Automation of Model Parameter Estimation for Random Telegraph Noise

Hirofumi SHIMIZU[†a)]**,** *Student Member***, Hiromitsu AWANO**[†]**,** *Nonmember***, Masayuki HIROMOTO**[†]**,**
and **Takashi SATO**[†]**,** *Members*

SUMMARY   The modeling of random telegraph noise (RTN) of MOS transistors is becoming increasingly important. In this paper, a novel method is proposed for realizing automated estimation of two important RTN-model parameters: the number of interface-states and corresponding threshold voltage shift. The proposed method utilizes a Gaussian mixture model (GMM) to represent the voltage distributions, and estimates their parameters using the expectation-maximization (EM) algorithm. Using information criteria, the optimal estimation is automatically obtained while avoiding overfitting. In addition, we use a shared variance for all the Gaussian components in the GMM to deal with the noise in RTN signals. The proposed method improved estimation accuracy when the large measurement noise is observed.

key words: *random telegraph noise (RTN), Gaussian mixture model (GMM), expectation-maximization (EM) algorithm, information criteria, model estimation*

1. Introduction

Random telegraph noise (RTN) is an increasing concern for LSI circuits designed using metal-oxide-semiconductor (MOS) transistors in an advanced technology node. Figure 1 shows a cross sectional view of a MOS transistor, illustrating a principle of RTN. RTN is observed as a temporal fluctuation of the threshold voltage (V_{th}) of a transistor. A capture and an emission of a carrier at the trap in the gate insulator film cause a change in surface potential, which leads to the fluctuation of threshold voltage. When two or more number of traps exist in a transistor, the effects of the individual RTNs are superimposed.

    The impact of RTN is fast growing as transistor dimensions become smaller because the impact of a charge carrier becomes bigger in small area devices. It is predicted that an RTN-induced fluctuation of the threshold voltage is expected to be more influential than random dopant fluctuation in the near future [1], [2]. The modeling of RTN thus becomes important for accurate prediction of the circuit performance and for studying stability of a circuit under the influence of RTN.

    A lot of studies exist for the modeling of RTN [2]–[5]. In these studies, the number of traps in a device, threshold voltage shift ΔV_{th} that corresponds to a trap, and the time constants of capture and emission, are the common param-
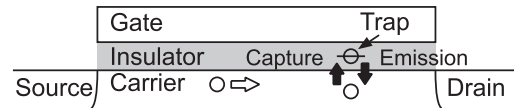
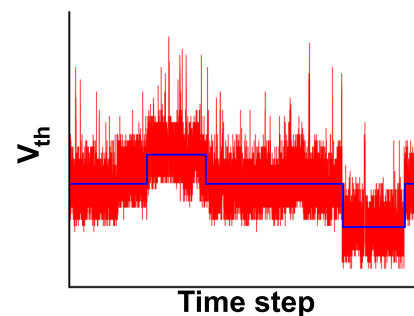**Fig. 1**   Principle of RTN.

Fig. 2   An example of measured time-change of threshold voltage.

eters. Characterization of these parameters from the measurement is necessary. Basically, the parameters have to be obtained based on the measurements over a large number of devices because RTN is a stochastic process. For that purpose, device arrays that are suitable to capture RTN are proposed, e.g., in [6], [7].

    In this paper, we focus on estimating two important model parameters — the number of interface-states and corresponding threshold voltage shift of each interface-state.

    When a trap captures or emits a carrier, the threshold voltage of a transistor changes accordingly. The number of interface-states, M, is simply related to the number of traps, K, with the following equation,

$$M = 2^K. \tag{1}$$

Here, K is a non-negative integer.

    In practice, estimation of the number of traps or the number of interface states is a difficult task because the voltage level is ambiguous in the measurement data. Let us take an example shown in Fig. 2. In this example, we notice three major voltage levels, which are traced by a solid line. Small measurement noise is superposed on the major voltage levels. Note that these noise-like threshold voltage shifts may also be the results of RTN having very small amplitudes. Furthermore, when noise becomes large as compared with the actual threshold voltage change due to RTN, it becomes

more difficult to accurately extract the number of states.

In this example, information on the number of internally existing states is lost in the measurement data. What we can observe is the threshold voltage fluctuation at each moment, which is a result of superposition of the activity of the traps. An accurate estimation of the number of states is thus an ill-conditioned, and hence inherently a difficult problem.

For the purpose of the model parameter characterization, we can obtain a large volume of RTN data measured over many devices, by using an array-like circuit [6]. We would like to estimate the model parameters from the measurements, hence the estimation has to be conducted automatically.

In this paper, we establish an automated method to estimate the number of interface-states in a device. In the proposed method, the threshold voltage levels are approximated using a Gaussian mixture model (GMM). By using the expectation maximization (EM) algorithm [8], [9], large voltage steps in the data are captured by the GMM. Considering that the small and constant noise is superposed on each voltage level, we limit the form of the GMM to share a single variance. Then, information criteria [10], [11] are used to avoid the problem of over-fitting that is caused by using too many Gaussian distributions. With the proposed method, human-subject estimation can be eliminated, and hence an objective and automated extraction of the RTN model parameters is achieved. The proposed method enhances the accuracy of the RTN model and will eventually improve the design quality of LSI circuits.

This paper is organized as follows. In Sect. 2, we review related work. In Sect. 3, a method for an interface-state and threshold-voltage level estimations are proposed. Then in Sect. 4, the proposed method is evaluated by artificial data having two or four interface-states. Finally, Sect. 5 summarizes this paper.

2. Related Works

For estimating the number of interface-states of MOS transistors, time lag plot (TLP) [12] and time dependent defect spectroscopy (TDDS) [13] are two well-known methods.

