A Low-Power High-Speed Spintronics-Based Neuromorphic Computing System Using Real Time Tracking Method

Hooman Farkhani, *Member, IEEE,* Mohammad Tohidi, Sadaf Farkhani, Jens Kargaard Madsen, and Farshad Moradi, *Senior Member, IEEE*

Abstract—In spintronic-based neuromorphic computing systems (NCS), the switching of magnetic moment in a magnetic tunnel junction (MTJ) is used to mimic neuron firing. However, the stochastic switching behavior of the MTJ and process variations effect lead to a significant increase in stimulation time of such NCSs. Moreover, current NCSs need an extra phase to read the MTJ state after stimulation which is in contrast with real neuron functionality in human body. In this paper, the read circuit is replaced with a proposed real-time sensing (RTS) circuit. The RTS circuit tracks the MTJ state during stimulation phase. As soon as switching happens, the RTS circuit terminates the MTJ current and stimulates the post neuron. Hence, the RTS circuit not only improves the energy consumption and speed, but also makes the operation of NCS similar to real neuron functionality. The simulation results in 65-nm CMOS technology confirm that the energy consumption and speed of the proposed RTS-based NCS are improved at least by 40% and 2.22X compared with a typical NCS, respectively. Finally, utilizing the RTS-based NCS in image processing applications such as character recognition and edge detection can lead to 90.3% improvement in energy delay products compared with the typical NCS.

Index Terms—Neuromorphic computing system, MTJ, memristor, energy consumption, spintronic

I. INTRODUCTION

During the past decade, enormous efforts have been put to design a computing system to be trained and adapted to communicate with environment in a similar way as human brain does. Human brain approximately consumes a power consumption of 20W to perform more than 10^{16} operations, which is giving the brain a 12 orders of magnitude advantage in operation/s/W/cm^3 in comparison to the state-of-the-art supercomputers with a power consumption of 18nW/operation [1]. Therefore, for future exascale computing (10^{18}), a paradigm shift facilitating the computation in an extremely low power density mode is essential, especially with the current explosion of data; the best known solution so far is to mimic the brain computing, so called "neuromorphic computing". During the last decade, neuromorphic computing has emerged as a mean to mimic the architecture of biological networks to overcome the limitation of the word-at-a-time thinking of conventional computers by processing data massively in parallel [2]. The IBM's TrueNorth and the Google's DeepMind are examples of such brain-inspired computers.

Hardware implementation of an NCS requires massively parallel components (act as neurons), which are capable of interacting with each other as well as the external world through adaptive or programmable devices (act as synapses). NSCs implemented using digital and analog CMOS have been reported previously [3-6]. However, the CMOS implementation of such systems is inefficient from the area and power perspectives [7-8]. Such inefficiencies have driven a significant effort to propose novel computing approaches using beyond-CMOS technologies. Thanks to the advances in nanotechnology research as well as the development of new materials, the combination of novel spin-based devices and electronic components has shown promising potentials for implementing low-power high-density NCSs. In spintronic-based NCSs, magnetic reversal or frequency locking of free layer magnetic moment in a Magnetic Tunnel Junction (MTJ) is used to mimic neuron firing. Although spintronic-based NCSs are more power-efficient and denser than their CMOS implementation, their power density is still far from the brain. This is due to the fact that the traditional way of changing the state of magnetic moment through bias current consumes high power. Hence, there is a crucial need for shortening the bias current in current spintronic-based NCSs.

To address this issue, a new real time sensing circuit is proposed that can simultaneously work with stimulation phase and cuts off the bias current immediately after MTJ switches. In this way, the overall energy consumption and delay of the NCS will improve, significantly. Moreover, using the proposed circuit, the NCS operation will be similar to real neuron functionality in the brain that fires a spike as soon as it receives sufficient activation.

II. STATE OF THE ART

A. Neuron Implementation

In order to mimic "neuronal" functionality in NCS applications, two types of a neuron are utilizing - called Step neuron and Non-Step neuron [9]. The Step neurons are implemented by MTJ, Lateral Spin Valve (LSV) and Spin Orbit Torque (SOT) [9]. All these devices try to mimic the neuron firing by switching the magnetic moment in a ferromagnetic layer named Free Layer (FL). This is done by passing a

relatively high polarized current from FL leading to a high power consumption. In the MTJ-based NCS, this polarized current flows through NCS inputs. However, for LSV- and SOT-based NCS, the magnetic switching in FL is done in two steps. First, the magnetization of FL is preset along the "hard axis" at the beginning of the neuronal operation. Then, the input currents will be applied to the FL. In this way, the input currents will be decreased significantly. However, they suffer from two major issues: 1) very high preset current and 2) large delay due to the two-steps neuronal operation [9].

Another problem which is common in all methods is that there is a need for an extra phase to check whether the magnetization of the FL has been switched (neuron fired) or not (neuron stayed unchanged) i.e. read operation. In fact, neuronal functionality is done in two phases in MTJ-based neurons (FL excitation and read operation) and three phases in LSV- and SOT-based neurons (preset phase, FL excitation and read operation). This is in contrast with real neuron functionality in the human body, which is done in one phase. In a biological neuron, as soon as the received excitation exceeds a specific threshold, an action potential will be generated at its output (i.e. firing). As mentioned above, in all methods, the FL switching of the MTJ is used to mimic neuronal firing. Hence, exploring the FL switching behavior in MTJs will help to overcome the above-mentioned issues.

B. MTJ Basis

The schematic of the MTJ is shown in Fig. 1 (a). It consists of a Pinned Layer (PL) with fixed magnetization direction and the FL with changeable magnetization direction, which are separated by a tunneling oxide layer or a non-magnetic material such as MgO or AlO$_x$ .The resistance of the MTJ is determined

(a) *Free layer, Tunneling oxide, Pinned layer, Current*
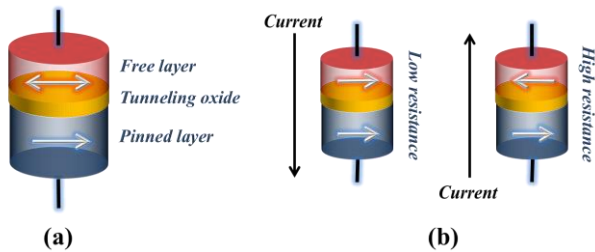**(b)** *Low resistance, High resistance, Current*

Fig. 1. (a) The schematic view of a MTJ as neuron and (b) the logic of MTJ according to the current direction.

by the relative magnetization direction of two ferromagnetic layers (Fig. 1 (b)). When the magnetization directions of the two magnetic layers are parallel (P-state) or anti-parallel (AP-state), MTJ resistance is low or high, respectively. In order to switch the magnetization direction of the FL, a spin-polarized current has to flow through the MTJ. This is done by applying an appropriate voltage to the terminals connected to the FL and PL and controlling the current direction. When the current flows from FL to PL, the magnetization direction of the FL and PL will be the same and the resistance of the MTJ is low. However, a current flowing from PL to FL will lead to an AP state with high MTJ resistance. In MTJ-based NCS, the magnetization reversal of the FL is used to mimic biological neuron firing [8-10]. Thus, the MTJ state can be determined by evaluating its resistance using a read circuitry.

C. MTJ-Based Neuromorphic Computing System

Fig. 2 shows the schematic view of an MTJ-based NCS. In order to perform synaptic functionality, a crossbar array of programmable memory devices (memristors) is employed to implement synapses to achieve a low-power operation. The resistance of the memristors can be tuned using electric signals. The crossbar array sums weighted input currents and the total current will pass through the MTJs, which will act as a neuron. When the input current of the MTJ is higher than its critical current, the magnetization of its FL will be switched. In the next phase, a peripheral circuit on the spintronic layer performs read operation to detect the state of the MTJ. The Switching of the MTJ means that the neuron has fired.

