

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

Manish Rana\*, Ramon Canal\*, Esteve Amat\*\* and Antonio Rubio\*\*\*

\*Department of Computer Architecture, UPC Barcelona

\*\*Centro Nacional de Microelectrónica (CNM), Institut de Microelectrónica de Barcelona (CSIC)

\*\*\*Department of Electronic Engineering, UPC Barcelona

**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- $V_{th}$  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. A full-factorial statistical analysis of e-DRAM cells is performed in presence of threshold voltage variations and the effect of up-sizing on mean MEP is reported. Finally, it is shown that the product of the read and write lengths provides a knob for trade-off between energy-efficiency and reliable MEP energy operation.

## 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,

$$I_{SUB} = I_0 e^{V_{GS} - V_{th} / nV_T}$$

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 presence of process variations lead to variability in the  $I_{on}$ ,  $I_{off}$  currents and the circuit delay. Consequently there are significant variations in MEP energy and delay and their means shift to higher values when the process variations increase. Moreover, the leakage current is highly dependent on the temperature and a shift in the operating temperature can cause significant increase in the MEP energy of the circuit [10].

Also, 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. Nevertheless, SRAM dominates the energy consumption among the components of a processor [11] and several alternative SRAM bit-cells have been proposed. These sub-threshold SRAM bit-cells have 8-transistors [12] or 10-transistors [13]–[15].

As an alternative to SRAM bit-cells, Meinerzhagen et.al. [16] 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.1, which also shows the waveforms for their read/write operation. Since the leakage current of the nMOS transistor is significantly higher than that of the pMOS transistor, alternate cell configurations that mix the transistor types (pMOS write transistor and nMOS transistors for the read path) achieve better memory cell performance than the nMOS-only design [2], [3]. The storage node capacitor (SN), formed by T2's gate capacitance and

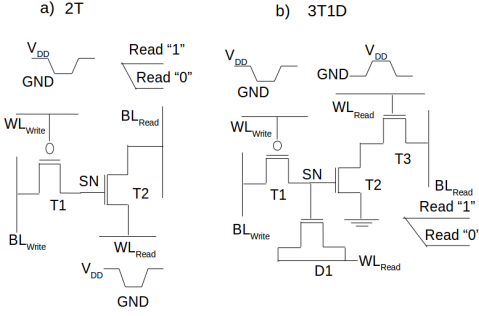


Fig. 1: Schematic of (a) 2T and (b) 3T1D gain cell. Read operation begins by pre-charging the read bitline. Subsequently read word-line is driven low for 2T and high for 3T1D gain cell to complete the read operation.

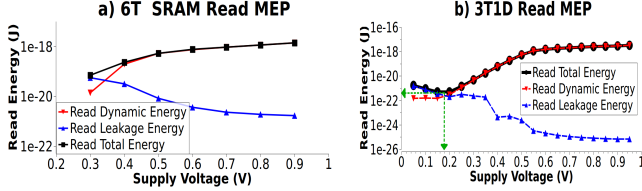


Fig. 2: Read minimum energy point (MEP) for a) 6T SRAM is  $7e-20$  J at 0.3V. b) 3T1D gain-cell is  $4e-22$  J at 0.2V.

T1's diffusion capacitance, stores the data as charge. To write data into the gain cell, T1 is turned on to transfer charge from  $BL_{Write}$  to SN. Fig.2 shows the MEP for read operation of 3T1D gain-cell and 6T SRAM bitcell, 6T has read MEP energy  $\sim 200\times$  that of 3T1D.

### 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  $WL_{Write}$ .

$\Delta r_{wl}$  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.

$\Delta w_{wl}$  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  $100^\circ C$

In order to estimate the energy at 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.3 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. The spice net-lists of the 2T and 3T1D gain-cells are simulated in HSPICE [18] circuit simulator. The e-DRAMs were shown to perform reliably in near-threshold region at 40nm node in [16]. So in this paper, e-DRAM

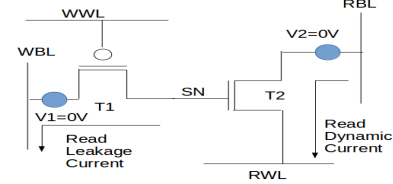


Fig. 3: The zero-voltage sources V1 and V2 are added to the write and read path. The current through these voltage sources is measured to estimate the leakage and dynamic energies during the read operation.

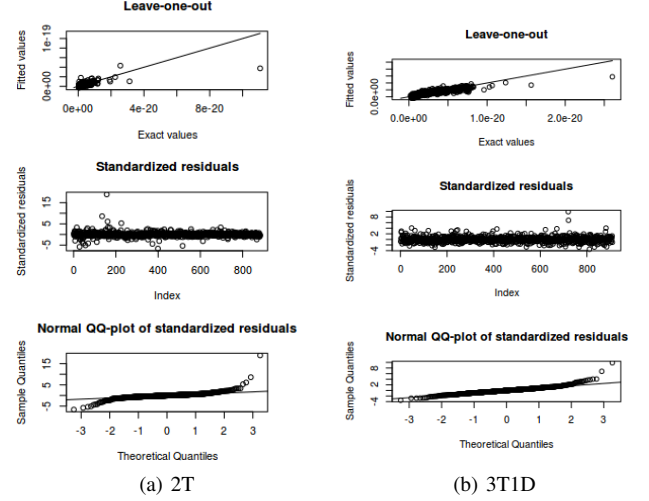


Fig. 4: Regression validation for kriging model of 2T and 3T1D MEP energy. The top plot shows the agreement between predicted and actual values using leave-one-out cross validation. The middle plot verifies the assumption that residuals are randomly distributed around zero without any drift. The bottom plot verifies the assumption that residuals are almost normally distributed.

gain-cells are studied at the next scaled technology node 32nm (using HP PTM models [19]) 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  $V_{th}$  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 [20] 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 [21]. 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.4.

