Time-Dependent Random Threshold Voltage Variation Due to Random Telegraph Noise

Gilson Wirth, *Senior Member, IEEE*

***Abstract*— With the downscaling of device dimensions, the variability of metal–oxide–semiconductor field-effect transistor (MOSFET) electrical behavior is produced by factors other than variations in physical dimensions and doping profiles, which are there since device fabrication and remain static over time. Besides these time-zero variability factors, factors that lead to performance variability from one instant in time to the other start playing a significant role. Random telegraph noise (RTN) is among these relevant time-dependent variability sources, causing the threshold voltage (V_T) of a transistor to change from one instant in time to the other. In this work, we extend the knowledge of the time-dependent random variability induced by RTN, by providing a statistical model for RTN-induced threshold voltage variance over time. The area scaling of threshold voltage variance is detailed and discussed, supporting designers in transistor sizing toward a more reliable design. Not only the threshold voltage variance expected in a single transistor is modeled but also its variability among transistors that by design should behave the same way. It is shown that with device size downscaling, the variability of RTN among devices increases faster than its mean value. The relationship between the time domain (RTN) and frequency domain (low-frequency noise) is studied. The modeling provides equations to extract trap amplitude contribution and trap densities from the moments of a measured low-frequency noise distribution or from the moments of a measured V_T variation distribution, without having to characterize individual step heights. Besides analytical modeling, Monte Carlo simulations are run, illustrating the RTN-induced time-dependent random V_T variation and model applicability.**

***Index Terms*— Metal–oxide–semiconductor field-effect transistor (MOSFET) scaling, random telegraph noise (RTN), time-dependent variability, timing jitter.**

I. Introduction

METAL–OXIDE–SEMICONDUCTOR field-effect transistors (MOSFETs) are the workhorses of integrated circuit (IC) design. Its reliable and stable operation is key for the success of semiconductor industry. Variability of MOSFET parameters and intrinsic noise lead to IC yield and reliability issues. Both stochastic variability and noise scale inversely with area. On the other hand, the cost increases with the area, and increasing the area may also increase the capacitive load, what decreases the performance and increases the power consumption. For the designer to be able to find the adequate balance between cost, reliability, and performance, adequate models are demanded.

There are sources of variability that are time-independent, present in the fresh (new) device due to imperfections in fabrication process and the discrete nature of matter. This includes line edge roughness, random dopant fluctuations, and metal (or poly) gate granularity. But there are also time-dependent—time-varying—sources of variability. Random telegraph noise (RTN) is among these time-varying sources of variability [1], [2].

There are well-established models for the time-zero variability. To a first-order, time-zero parameter variability (parameter statistical standard deviation) is considered to be inversely proportional to the square root of the area [1]. This is well-known and widely used by circuit designers. The time-zero (static) variability is not expected to induce jitter of signals. In digital circuits, the RTN chronological statistics, especially trap occupancy switching, has direct impacts on the circuit performance and reliability, as degradations like signal jitter happen when a trap switches state.

In this work, an analytical model to evaluate the RTN-induced threshold voltage variance over time [15] is detailed and extended. To the best of our knowledge, this is the first statistical model derivation of the RTN-induced threshold voltage variance. Properly modeling variability is of paramount relevance, since not only the amplitude of RTN-induced fluctuations increases with area downscaling but also the device-to-device variability increases. Variability increases even faster than the mean value, as shown in this work. This work is extended to show that it is possible to extract both the trap amplitude contributions and trap densities from the moments of a measured low-frequency noise distribution or from the moments of a measured V_T distribution, without having to characterize individual step heights. This is done in an approach similar to bias temperature instability (BTI) [14]. To the best of our knowledge, this is the first proposal to allow extraction of these parameters from both time-domain and frequency-domain noise measurements without having to characterize individual step heights.

II. Time-Domain Analytical Model

The alternate capture and emission of carriers at individual defect sites (charge traps) generates discrete fluctuations in

Manuscript received September 17, 2020; revised November 14, 2020; accepted November 16, 2020. Date of publication December 8, 2020; date of current version December 24, 2020. This work was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). The review of this article was arranged by Editor J. Franco.

The author is with the Department of Electrical Engineering, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre 90040-060, Brazil (e-mail: wirth@inf.ufrgs.br).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TED.2020.3039204.

Digital Object Identifier 10.1109/TED.2020.3039204

MOSFET current. These fluctuations, also called RTN, are the main source of low-frequency noise in deep-submicrometer MOSFETs. Considering the effect of a single trap, the drain current I_D alternates between a higher current state and a lower current state. The difference (fluctuation) in the drain current between states is δI_D. Usually, current fluctuations are measured. The current fluctuation δI_D may be translated into a threshold voltage fluctuation δV_T according to $\delta I_D = g_m \delta V_T$. Similarly, drain-current-referred noise power spectral density (PSD) may be translated into gate-referred voltage noise PSD using $S_{Id} = S_{Vg} g_m^2$ [1], [3], [7], [11], [12].

For compact modeling purposes, the capture and emission of electrons by a charge trap may be modeled as a two-state fluctuation of the threshold voltage V_T. If the trap is empty, it is considered that the device is at a lower V_T (hence higher current). If a charge carrier is trapped, it is considered that the device is at a higher V_T (hence lower current). The difference in V_T between the trapped and empty states of a single trap is then δV_T. Considering the ith trap, δV_{Ti} is the V_T fluctuation due to the ith trap. Please note that for traps that lead to a higher V_T if occupied (and hence a lower V_T if empty), the notation for the states may be exchanged, without any loss of generality and model validity.

