Journal of
Low Power Electronics and Applications

MDPI

Article

Phase Change Memory Drift Compensation in Spiking Neural Networks Using a Non-Linear Current Scaling Strategy

Joao Henrique Quintino Palhares [1,2,3,4], Nikhil Garg [3,4], Yann Beilliard [3,4], Lorena Anghel [2], Fabien Alibart [3,4], Dominique Drouin [3,4,*] and Philippe Galy [1,*]

1 STMicroelectronics, 850 rue Jean Monnet, 38920 Crolles, France; quij2205@usherbrooke.ca
2 CEA, CNRS, Grenoble INP, SPINTEC, Université Grenoble Alpes, 38000 Grenoble, France; lorena.anghel@grenoble-inp.fr
3 Institut Interdisciplinaire d'Innovation Technologique (3IT), Université de Sherbrooke, 3000 Bd de l'Université, Sherbrooke, QC J1K 0A5, Canada; nikhil.garg@usherbrooke.ca (N.G.); yann.beilliard@usherbrooke.ca (Y.B.); fabien.alibart@usherbrooke.ca (F.A.)
4 Laboratoire Nanotechnologies Nanosystèmes (LN2), CNRS IRL 3463, Université de Sherbrooke, 3000 Bd de l'Université Université, Sherbrooke, QC J1K 0A5, Canada
* Correspondence: dominique.drouin@usherbrooke.ca (D.D.); philippe.galy@st.com (P.G.)

Abstract: The non-ideality aspects of phase change memory (PCM) such as drift and resistance variability can pose significant obstacles in neuromorphic hardware implementations. A unique drift and variability compensation strategy is demonstrated and implemented in an FD-SOI SNN hardware unit composed of embedded phase change memories (ePCMs), current attenuators, and spiking neurons. The effect of drift and variability compensation on inference accuracy is tested on the MNIST dataset to show that our drift and variability mitigation strategy is effective in sustaining its accuracy over time. The variability is reduced by up to 5% while the drift coefficient is reduced by up to 57.8%. The drift is compensated and the SNN classification accuracy is sustained for up to 2 years with intrinsic control-free hardware that tracks the ePCM current over time and consumes less than 30 µW. The results are based on ePCM chip experimental data and pos-layout simulation of a test chip comprising the proposed circuit solution.

Keywords: PCM; PCM drift; drift mitigation; spiking neural networks; emerging non-volatile memory; 28 nm FD-SOI technology

Citation: Palhares, J.H.Q.; Garg, N.; Beilliard, Y.; Anghel, L.; Alibart, F.; Drouin, D.; Galy, P. Phase Change Memory Drift Compensation in Spiking Neural Networks Using a Non-Linear Current Scaling Strategy. *J. Low Power Electron. Appl.* **2024**, *14*, 50. https://doi.org/10.3390/jlpea14040050

Academic Editor: Davide Bertozzi

Received: 19 September 2024
Revised: 15 October 2024
Accepted: 19 October 2024
Published: 22 October 2024

1. Introduction

Implementing spiking neural networks (SNNs) with phase change memory (PCM) can potentially bypass the limitations of traditional computing and address energy concerns driven by growing computational demand. PCM's compatibility with CMOS, non-volatility, and programmability make it a promising candidate for dense, energy-efficient, non-von Neumann computing [1]. However, resistance drift and variability present significant challenges, particularly at room temperature, limiting PCM's potential in these applications [2]. Resistance drift changes the resistance in PCM cells over time, degrading SNN performance. While studies have shown that drift naturally diminishes at cryogenic temperatures (such as 77 K and 12 K), it remains a critical issue at room temperature, impacting the system's long-term accuracy [3]. Our previous work demonstrated that at cryogenic temperatures, the drift coefficient of PCM cells is significantly reduced, such that no hardware compensation is required to maintain system performance. However, this reduction does not extend to room temperature, where drift persists and leads to a performance degradation of up to 10–12% in classification tasks like MNIST. This makes the development of a solution to mitigate drift at room temperature crucial for practical neuromorphic systems.

Several methods have been proposed to address PCM drift. Hardware-based solutions, such as projected PCM architecture [4], introduce extra current to compensate for

drift, while 2-PCM Synapse employs binary PCM cells programmed through stochastic spike-timing-dependent plasticity (STDP) [5]. These approaches, while effective, often necessitate significant hardware redesigns or specific programming techniques. On the other hand, software-based methods for drift correction, commonly used in SNNs and artificial neural networks (ANNs), apply global gain factors to adjust for drift using average drift coefficients, but they typically require periodic recalibration or supervised retraining [6,7].

In contrast, we propose a novel non-linear (NL) current scaling strategy that provides real-time, continuous compensation for resistance drift, without the need for training, memory refresh, or PCM cell redesign. This strategy dynamically adjusts synaptic current in a non-linear manner, ensuring real-time adaptability and making it particularly suitable for neuromorphic hardware.

SNN hardware commonly utilizes one-transistor, one-resistor (1T1R) architecture, where each synaptic element consists of a transistor that controls access to a phase change memory (PCM) cell. This architecture enables precise control over the synaptic current by gating the PCM resistance, ensuring reliable read/write operations [8,9]. The 1T1R structure also facilitates scalability by allowing for dense arrays of synaptic connections. Current attenuators are integrated into this architecture to enhance neuron fan-in [10,11], which refers to the ability to handle multiple synaptic inputs per neuron, thereby supporting larger-scale neuromorphic systems.

Our proposed drift mitigation solution leverages the 1T1R structure and integrates current attenuators with analog neurons in a 28 nm FD-SOI-based SNN. This approach dynamically tracks resistance changes in the PCM cells and compensates for drift by non-linearly scaling down the synaptic current. By adjusting the current scaling factor in real-time based on the synaptic input, this method ensures continuous drift compensation without the need for training, memory refresh, or redesign of the PCM cells, as illustrated in Figure 1a.

