Tutorial: Analysis on 2T and 1T Array based

Compute in Memory (CIM) and Physical Unclonable Function (PUF)

Introduction:

The standard 8 transistors (8T) SRAM has been widely used in compute in-memory (CIM) and physical unclonable function (PUF) design. The 8T SRAM consists of a 2T array and 6T SRAM cells. Besides, 1T array is similar to 2T array and it can be designed as a PUF. In this paper, we make systematic analysis on the 2T and 1T array based CIM and PUF. Section 1 introduces the structure of 2T and 1T array and discuss about the relationship between the internal parameters and external control signals. Section 2 proposes metrics to evaluate the performance of CIM and PUF.

1 Preliminaries:

In a standard 2T cell, two NMOS are connected in series and the one connected to Bit Line (BL) is M1 while the other one connected to the ground is M2. M2's gate is generally connected to a 6T SRAM cell which stores the weights ($w_{i,j} \in \{0,1\}$) in CIM or directly controlled by challenges ($Ch_i \in \{0,1\}$) in PUF. M1's gate is controlled by the input pulse-width modulation (PWM) signal on the Word Line. We assume that the value of M1's gate voltage is V_{WL} and M2's is V_{WL}. The pre-charged voltage of Bit Line is V_{DD} and the capacitor connected on the Bit Line is C_{BL}. For the 1T Array, M2 is removed and M1's source is directly connected to the ground.

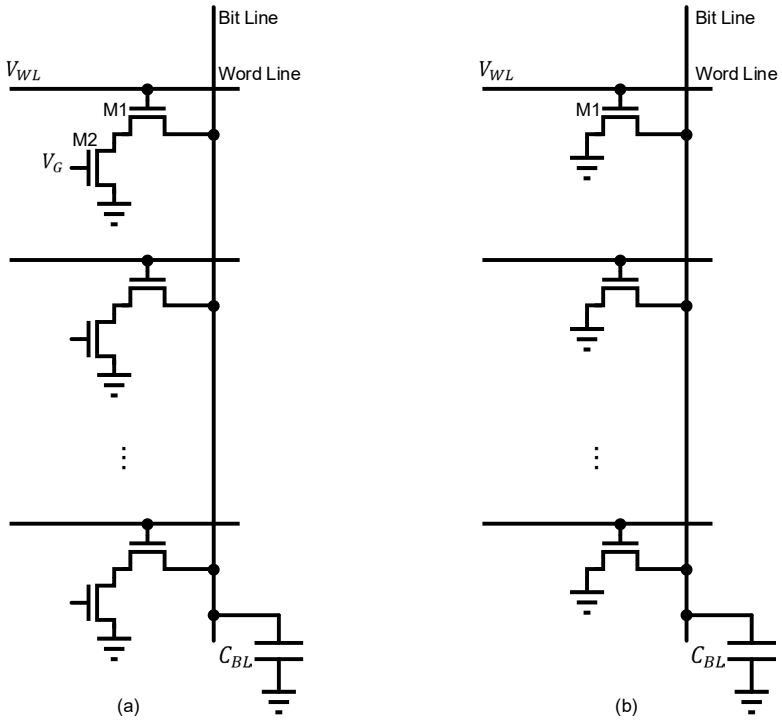

Fig. 1. The schematic of 2T Array (a) and 1T Array (b).

In this section, the operating region of the 2T cell is determined first and we derivate the relationship between the internal parameters and external control signals of both the 1T cell and 2T cell under the ideal condition.

1.1 The Operating Regions of 2 Transistors

2 transistors of the array should work in the linear region or saturation region so there are 4 possible combinations of 2T's operating regions. We assume the gate voltage of M1 (V_{WL}) is lower than the gate voltage of M2 (V_G):

$$V_{WL} \leq V_G \tag{1}$$

When the V_{BL} approaches to 0, M1 and M2 are both in the linear region. When the V_{BL} is growing and at least one transistor is in the saturation region, by the process of elimination, the only possibility of the working regions is that M2 is in the linear region and M1 is in the saturation region.

1.2 The Internal Parameters of the 1T and 2T Array

Generally, the discharge current of the 1T or 2T cell (I_{DS}) should be a constant when the Bit Line capacitor is discharging. Thus, for the 1T Array, the transistor should be in the saturation region. For the 2T Array, at least one of the two transistors should be in the saturation region, and under the condition of equation (1), M1 is in the saturation and M2 is in the linear region. Thus, for both 1T and 2T Array, the drain-source voltage of M1 (V_{DS1}) should be higher than $V_{GS1} - V_{th}$:

$$V_{DS1} \geq V_{GS1} - V_{th} \tag{2}$$

Eliminate the term V_{S1} on the both sides:

$$V_{D1} \geq V_{G1} - V_{th} = V_{WL} - V_{th} \tag{3}$$

Since V_{D1} is the Bit Line voltage V_{BL}, we can get the minimum value of V_{BL}:

$$V_{BL\_min} = V_{WL} - V_{th} \tag{4}$$

Than we can obtain the full scale of the Bit Line for 1T and 2T in the calculation phase:

$$V_{FS} = V_{DD} - V_{BL\_min} \tag{5}$$

And the least significant bit voltage V_{LSB} of the multiple-accumulate (MAC) output is:

