

A Monolithic Stochastic Computing Architecture for Energy and Area Efficient Arithmetic

Harikrishnan Ravichandran

Pennsylvania State University

Yikai Zheng

Pennsylvania State University

Thomas Schranghamer

Pennsylvania State University

Nicholas Trainor

Pennsylvania State University

Joan Redwing

Saptarshi Das (✉ sud70@psu.edu)

Pennsylvania State University <https://orcid.org/0000-0002-0188-945X>

Article

Keywords:

Posted Date: January 11th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1196783/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

A Monolithic Stochastic Computing Architecture for Energy and Area Efficient Arithmetic

*Harikrishnan Ravichandran¹, Yikai Zheng¹, Thomas F Schranghamer¹, Nicholas Trainor^{2,3},
Joan M Redwing^{2,3}, and Saptarshi Das^{2,3,4,5,*}*

¹*Engineering Science and Mechanics, Penn State University, University Park, PA 16802, USA*

²*Materials Science and Engineering, Penn State University, University Park, PA 16802, USA*

³*Materials Research Institute, Penn State University, University Park, PA 16802, USA*

⁴*Electrical Engineering and Computer Science, Penn State University, University Park, PA 16802, USA*

Abstract: As the energy and hardware investments necessary for conventional high-precision digital computing continues to explode in the emerging era of artificial intelligence, deep learning, and Big-data [1-4], a change in paradigm that can trade precision for energy and resource efficiency is being sought for many computing applications. Stochastic computing (SC) is an attractive alternative since unlike digital computers, which require many logic gates and a high transistor volume to perform basic arithmetic operations such as addition, subtraction, multiplication, sorting etc., SC can implement the same using simple logic gates [5, 6]. While it is possible to accelerate SC using traditional silicon complementary metal oxide semiconductor (CMOS) [7, 8] technology, the need for extensive hardware investment to generate stochastic bits (s-bit), the fundamental computing primitive for SC, makes it less attractive. Memristor [9-11] and spin-based devices [12-15] offer natural randomness but depend on hybrid designs involving CMOS peripherals for accelerating SC, which increases area and energy burden. Here we overcome the limitations of existing and

emerging technologies and experimentally demonstrate a standalone SC architecture embedded in memory based on two-dimensional (2D) memtransistors. Our monolithic and non-von Neumann SC architecture consumes a miniscule amount of energy < 1 nano Joules for s-bit generation and to perform arithmetic operations and occupy small hardware footprint highlighting the benefits of SC.

Introduction:

The aggressive downscaling of feature sizes in silicon based complementary metal-oxide-semiconductor (CMOS) technology over the past five decades has led to an exponential growth in the computing power of modern-day computers. Today, computers can fly jets, control industrial processes, and solve optimization problems. In fact, computers can also beat professional players in the game of ‘Go’ and predict complex structures of proteins thanks to the remarkable progress in the field of artificial intelligence (AI). The ongoing revolution in AI is directly linked to the unfathomable data processing power by computers enabling implementation of deep learning and various other sophisticated machine learning algorithms [1, 2]. However, there is significant infrastructure cost associated with advanced AI and computing systems. For example, any mathematical algorithm implemented using hardware requires arithmetic operations such as addition, subtraction, multiplication, sorting, etc., which are executed using logic circuits consisting of hundreds of transistors that occupy large area and consume significant amount of energy. Furthermore, the von Neumann architecture necessitate frequent data shuttling between the arithmetic and the memory units to run algorithms adding area and energy overheads. Needless to say, these challenges are aggravated as the data size grows exponentially for both AI and no-AI platforms [3, 4]. Therefore, a new paradigm that can drastically reduce the area and energy cost of

arithmetic operations can not only benefit cloud computing using supercomputers but also enable edge computing in resource-constrained internet of things (IoT) devices.

Stochastic computing (SC) is an attractive alternative, where arithmetic operations can be performed using simple logic gates yielding high energy and area efficiency [5, 6]. For example, a simple two-bit multiplication in a conventional CMOS based full adder circuit requires 78 transistors whereas a SC unit can execute the same operation using a single *AND* gate. Similarly, stochastic addition and subtraction can be performed using multiplexer (*MUX*) and *XOR* gates, respectively. The key difference is that unlike classical computing system which represents information in the form of binary logic ('1's and '0's), SC encodes information through stochastic bit (s-bit) streams that are interpreted as probabilities that fall in the interval [0,1]. For instance, the bit-stream $A = \{1\ 0\ 1\ 1\ 0\ 1\ 0\ 0\}$ encodes the value $p_A = 0.5$ since there are four 1's present within the bit-stream of length 8-bit. An attractive feature of SC is its resilience to error tolerance [16] since there is no distinction between the most- and the least significant bits or in other words, all s-bits carry equal weight. While promising, the application of SC has largely been limited to specialized domains such as image and audio processing where a finite amount of error or loss in precision is acceptable [17, 18]. Such limitation primarily stem from the requirement of having a much longer bit-stream for more accurate probability estimation that leads to a corresponding increase in the computation time and energy. Despite the shortcomings, SC is becoming popular for many AI applications, which deal with large volumes of audio-visual information. Note that the idea of SC is also rooted in bio-inspired computing since brain can process information in the presence of noise and can learn, adapt, and make right decisions to ensure the survival of the species at the cost of miniscule energy expenditure.