The value of the V_T fluctuation due to all traps at time t, here called $\Delta V_T(t)$, is then evaluated as the sum of the contribution of all N_{tr} traps found in a device that are active in the time window of interest, producing RTN:

$$\Delta V_T(t) = \sum_{i=1}^{\mathrm{Ntr}} \delta V_{Ti} S_i(t). \tag{1}$$

For simplicity and consistency with noise, we make the mean (expected) value of ΔV_T equal to zero, $E[\Delta V_T(t)] = 0$. Noise is assumed to have a mean value of zero. This is also convenient to lead to an elegant resolution for the expected V_T variance over time in a device. Zero mean value is achieved by making the mean value of the V_T fluctuation due to each trap equal to zero, by writing $S_i(t)$ as being

$$S_i(t) = \begin{cases} -\dfrac{1}{1+\beta_i}, & \text{if the trap is empty} \\ +\dfrac{\beta_i}{1+\beta_i}, & \text{if the trap is occupied.} \end{cases} \tag{2}$$

where $\beta_i = \tau_{Ci}/\tau_{Ei}$, with τ_{Ci} and τ_{Ei} being the capture and emission time constants, respectively.

Please note that this notation does not change the waveform of the RTN. Since the mean value the trap keeps empty is $\tau_{Ci}/(\tau_{Ci}+\tau_{Ei}) = \beta_i/(1+\beta_i)$ and the mean value the trap keeps occupied is $\tau_{Ei}/(\tau_{Ci}+\tau_{Ei}) = 1/(1+\beta_i)$, it just removes the mean value (dc component) of the waveform, for any value of β_i. The amplitude of $\Delta V_T(t)$ induced by the ith trap keeps being δV_{Ti}, since $\tau_{Ci}/(\tau_{Ci}+\tau_{Ei}) + \tau_{Ei}/(\tau_{Ci}+\tau_{Ei}) = \beta_i/(1+\beta_i) + 1/(1+\beta_i) = 1$. See Fig. 1, where this is exemplified. The mean value (dc component) of the waveform does not contribute neither to V_T variance over time nor to noise. Hence, it is appropriate to remove it, besides facilitating the statistical analysis.

As seen in Fig. 1 and described by (1) and (2) above, V_T is not constant, but it randomly fluctuates over time.

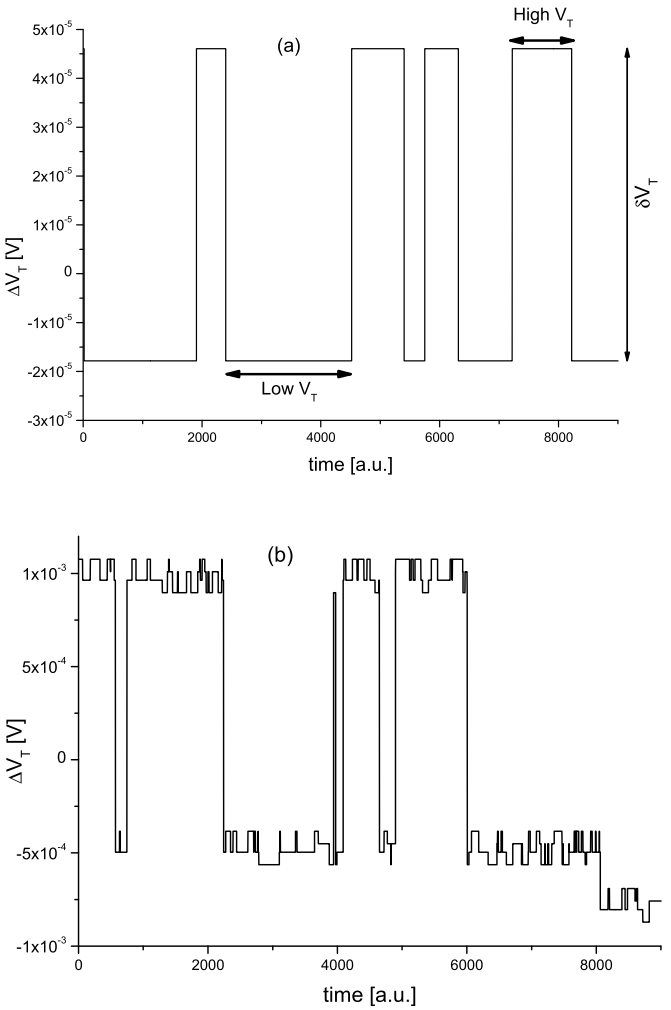

Fig. 1. MC simulation of threshold voltage variation over time $\Delta V_T(t)$ due to RTN. Trap occupancy switching leads to discrete fluctuations in V_T. The first 9000 points of the MC runs for two nMOSFETs of size 0.16 μm $\times$ 0.13 μm are shown. Due to random number of traps and related parameters, devices that by design should be identical show different V_T variance over time. Note that the formulation used here does lead to zero mean value, without changing the waveform.

To properly model this behavior, the variance of V_T taken over time, $\mathrm{Var}[\Delta V_T(t)]$, is evaluated. To simplify the notation, $\mathrm{Var}[\Delta V_T(t)]$ will be named V_{TVar}. Furthermore, due to random number of traps and random trap parameters, V_{TVar} may be different in devices that by design should be identical. Hence, it is also needed to model how V_{TVar} varies among devices. This is achieved by modeling $\mathrm{Var}[V_{TVar}]$, the variance of V_{TVar} among devices.

The variance of the threshold voltage fluctuation in a device taken over time is evaluated starting from

$$\mathrm{Var}[\Delta V_T(t)] = E\left[\Delta V_T(t)^2\right] - E[\Delta V_T(t)]^2 \tag{3}$$

with $E[\Delta V_T(t)] = 0$.

For evaluating $E[\Delta V_T(t)^2]$, we note that for a single trap

$$E\left[(\delta V_{Ti} S_i(t))^2\right] = E\left[\delta V_{Ti}{}^2 \cdot \frac{\beta_i}{(1+\beta_i)^2}\right] \equiv E\left[A_i^2\right]. \tag{4}$$

The amplitude contribution A_i of the ith trap to the variance of V_T is the same as its contribution to low-frequency noise.

See equation (40) in [3] and the discussion in Section III of this work.