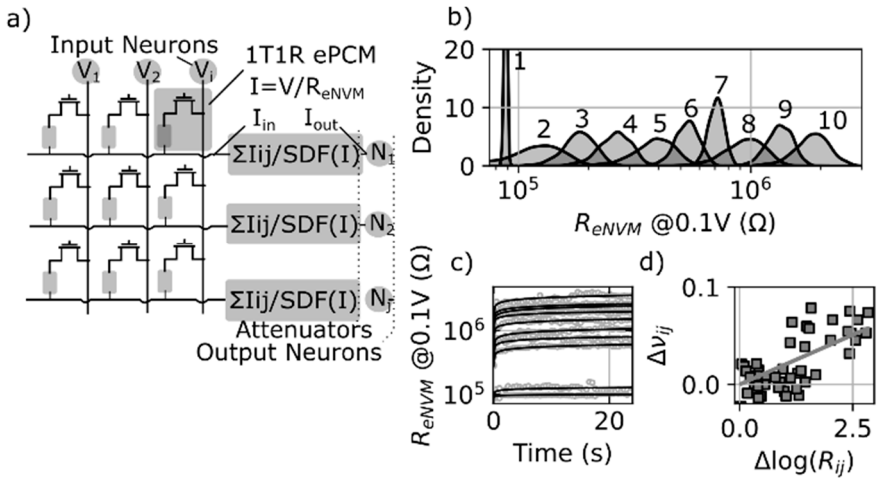

Figure 1. (**a**) Simplified illustration of the proposed SNN architecture with one-transistor and one-resistor (1T1R) ePCMs synaptic arrays, current attenuators, and neurons. (**b**) The ePCM is analog programmed to achieve up to 10 resistance states. (**c**) The ePCM resistance relaxation and drift coefficient are extracted and modeled over 25 s of retention test using Equation (1). (**d**) The corresponding drift versus resistance dependence is estimated by Equation (2) to draw meaningful conclusions about the behavior of the device, presenting a k coefficient of 0.02.

To evaluate our strategy, we combined experimental PCM data with post-layout SPICE simulations of a fully connected current attenuator and neuron. The simulations predict drift effects on SNN inference over a two-year period. A system-level MNIST classification task simulation confirms the effectiveness of our drift compensation, reducing drift by $50.58 \pm 6.10\%$ and variability by 5%, maintaining SNN accuracy for up to two years. Without this solution, performance degrades by 12.4%. This drift-resilient

approach provides significant advancements for neuromorphic computing, enabling robust real-world applications.

2. PCM Characterization Results

Drift refers to the structural relaxation of the amorphous phase in PCMs after programming [12]. Embedded Ge-rich GST-based PCM (ePCM) supplied by STMicroelectronics [13] was characterized and programmed using a Keysight B1500A semiconductor analyzer with a 200 Msample/s waveform generator module (WGFMU) to evaluate, quantify, and model the drift aspects.

2.1. PCM Multilevel Programming

A fully analog programming strategy was implemented to perform multilevel switching, consisting of 75 ns SET and RESET pulses applied at the top electrode. SET pulses were kept constant at 1.5 V, while RESET pulses increased in amplitude from 1.8 to 2.7 V. A DC voltage was applied to the selector's gate for varying current compliances between 0.8 and 1.2 V. The full sequence is repeated ten times for each compliance, resulting in a total of 1000 programming pulses. Resistance levels were read using 0.1 V pulses before and after programming. This programming approach yielded ten distinct resistance states ranging from 75 kΩ to 2 MΩ, as shown in Figure 1b.

2.2. Drift Quantification and Modeling

Post-programming, the resistance shift evolution is evaluated using a sequence of 0.1V over 25 s, sampled every 500 ms. The resistance progression over constant temperature is then fitted using the following equation [12]:

$$R_t = R_{t0}\left(\frac{t}{t_0}\right)^{v}, \quad (1)$$

R_t is the resistance at time t, R_{t0} the initial resistance, and v the drift coefficient. The drift response is shown in Figure 1c. The drift coefficients range from 0.011 to 0.086 for a low resistance (LRS) of 80 kΩ and a high resistance (HRS) of 1.5 MΩ. To predict and evaluate the effect of drift on the ePCM resistance distribution over time, the following model is used [14]:

$$v_{R_2} - v_{R_1} = k\left(\frac{R_2}{R_1}\right), \quad (2)$$

The k coefficient is 0.02 (see Figure 1d). This hybrid approach, combining experimental and modeling techniques, is crucial for drawing meaningful conclusions about the behavior of the ePCM at any given time regardless the resistance state. It was used to forecast the equivalent synaptic current I_{in} conveyed into the SNN inference hardware over time.

The drift quantification and modeling approach presented in this section is based on a hybrid experimental–modeling method that utilizes data from a PCM chip demonstrator provided by STMicroelectronics [3,13]. While long-term aging tests have not yet been conducted, the model incorporates well-established parameters from the state-of-the-art PCM literature [15–18]. This combination of experimental data and modeling provides a reliable foundation for predicting resistance shifts and their impact on synaptic current in SNN inference hardware, as previously demonstrated in our recent publication [3].

3. SNN Hardware

In spiking neural networks (SNNs), computation is performed by converting voltage spikes into current through ePCM synaptic resistances. An approach to augmenting analog SNN hardware involves scaling down the synaptic current integrated by the post-synaptic neuron [10,11,19,20]. This study utilizes current attenuators to increase the fan-in of the system and mitigate the synaptic current change over time caused by drift. The hardware used for simulating and evaluating this approach is based on a 28 nm FD-SOI-based current

attenuator connected to a low-power analog leaky integrate-and-fire (LIF) neuron [10]. It was designed and simulated using thick oxide high-k/metal gate, regular Vt 1.8 V-grade transistors. The inference hardware is virtually connected to the ePCM array as a lookup table to simulate and calculate the synaptic currents I_{in} and its downscaled version I_{out}. The key component in the present analysis is the current attenuator, and the output neuron serves as its load in the post-layout simulations (see Figure 2a). The system uses a low-dropout voltage regulator (LDO) to maintain a voltage reference (V_{ref}) in the common source node of the ePCM array, ensuring a drop voltage Vread in the ePCM memory of 0.1 V. Vread is calculated as V_i-V_{ref}, where V_i represents the voltage amplitude of the input neuron spike. The current I_{syn} is then duplicated by a current mirror and directed towards the current attenuator. The equivalent current Iin (or the ePCM synaptic current I_{syn}) is converted to voltage by MOS resistor N_9 (R_{N9}), which is then fed into the transconductance operational amplifier (TOA) at its differential input. The TOA generates a current proportional to the voltage difference, constrained by a 25 nA I_{bias}. The resulting Iout is a downscaled version of Iin, as defined by Equations (3) and (4). The ratio of Iin to Iout is known as the current scaling down factor (SDF) (5):

$$I_{in} = I_{syn} = \frac{V_i}{R_{eNVM}}, \tag{3}$$

$$I_{out} = \frac{I_{bias} k_n}{2U_t} \left(\frac{I_{in}.R_{N9}}{2} \right), \tag{4}$$