The concept of SC is well known and extensively studied. CMOS [7, 8], memristor [9-11] and spintronics [12-14] based SC architectures have already been demonstrated in the past. However, CMOS-based SC architectures require several hundred transistors to generate s-bits, which limits its area and energy efficiency [19]. Stochastic switching in memristors offer an excellent mechanism to generate fast and random bits [20] with the added benefits of high integration density since memristors can be scaled down to sub 10 nm. However, memristor-based SC architectures still require CMOS peripherals to control the probability of switching for the conversion of random bits into s-bits and for subsequent logic operations using those s-bits, which can ultimately limit the area and energy efficiency. Recently, spin-based magnetic random access memory (MRAM) devices [21] and spin-orbit torque magnetic tunnel junctions (SOT-MTJ)[22] have shown immense potential for SC since the probability of spin-flip can be controlled by externally driven current allowing seamless generation of s-bits. In addition spin-based devices offer high switching speed, a simpler structure, high throughput, and better area and energy efficiency and are therefore, fundamentally superior in performance to CMOS-based alternatives [23]. However, environmental, and electrical fluctuations can interfere and impact the spin-flip probability necessitating additional CMOS-based peripheral circuits to remove the bias. Although, recent demonstration of integer factorization [15] using spin-based MRAM devices is a milestone achievement, the SC architecture utilized for such demonstration involves extensive CMOS peripherals since two-terminal MRAMs suffer from similar limitations like the memristors.

Here, we overcome the above-mentioned limitations by introducing a standalone SC architecture embedded in memory, which is based on two dimensional (2D) memtransistors. Memtransistors are programmable field effect transistors (FETs) made from ultra-thin body semiconducting

channel material such as monolayer MoS₂ allowing aggressive channel length scaling owing to superior gate electrostatics. Our main contributions are 1) the realization of an area and energy efficient six-transistor (6T) s-bit generator circuit that exploits the inherent stochasticity in the carrier trapping and detrapping phenomena in the gate insulator of the 2D memtransistors and combines it with an inverting amplifier and a programmable thresholding inverter to obtain s-bits and 2) integration of s-bit generators with 2D memtransistor based logic gates such as *AND*, *MUX*, *XOR*, and *OR* gates to demonstrate arithmetic operations such as addition, subtraction, multiplication, and sorting.

Fabrication and characterization of 2D memtransistors

Fig. 1a shows the optical image of 2D memtransistors based hardware platform for the acceleration of the SC architecture and **Fig. 1b** shows the optical image and corresponding 3D schematic of a representative 2D memtransistor based on monolayer MoS₂, which are locally back-gated using a stack comprising of atomic layer deposition (ALD) grown 50 nm Al₂O₃ on sputter deposited 40/30 nm Pt/TiN. All back-gate islands were placed on a commercially purchased SiO₂/p⁺⁺-Si substrate. As we will discuss later, the stochastic conductance fluctuation in monolayer MoS₂ and analog and non-volatile programming capability offered by the Al₂O₃/Pt/TiN gate stack are central to our non-von Neumann SC architecture. Our monolayer MoS₂ was grown over large area via metal organic chemical vapor deposition (MOCVD) technique on sapphire substrate [24, 25] and subsequently transferred from the growth substrate to the SiO₂/p⁺⁺-Si substrate with predefined islands of Al₂O₃/Pt/TiN for 2D memtransistor fabrication. Details on monolayer MoS₂ synthesis, film transfer, and fabrication of the local back-gate gate islands, MoS₂ memtransistors, and SC architecture can be found in the *Methods* section.

2D memtransistors for hardware acceleration of stochastic computing (SC)

