Statistical Analysis and Comparison of 2T and 3T1D e-DRAM Minimum Energy Operation

Manish Rana, Ramon Canal
Department of Computer Architecture
Universitat Politecnica de Catalunya, Barcelona
Email: mrana@ac.upc.edu, rcanal@ac.upc.edu

Esteve Amat, Antonio Rubio
Department of Electrical Engineering
Universitat Politecnica de Catalunya, Barcelona
Email: esteve.amat@upc.edu, antonio.rubio@upc.edu

Abstract—Bio-medical wearable devices restricted to their small-capacity embedded-battery require energy-efficiency of the highest order. However, minimum-energy point (MEP) at sub-threshold voltages is unattainable with SRAM memory, which fails to hold below 0.3V because of its vanishing noise margins. This paper examines minimum-energy operation of 2T and 3T1D e-DRAM gain-cells as an alternative to SRAM at 32nm technology node with different design points: up-sizing transistors, using high-Vth transistors, read/write wordline assists and temperature. First, the e-DRAM cells are evaluated without considering any process variations. The design-space is explored by creating a kriging meta-model to reduce the number of simulations. Finally, a full-factorial statistical analysis of e-DRAM cells is performed in presence of threshold voltage variations. The effect on mean MEP is also reported.

I. INTRODUCTION

The emergence of Internet-of-Things (IOT) has opened up new opportunities to collect data for analysis in the cloud using wireless battery-operated wearable sensors. The number of these devices is expected to increase to 35 sextillion units in 2020 [1] finding use cases in many domains which were till yet silicon-free. Achieving a smaller form factor and higher energy-efficiency is of prime importance in a bio-medical wearable devices. Recently, embedded-DRAM (e-DRAM) caches have been advocated as the successors of SRAM [2]–[6] considering their higher densities (> 2X) [7] and smaller leakage, due to fewer number of transistor. 3T1D e-DRAM gain-cell is shown to be capable of achieving access speeds comparable to 6T SRAM [6] and with larger device density [3]. The maximum energy efficiency has been shown to exist at sub-threshold circuit operation [8], [9]. However, the 6-Transistor SRAM bit-cell cannot provide enough reliability because of its reduced noise margin at these ultra-low voltages. Operating e-DRAMs at sub-threshold/near-threshold region offers the next step in the direction of increasing energy-efficiency of wearable biomedical health-monitoring systems. This simulation-based exploratory paper makes the following contributions:

1) Comparison of the read energy at MEP considering up-sizing of transistors, word-line boosting, high-threshold-voltage transistors and temperature using kriging based regression modelling.

2) Statistical analysis of read energy at MEP in presence of threshold voltage variations.

II. BACKGROUND

The energy consumption in CMOS circuits is mainly constituted of the dynamic energy and leakage energy. The former is spent in switching capacitive loads and the later is consumed by sub-threshold leakage currents when the transistors are off. Dynamic energy of the circuit can be decreased quadratically by scaling supply voltage ($V_{DD}$). When the $V_{DD}$ is aggressively scaled down to sub-threshold voltages, the driving-current ($I_{on}$, $V_{GS} = V_{DD}$) and the off-current ($I_{off}$, $V_{GS} = 0$) are given by the equation,

$$I_{SUB} = I_{e}e^{\frac{V_{GS}}{nV_{T}}}$$

The delay ($t_d$) of the circuit increases exponentially when the supply voltage is scaled to sub-threshold region thereby increasing the leakage energy per operation of the circuit. The MEP of the circuit can be achieved at $V_{DD}$ in the sub-threshold region [8], [9]. However, the operating voltages for a processor are limited to the minimum-voltage required for the reliable operation of on-chip SRAM cache which fails when scaling down to ultra-low voltages because of its shrinking noise margins, Fig.1(a). Nevertheless, SRAM dominates the energy consumption among the components of a processor [10], (Fig.1(b)) and several alternative SRAM bit-cells have been proposed. These sub-threshold SRAM bit-cells have 8-transistors [11], 10-transistors [12]–[14] or more.

As an alternative to SRAM bit-cells, Meinerzhagen et.al. [15] investigated sub-threshold 2T e-DRAM gain-cells for ultra-low power medical applications. Their study showed reliable operation for 2kb e-DRAM array up to sub-threshold voltage of 0.4V at mature 0.18$\mu$m node and up to near-threshold voltage of 0.6V at scaled 40nm node. The gain cells 2T and 3T1D are fully compatible with the standard CMOS technology and do not need additional process steps to fabricate the cell capacitor such as in the case of 1T1C eDRAM cell. These gain cells being smaller than the SRAM bitcells, thus have promising potential to improve the energy efficiency and reduce the silicon cost. Further, Amat et.al. [2] observed that the 3T1D gain-cells exhibits better reliability in front of device variability and single event upsets than the 2T gain cell.