$E[\Delta V_T(t)^2]$ is then evaluated as

$$E\left[\Delta V_T(t)^2\right] = E\left[\sum_{i=1}^{Ntr}(\delta V_{Ti} S_i(t))^2\right] = E[N_{tr}]E\left[A_i^2\right] \quad (5)$$

which leads to

$$\mathrm{Var}[\Delta V_T(t)] = E[N_{tr}]E\left[A_i^2\right] \equiv E[V_{Tvar}] \quad (6)$$

where $E[N_{tr}] = n$ is the mean number of traps per device active in the time window of interest, considering an ensemble of geometrically identical devices, as discussed in Section III.

Due to the random trap activity, at each time instant the threshold voltage may be different. Equation (6) describes the expected variance of the threshold voltage taken over time in a single device, and we name it V_{Tvar}, relating it to the time-varying V_T due to RTN. It may be seen as a jitter of the threshold voltage, with V_T randomly fluctuating around its mean value.

Equation (6) is for the expected value of the V_T variance over time in a transistor, that is, the expected (mean) threshold voltage variation over time considering an ensemble of geometrically identical devices. Different transistors will have different number of traps N_{tr}, with different amplitude contributions A_i, and V_{Tvar} will vary among transistors that by design should be identical. Some transistors may have small amplitude traps—or even no traps—while others may have traps with large amplitude. Hence, it is also important to evaluate the variability of the expected value of V_{Tvar} among geometrically identical transistors. Each individual transistor may show V_{Tvar} different from its counterparts.

The variance of V_{Tvar} in an ensemble of devices may be evaluated using

$$\mathrm{Var}[V_{Tvar}] = E\left[V_{Tvar}^2\right] - E[V_{Tvar}]^2. \quad (7)$$

Assuming that the number of traps is Poisson-distributed, it is evaluated as being

$$\mathrm{Var}[V_{Tvar}] = E[N_{tr}]E\left[A_i^4\right]. \quad (8)$$

The Poisson distribution for the number of traps has been widely reported in the literature [14], [16]–[20].

III. Relation to 1/f Noise and Area Scaling

To detail the area scaling of V_{Tvar} and $\mathrm{Var}[V_{Tvar}]$, we must look at the area dependence of n and A_i. These are parameters well-studied in the literature. The mean number of traps is known to be proportional to area, $n \sim WL$. The mean amplitude due to a trap is known to scale inversely with area, $A_i \sim 1/WL$ [1], [3], [7]–[10]. From (6), the following relation between the device area and the mean (expected) V_{Tvar} can then be written

$$E[V_{Tvar}] = \mathrm{Var}[\Delta V_T(t)] \sim 1/WL. \quad (9)$$

For the variation in V_{Tvar} among devices, from (8) the following relation between device area and V_{Tvar} variability among devices can be written

$$\mathrm{Var}[V_{Tvar}] \sim 1/(WL)^3. \quad (10)$$

This means that with the area scaling not only the expected value of V_{Tvar} increases but also the variability of V_{Tvar} among devices increases. This increasing time-dependent random variability is a significant challenge for the device designer.

If the charge trapping and de-trapping is the dominant source of 1/f noise, the expected value of V_{Tvar} is related to the mean value of the 1/f noise power. Equation (6) has the same form and same area dependence as (34) in [3], which is replicated here for reader's convenience

$$E[S(f)] = \frac{E\left[A_i^2\right]N_{dec}WL}{f}\frac{\pi}{2}. \quad (11)$$

The trap density per unit area and frequency decade is $(N_{dec} ln 10)$ [3]. A frequency decade is the interval between two frequencies having a ratio of 10 to 1. Hence, the mean number of traps per device is [3]

$$E[N_{tr}] = n = N_{dec}WL\ln(10)\log\left(\frac{f_{max}}{f_{min}}\right) \quad (12)$$

where f_{min} and f_{max} are the lower and upper limits of the frequency window of interest, respectively.

Since we assume that the traps producing RTN and 1/f noise are the same, the equivalent equation in the time domain is

$$E[N_{tr}] = n = N_{dec}WL\ln(10)\log\left(\frac{t_{max}}{t_{min}}\right) \quad (13)$$

where t_{min} and t_{max} are the lower and upper limits of the time window of interest, respectively.

Both (6) and (11) are proportional to $E[A_i^2]$ and proportional to the mean number of traps in a given device size. The same area dependence is seen in (9) above and (35) in [3]. This is expected since the mean value of V_{Tvar} (and the mean value of phase noise) due to RTN is related to the mean 1/f noise due to RTN. Also, similar to 1/f noise (frequency domain), traps that have maximum contribution to V_{Tvar} (time domain) are the ones with capture time similar to the emission time, that is, $\beta \approx 1$.

Similarly, the variability of V_{Tvar} is related to the variability of the 1/f noise power. Equation (8) has the same form and same area dependence as (38) in [3], which is also replicated here for reader's convenience

$$\mathrm{Var}[S(f)] = \frac{E\left[A_i^4\right]N_{dec}WL}{2}\frac{1}{f^2}. \quad (14)$$

Both (8) above and (38) in [3] are proportional to $E[A_i^4]$ and proportional to the mean number of traps in a given device size. The same area dependence is seen in (10) above and (38) in [3]. This is expected, since the variability of V_{Tvar} (and phase noise variability) due to RTN is related to the variability of 1/f noise due to RTN.

Hence, as seen in 1/f noise, not only the mean value of V_{Tvar} is expected to increase with device area downscaling but also the scattering of V_{Tvar} among devices strongly increases with device downscaling. Each device has a random number of traps with random amplitudes and random time constants. The variation in V_T over time in a device is given by its particular number of traps and related parameters. The number of traps and related parameters vary among devices. The smaller the device size, the larger the RTN performance varies among devices that by design should be identical.

Now we address the parameter extraction using model equations. For both the time domain and frequency domain, the model parameters are trap amplitude contribution and trap density. These parameters can be extracted from measured distributions, either in the time domain or frequency domain.

From (11) and (14)

$$\frac{\mathrm{Var}[S(f)]}{E[S(f)]} = \frac{E\left[A_i^4\right]}{E\left[A_i^2\right]\pi f} \tag{15}$$

which does not depend on the trap density. It depends solely on the trap amplitude contribution.