D. Stochastic Behavior of MTJ Switching

The total switching time of an MTJ consists of the *incubation time* and *transit time*. The incubation time is defined as the time required for electrons to climb up the potential barrier in an MTJ, while the transit time denotes the time for the electrons to descend the potential barrier to the other states [11-12]. Incubation time contributes to more than 90% of the total write time and the stochasticity of switching results from fluctuations in incubation time [13]. As a result, the switching behavior of the MTJ is inherently stochastic. In fact, the switching time for an MTJ with a specific current varies dramatically with a distribution having a long tail [11]-[14]. This stochasticity (i.e.
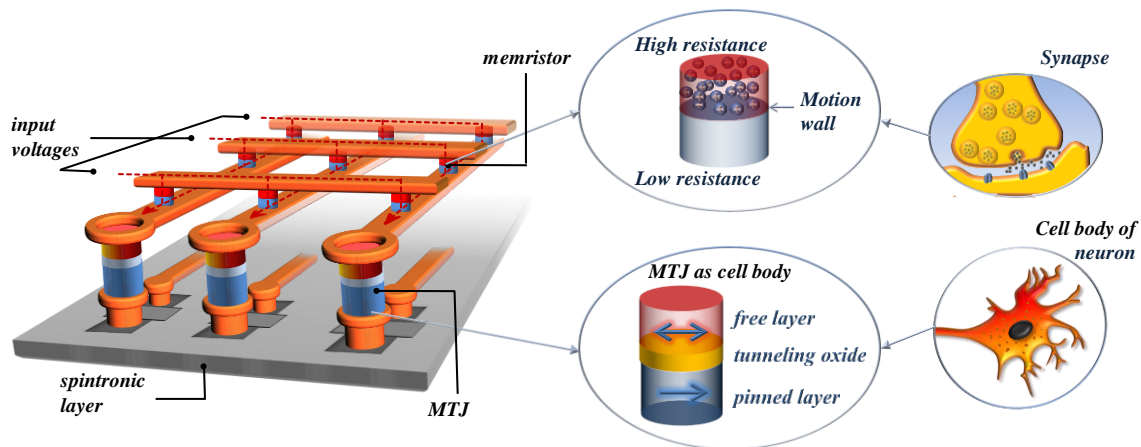

Fig. 2. The schematic view of a neuromorphic computing system. MTJs and memristors used to mimic neuronal and synaptic functionalities, respectively.

switching time variation) is hardly dependent on the MTJ current magnitude [15]. Higher MTJ current decreases the stochasticity at the cost of higher energy consumption and MTJ reliability degradation.

The MTJs in an NCS operate within dynamic reversal region and so the thermal fluctuations play an important role in their magnetization switching characteristics. Hence, the thermal fluctuations will lead to a large variation in switching time for a specific MTJ with constant switching current. Considering the switching current variations for different cells due to the process variations, the switching time varies severely for different MTJ-based neurons. In order to deal with this stochastic switching behavior of the MTJs, to guarantee a correct switching operation, a write pulse duration significantly longer than the average write time is required, which will lead to a large energy consumption and a low-speed operation of the MTJ-based NCSs.

E. Asymmetric Nature of MTJ Switching

As mentioned before, MTJ switching is done by applying a spin-polarized current in the appropriate direction. However, the switching behavior of the MTJ is asymmetric; thus switching to P and AP states take different time and energy. This asymmetry originates from two reasons. First, the MTJ resistance in P-state is lower than AP-state, which requires a higher switching current for P $\rightarrow$ AP transition compared with the AP $\rightarrow$ P transition. Second, the critical current for P $\rightarrow$ AP transition ($J_{P\rightarrow AP}$) is higher than that of AP $\rightarrow$ P transition ($J_{AP\rightarrow P}$) by 10%-50% [16-17]. Thus, in an MTJ, P $\rightarrow$ AP switching in comparison with AP $\rightarrow$ P switching takes longer time and higher energy. Hence, in an NCS, the MTJs are preset to the AP state before stimulation in order to decrease the energy consumption and delay [18].

III. PROPOSED REAL-TIME SENSING TECHNIQUE

As mentioned, there are two major issues associated with utilizing MTJs to mimic neuronal functionality in NCSs. First, the switching behavior of the MTJs is highly stochastic, which leads to a significant increase in the stimulation pulse in order to guarantee a correct switching operation. Utilizing longer stimulation pulse poses a penalty on speed and energy consumption. Second, the MTJ-based NCSs require two phases (write and read) and LSV- and SOT-based NCSs need three phases (preset, write and read) to complete the neuronal and synaptic actions. This is in contrast with real neuron functionality in the brain that fires a spike as soon as it receives sufficient activation. To this end, a novel Real-Time Sensing (RTS) technique is proposed in this section.

A. General Idea

Fig. 3 (a) shows the schematic of the NCS equipped with our proposed RTS technique. In this technique, the main idea is to perform real-time tracking of the MTJ state in order to cut off the MTJ current and stimulate the post neuron once the MTJ state is switched. Hence, the RTS technique not only acts similar to real neurons in the brain but also it decreases the overall energy consumption, significantly. Switching of the MTJ state from AP- to P-state reduces its resistance. Hence, the voltages of different nodes connected to the MTJ will change. In RTS technique, these voltage changes are used to sense the MTJ switching. As shown in Fig. 3 (b), there are two nodes n1 and n2 suitable for detecting the MTJ switching. After MTJ switching, V_{n1} increases and V_{n2} decreases. However, as it will be elaborated in the next sub-section, the voltage change of the
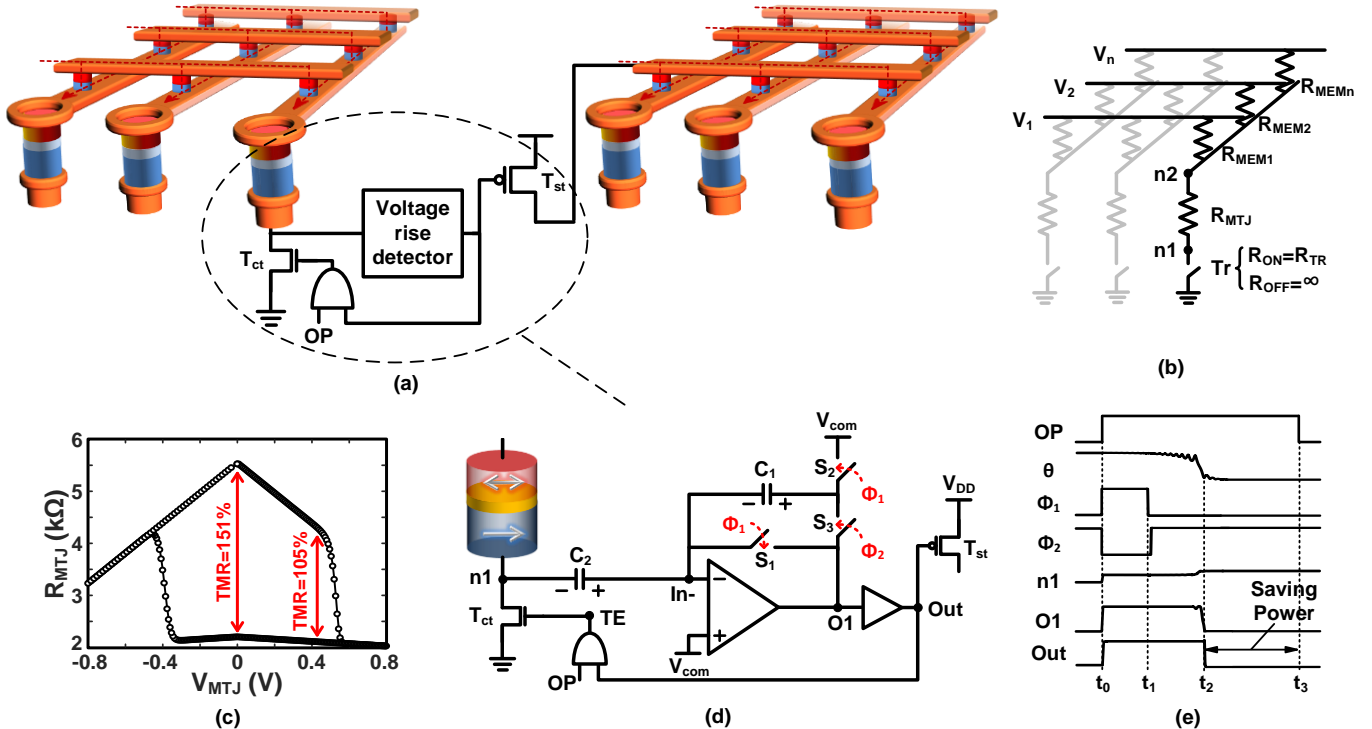

Fig. 3. (a) The schematic of an NCS and the proposed real-time sensing technique, (b) the resistive equivalent simplified circuit of NCS, (c) the MTJ resistance house curve, (d) the circuit implementation of proposed real-time sensing technique and (e) its timing diagram.

V_{n1} is higher than for the V_{n2}. Hence, the voltage rise on n1 is used for sensing the MTJ state switching in the proposed RTS circuit as shown in Fig. 3 (d). The voltage rise detector circuit connected to n1 senses MTJ switching and changes its output from '1' to '0'. The output of voltage drop detector circuit controls the gate of control transistor (T_{ct}) through the AND gate and the gate of stimulation transistor (T_{st}). As a result, the control transistor turns off (cuts off the MTJ current) and T_{st} turns on and stimulates post NCS.