A. 2T and 3T1D gain cells

2T and 3T1D gain cells are two-port memories with separate read and write paths as shown in Fig.2, which also shows the waveforms for their read/write operation. Since the leakage current of the nMOS...
III. METHODOLOGY

We study the energy-efficiency of 2T and 3T1D e-DRAM gain-cell within the following design space:

1) Different sizing of transistors: Nominal transistor sizes are taken from Lovin et al. [17]. The lengths and widths are increased in the range [1X, 2X, 3X, 4X] for each one of the e-DRAM cell transistors.

2) Wordline assist: A voltage offset in the range [0 to 0.2V] is applied to $WL_{Read}$ and $WL_{Write}$.

$\Delta_{wl}$ During a read operation, over-drive the $WL_{Read}$ for the 3T1D and under-drive the $WL_{Read}$ for the 2T. The effect is a faster read access and reduction in the read leakage energy. During standby (retention), under-drive the $WL_{Read}$ for the 3T1D and over-drive the $WL_{Read}$ for the 2T. The effect is a decrease in sub-threshold leakage through the read path.

$\Delta_{wuw}$ During read and standby (retention), over-drive $WL_{Write}$ to decrease the sub-threshold leakage through the write path.

3) High threshold voltage transistors: High threshold voltage transistors with $\Delta V_{th}$ in the range [0 to 0.2V].

4) Temperature: The operating temperature is varied from $-70^\circ C$ to $100^\circ C$.

These e-DRAM gain-cell designs are compared under the following metrics:

- Minimum-energy point (MEP): The dynamic and leakage energies of the gain-cell are estimated by measuring current flowing through the zero-voltage sources, V1 and V2, in the read and write path as shown in Fig.4 with 2T gain-cell as an example. The MEP read energy is defined as the sum of Read-0 and Read-1 energy at MEP voltage. The voltage sweep required to estimate MEP is performed down to 0.1V.

- Access Delay at MEP: The read delay is measured as the time from the instant the read word-line is activated till the read bitline voltage decreases by 0.03V, assuming sense amplifier can sense 30mV input voltage difference [18].

- Retention Time (RT) at MEP: In this paper, it is measured as the time it takes for the stored logic at SN to deteriorate till half of the supply voltage. This is different from its definition for above-threshold operation, where it is defined in terms of the threshold voltage of the read transistor T2 - Retention "0" (or "1") as the time it takes for $V_{SN}$ to rise (or fall) to $V_{th,T2}$. Since, the operating voltages in this paper are in sub-threshold region, we forgo the above-threshold definition and instead consider half-$V_{DD}$ as the limit for $V_{SN}$ in both Retention-"0" and "1" cases.

The spice net-lists of the 2T and 3T1D gain-cells are simulated in HSPICE [19] circuit simulator. The e-DRAMs were shown to perform reliably in near-threshold region at 40nm node in [15]. So in this paper, e-DRAM gain-cells are studied at the next scaled technology node 32nm (using HP PTM models [20]) which is going to be the technology node for the future sub-threshold circuit implementations.

A. Kriging meta-model for nominal(without-variation) case

In the design-space with four levels per parameter, there exists 262,144 ($4^9$) designs (2 lengths, 2 widths, 2 High Vth transistors, read and write wordline boosting and a temperature parameter) for 2T cell and 1,073,741,824 ($4^{15}$) designs for 3T1D cell. Furthermore, a voltage sweep needs to be performed at each of these design points to estimate the MEP. Design exploration with these many simulations can be very time expensive. Hence, a kriging meta-model [21] with matern kernel is first made for each of the metrics and then the subsequent analysis is done using these meta-models. To create these meta-models, 1000 points are sampled using the Latin-Hypercube-Sampling (LHS) method to produce a space-filling design. However, for a high-dimensional space, the distribution of points provided by LHS may deviate considerably from a uniform distribution (leading to high-discrepancy). Thus, an additional step of LHS optimization is performed, using the Enhanced Stochastic Evolutionary (ESE) algorithm provided in the DiceDesign package of R [22]. The kriging model trend is specified as a first order polynomial with a second order interactions. The model is cross validated by leave-one-out which gives coefficient of determination ($R^2$) 0.73 for 2T MEP energy. The validation plots for the regression model of 2T and 3T1D gain cells are shown in figure Fig.5.
Fig. 5: Regression validation for kriging model of 2T and 3T1D MEP energy. The top plot shows the agreement between predicted and actual values using leave-one-out cross validation. The middle plot verifies the assumption that residuals are randomly distributed around zero without any drift. The bottom plot verifies the assumption that residuals are almost normally distributed.