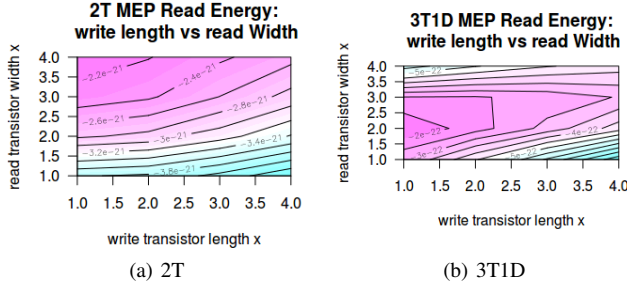


Fig. 5: Contour plots for MEP energy when up-sizing transistors. Increasing the write transistor length decreases MEP energy while increasing read transistor width increases MEP energy. Increasing both together keeps the MEP energy same. Colormap is blue for low MEP energy and pink for high MEP energy

#### 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 [22] 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:  $estimate \pm t_{\alpha/2, df} \sqrt{variance_{estimator}}$ , 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 [23].

### 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.5 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.6. 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.7 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

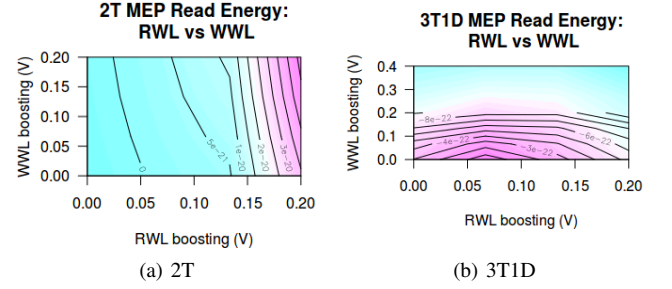


Fig. 6: Contour plots for MEP energy when wordline boosting is applied. Boosting read word line (RWL) is increasing MEP energy. Boosting write word line (WWL) is decreasing MEP energy. Colormap is blue for low MEP energy and pink for high MEP energy

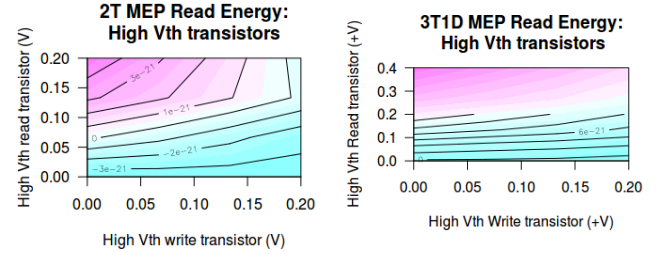


Fig. 7: Contour plots for MEP energy when high threshold transistors are used, with x-axis and y-axis as  $\Delta V_{th}$ . A high threshold voltage transistor in the write path decreases MEP energy. In contrast, using a high threshold transistor in the read path increases the MEP energy. Colormap is blue for low MEP energy and pink for high MEP energy

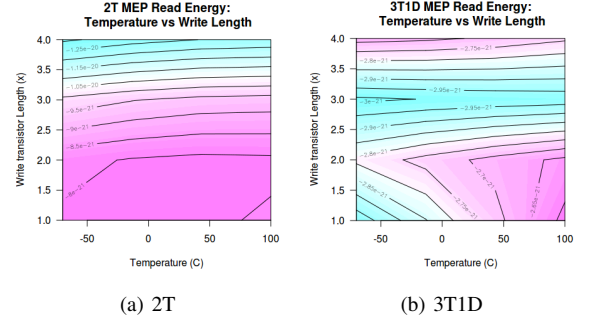
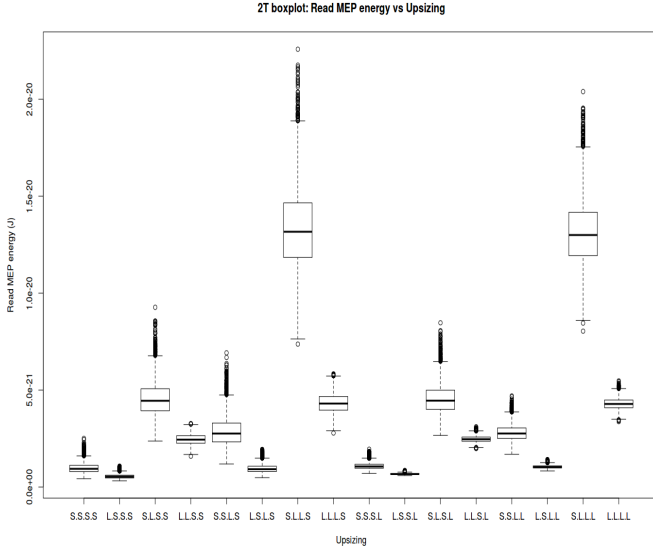


Fig. 8: Temperature increases MEP energy. This can be mitigated by increasing the write transistor length. Colormap is blue for low MEP energy and pink for high MEP energy

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 is seen in Fig.8. HSPICE simulations show that at  $100^\circ\text{C}$  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.

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.



(a) 2T

Fig. 9: 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.