A similar relation may be derived from the time-domain model. From (6) and (8)

$$\frac{\mathrm{Var}[V_{\mathrm{Tvar}}]}{E[V_{\mathrm{Tvar}}]} = \frac{E\left[A_i^4\right]}{E\left[A_i^2\right]} \tag{16}$$

which also depends solely on the trap amplitude contribution.

This implies that if the shape of the distribution is known, both $E[A_i]$ and trap densities may be extracted either from the moments of a measured low-frequency noise distribution or from the moments of a measured V_{Tvar} distribution, without having to characterize individual step heights.

Also of interest is that once the parameters are extracted from the frequency-domain measurement (low-frequency) noise, they may be used for modeling in the time domain (V_{T} variance over time). The reciprocal also holds: parameters extracted from the time-domain measurements may be used for modeling in the frequency domain.

This is similar to the approach used in BTI [14]. Considering $\Delta V_{T,\mathrm{BTI}}(t)$ to be the V_{T} shift due to BTI, the following equation may be derived [14]:

$$\frac{\mathrm{Var}\left[\Delta \boldsymbol{V_{T,BTI}(t)}\right]}{E\left[\Delta \boldsymbol{V_{T,BTI}(t)}\right]} = \left(\frac{E\left[\delta V_{Ti}^{\boldsymbol{2}}\right]}{E[\delta V_{Ti}]}\right). \tag{17}$$

Assuming the threshold voltage fluctuation (trap amplitude) δV_{Ti} to be exponentially distributed leads to [14]

$$E[\delta V_{\mathrm{Ti}}] = \frac{\mathrm{Var}\left[\Delta \boldsymbol{V_{T,BTI}(t)}\right]}{2E\left[\Delta \boldsymbol{V_{T,BTI}(t)}\right]}. \tag{18}$$

To allow the extraction of trap amplitude contribution A_i and trap density from low-frequency noise data or time-domain V_{T} variance data, it is also needed to know the shape of the distribution of A_i. If it is assumed that A_i is exponentially distributed, and noting that the moments of an exponentially distributed random variable X are given by $E[X^n] = n!/\lambda^n$, with $E[X] = 1/\lambda$ being the parameter of the exponential distribution [6], leads to

$$\frac{\mathrm{Var}[S(f)]}{E[S(f)]} = \frac{12E[A_i]^2}{\pi f}. \tag{19}$$

Allowing the extraction of $E[A_i]$ as being

$$E[A_i] = \sqrt{\frac{\pi f \mathrm{Var}[S(f)]}{12E[S(f)]}}. \tag{20}$$

This extraction of $E[A_i]$ was not done in [3] because there was no assumption made regarding the distribution of A_i.

Once $E[A_i]$ is determined, the trap density may be evaluated as

$$N_{\mathrm{dec}}\mathrm{WL} = \frac{E[S(f)]f}{E[A_i]^2}\frac{1}{\pi}. \tag{21}$$

And N_{tr} is calculated using (12). Parameter extraction from the experimental 1/f noise $S(f)$ is presented in Section IV.

Making the same assumption in the time domain–that A_i is exponentially distributed—leads to

$$\frac{\mathrm{Var}[V_{T\mathrm{var}}]}{E[V_{\mathrm{Tvar}}]} = \frac{E\left[A_i^4\right]}{E\left[A_i^2\right]} = 12E[A_i]^2. \tag{22}$$

Leading to

$$E[A_i] = \sqrt{\frac{\mathrm{Var}[V_{T\mathrm{var}}]}{12E[V_{\mathrm{Tvar}}]}}. \tag{23}$$

Once $E[A_i]$ is extracted, $E[A_i^2]$ can be evaluated using the properties of the exponential distribution, and the mean number of traps $E[N_{\mathrm{tr}}] = n$ may then be extracted using (6), which for an exponential distribution becomes $E[V_{\mathrm{Tvar}}] = 2E[N_{\mathrm{tr}}]E[A_i]^2$.

This means that for each device, the threshold voltage (V_{T}) is measured over time. For each device, the variance of V_{T} (V_{Tvar}) is evaluated in the measured time interval. The average value of V_{Tvar} of the ensemble of devices is calculated. This is $E[V_{\mathrm{Tvar}}]$. The variance of V_{Tvar} among the different devices is calculated. This is $\mathrm{Var}[V_{\mathrm{Tvar}}]$. Once $E[V_{\mathrm{Tvar}}]$ and $\mathrm{Var}[V_{\mathrm{Tvar}}]$ are obtained, $E[A_i]$ and $E[\mathrm{N_{tr}}]$ are obtained using (23) and (6), respectively. Due to lack of experimental data, parameter extraction from experimental threshold voltage fluctuation over time is not done in this work.

The time-domain analytical model derived in Section II and the frequency-domain model derived in [3] are valid for any statistical distribution of the V_{T} fluctuation induced by a single trap (no particular distribution is assumed in model derivation). Also, the area scaling does not depend on the shape of A_i distribution. Equations (15) and (16) are also valid for any distribution of A_i. However, to do parameter extraction, the shape of the distribution of A_i needs to be known, since the moments of A_i will be different for different distributions. The assumption of A_i being exponentially distributed leads to (19) to (23), used for parameter extraction in Section IV. If the assumption of A_i being exponentially distributed does not agree with the experimental data, the moments of A_i may be written as a function of the parameters of the distribution of interest, as discussed in Section IV.

IV. Experimental Results and Monte Carlo Simulations

The parameters trap amplitude contribution and trap density are extracted from the frequency-domain experimental results previously published in [3] for a 0.13-μm technology node, where charge trapping was found to be the major low-frequency noise source. The mean value of the gate-referred noise PSD $E[S_{\mathrm{VG}}]$ and the variance $\mathrm{Var}[S_{\mathrm{VG}}]$ are evaluated from the measured samples in the frequency range from 1 to 10 kHz. In this range, the noise PSD was measured

at 160 different frequencies f, approximately equally spaced in log scale. At each measured frequency f, the mean trap amplitude contribution $E[A_i]$ is extracted using (20), and the trap density is extracted using (21).