$$V_{LSB} = \frac{V_{FS}}{2^{N_y}} \tag{6}$$

where the N_y is the output precision of a single MAC operation. Besides, we can derive the relationship between the V_{BL_min} and ideal discharge current I_{DS0} in the calculation phase. For a 1T cell:

$$I_{DS0} = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{GS1} - V_{th})^2 \tag{7}$$

where the V_{GS1} refers to the gate-source voltage of M1:

$$V_{GS1} = V_{WL} \tag{8}$$

And substitute the term $V_{WL} - V_{th}$ with equation (4), we can obtain:

$$I_{DS0} \propto V_{BL\_min}^2 \tag{9}$$

For a 2T cell, both M1 and M2 are in the linear region and they can be regarded as resistors when:

$$V_{BL} \leq V_{BL\_min} \tag{10}$$

I_{DS} grows with the increase of V_{BL} and become a constant when V_{BL} is higher than V_{BL_min}, so the maximum I_{DS} in the linear region that is the ideal discharge current I_{DS0} in the computation phase can be calculated as:

$$I_{DS0} = \frac{V_{BL\_min}}{R_{M1} + R_{M2}} \tag{11}$$

The equivalent resistance is:

$$R_{Mi} = \frac{1}{\mu_n C_{ox} \frac{W_i}{L_i} (V_{GSi} - V_{th})} \tag{12}$$

We assume the resistance of M1 and M2 is a constant in the whole linear region and the source voltage V_{Si} equals to 0, and we can get the equation of I_{DS0}:

$$I_{DS0} = \frac{V_{BL\_min}}{\frac{1}{\mu_n C_{ox} \frac{W_1}{L_1} (V_{WL} - V_{th})} + \frac{1}{\mu_n C_{ox} \frac{W_2}{L_2} (V_G - V_{th})}} \tag{13}$$

When V_{BL_min} approaches to 0 or the R_{M2} is small due to the large $\frac{W_2}{L_2}$:

$$I_{DS0} \propto V_{BL\_min}^2 \tag{14}$$

When V_{BL_min} is higher:

$$I_{DS0} \propto V_{BL\_min} \tag{15}$$

The ideal discharge current is a constant. However, due to the Channel-length Modulation Effect (Early Effect), the current will increase in the saturation region when the drain-source voltage grows. Therefore, the I_{DS} can be written as:

$$I_{DS} = I_{DS0} + \frac{V_{DS1}(t) - V_{DS1_{sat}}}{r_0} \tag{16}$$

where $V_{DS1}(t) = V_{BL}(t) - V_{S1}$ is the drain-source voltage of M1 and $V_{DS1\_sat} = V_{BL\_min} - V_{S1}$ is the threshold of V_{DS1} between the linear region and the saturation region, the equation can be simplified into:

$$I_{DS} = I_{DS0} + \frac{V_{BL}(t) - V_{BL\_min}}{r_0} \tag{17}$$

in which the r_0 is the equivalent resistance of M1 in the saturation region:

$$r_0 = \frac{V_A}{I_{DS0}} \tag{18}$$

where the V_A is the early voltage and the I_{DS0} is ideal current. The discharge equation is:

$$C_{BL} \frac{dV_{BL}(t)}{dt} + I_{DS} = 0 \tag{19}$$

which is a differential equation and can be written in a standard form:

$$C_{BL}\frac{dV_{BL}(t)}{dt}+\frac{V_{BL}(t)}{r_0}+\left(I_{DS0}-\frac{V_{BL\_min}}{r_0}\right)=0 \tag{20}$$

We can solve the equation and obtain the drop voltage $\Delta V_{BL}(t)$, which is the output voltage and equals to $V_{BL}(0)-V_{BL}(t)$:

$$\Delta V_{BL}(t)=\left[V_{BL}(0)+r_0\left(I_{DS0}+\frac{V_{BL\_min}}{r_0}\right)\right]*\left(1-e^{-\frac{t}{\tau}}\right) \tag{21}$$

where $V_{BL}(0)=V_{DD}$ is the initial condition and $\tau=r_0C_{BL}$ is the RC time-constant of the discharge path. Assuming $\mathrm{t}\ll\tau$, we can simplify the equation with Taylor Series Expansion:

$$\Delta V_{BL}(t)=\frac{I_{DS0}\left(1+\frac{V_{DD}-V_{BL\_min}}{V_A}\right)t}{C_{BL}} \tag{22}$$

The term $I_{DS0}(1+\frac{V_{DD}-V_{BL\_min}}{V_A})$ can be simplified into I_{cell}:

$$\Delta V_{BL}(t)=\frac{I_{cell}t}{C_{BL}} \tag{23}$$

which has the same form of the voltage drop in with ideal constant discharge current. Therefore, if the time of the operation phase is much smaller than τ, the Early Effect has the linear impact on the output signal. Moreover, if there are more than one discharge paths for a Bit Line, this equation is still valid and I_{cell} should multiply the number of the paths.

Furthermore, the input of the MAC, which is the control signal of M1 (V_{WL}), is a Pulse Width Modulation (PWM) signal and the width of V_{WL} represents the value of the input x. The LSB width of V_{WL} (T_{LSB}) can be calculated as:

$$T_{LSB}=\frac{V_{FS}*C_{BL}}{I_{cell}N(2^{N_x}-1)} \tag{24}$$

where *N* is the number of rows in an array and N_x is the input precision.