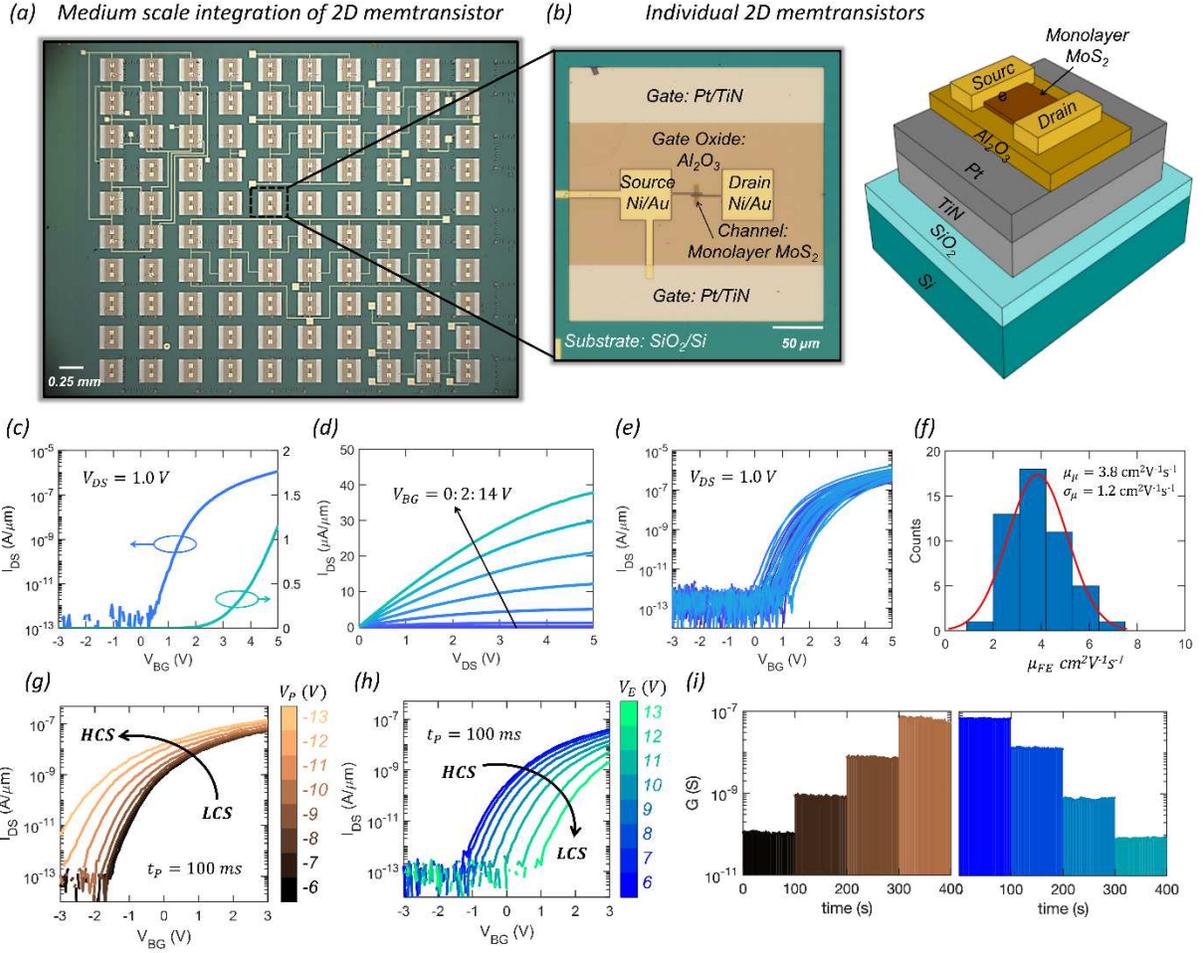


Figure 1. Fabrication and characterization of 2D memtransistors for acceleration of stochastic computing (SC). a) Optical image of a representative 2D memtransistors based medium scale integrated circuit for the hardware acceleration of SC. b) Optical image and corresponding 3D schematic of a representative 2D memtransistor based on monolayer MoS₂, which are locally back-gated using a stack comprising of atomic layer deposition (ALD) grown 50 nm Al₂O₃ on sputter deposited 40/30 nm Pt/TiN. All back-gate islands were fabricated on SiO₂/p⁺⁺-Si substrate. c) Transfer characteristics, i.e. source to drain current (I_{DS}) versus local back-gate voltage (V_{BG}) measured using source to drain bias, $V_{DS} = 1$ V for a representative MoS₂ memtransistor with channel length, $L = 1$ μm , and channel width, $W = 5$ μm in linear and logarithmic scale. d) Output characteristics, i.e. I_{DS} versus V_{DS} for different V_{BG} for the same MoS₂ memtransistor. e) Device-to-device variation in the transfer characteristics and f) corresponding histogram of extracted field effect mobility (μ_{FE}) distribution across 50 memtransistors. Analog g) programming and h) erase capability of 2D memtransistor when subjected to negative “Write” (V_P) and positive “Erase” (V_E) voltage pulses of different amplitudes ranging from 6 V to 13 V applied to the local back-gate electrode, each for a duration of $\tau_{P/E} = 100$ ms. i) Non-volatile retention for 4 representative programmed and erased states for 100 seconds.

Fig. 1c shows the transfer characteristics, i.e. source to drain current (I_{DS}) versus local back-gate voltage (V_{BG}) measured using source to drain bias, $V_{DS} = 1$ V, in linear and logarithmic scale for a representative MoS₂ memtransistor with channel length, $L = 1$ μm , and channel width, $W = 5$ μm .

As expected, n-type transport is observed in MoS₂, which is attributed to the pinning of the metal Fermi level near the conduction band [26-28]. Nevertheless, MoS₂ memtransistor exhibits excellent electrostatic gate control with current on/off ratio ($r_{\text{ON/OFF}}$) $\sim 10^6$, subthreshold slope (SS) ~ 370 mV/decade averaged over 4 orders of magnitude change in I_{DS} , minimal gate hysteresis when measured in air, and low gate leakage current. The threshold voltage (V_{TH}) was found to be ~ 2 V extracted at iso-current of 100 nA/ μm and the electron field effect mobility (μ_{FE}) extracted from the peak trans-conductance was found to be ~ 5 cm²/V-s. **Fig. 1d** shows the output characteristics, i.e. I_{DS} versus V_{DS} for different V_{BG} for the same MoS₂ memtransistor. The on current (I_{ON}) reached as high as ~ 40 $\mu\text{A}/\mu\text{m}$ for an inversion carrier density of $\sim 1.4 \times 10^{12}/\text{cm}^2$ at $V_{\text{DS}} = 5$ V. **Fig. 1e** shows the device-to-device variation in the transfer characteristics across 50 2D memtransistors and **Fig. 1f** show the corresponding histogram of extracted μ_{FE} with mean of ~ 3.8 cm²V⁻¹s⁻¹ and standard deviation of 1.2 cm²V⁻¹s⁻¹. These results indicate relatively high quality and uniform monolayer film growth using MOCVD, relatively damage-free film transfer, and clean memtransistor fabrication processes.