The experimental results for the devices of largest measured sample sizes are reported. The extracted values are as follows:

$E[A_i]$ equal to 9.8E-5 $\pm$ 2.4E-5 V and N_{dec}WL ln (10) equal to 0.53 $\pm$ 0.18 traps per frequency decade, for the 0.16 μm $\times$ 0.13 μm nMOSFETs, where 42 transistors were measured.

$E[A_i]$ equal to 2.1E-6 $\pm$ 0.62E-6 V and $N_{\text{dec}}WL$ ln (10) equal to 25.8 $\pm$ 9.2 traps per frequency decade, for the 10 μm $\times$ 0.13 μm nMOSFETs, where 27 transistors were measured.

$E[A_i]$ equal to 1.2E-4 $\pm$ 4.3E-5 V and $N_{\text{dec}}WL$ ln (10) equal to 1.43 $\pm$ 0.58 traps per frequency decade, for the 0.16 μm $\times$ 0.13 μm pMOSFETs, where 21 transistors were measured.

After parameter extraction from the frequency-domain experimental results, the frequency-domain Monte Carlo (MC) simulations were performed as follows:

The number of traps is assumed to be Poisson-distributed, with the mean number of traps in a device extracted from the frequency-domain measurements. A frequency window of 10^5 Hz is simulated, assuming $N_{\text{tr}} = N_{\text{dec}}$ WL ln(10) log(10^5). Hence, considering the values extracted from the frequency-domain, $E[N_{\text{tr}}]$ is 2.65 for 0.16 μm $\times$ 0.13 μm nMOSFETs, 129 for 10 μm $\times$ 0.13 μm nMOSFETs, and 7.15 for 0.16 μm $\times$ 0.13 μm pMOSFETs.

The corner frequency f_i of each trap in each device is chosen by lot according to a uniform distribution in log scale [3].

The amplitude contribution A_i of a single trap is a random variable, exponentially distributed, being the mean amplitude contribution extracted from the frequency-domain measurements. As extracted from the experimental data, the mean amplitude contributions is 9.8E-5 V for 0.16 μm $\times$ 0.13 μm nMOSFETs, 2.1E-6 V for 10 μm $\times$ 0.13 μm nMOSFETs, and 1.2E-4 V for 0.16 μm $\times$ 0.13 μm pMOSFETs.

For each device, $S(f)$ is calculated as $S(f) = \sum_{i=1}^{Ntr} A_i^2(1/f_i)(1/(1+(f/f_i)^2))$, which is (18) in [3].

Fig. 2 shows the quantile–quantile (Q–Q) plots of the experimental gate-referred noise PSD quantiles in log scale, $\log(S_{\text{VG}}(f))$ in V^2/Hz at a frequency of 10 Hz, against the quantiles of $\log(S_{\text{VG}}(f))$ at 10 Hz from MC simulations. A good agreement between the experimental data and MC simulations results is seen, showing that the model is appropriate and that the exponential distribution for A_i is a good assumption.

The exponential distribution has only one parameter and is uniquely determined by its moments, hence, being possible to determine its parameter analytically using (20)—frequency domain—or (23)—time domain, *without any fitting*, as done here. The exponential distribution has been widely reported in the literature [14], [16]–[20]. If the exponential distribution does not yield a good agreement with the experimental data, other distributions may be examined. The exponential distribution is a special case of the Weibull distribution with the shape parameter $k = 1$ [6]. Finer tuning (fitting) can be tried allowing the shape parameter of the Weibull distribution to vary. The lognormal distribution has also been widely reported in the literature [21]–[23]. Both lognormal and Weibull distributions have two parameters. While increasing the number of parameters may be favorable to bring data and