1.3 Conclusion

The external design parameters are V_{DD}, V_{WL}, V_G, C_{BL}, N, N_x and N_y. Equations of internal parameters can be defined as follows:

The lower edge of the Bit Line voltage:

$$V_{BL\_min}=V_{WL}-V_{th} \tag{25}$$

The full scale of the Bit Line voltage:

$$V_{FS}=V_{DD}-V_{BL\_min} \tag{26}$$

The *LSB* of the output voltage:

$$V_{LSB}=\frac{V_{FS}}{2^{N_y}} \tag{27}$$

The discharge current of a single cell:

$$I_{cell} \propto V_{BL\_min}^2 \tag{28}$$

The minimum width of the input signal:

$$T_{LSB} = \frac{V_{FS} * C_{BL}}{I_{cell} N(2^{N_x} - 1)} \tag{29}$$

2 Performance Trade-off in CIM and PUF

2.1 The Impact of the Process Variation on the Discharge Current

The ideal discharge current I_{cell0} without process variation are with the same value in all cells in an array. However, the process variation leads to the variation in the discharge current. According to the analysis in the previous section, M1 is the saturation current so the equation of the discharge current is:

$$I_{cell0} = \frac{1}{2}\mu_n C_{ox} \frac{W_1}{L_1} (V_{GS1} - V_{th})^2 \left(1 + \frac{V_{BL}(t) - V_{BL\_min}}{V_A}\right) \tag{30}$$

where $V_{BL}(t)$ can be substituted with V_{DD} when $\mathrm{t} \ll \tau$ and the V_{GS1} refers to the gate-source voltage of M1:

$$V_{GS1} = V_{WL} - V_{S1} \tag{31}$$

V_{S1} is 0 in 1T structure. For 2T structure, assuming the width-length ratio of M2 ($\frac{W_2}{L_2}$) is large compare to M1's, R_{M2} is small and therefore, $V_{S1} = V_{D2} = I_{DS} * R_{M2}$ is small. The equation can be simplified into:

$$I_{cell0} = \frac{1}{2}\mu_n C_{ox} \frac{W}{L} (V_{WL} - V_{th})^2 \left(1 + \frac{V_{DD} - V_{BL\_min}}{V_A}\right) \tag{32}$$

Commonly, there is process variation in the channel length (L) and the threshold voltage (V_{th}). V_A is directly proportional to L so the variation of the Early Voltage should be taken in to consideration. We use α to represent the variation part of these parameters.

$$L = L_0(1 + \alpha_L) \tag{33}$$

$$V_A = V_{A0}(1 + \alpha_L) \tag{34}$$

$$V_{th} = V_{th0}\left(1 + \alpha_{V_{th}}\right) \tag{35}$$

in which the L_0, V_{A0} and V_{th0} are the ideal value. Substitute the term of L, V_A and V_{th} with these equations, we can get the discharge current equation with variation:

$$I_{cell} = \frac{1}{2}\mu_n C_{ox} \frac{W}{L_0(1 + \alpha_L)} \left(V_{WL} - V_{th0}\left(1 + \alpha_{V_{th}}\right)\right)^2 \left(1 + \frac{V_{DD} - V_{BL\_min}}{V_{A0}(1 + \alpha_L)}\right) \tag{36}$$

Since $\alpha \ll 1$, Taylor Series Expansion is used:

$$I_{cell} = \frac{1}{2}\mu_n C_{ox} \frac{W}{L_0}(1 - \alpha_L) \left(V_{WL} - V_{th0}\left(1 + \alpha_{V_{th}}\right)\right)^2 \left(1 + \frac{V_{DD} - V_{BL\_min}}{V_{A0}}(1 - \alpha_L)\right) \tag{37}$$

α_L and $\alpha_{V_{th}}$ are two independent zero-mean Gaussian random variables so the sign of α terms can be either a minus or a plus. Expand the equation, neglect the high-order α terms and substitute the first part of the equation with I_{cell0} and simplify other terms:

$$I_{cell} = I_{cell0}\left(1 + \alpha_L + \frac{FS\alpha_L}{V_{A0} + FS} + \frac{2V_{th0}\alpha_{V_{th}}}{V_{BL\_min}}\right) \tag{38}$$

which can be rewritten as:

$$I_{cell} = I_{cell0}(1 + \alpha_I) \tag{39}$$

The α_I, which equals to $\alpha_L + \frac{FS\alpha_L}{V_{A0}+FS} + \frac{2V_{th0}\alpha_{V_{th}}}{V_{BL\_min}}$, is also a zero-mean Gaussian random variable:

$$\alpha_I \sim \mathcal{N}\left(0, \left(\frac{V_{A0} + 2FS}{V_{A0} + FS}\right)^2 \sigma_L^2 + \left(\frac{2V_{th0}}{V_{BL\_min}}\right)^2 \sigma_{V_{th}}^2)\right) \tag{40}$$

which can be simplified into:

$$\alpha_I \sim \mathcal{N}(0, \sigma_I^2) \tag{41}$$

2.2 The Performance Metrics of CIM

1.SNR: The output voltage of a single Bit Line is calculated as follows:

