

Double magnetic tunnel junction based $\Sigma\Delta\Sigma$ hardware neuron

Faiyaz E. Mullick¹, Rahul Sreekumar¹, Md Golam Morshed¹, Samiran Ganguly², Mircea Stan¹,
and Avik W. Ghosh¹

¹Dept. of ECE, University of Virginia, Charlottesville, VA, 22903, USA

²Dept. of ECE, Virginia Commonwealth University, Richmond, VA 23284, USA

Email: fm4fv@virginia.edu / Phone: (434)-326-8528

Introduction: Brain inspired spiking neurons have gained traction in edge machine learning (ML) applications as they encode information in temporal spikes improving energy efficiency [1]. However, replicating certain characteristics of spiking neurons, like their stochastic variations in firing threshold, is difficult in conventional silicon devices. Here we propose a novel double magnetic tunnel junction (MTJ) based spiking neuron which leverages the stochasticity of low barrier magnets (LBM) to emulate the firing threshold variability. The neural encoding schemes used in spiking neurons have recently started using schemes like Delta [2] and Sigma-delta modulation [3]. These facilitate the communication of features between neurons only when they surpass a certain threshold. As such, we use an encoding scheme termed as Sigma-Delta-Sigma (SDS) from our earlier work on temporal computation of information [4]. A unique feature of SDS encoding is its capability to *filter out noise-like* features from relevant information ensuring the noise-robustness of the network, a large improvement over previous LBM-based neurons.

Device and SDS Neuron Modeling: Our MTJ model is based upon the previous Binary Stochastic Neuron (BSN) [5]. To integrate the BSN as a stochastic threshold comparator within the SDS neuron, we must control the BSN's switching frequency to manage the dense stream of 'bits'. This ensures minimal degradation of the signal propagation within the SDS neuron's internal feedback loop. Utilizing SPICE-based LLG solvers [5], we discovered that incorporating a second MTJ in a BSN (Fig. 1(a)) results in four possible resistance combinations based on the alignment of low barrier magnets (LBM) with the fixed layer (FM): $R_P - R_P - R_P$ for HIGH voltage, $R_{AP} - R_P - R_{AP}$ for LOW voltage, and two intermediate states: $R_{AP} - R_P - R_P$ and $R_P - R_P - R_{AP}$. The presence of intermediate states prompts the NOT gate to adjust them, leading to a 'sparser' stream of 'bits' (Fig. 1(b)). Upon integration with the SDS modulation circuit (Fig. 1(c)), this exhibits typical $\Sigma\Delta$ action (Fig. 1(d)) with varying degrees of stochasticity.

To manipulate the stochastic behavior of our BSN-based SDS neuron (BSN-SDS), we change the Gilbert damping coefficient (α) of the LLG equations [5] which alters the switching frequency of the LBM. Hence, we change the α values in SPICE and analyze the resulting frequency signatures across a sweep of input current (I_{in}). As the duo MTJ BSN replaces the conventional comparator described in [4], the temporal characteristics of the BSN-SDS neuron align with the frequency and duty cycle relationships of the SD modulator circuit in Fig. 1(d). To implement the BSN-SDS neuron in a ML model requires converting the device behaviour into an analytical model. We base this on the BSN model [5] and the SD modulator model [2]. To ensure the accuracy of the analytical model, we calibrate it against our SPICE simulations in Fig. 2(a). We do this by measuring the hysteresis (δ_{hys}) of the BSN-SDS for different α and inputting that into (2) of fig. 2(b) to find f . This controls the internal delay within the SDS neuron which in turn alters the rate of change of $V_{IN}(t)$ to the BSN (1).

