


Probability Distribution of the Write-Error Rate of Voltage-Controlled Magnetoresistive Random-Access Memories

Hiroko Arai,^{1,*} Takahiro Hirofuchi,² and Hiroshi Imamura^{1,†}

¹*National Institute of Advanced Industrial Science and Technology (AIST), Research Center for Emerging Computing Technologies, Tsukuba, Ibaraki 305-8568, Japan*

²*National Institute of Advanced Industrial Science and Technology (AIST), Digital Architecture Research Center, Aomi, Koto-ku, Tokyo 135-0064, Japan*

 (Received 14 June 2021; revised 25 October 2021; accepted 3 December 2021; published 29 December 2021)

In this study, we theoretically explore the probability distribution of the write-error rate (WER) of a voltage-controlled magnetoresistive random-access memory (VC MRAM). A probability density function (PDF) of the WER is analytically obtained using the change-of-variable technique with the assumption that the anisotropy constant of the VC MRAM follows the normal distribution. Furthermore, we assume that the WER of a single memory cell forms a parabola as a function of the anisotropy constant, resulting from numerical simulations by solving the Landau-Lifshitz-Gilbert equation based on the macrospin model. Theoretical analysis facilitates in identifying types of the PDF: a monotonic decreasing function with a peak at the smallest WER, and a function with local maximum.

DOI: [10.1103/PhysRevApplied.16.064068](https://doi.org/10.1103/PhysRevApplied.16.064068)

I. INTRODUCTION

Reliability is a key characteristic of electronic devices [1–4]. In the present design of computing systems, memory devices ensure reliability with the help of error-correction systems. However, error correction requires the increase of circuit area, latency, and energy consumption. Recently, decreasing such additional overheads has emphasized the requirement for alternative designs of computing systems, in order to exclude error corrections in the hardware components [5–7].

One of the attempts is by the applications of machine learning or deep neural networks, which provides the output with probability or confidence. In other words, these techniques are principally error tolerant, that is, robust against bit errors [7–9]. Another way is by increasing the demand for edge devices. The edge devices limit the device size and operation energy, so that decreasing overheads of error-correction systems makes a large contribution.

Furthermore, a hybrid memory system has also drawn significant attention [10–12]. The system exhibits both conventional volatile memory systems, such as dynamic random-access memory (DRAM) and static RAM (SRAM), and nonvolatile memory systems, such as magnetic RAM (MRAM) [13–16], phase-change memory [17–19], and resistive RAM [20–22]. Compared with

the conventional DRAM and SRAM, nonvolatile memories have robustness for radiation-induced soft errors and energy efficiency. However, they have weakness for endurance, latency, and reliability. A hybrid memory system, combining two types of memory devices (e.g., erroneous memory device and reliable memory device), has a potential to expand the use of erroneous memory to more applications, which enables us to protect critical data by reliable memory. Practically, computing systems without error corrections on the hardware require an understanding of how data stored on a memory module include errors. The errors would occur with any spatial distribution on the memory module. The assumption with uniform probability for the error is not accurate.

For efficient design of these emerging devices based on nonvolatile memories, variation-aware simulations are necessary. Although several groups have developed the variation-aware model of the spin-transfer torque (STT) MRAM [23–26], little attention has been paid to the variation-aware model of the voltage-control (VC) MRAM, which is a promising candidate for high-performance and ultralow power nonvolatile memory.

VC MRAM is sensitive to manufacturing variations. MRAM stores information in the direction of magnetization in a magnetic tunnel junction (MTJ). In VC MRAM, the writing operation, that is, the magnetization switching, occurs with voltage control of magnetic anisotropy (VCMA) [27–31]. Each MTJ cell has manufacturing variation owing to fabrication processes, implying that a MRAM chip is made by a set of MTJ cells with slightly

*arai-h@aist.go.jp

†h-imamura@aist.go.jp

different material parameters. The effect of VCMA results in variation for each MTJ cell even if exactly constant voltage is applied to all MTJ cells. Thus, the performance of the MRAM chip results from the statistical distribution of MTJ cells. Particularly, the write process is the most affected process due to the manufacturing variation rather than the read and the retention processes. However, most previous studies have focused on improving the single-bit write-error rate (WER) [32–39]. The effects of manufacturing variations on the WER are still unclear.

In this study, a distribution of the WER of VC MRAM due to the manufacturing variation is theoretically explored. By assuming the variation of magnetic anisotropy, a probability density function (PDF) of the WER is analytically obtained, and the condition to determine the shape of the PDF is established. The frequency distribution of the WER is also obtained from the PDF, which strongly depends on the variation of magnetic anisotropy. The remainder of the paper is organized as follows. Section II describes the numerical simulations of the WER for a single MTJ. Section III explains the analytical derivation of the PDF of the WER. Section IV presents the general property of the WER distribution, while Sec. V focuses on the contribution of the variation of magnetic anisotropy. Finally, the paper is summarized in Sec. VI.

II. WRITE-ERROR RATE OF A SINGLE MAGNETIC TUNNEL JUNCTION

First, the WER of a single MTJ is investigated to clarify the relationship between the WER and magnetic anisotropy. The MTJ nanopillar considered is schematically shown in Fig. 1(a). The nonmagnetic insulating layer is sandwiched by the two ferromagnetic layers: the free layer and the reference layer. A static external field (\mathbf{H}_{ext}) is applied in the positive x direction. Application of the voltage (V) pulse reduces the magnetic anisotropy through the VCMA effect and induces the precessional motion of the magnetization around the external magnetic field. The pulse width is set to be about a half of the precession period to switch the magnetization. The direction of the magnetization in the free layer is represented by the unit vector $\mathbf{m} = (m_x, m_y, m_z)$. The magnetization unit vector in the reference layer \mathbf{p} is fixed to align with the positive z direction, $\mathbf{p} = (0, 0, 1)$. The x and y axes are considered as the in-plane directions, while the z axis is considered as the out-of-plane direction. The size of the nanopillar is assumed to be significantly small such that the magnetization dynamics can be described by the macrospin model.

Figure 1(b) shows the V dependence of the total anisotropy constant (K_u) of the free layer, comprising the crystalline anisotropy, interfacial anisotropy, and shape anisotropy. Hereafter, K_u is called the “anisotropy constant” for simplicity. Without application of the voltage,

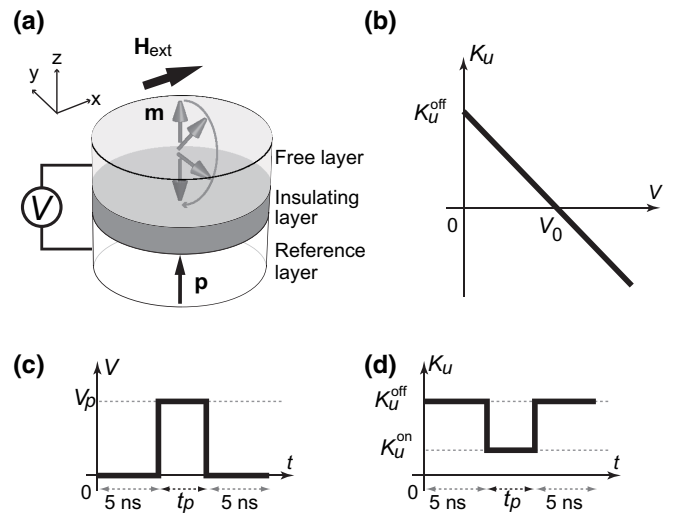


FIG. 1. (a) Magnetic tunnel junction and Cartesian definitions coordinates (x, y, z). The nonmagnetic insulating layer is sandwiched by the two ferromagnetic layers: the free layer and the reference layer. A static magnetic field (\mathbf{H}_{ext}) is applied along the positive x direction. The application of the voltage (V) pulse can induce precessional motion of the magnetization in the free layer through the VCMA effect. (b) Voltage (V) dependence of the uniaxial anisotropy constant, K_u . K_u is a linearly decreasing function of V , which assumes the value of K_u^{off} at $V = 0$ and vanishes at $V = V_0$. (c) Shape of the voltage pulse with amplitude V_p and duration t_p . A relaxation time of 5 ns is set before and after the pulse. (d) Time dependence of the anisotropy constant K_u . The anisotropy constants at $V = 0$ and V_p are represented by K_u^{off} and K_u^{on} , respectively.

the free layer is assumed to have out-of-plane uniaxial anisotropy, which is characterized by K_u^{off} . The variation of K_u^{off} is assumed to follow the normal distribution. We assume that K_u is a linear decreasing function of V and vanishes at $V = V_0$, that is, $K_u = (1 - V/V_0) K_u^{\text{off}}$. For $V > V_0$, K_u becomes negative, and therefore, the free layer has an easy-plane anisotropy.