$$V_{OUT} = V_{DD} - \Delta V_{BL} \tag{42}$$

The ideal voltage drop ΔV_{BL0} is used as the output signal:

$$\Delta V_{BL0,j} = \sum_{i=1}^{N} \frac{I_{cell0} \cdot (x_i \cdot T_{LSB}) \cdot w_{i,j}}{C_{BL}} \tag{43}$$

where $\Delta V_{BL0,j}$ means the ideal voltage drop of the *j*th column, x_i refers to the input and $w_{i,j}$ is the weight of the MAC operation. The equation can be written as:

$$\Delta V_{BL0,j} = \frac{I_{cell0} T_{LSB}}{C_{BL}} \sum_{i=1}^{N} x_i \cdot w_{i,j} \tag{44}$$

Generally, the output noise is the thermal noise originated from the resistive element in the discharge path:

$$P_{n\_static} = \frac{kT}{C_{BL}} \tag{45}$$

where the *k* is the Boltzmann Constant and *T* is the thermodynamic temperature. It is necessary to note that the number of discharge path doesn't change the power of the output noise:

When there is only one discharge path as Fig. 2. shows:

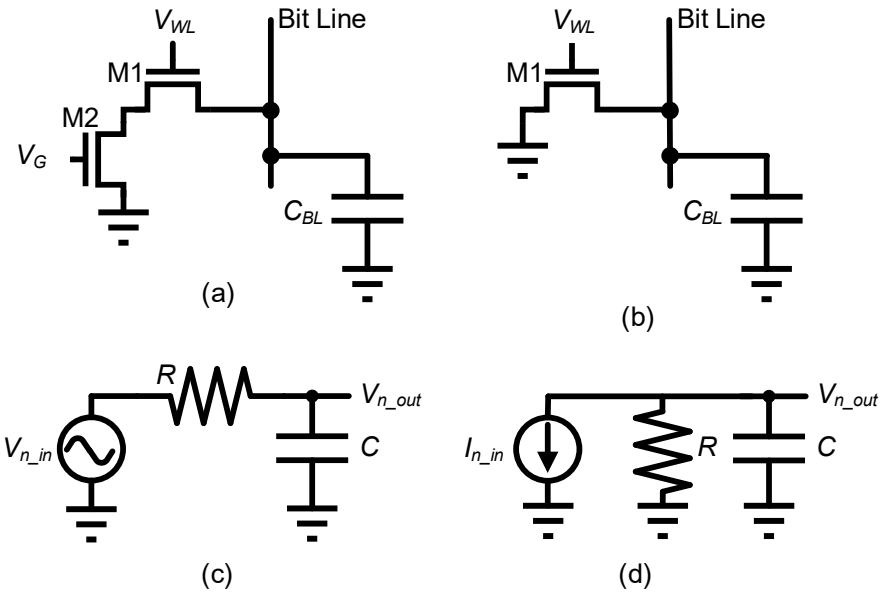

Fig. 2. The circuit of one discharge path (a) (b) and its equivalent model (c) (d).

R is the equivalent resistance of the discharge path. To simplify the analysis, the current form of the noise source is used (Fig. d). The input thermal noise $I_{n\_in}(t)$ is Gaussian noise:

$$I_{n\_in}(t) \sim \mathcal{N}(0, \sigma_n^2) \tag{46}$$

The power spectral density of the input thermal noise (PSD_{n_in}) is:

$$PSD_{n\_in} = \frac{4kT}{R} \tag{47}$$

The output noise power is the integral of the output noise after the channel over all frequencies:

$$P_{n\_out} = \overline{V^2}_{n\_out} = \int_0^{\infty} PSD_{n\_in} \left| \frac{R\frac{1}{sC}}{R + \frac{1}{sC}} \right|^2 df \tag{48}$$

where $\frac{R\frac{1}{sC}}{R+\frac{1}{sC}}$ equals to $\frac{R}{1+j2\pi fRC}$ is the transfer function of the discharge path. The equation can be written as:

$$P_{n\_out} = \int_0^{+\infty} \frac{4kT}{R} \frac{R^2}{1 + (2\pi RC)^2 f^2} df = 4kTR \int_0^{\infty} \frac{1}{1 + (2\pi RC)^2 f^2} df \tag{49}$$

The integral part can be solved with Trigonometric Substitution and we obtain:

$$P_{n\_out} = 4kTR * \frac{1}{4RC} = \frac{kT}{C} \tag{50}$$

When there are $n > 1$ discharge paths, the model is shown in Fig. 3.