B. Mathematical Analysis of Different Nodes Voltage Change

An NCS can be modeled by the resistive equivalent simplified circuit as shown in Fig. 3 (b). R_{MEMx} and R_{MTJ} (Fig. 3 (c)) are used to model the memristors and the MTJ, respectively, and the control transistor is modeled by a switch with $R_{ON}=R_{Tr}$ and $R_{OFF}=\infty$. The voltage of the common node between MTJ and memristors (V_{n2}) can be calculated as follows:

$$V_{n2} = \frac{(R_{MTJ}+R_{Tr})||(R_{MEM2}||\cdots||R_{MEMn})}{(R_{MTJ}+R_{Tr})||(R_{MEM2}||\cdots||R_{MEMn})+R_{MEM1}}V_1 +$$
$$\frac{(R_{MTJ}+R_{Tr})||(R_{MEM1}||R_{MEM3}||\cdots||R_{MEMn})}{(R_{MTJ}+R_{Tr})||(R_{MEM1}||R_{MEM3}||\cdots||R_{MEMn})+R_{MEM2}}V_2 +$$
$$\vdots$$
$$+ \frac{(R_{MTJ}+R_{Tr})||(R_{MEM1}||\cdots||R_{MEMn-1})}{(R_{MTJ}+R_{Tr})||(R_{MEM1}||\cdots||R_{MEMn-1})+R_{MEMn}}V_n \quad (1)$$

When an MTJ switches from AP-state to P-state (neuron fires), the resistance of the MTJ will decrease while the resistance of the memristors remains unchanged. Considering the constant V_{GS} of the control transistor before and after switching, R_{Tr} is somehow constant. Hence, based on (1), switching of the MTJ to P-state will degrade V_{n2} voltage. This voltage drop can be used to sense MTJ switching. Another node that can be used for sensing MTJ switching is the common node between MTJ and control transistor (n1). The voltage of n1 (V_{n1}) can be calculated based on V_{n2} as follows:

$$V_{n1} = \frac{R_{Tr}}{R_{Tr}+R_{MTJ}}V_{n2} \quad (2)$$

Noted, between n1 and n2 nodes, the node that has higher voltage change is better for sensing the MTJ switching. In order to compare the absolute change of the V_{n1} and V_{n2} due to the MTJ switching, two assumptions have been made to simplify the equations. First, all input voltages are supposed to be equal ($V_1=V_2=\ldots=V_n=V_i$). Second, similar resistance is considered for all memristors ($R_{MEM1}=R_{MEM2}=\ldots=R_{MEMn}=R_{MEM}$). According to the above-mentioned assumptions, (1) and (2) are rewritten as follows:

$$V_{n2} = \frac{(R_{MTJ}+R_{Tr})||\left(\frac{R_{MEM}}{n-1}\right)}{(R_{MTJ}+R_{Tr})||\left(\frac{R_{MEM}}{n-1}\right)+R_{MEM}}V_i \times n \quad (3)$$

$$V_{n1} = \frac{(R_{MTJ}+R_{Tr})||\left(\frac{R_{MEM}}{n-1}\right)}{(R_{MTJ}+R_{Tr})||\left(\frac{R_{MEM}}{n-1}\right)+R_{MEM}}V_i \times n \times \frac{R_{Tr}}{R_{Tr}+R_{MTJ}} \quad (4)$$

When the number of inputs (n) increases the term $R_{MEM}/(n-1)$ decreases. Hence, for sufficiently large number of inputs, (3) and (4) can be approximated by following equations:

$$V_{n2} = \frac{\left(\frac{R_{MEM}}{n-1}\right)}{\left(\frac{R_{MEM}}{n-1}\right)+R_{MEM}}V_i \times n \quad (5)$$

$$V_{n1} \cong V_i \times \frac{R_{Tr}}{R_{Tr}+R_{MTJ}} \quad (6)$$

Equations (5) shows that for a large number of inputs, the voltage of n2 is independent of R_{MTJ}. Hence, the MTJ switching does not change V_{n2}. However, based on (6), V_{n1} can change effectively by R_{MTJ}. Based on our simulations, if the number of active inputs is equal or greater than three, the voltage change of n1 will be larger. Noted, in real cases that the memristor resistances are different, R_{MEM} will be smaller, which leads to a smaller voltage change on V_{n2}. Considering the fact that the number of inputs in an NCS is usually more than three, node n1 is used in the proposed RTS circuit to sense the switching of the MTJ. In rare cases, if the number of inputs will be lower than three or the equivalent resistance of memristor network (R_{MEM}) will be higher than MTJ resistance (R_{MTJ}), V_{n2} can be used as the input of voltage rise detector circuit.

C. Voltage Rise Detector Circuit

In order to detect the voltage rise of the node n1, a new voltage rise detector circuit shown in Fig. 3 (d) is proposed and its timing diagram is shown in Fig. 3 (e). In order to decrease the write power consumption of Spin-Transfer-Torque Random Access Memories (STT-RAMs), techniques to detect the MTJ switching and terminate the current through it during write have been previously proposed in [12, 19-22]. However, to our knowledge, this approach is used in NCSs for the first time. Moreover, previously proposed techniques are not robust against Process-Voltage-Temperature variations (PVT). These methods use the MTJ voltage change [12, 19-21] or current change [22] due to MTJ switching in order to detect the switching and then terminating the current. The high sensitivity of techniques proposed in [12, 19-20] to process variations discussed in [21]. In [21], the sensitivity of STT-RAM array due to process variations decreases using a Self-referenced Differential Write Termination (SDWT) block [21]. On the other hand, the current change method from [22] suffers from extra power consumption in the reference MTJ as well as area overhead due to the used dummy blocks. However, in the proposed RTS circuit, the problem of variations in the decision point of comparator due to PVTs is solved, which significantly increases its robustness against PVTs.

Fig. 4 (a) shows the circuit implementation of the proposed RTS circuit of Fig. 3 (d), while the S1 and S3 are replaced with transmission gates and S2 is replaced by PMOS transistor. In order to minimize the sensitivity of the RTS-based NCS to PVTs, two techniques are utilized.

1) Self-Referenced Sense Scheme: The idea is to sample the primary voltage of node n1 on C_2 as a reference voltage instead of using a constant reference voltage. As a result, the Vn1 change will be compared with its primary voltage stored on C_2. Hence, the V_{n1} variations due to PVTs has no effect on tracking the MTJ switching due to the fact that V_{n1} changes are compared with its initial value.

2) Auto-Zeroing Technique: The idea is to sample the input referred offset of the amplifier on C_2 capacitor at the beginning of NCS stimulation time, and then subtracting it from n1 voltage for the rest of the stimulation time. In this way, the
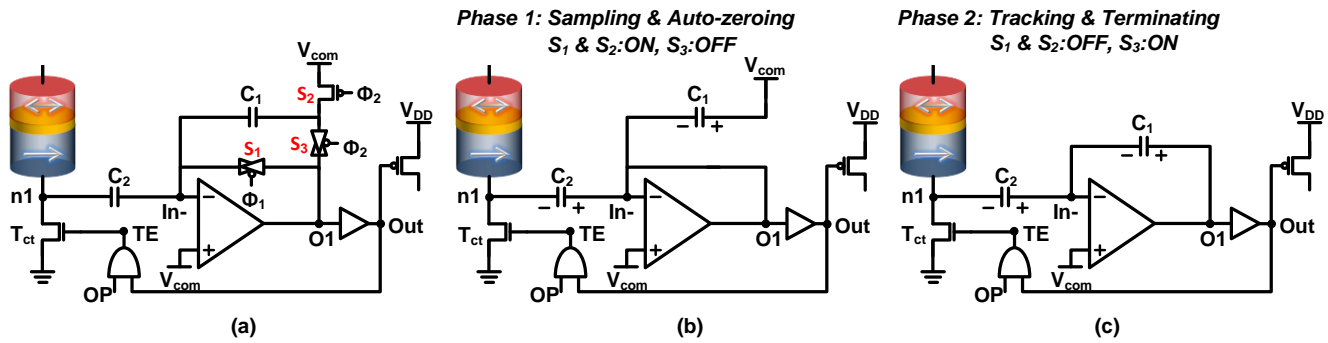
Fig. 4. (a) Circuit implementation of voltage rise detector block in (b) sampling and auto-zeroing phase and (c) tracking and terminating phase.

amplifier offset due to PVTs will be canceled, which increases the RTS robustness against PVTs, significantly.

The RTS operation time can be divided into two phases. Phase 1 that includes sampling and auto-zeroing and Phase 2, which includes tracking and terminating. Fig. 4 (b) and (c) show the RTS circuit in Phase 1 and Phase 2, respectively. In Phase 1, $t_0 < t < t_1$ in Fig. 3 (e), the signals Φ_1 and Φ_2 are '1' and '0', respectively. As a result, S_1 and S_2 are closed and S_3 is open as shown in Fig. 4 (b). Hence, the amplifier is in unity-gain configuration and V_{In-} will be settled at $V_{com} + V_{offset}$. So, the capacitors C_2 and C_1 are charged to $V_{com} + V_{offset} - V_{n1}$ and V_{offset}, respectively. As a result, in Phase 1, not only the primary voltage of the node n1 will be stored on C_2 (sampling), but also the input offset voltage of the amplifier will be stored on C_1 and C_2 (auto-zeroing). The combination of these two techniques makes the RTS-based NCS robust against PVTs. It is worth noting that Phase 1 ($t_0 < t < t_1$) will be chosen shorter than the minimum time required for the MTJ to switch (i.e. 1.12ns, as will be discussed in Section V). Hence, it will not degrade the operation of the NCS. Then, at Phase 2 ($t > t_1$), the signals Φ_1 and Φ_2 become '0' and '1', respectively. Hence, S_1 and S_2 are open and S_3 is closed. Note that the rising edge of Φ_2 is slightly after the falling edge of Φ_1 (bottom plate sampling) to eliminate the charge injection effect on V_{C1} due to the turned off S_2. As a result, in Phase 2, C_1 will create a negative feedback on the amplifier and In- becomes a high impedance node as shown in Fig. 4 (c). This negative feedback keeps V_{In-} at $V_{com} + V_{offset}$. When MTJ switches at t_2, V_{n1} will rise (tracking). On the other hand, the amplifier tries to keep the V_{In-} at $V_{com} + V_{offset}$. As a result, some positive charge from C_2 will be transferred to C_1. This will decrease the voltage of C_1, which leads to a voltage drop on node O_1. In case the capacity of C_2 will be sufficiently higher than C_1, the voltage drop on C_1 will be large enough to change V_{O1} from V_{DD} to ground. This will lead to turned off T_{ct} through AND gate (terminating).