The shape of the assumed voltage pulse is illustrated in Fig. 1(c), where V_p and t_p represent the amplitude and duration of the pulse, respectively. A relaxation time of 5 ns is set before and after the pulse. As shown in Fig. 1(d), the anisotropy constant is assumed to decrease to the value of K_u^{on} with the application of V_p , that is, $\Delta K_u = K_u^{\text{on}} - K_u^{\text{off}}$.

The dynamics of the magnetization unit vector in the free layer is obtained by solving the Landau-Lifshitz-Gilbert (LLG) equation,

$$\frac{d\mathbf{m}}{dt} = -\gamma \mathbf{m} \times \mathbf{H}_{\text{eff}} + \alpha \mathbf{m} \times \frac{d\mathbf{m}}{dt}, \quad (1)$$

where γ is the gyromagnetic constant and α is the Gilbert damping factor. The first and second terms on the right-hand side represent the torque resulting from the effective

field \mathbf{H}_{eff} and the damping torque, respectively. The effective field comprises the external field, anisotropy field (\mathbf{H}_{anis}), and thermal agitation field ($\mathbf{H}_{\text{therm}}$) as

$$\mathbf{H}_{\text{eff}} = \mathbf{H}_{\text{ext}} + \mathbf{H}_{\text{anis}} + \mathbf{H}_{\text{therm}}. \quad (2)$$

The anisotropy field is defined as

$$\mathbf{H}_{\text{anis}} = \frac{2K_u(t)}{\mu_0 M_s} m_z(t) \mathbf{e}_z, \quad (3)$$

where μ_0 is the vacuum permeability, M_s is the saturation magnetization of the free layer, and \mathbf{e}_z is the unit vector in the positive z direction. The thermal agitation field is determined by the fluctuation-dissipation theorem [40–44] and satisfies the following relations:

$$\langle H_{\text{therm}}^i(t) \rangle = 0, \quad (4)$$

$$\langle H_{\text{therm}}^i(t) H_{\text{therm}}^j(t') \rangle = \xi \delta_{ij} \delta(t - t'), \quad (5)$$

where the indices i and j denote the x , y , and z components of the thermal agitation field. δ_{ij} represents Kronecker's δ , and $\delta(t - t')$ represents the Dirac's δ function. The coefficient ξ is given by

$$\xi = \frac{2\alpha k_B T}{\gamma \mu_0 M_s \Omega}, \quad (6)$$

where k_B is the Boltzmann constant, T is temperature, and Ω is the volume of the free layer.

The following parameters are assumed for numerical calculations: $\alpha = 0.1$, and $M_s = 0.955$ MA/m [36]. The magnitude of the external field is $H_{\text{ext}} = 1$ kOe. The diameter of the free layer is 40 nm. The thickness of the free layer is 1.1 nm. We assume the constant voltage is applied with the magnitude where $K_u^{\text{off}} = 1.1 \times 10^5$ J/m³ [36] vanishes at $V = V_0$, that is, $\Delta K_u = -1.1 \times 10^5$ J/m³. The initial state of the magnetization is set $\mathbf{m} = (0, 0, 1)$, i.e., up state. The WERs are calculated from 10^7 trials for $T = 300$ K.

Figure 2 shows a color map of the WER on the t_p - K_u plane. The values of both K_u^{off} and K_u^{on} are shown in the plot. The value of K_u changes in the range of $0.6 \times 10^5 \leq K_u^{\text{off}} \leq 2.0 \times 10^5$ J/m³ at $V = 0$, corresponding to $-0.5 \times 10^5 \leq K_u^{\text{on}} \leq 0.9 \times 10^5$ J/m³ at $V = V_0$. The area with dark color is located around $t_p = 0.20$ ns and $K_u^{\text{on}} > 0$. The vertical dashed line expresses at $t_p = 0.18$ ns, which is a half period of the magnetization precession at $K_u^{\text{on}} = 0$. Along the dashed line, the K_u dependence of the WER is a downward convex function of K_u as shown by the blue curve with markers in Fig. 3.

The WER is minimized at the optimum value of $K_u^{\text{off}} = 1.23 \times 10^5$ J/m³, where the direction of the magnetization at the end of the pulse is close to the equilibrium direction of the down state at $V = 0$. Decrease of K_u^{off}

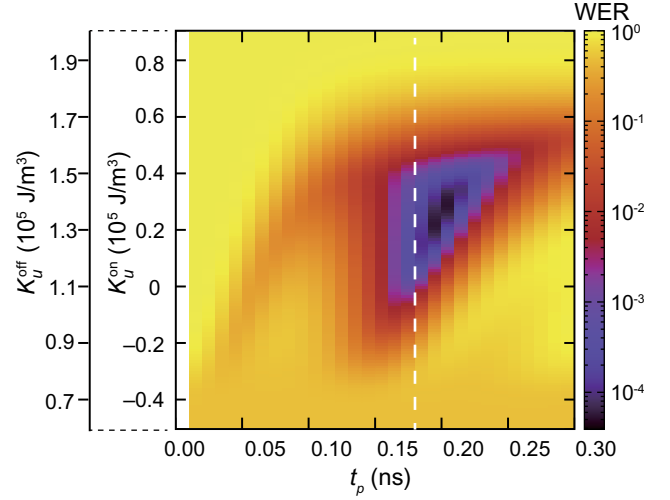


FIG. 2. Color map of the WER- K_u relation on the t_p - K_u plane with both scale of K_u^{on} and K_u^{off} , keeping ΔK_u constant as -1.1×10^5 J/m³. The WER along the dashed line at $t_p = 0.18$ is shown in Fig. 3.

from the optimal value induces the in-plane magnetic anisotropy, i.e., $K_u^{\text{on}} < 0$. The trajectory of the precession is squeezed to the equator, $m_z = 0$. The direction of the magnetization at the end of the pulse is close to the equator and has large probability to switch back to the up state during relaxation. Therefore, the WER increases with decrease of K_u^{off} for $K_u^{\text{off}} < 1.23 \times 10^5$ J/m³. If K_u^{off} is much smaller than the optimum value, the magnetization direction at the end of the pulse is almost in plane. After the pulse, the magnetization rotates to the up or down state with equal probability of 0.5.

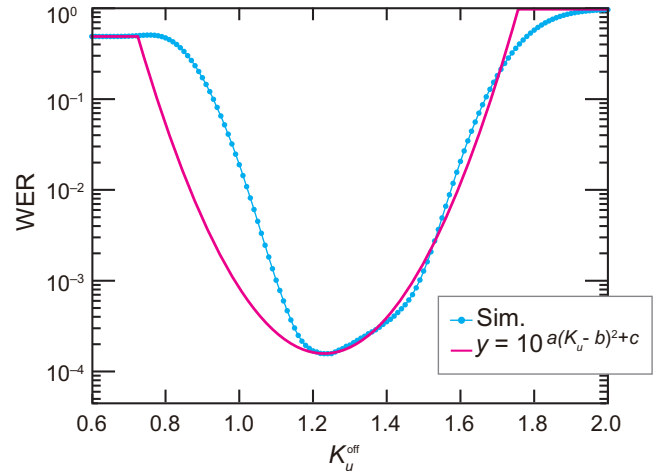


FIG. 3. K_u dependence of the WER at a pulse duration of 0.18 ns, shown by the (blue) line with markers. The solid (red) line shows the fitting result obtained using Eq. (8), with $a = 13.73$, $b = 1.23$, and $c = -3.81$. The vertical (gray) lines are a guide for the eyes. The unit for K_u is omitted for simplicity.

For $K_u^{\text{off}} > 1.23 \times 10^5 \text{ J/m}^3$, the WER is an increasing function of K_u^{off} and converges to unity at $K_u^{\text{off}} = 2.0 \times 10^5 \text{ J/m}^3$. In this region, the K_u^{on} is positive and the magnetization feels the uniaxial anisotropy field along the perpendicular to plane direction during the voltage pulse. The increase of the perpendicular anisotropy field has two effects on the WER. One is the change of the precession period, which shifts the direction of the magnetization at the end of the pulse from the equilibrium direction of the down state and increases the WER. The shift and therefore the WER increases with increase of K_u^{off} . The other is the enhancement of the thermal stability at $V = 0$, which reduces the WER because the thermal fluctuation of the magnetization is suppressed. The suppression of the thermal fluctuation is the origin of the slope around $K_u^{\text{off}} = 1.4 \times 10^5 \text{ J/m}^3$. If K_u^{off} is much larger than the optimal value, the magnetization stays along the initial direction and therefore the WER is unity.