#### B. 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.9. For both 2T and 3T1D gain cells, the design with 4x up-sized length for read transistors and width for write transistors (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.9 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 ???. 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 Tukey's honest significant differences test [24] is then used to estimate the set of 95% confidence intervals (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 II and

TABLE I: p-values for different 2T and 3T1D 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.

2T	Write Length	Write Width	Read Length		Read Width 0.68585			
p-value	2e-16	2e-16	2e-16					
3T1D	Write Length	Write Width	Read Length (T2)	Read Length (T3)	Read Width (T2)	Read Width (T3)	Diode Length	Diode Width
p-value	2e-16	2e-16	2e-16	2e-16	2e-16	2e-16	0.8693	0.7525

TABLE II: 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.

Factor	difference means of levels	lower 95% CI	upper 95% CI	summary
write length	$\mu(L) - \mu(S)$	-3.31e-21	-3.28e-21	atleast 60% dec
write width	$\mu(L) - \mu(S)$	3.14e-21	3.17e-21	atleast 140% inc
read length	$\mu(L) - \mu(S)$	4.75e-21	4.78e-21	atleast 349% inc

III. 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 3T1D 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 3T1D gain cell is the read transistor length with at least 349% increase for 2T and at least 215% increase for 3T1D.

#### C. Fitting a Log-Normal distribution to MEP energy under process variations

The previous section provided comparison of the empirical mean MEP energies among the up-sized gain-cell designs. When designing memories for IoT devices, it is also necessary to compare the higher quantiles of the MEP energy distribution to estimate the probability of meeting the energy constraints. In the case of a Normally distributed data, the 0.998 quantile (which is  $\mu \pm 3\sigma$  under Normal distribution) or higher estimates are used when comparing the yield. However, as the boxplots in Fig.9 show, the distribution of the MEP energy is not a Normal distribution rather have a long right-tail. Hence, to accurately compare the MEP energy distributions of the up-sized designs at higher quantiles similar to the  $(\mu + 3\sigma)$  under Normal distribution, the energy values at the 0.998 and higher quantiles must be first found. This requires fitting a long-tail distribution to the empirical MEP energy distribution.

The subthreshold current is exponentially dependent on the threshold voltage of the transistors. Since, the threshold voltages of the transistors are considered to be Normally distributed random variables due to the presence of process variations. Hence, the subthreshold current (both the on-current,  $I_{ON}$ , and the off-current  $I_{OFF}$ ) is a Log-Normal random variable, and as a consequence so is delay,  $t_d$ , which is dependent on the random variable  $I_{ON}$  [25]. The leakage energy given by:  $E_{LEAK} \propto I_{OFF} * t_d$ , is a product of two Log-Normal random variables, which are however not independent because the voltage at storage node, SN, influences both the  $I_{ON}$  and  $I_{OFF}$ . TABLE III: 3T1D: 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.

Factor	difference means of levels	lower 95% CI	upper 95% CI	summary
write length	$\mu(L) - \mu(S)$	-3.87e-21	-3.86e-21	atleast 63% dec
write width	$\mu(L) - \mu(S)$	3.66e-21	3.68e-21	atleast 160% inc
read length (T2)	$\mu(L) - \mu(S)$	1.01e-21	1.027e-21	atleast 27% inc
read length (T3)	$\mu(L) - \mu(S)$	4.30e-21	4.32e-21	atleast 215% inc
read width (T2)	$\mu(L) - \mu(S)$	-9.02e-22	-8.86e-22	atleast 19% dec
read width (T3)	$\mu(L) - \mu(S)$	8.62e-22	8.78e-22	atleast 24% inc



flowing through the read path and  $I_{OFF}$  flowing through the write path (Fig.3). The leakage energy is thus not an exact Log-Normal distribution. Still, it is seen from the Fig.10 that CDF plots of fitted Log-Normal distributions better approximate the empirical MEP energy CDF than the fitted Normal distribution. The Normal distribution parameters are estimated using the maximum likelihood estimate (MLE). The parameters for Log-Normal distribution are estimated using three different estimation approaches to compare their accuracy at the longer right tail: (1) maximum-likelihood (MLE), (2) quantile-matching (QME) and (3) maximum-goodness-of-fit. The quantile-matching method estimates the Log-Normal distribution parameters by minimizing the distance between the observed and theoretical quantiles. We specify one of the quantile probability = 0.998, which is the quantile probability of  $\mu + 3\sigma$  under Normal distribution. This gives better fit at higher values of the MEP energy distribution. The maximum-goodness-of-fit method provides the estimates of the Log-Normal distribution parameters for which the distance between the CDF of the fitted distribution and the CDF of the empirical MEP read energy distribution is minimum. Furthermore, achieving correctness of the estimated CDF at the right-tail (higher energy values) of the MEP read energy is important because the empirical distribution of the MEP energy has a longer right tail. Hence, the following Anderson-Darling statistic [26], [27] are considered for fitting the Log-Normal distribution to the empirical MEP energy distribution with higher accuracy at the right-tail:

$$ADR = \int_{-\infty}^{\infty} \frac{(CDF_{Empirical}(x) - CDF_{Log-Normal}(x))^2}{(1 - CDF_{Log-Normal}(x))} dx$$

$$ADR2 = \int_{-\infty}^{\infty} \frac{(CDF_{Empirical}(x) - CDF_{Log-Normal}(x))^2}{(1 - CDF_{Log-Normal}(x))^2} dx \quad (1)$$