Different amplifier topologies such as two-stage, telescopic and folded-cascode amplifiers can be used here. Amongst them, the two-stage amplifier offers the highest gain while the worst settling time in closed-loop configuration [23-24]. Considering the fact that increasing the settling time of the amplifier increases the delay of the voltage detector circuit, utilizing the two-stage amplifier decreases the advantage of power improvement of the proposed RTS circuit. Among other two amplifiers, the telescopic amplifier is faster than the folded-cascode amplifier. However, due to the higher number of

stacked transistors in the telescopic amplifier, the minimum supply voltage will be limited, which can lead to higher power consumption. Hence, the folded-cascode amplifier is utilized in our proposed RTS circuit. The folded-cascode amplifier is designed in sub-threshold region to achieve the least power consumption.

The advantage of the proposed RTS circuit is that it can be used in different MTJ-based architectures such as step-neuron MTJ-based NCS, LSV- and SOT-based NCS [9], non-step stochastic MTJ-based NCSs [25] and STT-RAMs [19-22] in order to improve the energy consumption and speed. In all above-mentioned architectures, MTJ switching leads to a resistance change, which changes the voltage of different nodes of the circuit and the proposed RTS circuit can detect this voltage change and cut off the MTJ current.

IV. SIMULATION RESULTS

The simulation results for the RTS-based NCS in 65nm CMOS technology are presented in this section. The sampling capacitors, C1 and C2, are implemented using MOSCAP and MIMCAP, respectively. Simulations are performed at the supply voltage of 1V and the temperature of 25°C in HSPICE simulator. Different SPICE-compatible MTJ models have been proposed, previously [26-29]. In this paper, the compact modular MTJ model presented in [29] with stochastic LLG solver block is used. This is a flexible model and allows the user to define the physical device dimensions such as T_{MgO}, W_{MTJ} and L_{MTJ}. The MTJ parameters used in this paper are listed in Table I. Several memristor models including linear ion drift model [30], nonlinear ion drift model [31], Simmons tunnel barrier model [32], and Threshold Adaptive Memristor Model (TEAM) [33] have been proposed in literature. Among them,

TABLE I
NCS CHARACTERISTICS

MTJ [29]	Free Layer Dimensions	40nm×116nm×1.5nm
	Oxide Thickness (T_{MgO})	1.5nm
	Saturation Magnetization (M_S)	800 emu/cm^3
	Damping Factor (α)	0.01
	Gyromagnetic Factor (γ)	17.6 GHz/Oe
Transistor	Supply Voltage (V_{DD})	1V
	Technology	65nm
	dimensions: W/L	200nm/70nm
Memristor [33]	R_{ON}/R_{OFF}	100Ω/10000Ω
	Thin film thickness	3nm
	K_on	-8x10^{-13}
	K_off	8x10^{-13}
	α	3

the TEAM model presented in [33] is a very flexible and accurate model offering the capability of simulating memristors with different physical structure. In this paper, we use TEAM model.

A. Effect of Stochasticity and Proccess Variation on MTJ Switching Time

As mentioned in section II, the MTJ switching time is stochastic due to thermal fluctuations. Moreover, process variations can change the characteristics of MTJs, memristors, and transistors of the NCS that leads to variation in MTJ switching time. In this subsection, the effects of process variations and thermal fluctuations on MTJ Switching Time (ST) are explored. To this end, a Monte Carlo (MC) simulation of 1000 iterations is run for different MTJ current values above the threshold current for which the corresponding ST values are calculated. Fig. 5 (a) shows the mean value and standard deviation of the ST values for different MTJ currents. The ST histogram and its fitted positively skewed function for I_{MTJ}=90μA are shown in Fig. 5 (a). As a result of thermal fluctuations, the distribution of the ST is positively skewed for different MTJ currents, which means that ST is more dispersed above its mean value. Noted, the thermal fluctuations effect decreases the ST mean value on one hand and on the other hand, it increases the stochasticity of the ST.

To calculate the maximum and minimum of ST, the 6σ worst case of the ST is calculated for different MTJ currents as shown

in Fig. 5 (b). The longest ST of 16.92ns is achieved at I_{MTJ}=70μA and the minimum ST is calculated as 1.12ns at I_{MTJ}=140μA. Note that the minimum 6-sigma worst case of ST is almost constant for higher MTJ currents. In order to guarantee the MTJ switching in typical NCS, a pulse period higher than 16.92ns is required. On the other hand, the sampling period of the proposed RTS circuit (Phase 1) should be shorter than 1.12ns. This will ensure the correct operation of the RTS circuit.

B. Transient Simulation

The impact of utilizing the proposed RTS circuit on detecting MTJ state switching and terminating its current in an NCS is explored through transient simulation for I_{MTJ}=90μA as shown in Fig. 6. θ is the angle between FL and PL magnetization vectors. When $\theta = 0$, the two layers are in the P-state and $\theta = \pi$ means they are in the AP-state. The NCS stimulation starts at t=1ns by enabling OP signal. The signal Φ_1 is also enabled at the same time with OP (t=1ns) for 0.6ns (the Φ_2 signal goes '0' during this period). As a result, the C_1 and C_2 capacitors of RTS circuit are precharged to V_{offset} and $V_{com}+V_{offset}-V_{n1}$, respectively. As mentioned in the previous subsection, the activation time of Φ_1 should be chosen in a way that it will be shorter than the shortest MTJ switching time (i.e. 1.12ns). However, if in a rare case switching occurs earlier than 0.6ns, it will not affect the NCS operation. In this case, the RTS circuit cannot terminate the MTJ stimulation and the MTJ current will cut off when the OP signal goes '0' (like a typical NCS).
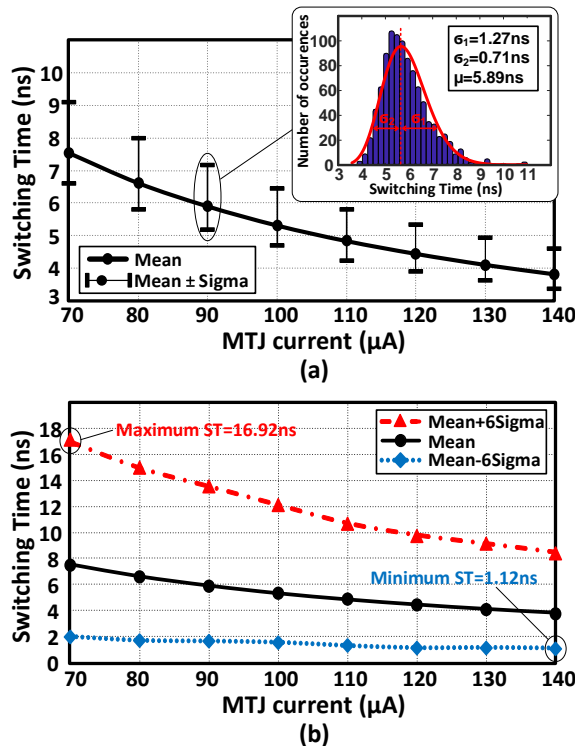

Fig. 5. (a) The mean value and standard deviation of MTJ switching time for different MTJ currents above switching threshold current. In each current point a Monte Carlo simulation with 1000 iterations is run. The histogram and its fitted positively skewed function of MTJ switching time for I_{MTJ}=90μA is depicted. (b) The mean value of MTJ switching time and its 6-sigma variation for different MTJ currents. The maximum and minimum switching time are calculated as 16.92ns at I_{MTJ}=70μA and 1.12ns at I_{MTJ}=140μA.
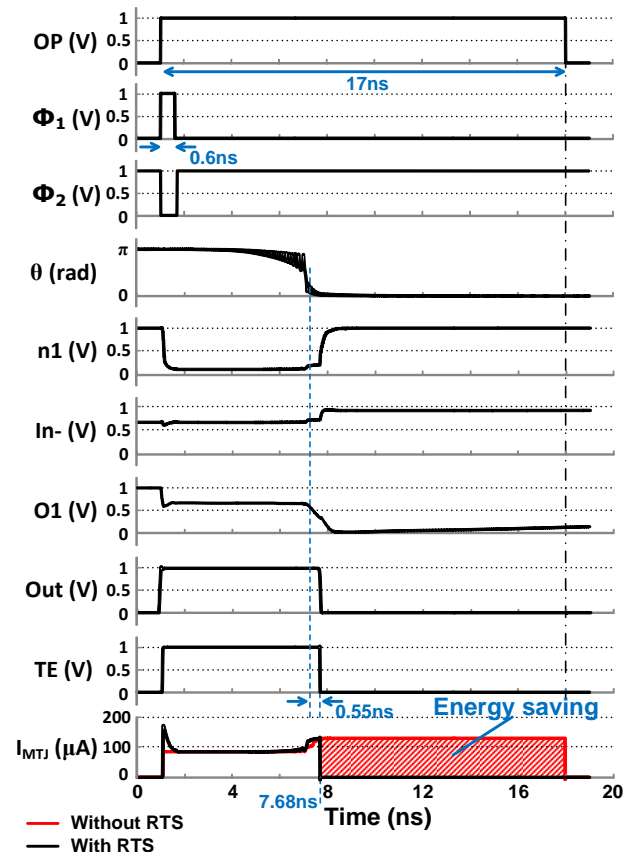

Fig. 6. Transient simulation of a NCS equipped with proposed real-time sensing technique.