$$SDF = \frac{I_{in}}{I_{out}}, \tag{5}$$

where K_n is the subthreshold slope and U_t the thermal voltage. In SNNs, neurons receive input spikes sparsely from multiple pre-synaptic neurons [21]. Therefore, it is reasonable to operate the current attenuator with an I_{in} margin limited to the readout range of one average ePCM, excluding the range beyond the limits of Isyn, as indicated in Figure 2b by the gray dashed lines. Consistency in SDF within the crossbar readout range ensures accurate network responses. The stability of SDF within this physical interpretation range relies on the common mode rejection ratio of the current attenuator, influenced by the transistor saturation generating I_{bias} [22]. Failure to meet the transistor saturation condition for Ibias can lead to a strong dependency of Iout on the common mode voltage, the voltage drop of R9, and Iin. Adjusting the values of VsN10 and VsN11 (Vs) is necessary to achieve this condition (refer to Figure 2b). The scaling methods, SDF-NL (1) and SDF-NL (2), stand for different levels of non-linear (NL) current scaling down factor (SDF) in relation to the input current (I_{in}), each tuned to exhibit varying degrees of non-linearity in the current attenuation process, controlled by the vs. terminal of the current attenuator. In contrast, the constant SDF represents a fixed scaling factor, effectively neglecting the drift compensation mechanism, as though the solution were not applied. These strategies are illustrated in Figure 2b, with SDF-NL (1) shown in blue and SDF-NL (2) in red.

Regarding area overhead, the respective area budgets of the current attenuator and neuron circuit blocks are 774 μm^2 and 466 μm^2. Note that the neuron fan-in, which is crucial for large-scale networks, is proportional to the membrane capacitance and inversely proportional to the synaptic current [10,11,23]. The current attenuator enables an at least 4000$\times$ increase in neuron fan-in by reducing the input current (i.e., for an SDF of 4000), which is far more area-efficient compared to scaling the membrane capacitance, as achieving the same result through capacitance increase would require a 4000-fold larger membrane capacitor area (equivalent to a 118-fold increase in the combined area of the neuron and current attenuator).

In terms of power consumption, the error amplifier, current attenuator, and neuron consume an average dynamic power of 25 pW, 3.66 μW, and 25.9 μW, respectively. This corresponds to an input current I_{in} of 1.33 μA and a neuron spike rate of 13.4 kHz. The

dynamic power across the entire resistance range of the PCM, along with the corresponding neuron spike rate, is shown in Figure 2c. The static power consumption is 0.11 µW.

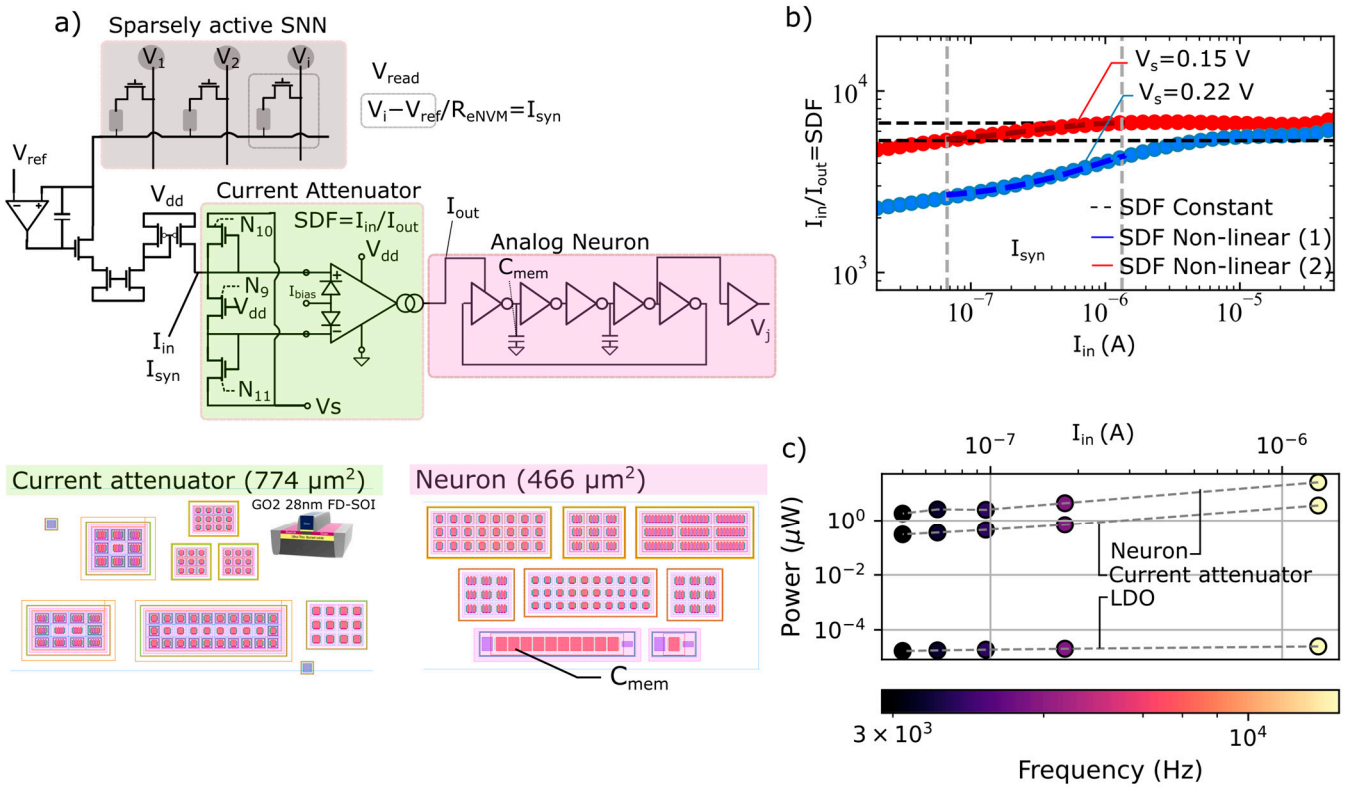

Figure 2. (**a**) Simplified schematics of the SNN circuit for scaling down ePCM reading current I_{in}. The circuit includes a 28 nm FD-SOI current attenuator connected to a leaky integrate-and-fire (LIF) neuron (serving as a load in the post-layout simulation). The respective area budgets of the current attenuator and neuron are 774 µm^2 and 466 µm^2. (**b**) The current scaling down factor (SDF) versus I_{in}. SDF(I_{in}) is modeled using a one-phase exponential decay function (7–8). SDF non-linear (1) and (2) are achieved with vs. set at 0.15 (red) and 0.22 V (blue), respectively, while the constant SDF is represented by a horizontal line. (**c**) The dynamic power consumption of the neuron, current attenuator, and voltage regulator error amplifier (LDO) versus I_{in}, along with the corresponding neuron spike rate.