Both of the above right-tail AD statistics, give larger weights to observations at the right tail, thereby, decrease the discrepancy between the right tails of the empirical MEP energy distribution and the fitted Log-Normal distribution. Moreover, among these two right-tail AD statistics, the ADR2 has larger weights at the right tail. The Fig.10 also shows the discrepancy at the tails using the QQ plots of the fitted Normal and Log-Normal distributions and the empirical MEP energy distribution. The Log-Normal distribution fitted using AD statistic provides higher accuracy of quantiles than others at the right-tail of the MEP energy distribution (top area in the QQ-plot). This accuracy comes at the cost of larger discrepancy at the left-tail of the distribution (lower-left area in CDF plot). Since, we are only interested in comparing quantiles for higher MEP energy values (which are on the right-tail), the use of the maximum-goodness-of-fit estimation method based on the ADR2 statistic is used for fitting Log-Normal distribution to the empirical distribution of MEP energy. Tables IV shows that using Normal distribution to fit the empirical MEP energy distribution results in under-estimating the MEP energy at higher quantiles. For instance, the 2T MEP energy at 0.998 quantile under the Log-Normal distribution, is 22.43% higher than the 0.998 quantile ( $\mu + 3\sigma$ ) under Normal distribution. Thus, 99.7% confidence interval of the MEP energy would be under-estimated if it is calculated as  $\mu \pm 3\sigma$ .

Moreover, the Log-Normal distribution fitted using MLE estimates of parameters, also under-estimates the quantiles of the MEP energy compared to those fitted using right-tail AD statistic. This is also seen in the QQ plot of Fig.10, where the Log-Normal distribution fitted using MLE estimates assigns MEP energy values at higher empirical quantiles to lower theoretical quantiles (that is, for observations lying above the straight line in the QQ plot).

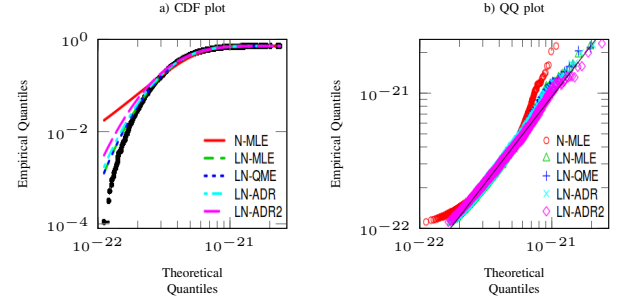


Fig. 10: CDF and QQ-plot comparing the accuracy of the fitted Normal distribution (N-MLE) and Log-Normal distributions (LN-MLE, LN-QME, LN-ADR, LN-ADR2) with respect to the empirical MEP energy data. The Log-Normal distribution fitted using maximum-goodness-of-fit estimate (LN-ADR2) predicts the tail better than the other fitted distributions

TABLE IV: 2T gain-cell: Fitting Normal distribution to MEP energy underestimates the quantiles at the right tail of the distribution (quantiles at same probability as  $\mu + 1\sigma$ ,  $\mu + 3\sigma$ ,  $\mu + 5\sigma$  under normal distribution) in comparison with the fitted Log-Normal distribution

Probability of Quantile Under Normal distribution	Relative Error in the Quantile estimates of fitted Normal distribution compared to the fitted Log-Normal estimates			
	Parameter estimation methods for fitting Log-Normal distribution			
	MLE	QME	AD	AD 2nd Order
Prob = $CDF(\mu + 1\sigma) = 0.8413447$	-0.89%	-1.55%	-1.57%	-0.82%
Prob = $CDF(\mu + 3\sigma) = 0.9986501$	12.24%	16.13%	17.65%	22.43%
prob = $CDF(\mu + 5\sigma) = 0.9999997$	37.06%	50.66%	54.66%	66.19%

#### D. Shrinking the right tail of MEP energy distribution

The longer tails of the MEP energy restrict from achieving a reliable minimum-energy operation. The higher quantiles are also at higher energy values at the far end of the right-tail of MEP energy distribution. As such, reducing the right tail of the MEP energy distribution would be highly beneficial for the energy-efficient operation. Here onwards,  $\mu$  and  $\sigma$  refer to the mean and sigma of the fitted MLE Normal distribution to the empirical MEP energy distribution (In this case,  $\mu$  then is equal to the empirical mean of energy distribution). Fig.11 compares the QQ plot for five up-sizing options: width of the read transistor, length of write transistor, both read width and write length together, and the lengths read and write transistors together. Upsizing the lengths of both read and write transistors, provides an almost Normal distribution of MEP energy. For this up-sized 2T design, the difference between the 0.998 quantile of fitted Normal distribution (that is  $\mu + 3\sigma$ ) and 0.998 quantile of the fitted Log-Normal distribution is only 1.46%, compared to the 66% relative difference for nominal size design. Thus, it provides shorter right tails than the other designs. However, this shrinkage in MEP energy distribution tail comes at the cost of 150% increase in mean MEP energy.

In order to compare upsized designs in terms of their right-tail length, a kriging model is made to predict the discrepancy between the  $\mu + 3\sigma$  of the fitted Normal MEP energy distribution and the 0.998-quantile of the fitted Log-Normal MEP energy distribution, with transistor sizes as the predictors of the model. The upsized designs which have longer right tails will also show larger discrepancy in the 0.998 quantile from the  $\mu + 3\sigma$  estimate.