The TLP records the transitions of the time-series signal in a two-dimensional plane so that the x- and y-axes become the threshold voltages of consecutive measurement time. When the value of the current threshold voltage equals to the previous time, the symbol in the TLP is located on a diagonal. When the values of two consecutive times are different, then the symbol will be plotted on off-diagonal areas. By repeating this process for entire measurement period, threshold voltage changes, caused by a single trap and thus having a constant amplitude, will form a square in the plot. By counting the number of squares, we can find the number of traps in the RTN waveform.

In the TDDS, interface-states of a MOS transistor are observed through repetitive bias-temperature instability (BTI) measurements. More specifically, it focuses on the recovery period of the device after being kept in stress conditions. The change in threshold voltage is plotted as y-axis being the x-axis an elapsed time since the recovery period has started. By repetitively applying the stress and recovery periods, interface-states and their emission-time constants will be represented as a spectrum-like figure.

The above two methods are useful for visualizing the existence of the interface-states. However, if we want to estimate the number of interface-states, these methods require human effort to find specific features that represent interface-states. This process has not yet been automated. Also, when signal-to-noise (S/N) ratio becomes worse, it becomes difficult to find the features even for a human.

More recently, a new approach that uses Bayesian inference to obtain RTN model parameters has been proposed in [14]. This method simultaneously and automatically estimates the number of traps and hidden activity of each trap. However, estimation takes very long time since the method is based on Monte Carlo estimation of probabilistic model. Proposed method can also be used for the preprocessing of [12], [13] or [14], by quickly selecting the important samples that has many traps or that suffers a large threshold voltage shift.

3. Automated Estimation of the Number of Interface-States

3.1 Algorithm Overview

In this section, we propose a new method to estimate the number of MOS transistor's interface-states from the time domain RTN waveform. In our method, the GMM is used to fit the threshold voltage histogram. The GMM is selected so that it best describes the voltage histogram with the minimum number of model components.

An example of the time dependent threshold voltage shift and the corresponding threshold voltage histogram is presented in Fig. 3. On the right axis of the time domain graph, the threshold voltages are projected and the voltage histogram is plotted. In the time domain graph, there are two distinctive voltage levels that look like two bands because of the heavy noise. We assume that the noise is caused by the sources that have constant variance, such as measurement noise or inseparably small RTN. Hence, we consider that the voltage histogram is represented by the sum of Gaussian distributions. Under the above assumption, it is appropriate to approximate the voltage histogram using a Gaussian mixture model (GMM) [15], [16]. Once the GMM is obtained, then the mean of each Gaussian distribution gives the threshold voltage level, and the variance represents the spread of the voltage level due to either noise or smaller RTN components. Also, the mixing coefficient, i.e., the contribution of a Gaussian component is related to the duration in which a trap is in a particular state.

Trying to automate the estimation process, we have proposed a method that uses the EM algorithm [17]. In this method, starting from a relatively large number of Gaus-
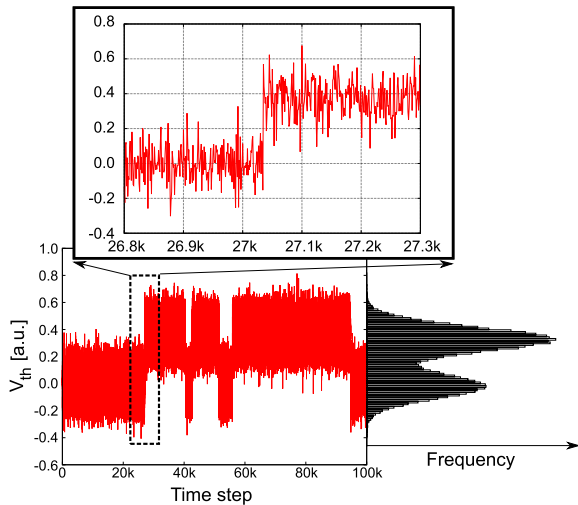
Fig. 3 Example of generating a voltage histogram from the time-domain threshold voltage. The threshold voltages are projected to form the voltage histogram.

sian distributions, the number of models is being reduced until the number of models converges. The reduction of the model uses the following rule. The two Gaussian models are considered to be equal and thus forced to merge into one when the mean values of the distributions are very close. A closeness parameter is used to control the threshold to judge whether the distributions are close or not. The estimation of the number of interface-states heavily depends on the closeness parameter. The value of closeness parameter should be set adaptively to obtain a good accuracy, depending on the noise in the measurement.

Therefore, the following procedures are additionally introduced in the proposed method in order to achieve better adaptability of the model.

- We propose to use a single variance for all Gaussian models in a GMM.
- We propose to use information criteria to obtain the best model instead of using the closeness parameter.

The first procedure is based on the assumption that the measurement variation comes from a single source or the sum of sources, which should yield a constant variance. In this paper, to simulate this situation, variances of the Gaussian distribution in the GMM will be constrained to be the same. We call this bounded variance model as GMM with *single variance* or that with *shared variance*, or that with *constrained variance*. The second procedure is introduced to alleviate the difficulty to determine the closeness parameter.

The proposed flow for estimating the optimal GMM is presented in Algorithm 1, which can be summarized as follows.