B. Full Factorial Analysis in presence of process variations

In the presence of process variations, it is necessary to find statistically significant design parameters. To compare each of these parameters of significance, their confidence intervals for improvement in MEP are needed. For this, a $2^k$ full factorial design experiment with 5000 replications is done for up-sized designs (lengths and widths of transistors with two levels [1x, 4x]). The p-values from ANOVA test [23] are then used to identify statistically significant design parameters with significance level of 0.001. The 95% confidence intervals for each design parameter in the effects-model are estimated as:

\[ \text{estimate} \pm t_{\alpha/2, df} \sqrt{\text{variance estimator}}, \]

where $\alpha = 0.05$ and $df$ is the degrees of freedom of error term. The variability in threshold voltage is assumed to be 6% following the EU project statement [24].

IV. RESULTS

A. Nominal Analysis (without process variations)

1) Sizing: The width of the read transistor is typically up-sized to increase the retention time. This however increases the MEP energy. The contour plot in Fig.6 shows that it is possible to decrease MEP energy when up-sizing the write transistor length while also up-sizing the read transistor width. The HSPICE simulation of 4x write transistor length design shows a decrease in MEP energy by 29% for 2T and 26% for 3T1D.

2) Wordline Boosting: Applying read wordline boosting increases the MEP energy. In contrast, the effect of write wordline boosting is to reduce the MEP energy. This can be seen in Fig.7. HSPICE simulations of 0.2V read wordline boosting design shows MEP energy is higher by 564% for 2T and 61% for 3T1D. While HSPICE simulation of 0.2V write wordline boosting design shows MEP energy is lower by 34% for 2T and 41% for 3T1D.

3) High Threshold Voltage Transistors: Using high threshold voltage transistors in the read and write paths to decrease leakage current has opposite effects on the MEP energy. While using high threshold transistors on the write path is reducing MEP energy, using high threshold transistors in the read path increases the MEP energy. This effect can be explained by the increase in the read delay which would consequently increase the read leakage energy. The contour plots in Fig.8 suggest that designs with high threshold transistors on both read and write path have lower MEP energy than designs with only high threshold read transistors. The HSPICE simulation of 0.2V higher threshold voltage for write transistor shows a decrease in MEP energy by 35% for 2T and 25% for 3T1D. The HSPICE simulation of the design with 0.2V higher threshold voltage read transistors shows an increase in the MEP energy by 860% for 2T and 293% for 3T1D.

4) Temperature: Increase in temperature increases the read MEP energy. However, the increase in energy can be reduced by also increasing the write length as in seen in Fig.9. HSPICE simulations show that at 100°C the increase in MEP energy is 116.9% for 2T and 130% for 3T1D. This increase is then reduced with the 4x up-sizing of write transistor length to only 12% for 2T and 23% for 3T1D.
Fig. 9: Temperature increases MEP energy. This can be mitigated by increasing the write transistor length. Colormap is blue for low MEP energy and pink for high MEP energy.

Fig. 10: Contour plots for product of Read MEP energy and read delay. The design with smallest product would be the optimum point with less MEP energy and smaller read delay. Colormap is blue for low values and pink for high values.

In summary, the read MEP energy is reduced by either write wordline boosting or using write transistor with high threshold voltage or by up-sizing write transistor length for both 2T and 3T1D gain cells. Thus reducing leakage current through write path is necessary to reduce MEP energy, especially at higher temperatures. On the contrary, reducing read delay by either up-sizing read transistor width or read wordline boosting increases the read MEP energy.

B. Joint Optimization of read energy with read delay, Retention time

The designs with a smaller Read MEP energy and also smaller read delay are found by considering designs with least energy-delay product. The contour plot for this product is shown in Fig.10, which shows that up-sizing the write transistor length decreases the energy-delay product. In contrast, up-sizing the read-transistor width increases the energy-delay product. The HSPICE simulation of 4x write transistor length design shows that the energy-delay product is reduced by 30% for 2T and 26.3% for 3T1D.

