる限り小さいことが望ましい。すべてのトランジスタが動作不良の原因であることから、すべてのトランジスタを等しく大きくしていき、上述の原因による動作不良が 5σ の歩留まりを満たすトランジスタサイズを探索する。

FDM の読み出し遅延は図 1 のマルチプレクサ部分が支配的である。ニアスレッシュド領域ではしきい値電圧ばらつきに起因する論理ゲートの遅延のばらつきが対数正規分布で近似できることが知られている [14]。従って、FDM の読み出し遅延が対数正規分布に従うと仮定し、得られたデータからフィッティングを行い、 5σ 歩留まりを満たす遅延を見積る。図 5 に見積り結果を示す。横軸は遅延の常用対数、縦軸は累積密度関数 (Cumulative Density Function: CDF) である。マルチプレクサは面積の観点から最小サイズにしている。

5 メモリの比較実験

本章では、設計を行った FDM と、比較対象のため、文献 [2] に紹介されている FDM と SRAM の読み出しエネルギーを商用 28 nm プロセストランジスタモデルを用いて評価する。

5.1 シミュレーションセットアップ

評価項目のメモリとして、以下のメモリを評価する。

1. ビット線を使用する従来型 SRAM (図 8)
2. 文献 [2] で提案されている MUX とラッチを用いたメモリ
3. 消費エネルギーを小さくした FDM (図 2)

評価するプロセスは商用 28 nm プロセス、電源電圧はニアスレッシュド領域として 0.4 V、(しきい値電圧は 0.35 V 程度)、想定するメモリ容量は 1 KB とし、それぞれのメモリの読み出しエネルギーを評価する。各ビットセルはランダムに 0/1 を記憶しており、トランジスタに特性ばらつきを与えない条件下で、読み出し操作で生じるエネルギーの平均値を評価する。動作速度は、各ビットセルに対し 5σ の歩留まりを保證する遅延から決定を行った。ただし、アドレスデコーダ回路を構成するトランジスタのサイズは十分大きく、遅延がばらつかないと仮定し、この部分には一定の遅延を与えている。

FDM の比較対象として、文献 [6] でニアスレッシュド動作のために用いられている 10T SRAM のビットセルから、図 8 の破線で示す部分に 2 個のトランジスタを追加した 12T SRAM 構造を使用する。10T SRAM は低電圧動作の安定性のため、ラッチ構造のビットセルになっている。この構造に、差動センス構造を用いるため、2 個のトランジスタを追加している。大きな容量が付加されたビット線を充放電し、センスアンプで電位差を増幅する構造は SRAM で共通のため、この安定したビットセル構造を用いる。

まず、読み出し速度の評価を行う FDM のマルチプレクサ部分の読み出し遅延と、SRAM のビット線の電位がセンスアンプで判定可能になるために必要な時間の CDF の関係を図 7 に示す。トランジスタに特性ばらつきを与えない場合に対応する 0σ の遅延は SRAM の方が FDM より高速であるが、 5σ の最悪遅延では、FDM の方が遅延が小さい結果になった。これは、以下に示す理由が考えられる。図

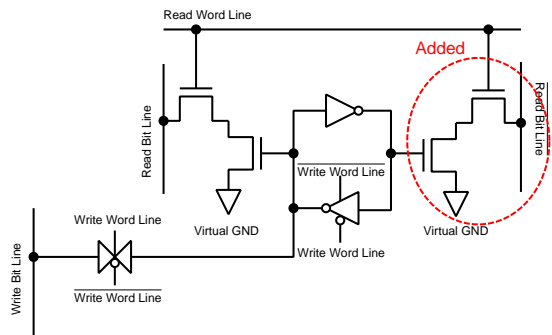


図 6: 12T ビットセル構造

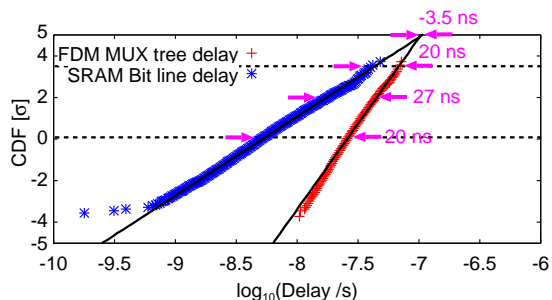


図 7: SRAM の読み出し速度と FDM 読み出し速度の違い

8(a) の SRAM 読み出し構造で示すように, SRAM では単体のビットセルが大きな容量を持ったビット線の電荷を放電するため, ビットセルの特性ばらつきがビットセルの容量で拡大されてしまう. 一方, 図 8(b) で示す FDM の読み出し構造のように小さな出力容量を持った MUX が多段に接続された構造では, それぞれの MUX の特性ばらつきが足し合わさって遅延が決定されるため, 図 8(a) に示す構造より, 遅延のばらつきが小さい. したがって, 高い性能歩留まりでは, 遅延の逆転が発生したと考えられる. メモリの動作速度が異なる値になったため, 動作速度が一致するように, それぞれのメモリでトランジスタのしきい値電圧を調整する.