1. Initialize M, number of interface-states, to be 1.
2. Fit GMM to the voltage histogram.
3. Fit GMM to the voltage histogram with shared variance.
4. Calculate information criteria (IC) to evaluate the

Algorithm 1 *Two step algorithm*: obtain the best GMM with shared variance, which approximates a voltage histogram.

input: $X = (x_1, \ldots, x_N)$
output: M_{icmin}

$M \leftarrow 1$
$IC[0] \leftarrow \infty$
while (M_{icmin} != Null) **do**
    $\mu_m^{\text{init}}, \sigma_m^{\text{init}} \leftarrow$ Clustering by k-means++
    Calculate π_m^{init}
    Calculate $(\ln L)^{\text{init}}$ using Eq. (12)
    $(\ln L)^{\text{new}} \leftarrow (\ln L)^{\text{init}}$
    while $r_{\text{err}} >= r_{\text{thresh}}$ **do**
        $(\ln L)^{\text{old}} \leftarrow (\ln L)^{\text{new}}$
        E step
        M step
        Calculate $(\ln L)^{\text{new}}$ using Eq. (12)
        Calculate r_{err} using Eq. (13)
    end while
    while $r_{\text{err}} >= r_{\text{thresh}}$ **do**
        $(\ln L)^{\text{old}} \leftarrow (\ln L)^{\text{new}}$
        E step
        M step
        Equalize σ_m^{new} using Eqs. (10) and (11)
        Calculate $(\ln L)^{\text{new}}$ using Eq. (12)
        Calculate r_{err} using Eq. (13)
    end while
    $IC[M] \leftarrow$ Calculate IC
    if (M_{icmin} == Null) and ($M >= 2$) **then**
        if ($IC[M] >= IC[M-1]$) and ($IC[M-1] >= IC[M-2]$) **then**
            $M_{\text{icmin}} = M - 2$
        end if
    end if
    $M \leftarrow M + 1$
end while
return M_{icmin}

---

    goodness of fit of the GMM at the number of interface-states M.

5. If ($M \geq 2$ and $IC[M] \geq IC[M-1] \geq IC[M-2]$) is satisfied, then return ($M - 2$) as the best estimation of the number of interface-states. Otherwise, increment M by one and return to Step 2.

In Steps 2 and 3, two EM algorithms are successively conducted. We here call the proposed algorithm as the *two step algorithm*. In the first step, a normal EM algorithm that does not use shared variance is used, and in the second step, an EM algorithm that uses GMM with shared variance is applied. When the variance of the GMM is equalized, the EM algorithm tends to be trapped by local minima because of the reduced degree of freedom due to variance constraint. Hence we first apply the normal EM algorithm to roughly fit the model, and then conduct the second EM algorithm by using GMM to equalize variance.

Each step of the proposed algorithm will be described in the following subsections.

3.2 Obtaining the Optimal Gaussian Mixture Model

In the proposed flow, estimation of GMM is repeated by incrementing the number of interface-states M. Starting from

$M = 1$, we fit GMM to the given threshold voltage data by using the EM algorithm. The EM algorithm is a method that estimates the parameters of a probabilistic model with a maximum likelihood, i.e., it finds the most reasonable parameters of GMM. The parameters of GMM are mean, variance and contribution of each Gaussian component. The k-means++ [18] is used to determine the initial parameter values for the EM algorithm.

We have to find the optimal number of Gaussian models that best approximates the voltage histogram. Special attention has to be paid when evaluating the goodness of the model. If we increase the complexity of the model, which can be realized by using larger M, we can always achieve smaller error. However, the smaller error does not always mean the model is good. Too complex model having too many Gaussian components may suffer from overfitting problem. The model may adapt to the measurement noise that should not be included in the model. We have to evaluate the fitness while avoiding this overfitting problem. For this purpose, we utilize the idea of *information criteria*.

In the following subsections, we review GMM and EM algorithms.

3.2.1 Gaussian Mixture Model

Let an observed data sequence and its length be $X = (x_1, \ldots, x_N)$ and N, respectively. The GMM has the following form:

$$p(x_n) = \sum_{m=1}^{M} \pi_m \mathcal{N}\left(x_n | \mu_m, \sigma_m^2\right), \tag{2}$$

where $\mathcal{N}\left(x_n | \mu_m, \sigma_m^2\right)$ is m-th independent normal distribution, and M is the number of the normal distributions. μ_m, σ_m^2, and π_m are also the mean, variance, and mixing coefficient of the m-th normal distribution. The m-th independent normal distribution is

$$\mathcal{N}\left(x_n | \mu_m, \sigma_m^2\right) = \frac{1}{\sqrt{2\pi\sigma_m^2}} \exp\left\{-\frac{1}{2\sigma_m^2}(x_n - \mu_m)^2\right\}. \tag{3}$$

The mixing coefficients satisfy Eq. (4),

$$\sum_{m=1}^{M} \pi_m = 1, \quad \text{and} \quad 0 \le \pi_m \le 1. \tag{4}$$

3.2.2 EM Algorithm

The EM algorithm is a method that estimates a maximum likelihood solution of the parameters of a probabilistic model. We apply the EM algorithm to estimate the GMM that most likely explains measurement data. The EM algorithm is conducted through the following steps.

1. Initialization: initializes the mixture of Gaussian distributions (the number of Gaussian distributions: M) and calculates the log-likelihood for the initial distribution. In order to automatically set initial values for

the parameters of the GMM (μ_m, σ_m^2, and π_m), we apply k-means++ algorithm for the observed threshold voltages to form M clusters.

2. E step: given the current parameters, calculate responsibility $\gamma(z_{nm})$, which is the contribution of a Gaussian model in the GMM to a voltage sample x_n using

$$\gamma(z_{nm}) = \frac{\pi_m \mathcal{N}\left(x_n | \mu_m, \sigma_m^2\right)}{\sum_{j=1}^{M} \pi_j \mathcal{N}\left(x_n | \mu_j, \sigma_j^2\right)}. \tag{5}$$

3. M step: update the parameters for the GMM using the present responsibility,

$$\mu_m^{\text{new}} = \frac{1}{N_m} \sum_{n=1}^{N} \gamma(z_{nm}) x_n, \tag{6}$$

$$(\sigma_m^{\text{new}})^2 = \frac{1}{N_m} \sum_{n=1}^{N} \gamma(z_{nm}) (x_n - \mu_m^{\text{new}})^2, \tag{7}$$

and

$$\pi_m^{\text{new}} = \frac{N_m}{N}, \tag{8}$$

where N_m is

$$N_m = \sum_{n=1}^{N} \gamma(z_{nm}). \tag{9}$$

When it is directed to use the GMM with shared variance, σ_m^{new} will be equalized through the following averaging procedure after calculation of σ_m^{new} in Eq. (7).