After disabling Φ_1, the In- node turns to a high impedance node and its voltage remains at $V_{com}+V_{offset}$. When switching occurs at t=7.13ns, V_{n1} increases leading to an increase in V_{In-}. However, the amplifier tries to keep V_{In-} at $V_{com}+V_{offset}$. As a result, the output voltage of amplifier (V_{O1}) starts decreasing from $V_{com}+V_{offset}$ to ground. The speed of V_{O1} reduction determines the delay of RTS circuit, which depends on amplifier gain and the total capacitance of the O1 node. Hence, increasing the gain of amplifier decreases the total delay of RTS at the cost of extra energy consumption. The node O1 is followed by a low-skewed buffer. Hence, when V_{O1} drops below 0.3V, the output of low-skewed buffer (node Out) switches from '1' to '0' at t=7.68ns. As a result, the TE signal goes '0' and the control transistor (T_{cr}) turns off. Hence, the MTJ current becomes zero leading to a significant energy saving as shown in Fig. 6.

C. Energy-Delay Optimization

The energy consumption of the RTS-based NCS can be calculated by summing the energy consumption of RTS circuit and the energy consumption of NCS during stimulation phase. On the other hand, the delay of RTS circuit can affect the total energy consumption in two ways. First, decreasing the RTS delay leads to lower energy consumption in NCS circuit due to the faster MTJ current cut-off after switching. Second, the lower delay in RTS circuit is achieved by a higher bias current of the amplifier that leads to a higher energy consumption of RTS circuit. Hence, there is a need to find an optimal delay that minimizes the total energy consumption. To do this, the total energy consumption of RTS-based NCS, the energy consumption of RTS circuit and the energy consumption of NCS are calculated for different RTS delays and different MTJ currents as shown in Fig. 7. The RTS delay is calculated as the time difference between when θ reaches to 10% of its final state and the TE signal drops to 0.1V. As expected, by increasing the delay, the energy consumption of RTS circuit and NCS are decreased and increased respectively. Hence, there is an optimal point that makes the total energy consumption

minimum. As illustrated in Fig. 7, the optimal RTS delay of 0.55ns is achieved for different MTJ currents. The minimum total energy consumption at optimal delay point varies from 950fJ for I_{MTJ}=70μA to 833fJ for I_{MTJ}=140μA.

D. Energy Reduction

The total energy consumption of the typical NCS includes the energy consumes in the array (memristors and MTJs) and the energy consumption of the read circuit. After stimulating the target MTJ (OP goes '0'), the read circuit reads its state. If the MTJ state is switched, the read circuit stimulates the post MTJ-based neuron. However, in the RTS-based NCS, the read circuit is replaced by the RTS circuit. In fact, if the MTJ switching happens, the RTS circuit not only terminates MTJ current but also stimulates the post MTJ-based neuron. Hence, in order to compare the energy consumption of typical NCS with the RTS-based NCS, the total energy consumption of NCS and read circuit should be compared with the total energy of the NCS and the RTS circuit. In order to calculate the energy consumption, the average switching times calculated in section IV. A are used.

Fig. 8 shows the total energy consumption of the typical and the RTS-based NCSs at different MTJ currents. The energy consumption of read circuit in typical NCS and RTS circuit in RTS-based NCS are shown by dashed colors. In the typical NCS, the MTJ current will flow after MTJ switching until the end of stimulation that is determined by the OP signal (17ns). As a result, increasing the MTJ current will lead to an increase in the energy consumption of the typical NCS. Noted, after MTJ switching, the MTJ current increases due to the MTJ resistance reduction. The average read power of 93μW with the average read duration of 1ns are considered for read circuit in typical NCS [18].

In the RTS-based NCS, the MTJ current stops after MTJ switching as shown in Fig. 6 that leads to a significant energy reduction. The energy consumption of RTS circuit is related to charging C_1 and C_2 capacitors into suitable voltages, and the amplifier bias current. The designed amplifier and its bias circuitry consume 64.45μW and 6.02μW of power, respectively i.e. the total power consumption of 70.47μW. The energy consumption of the RTS-based NCS is equal to the total energy consumption of the NCS and the RTS circuit, which decreases
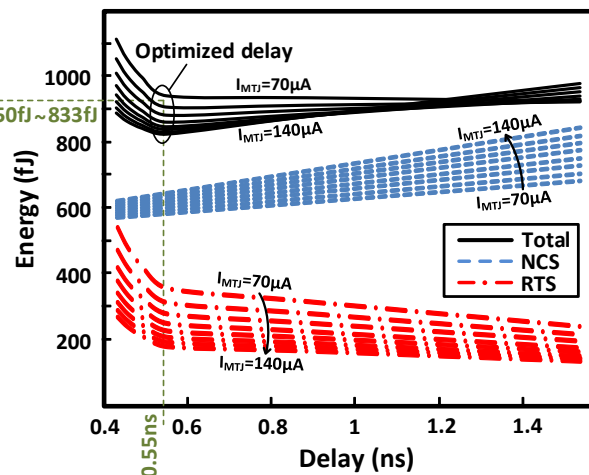

Fig. 7. The total energy consumption of RTS-based NCS, the energy consumption of RTS circuit and energy consumption of NCS versus delay of RTS circuit for different MTJ currents. When the delay of RTS circuit is 0.55ns, the total energy consumption is minimized at 950fJ for I_{MTJ}=70μA up to 833fJ for I_{MTJ}=140μA.
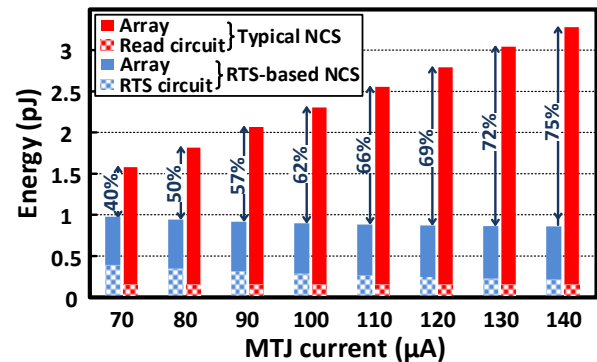

Fig. 8. The total energy consumption of RTS-based NCS and typical NCS versus different MTJ currents. The percentage of energy improvement of RTS-based NCS is shown. The energy consumption of read and RTS circuits in typical and RTS-based NCS are shown by dashed color.

from 950fJ at I_{MTJ}=70μA to 833fJ at I_{MTJ}=140μA. This energy reduction is attributed to lower RTS energy consumption at higher MTJ currents. In fact, by increasing the MTJ current, the switching time will decrease. As a result, the RTS circuit will be turned on for a shorter period of time, which reduces the energy. As illustrated in Fig. 8, by increasing the MTJ current, the energy consumption improvement of the RTS-based NCS increases from 40% at I_{MTJ}=70μA to 75% at I_{MTJ}=140μA. Hence, the energy improvement of the RTS-based NCS in comparison with the typical NCS is at least 40% and further reduction will depend on the application.

E. Speed Improvement

The NCS delay can be calculated as the difference between the time the neuron is stimulated and the time at which the neuron stimulates the post neuron. In the typical NCS, the delay is constant and it is equal with the sum of the required time to guarantee switching of the MTJ (OP signal duration) and the delay of read circuit. As mentioned in subsection IV.A, the minimum OP duration while considering 6σ variations is 16.92ns. The overall delay of a read circuit can be estimated as 1ns [18]. As a result, the NCS delay is 17.92ns, which leads to a maximum frequency of 55.8MHz.

In the RTS-based NCS, as shown in Fig. 9, the NCS delay reduces with the increase of MTJ current. This is attributed to 1. the MTJ switching time decreases by increasing the MTJ current and 2. the RTS circuit will stimulate the post neuron immediately after MTJ switching. The frequency of the RTS-based NCS increases from 124MHz at I_{MTJ}=70μA to 229MHz at I_{MTJ}=140μA. Hence, at the worst case (i.e. I_{MTJ}=70μA), 2.22X speed improvement will be achieved by the proposed RTS circuit.