Finally, **Fig. 1g-i**, respectively, show the analog programming, erase, and non-volatile retention capability of our 2D memtransistor. When the 2D memtransistor is subjected to negative “Write” (V_p) and positive “Erase” (V_E) voltage pulses of different amplitudes ranging from 6 V to 13 V applied to the local back-gate electrode, each for a duration of $\tau_{p/E} = 100$ ms, the transfer characteristics show shift in V_{TH} , which can be attributed to charge trapping/detrapping at and near the MoS₂/Al₂O₃ interface. Negative shift in the in the transfer characteristics with increasing magnitude of V_p and positive shift with increasing magnitude of V_E are indicative of electron trapping and de-trapping in the local back-gate stack, respectively. Interestingly, the trapping and

de-trapping processes were found to be non-volatile as shown in **Fig. 1i** for 4 representative programmed and erased states for 100 seconds. We also found that the device is capable of retaining programmed conductance states for more than 10 hours. While it is desirable to improve the memory retention, for the application sought for in this work, i.e. SC, the memory retention was found to be adequate.

Programming stochasticity in 2D memtransistor and s-bit generation

Generation of high-quality random bits is a pre-requisite for reducing computational inaccuracies at the output of any stochastic operation. Here, we exploit the inherent stochasticity in the carrier trapping and detrapping phenomena in the gate oxide of our 2D memtransistor as the source of true randomness. **Fig. 2a** shows the transfer characteristics of a representative MoS₂ memtransistor, measured each time after the application of $V_P = -10$ V and $V_E = 10$ V each for $\tau_s = 100$ ms, for a total of 100 cycles and **Fig. 2b** shows the distribution of G_{MT} measured using $V_{BG} = 0$ V. Clearly, the cycle-to-cycle variability in post-programmed and post-reset G_{MT} follow Gaussian random distributions. While programming stochasticity is detrimental for conventional computing, it offers unique opportunity for SC.

Next, in order to translate the conductance fluctuation into s-bits, we deploy a module consisting of six memtransistors ($MT1, MT2, MT3, MT4, MT5,$ and $MT6$) as shown using the optical image and corresponding circuit diagram in **Fig. 2c-d**, respectively. The voltage waveforms applied to the nodes, $N1, N2$, i.e., V_{N1}, V_{N2} , respectively, are shown in **Fig. 2e**. Note that during each clock cycle (τ_{clk}), V_{N1} toggles between 0 V, 0 V, and $V_{DD} = 2$ V and V_{N2} toggles between $V_P = -7$ V, $V_E = 10$ V, and $V_R = 1$ V. Voltages applied to nodes, $N3,$ and $N4$, i.e., $V_{N3},$ and V_{N4} are held constant