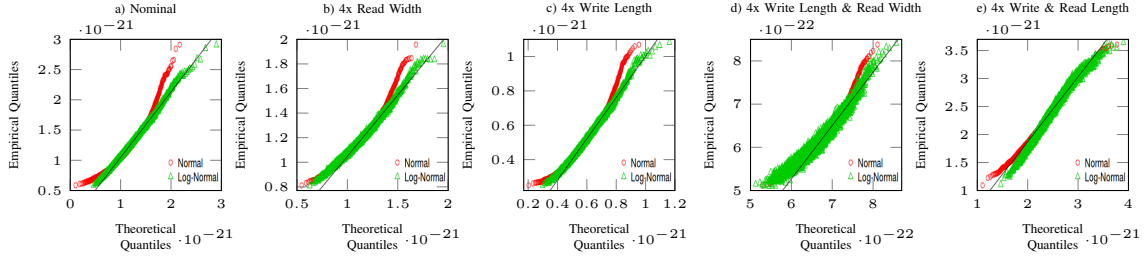


Fig. 11: Upsizing lengths of both read and write transistors decreases the right tail of MEP energy distribution

In Fig.12(a), it is seen in the QQ plot that the resulting model has non-Normal residuals and moreover, the distribution of standardized residual (middle sub-figure) shows that these residuals have an unequal variances. The prediction given by the kriging model for a given transistor upscaling is the mean of Normally distributed predictions for that upsized design. So, in this case, because the residuals are not Normally distributed, the mean of the predictions by the kriging model will not be accurate. Furthermore, The un-equal variances of the residuals violates the assumption of homogeneous variance of the random error which results in larger variance (uncertainty) of the model predictions. Hence, addressing both these problems is necessary to build an accurate model for predicting the 0.998-quantile discrepancy. Since, the QQ plot in Fig.12(a) shows that the residuals have higher values at the tails than if they were normally distributed, a log transformation of the target variable, that is  $\log(D_{0.998Q}) = \log(\text{Quantile}(0.998) - (\mu + 3\sigma))$  is modeled using kriging with transistor lengths and widths as predictors. The analysis of the residuals of the consequent model is shown in Fig.12(b). The residuals with this model have normal distribution with nearly same variance, as such it satisfies the normality and uniform-variance assumptions of the kriging model. Thus the consequent model to predict the 0.998 quantile discrepancy is of the form:

$$\log(D_{0.998Q}) = G(\text{Length}_{\text{write}}, \text{Length}_{\text{read}}, \text{Width}_{\text{write}}, \text{Width}_{\text{read}})) + \epsilon \quad (2)$$

where,  $G$  is a stationary Gaussian random field with covariance between vectors in the field defined by a kernel function (in this case, matern kernel is used), and  $\epsilon$  is the *i.i.d* random error.

Altogether, following models are created using kriging regression:

- 1)  $\log(D_{0.998Q})$  - To predict the discrepancy in the fitted Log-Normal 0.998 Quantile of the MEP energy from that of fitted Normal distribution (that is  $\mu + 3\sigma$ ). It is used as a measure of the energy distribution's right-tail length because larger right-tail will increase the 0.998 quantile value compared to  $(\mu + 3\sigma)$ . Log-transformation is needed to satisfy the normality and uniform variance assumption of the model residuals.
- 2)  $\mu_{\lognormal}$  - To predict the mean MEP energy predicted by the fitted Log-Normal distribution. It is used to compare the energy-efficiency among the upsized designs at the mean of their energy distributions. Note that the mean under the fitted Log-Normal distribution will be smaller than the empirical mean of the MEP energy distribution because Log-Normal distribution takes into consideration that some data lies at right-tail. For instance, for the nominal sized design, it's empirical mean is  $7.26e-22J$  while the mean of the fitted Log-Normal distribution is  $7.06e-22J$ . No transformation is needed as the residuals satisfy the normality and uniform variance assumptions.
- 3)  $\log(Q(0.998))$  - To predict the MEP energy at the 0.998-Quantile,  $Q(0.998)$ , under the fitted Log-Normal distribution. It

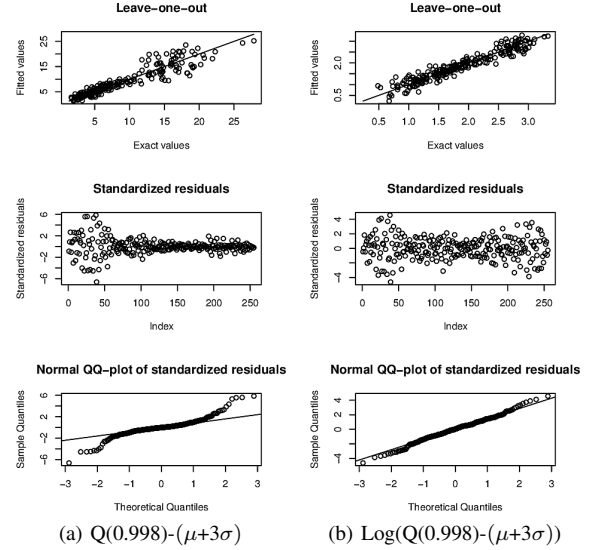


Fig. 12: Model diagnostic plots: Convalidation, homoscedasticity (Equal variance of residuals) and Normality of residuals, for the model of deviation between 0.998 Quantile under fitted Log-Normal and Normal MEP energy distribution

is used to compare the energy-efficiency at higher quantile than mean. Upsized designs with smaller 0.998-Quantile MEP energy also have smaller probability of exceeding energy constraints of the IoT device, than those designs with larger energy values. Log-transformation is needed to satisfy the normality and uniform variance assumptions of the residuals in the model.