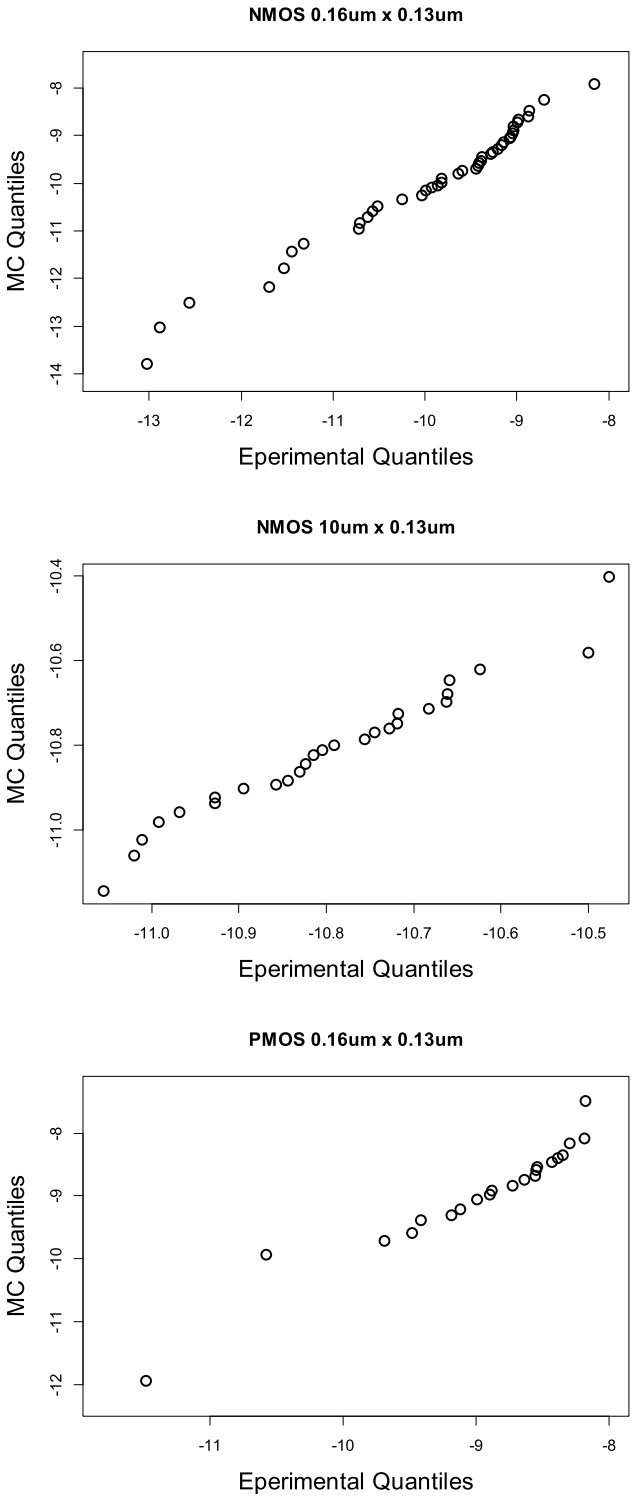

Fig. 2. Q–Q plots of the experimental gate-referred noise PSD quantiles in log scale, $\log(S_{\text{VG}}(f))$, at the frequency of 10 Hz against the quantiles of $\log(S_{\text{VG}}(f))$ at 10 Hz from MC simulations.

model into agreement, the solution using (15) or (16) will not be unique anymore, and fitting will be needed. Equations (15) and (16) relate the measured $S(f)$ or V_{TVar} solely to A_i (it does not depend on other parameters) and are valid for any distribution of A_i. The moments of A_i may be written as a function of the parameters of the candidate distribution, and fitting performed. For instance, in the case of the lognormal distribution, (15) may be used to relate the two lognormal distribution parameters (μ and σ) to the moments of A_i, and fitting performed to properly fit the different experimental quantiles to the theoretical ones, determining the distribution parameters.

MC simulations were also run in the time-domain to confirm the behavior predicted by the analytical model presented in Section III, and the results for the nMOSFETs are presented and discussed. The time-domain MC simulations were run using the parameters extracted from the frequency-domain and assuming that:

Charge trapping and de-trapping are the stochastic events governed by the characteristic time constants, which are uniformly distributed on a log scale.

The number of traps is assumed to be Poisson-distributed, with the mean number of traps in a device extracted from the frequency-domain measurements. A time window of 10^5 s is simulated, assuming an $N_{\mathrm{tr}} = N_{\mathrm{dec}}$ WL $\ln(10^5)$. Hence, considering the values extracted from the frequency domain, $E[N_{\mathrm{tr}}]$ is 2.65 for the 0.16 μm $\times$ 0.13 μm nMOSFETs, 129 for the 10 μm $\times$ 0.13 μm nMOSFETs, and 7.15 for the 0.16 μm $\times$ 0.13 μm pMOSFETs.

The amplitude contribution A_i of a single trap is a random variable, exponentially distributed, being the mean amplitude contribution extracted from the frequency-domain measurements. As extracted from the experimental data, the mean amplitude contribution is 9.8E-5 V for the 0.16-μm $\times$ 0.13 μm nMOSFETs, 2.1E-6 V for the 10-μm $\times$ 0.13 μm nMOSFETs, and 1.2E-4 V for the 0.16-μm $\times$ 0.13 μm pMOSFETs.

For each device size, 1000 MC simulations are run. At each simulation time step, the trap switching probability (capture or emission of a charge carrier by a trap) is evaluated according to the trap capture or emission time constant. As an example, Fig. 1 shows the first 9000 points of the MC runs of two nMOSFETs with WL of 0.16 μm $\times$ 0.13 μm. The device in Fig. 1(b) shows a much larger V_{T} fluctuation than the one in Fig. 1(a), illustrating the variability among devices that by design should behave identically. For the transistor in Fig. 1(a), the threshold voltage variance is $V_{\mathrm{Tvar}} = 8.21\text{E-}10\ V^2$, which means a standard deviation of V_{T} over time of 2.86E-5 V. For the transistor in Fig. 1(b), the threshold voltage variance is $V_{\mathrm{Tvar}} = 5.20\text{E-}7\ V^2$, which means a standard deviation of V_{T} over time of 7.21E-4 V. This variation in V_{T} over time could be seen as jitter of the threshold voltage, randomly varying around its mean value.

The 1000 MC runs for the 10 μm $\times$ 0.13 μm nMOSFETs yielded a mean value of $V_{\mathrm{Tvar}} = 1.12\text{E-}9\ V^2$, and $\mathrm{Var}[V_{\mathrm{Tvar}}] = 6.27\text{E-}20\ \mathrm{V}^4$. The analytical results from (6) and (8) are 1.14E-9 V^2 and 6.02E-20 V^4, respectively. The 1000 MC runs for the 0.16-μm $\times$ 0.13 μm nMOSFETs yielded a mean value of $V_{\mathrm{Tvar}} = 5.16\text{E-}8\ \mathrm{V}^2$, and $\mathrm{Var}[V_{\mathrm{Tvar}}] = 5.44\text{E-}15\ V^4$. The analytical results from (6) and (8) are 5.09E-8 V^2 and 5.87E-15 V^4, respectively. The 1000 MC runs for the 10 μm $\times$ 0.13 μm pMOSFETs yielded a mean value of $V_{\mathrm{Tvar}} = 2.01\text{E-}7\ V^2$, and $\mathrm{Var}[V_{\mathrm{Tvar}}] = 3.23\text{E-}14\ \mathrm{V}^4$. The analytical results from equations (6) and (8) are 2.06E-7 V^2 and 3.56E-14 V^4, respectively. The histograms in Fig. 3 show the MC simulation results for the nMOSFETs. Due to random number of traps, random trap amplitude, and random trap activity, the variance of V_{T} taken over time (V_{Tvar}) is different for each device. Black diamonds show the mean V_{T} variance obtained in the MC simulations for each device size. The error bars have a total length equal to two times $(\mathrm{Var}[V_{\mathrm{Tvar}}])^{0.5}$ of each device size, as obtained from the MC simulations. Please note that for the 0.16 μm $\times$ 0.13 μm nMOSFETs, the error bar is larger than the mean value. Due to the area scaling of $\mathrm{Var}[V_{\mathrm{Tvar}}]$, in Fig. 3 the error bar increases rapidly as the area becomes smaller. With nMOSFET area decreasing from 10 μm $\times$ 0.13 μm to 0.16 μm $\times$ 0.13 μm, the mean value of V_{Tvar} increased from 1.12E-9 V^2 to 5.16E-8 V^2, while the variance of V_{Tvar} increased from 6.02E-20 V^4 to 5.87E-15 V^4. The variance increases much faster than the mean value, leading