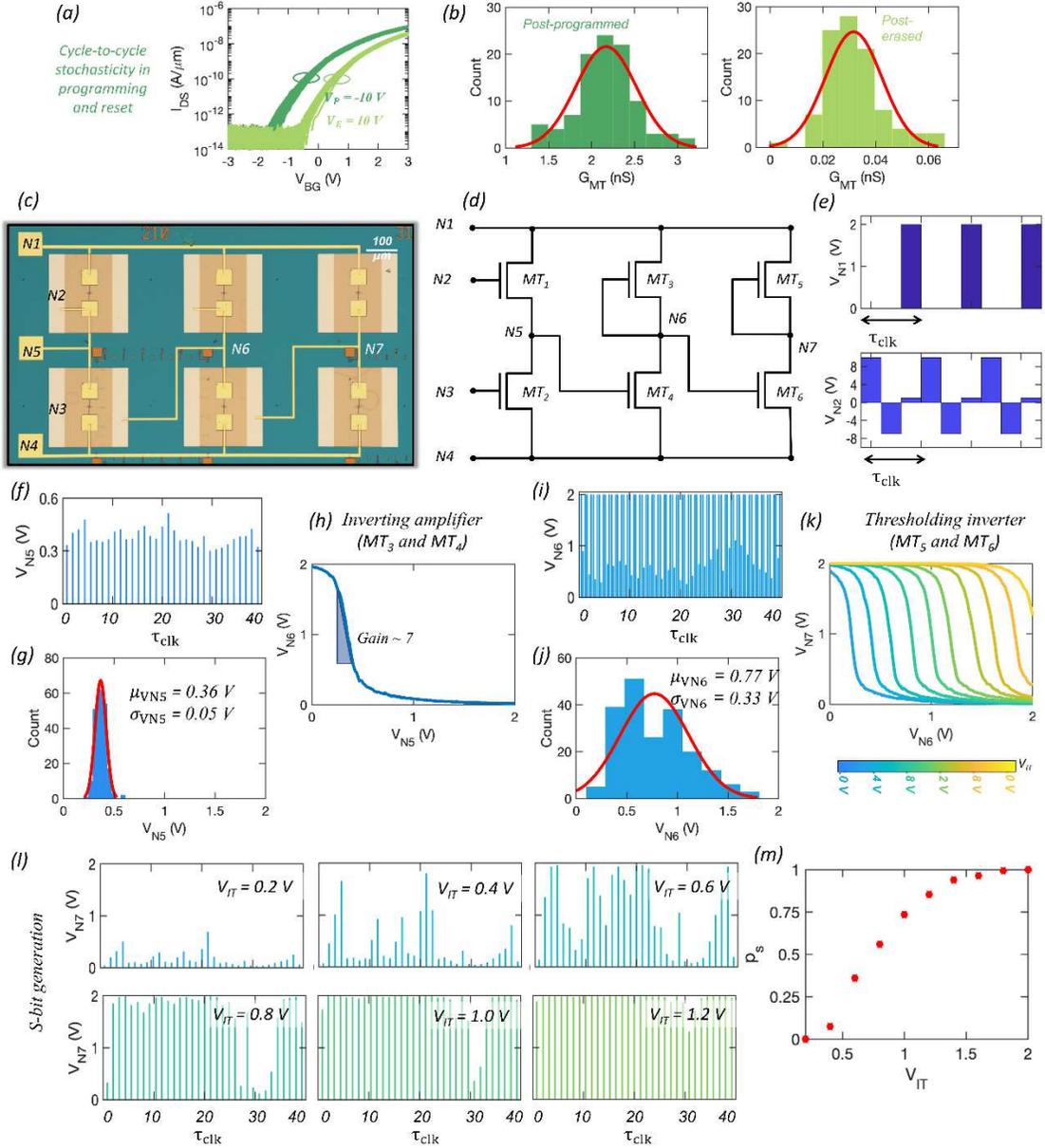


Figure 2. Programming stochasticity in 2D memtransistor and s-bit generation. a) Transfer characteristics of a representative 2D memtransistor, measured each time after the application of $V_P = -10$ V and $V_E = 10$ V each for $\tau_s = 100$ ms, for a total of 100 cycles. b) Distribution of post-programmed and post-erased conductance (G_{MT}) measured using $V_{BG} = 0$ V follow Gaussian random distributions, which forms the basis for high-quality randomness for s-bit generation. c) optical image and d) corresponding circuit diagram for the proposed s-bit generator consisting of six memtransistors (MT_1 , MT_2 , MT_3 , MT_4 , MT_5 , MT_6). e) Voltage waveforms applied to the nodes, N1, N2, i.e., V_{N1} , V_{N2} . During each clock cycle (τ_{clk}), V_{N1} toggles between 0 V, 0 V, and $V_{DD} = 2$ V and V_{N2} toggles between $V_P = -7$ V, $V_E = 10$ V, and $V_R = 1$ V. Voltages applied to nodes, N3, and N4, i.e., V_{N3} , and V_{N4} are held constant at 1V and 0 V, respectively. f) Voltage readout at node, N5, i.e., V_{N5} . Since the memtransistors, MT_1 and MT_2 are connected in series and G_{MT1} fluctuates due to programming and reset every τ_{clk} , so does V_{N5} . g) Distribution of V_{N5} over 200 τ_{clk} follows a random Gaussian distribution with mean, $\mu_{VN5} = 0.36$ V and standard deviation, $\sigma_{VN5} = 0.05$ V. h) Output, V_{N6} , of an inverting amplifier constructed using MT_3 and MT_4 as a function of the input, V_{N5} with a gain of ~ 7 . i) V_{N6} corresponding to V_{N5} shown in (f). j) Distribution of V_{N6} which follows a random Gaussian distribution with mean, $\mu_{VN6} = 0.77$ V and an increased standard deviation of $\sigma_{VN6} = 0.33$ V. k) Output, V_{N7} , of a thresholding inverter constructed using MT_5 and MT_6 as a function of the input, V_{N6} for different inversion threshold, V_{IT} . l) V_{N7} corresponding to V_{N6} shown in (i) for different V_{IT} . m) Probability of obtaining '1' in the bit stream (p_s) as a function of V_{IT} . This clearly shows the ability of the proposed circuit to transform the cycle-to-cycle conductance fluctuations in 2D memtransistor into s-bits with reconfigurable p_s that lie between $[0, 1]$.

at 1V and 0 V, respectively. This is done to program and reset $MT1$ and then readout the voltage at node, $N5$, i.e., V_{N5} during each τ_{clk} . Since the memtransistors, $MT1$ and $MT2$ are connected in series, V_{N5} is determined by their corresponding conductance values, i.e., G_{MT1} and G_{MT2} . As G_{MT1} fluctuates from cycle to cycle, so does V_{N5} as shown in **Fig. 2f**. **Fig. 2g** shows the histogram of V_{N5} , which follows a random Gaussian distribution with mean, $\mu_{VN5} = 0.36$ V and standard deviation, $\sigma_{VN5} = 0.05$ V.

Next the Gaussian distribution is broadened by using an inverting amplifier constructed using $MT3$ and $MT4$. Note that the local back-gate of $MT3$ is shorted to its source at node, N_6 . This ensures that $MT3$ operates as a depletion mode (normally on) transistor or as a load resistor. **Fig. 2h** shows the output, V_{N6} , as a function of the input, V_{N5} . The slope of the curve is referred to as the gain of the amplifier, and higher the gain wider is the broadening of the Gaussian. We achieved a gain of ~ 7 , which was sufficient for the hardware acceleration of SC. The gain can be increased by cascading multiple amplifiers; however it adds area and energy overhead. **Fig. 2i** shows V_{N6} corresponding to V_{N5} in **Fig. 2f** and **Fig. 2j** shows the histogram of V_{N6} which follows a random Gaussian distribution with mean, $\mu_{VN6} = 0.77$ V and an increased standard deviation of $\sigma_{VN6} = 0.33$ V.

To transform the analog fluctuations seen in V_{N6} into s-bits, we use a thresholding inverter constructed using $MT5$ and $MT6$. **Fig. 2k** shows the output, V_{N7} , as a function of the input, V_{N6} for different inversion threshold, V_{IT} , which is defined as the magnitude of V_{N6} at which V_{N7} reaches $V_{DD}/2$. Note that the programmability of V_{IT} is a critical feature that distinguishes 2D memtransistors based inverters from conventional CMOS-based inverters and allows us to

seamlessly obtain the s-bits. **Fig. 2l** shows V_{N7} corresponding to V_{N6} in **Fig. 2i** for different V_{IT} and **Fig. 2m** shows the probability of obtaining ‘1’ in the bit stream (p_s) as a function of V_{IT} . As expected, if V_{IT} is too low, then almost all V_{N6} values corresponding to the Gaussian distribution in **Fig. 2j** translate into $V_{N7} \approx 0$ V, which is reflected as near zero p_s . Similarly, if V_{IT} is too high, then almost all V_{N6} values translate into $V_{N7} \approx 2$ V leading to $p_s = 1$. Between these two extremes, p_s increases monotonically with V_{IT} . This clearly shows that we are able to convert the cycle-to-cycle random conductance fluctuations in 2D memtransistor into s-bits with reconfigurable p_s that lie between [0,1] using the circuit based on 6 2D memtransistors.