III. ANALYTICAL DERIVATION OF PROBABILITY DENSITY FUNCTION OF THE WRITE-ERROR RATE

Let us consider the MTJ nanopillars having a manufacturing variation of K_u^{off} , which follows the normal distribution. The other material parameters are assumed to be constant. Hereafter, the superscript “off” and the unit of 10^5 J/m^3 for K_u are omitted for simplicity. The PDF for K_u is written as

$$f(K_u) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(K_u - \mu)^2}, \quad (7)$$

where μ and σ are the mean and standard deviation, respectively. We also assume that the K_u dependence of the WER is approximated by the following function:

$$y(K_u) = 10^{a(K_u - b)^2 + c}, \quad (8)$$

with a , b , and c representing constants. Because the WER is a downward convex function of K_u as shown in Fig. 3,

a is assumed to be positive. The value c requires $c < 0$ because y represents the WER, which satisfies $0 \leq y \leq 1$.

We model the WER shown by the blue curve with markers by the red parabola with $a = 13.73$, $b = 1.23$, and $c = -3.81$, as shown in Fig. 3. The parabola is assumed to have two cuts at WER = 0.5 (left) and WER = 1 (right). The values of a , b , and c are obtained as follows. The values of b and c are set to reproduce the minimum value of the WER. Then the value of a is obtained by fitting the WER using the least-squares method.

We need to find the PDF of the WER when K_u follows the normal distribution. Practically, one can independently control the anisotropy distribution and the operation condition. All the parameters, a , b , c , μ , and σ , are variable. To simplify the problem, the distribution of K_u is transformed to the standard normal distribution, implying that two parameters are fixed as $\mu = 0$ and $\sigma = 1.0$. Introducing a variance $z = (K_u - \mu)/\sigma$, one obtains the PDF of the standard normal distribution

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}. \quad (9)$$

The WER- K_u relation is transformed to

$$y(z) = d \times 10^{A(z-B)^2}, \quad (10)$$

where $d = 10^c$, $A = a\sigma^2$, and $B = (b - \mu)/\sigma$, i.e., d corresponds to the best value of the WER. Figure 4(a) shows the WER as a function of z modeled by Eq. (10), where the parameters are set $A = 1.0$, $B = 0.5$, and $d = 10^{-4}$ ($c = -4$) as an example, while Fig. 4(b) shows Eq. (9). To simplify the problem, we assume that $y(z)$ is symmetric with respect to $z = B$: $y(z)$ also converges to 1.0 as $z \rightarrow 0$. The problem is now aggregated to investigate the influence of the parameters A , B , and d , which represent the degree of curvature, vertex, and the depth of the parabola, respectively.

The PDF is obtained using the change-of-variable technique as

$$\begin{cases} g(y) = \frac{N}{2} \left(\frac{\ln \bar{y}}{A \ln 10} \right)^{-\frac{1}{2}} \frac{1}{\bar{y} A d \ln 10} \frac{1}{\sqrt{2\pi}} \left[e^{-\frac{1}{2} \left(B + \sqrt{\frac{\ln \bar{y}}{A \ln 10}} \right)^2} + e^{-\frac{1}{2} \left(B - \sqrt{\frac{\ln \bar{y}}{A \ln 10}} \right)^2} \right], & \left(1 < \bar{y} \leq \frac{1}{d} \right), \\ g(y) = 0 & , \quad (\bar{y} \leq 1) \end{cases} \quad (11)$$

where $\bar{y} = y/d$, and \ln denotes the natural logarithm (\log_e). N is a normalization constant written using the Gauss error function, $\text{erf}(x)$,

$$N = \frac{2}{\text{erf} \left(\sqrt{\frac{\ln(1/d)}{2A \ln 10}} + \frac{B}{\sqrt{2}} \right) + \text{erf} \left(\sqrt{\frac{\ln(1/d)}{2A \ln 10}} - \frac{B}{\sqrt{2}} \right)}. \quad (12)$$

The normalization constant is needed because the range of y is limited to be $y \leq 1$. Apparently, $g(y)$ is an even function for B . One can assume $B > 0$ or $b - \mu > 0$ without loss of generality.

To investigate the property of $g(y)$, let us differentiate $g(y)$ with y .

$$g'(y) = \frac{N}{4} \left(\frac{\ln \bar{y}}{A \ln 10} \right)^{-\frac{3}{2}} \left(\frac{1}{\bar{y} A d \ln 10} \right)^2 \frac{1}{\sqrt{2\pi}} \times \left\{ \left[e^{-\frac{1}{2} \left(B + \sqrt{\frac{\ln \bar{y}}{A \ln 10}} \right)^2} + e^{-\frac{1}{2} \left(B - \sqrt{\frac{\ln \bar{y}}{A \ln 10}} \right)^2} \right] \left[-1 - 2A \ln 10 \left(\frac{\ln \bar{y}}{A \ln 10} \right) \right] + \left(\frac{\ln \bar{y}}{A \ln 10} \right)^{\frac{1}{2}} \left[-e^{-\frac{1}{2} \left(B + \sqrt{\frac{\ln \bar{y}}{A \ln 10}} \right)^2} \left(B + \sqrt{\frac{\ln \bar{y}}{A \ln 10}} \right) + e^{-\frac{1}{2} \left(B - \sqrt{\frac{\ln \bar{y}}{A \ln 10}} \right)^2} \left(B - \sqrt{\frac{\ln \bar{y}}{A \ln 10}} \right) \right] \right\}. \quad (13)$$

Designating $\sqrt{\frac{\ln \bar{y}}{A \ln 10}}$ as m and using the expression

$$\frac{e^{-\frac{1}{2}(B+m)^2}}{e^{-\frac{1}{2}(B-m)^2}} = e^{-2Bm}, \quad (14)$$

one can obtain the following equation from $g'(y) = 0$:

$$(1 + e^{-2Bm}) (1 + 2Am^2 \ln 10) + m [e^{-2Bm} (B + m) - B + m] = 0. \quad (15)$$

By taking the limit $e^{-2Bm} \rightarrow 0$, it follows

$$1 + 2Am^2 \ln 10 - Bm + m^2 = 0. \quad (16)$$

Equations (15) and (16) can be plotted as a contour plot on the m - B plane for a given A . Figure 5 shows the plot for $A = 1.0$. Both curves are quite similar. Equation (16) is to be analyzed instead of Eq. (15). Furthermore, the maximum value of m exists, and it is given as

$$m_{\max} = \sqrt{\frac{-c}{A}}, \quad (17)$$

which corresponds to $\bar{y} = 1/d$ or $y = 1$. We obtain $m_{\max} = 2.0$ for the case where $A = 1.0$ and $c = -4$.

Let us analyze Eq. (16). First, we obtain the minimum of B , B_{\min} . By representing Eq. (16) in the form of a function of m :

$$B = m(1 + 2A \ln 10) + \frac{1}{m} \quad (18)$$

is obtained. By the first and second derivative test, B has a local minimum

$$B_{\min} = 2\sqrt{1 + 2A \ln 10} \quad (19)$$

at $m_0 = (\sqrt{1 + 2A \ln 10})^{-1}$.

If a given B satisfies $B_{\min} < B$, Eq. (16) has a solution. Solving Eq. (16) for m , we obtain the solutions of m at which the critical points of $g(y)$ are located,

$$m_{\pm} = \frac{B \pm \sqrt{B^2 - 4(1 + 2A \ln 10)}}{2(1 + 2A \ln 10)}. \quad (20)$$

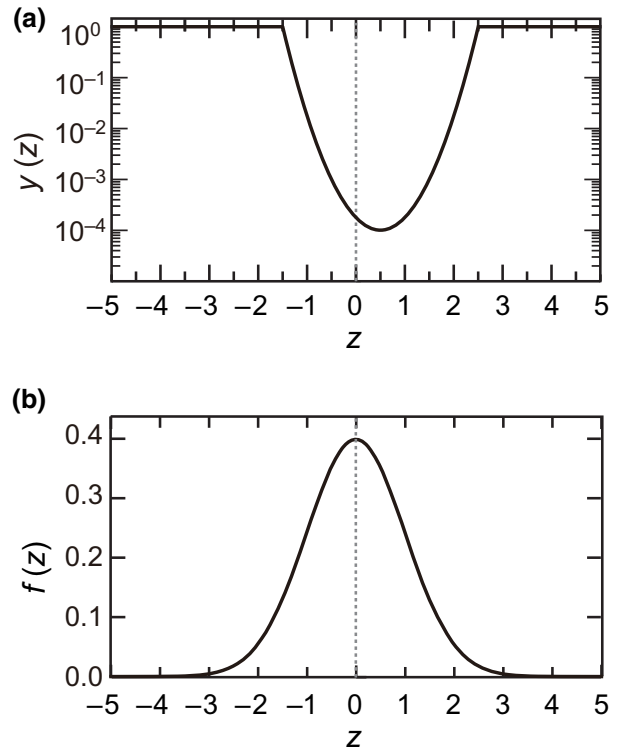


FIG. 4. (a) WER- K_u relation given by Eq. (10), with $A = 1.0$, $B = 0.5$, and $d = 10^{-4}$. (b) Standard normal distribution given by Eq. (9). The mean and standard variance for z , that is, K_u , is fixed to 0 and 1.0, respectively.