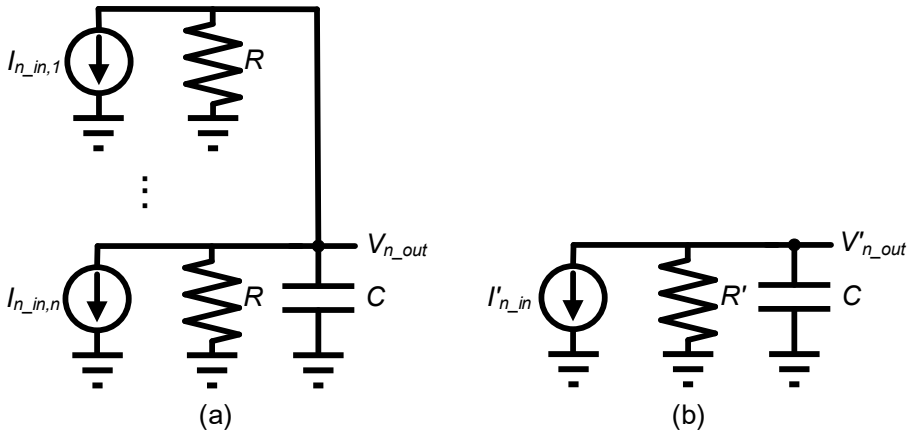

Fig. 3. The equivalent circuit of multiple discharge paths (a) and the simplified model (b).

Different cells are connected on the same Bit Line so the noise sources and resistor are connected in parallel (Fig. 3(a)). The equivalent circuit is shown in Fig. 3(b) and the equivalent input noise $I'_{n\_in}(t)$ is:

$$I'_{n\_in}(t) = \sum_{i=1}^{n} I_{n\_in,i}(t) \tag{51}$$

Thus:

$$I'_{n\_in}(t) \sim \mathcal{N}(0, n\sigma_n^2) \tag{52}$$

According to the Parseval's Theorem, the power in the frequency domain equals to that in the time domain:

$$P_F = \int_{-\infty}^{+\infty} PSDdf = P_T = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{+T} E\{I^2(t)\}dt \tag{53}$$

From this equation we can obtain:

$$PSD \propto E\{I^2(t)\} \tag{54}$$

$E\{I^2(t)\}$ is the average power in the time domain:

$$E\{I'^2_{n\_in}(t)\} = E^2\{I'_{n\_in}(t)\} + D\left(I'_{n\_in}(t)\right) = 0 + n\sigma_n^2 = nE\{I^2_{n\_in}(t)\} \tag{55}$$

Thus:

$$PSD'_{n\_in} = nPSD_{n\_in} \tag{56}$$

The equivalent resistance is:

$$R' = \frac{1}{\sum_{i=1}^{n} \frac{1}{R}} = \frac{R}{n} \tag{57}$$

Than the average power of the output noise is:

$$P'_{n_{out}} = \overline{V'^2_{n_{out}}} = \int_0^{+\infty} PSD'_{n_{in}} \frac{R'^2}{1 + (2\pi R'C)^2 f^2} df = \frac{4kTR}{n} * \frac{n}{4RC} = \frac{kT}{C} \tag{58}$$

This feature can be illustrated that although the increase of the number of the discharge paths raises the PSD of noise, the bandwidth of the path is also changed, which results in the stationary power of the output noise. Thus, the power of the dynamic noise on the Bit Line is kT/C_{BL} and it is very small compared to the signal. Therefore, the main cause accounting for the error of the output is the static noise attributed to the variation of the discharge current:

Because of the process variation, the discharge current of a specific cell is different in a Bit Line:

$$I_{cell,i} = I_{cell0}(1 + \alpha_{I,i}) \tag{59}$$

where $\alpha_{I,i}$ represents the variation part of the current and it is a Gaussian random variable. It causes the mismatch and variation in the output signal:

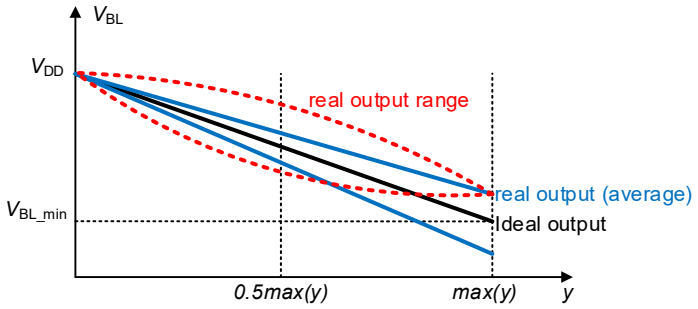

Fig. 4.

As Fig. 4 shows, the black full line is the ideal output signal on a Bit Line. The red curve is the envelop of the real output signal and when the output is half of the max output, the variation is the largest. The blue line is the average of the real output signal, which can be described as the ideal output signal with mismatch on the current. Take a basic model as an instant: $N_x = N_w = 1,\ N = 4$. The discharge current of different cells is different and they are $I_{cell,1}$, $I_{cell,2}$, $I_{cell,3}$ and $I_{cell,4}$. We assume T_{LSB}/C_{BL} is a constant A.