The average energy expenditure for s-bit generation (E_{s-bit}) was calculated using Eq. 3.

$$E_{s-bit} = \frac{1}{2} C_G [V_P^2 + V_E^2 + V_{DD}^2]; C_G = \frac{\epsilon_0 \epsilon_{ox} W L}{t_{ox}} \quad [3]$$

In Eq. 3, C_G is the gate capacitance, $\epsilon_0 = 8.85 \times 10^{-12}$ F/m is the vacuum permittivity, $\epsilon_{ox} = 10$, and $t_{ox} = 50$ nm are, respectively, the relative permittivity and thickness of Al_2O_3 . We found that $E_{s-bit} < 2$ pJ/clock-cycle, which supports our claim on energy efficient s-bit generation. Also note that each memtransistor has an active device area is $\sim 5 \mu m^2$ excluding the large contact pads. Therefore, the active footprint of the s-bit generator is only $30 \mu m^2$. Given that monolayer 2D materials offer aggressive dimensional scalability, it is possible to reduce the active footprint significantly without compromising the quality of the s-bits.

Stochastic arithmetic modules

Multiplication: Stochastic multiplication can be accomplished using a simple *AND* gate as shown in **Fig. 3a**. The stochastic output, $C(p_C)$, of an *AND* gate with two stochastic input variables, $A(p_A)$ and $B(p_B)$, is given by Eq. 4.

$$C = AB \quad [4a]$$

$$p_C = p_A p_B \quad [4b]$$

Where, p_A , p_B , and p_C , are the probabilities associated with the random variables, A , B , and C respectively. Note that Eq. 4 is valid if and only if the random variables, A and B , are mutually independent or uncorrelated. **Fig. 3b-c**, respectively, show the optical image and corresponding circuit configuration of a stochastic multiplier consisting of 2 s-bit generator and an *AND* gate with a total of 15 memtransistors. The *AND* gate consists of 3 memtransistors, $MT7$, $MT8$, and $MT9$. Inputs, A and B , are applied to the local back-gates of $MT7$ and $MT8$, which are connected in series with $MT9$ at node $N8$. The source and gate terminals of $MT9$ are shorted and connected to the ground. As such, $MT9$ operates as a load resistor. The output, C , of the *AND* gate is obtained at node $N8$. **Fig. 3d** shows the representative stochastic bit-streams for the random variables, $A(p_A = 0.6)$ and $B(p_B = 0.74)$ obtained from their respective s-bit generators by programming V_{IT} and the corresponding output bit-stream for C with $p_C = 0.46$. **Fig. 3e** shows the colormaps of percentage errors (ϵ) for multiplication following Eq. 5 obtained for different combinations of p_A and p_B . We have used bit-streams of length 200-bit to evaluate the corresponding probability values.

$$\epsilon = \left| 1 - \frac{(p_C)_{obtained}}{(p_C)_{expected}} \right| \times 100\% \quad [5]$$