F. Effect of Process Variations on RTS Operation

As fully discussed in sub-section III.C, the self-referenced sense scheme and auto-zeroing technique are utilized to minimize the sensitivity of the RTS-based NCS to PVTs. Here, the effectiveness of these techniques on the robustness of RTS-based NCS in the presence of process variations is explored. To this end, a MC simulation with 1000 iterations is run on the RTS-based NCS. Then, the amplifier output voltage (V_{O1}) variations and the buffer decision point variations are calculated (Fig. 10). When V_{O1} drops below buffer Decision Point (DP)

due to process variations (before MTJ switching), TE signal goes to '0'. This will cut off the corresponding MTJ current in a wrong way. We call it *hard error* due to the fact that this makes the NCS output wrong.

As shown in Fig. 10 (a), in the RTS-based NCS, the DP has a normal distribution with 8.8mV standard deviation while the V_{O1} distribution is negatively skewed with a right tail standard deviation (σ1) of 8.6mV and left tail standard deviation (σ2) of 4.4mV. Hence, the chance of fault in RTS-based NCS is almost 0 (18.9σ variation on V_{O1} and DP voltages is needed to make a failure at the output of NCS as shown in Fig. 10 (a)).

G. Comparison with SDWT

In order to explore the effectiveness of the proposed RTS technique over previously proposed circuits, a comprehensive comparison is done between the RTS circuit and the best previously proposed circuit (SDWT). Considering the fact that the SDWT circuit in [21] is designed to work with STT-RAM array, the optimized version of it is used here. Fig. 10 (b) shows the histogram and fitted Gaussian function of V_{O1} and DP while using SDWT circuit of [21] as MTJ current termination block. The DP has a normal distribution with 7.6mV standard deviation and V_{O1} has a negatively skewed distribution with a right tail standard deviation (σ1) of 18.8mV and a left tail standard deviation (σ2) of 62.7mV. The large increase of V_{O1} standard deviation when using SDWT compared with the RTS circuit is due to the comparator offset. This will significantly increase the probability of erroneous output of NCS (Only 4.5σ variation on V_{O1} and DP voltages can make NCS output wrong). As shown in Fig. 10 (b), in the MC simulation with 1000 iterations, a *hard error* happens in five iterations when
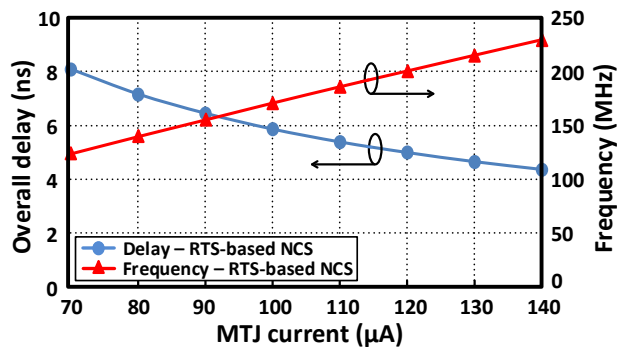

Fig. 9. The overall delay and frequency of RTS-based NCS versus different MTJ currents. By increasing the MTJ current, the overall delay of RTS-based NCS reduces which leads to its frequency increase.
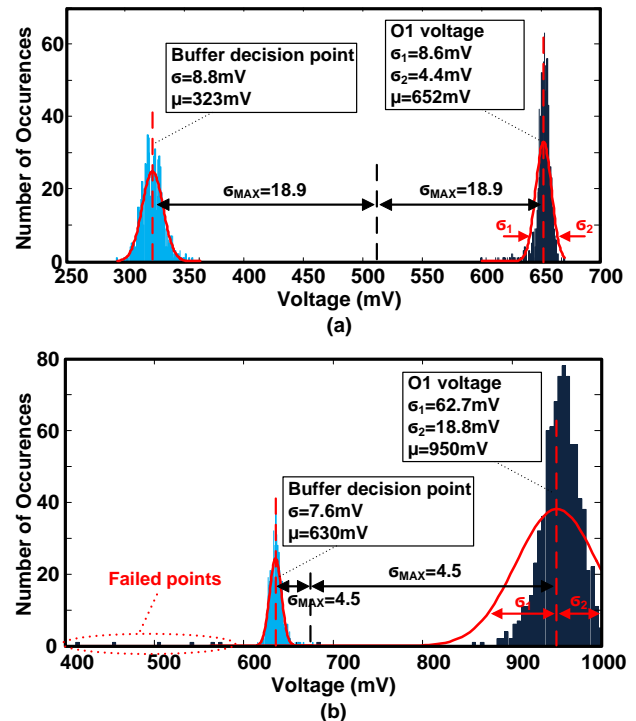

Fig. 10. The histogram and fitted Gaussian function of V_{O1} and DP due to process variations for a MC simulation with 1000 iterations while using (a) RTS-based NCS and (b) SDWT-based NCS.

TABLE II
OPERATING TEMPERATURE RANGE, MINIMUM SUPPLY VOLTAGE, DELAY AND POWER CONSUMPTION OF RTS AND SDWT CIRCUITS

Corners Parameters	Proposed RTS					SDWT [21]				
	TT	FF	FS	SF	SS	TT	FF	FS	SF	SS
Temperature (°C) [a]	> -29	> -54	> -19	> -22	> 6	> 0	> -14	> 14	> 8	Failed
V_{DDmin} (V) [b]	0.88	0.78	0.9	0.88	0.96	0.94	0.84	0.95	0.95	Failed
Delay (ps) [c]	546	454	890	338	1468	797	423	948	672	Failed
Power (µW) [c]	43.34	55.68	41.46	45.42	34.3	44.04	58.04	43.6	45.06	Failed

[a] The operating temperature range is calculated at nominal supply voltage (1V) for different corners.
[b] The minimum supply voltage is calculated at room temperature (25°C) for different corners.
[c] The delay and power are calculated at room temperature (25°C) and nominal supply voltage (1V) for different corners.

using SDWT circuit.

The process corners are the worst process variations conditions that a circuit should work correctly in them. In Table II, the operating temperature range, minimum supply voltage, delay and power consumption of RTS and SDWT circuits are tabulated for different process corners. The SDWT circuit failed to track MTJ state and cut off its current at SS corner. However, the RTS circuit works correctly at all process corners. The operating temperature range is calculated at the nominal supply voltage of 1V. There is no temperature upper limit for both circuits. However, the temperature lower limit of the RTS circuit is lower than the SDWT circuit at all corners. The minimum supply voltages of the RTS and the SDWT circuits are calculated at room temperature for all process corners. In all process corners, the minimum supply voltage of the RTS circuit is lower than the SDWT circuit. The delay of the RTS circuit is less than the SDWT circuit at all corners except the FF corner that the SDWT is faster. Finally, the power consumption of two circuits is compared. Again, the RTS circuit has a lower power consumption than the SDWT at all process corners. However, the difference is negligible. This is due to the fact that the most power consuming component is the comparator in tracking phase, which is the same for both circuits.

The area of the RTS circuit is estimated as 55.08 µm², which is 35% lower than the SDWT circuit. The capacitor C1 in the RTS circuit and both sampling capacitors of the SDWT circuit are implemented using MIMCAP on top of the rest of the circuit in order to decrease the area overhead. For both SDWT and RTS circuits, the area of sampling capacitors is dominant. The improved area overhead of the RTS circuit is related to the smaller sampling capacitors in comparison with the SDWT circuit. Noted, in the RTC-based NCS, the read circuit of the typical NCS is replaced by the RTS circuit. The read operation in the typical NCS is similar to STT-RAM where the read circuit senses the MTJ state by passing a read current through the MTJ and calculating its resistance. However, the read current should be sufficiently lower than the critical current to prevent MTJ switching. The low read current significantly increases the sensitivity of the read circuit to PVT [34-35]. The proposed RTS circuit occupies 29% lower area and 24% higher area compared with the OCCS-SA [34] and CSB-SA [35], respectively.

H. Image Processing Applications

The effect of RTS circuit on NCS energy consumption is explored in two common image processing applications

including edge detection and character recognition. To do this, first, the behavioral model of the NCS and the RTS circuit is obtained [36-37]. It is done by extracting the mean value and standard deviations of the MTJ switching time for different MTJ currents through Hspice simulation results and fitting the equations to them. Second, the relation between input pattern and the MTJ current is modeled for edge detection and character recognition. Then, the behavioral models (fitted equations) are used in Matlab to determine the switching time based on the MTJ current. Finally, knowing the MTJ current and its switching time, the energy consumption and the delay of the RTS-based NCS and the typical NCS can be calculated in each application.