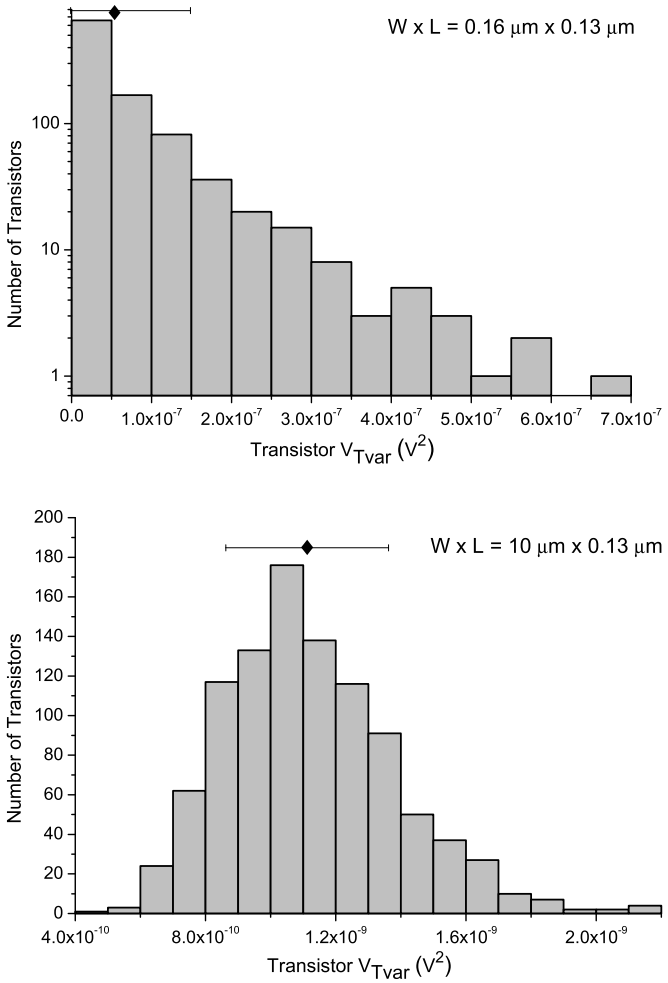

Fig. 3. Histogram of V_{Tvar} from MC simulations for two different nMOSFET sizes. For each size, 1000 devices were simulated. Black diamonds show the average V_{Tvar} for each device size. The error bars have a total length equal to two times $(\mathrm{Var}[V_{\mathrm{Tvar}}])^{0.5}$ of each device size. Please note that the *y*-axis of the first histogram is in log scale.

to the larger spread seen in the histogram of Fig. 3. For the small area device, the spread of the x-axis in Fig. 3 is much larger than for the large area device, showing that with area downscaling the probability of obtaining extreme values increases much faster than the mean value. This behavior has been experimentally observed, but analytical modeling was lacking. The RTN distributions clearly show a nonGaussian distribution, with variation increasing as the device area scales down, as observed, for example, in [2] and [4]. Furthermore, it is seen to be a-heavy tailed distribution, as observed, for example, in [2] and [5].

Equation (8) clearly shows that with area downscaling, the variance increases faster than the mean value—variance increases with $1/(WL)^3$. For ICs, this may lead to increased jitter of signals with device size downscaling. In [13], the authors observed that by slightly increasing the transistor size, more than 50% reduction of the ring oscillator frequency uncertainty can be achieved. The ring oscillator frequency uncertainty was related to the transistor delay uncertainty due to RTN.

In such situation, even if the mean value of the V_T fluctuation over time may not be considered an issue, its increasing variability with device downscaling may be a challenge. Besides becoming a potential reliability issue, such device-to-device variations may require repeating tests and characterization procedures over many devices to obtain the statistical properties, becoming time-consuming and costly. Hence, proper statistical modeling is highly demanded.

V. Conclusion

We extend the knowledge of the time-dependent random variability induced by RTN, by providing an analytical model for the threshold voltage variance V_{Tvar} produced by RTN. We addressed not only the mean (expected) value but also its variability among devices. The area scaling of RTN-induced V_T variance and its variability among devices is detailed and discussed, supporting designers in transistor sizing toward a more reliable design. We discuss the behavior in the time domain (RTN) and frequency (1/f noise) domain, showing the relationship between models and parameters. Similar to the case of 1/f noise, with area downscaling the device-to-device variability of V_{Tvar} increases faster than its mean value. MC simulations are run, illustrating the RTN-induced V_T variability and model applicability. We provide equations to extract the model parameters noise amplitude contribution and trap density without the need to measure individual step heights.

References

[1] M. Simicic, P. Weckx, B. Parvais, P. Roussel, B. Kaczer, and G. Gielen, "Understanding the impact of time-dependent random variability on analog ICs: From single transistor measurements to circuit simulations," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 27, no. 3, pp. 601–610, Mar. 2019, doi: 10.1109/TVLSI.2018.2878841.

[2] T. Matsumoto, K. Kobayashi, and H. Onodera, "The impact of RTN-induced temporal performance fluctuation against static performance variation," in *Proc. IEEE Electron Devices Technol. Manuf. Conf. (EDTM)*, Feb. 2017, pp. 31–32, doi: 10.1109/EDTM.2017.7947496.

[3] G. I. Wirth, J. Koh, R. daSilva, R. Thewes, and R. Brederlow, "Modeling of statistical low-frequency noise of deep-submicrometer MOSFETs," *IEEE Trans. Electron Devices*, vol. 52, no. 7, pp. 1576–1588, Jul. 2005, doi: 10.1109/TED.2005.850955.