The contour plots in the Fig.13 show a) the trend in the mean of MEP energy distribution, b) its 0.998 quantile, and c) its deviation from Normal distribution against the read and write length upscaling. It is seen that increasing the write length decreases  $\mu_{\lognormal}$  initially and then increases them. For instance, in the case of design with 3x read length, the mean MEP energy is around  $108 \times 10^{-23}J$  for 1x write length,  $91 \times 10^{-23}J$  for 2x write length,  $86 \times 10^{-23}J$  for 3x write length and  $104 \times 10^{-23}J$  for 4x write length. Similar trend is seen in Fig.13(b) for  $Q(0.998)$ , which is the energy at the 0.998 quantile of the MEP energy distribution. The contour plot, Fig.13(c) shows that the discrepancy from Normal distribution is least when both the read length and write length are increased.

The mean of MEP distribution and the discrepancy of its 0.998 quantile from the Normal distribution are related to the product and the ratio of the upscaling in write and read lengths. This is seen in the graphs of Fig.14(a),(b) where the increase in the write length to read length ratio (0.25 to 4) increases the energy-efficiency by decreasing the mean MEP energy ( $117 \times 10^{-23}J$  to  $51 \times 10^{-23}J$ , 56% decrease) and the 0.998 quantile ( $161 \times 10^{-23}J$  to  $92 \times 10^{-23}J$ , 42%

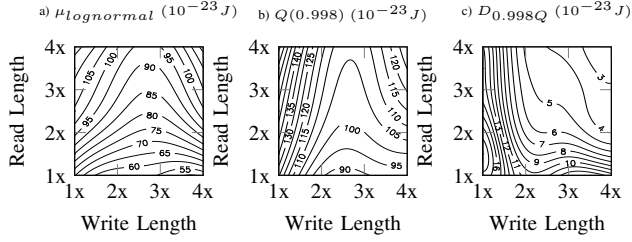


Fig. 13: Contour plots comparing upsizing of read and write lengths. All number labeling contours are in the order of  $10^{-23}$ .

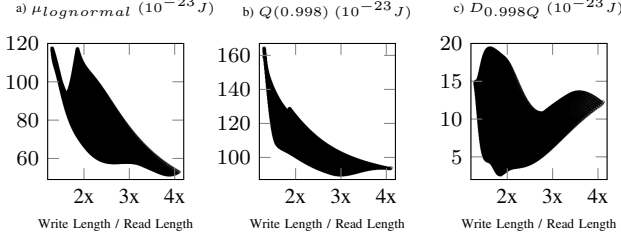


Fig. 14: Write-Length/Read-Length ratio vs- (a) lognormal mean (b) 0.998 quantile (c) Discrepancy of 0.998 quantile from  $\mu + 3\sigma$ , for 4000 grid points (200x200) with the write and read length between [1x,4x]

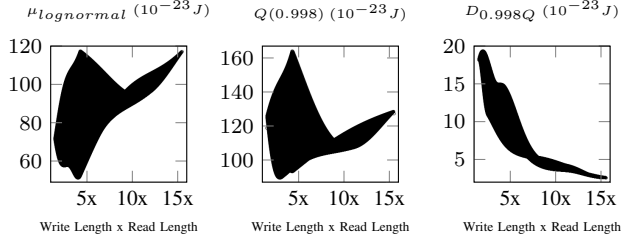


Fig. 15: Product of write length and read length vs- (a) lognormal mean (b) 0.998 quantile (c) Discrepancy of 0.998 quantile from  $\mu + 3\sigma$ , for 4000 grid points (200x200) with the write and read length between [1x,4x]

decrease). However, Fig.14(c) shows that the discrepancy from the Normal distribution is also larger when write length is larger than read length ( $2 \times 10^{-23} \text{J}$  at ratio=4/4=1 to  $14 \times 10^{-23} \text{J}$  at ratio=4/1=4, 600% increase). On the other hand, a larger write length and read length product (1 to 16), decreases the discrepancy from Normal distribution ( $17 \times 10^{-23} \text{J}$  to  $2 \times 10^{-23} \text{J}$ , 88% decrease) but at the cost of lower energy-efficiency (mean increases from  $70 \times 10^{-23} \text{J}$  to  $117 \times 10^{-23} \text{J}$ , 67% increase), seen in Fig.15.

In order to find trade-offs between achieving higher energy-efficiency and achieving normally distributed MEP energy, the product ( $\text{MxD} = \mu_{\lognormal} * D_{0.998Q}$ ) of the mean MEP energy and the discrepancy from Normal distribution is taken. The values of  $\mu_{\lognormal}$  and  $D_{0.998Q}$  are scaled to [1, 2] before taking the product to make it independent of units. A design with lower mean MEP energy and smaller discrepancy from Normal distribution will also have smaller value of MxD. The Fig.16 shows that MxD decreases (from 3.43 to 1.64) with the increase in the ratio of write length and read length (0.25 to 4). Moreover, the decrease in MxD (from 1.74 to 1.64) is not comparatively significant when ratio is increased from 2 to 4. The Fig.16(d) shows that MxD remains almost constant (from 1.67 to 1.74) with increase in the product of write and read lengths (from 1 to 16), on the condition that the write length is at-least twice the read length. This suggests that the ratio of write and read length, and their product can be used as knobs for trade-offs between MEP energy-efficiency and Normal distribution of MEP energy while keeping MxD constant. To illustrate this trade-