3.1. Drift Compensation and the Effect of Non-Linear Current Scaling

The current attenuator adjusts the output current based on the input reading current Iin, especially when the SDF is influenced by the common mode voltage. As I_{syn} (or equivalent I_{in}) decreases over time following power law Equation (1), the strategy to counteract drift involves adapting the SDF to compensate for the decline in Iin due to drift. This adjustment results in an SDF that dynamically tracks the input current, ensuring responsiveness to the input reading current through an exponential decay function. Essentially, this mechanism acts as a current bias compensation that reduces the SDF proportionally to the reading current Iin, maintaining appropriate compensation for I_{out}. The current scaling down factor, denoted as SDF(Iin) and dependent on Iin, is modeled using SPICE simulations with a one-phase exponential decay function:

$$\mathrm{SDF}(\mathrm{I_{in}}) = \mathrm{SDF_{max}} - \mathrm{SDF_{min}}.\mathrm{e}^{-(\lambda(\mathrm{I_{in}}))}, \quad (6)$$

SDF represents the current scaling down factor, where SDF_{max} denotes the upper bound plateau corresponding to the maximum I_{in} and SDF_{min} represents the lower bound plateau corresponding to the minimum I_{in}. λ signifies the decay rate influenced by I_{in} and

R_{eNVM}, the ePCM resistance state. The fitting process is conducted based on post-layout circuit SPICE simulation outcomes, with the ePCM readout range I_{syn} serving as I_{in}, as indicated by dashed lines in Figure 2b. Finally, combining Equations (1)–(6):

$$I_{out}(t) = \frac{I_{in}}{SDF(I_{in})} = \frac{\frac{V_{read}}{R_{eNVM}(t)}}{SDF_{max} - SDF_{min}.e^{-(\lambda(\frac{V_{read}}{R_{eNVM}(t)}))}}, \quad (7)$$

$$I_{out}(t) = \frac{\frac{V_{read}}{R_{t0}\left(\frac{t}{t_0}\right)^{k.\log(\frac{R_{t0}}{R_{R1}})+v_{R1}}}}{SDF_{max} - SDF_{min}.e^{-(\lambda(\frac{V_{read}}{R_{t0}(\frac{t}{t_0})^{k.\log(\frac{R_{t0}}{R_{R1}})+v_{R1}}}))}}, \quad (8)$$

With V_{read} set at 0.1 V, the SDF parameter essentially embodies a decay rate dependent on R_{eNVM}, denoted as $\lambda(R_{eNVM})$, which inherently adapts to the reading current. Consequently, this adjustment extends to each resistance state and its associated drift coefficient. To benchmark and compare different tuning conditions, the net output current is normalized using the following formula:

$$X_{Normalized} = \frac{X - X_{min}}{X_{max} - X_{min}}, \quad (9)$$

X_{min} is the minimum I_{out} and X_{max} the maximum I_{out} It is important to note that the normalized net output current is equivalent to conductance, as the current is simply the conductance multiplied by a constant reading voltage of 0.1 V. To evaluate the effect of the SDF tunning, the calculation is carried out with a constant SDF (black dashed line—SDF_{max}), and a variable SDF (I_{in}) (red and blue curves), as shown in Figure 2. The calculation is performed at 6 s and progresses logarithmically up to approximately two years, reaching 6×10^7 s. The I_{out} evolution overtime for each of the 10 ePCM states distributions are shown in Figure 3a–d at different times: 1×10^{-3}, 6×10^1, 6×10^4, and 6×10^6 s. As shown in Figure 3, while drift tends to shift the normalized current distributions away from their initial state over time, the application of the non-linear SDF counteracts it.

The more non-linear the SDF is with respect to I_{in}, the closer the normalized current distribution remains to its pre-drift state (as shown by the gray dashed lines). Non-linear SDF (2) demonstrates higher efficiency compared to non-linear SDF (1), compensating for drift more effectively. Regardless of the degree of non-linearity, this method robustly counteracts drift, suggesting a reliable strategy that accommodates variability and mismatch. This underscores the effectiveness of the non-linear SDF in mitigating drift.

While this study focuses on PCM drift, this method has the potential to be compatible with other memory technologies that, although governed by different physical phenomena, exhibit various forms of resistance relaxation over time [23,24]. These different manifestations of resistance change, while rooted in distinct mechanisms, lead to equivalent forms of current degradation similar to drift in PCM [23,24]. Such technologies could potentially benefit from the continuous current adjustment mechanism provided by the non-linear scaling strategy, helping to mitigate performance degradation. Future research could also explore the applicability of the non-linear current scaling strategy to other forms of signal degradation, such as series resistance and crosstalk in larger systems [25]. Further studies will be necessary to validate these aspects and investigate the robustness of this approach in mitigating different sources of performance degradation across diverse memory technologies and crossbar structures as they scale. In the next section, the equivalent reduction in the drift coefficient is estimated.