a. Obtain σ_m^{new} using Eq. (7)
b. Calculate the common variance σ^{new} by

$$\sigma^{\text{new}} = \sqrt{\frac{1}{M} \sum_{m=1}^{M} (\sigma_m^{\text{new}})^2}. \tag{10}$$

c. Update variance of each GMM by

$$\sigma_m^{\text{new}} = \sigma^{\text{new}}. \tag{11}$$

4. Convergence test: calculate the log-likelihood using

$$
\begin{aligned}
\ln L &= \ln p(X | \mu, \sigma, \pi) \\
&= \sum_{n=1}^{N} \ln \left\{ \sum_{m=1}^{M} \pi_m \mathcal{N}\left(x_n | \mu_m, \sigma_m^2\right) \right\},
\end{aligned}
\tag{12}
$$

and evaluate its change by

$$r_{\text{err}} = \frac{\left|(\ln L)^{\text{new}} - (\ln L)^{\text{old}}\right|}{(\ln L)^{\text{old}}}. \tag{13}$$

Here, $(\ln L)^{\text{old}}$ and $(\ln L)^{\text{new}}$ are the log likelihood before and after the M step, respectively. The convergence criterion is

$$r_{\text{err}} < r_{\text{thresh}}. \tag{14}$$

r_{thresh} is the threshold to judge convergence. The threshold can be determined to be a smaller value than the penalty terms in information criteria described in the subsequent subsection. The E and M steps are repeated until the convergence criterion is satisfied.

3.3 GMM with Shared Variance

Reflecting the physical assumption that the variance of noise is constant, the variances of the GMM components are constrained to be equal by using Eq. (10). The effectiveness of the shared variance is better understood through examples. Figures 4(a) and 4(b) show the distributions of the GMM with and without shared variance, respectively. This example is taken from the case when there are two voltage levels. The correct voltage levels for the two distributions, $p_1(x)$ and $p_2(x)$, are about 0 and 1, respectively. The distributions of $\hat{p}_1(x)$ and $\hat{p}_2(x)$ are the Gaussian components of the optimized GMM after the application of the EM algorithm. The means of the distributions are indicated by markers. In Fig. 4(a), when the variance is not constrained, there are large discrepancies between the true and estimated distributions. Variance of $\hat{p}_2(x)$ is larger than that of $\hat{p}_1(x)$ at the end of the EM algorithm.

Although the estimated components in the GMM are different from the true distributions, the sum of the two estimated distributions may become very similar to the true distribution as seen in Fig. 4(c). The two different GMMs and the true distribution are almost indistinguishable. The GMM in Fig. 4(a) happens to be a good approximation of the entire distribution. We know from this example that it is extremely important to match each Gaussian component to the true distribution that is unknown. This example suggests that the approximation of histogram using GMM can have several local optimums. With the aid of physical assumption, the local optimums can be avoided.

Figure 4(b) shows a result when a single variance is used. In this case, not just the entire distribution but each Gaussian distribution matches to the true distribution very well, giving us the correct estimation results of the voltage levels. By the shared variance, the number of local optimum may be reduced and there are increased chances to achieve successful estimation. The shared variance is very effective to improve the accuracy of estimations.
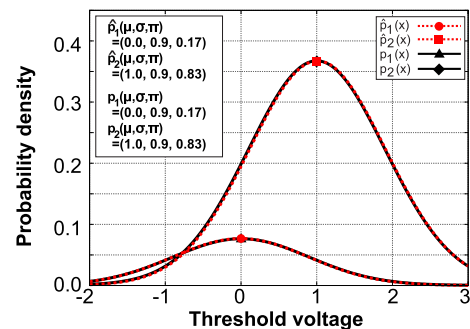
3.4 Information Criteria

To select the optimal GMM, we use the idea of information criteria. There exist many information criteria that can be used for the model selection. In this work, we apply Akaike information criterion (AIC) [10] and Bayesian information criterion (BIC) [11]. AIC and BIC are given by the following equations:
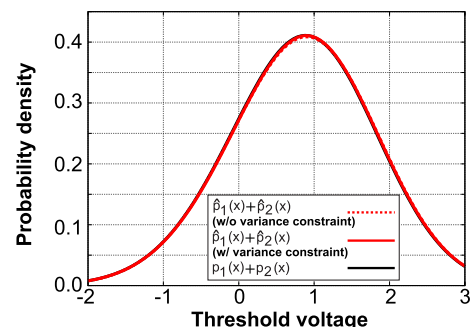
$$\text{AIC} = -2\ln L + 2k \tag{15}$$
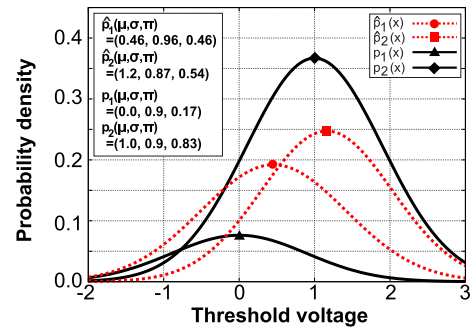
(a) Without shared variance.

(b) With shared variance.

(c) Total distributions.

Fig. 4 Approximation example showing the effectiveness of shared variance in the GMM. p_1 and p_2 are the correct distributions. $\hat{p}_1$ and $\hat{p}_2$ are the estimated distributions.

and

$$\text{BIC} = -2\ln L + k\ln N, \tag{16}$$

where L, k, and N are the maximum likelihood, the number of free parameters, and the length of the measurement sequence. In our case, the free parameters are: μ, σ, and π of each state. Hence, the number of free parameters k becomes $(3M-1)$ when different variance is used for each Gaussian model. In the case that a shared variance is used for each model, then the number of free parameters k becomes $2M$.