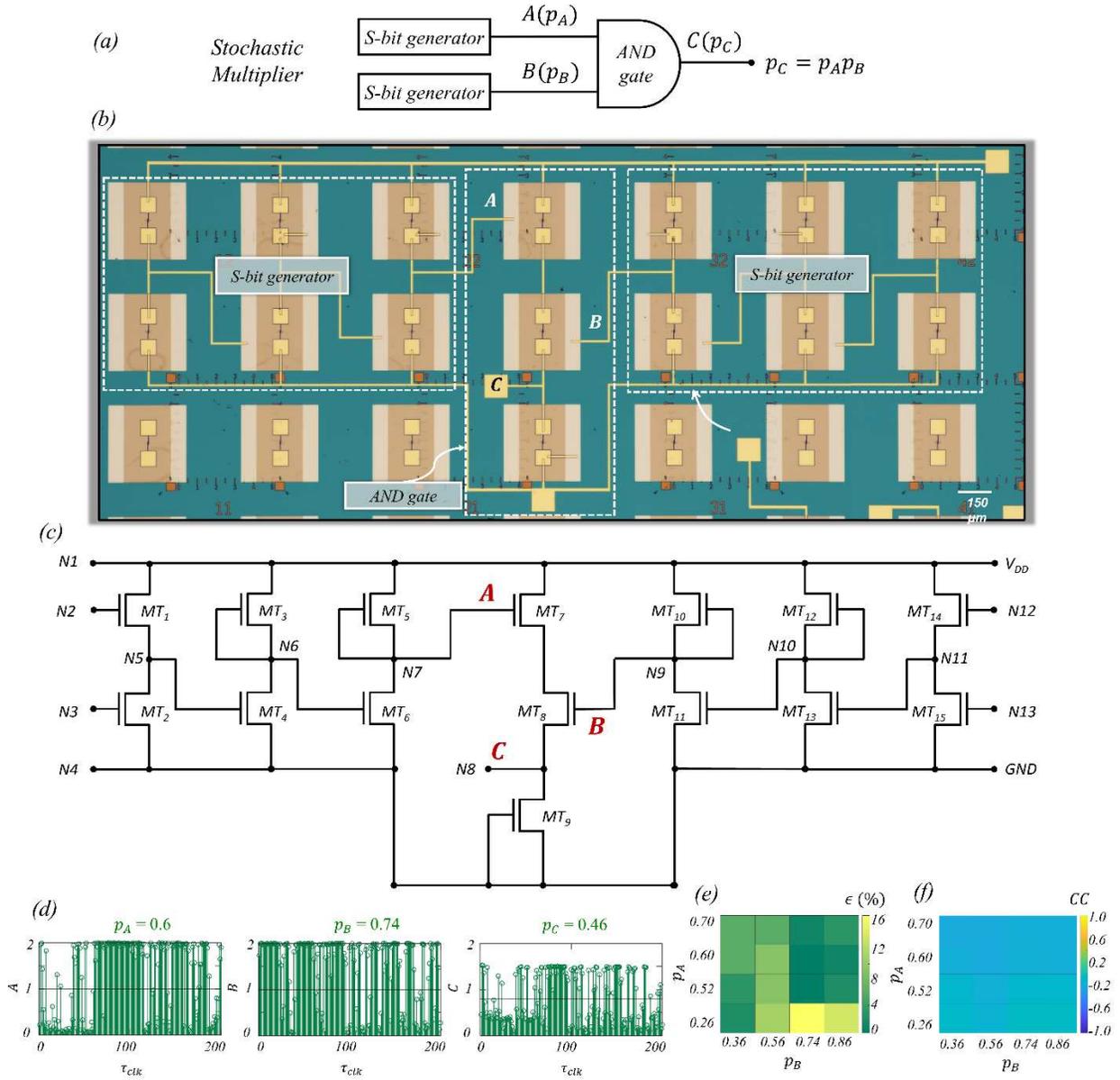


Figure 3. Stochastic multiplier. a) Schematic, b) optical image, and c) corresponding circuit configuration of a stochastic multiplier consisting of 2 s-bit generator and one AND gate with a total of 15 memtransistors. d) Representative stochastic bit-streams for the random variables, $A(p_A)$ and $B(p_B)$ obtained from their respective s-bit generators and the corresponding output bit-stream for $C(p_C)$. Colormaps of e) percentage errors (ϵ) for multiplication and f) corresponding correlation coefficient (CC) for different combinations of p_A and p_B . Lower values of ϵ is a direct consequence of near ideal CC values close to zero indicating mutual independence of A and B , which is critical for accurate multiplication. We have used bit-streams of length 200-bit to evaluate the probability values associated with the random variable.

In Eq. 5, $(p_C)_{obtained}$, and $(p_C)_{expected}$ are the experimentally obtained and theoretically predicted output of the stochastic computation following Eq. 4. As mentioned earlier, to obtain accurate multiplication product, A and B must be mutually independent. **Fig. 3f** shows the

colormap of correlation coefficient (CC) between the s-bit streams used as A and B . Low CC values close to zero confirm mutual independence of A and B , which translate into accurate multiplication results obtained in **Fig. 3e**. Clearly, our 15 memtransistor circuit is able to perform stochastic multiplication with high accuracy. Note that the accuracy can be increased by increasing the length of s-bit streams at the expense of longer computation time since one s-bit is generated every τ_{clk} . The average energy expenditure for the multiplication operation is ~ 0.8 nJ, when 200 τ_{clk} are used. Certainly, the energy expense can be reduced by reducing the length of the s-bit streams at the cost of reduced precision.

Addition: Stochastic addition operation can be accomplished using a MUX as shown in **Fig. 4a**. The stochastic output, $C(p_C)$, of a MUX with two stochastic input variables, $A(p_A)$ and $B(p_B)$, and a stochastic select line, $S(p_S)$ is given by Eq. 6.

$$C = SA + S^c B \quad [6a]$$

$$p_C = p_S p_A + (1 - p_S) p_B \quad [6b]$$

$$p_C = 0.5(p_A + p_B); \text{ if } p_S = 0.5 \quad [6c]$$

Clearly, for $p_S = 0.5$, one can achieve scaled addition. **Fig. 4b-c**, respectively, show the optical image and corresponding circuit configuration of a stochastic adder consisting of 3 s-bit generator modules and one 2×1 MUX with a total of 22 memtransistors. The 2×1 MUX consists of 4 memtransistors, $MT19$, $MT20$, $MT21$, and $MT22$. Note that, memtransistors, $MT19$ and $MT20$ form a NOT gate with stochastic variable S as the input and S^c as the output. S and S^c are applied to the local back-gates of $MT21$ and $MT22$, respectively, which are connected in series at node $N19$. The stochastic variable, A is connected to the source terminal of $MT21$ at node $N12$, whereas, the stochastic variable, B is connected to the drain terminal of $MT22$ at node $N13$. The