The HSPICE simulations showed that the retention time of 2T for stored value of ‘1’ and of 3T1D for stored value of ‘0’ is greater than 1ms for all up-sizing design options. The contour plots showing retention time for a stored value of ‘0’ for 2T and a stored value ‘1’ for 3T1D are shown in Fig.11.

In the case of 3T1D gain cell, the retention time of ‘1’ increases with up-sizing of write transistor length up to 2x and then starts decreasing. This is because the MEP supply voltage starts decreasing from 0.18V at 2x length to 0.14V at 4x length write transistor. Though the up-sizing of the read transistor width increases the retention time at a fixed supply voltage, it however decreases the read MEP supply voltage which is 0.18V for 1x width, 0.16V for 2x and 3x width, and 0.14V for 4x read transistor width. The effect of this on retention time is seen in the contour plot in Fig.11(a) where the retention time of ‘1’ at MEP decreases with up-sizing of read transistor width. The HSPICE simulation of the 3T1D design with 2x write transistor length shows 6.6% increase in retention time of ‘1’. The contour plot shows that the ‘energy * 1/retention time’ product for 3T1D decreases with up-sizing of write transistor length. The HSPICE simulation of design with 4x write transistor length shows 21% decrease in the ‘energy * 1/retention time’ product.

In contrast to the retention time of ‘1’ in 3T1D, the retention time of ‘0’ of 2T increases as MEP supply voltage decreases. The up-sizing of the read transistor width or the write transistor length decreases the MEP supply voltage from 0.18V to 0.1V. The HSPICE simulation of the design with both read transistor width and write transistor length up-sized by 4x shows 25% increase in 2T’s retention time of ‘0’. The product ‘energy * 1/retention time’ for 2T is higher for the up-sized read transistor width and decreases with up-sizing of write transistor length. The HSPICE simulation of design with 4x up-sized write transistor length shows 44% decrease in this product. Similar results are reported at the near-threshold voltage of 0.4V for 2T all-PMOS gain cell (i.e. higher worst case retention time of ‘1’ compared to ‘0’) [25]. The minimum energy-retention time trade-off can further be improved by tolerating some retention time failures in the presence of process variations [25].

Thus, reducing leakage current through write path by up-sizing the write transistor length also reduces the energy-delay product and the energy-1/retention product. While up-sizing read transistor width to decrease the read delay and increase the retention time, contrarily, increases the energy-delay product and the energy-1/retention product.

C. Full-Factorial analysis in presence of threshold voltage variations

In presence of process variations, the difference in median MEP energy of different read and write path transistor up-sizing is shown in boxplot Fig.12. For both 2T and 3T1D gain cells, the design with 4x up-sized length for read transistors and width for write transistors...
Fig. 12: Boxplot for MEP energy vs Up-sizing. X-labels are the different up-sizing combinations with first two symbols for read and write transistor lengths and last two for widths. "S" is 1x and "L" is 4x increase. For instance, SSSS is the design with all read and write lengths/widths of 1x size. In 2T gain cell, the distribution of MEP energy for designs with 1x read transistor width (first 8 designs from left) have similar distribution as the designs with 4x read transistor width (last 8 designs). Up-sized designs of 3T1D gain cells show larger standard deviation than 2T gain cells.

(S.L.L.S design) has the maximum median MEP energy. In the case of the 2T gain cell, up-sizing the width of the read transistor has only 12% increase in median MEP energy. The comparison of the 2T gain cell’s median MEP energy of the first 8 designs (designs with 1x read transistor width) with the last 8 designs (designs with 4x read transistor width) in the Fig.12 suggests that up-sizing read transistor width does not have significant effect on the median MEP energy. To verify this, the p-values from the ANOVA test are calculated for the main effects model. The results are shown in Tables I and II. The p-value in this analysis is interpreted as the probability of observing a difference in the mean MEP energy for an up-sized design with a sample size of 5000 when there is no actual change in MEP energy (i.e. the probability of observing different means when the null hypothesis is true). The effect of an up-sized design on MEP energy is considered to be statistically significant if its p-value is small. Considering the significance level of 0.001 (i.e. less than one in thousand chance of being wrong), since the p-value for up-sizing of read transistor width is greater than this significance level, the null hypothesis that up-sizing read transistor width has no statistically significant effect on MEP energy in presence of Vth variations cannot be rejected. The same conclusion is also reached from the main-effects plot in Fig.13 where the 95% confidence intervals of MEP energy for 4x up-sized read transistor width overlap with those of 1x read transistor width. The Tukey’s honest significant differences test [26] is then used to estimate the set of 95% confidence intervals

Table I: p-values for different 2T up-sizing. Smaller p-value means that factor has statistically significant effect. A p-value larger than 0.001 is considered to have no strong statistically significant effect on the response variable.