y	0	1	2	3	4
ΔV_{BL0}	0	$1I_{cell0}A$	$2I_{cell0}A$	$3I_{cell0}A$	$4I_{cell0}A$
ΔV_{BL}	0	$I_{cell,1}A$ $I_{cell,2}A$ $I_{cell,3}A$ $I_{cell,4}A$	$I_{cell,1}A + I_{cell,2}A$ $I_{cell,1}A + I_{cell,3}A$ $I_{cell,1}A + I_{cell,4}A$ $I_{cell,2}A + I_{cell,3}A$ $I_{cell,2}A + I_{cell,4}A$ $I_{cell,3}A + I_{cell,4}A$	$I_{cell,1}A + I_{cell,2}A + I_{cell,3}A$ $I_{cell,1}A + I_{cell,2}A + I_{cell,4}A$ $I_{cell,1}A + I_{cell,3}A + I_{cell,4}A$ $I_{cell,2}A + I_{cell,3}A + I_{cell,4}A$	$A\sum_{i=1}^{N} I_{cell,i}$
Count(ΔV_{BL})	$1=\binom{4}{0}$	$4=\binom{4}{1}$	$6=\binom{4}{2}$	$4=\binom{4}{3}$	$1=\binom{4}{4}$
$E(\Delta V_{BL})$	0	$\frac{1}{4}A\sum_{i=1}^{N} I_{cell,i}$	$\frac{2}{4}A\sum_{i=1}^{N} I_{cell,i}$	$\frac{3}{4}A\sum_{i=1}^{N} I_{cell,i}$	$A\sum_{i=1}^{N} I_{cell,i}$
$D(\Delta V_{BL})$	0	$\frac{A^2 \cdot B}{16}$	$\frac{A^2 \cdot B}{12}$	$\frac{A^2 \cdot B}{16}$	0

Table 1.

In the Table 1, the ΔV_{BL0} refers to the black line, ΔV_{BL} refers to the red curve, $E(\Delta V_{BL})$ refers to the blue line and $B = \sum(I_{cell,i} - I_{cell,j})_{i\neq j}{}^2 = {I_{cell0}}^2 \sum(\alpha_{I,i} - \alpha_{I,j})_{i\neq j}{}^2$ is the sum of the square of every two distinct current's difference. In general, since $\binom{N}{0.5N}$ is the largest among all $\binom{N}{i}$, output voltage at 0.5max(y) has the largest variation which can be regarded as static noise. Considering a more general condition:

$$D(\Delta V_{BL})_{y=0.5\max(y)} = \frac{\sum_{i=1}^{N} {\alpha_{I,i}}^2 - \frac{2}{N-1}\sum \alpha_{I,i} \cdot \alpha_{I,j}}{4} \left(\frac{I_{cell0} T_{LSB}}{C_{BL}}\right)^2 \tag{60}$$

$\alpha_{I,i}$ is a random variable, we can use the expect of the $D(\Delta V_{BL})_{y=0.5\max(y)}$ as the average power of noise:

$$P_{n\_static} = E[D(\Delta V_{BL})_{y=0.5\max(y)}] \tag{61}$$

Because of:

$$E\left({\alpha_{I,i}}^2\right) = D\left(\alpha_{I,i}\right) + E\left(\alpha_{I,i}\right)^2 = D\left(\alpha_{I,i}\right) \tag{62}$$

$$E\left(\alpha_{I,i} \cdot \alpha_{I,j}\right) = E\left(\alpha_{I,i}\right) \cdot E\left(\alpha_{I,j}\right) = 0 \tag{63}$$

We obtain:

$$P_{n\_static} = \frac{N\sigma_I^2}{4} \left(\frac{I_{cell0} T_{LSB}}{C_{BL}}\right)^2 \tag{64}$$

When the precision of inputs and weights is not 1, ${V_{n\_static}}^2$ can be written in the form of:

$$P_{n\_static} = f(N, N_x, N_w)\sigma_I^2 \left(\frac{I_{cell0} T_{LSB}}{C_{BL}}\right)^2 \tag{65}$$

Besides, $E(\Delta V_{BL})$ is the average of all the possible output voltage with the same y:

$$E(\Delta V_{BL}) = y * \frac{1}{N} \sum_{i=1}^{N} \frac{I_{cell,i} T_{LSB}}{C_{BL}} \tag{66}$$

This equation can be changed into:

$$E(\Delta V_{BL}) = \Delta V_{BL0} * \left(1 + \frac{1}{N} \sum_{i=1}^{N} \alpha_{I,i}\right) \tag{67}$$

which is combined with the ideal output (ΔV_{BL0}) and the mismatch ($\Delta V_{BL0} * \frac{1}{N}\sum_{i=1}^{N} \alpha_{I,i}$). To summarize, there are 4 kinds of mismatch or noise in the output:

1. The mismatch caused by the Early Effect when $t \ll \tau$ and it is with the same value for all columns.
2. The mismatch caused by the process variation. This mismatch refers to difference between the blue and the black line in Fig. 4. This mismatch is different from column to column and equals to:

$$E(\Delta V_{BL}) - \Delta V_{BL0} = \Delta V_{BL0} * \frac{1}{N}\sum_{i=1}^{N} \alpha_{I,i} \tag{68}$$

3. Dynamic noise (thermal noise):

$$P_{n\_dynamic} = \frac{kT}{C_{BL}} \tag{69}$$

4. Static noise caused by the process variation:

$$P_{n\_static} = f(N, N_x, N_w)\sigma_I^2 \left(\frac{I_{cell0} T_{LSB}}{C_{BL}}\right)^2 \tag{70}$$

The power of signal is:

$$P_{signal} = V_{FS}^2 \tag{71}$$

The SNR can be defined as:

$$SNR = 10\log_{10}\left(\frac{P_{signal}}{P_{n_{static}}}\right) = 20\log_{10}\left[\frac{N(2^{N_{IN}} - 1)}{\sqrt{f(N, N_x, N_w)}\sigma_I}\right] \tag{72}$$