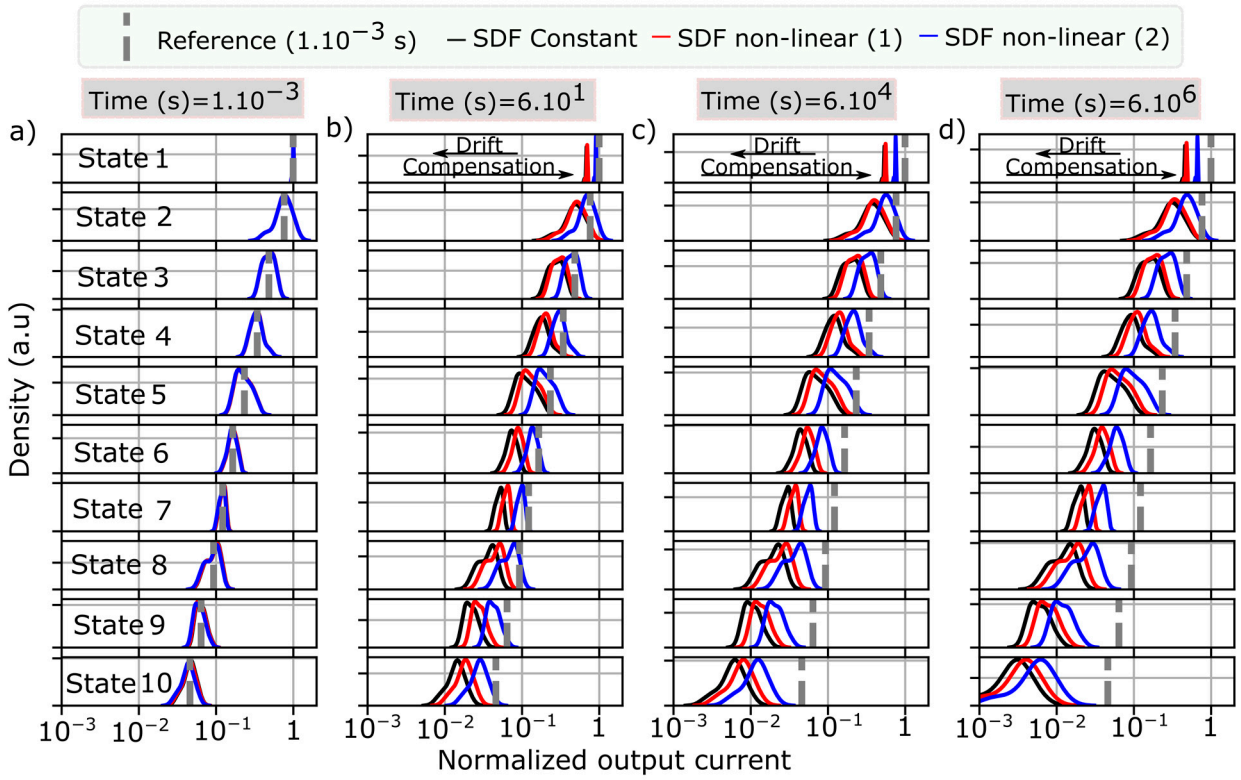

Figure 3. (**a**) Impact of non-linear SDF on ePCM state distribution and I_{out}. (**a**) Initial net output current distributions at 1 ms post-programming (average values marked by gray dashed line) and drifted counterparts at subsequent time intervals: (**b**) 6.10^1 s, (**c**) 6.10^4 s, and (**d**) 6.10^6 s. Drift shifts the distribution from the initial state, countered by non-linear SDF.

Variability and Drift Mitigation

The calculation is performed with different tunned SDFs to demonstrate varying degrees of drift compensation. To assess and quantify the reduction in the drift coefficient resulting from the non-linear SDF, a reverse calculation is conducted using Equation (8) to estimate the resistance evolution over time from Iout for each of the 10 average ePCM resistance states, considering a constant SDF across all scenarios

The derived equivalent resistance over time is then modeled using Equation (1) to estimate the drift coefficient, as depicted in Figure 4a. The drift coefficients corresponding to the equivalent resistances for each SDF (linear and non-linear 1, 2) are shown in Figure 4b. An additional aspect of the non-linear SDF is its tendency to increase as Iin increases, leading to a subtle distortion in the shape of the normalized net current by shifting the current distribution from its minimum towards its maximum value. This distortion impacts the average and standard deviation of the normalized current for each ePCM state over time, as shown in Figure 5. In Figure 4c, it is shown that SDF non-linear 1 results in an average reduction in the drift coefficient of 13.94 $\pm$ 3.79%, while SDF non-linear 2 reduces the drift coefficient by 50.58 $\pm$ 6.10% compared to SDF linear. To evaluate the effect of it in the variability of the state's distribution, the resistance and corresponding reading current variability I_{in} were estimated by the coefficient of variability (C.V):

$$C.V = \frac{\sigma}{\mu} \tag{10}$$

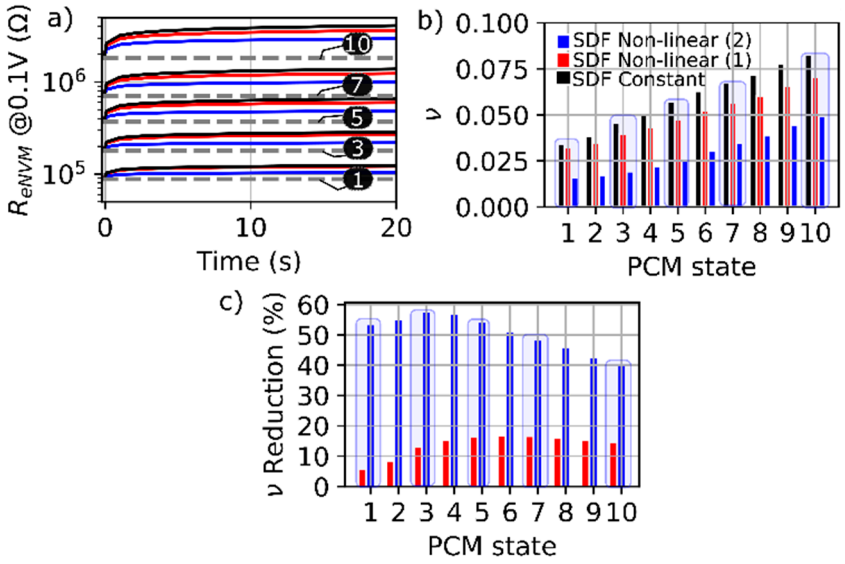

Figure 4. (**a**) Forecasted evolution of equivalent resistance for states 1, 3, 5, 7, and 10 predicted through I_{out} for both SDF constant and non-linear (1, 2). (**b**) Drift coefficients calculated using Equation (1). (**c**) Reduction in drift coefficient achieved by implementing non-linear SDF (1, 2) compared to SDF constant, relative to the original drift before any compensation. The results show a maximum drift reduction of 57.8% for SDF non-linear (2) and 16.9% for SDF non-linear (1).