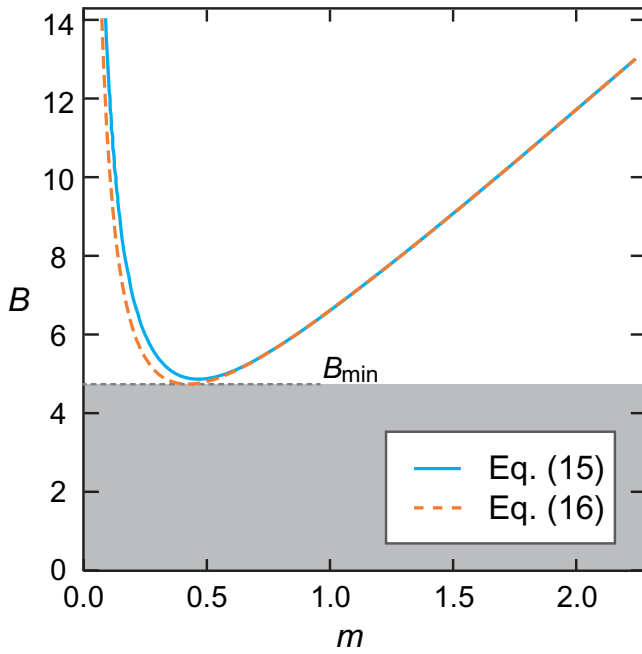


FIG. 5. Curves obtained from contour plots solved by Eqs. (15) and (16), with $A = 1.0$ on the m - B plane. B_{\min} describes the minimum B in Eq. (16). In the shaded area satisfying $B < B_{\min}$, the PDF is a monotonically decreasing function of the WER.

Thus, the critical points of $g(y)$ are located at

$$y_{\pm} = d \times 10^{Am_{\pm}^2}. \quad (21)$$

In the shaded area in Fig. 5, satisfying $B < B_{\min}$, the PDF is a monotonic decreasing function of WER within the range of considered parameters as shown below.

A cumulative distribution function (CDF) for $g(y)$ is given:

$$F(x) = \frac{N}{2} \left[\operatorname{erf} \left(\sqrt{\frac{\ln(x/d)}{2A \ln 10}} + \frac{B}{\sqrt{2}} \right) + \operatorname{erf} \left(\sqrt{\frac{\ln(x/d)}{2A \ln 10}} - \frac{B}{\sqrt{2}} \right) \right]. \quad (22)$$

The details of the derivation of PDF and CDF are summarized in the Appendix.

IV. GENERAL PROPERTY FOR THE PROBABILITY DENSITY FUNCTION

Let us investigate the effects of some parameters on the PDF of the WER. Figure 6 shows the curve of B_{\min} as a function of A , given in Eq. (19). If the values of A and B are present in the shaded area under the line, $g(y)$ has no extreme values. However, if the values are both in the area above the line, $g(y)$ has extreme values.

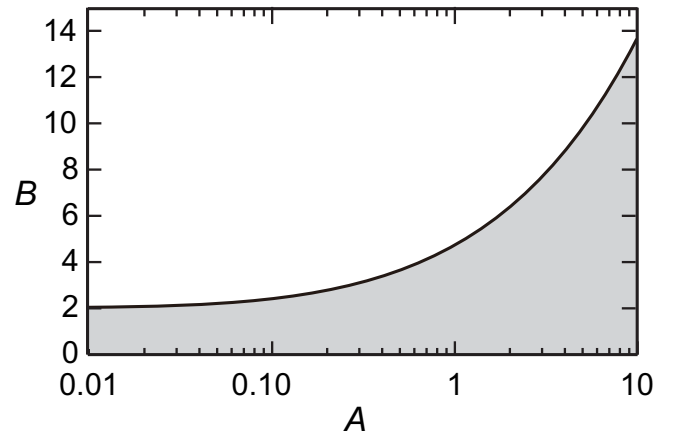


FIG. 6. Curve of B_{\min} as a function of A , given in Eq. (19). When the parameters A and B are both in the shaded area, the PDF is a monotonically decreasing function, whereas when A and B are both in the area above the line, the PDF has some extreme values.

The typical shapes of $g(y)$ are summarized in Fig. 7 with different B . The parameters except for B are set as $A = 1.0$ and $d = 10^{-4}$. At $B = 0$, that is, $b = \mu$, it corresponds that the majority of MTJs can operate with the best WER. In this case, as shown in Fig. 7(a), the shape of the PDF has a sharp peak at $y = d$. Figure 7(b) shows the case $B = 4.0$, which still satisfies $B < B_{\min} = 4.735$. The shape of the PDF is similar to that shown in Fig. 7(a), although the contribution of the probability density at higher WERs increases. In the case of Figs. 7(c) and 7(d), the PDFs have extreme values because $B > B_{\min}$ is satisfied. The dashed and dotted vertical lines express the positions at y_- and y_+ given in Eq. (21), respectively. Note that the increase of B implies the value of K_u with the best WER deviates from μ . As a result, the local maximum of the PDF shifts to a higher WER side with the increase of B . With a further increase in B , one extreme value of the PDF disappears as shown in Fig. 7(e).

The range of B showing the local maximum in $g(y)$ depends on A . Figure 8 shows y_+ as a function of B with different A . The dashed vertical line expresses the position at B_{\min} , which is 2.4 for Fig. 8(a), 4.7 for Fig. 8(b), and 13.7 for Fig. 8(c). With the increase of A , the range of B giving $d < y_+ < 1$ becomes wider.

V. PRACTICAL PROPERTY FOR THE DISTRIBUTION OF THE WRITE-ERROR RATE

A. Frequency distribution

Practically, the shape of the frequency distribution of the WER, that is, a histogram for the WER, is more relevant than that of the PDF. In addition, the shapes of the frequency distribution would be different from that of the PDF owing to a logarithmic scale in the horizontal axis. In this

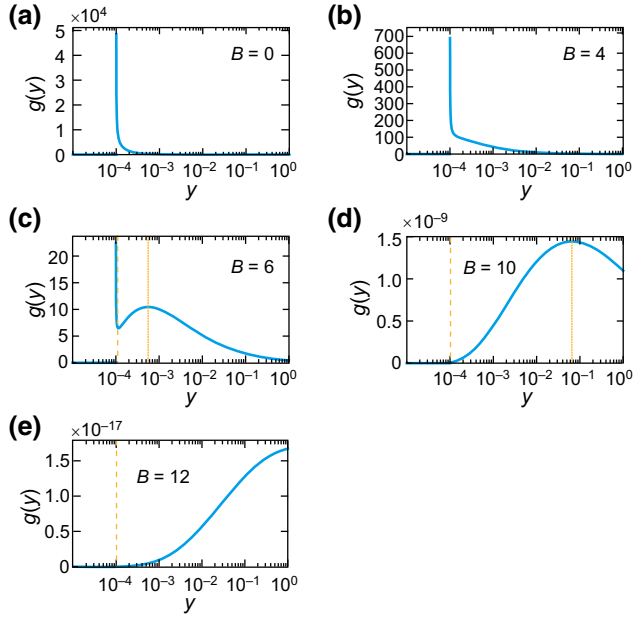


FIG. 7. Probability density functions of WER with different B . Other parameters are set as $A = 1.0$, and $d = 10^{-4}$. Dashed and dotted vertical lines express y_- and y_+ , given by Eq. (21), respectively.

section, the frequency distribution of the WER is investigated. One way to produce a frequency distribution from a PDF is to multiply the sample number and the probability obtained by integrating the PDF for every interval corresponding to a bin width. Another way is to use the inverse transform method, which produces a distribution of random variables with any PDF. The algorithm of the inverse transform method is expressed as follows:

1. Generate a random number u from the uniform distribution in the interval $[0, 1]$ ($\mathbf{u} = \{u_1, u_2, \dots, u_n\}$).
2. Solve $F(x) = u$ for x numerically for each element of \mathbf{u} , then obtain a set $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$.

The obtained $\{x_1, x_2, \dots, x_n\}$ shows the distribution of random variables with the given PDF.