2.Energy: The energy consumption can be defined as the maximum energy drop on the capacitor after the discharge:

$$E = \frac{1}{2} C_{BL} {V_{DD}}^2 - \frac{1}{2} C_{BL} {V_{BL\_min}}^2 \tag{73}$$

3. Delay: The propagation delay of the computation phase is directly proportional to the *LSB* width of input signal, so T_{LSB} can be used as a metrics to evaluate or estimate the delay:

$$T_{LSB} = \frac{V_{FS} * C_{BL}}{I_{cell} N(2^{N_{IN}} - 1)} \tag{74}$$

2.3 The Performance Metrics of PUF

1. The Introduction of the Systematic Variation: The process variation of *L* contains not only random variation but also systematic variation which distributes multivariate Gaussian among all cells in an array. For CIM, the systematic variation can be regarded as part of the random variation because it will not change the way of analysis and the anaylsis reults. But for PUF, the output of the PUF should be fully random and with no determined or predictable part. The systematic variation reduce the randomness (Entropy) of the PUF. Futhermore, the temperature variation is a kind of systematic variation because it is with strong spatial correlation. These two kinds of systematic variation can be described with polynomial and the I_{cell} with sysrematic variation can be described as follows:

$$I_{cell,x,y} = I_{cell0}\left(1 + \alpha_{I,x,y} + f(x,y)\right) \tag{75}$$

α_I is the random variation part and $f(x,y)$ is the systematic variation part:

$$f(x,y) = \mathrm{a}_o + \mathrm{a}_1 x + \mathrm{a}_2 y + \mathrm{a}_3 x^2 + \mathrm{a}_4 xy + \mathrm{a}_5 y^2 + \mathrm{a}_6 x^3 + \mathrm{a}_7 x^2 y + \mathrm{a}_8 xy^2 + \mathrm{a}_9 y^3 \ldots \tag{76}$$

The typical number of $f(x,y)$'s order is 4.

2. The Output of PUF: Voltage or discharge delay of the cell are commonly used as the output of the PUF (PUF_{OUT}). In generall, the differential pair is used to raise the randomness of PUF and thus, two chosen cells are on the same Word Line and their y are the same. For voltage output, the discharge time T_0 is a constant:

$$V_{out} = V_{out,1} - V_{out,2} = \frac{I_{cell,x1,yo}T_0}{C_{BL}} - \frac{I_{cell,x2,y0}T_0}{C_{BL}}$$

$$= \frac{I_{cell0}T_0}{C_{BL}}\left[\left(\alpha_{I,x1,yo} - \alpha_{I,x2,yo}\right) + f(x1) - f(x2)\right] \quad (77)$$

$$f(x) = \mathrm{a}_o{}' + \mathrm{a}_1{}'x + \mathrm{a}_2{}'x^2 + \mathrm{a}_3{}'x^3 \cdots \quad (78)$$

For delay output, the voltage drop ΔV_{BL} is a constant:

$$T_{OUT} = T_{out,1} - T_{out,2} = \frac{\Delta V_{BL}C_{BL}}{I_{cell,x1,yo}} - \frac{\Delta V_{BL}C_{BL}}{I_{cell,x2,yo}}$$

$$= \frac{\Delta V_{BL}C_{BL}}{I_{cell0}}\left[\left(\alpha_{I,x2,yo} - \alpha_{I,x1,yo}\right) + f(x2) - f(x1)\right] \quad (79)$$

These two kind of put can be described in the same way:

$$PUF_{out} = k\left[\left(\alpha_{I,x1,yo} - \alpha_{I,x2,yo}\right) + f(x1) - f(x2)\right] \quad (80)$$

k is a constant and for 1 bit output, when $PUF_{out} > 0$, the output is 1. Otherwise, the output is 0.

$$PUF_{out} \sim \mathcal{N}(k[f(x1) - f(x2)], 2k^2\sigma_I^2) \quad (81)$$

3. The Randomness of PUF: The metric of randomness is the Entropy, which is definded as:

$$H = \sum -p(y_i)\log_2 p(y_i) \quad (82)$$

When the output is 1 bit, $y_i \in \{0,1\}$. When $p(0) = p(1)$, the Entropy is the largest.

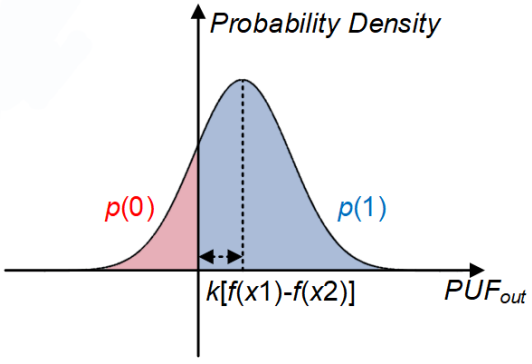

Fig. 5.