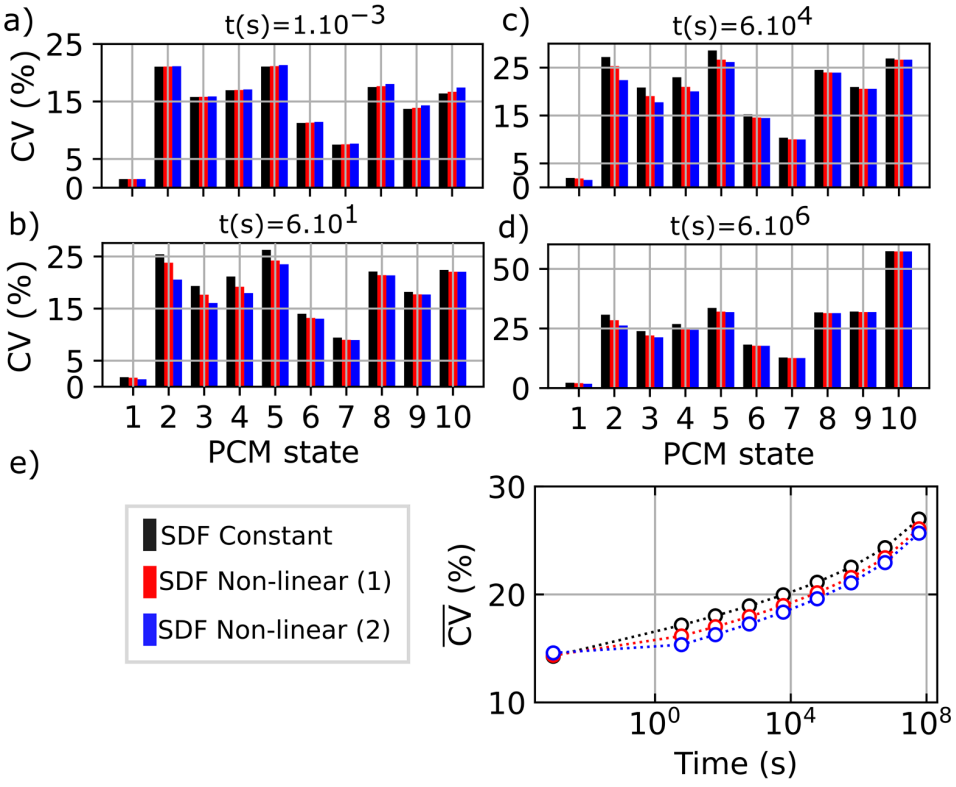

Figure 5. Impact of non-linear current scaling (SDF non-linear (1) in red, (2) in blue) on the coefficient of variation (CV) for ePCM normalized net current. The analysis includes initial state pre-drift at Forecasted evolution of equivalent resistance for states 1, 3, 5, 7, and 10 predicted through I_{out} for both SDF constant and non-linear (**a**) 1.10^{-3} s with constant SDF (black), followed by each time step in the analysis (**b**) 6.10^1, (**c**) 6.10^4, and (**d**) 6.10^6 s. (**e**) Average coefficient of variation of the 10 ePCM states over time post-non-linear current scaling. The average and maximum CV reduction is 1.6 ± 1.5% and 5%, respectively.

The C.V of the I_{out} distributions calculated for a constant SDF and non-linear SDF indicates that SDF non-linear 2 reduces variability by up to 5%. The average reduction across the ten ePCM states is 1.6 ± 1.5%, as shown in Figure 5d. This reduction is particularly notable in the more conductive states 1 to 5. Previous studies have highlighted the significant role of variability in neural network performance [21]. Therefore, the proposed hardware configuration offers both drift and variability reduction. The subsequent section evaluates the impact of this strategy in a fully connected SNN through system-level simulation.

4. System-Level Simulation

The impact of drift and variability compensation on inference accuracy is simulated using Python with an SNN implementation to evaluate the effect of non-linear SDF on the MNIST classification task. The input layer features 784 LIF neurons and 50 output neurons for 9 MNIST classes. Pixel intensity is encoded through frequency coding, aligning the spiking frequency with the pixel value.

The drift compensation strategy is applied during inference, independent of the training method. For instance, the neural network is trained using voltage-dependent synaptic plasticity (VDSP) in software [24]. The trained network is then transferred to ePCM synapses based on the nearest resistance distribution for inference on 10,000 grayscale images. Accuracy is computed with 10 different seeds for each time step and weight distribution. The initial ePCM weights, before drift, correspond to the transferred trained weights. The weight distribution relaxes according to drift after 1.10^{-3}, up to 6.10^{7} s and the inference accuracy is computed over time to validate our drift mitigation strategy.

As shown in Figure 6b, the baseline accuracy pre-weight transfer was 72.7%. The accuracies post-weight transfer were 71.54%, 71.55%, and 71.55% for SDF constant, non-linear 1, and non-linear 2, respectively. After two years, accuracies shifted to 59.15%, 66%, and 70%. Resistance drift led to a 12.4% accuracy reduction, mitigated by both non-linear SDFs for at least up to $6 \times 10_5$ s. SDF non-linear 2 sustained accuracy for up to 6×10^{7} s (approximately 2 years). The accuracy achieved in this study aligns with state-of-the-art results for similar neuromorphic architectures using PCMs and fully connected networks [26,27]. It is important to note that accuracy scales with the number of output neurons. For instance, 90% accuracy can be attained in a VDSP-based SNN with 500 output neurons [24]. However, 50 neurons are effectively sufficient to demonstrate the impact of drift over time with and without the real-time compensation strategy, while reducing computational costs.

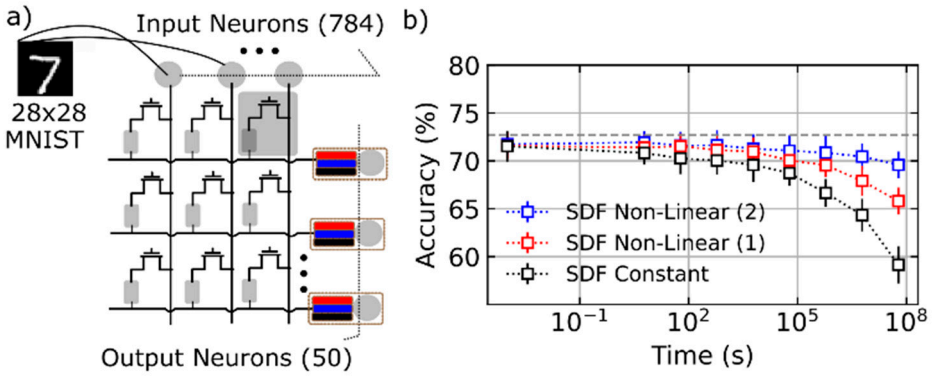

Figure 6. (**a**) SNN system-level simulation assessing the impact of non-linear SDF on the MNIST classification task, with 784 (28×28) input LIF neurons and 50 output neurons. (**b**) Average accuracy over time from 1.10^{-3} to 6.10^{7} s, based on 10 different seeds, for each ePCM net normalized output current distribution at constant SDF, SDF non-linear (1), and SDF non-linear (2). The baseline average accuracy pre-weight transfer is 72.7% for 50 output neurons. Initial accuracies are 71.54%, 71.55%, and 71.55% for SDF constant, non-linear 1, and non-linear 2, respectively. After 2 years, accuracies are 59.15%, 66%, and 70%, respectively.