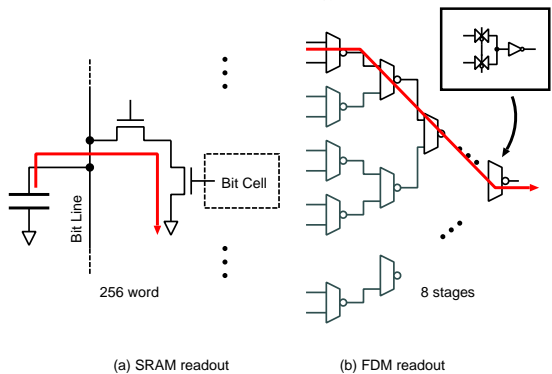


図 8: SRAM の読み出し構造と FDM の読み出し構造の違い

表 1: メモリ評価結果

| メモリ | ビット線または MUX でのエネルギー |
|-----------------|---------------------|
| 1. SRAM | 116 fJ |
| 2. 文献 [2] の FDM | 201 fJ |
| 3. 提案した FDM | 16 fJ |
| メモリ | 合計消費エネルギー |
| 1. SRAM | 168 fJ |
| 2. 文献 [2] の FDM | 374 fJ |
| 3. 提案した FDM | 98 fJ |

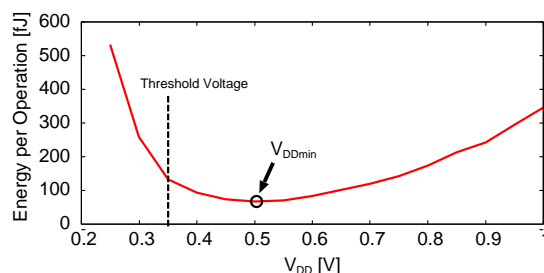


図 9: 電源電圧を変化させた際の FDM (メモリ 3) の動作周期あたりの消費エネルギー

評価結果を表 1 に示す. 文献 [2] の FDM (メモリ 2) は設計した FDM (メモリ 3) より 2 倍以上エネルギー効率が良いことがわかった. SRAM 構造 (メモリ 1) と比較して, 設計した FDM (メモリ 3) は 42%消費エネルギーが小さいことがわかった. 以下に示す原因が考えられる. 図 8(a) は一般的な SRAM の読み出し構造である. ビットセルが読み出し操作の際に, ビット線に付加された大きな容量を毎読み出しサイクル読み出すことがわかる. 図 8(b) は設計した FDM (メモリ 3) の読み出し構造である. 小さな出力容量が付加された MUX を局所的に稼働する構成のため, 消費エネルギーが小さい. 一方, 文献 [2] の FDM (メモリ 2) は毎読み出しサイクル, すべての MUX を稼働するため, SRAM より大きなエネルギーが消費されていることがわかる.

今回設計した FDM がニアスレッシュホールド領域でエネルギー効率が良くなることを述べる. 図 9 は, 電源電圧を変化させながら FDM の動作周期あたりの消費エネルギーをプロットしたグラフである. グラフの最小点がしきい値電圧 (約 0.35 V) より大きな点に位置していることから, FDM のように, 活性化率が低いメモリ構造の V_{DDmin} はサブスレッシュホールド領域ではなく, ニアスレッシュホールド領域側に存在することが実験的にわかる.

スタンダードセルベースのメモリ構造は, 低電圧で安定して動作するが, 従来型の SRAM と比べて

面積効率が悪いことが欠点の1つである。文献 [10] では、複数のメモリに対して面積オーバーヘッドの見積りが行われている。文献中では、商用 40 nm プロセスのメモリと比較して、スタンダードセルベースのメモリは 480%から 1800%の面積オーバーヘッドであることが示されている。今回設計した FDM はスタンダードセルベースではなく、セルを独自に最適化して配置できる構造であるため、上述の面積オーバーヘッドより小さい値でメモリを実装することができると思われる。

6 結論

本稿では、ニアスレッショルド領域で高いエネルギー効率を実現するメモリ構造である完全デジタル型メモリ (FDM) の設計と、読み出しエネルギーの評価を行った。FDM は 3σ の歩留まりを確保するように設計された。FDM は、読み出しに最小限必要な MUX のみを稼働するため、従来型の SRAM と比較して 42%消費エネルギーが小さいことがわかった。また、高い性能歩留まりを考えると、MUX を多段に接続した読み出し構造の読み出し遅延が、従来型の SRAM 読み出し構造の読み出し遅延より小さくなることがわかった。