As Fig. 5 shows, when $f(x1) - f(x2) = 0$, the distribution of PUF's output is zero-mean Gaussian and $p(0) = p(1)$. When $f(x1) - f(x2) \neq 0$, $p(0) \neq p(1)$ and the Entropy will decrease. We can use the possibilites of 0 or 1 as the metric of the entropy. Assuming $f(x1) - f(x2) > 0$:

$$p(0) = \Phi_{k[f(x1)-f(x2)],\sqrt{2}k\sigma_I}(0) = \Phi\left(\frac{0 - k[f(x1) - f(x2)]}{\sqrt{2}k\sigma_I}\right) = \Phi\left(\frac{-[f(x1) - f(x2)]}{\sqrt{2}\sigma_I}\right) \quad (83)$$

The term $\frac{-[f(x1)-f(x2)]}{\sqrt{2}\sigma_I}$ should approach to 0 and therefore, σ_I is the larger the better. Besides, the two chosen cells should be close to each other to reduce the term $f(x1) - f(x2)$.

4. The Robutness of PUF: The robutness of PUF can be evaluated by Bit Error Rate (BER). The output of the PUF should be a constant when the challenge is the same but the noise in the output voltage will cause the error in the output. To simplify the analysis, assuming two cells are near the each other and the systematic variation of the output is 0:

$$PUF_{out} \sim \mathcal{N}(0, 2k^2\sigma_I^2) \quad (84)$$

Specifically:

$$V_{out} = \frac{I_{cell0}T_0}{C_{BL}}[(\alpha_{I,x1,yo} - \alpha_{I,x2,yo}) \quad (85)$$

$$V_{out} \sim \mathcal{N}\left(0, 2\left(\frac{I_{cell0}T_0}{C_{BL}}\right)^2 \sigma_I^2\right) \quad (86)$$

$$T_{out} = \frac{\Delta V_{BL}C_{BL}}{I_{cell0}}[(\alpha_{I,x2,yo} - \alpha_{I,x1,yo}) \quad (87)$$

$$T_{out} \sim \mathcal{N}\left(0, 2\left(\frac{\Delta V_{BL}C_{BL}}{I_{cell0}}\right)^2 \sigma_I^2\right) \quad (88)$$

The noise is Gaussian noise and because of the differential structure, the distribution of the noise is:

$$V_n \sim \mathcal{N}\left(0, 2\frac{kT}{C_{BL}}\right) \quad (89)$$

$$T_n \sim \mathcal{N}\left(0, 2\frac{\Delta V_{BL}C_{BL}q}{{I_{cell0}}^2}\right) \quad (90)$$

If the output is 1 bit and the threshold is 0, when the output voltage with noise crosses the threshold, the digital output is with error. Thus, the BER is:

$$p(error) = \Phi_{PUF_{OUT},\sqrt{2}\sigma_n}(0) = \Phi\left(\frac{0 - PUF_{out}}{\sqrt{2}\sigma_n}\right) = \Phi\left(\frac{-|PUF_{out}|}{\sqrt{2}\sigma_n}\right) \quad (91)$$

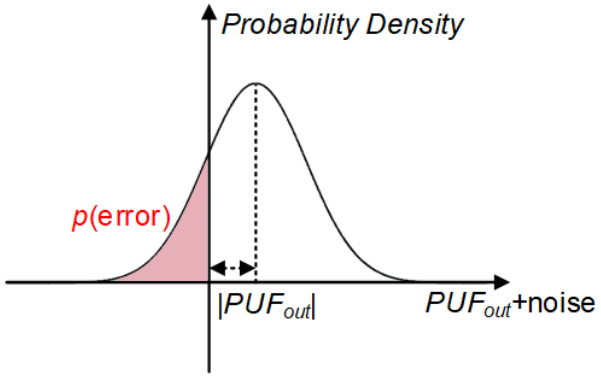

Fig. 6.

Since $p(error)$ should be as small as possible, $-|PUF_{out}|/\sqrt{2}\sigma_n$ should approach to $-\infty$. So $|PUF_{out}|$ should be large and σ_n should be small. We cannot determine the value of PUF_{out} because it is a random variable. We can use $0.05k\sigma_I$ or $0.01k\sigma_I$ to substitute PUF_{out}. According to the characteristics of Gaussian Distribution, 96% of $|PUF_{out}|$ is larger than $0.05k\sigma_I$ and 99.2% of $|PUF_{out}|$ is larger than $0.01k\sigma_I$. We use $0.05k\sigma_I$ to estimate the performance:

$$p(error)_{V_{out}} = \Phi\left(\frac{-0.05\sqrt{2}\frac{I_{cell0}T_0}{C_{BL}}\sigma_I}{\sqrt{2}\sqrt{\frac{kT}{C_{BL}}}}\right) = \Phi\left(\frac{-0.05I_{cell0}T_0\sigma_I}{\sqrt{kTC_{BL}}}\right) \tag{92}$$

From the equation we can find that for voltage type output, raising the discharge time or increasing the current can improve the robustness.

$$p(error)_{T_{out}} = \Phi\left(\frac{-0.05\sqrt{2}\frac{\Delta V_{BL}C_{BL}}{I_{cell0}}\sigma_I}{\sqrt{2}\sqrt{\frac{\Delta V_{BL}C_{BL}q}{{I_{cell0}}^2}}}\right) = \Phi\left(\frac{-0.05\sqrt{\Delta V_{BL}C_{BL}}\sigma_I}{\sqrt{q}}\right) \tag{93}$$

From this equation we can find that for increasing the current has no impact on the robustness for time type output. Raising the voltage drop can improve the performance.