This shows that our drift mitigation strategy is effective to compensate and counteract the accuracy drop caused by drift at room temperature. Please note that feedforward fully connected networks serve as the foundation for deep learning architectures and more complex models, such as convolutional (CNNs) and recurrent neural networks (RNNs). While the primary goal of this study is to analyze the effects of drift and the novel non-linear current scaling strategy on a simplified network, these findings may be generalized to other network topologies in future studies. To date, this is the first demonstration of a FD-SOI-based SNN-oriented fully analog solution that mitigates ePCM drift and variability to sustain accuracy for up to 2 years in a MNIST classification task at room temperature.

5. Conclusions

In summary, fully analog SNN hardware implementations using ePCM and FD-SOI technology require a current scaling block to enhance neuron fan-in and manage the neuron membrane capacitor more effectively. We introduce a novel strategy to compensate for ePCM drift at room temperature through current scaling. This drift mitigation technique, combined with current attenuation, is crucial for successfully integrating eNVM into low-power CMOS SNNs. Unlike traditional methods that rely on average ePCM drift behavior, training, memory refresh, or PCM cell redesign, our non-linear current scaling factor (SDF) continuously tracks the synaptic current in real time, eliminating the need for feedback control or additional circuitry. Our solution demonstrates robustness to system variability and mismatch, offering a flexible range of SDF profiles to accommodate different levels of drift compensation, as shown in our post-layout simulations for non-linear SDF (1) and (2). The proposed approach can reduce drift by 57.8%, maintaining accuracy for up to two years. Additionally, the total power consumption of the circuit, including the LDO error amplifier, current attenuator, and neuron, is kept under 30 μW. To date, no fully analog embedded SNN hardware solution has addressed both ePCM drift and resistance variability simultaneously. Our approach paves the way for ePCM/CMOS-based SNN implementations that can compensate for eNVM non-idealities, advancing the potential for neuromorphic applications.

6. Patents

The authors declare that parts of this work are included in a pending patent [Internal Reference: 23-GR1CO-0718].