<table>
<thead>
<tr>
<th>Up-sizing</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>write length</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>write width</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>read length</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>read width</td>
<td>0.6585</td>
</tr>
</tbody>
</table>

Table II: p-values for different 3T1D up-sizing.

<table>
<thead>
<tr>
<th>Up-sizing</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>write length</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>read length (T2)</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>read length (T3)</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>diode length</td>
<td>0.8693</td>
</tr>
<tr>
<td>write width</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>read width (T2)</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>read width (T3)</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>diode width</td>
<td>0.7525</td>
</tr>
</tbody>
</table>

Table III: 2T: 95% CI for difference in means of MEP energy between levels: small (1x) and large (4x), for read and write transistors up-sizing. "L" is for large and "S" is for small.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Difference between means of levels</th>
<th>Lower 95% CI</th>
<th>Upper 95% CI</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>write length</td>
<td>$\mu(L) - \mu(S)$</td>
<td>-3.31e-21</td>
<td>-3.28e-21</td>
<td>atleast 60% dec</td>
</tr>
<tr>
<td>write width</td>
<td>$\mu(L) - \mu(S)$</td>
<td>3.14e-21</td>
<td>3.17e-21</td>
<td>atleast 140% inc</td>
</tr>
<tr>
<td>read length</td>
<td>$\mu(L) - \mu(S)$</td>
<td>4.75e-21</td>
<td>4.78e-21</td>
<td>atleast 349% inc</td>
</tr>
</tbody>
</table>
TABLE IV: 3TID: 95% CI for difference in means of MEP energy between levels :small (1x) and large (4x), for read and write transistor up-sizing. "L" is for large and "S" is for small.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Difference between means of levels</th>
<th>95% CI</th>
<th>summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>write length</td>
<td>µ(L) − µ(S)</td>
<td>-3.8e-21</td>
<td>atleast 3% dec</td>
</tr>
<tr>
<td>read length (T2)</td>
<td>µ(L) − µ(S)</td>
<td>1.9e-21</td>
<td>atleast 27% inc</td>
</tr>
<tr>
<td>read width (T2)</td>
<td>µ(L) − µ(S)</td>
<td>4.3e-21</td>
<td>atleast 3% inc</td>
</tr>
<tr>
<td>read length (T3)</td>
<td>µ(L) − µ(S)</td>
<td>-9.0e-22</td>
<td>atleast 19% dec</td>
</tr>
<tr>
<td>read width (T3)</td>
<td>µ(L) − µ(S)</td>
<td>8.6e-22</td>
<td>atleast 24% inc</td>
</tr>
</tbody>
</table>

(CI) of differences between the mean MEP energy between 1x and 4x levels of statistically significant up-sizing factors. The results are shown in Tables III and IV. The increase (decrease) in the mean MEP energy at the 4x up-sizing level is calculated as the percentage relative difference between the lower (upper) level value of its 95% CI and the mean at 1x up-sizing level. Up-sizing the write transistor length reduces the mean MEP energy by at-least 60% for 2T and 63% for 3TID gain cells in presence of threshold voltage variations. The up-sizing factor with largest increase in mean MEP energy in presence of vth variations for both 2T and 3TID gain cell is the read transistor length with at least 34% increase for 2T and at least 21% increase for 3TID.

V. CONCLUSION

This paper investigates the minimum read energy operation of 2T and 3TID gain cell in order to be candidates to substitute SRAM bitcells in sub-threshold memories. Results show that read MEP energy can be reduced by either increasing the length of write transistor (> 20% decrease), or by providing write word-line boosting during read (> 34% decrease), or using high-threshold voltage write transistor (> 25% decrease). In presence of process variations, the p-values from ANOVA show that up-sizing of read transistor width for 2T and up-sizing of diode transistor for 3TID are not statistically significant factors influencing read MEP energy. The factor resulting in largest increase in read MEP energy for both 2T and 3TID gain cell is the read transistor length (> 21% increase).

ACKNOWLEDGMENTS

This work has been supported by the Spanish Ministry of Economy and Competitiveness (MINECO) and ERDF through the project TEC2013-45638-C3-2-R, and the Spanish Ministry of Education and Science under grant TIN2010-18368 and TIN2013-44375-R, Generalitat of Catalonia under grant 2009SGR1250. Manish Rana is supported by FI-DGR-2013-638 grant.

REFERENCES