In general, the value of M that minimizes information criteria gives the optimal model. The information criteria usually have a single minimum point as in Fig. 6. As we increase M, the criteria first decreases sharply because of the decrease in the first term of Eq. (15) or (16). This is because the more complex model matches better than the simpler one and thus increases log-likelihood. Then, beyond

the minimum point, the criteria start to increase due to the second term that gives penalty for using a complex model.

With this observation, we can basically increase M to find the first kink point at which the information criteria start to increase. This will enable us to automatically determine the optimal GMM. However, we sometimes observe the local minimum point at a smaller M value. The reason of having the local minimum has not yet been clearly understood, but the use of single shared variance described in the previous subsection, insufficient number of iterations in the EM algorithm, etc. are the possible reasons. To avoid this local minimum point, we judge the minimum point by two consecutive increases in information criteria ($\text{IC}[M] \geq \text{IC}[M-1] \geq \text{IC}[M-2]$).

4. Numerical Evaluations

In this section, we demonstrate that the parameters of the GMM, i.e., the number of interface-states and threshold voltage levels are estimated correctly and automatically using the proposed method. The effectiveness of using the GMM with shared variance is also verified.

Throughout the evaluation, we use artificial RTN data because the ground truth is unable to know when we use the measured RTN data. The artificial RTNs are generated using the parameters listed in Table 1. Time-changing threshold voltage without noise is first generated as the superposition of two-state Markov models. The duration of each state is determined randomly following the states of the Markov model that is characterized by time-constants τ_e and τ_c for each trap [5]. Then, Gaussian noise of zero mean and the variance of σ_d^2 are added for each time step. Because the noise variance will be swept, the differences of threshold voltage levels are normalized to be 1.0 without losing the generality.

4.1 Examples of the Proposed Estimation Method

Examples of the generated RTNs are shown in Fig. 5 for $M_s = 2, \sigma_d = 0.3$ and Fig. 7 for $M_s = 4, \sigma_d = 0.4$, respectively. Here, M_s represents the true number of states used in the data generation step, which is equal to 2^K. The generated RTN waveforms are very similar to the measurements found in literatures, e.g., in [7]. Large measurement noise makes it difficult to estimate the exact voltage levels for each state.

The proposed method is applied for the generated RTNs. The GMMs with different number of Gaussian components M are approximated through the EM algorithm. The values of information criteria for the optimized GMMs are plotted in Figs. 6 and 8. The log-likelihood decreases rapidly as M is increased. It saturates at about $M = 2$ in Fig. 6 and $M = 4$ in Fig. 8. Because of the second terms in Eqs. (15) and (16), both AIC and BIC start to increase at these values. According to the proposed method shown in Algorithm 1, M that gives the minimum value of information criteria is the best estimation for the number of

Table 1 Parameters for generating two-state RTN data ($K = 1$ is used for the experiments of Figs. 5, 6 and 9. $K = 2$ is used for Figs. 7, 8 and 10.

| Parameter	Value
Number of traps K	1 or 2
Data length N	100,000
ΔV_{th} amplitude	1.0
Time constants τ_c, τ_e	Log-normal random numbers (mean, sigma)= (9, 1)
Standard deviation of measurement noise σ_d	0.0 to 0.9
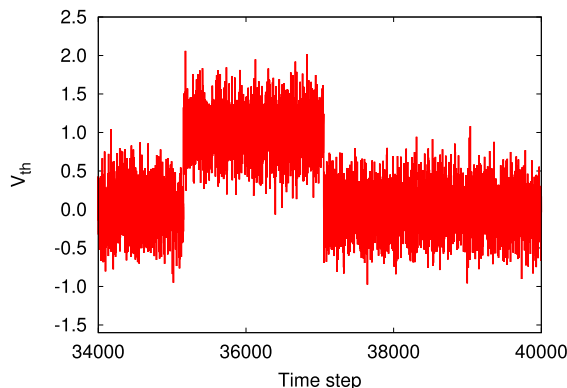
Fig. 5 Example data of artificial RTN (M_s=2, σ_d=0.3). Samples in between 34 k to 40 k are plotted for clarity out of 100 k samples in total.
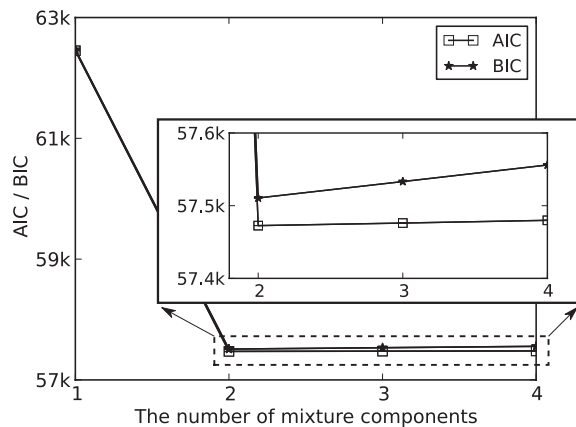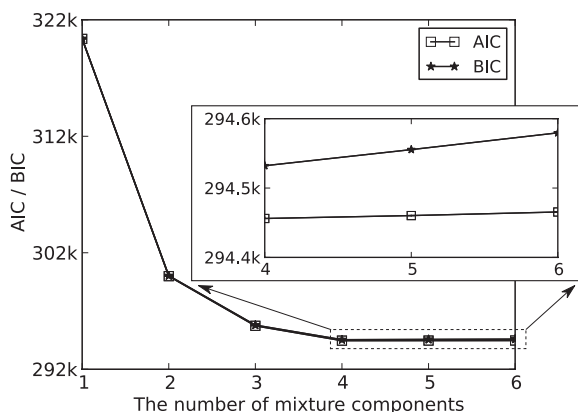

Fig. 6 Changes of AIC and BIC when estimating the number of interface-states for the data in Fig. 5.

interface-states. In these examples, we see that the proposed method estimates the number of states correctly.