To avoid any trouble on numerical integration because of a small number on $g(y)$ and the WER value, the histogram of the WER is directly produced by the inverse transform method. Figure 9 summarizes the PDFs, $g(y)$, in the left column, the corresponding CDFs, $F(x)$, in the middle column, and the frequency distributions of the WER obtained by inverse transform method in the right column with a different parameter set. The parameters for Fig. 9(a) are set $a = 13.73$, $b = 1.23$, $c = -3.81$, $\mu = 1.1$, and $\sigma = 0.11$, which corresponds to $A = 0.166$, $B = 1.182$. The parameters for Fig. 9(b) are similar to those in Fig. 9(a), but $b = 1.45$, i.e., $B = 3.18$. The parameters for Fig. 9(c) are similar to those in Fig. 9(a), but $\sigma = 0.011$, i.e., $A = 1.66 \times 10^{-3}$ and $B = 11.8$. Because the number of random

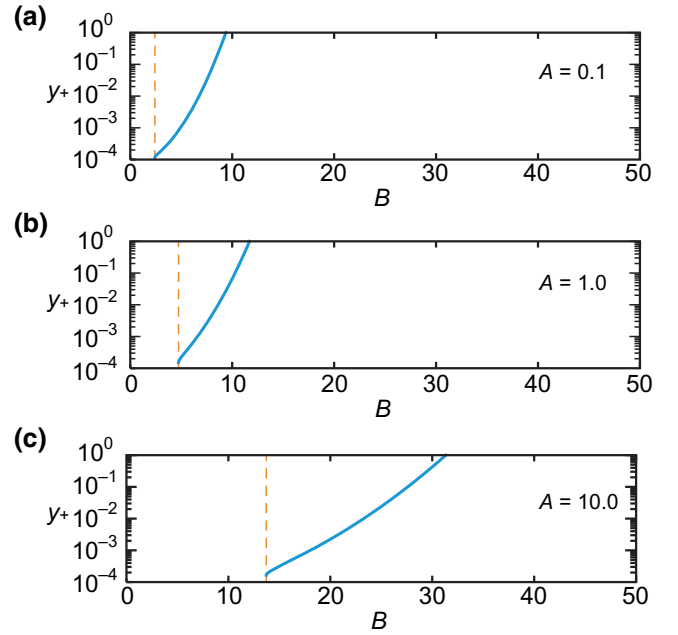


FIG. 8. Curves of y_+ as a function of B with different A . The value of d is fixed, $d = 10^{-4}$. The dashed line expresses the position at B_{\min} , which is given by Eq. (19).

numbers from the uniform distribution are set $n = 1000$, there are 1000 elements in \mathbf{x} . In the middle column of Fig. 9, the gray solid line shows Eq. (22), and the red points show a part of \mathbf{x} . One obtains $B_{\min} = 2.66$ for $A = 0.166$, and $B_{\min} = 2.01$ for $A = 1.66 \times 10^{-3}$. As aforementioned, the PDF is a monotonic decreasing function as shown in Fig. 9(a), while the PDF has local maximum in Figs. 9(b) and 9(c).

It is different between the shape of the PDF in the left column and the shape of the frequency distribution in the right column of Fig. 9. As the horizontal axis of the PDF is a logarithmic scale, the left side of the PDF has a smaller contribution than the right side. We note that the contribution of larger WER is more emphasized in Fig. 9(b), resulting from the value range of the PDF is relatively smaller than other cases in Fig. 9. For the application to conventional memory system, it would be expected the WER distribution biased to the left or bell-shaped distribution with narrow width. On the other hand, the WER distribution with wider width is applicable for a hybrid memory system. We also note that the difference between Figs. 9(a) and 9(c) results from the value of σ . By reducing σ , the spread of the histogram is significantly suppressed, with changing the shape of the histogram.

B. Contribution of the standard deviation

In the previous section, it is found that the frequency distribution of the WER significantly depends on the value

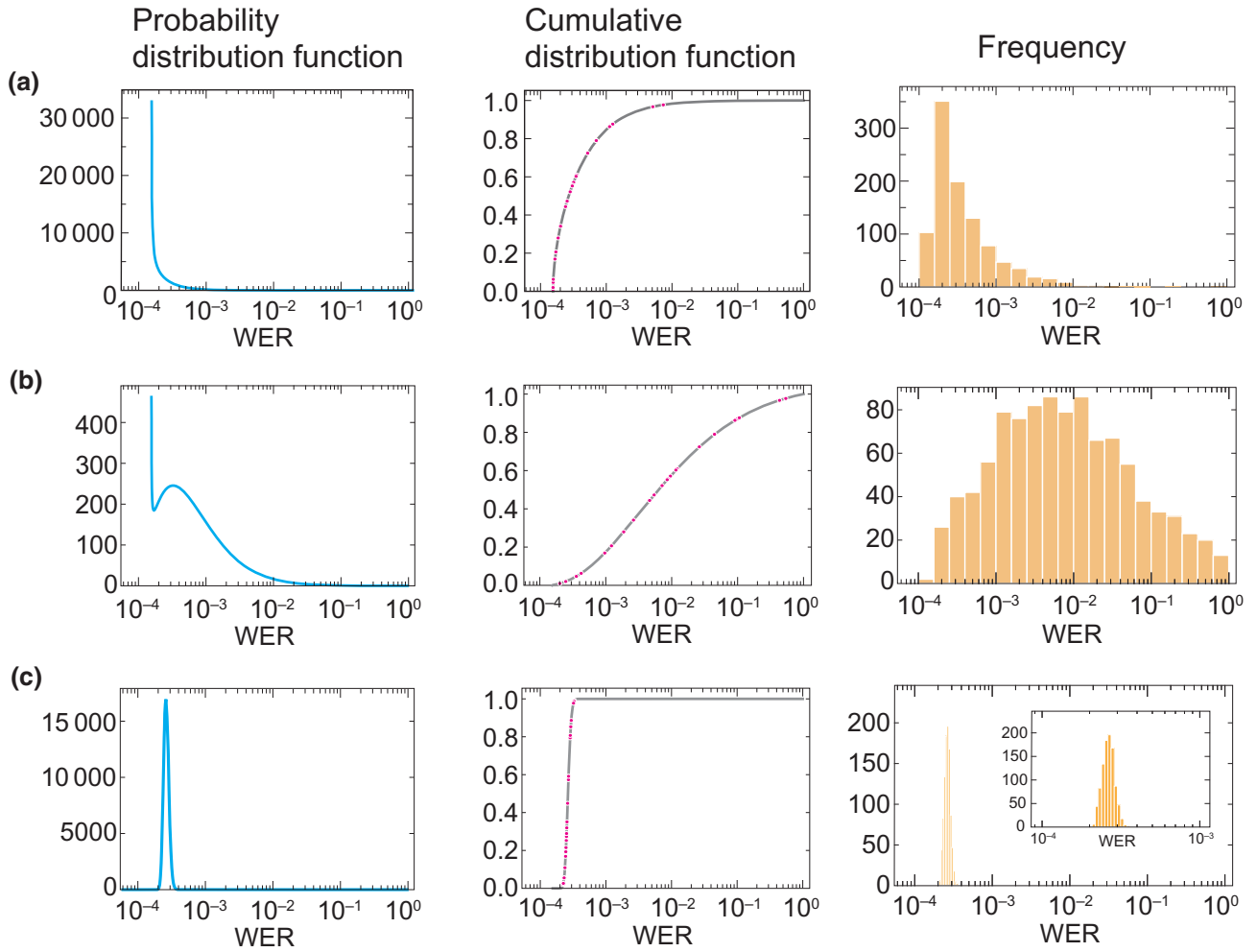


FIG. 9. PDFs in the left column, CDFs in the middle column, and frequency distributions in the right column. Parameters for (a) are set $a = 13.73$, $b = 1.23$, $d = 1.564 \times 10^{-4}$, i.e., $c = -3.81$, $\mu = 1.1$, and $\sigma = 0.11$, corresponding, $A = 0.166$ and $B = 1.182$. In (b), $b = 1.45$, i.e., $B = 3.18$. In (c), $\sigma = 0.011$, i.e., $A = 1.66 \times 10^{-3}$ and $B = 11.8$. The other parameters are the same in (a). In the middle column, the points show a part of data that is prepared to draw the histogram shown in the right column. All histograms are illustrated with 20 bins. In (c), the enlarged view is also shown.

of σ . The other parameters can contribute to the frequency distribution of the WER. However, other parameters, such as a , b , and c result from magnetization dynamics, i.e., the condition of the magnetization switching operation. The operation condition is to be optimized for the aim of applications. This implies that the most useful parameter to control the frequency distribution is σ . In this section, we focus on the contribution of σ .

First, we note that the upper limit of σ is required in the range of K_u to ensure the perpendicularly magnetized MTJs under the assumption of the normal distribution of K_u . For example, if $\sigma = 0.33$ and $\mu = 1.1$, the range of “5-sigma” corresponds to $-0.55 \leq K_u \leq 2.75$. It reaches $K_u < 0$, that is, in-plane magnetization state. If one would like to prepare perpendicularly magnetized MTJs within the range of 5-sigma, the upper limit of the variance is to be $\sigma = 0.22$.