Results and Discussion: Most models are trained on clean data before deployment for inference or prediction tasks, however, the input test data might have noise which could reduce the performance of the model. Our encoding schema addresses this by enabling neurons to filter out noise components from incoming features, ensuring robust feature extraction even with low-quality raw data. The feedback mechanism with the SDS neuron incorporates a low-pass filter to identify noise-like features and removing them, allowing test on noisy data. To test out our model we use a liquid state machine (LSM) implementation of the echo state network in [6]. The network has a reservoir like in fig. 3(a) and we test out different reservoir sizes of 10, 20 and 50. For parameters for the analytical equations of fig. 2 (b) we used the ones in [4-6]. Our simulations, illustrated in Fig. 3 (b) - (d), demonstrate an average improvement of 1.2x with a reservoir size of 10 neurons, with robustness increasing as the reservoir size grows. The peak improvement of 1.56x is observed with a reservoir size of 50 neurons. A significant aspect of our encoding schema is its enhanced resilience to noise compared to the baseline, particularly notable for specific reservoir sizes. Our simulation results indicate that the percentage increase in Normalized Mean Squared Error (NMSE) across varying signal qualities is < 11.2%, roughly 2.1x more robust than the baseline BSN model (reservoir size = 10). This resilience diminishes as the reservoir size increases, indicating the model's capacity to learn and reconstruct features under adverse conditions. Specifically, we observe a 1.32x more robust performance with a reservoir size of 50 neurons. Thus, we show a viable double MTJ based spiking neuron and demonstrate its ability to work in very small models with very good resilience to noise.

[1] Rath et al., ACM Computing Surveys, vol. 55, p. 1-49, [4] Sreekumar et al., 28th ICECS., pp. 1-6, (2021). (2023).

[2] Yang et al., IEEE ISSCC, p. 388-389, (2016).

[5] Hassan et al., Phys. Rev. Applied, vol 15, 064046, (2021)

[3] Harbor et al., IEEE DASC, p. 1-6, (2023).

[6] Morshed et al., Front. Nanotechnol., vol. 5, (2023).

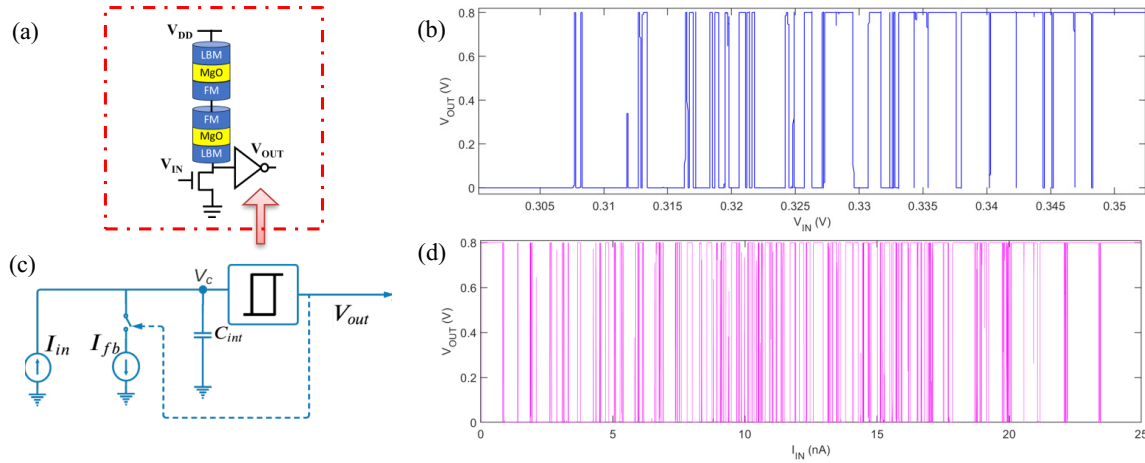


Fig.-1 Double MTJ based BSN (a) Circuit with double MTJ. P = parallel, AP = antiparallel. We have **two** scenarios when both free layers are **P** or **AP** with the fixed layer (FM) which lead to a logical HIGH or LOW respectively. Anything in between gets pushed to rails by NOT gate in (b) Output Characteristics. MTJ based Δ - Σ Modulator (c) Circuit (d) Transfer characteristics showing Δ - Σ action of wider pulses with increasing input current.