In order to check the robustness of the estimation, we here confirm the stability of the estimation. After initializing the model parameters using k-means++, means of the Gaussian models in the GMM are modified by adding a random number drawn from a Gaussian distribution that have zero mean and the variance of σ_d^2. Then the parameter estimations are conducted. This process has been repeated for 100 times and evaluated the log likelihood. If, depending on the change of initial values for the EM algorithm, log likelihood largely fluctuates in the order of or larger than the second terms in Eqs. (15) and (16), then the evaluations
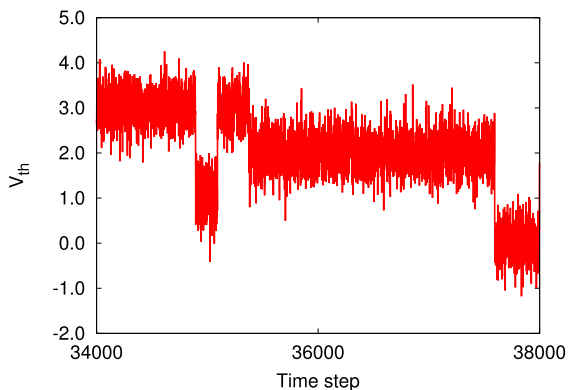
Fig. 7 Example data of artificial RTN (M_s=4, σ_d=0.4). Samples between 34 k to 38 k out of 100 k samples are plotted.

Fig. 9 Success rate of estimations (M_s=2). The estimation accuracy has been improved by the proposed two step algorithm.

Fig. 8 Changes of information criteria for the data of Fig. 7.

using AIC and BIC can be considered unreliable. We calculated the maximum fluctuation of AIC and BIC is as high as 0.714, which is smaller than the smallest possible value that the penalty terms in Eqs. (15) and (16) can take. Hence, we can consider the convergence criteria are sufficiently good, and the use of information criteria is effective.
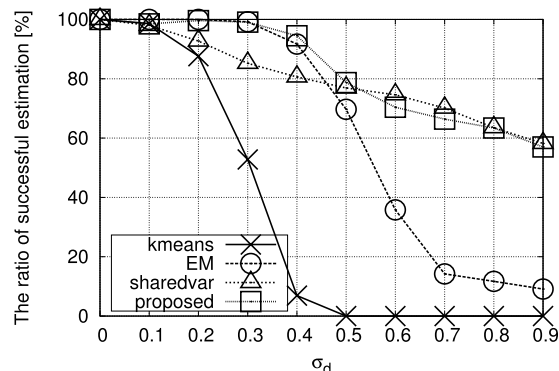
Since both AIC and BIC give very close estimations in almost all examples, we hereafter use BIC as an information criterion to evaluate the number of states.

4.2 State Number and Voltage Level Estimation

In order to validate the proposed method, 1000 artificial RTNs are generated for M_s = 2, and 100 artificial RTNs are generated for M_s = 4. The artificial RTNs are generated by using the parameters listed in Table 1. Then, the interface-state numbers are estimated. The condition for the successful threshold voltage level estimation is to satisfy

$$|\hat{v}_m - v_m| < 0.05 \tag{17}$$

for all voltage levels. Here, $\hat{v}_m$ is the estimated threshold voltage level, or the mean of the m-th Gaussian component in the GMM, and v_m is the true threshold voltage levels. Because voltage levels are normalized to have 1 V intervals,

this condition corresponds to a 5% of the voltage level difference.

Figure 9 compares the ratio of successful estimations for the case of M_s = 2. Four procedures are compared: 1) k-means++ algorithm only (kmeans), 2) after the k-means++, the normal EM algorithm which does not equalize variance of GMM is conducted (EM), 3) after the k-means++, the EM algorithm which equalize variance of GMM is conducted (sharedvar), and 4) the proposed two-step algorithm in which the normal EM algorithm and the EM algorithm that uses GMM with shared variance are successively conducted. Note that an identical RTN is used as the inputs for fair comparison. In the experiment, signal-to-noise ratio is swept by changing the standard deviation of the measurement noise. When the noise is small, just applying k-means++ gives the correct estimations for the number of interface-states and for voltage levels. As the noise becomes larger, it becomes difficult for k-means++ to yield correct estimation. The use of the normal EM algorithm after the initialization of k-means++ greatly improves estimation accuracy. When GMM with shared variance is used in the EM algorithm, estimation rate becomes worse for the smaller noise $\sigma_d \leq 0.46$ but becomes better for the larger noise $\sigma_d > 0.46$. The proposed two-step algorithm can incorporate the benefits of the two. For example, when σ_d is 0.6, the ratio of the correct estimations for both the number of states and the threshold voltage levels is improved from 35% to 70%.

Figure 10 shows the ratio of the successful estimation for the case of M_s = 4. Very similar trend to the case of M_s = 2 can be observed. The estimation performance has been improved by the shared variance in the GMM. For example, when σ_d is 0.5, the ratio of the correct estimations for both the number of states and the threshold voltage levels is improved from 5% to 29%