If the parameters a , b , c , and μ are fixed, the condition $B_{\min} < B$ is rewritten in terms of σ :

$$2\sqrt{1 + 2A \ln 10} < \frac{b - \mu}{\sigma}. \quad (23)$$

Solving the equation for σ with $\sigma > 0$, one obtains

$$\sigma = \sqrt{\frac{-4 + \sqrt{16 + 32a(\ln 10)(b - \mu)^2}}{16a \ln 10}} \equiv \sigma_{\text{th}}. \quad (24)$$

Thus, one obtains $\sigma < \sigma_{\text{th}}$, instead of $B > B_{\min}$. Figure 10 shows σ_{th} as a function of $b - \mu$ with $a = 13.73$. The crosses describe the parameter set of $(b - \mu, \sigma)$, as mentioned in Figs. 9(a)–9(c). When $b - \mu = 0.13$, corresponding to Figs. 9(a) and 9(c), $\sigma_{\text{th}} = 0.059$. The shaded area corresponds to the shaded area in Fig. 6.

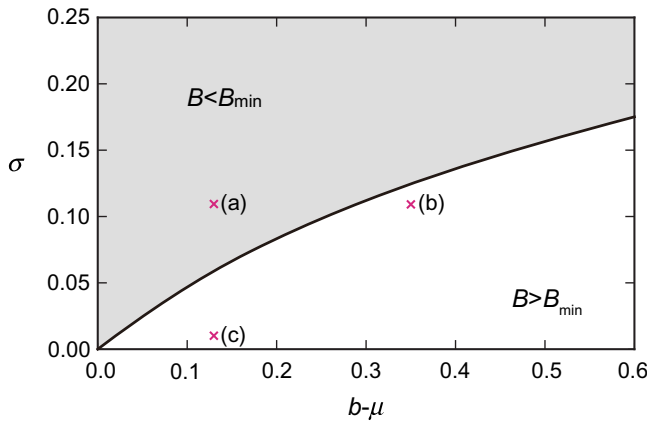


FIG. 10. Curve of σ_{th} as a function of $b - \mu$, given by Eq. (24). The parameters are fixed as $a = 13.73$. The area under the line satisfies $B > B_{min}$, while the shaded area above the line satisfies $B < B_{min}$. The crosses with an alphabet indicate the set of the parameters used in Fig. 9.

Finally, the analysis for the shape of the frequency distribution of the WER is presented. Figure 11 shows the property for the distribution of the WER. The color bars in each figure show the mean, standard deviation, skewness, and kurtosis, respectively. They are the average values for the ten trials with 1000 data obtained by the inverse transform method. The value of $(b - \mu)$ changes to 0, 0.13 and from 0.25 to 0.55 with 0.1 steps. The value of σ changes to 0.011, 0.03, 0.05, 0.07, 0.09, 0.11, and 0.22. The other parameters are fixed: $a = 13.73$ and $c = -3.81$. The dashed line shows σ_{th} given in Eq. (24). With increasing $b - \mu$, and also increasing σ , the mean of the WER increases, as shown in Fig. 11(a). A similar behavior is found on the standard deviation of the WER, as shown in Fig. 11(b). Skewness shown in Fig. 11(c) and kurtosis shown in Fig. 11(d) describe the histogram's shape. If the skewness is positive, it indicates that the tail of the histogram is on the right. Since the value of skewness at $(b - \mu, \sigma) = (0.13, 0.11)$ is 1.96, the tail of the histogram is on the right, as seen in Fig. 9(a). Conversely, since the value of skewness at $(0.35, 0.11)$ is 0.322, the histogram shows almost symmetric shape, as shown in Fig. 9(b). Since the value of skewness at $(0.13, 0.011)$ is 0.282, as shown in Fig. 9(c), the histogram of the WER has a bell shape. Kurtosis describes whether the tail of the distribution is heavy or light. In the definition adopted here, kurtosis equals 0 for the normal distribution. If kurtosis is positive, the histogram has heavier tails compared with the normal distribution. If kurtosis is negative, the histogram has lighter tails. The value of kurtosis at $(0.13, 0.011)$ corresponding to Fig. 9(c) is 0.11. We note that even if $b - \mu = 0$, that is, anisotropy constants for the axis of symmetry of the parabola of the WER for a single memory cell matches the value of the mean of the normal distribution of the anisotropy constant for VC MRAM, the

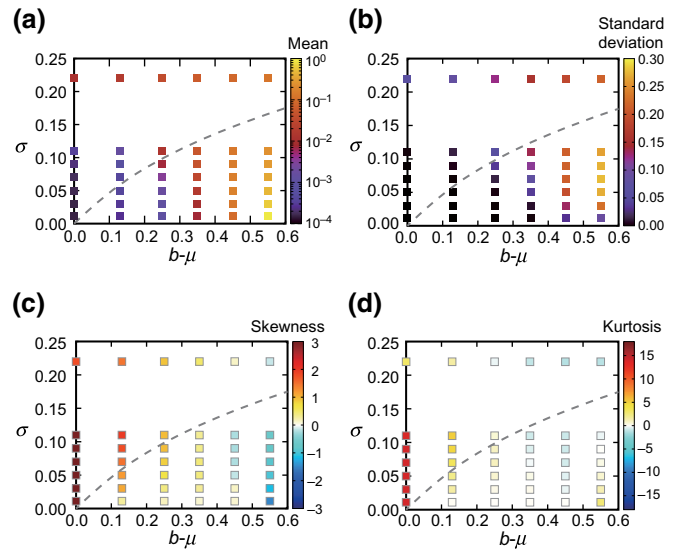


FIG. 11. Property for the distribution of the WER on the $(b - \mu) - \sigma$ plane. (a) Mean of data, (b) standard deviation, (c) skewness, and (d) kurtosis. The dashed line shows σ_{th} , given by Eq. (24).

frequency distribution of the WER shows skewed distribution, which has the largest frequency at the smallest WER, and not bell-shape distribution. Furthermore, even if the distributions have similar means and standard deviations, these shapes are not always similar. For example, let us focus on the following two cases, at $(0, 0.09)$, and at $(0.25, 0.011)$. The values of mean are 2.3×10^{-4} at $(0, 0.09)$, and 1.1×10^{-3} at $(0.25, 0.011)$. The values of standard deviation are 2.8×10^{-4} at $(0, 0.09)$, and 2.0×10^{-4} at $(0.25, 0.011)$. Although these values are similar to each other, the shapes of the frequency distribution of the WER are different. The value of skewness is 2.97 at $(0, 0.09)$, where the tail of the histogram is on the right. Conversely, the value of skewness is 0.17 at $(0.25, 0.011)$, where the histogram resembles a bell shape. These results help us to design a specification depending on applications, for example, DRAM replacement or approximate computing. For the application to conventional memory system which is expected to replace DRAM, the WER is expected to be uniform with low WER. It requires the WER distribution with narrow width, i.e., positive kurtosis. It also requires that most of the memory cells have low WER, i.e., 0 or positive skewness. On the other hand, error permissive applications such as erroneous memory and approximate computing loosens those requirements. The WER distribution itself becomes more relevant.

VI. SUMMARY

In this study, the distribution of the WER of the VC-MRAM owing to the manufacturing variation of K_u is theoretically explored. An analytical expression for the

PDF is obtained, and then analyzed to characterize the shape of the PDF. The frequency distribution of the WER is also obtained numerically. The shape of the frequency distribution or histogram is completely different from the normal distribution of material parameter K_u we assume, and strongly depends on the operation conditions of the VC MRAM and the variation of K_u . Nonvolatile memories, including the VC MRAM have the distribution of the errors owing to the manufacturing variation. This study ushers in a basic understanding for alternative design of computing systems exploiting no or less error-corrected memories.

ACKNOWLEDGMENTS

This work is partly supported by JSPS KAKENHI Grants No. JP19H01108 and No. JP20K12003.

APPENDIX: MATHEMATICAL SUPPLEMENTS

1. Change-of-variable technique

Let X be a continuous random variable with a PDF $f_X(x)$. A probability is obtained by integrating the PDF,

$$P(X \in A) = \int_A f_X(x) dx, \quad (\text{A1})$$

where A denotes a possible set of X . The cumulative distribution function of X is defined as

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(z) dz, \quad (\text{A2})$$

if f_X is continuous at x . Therefore, one can obtain

$$\frac{d}{dx} F_X(x) = f_X(x). \quad (\text{A3})$$

Given a new variable $Y = h(X)$, let us consider the PDF of Y , $g_Y(y)$. The PDF can be obtained using Eq. (A3), such that one should calculate $P(Y \leq y)$.