Edge detection: Fig. 11 shows the simulation results of edge detection on a 512×512 image in Matlab. The edge detection is done through the 3×3 Sobel operator edge detector [38]. For each pixel, a 3×3 neighborhood of the pixel is used as the 9 inputs of the crossbar array of NCS. If the resulting current will be higher than a threshold, the MTJ switches and the pixel is an edge. The Sobel operator is applied to each pixel to decide whether it is an edge or not. In each case, the energy consumption and delay is calculated for the RTS-based NCS and the typical NCS. The total energy consumption and the

Fig. 11. Simulation results of edge detection on an 512×512 image with the best suited threshold. The simulation is done through behavioral modeling of the NCS and RTS circuit in Matlab.

TABLE III
ENERGY CONSUMPTION AND DELAY
OF NCS FOR EDGE DETECTION APPLICATION

Method	Energy consumption		Delay		EDP [c] (nJ×ms)
	Total (nJ)	Pixel [a] (fJ)	Total (ms)	Pixel [b] (ns)	
Typical NCS	396.5	1512.6	4.46	17	1768
RTS-based NCS	151.3	577.1	3.8	14.5	575
Improvement	61.8 %		14.7 %		67.5 %

[a] Energy consumption per pixel = Total energy / (rows × columns)
[b] Delay per pixel = Total delay / (rows × columns)
[c] Energy delay products

TABLE IV
ENERGY CONSUMPTION AND DELAY
OF NCS FOR CHARACTER RECOGNITION APPLICATION

Test images / Train images								
Similarity(a) (%)	100	92.9	86.5	78.6	75.4	71.4	69	64.3
	14.3	19.8	26.2	19.8	5.6	9.5	7.1	11.9
	28.6	21.4	15.1	7.1	8.7	6.3	10.3	8.7
	46	38.9	32.5	24.6	37.3	33.3	31	18.3
	0.1	0.2	6.3	6.4	0.2	2.4	7.9	11.1
	41.3	37.3	34.1	27.8	40.5	36.5	32.5	26.2
	61.1	54	49.2	41.3	50.8	48.4	47.6	36.5
	0.06	0.04	0.01	0.04	0.05	0.05	0.06	0.05
	49.2	42	35.7	27.8	35.7	31.7	32.5	26.2
	58.7	51.6	45.2	42.1	40.5	36.5	37.3	35.7
Energy cons. (pJ) Typical	10.42	9.41	8.7	7.33	7.66	7.18	7.11	6.21
RTS	2.55	2.38	2.3	2.04	2.25	2.19	2.24	2.04
Average improvement	71.9 %							
Delay (ns) Typical	17	17	17	17	17	17	17	17
RTS	5.07	5.3	5.53	5.89	6.07	6.3	6.44	6.76
Average improvement	65 %							
EDP (pJ×ns) Typical	177.2	159.9	147.9	124.6	130.3	122.1	120.8	105.5
RTS	12.9	12.6	12.7	12.1	13.6	13.8	14.4	13.8
Average improvement	90.3 %							

(a) Similarity indicates the percentage of similarity between the train images of 0-9 and test images of 0

delay are calculated through summing up the calculated energy consumptions and the delays of all pixels. The total energy consumption and the delay, the energy consumption and delay per pixel and the Energy-Delay Product (EDP) of the edge detection application in Fig. 11 are tabulated in Table III. The energy consumption and the delay of NCS are improved by 61.8% and 14.7%, respectively, by the use of the proposed RTS circuit compared to the typical NCS. Noted, during edge detection, MTJ switching is only happening for edge pixels. For non-edge pixels, the MTJ current is less than the minimum required switching current. Hence, considering the fact that the RTS circuit detects MTJ switching and cuts off its current, it can only improve the energy consumption and delay for the edge pixels. As a result, the greater the number of edge pixels, the efficiency of the RTS circuit will increase. Considering the higher number of non-edge pixels compared with edge pixels and the long delay of non-edged pixels (17ns), the total delay will be mostly determined by non-edge pixels. However, the total energy consumption is more affected by edge pixels. This is due to higher MTJ current of edge pixels. Finally, the EDP of the RTS-based NCS and the typical NCS are calculated. The simulation results show 67.5% decrease in EDP for the proposed NCS.

Character recognition: The simulation results of numerical digit character recognition using trained Hamming method [39] are tabulated in Table IV. The size of training and test images in the data set are 21x12. The simulations are performed in Matlab using the behavioral models of the RTS-based NCS and the typical NCS. The NCS is simulated using a crossbar array with 10 output MTJs in order to mimic neuronal functionality for numerical digits of 0-9. For each test image, an input pattern

with 252 samples is extracted and used as the inputs of NCS. The energy consumption, the delay and the EDP of the RTS-based NCS and the typical NCS are calculated for each test image. The parameter similarity in Table IV indicates the percentage of similarity between train and test images. The higher similarity causes a higher current in corresponding MTJ that leads to shorter switching time. This trend is verified through simulation results of Table IV, where for the first test image with 100% similarity, the delay is minimum and the delay increases with similarity degradation.

The energy consumption of typical NCS is higher for test images with higher similarity that is due to a higher MTJ current. However, while using RTS circuit, the energy consumption is almost constant. This is due to the fact that in the RTS-based NCS, the MTJ currents will be cut off after switching. Hence, for the RTS-based NCS, the energy consumption is calculated through the product of the supply voltage, the MTJ current and the delay. As the similarity decreases, the MTJ current decreases and the delay increases which resulted in almost constant energy consumption.

The average energy consumption and delay improvements are calculated 71.9% and 65%, repectively, using proposed RTS circuit compared with the typical NCS. As a result, one order of magnitude (90.3%) improvement in EDP is achieved through utilizing RTS circuit.

V. CONCLUSION

The read circuit of typical NCSs is replaced with a proposed real time sensing circuit in order to improve the energy consumption and speed of such computing systems. The improvements are achieved through tracking the MTJ state and terminating its current right after MTJ switching using a proposed RTS circuit during stimulation phase. Another side-benefit of the RTS circuit is that the operation of the RTS-based NCS will be similar to real neuron functionality in the human body. The simulation results confirm that the energy consumption and overall speed of the RTS-based NCS are improved by 40% and 2.22X in comparison with the typical NCS. Finally, the effect of RTC circuit on energy delay products of the NCS in image processing applications such as character recognition and edge detection shows up to 90.3% improvement in EDP compared with the typical NCS.

REFERENCES

[1] https://www.top500.org/system/177999

[2] Europe's Human Brain Project (HBP): https://www.humanbrainproject.eu/

[3] A. Basu et. al., "Neural dynamics in reconfigurable silicon," IEEE Trans. on Biomedical Circuits and Systems, vol. 4, no. 5, pp. 311–319, Oct. 2010.

[4] S. Ramakrishnan, P. E. Hasler, and C. Gordon, "Floating gate synapses with spike-time-dependent plasticity," IEEE Trans. on Biomedical Circuits and Systems, vol. 5, no. 3, pp. 244–252, Jun. 2011.

[5] M. Sharad, D. Fan, and K. Roy, "Spin-neurons: A possible path to energy-efficient neuromorphic computers," Journal of Applied Physics, vol. 114, no. 23, p. 234906, Nov. 2013.

[6] D. Fan, Y. Shim, A. Raghunathan, and K. Roy, "STT-SNN: A spin-transfer-torque based soft-limiting non-linear neuron for low-power artificial neural networks," IEEE Trans. on Nanotechnology, vol. 14, no. 6, pp. 1013-1023, Jun. 2015.

[7] A. Sengupta, Y. Shim, and K. Roy, "Simulation studies of an all-spin artificial neural network: Emulating neural and synaptic functionalities through domain wall motion in ferromagnets," IEEE Trans. on Biomedical Circuits and Systems, vol. 10, no. 6, pp. 1152–1160, May 2016.

[8] X. Fong et. al., "Spin-transfer torque devices for logic and memory: prospects and perspectives," IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems (TCAD), vol. 35, no. 1, pp. 1-22, Jan. 2016.

[9] A. Sengupta and K. Roy, "A vision for all-spin neural networks: a device to system perspective," IEEE Trans. on Circuits and Systems–I: Regular Papers, vol. 63, no. 12, pp. 2267-2277, Dec. 2016.

[10] A. Sengupta and K. Roy, "Spin-transfer torque magnetic neuron for low power neuromorphic computing," IEEE International Joint Conference on Neural Networks (IJCNN), 2015.

[11] F. Iga et. al., "Time-resolved switching characteristic in magnetic tunnel junction with spin transfer torque write scheme," Japanese Journal of Applied Physics, vol. 51, no. 2, pp. 02BM02, Feb. 2012.

[12] T. Zheng, J. Park, M. Orshansky, and M. Erez, "Variable-energy write STT-RAM architecture with bit-wise write-completion monitoring," Symposium on Low Power Electronics and Design, Beijing China, pp. 229-234, Sep. 2013.

[13] T. Devolder et al., "Single-shot time-resolved measurements of nanosecond-scale spin-transfer induced switching: stochastic versus deterministic aspects," Physical Review Letters, vol. 100, no. 5, pp. 057206, Feb. 2008.