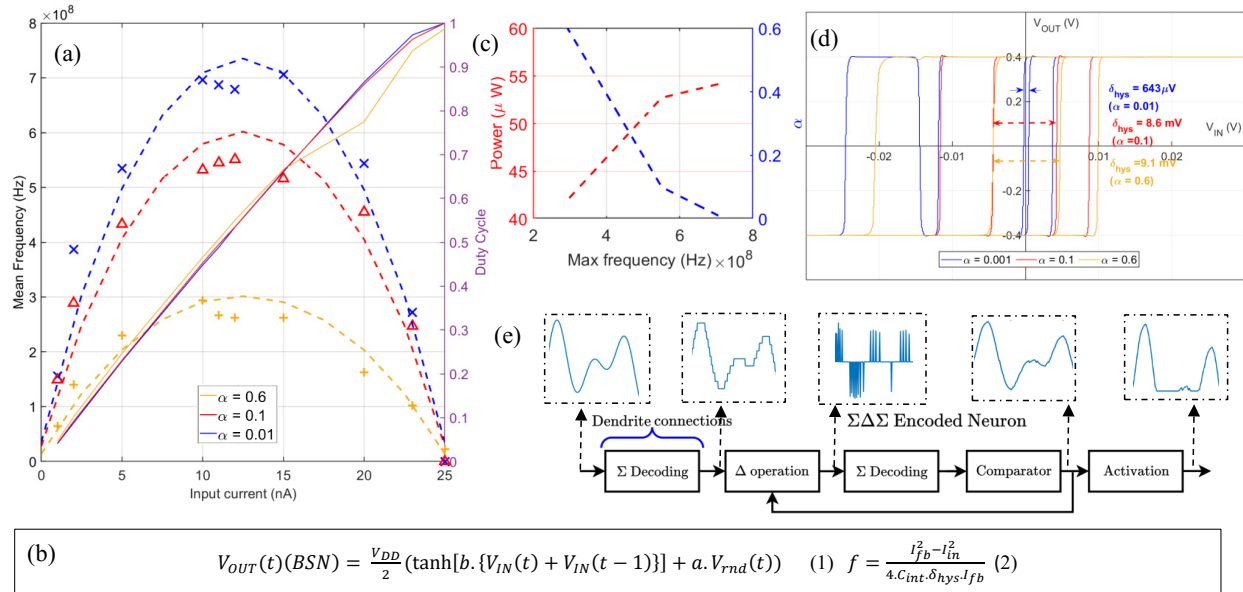


Fig.-2 (a) Temporal characteristics of BSN SDS neuron for different α (scatter for SPICE and dotted for analytical) and solid lines for duty cycle. (b) analytical relationship between hysteresis and α (c) power and α trade-off for different max frequencies. (d) plot of hysteresis for different α . (e) SDS modeling with intermediate outputs

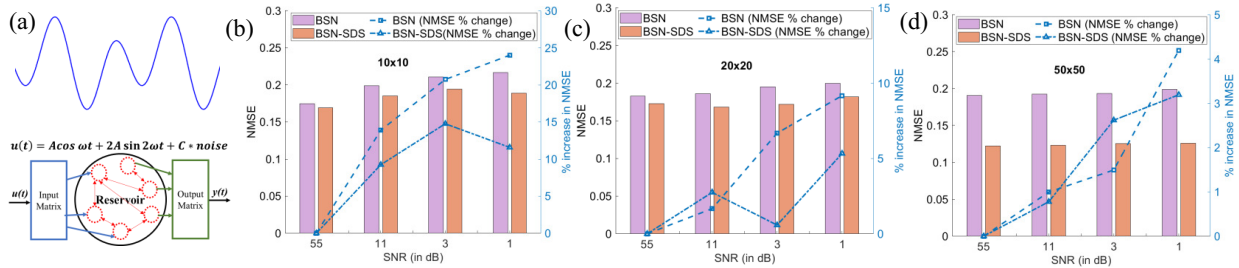


Fig.-3 (a) Input signal and reservoir structure for LSM and NMSE plots for reservoir sizes, we alter SNR by changing the ratio A/C (b) 10x10 (3) 20x20 and (d) 50x50 respectively.