If $h(x)$ is an increasing function,

$$\begin{aligned} P(Y \leq y) &= P[X \leq h^{-1}(y)] \\ &= \int_{-\infty}^{h^{-1}(y)} f_X(x) dx. \end{aligned} \quad (\text{A4})$$

Therefore, one can obtain

$$g_Y(y) = f_X[h^{-1}(y)] \times \frac{d}{dy} h^{-1}(y). \quad (\text{A5})$$

Conversely, if $h(x)$ is a decreasing function,

$$\begin{aligned} P(Y \leq y) &= P[h^{-1}(y) \leq X] \\ &= 1 - P[X \leq h^{-1}(y)] \\ &= 1 - \int_{-\infty}^{h^{-1}(y)} f_X(x) dx. \end{aligned} \quad (\text{A6})$$

Therefore, one can obtain

$$g_Y(y) = -f_X[h^{-1}(y)] \times \frac{d}{dy} h^{-1}(y). \quad (\text{A7})$$

In the case of Eq. (10),

$$h^{-1}(y) = z = B \pm \sqrt{\frac{\ln(y/d)}{A \ln 10}}, \quad (\text{A8})$$

where $h^{-1}(y)$ with the upper (lower) sign corresponds to that in Eq. (A5) [Eq. (A7)], that is, for the range $B < z$ ($z < B$). Finally, one obtains

$$\begin{aligned} g(y) &= \frac{1}{2} \left(\frac{\ln \bar{y}}{A \ln 10} \right)^{-\frac{1}{2}} \frac{1}{\bar{y} A d \ln 10} \frac{1}{\sqrt{2\pi}} \\ &\times \left[e^{-\frac{1}{2} \left(B + \sqrt{\frac{\ln \bar{y}}{A \ln 10}} \right)^2} + e^{-\frac{1}{2} \left(B - \sqrt{\frac{\ln \bar{y}}{A \ln 10}} \right)^2} \right], \quad (\bar{y} > 1) \end{aligned} \quad (\text{A9})$$

without the normalization constant.

2. Cumulative distribution function

A CDF is calculated from the definition:

$$\begin{aligned} F(x) &= \int_{-\infty}^x g(y) dy \\ &= \int_{-\infty}^d g(y) dy + \int_d^x g(y) dy \\ &= \int_1^{x/d} g(\bar{y}) d \times d\bar{y} \\ &= \int_0^{m_1} g(m) d 2m \ln 10 \times 10^{Am^2} dm, \end{aligned}$$

where $m_1 = \sqrt{\frac{\ln(x/d)}{A \ln 10}}$. Rearranging the equation, we obtain

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_0^{m_1} \left[e^{-\frac{1}{2}(m+B)^2} + e^{-\frac{1}{2}(m-B)^2} \right] dm. \quad (\text{A10})$$

Introducing the Gauss error function:

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \quad (\text{A11})$$

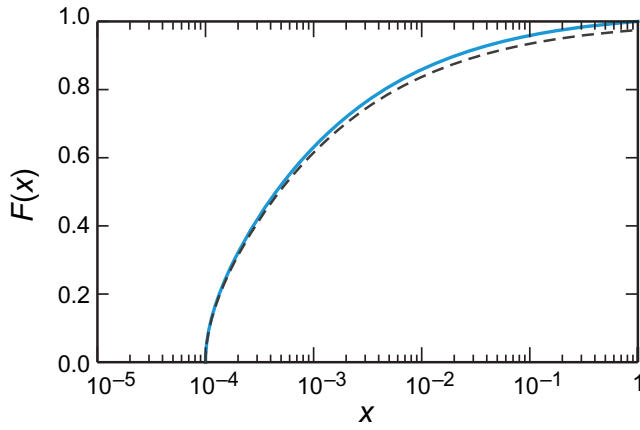


FIG. 12. Cumulative distribution functions given by Eq. (A12) shown as the dashed line, and given by Eq. (22), as shown by the solid (blue) line. The parameters are mentioned in the text.

and using the property $\text{erf}(-x) = -\text{erf}(x)$, one obtains the CDF,

$$F(x) = \frac{1}{2} \left[\text{erf} \left(\frac{m_1 + B}{\sqrt{2}} \right) + \text{erf} \left(\frac{m_1 - B}{\sqrt{2}} \right) \right] \quad (\text{A12})$$

without the normalization constant. The CDF is required to reach 1 at $x = 1$, because the variable x implies the WER. However, Eq. (A12) does not satisfy the requirement, because $y(z)$ is cut off above $y \geq 1$, as shown in Fig. 4(a). To satisfy the requirement on the CDF, a normalization constant is derived from

$$N \int_{-\infty}^{+\infty} g(y) dy = 1. \quad (\text{A13})$$

Then, one obtains

$$\begin{aligned} N &= \frac{1}{F(x=1)} \\ &= \frac{2}{\text{erf} \left(\sqrt{\frac{\ln(1/d)}{2A \ln 10} + \frac{B}{\sqrt{2}}} \right) + \text{erf} \left(\sqrt{\frac{\ln(1/d)}{2A \ln 10} - \frac{B}{\sqrt{2}}} \right)}. \end{aligned} \quad (\text{A14})$$

Figure 12 shows the CDFs. The dashed line shows Eq. (A12), and the solid (blue) line shows Eq. (22). The parameters are set $A = 0.363$, $B = 1.364$, and $d = 10^{-4}$ ($a = 30$, $b = 1.25$, $c = -4$, $\mu = 1.1$, and $\sigma = 0.11$) as an example. By making the normalization constant, Eq. (22) reaches 1 at $x = 1$.

3. Analysis of histogram

As aforementioned, a set of data $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ is obtained through the inverse transform method. The mean

for a set of data is defined as

$$\bar{x} = \frac{1}{n} \sum_i^n x_i. \quad (\text{A15})$$

The standard deviation is defined as

$$\bar{\sigma} = \sqrt{\frac{1}{n} \sum_i^n (x_i - \bar{x})^2}. \quad (\text{A16})$$

There are some definitions for skewness and kurtosis. The following definition of skewness is used in this study:

$$\text{skewness} = \frac{\mu_3}{\mu_2^{3/2}}, \quad (\text{A17})$$

where μ_r denotes r th central moment defined as

$$\mu_r = \frac{1}{n} \sum_i^n (x_i - \bar{x})^r. \quad (\text{A18})$$

Note that μ_2 is the variance, denoted by $\bar{\sigma}^2$. The following definition of kurtosis is used in this study:

$$\text{kurtosis} = \frac{\mu_4}{\mu_2^2} - 3, \quad (\text{A19})$$

which is equal to *excess* kurtosis. We note that to analyze the histogram of the WER with logarithmic scale, skewness and kurtosis are calculated for the data of $\ln \mathbf{x} = \{\ln x_1, \ln x_2, \dots, \ln x_n\}$.

Using ten different seeds to generate a set of random variables from a uniform distribution, ten sets of data, $\mathbf{x}_1, \mathbf{x}_2, \dots$, and \mathbf{x}_{10} , are obtained. The values of mean, standard deviation, skewness, and kurtosis are calculated for each of the ten trials. The values for the average of ten trials are then obtained and summarized in Fig. 11.

- [1] R. Baumann, Soft errors in advanced computer systems, *IEEE Design Test Comput.* **22**, 258 (2005).
- [2] B. Schroeder, E. Pinheiro, and W.-D. Weber, in *Proceedings of the eleventh international joint conference on Measurement and modeling of computer systems*, SIGMETRICS '09 (Association for Computing Machinery, New York, NY, USA, 2009), p. 193.
- [3] W. Zhao, Y. Zhang, T. Devolder, J. Klein, D. Ravelosona, C. Chappert, and P. Mazoyer, Failure and reliability analysis of STT-MRAM, *Microelectron. Reliab.* **52**, 1848 (2012).
- [4] Y. Ibrahim, H. Wang, J. Liu, J. Wei, L. Chen, P. Rech, K. Adam, and G. Guo, Soft errors in DNN accelerators: A comprehensive review, *Microelectron. Reliab.* **115**, 113969 (2020).
- [5] S. Mittal, A survey of techniques for approximate computing, *ACM Comput. Surv.* **48**, 62:1 (2016).