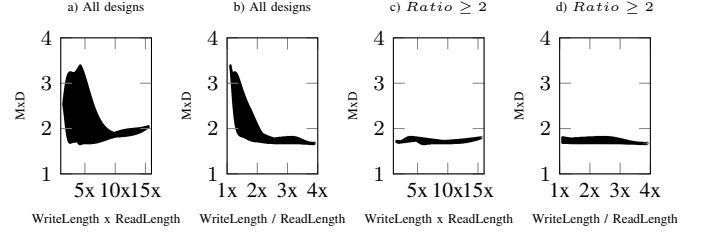


Fig. 16: Product of mean MEP energy and discrepancy from Normal distribution vs- product and ratio of write length and read length, for 4000 grid points (200x200) with the write and read length between [1x,4x]

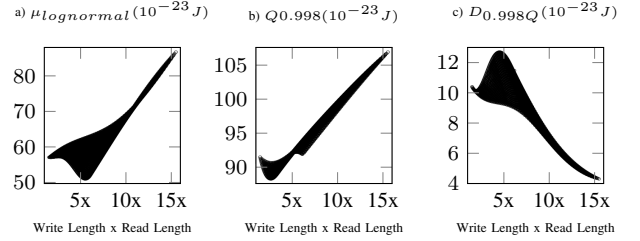


Fig. 17: Product of write and read lengths vs- (a) Mean MEP energy, (b) 0.998 quantile, and (c) Discrepancy from Normal distribution, when the write length is at-least twice the read length

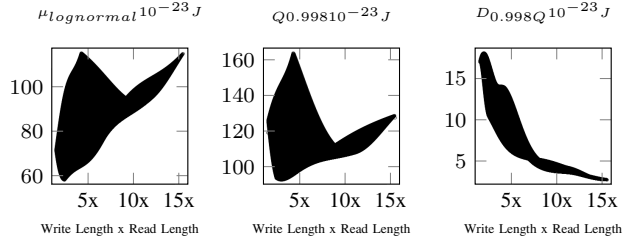


Fig. 18: Product of write and read lengths vs- (a) Mean MEP energy, (b) 0.998 quantile, and (c) Discrepancy from Normal distribution, when the write length is less than twice the read length

off, Fig.17 shows that when the write length is at-least twice the read length, then the designs with higher product of write and read length show smaller discrepancy from Normal distribution at the cost of higher MEP energy and its 0.998 Quantile. That is, for these designs, increasing the product of the length up-sizing from 1 to 16, decreases  $D_{0.998Q}$  from  $10.5 \times 10^{-23} \text{J}$  to  $4 \times 10^{-23} \text{J}$  (62% decrease) and increases the  $\mu_{\lognormal}$  from  $56 \times 10^{-23} \text{J}$  to  $88 \times 10^{-23} \text{J}$  (57% increase). On the other hand, when the product of the the lengths is smaller (with write length still at-least twice the read length) then energy-efficiency increases at the cost of higher discrepancy from Normal distribution. However, it is also seen in Fig.15 that the MEP energy is higher for some designs with smaller product of lengths. The plots in Fig.18 show that these designs with higher energy values at lower product were only those designs with write length smaller than the twice of read length. These designs exhibit larger spread in the MEP energy among themselves. Consequently, no simple trade-off mechanism between the MEP energy-efficiency and achieving normally distributed MEP energy, could be found similar to product of write and read lengths which is available when the write length at-least twice the read length. It should also be noted, that for designs with write-length/read-length  $> 2x$ , the increase in the product of the write and read lengths also increases the 0.998 quantile (from  $91 \times 10^{-23} \text{J}$  to  $107 \times 10^{-23} \text{J}$ , 17.5% increase). The relative difference between the mean MEP energy and 0.998 quantile, is an estimate of the coefficient-of-variation. That is, it provides an estimate of the variance in the MEP energy distribution relative to its mean, similar to

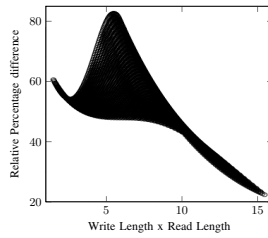


Fig. 19: Relative difference in percentage between 0.998 quantile and  $\mu_{\lognormal}$  -vs- product of read and write lengths, when the write length is at-least twice the read length

the  $3\sigma/\mu$  ratio used for Normally distributed data. Fig.19 shows that this relative difference decreases from 62% to 22% with increase in the product of write and read lengths upsizing from 1 to 16. Therefore, using write and read lengths which have higher product, increases the reliability of MEP energy (coefficient-of-variation decreases to 22%) and also provides a more symmetric MEP energy distribution (62% decrease in  $D_{0.998Q}$ ) using which it is easy to find the 0.998 quantile using  $\mu + 3\sigma$  and 99.5 % confidence intervals using  $\mu \pm 3\sigma$ . However, this comes at the cost of 57% increase in the mean MEP energy.

## V. CONCLUSION

This paper investigates the minimum read energy operation of 2T and 3T1D 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 ( $> 26\%$  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 3T1D are not statistically significant factors influencing read MEP energy, while largest increase in read MEP energy is caused by the upsizing of read transistor length ( $> 215\%$  increase). Moreover, the efficacy of using Log-Normal distribution to fit the empirical MEP energy distribution using AD statistic is also shown. Furthermore, it is shown that the product of read and write length upsizing can be used as knob for trade-off between reliable MEP energy operation or energy-efficient operation.

## 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 Catalunya under grant 2009SGR1250. Manish Rana is supported by FI-DGR-2013-638 grant.