[4] N. Tega *et al.*, "Impact of HK / MG stacks and future device scaling on RTN," in *Proc. Int. Rel. Phys. Symp.*, Apr. 2011, pp. 6A.5.1–6A.5.6, doi: 10.1109/IRPS.2011.5784546.

[5] S. Dongaonkar, M. D. Giles, A. Kornfeld, B. Grossnickle, and J. Yoon, "Random telegraph noise (RTN) in 14nm logic technology: High volume data extraction and analysis," in *Proc. IEEE Symp. VLSI Technol.*, Jun. 2016, pp. 1–2, doi: 10.1109/VLSIT.2016.7573424.

[6] L. G. Parrat, *Probability and Experimental Errors on Science*. New York, NY, USA: Wiley, 1961.

[7] M. J. Deen, "Low-frequency noise in semiconductor devices–state-of-the-art and future perspectives plenary paper," in *Proc. Int. Conf. Noise Fluctuations (ICNF)*, Jun. 2017, pp. 1–4, doi: 10.1109/ICNF.2017.7985925.

[8] H. Reisinger, "The time-dependent defect spectroscopy," in *Bias Temperature Instability for Devices and Circuits*, T. Grasser, Ed. New York, NY, USA: Springer, 2014, pp. 75–109.

[9] N. Ashraf, D. Vasileska, G. Wirth, and P. Srinivasan, "Accurate model for the threshold voltage fluctuation estimation in 45-nm channel length MOSFET devices in the presence of random traps and random dopants," *IEEE Electron Device Lett.*, vol. 32, no. 8, pp. 1044–1046, Aug. 2011, doi: 10.1109/LED.2011.2158287.

[10] S. M. Amoroso *et al.*, "Investigation of the RTN distribution of nanoscale MOS devices from subthreshold to on-state," *IEEE Electron Device Lett.*, vol. 34, no. 5, pp. 683–685, May 2013, doi: 10.1109/LED.2013.2250477.

[11] E. Simoen, R. Ritzenthaler, T. Schram, H. Arimura, N. Horiguchi, and C. Claeys, "On the low-frequency noise of high-κ gate stacks: What did we learn?" in *Proc. 14th IEEE Int. Conf. Solid-State Integr. Circuit Technol. (ICSICT)*, Oct. 2018, pp. 1–4, doi: 10.1109/ICSICT.2018.8565820.

[12] N. Mavredakis, N. Makris, P. Habas, and M. Bucher, "Charge-based compact model for bias-dependent variability of $1/f$ noise in MOSFETs," *IEEE Trans. Electron Devices*, vol. 63, no. 11, pp. 4201–4208, Nov. 2016, doi: 10.1109/TED.2016.2608722.

[13] T. Matsumoto, K. Kobayashi, and H. Onodera, "Impact of random telegraph noise on CMOS logic delay uncertainty under low voltage operation," in *IEDM Tech. Dig.*, Dec. 2012, p. 25, doi: 10.1109/IEDM.2012.6479104.

[14] B. Kaczer, M. Toledano-Luque, J. Franco, and P. Weckx, "Statistical distribution of defect parameters," in *Bias Temperature Instability for Devices and Circuits*, T. Grasser, Ed. New York, NY, USA: Springer, 2014.

[15] G. Wirth, "Time dependent threshold voltage variability due to random telegraph noise," in *Proc. IEEE Latin Amer. Electron Devices Conf. (LAEDC)*, Feb. 2020, pp. 1–4, doi: 10.1109/LAEDC49063.2020.9073286.

[16] J. Franco *et al.*, "RTN and PBTI-induced time-dependent variability of replacement metal-gate high-k InGaAs FinFETs," in *IEDM Tech. Dig.*, Dec. 2014, p. 20, doi: 10.1109/IEDM.2014.7047087.

[17] A. Ghetti, C. Monzio Compagnoni, A. S. Spinelli, and A. Visconti, "Comprehensive analysis of random telegraph noise instability and its scaling in deca–nanometer flash memories," *IEEE Trans. Electron Devices*, vol. 56, no. 8, pp. 1746–1752, Aug. 2009, doi: 10.1109/TED.2009.2024031.

[18] K. Takeuchi, T. Nagumo, S. Yokogawa, K. Imai, and Y. Hayashi, "Single-charge-based modeling of transistor characteristics fluctuations based on statistical measurement of RTN amplitude," in *Proc. Symp. VLSI Technol.*, Honolulu, HI, USA, Jun. 2009, pp. 54–55.

[19] A. Asenov, R. Balasubramaniam, A. R. Brown, and J. H. Davies, "RTS amplitudes in decananometer MOSFETs: 3-D simulation study," *IEEE Trans. Electron Devices*, vol. 50, no. 3, pp. 839–845, Mar. 2003.

[20] A. C. J. Rossetto, V. V. A. Camargo, T. H. Both, D. Vasileska, and G. I. Wirth, "Statistical analysis of the impact of charge traps in p-type MOSFETs via particle-based Monte Carlo device simulations," *J. Comput. Electron.*, vol. 19, no. 2, pp. 648–657, Jun. 2020, doi: 10.1007/s10825-020-01478-6.

[21] Z. Zhang *et al.*, "New insights into the amplitude of random telegraph noise in nanoscale MOS devices," in *Proc. IEEE Int. Rel. Phys. Symp. (IRPS)*, Apr. 2017, pp. 3C-3.1–3C-3.5, doi: 10.1109/IRPS.2017.7936288.

[22] N. Tega *et al.*, "Increasing threshold voltage variation due to random telegraph noise in FETs as gate lengths scale to 20 nm," in *Proc. Symp. VLSI Technol.*, Honolulu, HI, USA, Jun. 2009, pp. 50-51.

[23] B. Zimmer, O. Thomas, S. O. Toh, T. Vincent, K. Asanovic, and B. Nikolic, "Joint impact of random variations and RTN on dynamic writeability in 28nm bulk and FDSOI SRAM," in *Proc. 44th Eur. Solid State Device Res. Conf. (ESSDERC)*, Sep. 2014, pp. 98–101, doi: 10.1109/ESSDERC.2014.6948767.