- [6] Q. Xu, T. Mytkowicz, and N. S. Kim, Approximate computing: A survey, *IEEE Design Test* **33**, 8 (2016).
- [7] C. Torres-Huitzil and B. Girau, Fault and error tolerance in neural networks: A review, *IEEE Access* **5**, 17322 (2017).
- [8] M. Qin, C. Sun, and D. Vucinic, Improving robustness of neural networks against bit flipping errors during inference, *J. Image Graphics* **6**, 181 (2018).
- [9] A. Bosio, P. Bernardi, A. Ruospo, and E. Sanchez, in *2019 IEEE Latin American Test Symposium (LATS)* (2019), p. 1, iSSN: 2373-0862.
- [10] S. Mittal and J. S. Vetter, Reliability tradeoffs in design of volatile and nonvolatile caches, *J. Circuits Syst. Comput.* **25**, 1650139 (2016).
- [11] T. Hirofuchi and R. Takano, RAMinate: Hypervisor-based Virtualization for Hybrid Main Memory Systems, SoCC '16: Proceedings of the Seventh ACM Symposium on Cloud Computing, 112 (2016).
- [12] R. Salkhordeh, O. Mutlu, and H. Asadi, An analytical model for performance and lifetime estimation of hybrid DRAM-NVM main memories, *IEEE Trans. Comput.* **68**, 1114 (2019).
- [13] A. V. Khvalkovskiy, D. Apalkov, S. Watts, R. Chepulskaa, R. S. Beach, A. Ong, X. Tang, A. Driskill-Smith, W. H. Butler, P. B. Visscher, D. Lottis, E. Chen, V. Nikitin, and M. Krounbi, Basic principles of STT-MRAM cell operation in memory arrays, *J. Phys. D: Appl. Phys.* **46**, 074001 (2013).
- [14] D. Apalkov, B. Dieny, and J. M. Slaughter, Magnetoresistive random access memory, *Proc. IEEE* **104**, 1796 (2016).
- [15] S. Bhatti, R. Sbiaa, A. Hirohata, H. Ohno, S. Fukami, and S. Piramanayagam, Spintronics based random access memory: A review, *Mater. Today* **20**, 530 (2017).
- [16] T. Nozaki, T. Yamamoto, S. Miwa, M. Tsujikawa, M. Shirai, S. Yuasa, and Y. Suzuki, Recent progress in the voltage-controlled magnetic anisotropy effect and the challenges faced in developing voltage-torque MRAM, *Micromachines* **10**, 327 (2019).
- [17] G. W. Burr, M. J. Breitwisch, M. Franceschini, D. Garetto, K. Gopalakrishnan, B. Jackson, B. Kurdi, C. Lam, L. A. Lastras, A. Padilla, B. Rajendran, S. Raoux, and R. S. Shenoy, Phase change memory technology, *J. Vac. Sci. Technol. B* **28**, 223 (2010).
- [18] G. W. Burr, R. M. Shelby, A. Sebastian, S. Kim, S. Kim, S. Sidler, K. Virwani, M. Ishii, P. Narayanan, A. Fumarola, L. L. Sanches, I. Boybat, M. Le Gallo, K. Moon, J. Woo, H. Hwang, and Y. Leblebici, Neuromorphic computing using non-volatile memory, *Adv. Phys.: X* **2**, 89 (2017).
- [19] M. Le Gallo and A. Sebastian, An overview of phase-change memory device physics, *J. Phys. D: Appl. Phys.* **53**, 213002 (2020).
- [20] A. Sawa, Resistive switching in transition metal oxides, *Mater. Today* **11**, 28 (2008).
- [21] H. Akinaga and H. Shima, Resistive random access memory (ReRAM) based on metal oxides, *Proc. IEEE* **98**, 2237 (2010).
- [22] E. Carlos, R. Branquinho, R. Martins, A. Kiazadeh, and E. Fortunato, Recent progress in solution-based metal oxide resistive switching devices, *Adv. Mater.* **33**, 2004328 (2021).
- [23] D. Suzuki, M. Natsui, H. Ohno, and T. Hanyu, in *Extended Abstracts of the 2010 International Conference on Solid State Devices and Materials* (2010).
- [24] E. Eken, L. Song, I. Bayram, C. Xu, W. Wen, Y. Xie, and Y. Chen, in *Proceedings of the 53rd Annual Design Automation Conference* (2016), p. 1.
- [25] R. De Rose, M. Lanuzza, F. Crupi, G. Siracusano, R. Tomasello, G. Finocchio, M. Carpentieri, and M. Alioto, A variation-aware timing modeling approach for write operation in hybrid CMOS/STT-MTJ circuits, *IEEE Trans. Circuits Syst. I: Regular Papers* **65**, 1086 (2018).
- [26] F. Razi, M. H. Moaiyeri, R. Rajaei, and S. Mohammadi, A variation-aware ternary spin-hall assisted STT-RAM based on hybrid MTJ/GAA-CNTFET logic, *IEEE Trans. Nanotechnol.* **18**, 598 (2019).
- [27] M. Weisheit, S. Fahler, A. Marty, Y. Souche, C. Poinignon, and D. Givord, Electric field-induced modification of magnetism in thin-film ferromagnets, *Science* **315**, 349 (2007).
- [28] T. Maruyama, Y. Shiota, T. Nozaki, K. Ohta, N. Toda, M. Mizuguchi, A. A. Tulapurkar, T. Shinjo, M. Shiraishi, S. Mizukami, Y. Ando, and Y. Suzuki, Large voltage-induced magnetic anisotropy change in a few atomic layers of iron, *Nat. Nanotechnol.* **4**, 158 (2009).
- [29] T. Nozaki, Y. Shiota, M. Shiraishi, T. Shinjo, and Y. Suzuki, Voltage-induced perpendicular magnetic anisotropy change in magnetic tunnel junctions, *Appl. Phys. Lett.* **96**, 022506 (2010).
- [30] Y. Shiota, T. Nozaki, F. Bonell, S. Murakami, T. Shinjo, and Y. Suzuki, Induction of coherent magnetization switching in a few atomic layers of FeCo using voltage pulses, *Nat. Mater.* **11**, 39 (2012).
- [31] S. Miwa, M. Suzuki, M. Tsujikawa, K. Matsuda, T. Nozaki, K. Tanaka, T. Tsukahara, K. Nawaoka, M. Goto, Y. Kotani, T. Ohkubo, F. Bonell, E. Tamura, K. Hono, T. Nakamura, M. Shirai, S. Yuasa, and Y. Suzuki, Voltage controlled interfacial magnetism through platinum orbits, *Nat. Commun.* **8**, 15848 (2017).
- [32] J. J. Nowak, R. P. Robertazzi, J. Z. Sun, G. Hu, D. W. Abraham, P. L. Trouilloud, S. Brown, M. C. Gaidis, E. J. O'Sullivan, W. J. Gallagher, and D. C. Worledge, Demonstration of ultralow bit error rates for spin-torque magnetic random-access memory with perpendicular magnetic anisotropy, *IEEE Magn. Lett.* **2**, 3000204 (2011).
- [33] Y. Shiota, T. Nozaki, S. Tamaru, K. Yakushiji, H. Kubota, A. Fukushima, S. Yuasa, and Y. Suzuki, Evaluation of write error rate for voltage-driven dynamic magnetization switching in magnetic tunnel junctions with perpendicular magnetization, *Appl. Phys. Express* **9**, 013001 (2016).
- [34] Y. Shiota, T. Nozaki, S. Tamaru, K. Yakushiji, H. Kubota, A. Fukushima, S. Yuasa, and Y. Suzuki, Reduction in write error rate of voltage-driven dynamic magnetization switching by improving thermal stability factor, *Appl. Phys. Lett.* **111**, 022408 (2017).
- [35] T. Ikeura, T. Nozaki, Y. Shiota, T. Yamamoto, H. Imamura, H. Kubota, A. Fukushima, Y. Suzuki, and S. Yuasa, Reduction in the write error rate of voltage-induced dynamic magnetization switching using the reverse bias method, *Jpn. J. Appl. Phys.* **57**, 040311 (2018).
- [36] T. Yamamoto, T. Nozaki, H. Imamura, Y. Shiota, T. Ikeura, S. Tamaru, K. Yakushiji, H. Kubota, A. Fukushima, Y.

- Suzuki, and S. Yuasa, Write-Error Reduction of Voltage-Torque-Driven Magnetization Switching by a Controlled Voltage Pulse, *Phys. Rev. Appl.* **11**, 014013 (2019).
- [37] T. Yamamoto, T. Nozaki, H. Imamura, Y. Shiota, S. Tamaru, K. Yakushiji, H. Kubota, A. Fukushima, Y. Suzuki, and S. Yuasa, Improvement of write error rate in voltage-driven magnetization switching, *J. Phys. D: Appl. Phys.* **52**, 164001 (2019).
- [38] R. Matsumoto and H. Imamura, Methods for reducing write error rate in voltage-induced switching having prolonged tolerance of voltage-pulse duration, *AIP Adv.* **9**, 125123 (2019).
- [39] T. Yamamoto, T. Nozaki, H. Imamura, S. Tamaru, K. Yakushiji, H. Kubota, A. Fukushima, Y. Suzuki, and S. Yuasa, Voltage-Driven Magnetization Switching Using Inverse-Bias Schemes, *Phys. Rev. Appl.* **13**, 014045 (2020).
- [40] W. F. Brown, Thermal fluctuations of a single-domain particle, *Phys. Rev.* **130**, 1677 (1963).
- [41] H. B. Callen and T. A. Welton, Irreversibility and generalized noise, *Phys. Rev.* **83**, 34 (1951).
- [42] H. B. Callen and R. F. Greene, On a theorem of irreversible thermodynamics, *Physical Review* **86**, 702 (1952).
- [43] H. B. Callen, M. L. Barasch, and J. L. Jackson, Statistical mechanics of irreversibility, *Phys. Rev.* **88**, 1382 (1952).
- [44] R. F. Greene and H. B. Callen, On a theorem of irreversible thermodynamics. II, *Phys. Rev.* **88**, 1387 (1952).