[14] X. Wang, Y. Zheng, H. Xi, and D. Dimitrov, "Thermal fluctuation effects on spin torque induced switching: Mean and variations," Journal of Applied Physics, vol. 103, no. 3, pp. 034507, Feb. 2008.

[15] Z. Diao et al., "Spin-transfer torque switching in magnetic tunnel junctions and spin-transfer torque random access memory," J. Phys. Condens. Matter, vol. 19, no. 16, pp. 165209, Apr. 2007.

[16] Y. Zhang, X. Wang, Y. Li, A.K. Jones, and Y. Chen, "Asymmetry of MTJ switching and its implication to STT-RAM designs," Design, Automation & Test in Europe Conference & Exhibition (DATE), Dresden, pp. 1313-1318, 2012.

[17] D. D. Tang and Y. J. Lee, Magnetic memory fundamentals and technology, New York: Cambridge University Press, 2010, pp. 122-164.

[18] A. Sengupta and K. Roy, "Spin-transfer torque magnetic neuron for low power neuromorphic computing," in Proc. Int. Joint Conf. Neural Netw. (IJCNN), pp. 1–7, 2015.

[19] P. Zhou, B. Zhao, J. Yang, and Y. Zhang, "Energy reduction for STT-RAM using early write termination," in Proc. ICCAD, pp. 264–268, 2009.

[20] X. Bi, Z. Sun, H. Li, and W. Wu, "Probabilistic design methodology to improve run-time stability and performance of STT-RAM caches," in Proc. ICCAD, pp. 88–94, 2012.

[21] H. Farkhani et al., "STT-RAM energy reduction using self-referenced differential write termination technique," IEEE Trans. on Very Large Scale Integration (VLSI), vol. 25, no. 2, pp. 476-487, Feb. 2017.

[22] R. Bishnoi, M. Ebrahimi, F. Oboril, and M. B. Tahoori, "Asynchronous asymmetrical write termination (AAWT) for a low power STTMRAM," in Proc. Design, Autom. Test Eur. Conf. Exhibit. (DATE), pp. 1–6, 2014.

[23] B. Razavi, Design of analog CMOS integrated circuits. New York: McGraw-Hill, 2001.

[24] P. R. Gray, P. J. Hurst, S. H. Lewis and R. G. Meyer, Analysis and design of analog integrated circuits. New York: Wiley publication, 2001.

[25] A. Sengupta, M. Parsa, B. Han, and K. Roy, "Probabilistic deep spiking neural systems enabled by magnetic tunnel junction," IEEE Trans. On Electron Devices, vol. 63, no. 7, pp. 2963-2970, Jul. 2016.

[26] G. D. Panagopoulos, C. Augustine, and K. Roy, "Physics-based SPICE-compatible compact model for simulating hybrid MTJ/CMOS circuits," IEEE Trans. on Electron Device, vol. 60, no. 9, pp. 2808-2814, Sep. 2013.

[27] W. Guo et al., "SPICE modeling of magnetic tunnel junctions written by spin-transfer torque," Journal of Applied Physics, vol. 43, no. 21, p. 215001-1–215001-8, Jun. 2010.

[28] M. Madec et al., "Compact modeling of magnetic tunnel junction," In Proc. 6th Int. IEEE Northeast Workshop on Circuits Syst. TAISA Conf., pp. 229–232, 2008.

[29] K. Y. Camsari, S. Ganguly, and S. Datta, "Modular Spintronics Library," https://nanohub.org/resources/17831, 2013.

[30] D. B. Strukov, G. S. Snider, D. R. Stewart, and R. S. Williams, "The missing memristor found," Nature, vol. 453, pp. 80–83, May 2008.

[31] E. Lehtonen and M. Laiho, "CNN using memristors for neighborhood connections," in Proc. Int. Workshop Cell. Nanoscale Netw. Their Appl., pp. 1–4, Feb. 2010.

[32] M. D. Pickett et al., "Switching dynamics in titanium dioxide memristive devices," Journal of Applied Physics, vol. 106, no. 7, pp. 1–6, Oct. 2009.

[33] S. Kvatinsky, E. G. Friedman, A. Kolodny, and U. C. Weiser, "TEAM: threshold adaptive memristor model," IEEE Trans. on Circuits and Systems–I: Regular Papers, vol. 60, no. 1, Jan. 2013.

[34] T. Na et al., "Offset-canceling current-sampling sense amplifier for resistive nonvolatile memory in 65 nm CMOS IEEE J. Solid-State Circuits, vol. 52, no. 2, pp. 496-504, Feb. 2017.

[35] M.-F. Chang et al., "An offset-tolerant fast-random-read current-sampling-based sense amplifier for small-cell-current nonvolatile memory," IEEE J. Solid-State Circuits, vol. 48, no. 3, pp. 864–877, Mar. 2013.

[36] M. Sharad, C. Augustine, G. Panagopoulos, and K. Roy, "Ultra low energy analog image processing using spin neurons," IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH), pp. 211-217, Jul. 2012.

[37] M. Sharad, C. Augustine, and K. Roy, "Boolean and non-Boolean computation with spin devices," IEEE International Electron Devices Meeting (IEDM), pp. 11.6.1-4, Dec. 2012.

[38] W. K. Pratt, digital image processing, fourth edition, Wiley publication, 2007.

[39] A. Namane, A. Guessoum, E.H. Soubari, and P. Meyrueis, "CSM neural network for degraded printed character optical recognition," Journal of Visual Communication and Image Representation, Elsevier, vol. 25, no. 5, pp. 1171-1186, Jul. 2014.

Hooman Farkhani received the B.Sc. degree in electrical engineering from Kashan University, Kashan, Iran, in 2004, and the M.Sc. and Ph.D. degrees in electronic engineering from Ferdowsi university of Mashhad, Mashhad, Iran, in 2008 and 2014, respectively. He was a research assistant in Integrated Circuit Design (ICD) laboratory in Ferdowsi university of Mashhad. He worked at Aarhus University in Denmark for four months on spin-transfer torque random access memory (STT-RAM). He was an assistant professor at the department of electrical engineering, Najafabad branch, Azad University, Isfahan, Iran from 2015 to 2017. He is currently an MSCA-IF postdoc at Aarhus University in Denmark where he collaborates with ICE-LAB on designing low power neuromorphic computing systems. His other fields of interest are low power and low voltage STT-RAM design, SRAM design and fully digital ADCs.

Mohammad Tohidi received the B.Sc. and M.Sc. degrees in electronics engineering from Urmia University, Iran, in 2010 and 2013, respectively (with honor). He is currently pursuing the Ph.D. degree in electronics engineering with the ICE-Lab, Aarhus University, Aarhus, Denmark. During his

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JETCAS.2018.2813389, IEEE Journal on Emerging and Selected Topics in Circuits and Systems

12

Ph.D. studies, he is working on the design of low power device for seizure-detection, biomedical applications. His current research interests include spintronics, memory cells and the design of mixed-signal integrated circuits, data-converters and low-power biomedical circuits and systems.

Sadaf Farkhani received the B.S. and the M.S. degrees in Telecommunication Engineering from the Islamic Azad University of Najafabad, Isfahan, Iran in 2014 and 2018. Her research interests lie in the broad area of neural network, sparse representation theory, digital watermarking and alphanumeric character recognition.

Jens Madsen completed his Master degree in Electrical Engineering in 1990 at DTU in electronics and in 1993 achieved his PhD degree in electrical engineering also from DTU on high-speed IC design. Since 2010 Jens has been at Engineering College of Aarhus (IHA) and Aarhus School of Engineering (ASE), Aarhus University, and from 2011 as Professor (Docent) and Head of the Electrical and Computer Engineering department. Prior to this position he has 20 years of R&D experience in Electronic and Semiconductor academia and industry target Data and Telecommunications equipment with more than 10 years of management experience within R&D organizations. He served latest as V.P. of engineering at Enigma Semiconductor, Inc., and before that as Director of Switch Product Development at Vitesse Semiconductor Corp., where he established the Danish Design center in 1998. Before joining Vitesse, he worked as design engineer in the ASIC group at DSC Communications and before that as an assistant research professor at the Broadband Telecommunication Institute at DTU. In 1991, he was a research visitor at the ECE department at UCSB, California. Currently, his main research interests are within integrated Analogue/Digital mixed signal designs and SoC systems including low power design techniques and methodologies for future sensor circuits, systems and networks, e.g. to be applied in Biomedical applications.

Farshad Moradi (M'11, SM'17) received his B.Sc. and M.Sc. in electrical engineering from Isfahan University of Technology and Ferdowsi University of Mashhad, respectively. He received his Ph.D. degree in Electrical Engineering from University of Oslo, Norway, in 2011. From 2009 to 2010, he visited the Nanoelectronics Research Laboratory at Purdue University, IN, USA. He is currently an Associate Professor with the Integrated Circuit and Electronics Laboratory, at the department of Engineering, Aarhus University, Denmark. He is an associate editor of Integration, the VLSI and VLSI Journal. He is the author/co-author of more than70 Journal and Conference papers. His current research interests include ultralow-power digital/memory circuit/device design.