Author Contributions: Conceptualization, J.H.Q.P. and P.G.; methodology, J.H.Q.P.; software, J.H.Q.P. and N.G.; validation, J.H.Q.P., N.G., P.G., D.D., F.A. and L.A.; formal analysis, J.H.Q.P., P.G., D.D. and F.A.; investigation, J.H.Q.P.; resources, P.G., D.D., L.A. and F.A.; data curation, J.H.Q.P. and N.G.; writing—original draft preparation, J.H.Q.P.; writing—review and editing, J.H.Q.P. and P.G.; visualization, J.H.Q.P., P.G., D.D., F.A., L.A., Y.B. and N.G.; supervision, P.G., D.D., F.A., L.A. and Y.B.; project administration, P.G., D.D., L.A. and F.A.; funding acquisition, P.G., D.D., L.A. and F.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the CIFRE-Conventions Industrielles pour la Formation par la Recherche N°2020/1230.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: The authors declare that parts of this work are included in a pending patent/Invention [Internal Reference: 23-GR1CO-0718]. This work was financially supported by STMicroelectronics SA and ANRT (CIFRE-Conventions Industrielles pour la Formation par la Recherche). We acknowledge the Laboratoire Nanotechnologies et Nanosystèmes (LN2)/International Research Laboratory and the SPINTEC IC design team.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Rajendran, B.; Sebastian, A.; Schmuker, M.; Srinivasa, N.; Eleftheriou, E. Low-Power Neuromorphic Hardware for Signal Processing Applications: A Review of Architectural and System-Level Design Approaches. *IEEE Signal Process Mag.* **2019**, *36*, 97–110. [CrossRef]
2. Burr, G.W.; Brightsky, M.J.; Sebastian, A.; Cheng, H.Y.; Wu, J.Y.; Kim, S.; Lam, C.H. Recent Progress in Phase-Change Memory Technology. *IEEE J. Emerg. Sel. Top. Circuits Syst.* **2016**, *6*, 146–162. [CrossRef]
3. Palhares, J.H.Q.; Garg, N.; Mouny, P.-A.; Beilliard, Y.; Sandrini, J.; Arnaud, F.; Anghel, L.; Alibart, F.; Drouin, D.; Galy, P. 28 nm FDSOI embedded PCM exhibiting near zero drift at 12 K for cryogenic SNNs. *NPJ Unconv. Comput.* **2024**, *1*, 8. [CrossRef]
4. Koelmans, W.W.; Sebastian, A.; Jonnalagadda, V.P.; Krebs, D.; Dellmann, L.; Eleftheriou, E. Projected phase-change memory devices. *Nat. Commun.* **2015**, *6*, 8181. [CrossRef] [PubMed]
5. Suri, M.; Garbin, D.; Bichler, O.; Querlioz, D.; Vuillaume, D.; Gamrat, C.; DeSalvo, B. Impact of PCM resistance-drift in neuromorphic systems and drift-mitigation strategy. In Proceedings of the 2013 IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH) 140–145, Brooklyn, NY, USA, 15–17 July 2013; pp. 140–145. [CrossRef]
6. Nandakumar, S.R.; Boybat, I.; Le Gallo, M.; Eleftheriou, E.; Sebastian, A.; Rajendran, B. Experimental Demonstration of Supervised Learning in Spiking Neural Networks with Phase-Change Memory Synapses. *Sci. Rep.* **2020**, *10*, 8080. [CrossRef] [PubMed]
7. Ambrogio, S.; Kumar, A.; Chen, A.; Burr, G.W.; Gallot, M.; Spoon, K.; Tsai, H.; Mackin, C.; Wesson, M.; Kariyappa, S.; et al. Reducing the Impact of Phase-Change Memory Conductance Drift on the Inference of large-scale Hardware Neural Networks. In Proceedings of the 2019 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 7–11 December 2019; pp. 6.1.1–6.1.4. [CrossRef]
8. Li, H.; Wang, S.; Zhang, X.; Wang, W.; Yang, R.; Sun, Z.; Yao, Y. Memristive Crossbar Arrays for Storage and Computing Applications. *Adv. Intell. Syst.* **2021**, *3*, 2100017. [CrossRef]
9. Xia, Q.; Yang, J.J. Memristive crossbar arrays for brain-inspired computing. *Nat. Mater.* **2019**, *18*, 309–323. [CrossRef] [PubMed]
10. Palhares, J.H.Q.; Beilliard, Y.; Sandrini, J.; Arnaud, F.; Garello, K.; Prenat, G.; Anghel, L.; Alibart, F.; Drouin, D.; Galy, P. A tunable and versatile 28 nm FD-SOI crossbar output circuit for low power analog SNN inference with eNVM synapses. *Solid. State Electron.* **2023**, *209*, 108779. [CrossRef]
11. Ahmadi-Farsani, J.; Ricci, S.; Hashemkhani, S.; Ielmini, D.; Linares-Barranco, B.; Serrano-Gotarredona, T. A CMOS–memristor hybrid system for implementing stochastic binary spike timing-dependent plasticity. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2022**, *380*, 20210018. [CrossRef] [PubMed]
12. Ielmini, D.; Lavizzari, S.; Sharma, D.; Lacaita, A.L. Physical interpretation, modeling and impact on phase change memory (PCM) reliability of resistance drift due to chalcogenide structural relaxation. In Proceedings of the 2007 IEEE International Electron Devices Meeting, Washington, DC, USA, 10–12 December 2007; pp. 939–942. [CrossRef]
13. Arnaud, F.; Zuliani, P.; Reynard, J.; Gandolfo, A.; Disegni, F.; Mattavelli, P.; Gomiero, E.; Samanni, G.; Jahan, C.; Berthelon, R.; et al. Truly Innovative 28 nm FDSOI Technology for Automotive Micro-Controller Applications embedding 16MB Phase Change Memory. In Proceedings of the 2018 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 1–5 December 2018; pp. 18.4.1–18.4.4. [CrossRef]
14. Braga, S.; Cabrini, A.; Torelli, G. Dependence of resistance drift on the amorphous cap size in phase change memory arrays. *Appl. Phys. Lett.* **2009**, *94*, 092112. [CrossRef]
15. Boniardi, M.; Ielmini, D.; Lavizzari, S.; Lacaita, A.L.; Redaelli, A.; Pirovano, A. Statistics of Resistance Drift Due to Structural Relaxation in Phase-Change Memory Arrays. *IEEE Trans. Electron. Devices* **2010**, *57*, 2690–2696. [CrossRef]
16. Ielmini, D.; Sharma, D.; Lavizzari, S.; Lacaita Andrea, L. Reliability Impact of Chalcogenide-Structure Relaxation in Phase-Change Memory (PCM) Cells—Part I: Experimental Study. *IEEE Trans. Electron. Devices* **2009**, *56*, 1070–1077. [CrossRef]
17. Boniardi, M.; Redaelli, A.; Pirovano, A.; Tortorelli, I.; Ielmini, D.; Pellizzer, F. A physics-based model of electrical conduction decrease with time in amorphous Ge2Sb2Te5. *J. Appl. Phys.* **2009**, *105*, 084506. [CrossRef]
18. Ielmini, D.; Sharma, D.; Lavizzari, S.; Lacaita, A.L. Physical mechanism and temperature acceleration of relaxation effects in phase-change memory cells. In Proceedings of the 2008 IEEE International Reliability Physics Symposium, Phoenix, AZ, USA, 27 April–1 May 2008; pp. 597–603. [CrossRef]
19. Ahmadi-Farsani, J.; Linares-Barranco, B.; Serrano-Gotarredona, T. A Current-Attenuator for Performing Read Operation in Memristor-Based Spiking Neural Networks. In Proceedings of the 2020 XXXV Conference on Design of Circuits and Integrated Systems (DCIS), Segovia, Spain, 18–20 November 2020; pp. 1–4. [CrossRef]
20. Mohan, C.; de la Rosa, J.M.; Vianello, E.; Perniola, L.; Reita, C.; Linares-Barranco, B.; Serrano-Gotarredona, T. A Current Attenuator for Efficient Memristive Crossbars Read-Out. In Proceedings of the 2019 IEEE International Symposium on Circuits and Systems (ISCAS), Sapporo, Japan, 26–29 May 2019; pp. 1–5. [CrossRef]
21. Moro, F.; Esmanhotto, E.; Hirtzlin, T.; Castellani, N.; Trabelsi, A.; Dalgaty, T.; Molas, G.; Andrieu, F.; Brivio, S.; Spiga, S.; et al. Hardware calibrated learning to compensate heterogeneity in analog RRAM-based Spiking Neural Networks. In Proceedings of the 2022 IEEE International Symposium on Circuits and Systems (ISCAS), Austin, TX, USA, 27 May–1 June 2022; pp. 380–383. [CrossRef]
22. Liu, S.-C.; Kramer, J.; Indiveri, G.; Delbrück, T.; Douglas, R. *Analog VLSI: Circuits and Principles*; MIT Press: Cambridge, MA, USA, 2002.

23. Aldana, S.; Zhang, H. Unravelling the Data Retention Mechanisms under Thermal Stress on 2D Memristors. *ACS Omega* **2023**, *8*, 27543–27552. [CrossRef] [PubMed]
24. Garg, N.; Balafrej, I.; Stewart, T.C.; Portal, J.-M.; Bocquet, M.; Querlioz, D.; Drouin, D.; Rouat, J.; Beilliard, Y.; Alibart, F. Voltage-dependent synaptic plasticity: Unsupervised probabilistic Hebbian plasticity rule based on neurons membrane potential. *Front. Neurosci.* **2022**, *16*, 983950. [CrossRef] [PubMed]
25. Shin, T.; Son, K.; Kim, S.; Cho, K.; Park, S.; Kim, S.; Park, G.; Sim, B.; Kim, J. Impact of On-Chip Interconnection in a Large-Scale Memristor Crossbar Array for Neural Network Accelerator and Neuromorphic Chip. In Proceedings of the 2019 IEEE 28th Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS), Montreal, QC, Canada, 6–9 October 2019; pp. 1–3. [CrossRef]
26. Boybat, I.; Le Gallo, M.; Nandakumar, S.R.; Moraitis, T.; Parnell, T.; Tuma, T.; Rajendran, B.; Leblebici, Y.; Sebastian, A.; Eleftheriou, E. Neuromorphic computing with multi-memristive synapses. *Nat. Commun.* **2018**, *9*, 2514. [CrossRef] [PubMed]
27. Nandakumar, S.R.; Le Gallo, M.; Boybat, I.; Rajendran, B.; Sebastian, A.; Eleftheriou, E. A phase-change memory model for neuromorphic computing. *J. Appl. Phys.* **2018**, *124*, 152135. [CrossRef]