従来型 SRAM と比較した時の FDM の面積オーバーヘッドの正確な見積もりと、今回の議論で得られた FDM の特性から、FDM の適切な使用ケースを提案することが今後の課題である。

謝辞

本研究は科研費 (B-25280014 および B-26280013 および B-26540021) による支援によっておこなわれた。また設計実験は、東京大学大規模集積システム設計教育研究センターを通し、シノプシス株式会社、日本ケイデンス株式会社、メンター株式会社の協力で行われた。

参考文献

- [1] M. Qazi, M. Sinangil, and A. Chandrakasan, "Challenges and Directions for Low-Voltage SRAM," *IEEE Design and Test of Computers*, vol. 28, no. 1, pp. 32–43, Jan. 2011.
- [2] A. Wang and A. Chandrakasan, "A 180-mV sub-threshold FFT processor using a minimum energy design methodology," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 1, pp. 310–319, Jan. 2005.
- [3] O. Akgun and V. Öwall, "A < 1 pJ sub-VT Cardiac Event Detector in 65 nm LL-HVT CMOS," in *Proc. of VLSI System on Chip Conference (VLSI-SoC)*, Sept. 2010, pp. 27–29.
- [4] P. Meinerzhagen, S. M. Y. Sherazi, A. Burg, and J. N. Rodrigues, "Benchmarking of Standard-Cell Based Memories in the Sub-VT Domain in 65-nm CMOS Technology," *IEEE Trans on Emerging and Selected Topics in Circuits and Systems*, vol. 1, no. 2, pp. 173–182, June 2011.
- [5] G. Chen, D. Blaauw, T. Mudge, D. Sylvester, and N. S. Kim, "Yield-driven near-threshold SRAM design," in *IEEE/ACM International Conference on Computer-Aided Design*, Nov 2007, pp. 660–666.
- [6] S. Jain, S. Khare, S. Yada, V. Ambili, P. Sali-hundam, S. Ramani, S. Muthukumar, M. Srinivasan, A. Kumar, S. Gb, R. Ramanarayanan, V. Erraguntla, J. Howard, S. Vangal, S. Dighe, G. Ruhl, P. Aseron, H. Wilson, N. Borkar, V. De, and S. Borkar, "A 280mV-to-1.2V wide-operating-range IA-32 processor in 32nm CMOS," in *IEEE International Solid-State Circuits Conference*, Feb. 2012, pp. 66–68.
- [7] K. Takeda, Y. Hagihara, Y. Aimoto, M. Nomura, Y. Nakazawa, T. Ishii, and H. Kobatake, "A read-static-noise-margin-free SRAM cell for low-VDD and high-speed applications," *IEEE Journal of Solid-State Circuits*, vol. 41, no. 1, pp. 113–121, Jan. 2006.
- [8] L. Chang, D. Fried, J. Hergenrother, J. Sleight, R. Dennard, R. Montoye, L. Sekaric, S. McNab, A. Topol, C. Adams, K. Guarini, and W. Haensch, "Stable SRAM cell design for the 32 nm node and beyond," in *Symposium on VLSI Technology*, June. 2005, pp. 128–129.
- [9] I.-J. Chang, J.-J. Kim, S. P. Park, and K. Roy, "A 32 kb 10T Sub-Threshold SRAM Array With Bit-Interleaving and Differential Read Scheme in 90 nm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 2, pp. 650–658, Feb. 2009.
- [10] T. Gemmeke, M. Sabry, J. Stuijt, P. Raghavan, F. Cattoor, and D. Atienza, "Resolving the memory bottleneck for single supply near-threshold computing," in *Design, Automation and Test in Europe Conference and Exhibition (DATE), 2014*, March 2014, pp. 1–6.
- [11] B. Zhai, D. Blaauw, D. Sylvester, and K. Flautner, "Theoretical and practical limits of dynamic voltage scaling," in *Design Automation Conference*, July 2004, pp. 868–873.
- [12] N. Weste and D. Harris, *CMOS VLSI Design: A Circuits and Systems Perspective*, 4th ed. USA: Addison-Wesley Publishing Company, 2010.
- [13] M. Pelgrom, A. C. J. Duinmaijer, and A. Welbers, "Matching properties of MOS transistors," *IEEE Journal of Solid-State Circuits*, vol. 24, no. 5, pp. 1433–1439, Oct 1989.
- [14] S. Keller, D. Harris, and A. Martin, "A Compact Transregional Model for Digital CMOS Circuits Operating Near Threshold," *IEEE Transactions on Very Large Scale Integration Systems*, vol. PP, no. 99, p. 1, 2013.