Here, the validity of using two consecutive increases in information criteria as the stop-condition is evaluated. Figure 11 shows the ratio of the successful estimation for M_s = 4 when different number of *consecutive increases* is used. The RTN signals used in this evaluation are similar to those of Fig. 10. The number of RTN signals used is 100.
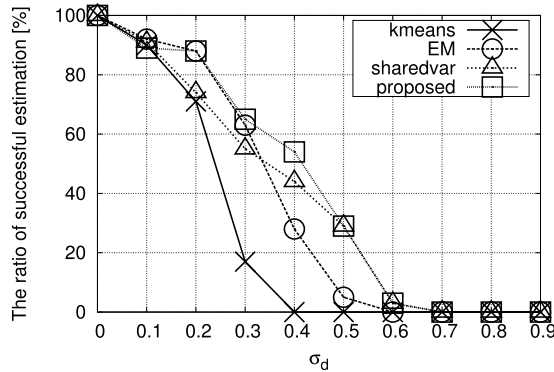
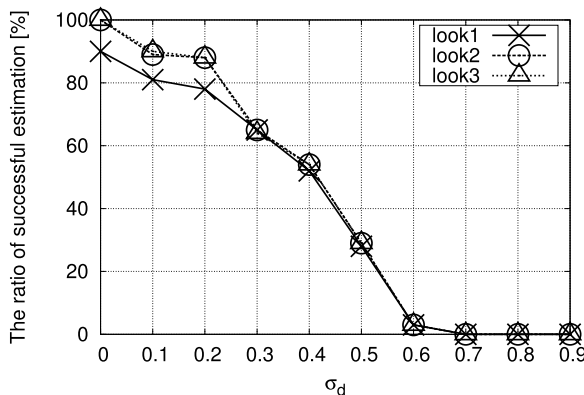**Fig. 10**   Success rate of estimations (M_s=4).

Fig. 11   Success rate of estimations for $M_s = 4$ for different number of *consecutive increases*. The consecutive increase means how many consecutive increases in information criteria are required to terminate the proposed two step algorithm. "look1," "look2," and "look3" requires one, two, and three consecutive increases, respectively.

Fig. 12   Median estimation time for successful estimations (M_s=2). Although the conventional EM algorithm is faster than the proposed method, its estimations are very unreliable as seen in Fig. 9.

Fig. 13   Median estimation time for successful estimations (M_s=4). There are no correct estimations beyond $\sigma_d = 0.6$ for the conventional and the proposed methods.

"look1," in which one increasing step is necessary, is inferior to "look2" which requires two consecutive increases to stop. Only a slight difference could be found between "look2" and "look3" which requires three consecutive increases. Therefore, it is reasonable to use two consecutive increases.

   Computational time required for these methods are also evaluated. Figure 12 compares median estimation time for $M_s = 2$. Circles represent the conventional method, which uses the EM algorithm without shared variance. Triangles represent the method which uses the EM algorithm with shared variance and squares represent the proposed two step algorithm in which two EM algorithms are successively conducted. Note that the estimation time is defined by the sum of the processing times for k-means++ and the EM algorithm. Also note that the estimation time is calculated only for the samples in which successful estimation was made. Because the processing time for k-means++ is much shorter than that of the EM algorithm, the estimation time is mostly determined by the EM algorithm. Because the proposed algorithm consists of two EM algorithms runs, the proposed algorithm is always slower than the one that uses the normal EM algorithm after k-means++. Considering that the
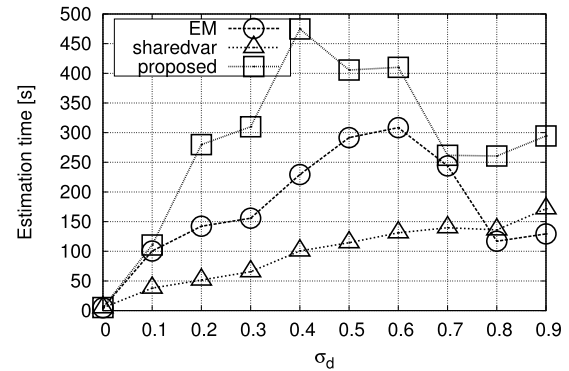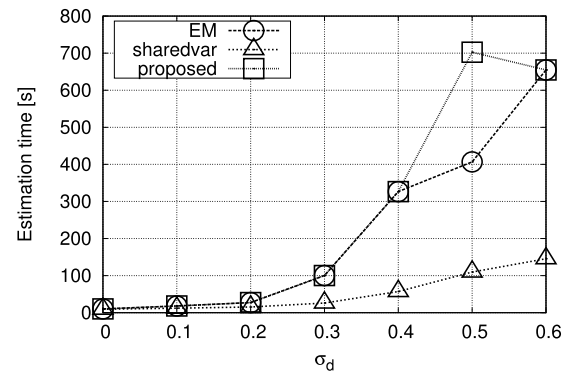
proposed algorithm yields better accuracy than the normal EM algorithm, the time required for the proposed algorithm, which is approximately equal to or less than twice of the normal EM algorithm, can be acceptable. Reduction of the computational time is one of our future works.

   In a general trend, it takes more time to estimate the interface-states without shared variance than to estimate the interface-states with shared variance because of large noise. Figure 13 compares the median of estimation time for $M_s = 4$. As we have seen in Fig. 10, because of the difficulties to obtain correct estimation results, there is no valid timing in the cases of $\sigma_d > 0.6$.

5.   Conclusion

A novel method that realizes automated estimation of the number of interface-states from RTN waveform is proposed. The proposed method utilizes a Gaussian mixture model to represent voltage distributions of the threshold voltage levels. The threshold voltage shift associated with the change of the states is also accurately estimated as the means of the Gaussian distributions. By using information criteria with the EM algorithm, the appropriate number of the Gaussian models to fit the voltage distribution is also determined al-

gorithmically, while avoiding over-fitting. In addition, by adding the constraint to share a variance among the Gaussian models in the GMM, the proposed method has successfully estimated the number of interface-states even under the severe conditions where the large measurement noise is observed.

Through numerical evaluations, it has been proven that the proposed method provides accurate and fast auto-estimation of the model parameters. By the proposed two step algorithm, successful estimation ratio has been improved from 35% to 70% for two-state case when large noise ($\sigma_d = 0.6$) is observed with the RTN waveform. The proposed method is considered useful for improving design quality of LSI circuits.

Acknowledgements

References

[1] N. Tega, H. Miki, F. Pagette, D. Frank, A. Ray, M. Rooks, W. Haensch, and K. Torii, "Increasing threshold voltage variation due to random telegraph noise in FETs as gate lengths scale to 20 nm," Proc. Symp. VLSI Technol., pp.50–51, June 2009.