## REFERENCES

- [1] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers, "Big data: The next frontier for innovation, competition, and productivity," 2011.
- [2] E. Amat, A. Calomarde, F. Moll, R. Canal, and A. Rubio, "Feasibility of the embedded dram cells implementation with finfet devices," *Computers, IEEE Transactions on*, 2014.
- [3] K. C. Chun, P. Jain, J. H. Lee, and C. H. Kim, "A sub-0.9 v logic-compatible embedded dram with boosted 3t gain cell, regulated bit-line write scheme and pvt-tracking read reference bias," in *VLSI Circuits, Symposium on*, 2009.
- [4] K. C. Chun, P. Jain, J. H. Lee, and C. H. Kim, "A 3t gain cell embedded dram utilizing preferential boosting for high density and low power on-die caches," *Solid-State Circuits, IEEE Journal of*, 2011.

- [5] R. Iqbal, P. Meinerzhagen, and A. Burg, "Two-port low-power gain-cell storage array: voltage scaling and retention time," in *ISCAS*. IEEE, 2012.
- [6] X. Liang, R. Canal, G.-Y. Wei, and D. Brooks, "Replacing 6t srams with 3t1d drams in the l1 data cache to combat process variability," *IEEE micro*, 2008.
- [7] A. Teman, P. Meinerzhagen, A. Burg, and A. Fish, "Review and classification of gain cell edram implementations," in *Electrical & Electronics Engineers in Israel (IEEEI), IEEE 27th Convention of*, 2012.
- [8] B. H. Calhoun, A. Wang, and A. Chandrakasan, "Modeling and sizing for minimum energy operation in subthreshold circuits," *Solid-State Circuits, IEEE Journal of*, vol. 40, no. 9, pp. 1778–1786, 2005.
- [9] S. Hanson, B. Zhai, D. Blaauw, D. Sylvester, A. Bryant, and X. Wang, "Energy optimality and variability in subthreshold design," in *Proc. ISLPED*. ACM, 2006, pp. 363–365.
- [10] E. Amat, J.-F. Christmann, O. Billoint, I. Miro, and E. Beigne, "Fdsoi suitability for asynchronous circuits at sub-vt," *ECS Transactions*, vol. 66, no. 5, pp. 315–322, 2015.
- [11] A. Y. Dogan, J. Constantin, M. Ruggiero, A. Burg, and D. Atienza, "Multi-core architecture design for ultra-low-power wearable health monitoring systems," in *DATE*, 2012, pp. 988–993.
- [12] N. Verma and A. P. Chandrakasan, "A 256 kb 65 nm 8t subthreshold sram employing sense-amplifier redundancy," *Solid-State Circuits, IEEE Journal of*, vol. 43, no. 1, pp. 141–149, 2008.
- [13] J. P. Kulkarni, K. Kim, and K. Roy, "A 160 mv robust schmitt trigger based subthreshold sram," *Solid-State Circuits, IEEE Journal of*, 2007.
- [14] B. H. Calhoun and A. P. Chandrakasan, "A 256-kb 65-nm sub-threshold sram design for ultra-low-voltage operation," *Solid-State Circuits, IEEE Journal of*, vol. 42, no. 3, pp. 680–688, 2007.
- [15] 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," *Solid-State Circuits, IEEE Journal of*, vol. 44, no. 2, pp. 650–658, 2009.
- [16] P. Meinerzhagen, A. Teman, R. Gitterman, A. Burg, and A. Fish, "Exploration of sub-vt and near-vt 2t gain-cell memories for ultra-low power applications under technology scaling," *Journal of Low Power Electronics and Applications*, 2013.
- [17] K. Lovin, B. C. Lee, X. Liang, D. Brooks, and G.-Y. Wei, "Empirical performance models for 3t1d memories," in *ICCD*. IEEE, 2009, pp. 398–403.
- [18] [Online]. Available: <https://www.synopsys.com/Tools/Verification/AMSVVerification/CircuitSimulation/HSPICE/Pages/default.aspx>
- [19] [Online]. Available: <http://ptm.asu.edu/>
- [20] J. D. Martin and T. W. Simpson, "Use of kriging models to approximate deterministic computer models," *AIAA journal*, vol. 43, no. 4, pp. 853–863, 2005.
- [21] D. Dupuy, C. Helbert, and J. Franco, "DiceDesign and DiceEval: Two R packages for design and analysis of computer experiments," *Journal of Statistical Software*, vol. 65, no. 11, pp. 1–38, 2015. [Online]. Available: <http://www.jstatsoft.org/v65/i11/>
- [22] G. E. Box, W. G. Hunter, J. S. Hunter *et al.*, "Statistics for experimenters," 1978.
- [23] J. A. Rubio Sola, J. Figueras Pàmies, E. I. Vatajelu, and R. Canal Corretger, "Process variability in sub-16nm bulk cmos technology," 2012.
- [24] H. Abdi and L. J. Williams, "Tukeys honestly significant difference (hsd) test," *Encyclopedia of Research Design*. Thousand Oaks, CA: Sage, pp. 1–5, 2010.
- [25] J. Kwong and A. P. Chandrakasan, "Variation-driven device sizing for minimum energy sub-threshold circuits," in *Proceedings of the 2006 international symposium on Low power electronics and design*. ACM, 2006, pp. 8–13.
- [26] T. W. Anderson and D. A. Darling, "Asymptotic theory of certain goodness of fit" criteria based on stochastic processes," *The annals of mathematical statistics*, pp. 193–212, 1952.
- [27] M. L. Delignette-Muller and C. Dutang, "fitdistrplus: An R package for fitting distributions," *Journal of Statistical Software*, vol. 64, no. 4, pp. 1–34, 2015. [Online]. Available: <http://www.jstatsoft.org/v64/i04/>