[2] N. Tega, H. Miki, M. Yamaoka, H. Kume, T. Mine, T. Ishida, Y. Mori, R. Yamada, and K. Torii, "Impact of threshold voltage fluctuation due to random telegraph noise on scaled-down SRAM," Proc. IEEE Int. Rel. Physics Symp., pp.541–546, April 2008.

[3] N. Tega, H. Miki, T. Osabe, A. Kotabe, K. Otsuga, H. Kurata, S. Kamohara, K. Tokami, Y. Ikeda, and R. Yamada, "Anomalously large threshold voltage fluctuation by complex random telegraph signal in floating gate flash memory," Proc. IEEE Int. Electron Devices Meeting, pp.1–4, Dec. 2006.

[4] 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," Proc. Symp. VLSI Technol., pp.54–55, June 2009.

[5] K. Ito, T. Matsumoto, S. Nishizawa, H. Sunagawa, K. Kobayashi, and H. Onodera, "Modeling of random telegraph noise under circuit operation — simulation and measurement of RTN-induced delay fluctuation," Proc. IEEE Int. Symp. Quality Electronic Design, pp.1–6, March 2011.

[6] T. Sato, T. Kozaki, T. Uezono, H. Tsutsui, and H. Ochi, "A device array for efficient bias-temperature instability measurements," Proc. European Solid-State Device Research Conf., pp.143–146, Sept. 2011.

[7] K. Abe, T. Fujisawa, A. Teramoto, S. Watabe, S. Sugawa, and T. Ohmi, "Anomalous random telegraph signal extractions from a very large number of n-metal oxide semiconductor field-effect transistors using test element groups with 0.47 Hz–3.0 MHz sampling frequency," Jpn. J. Appl. Phys., vol.48, no.4, 04C044, 2009.

[8] A.P. Dempster, N.M. Laird, and D.B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," J. The Roy. Statistical Soc., Series B, vol.39, no.1, pp.1–38, 1977.

[9] G.J. McLachlan and T. Krishnan, The EM algorithm and extensions, Wiley, 1997.

[10] H. Akaike, "A new look at the statistical model identification," IEEE Trans. Autom. Control, vol.19, no.6, pp.716–723, Dec. 1974.

[11] G. Schwarz, "Estimating the dimension of a model," Ann. Stat., vol.6, no.2, pp.461–464, March 1978.

[12] T. Nagumo, K. Takeuchi, S. Yokogawa, K. Imai, and Y. Hayashi, "New analysis methods for comprehensive understanding of random telegraph noise," Proc. IEEE Int. Electron Devices Meeting, pp.1–4, Dec. 2009.

[13] T. Grasser, H. Reisinger, P.-J. Wagner, F. Schanovsky, W. Goes, and B. Kaczer, "The time dependent defect spectroscopy (TDDS) for the characterization of the bias temperature instability," Proc. IEEE Int. Rel. Physics Symp., pp.16–25, May 2010.

[14] H. Awano, H. Tsutsui, H. Ochi, and T. Sato, "Bayesian estimation of multi-trap RTN parameters using Markov chain Monte Carlo method," IEICE Trans. Fundamentals, vol.E95-A, no.12, pp.2272–2283, Dec. 2012.

[15] G. McLachlan and K. Basford, Mixture models: Inference and applications to clustering, Marcel Dekker, 1988.

[16] G. McLachlan and D. Peel, Finite Mixture Models, Wiley, 2004.

[17] H. Shimizu, H. Tsutsui, H. Ochi, and T. Sato, "An automated estimation of MOS transistor's interface-state numbers using EM algorithm," IEICE Society Conference, C-12-18, p.93, Sept. 2011. (in Japanese)

[18] D. Arthur and S. Vassilvitskii, "K-means++: The advantages of careful seeding," Proc. ACM-SIAM Symp. Discrete Algorithms, pp.1027–1035, 2007.

Hirofumi Shimizu received the B.E. degree in Electrical and Electronic Engineering from Kyoto University in 2012. Presently, he is a master course student at Department of Communications and Computer Engineering, Kyoto University. He is a student member of IEICE.

Hiromitsu Awano received his B.E. degree in Informatics and his master degree in Communications and Computer Engineering from Kyoto University in 2010 and 2012, respectively. Presently, he is a doctor course student at Department of Communications and Computer Engineering, Kyoto University. He is a research fellow of japan society for the promotion of science and a student member of IPSJ.

Masayuki Hiromoto received B.E. degree in Electrical and Electronic Engineering and M.Sc. and Ph.D. degrees in Communications and Computer Engineering from Kyoto University in 2006, 2007 and 2009 respectively. He was a JSPS research fellow from 2009 to 2010, and with Panasonic Corp. from 2010 to 2013. In 2013, he joined the Graduate School of Informatics, Kyoto University, where he is currently an assistant professor. His research interests include VLSI design methodology, image processing and pattern recognition. He is a member of IEEE, IEICE, and IPSJ.

Takashi Sato received B.E. and M.E. degrees from Waseda University, Tokyo, Japan, and a Ph.D. degree from Kyoto University, Kyoto, Japan. He was with Hitachi, Ltd., Tokyo, Japan, from 1991 to 2003, with Renesas Technology Corp., Tokyo, Japan, from 2003 to 2006, and with the Tokyo Institute of Technology, Yokohama, Japan. In 2009, he joined the Graduate School of Informatics, Kyoto University, Kyoto, Japan, where he is currently a professor. He was a visiting industrial fellow at the University of California, Berkeley, from 1998 to 1999. His research interests include CAD for nanometer-scale LSI design, fabrication-aware design methodology, and performance optimization for variation tolerance. Dr. Sato is a member of the IEEE and the Institute of Electronics, Information and Communication Engineers (IEICE). He received the Beatrice Winner Award at ISSCC 2000 and the Best Paper Award at ISQED 2003.