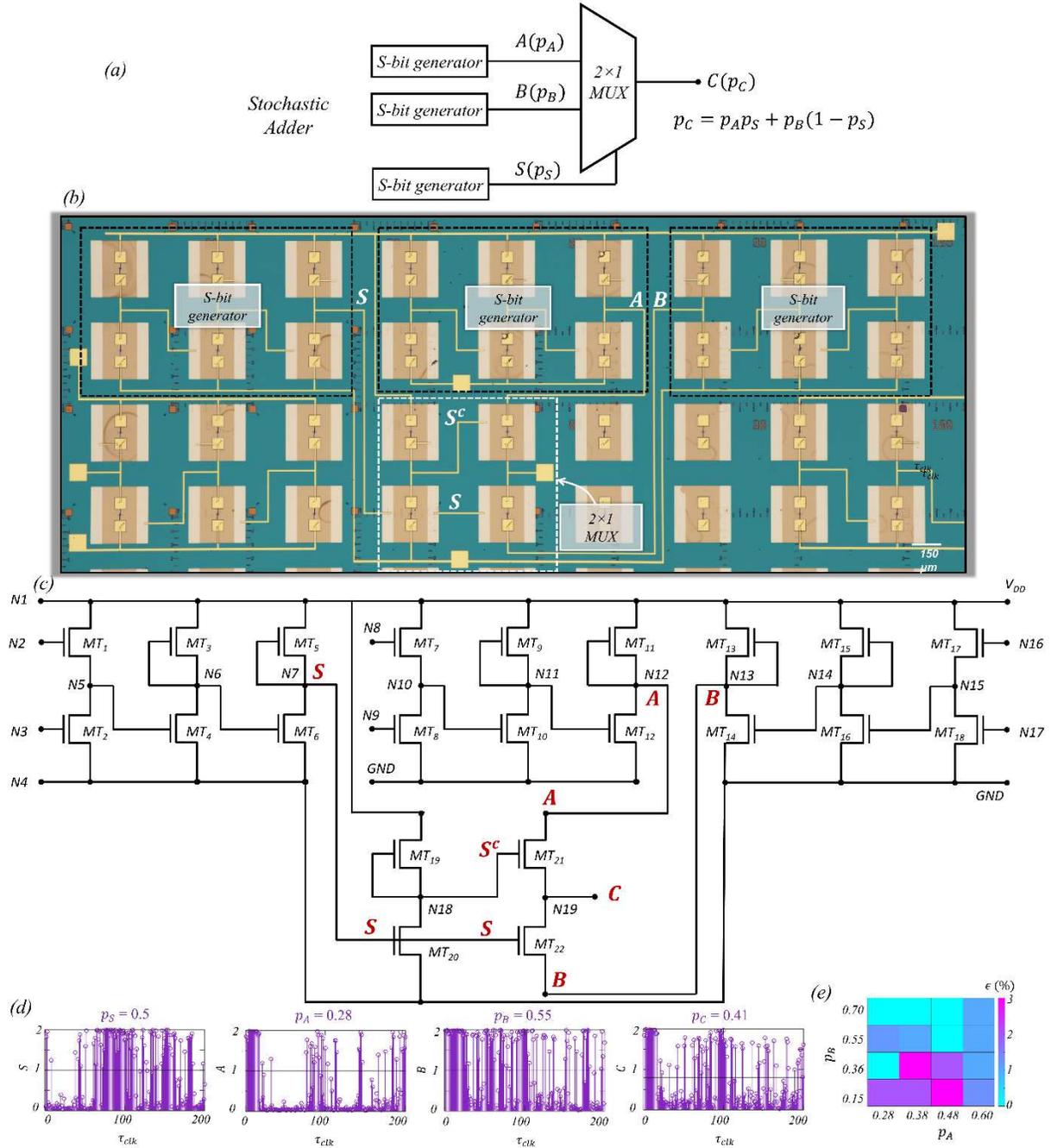


Figure 4. Stochastic adder. a) Schematic, b) optical image, and c) corresponding circuit configuration of a stochastic adder consisting of 3 s-bit generator and one 2×1 MUX gate with a total of 22 memtransistors. d) Representative stochastic bit-streams for the random variables $S(p_S)$, $A(p_A)$, and $B(p_B)$ obtained from their respective s-bit generation modules at nodes N7, N12, and N13 and the corresponding output bit-stream for $C(p_C)$. e) Colormaps of percentage errors (ϵ) for scaled addition for different combinations of p_A , p_B , for $p_S \approx 0.5$.

output of the MUX, i.e., C is obtained at node N19. **Fig. 4d** shows the representative stochastic bit-streams for the random variables $S(p_S = 0.5)$, $A(p_A = 0.28)$, and $B(p_B = 0.55)$ obtained

from their respective s-bit generation modules at nodes $N7$, $N12$, and $N13$ and the corresponding output bit-stream for C with $p_C = 0.41$. **Fig. 4e** shows the colormaps of percentage errors (ϵ) for scaled addition (Eq. 5) for different combinations of p_A , p_B , for $p_S \approx 0.5$. Clearly, our 22 memtransistor module is able to perform stochastic addition with high accuracy. The average energy expenditure for the scaled addition operation is ~ 1.2 nJ.

Subtraction: While the circuits used for stochastic multiplication and addition require the stochastic inputs to be independent or uncorrelated to achieve accurate results, stochastic subtraction benefits greatly from the correlation between the stochastic inputs. In fact, Alaghi and Hayes [29] have shown that correlated inputs can drastically alter the functionality of a stochastic circuit thereby simplifying the hardware acceleration of specific arithmetic operations. For example, if a *XOR* gate (**Fig. 5a**) is implemented using two uncorrelated stochastic input variables, $A(p_A)$ and $B(p_B)$, the stochastic output, $C(p_C)$ will be given by Eq. 8.

$$C = AB^c + A^cB \quad [8a]$$

$$p_C = p_A(1 - p_B) + p_A(1 - p_B) \quad [8b]$$

However, when A and B are highly correlated, it implements absolute-valued subtraction following Eq. 9.

$$p_C = |p_A - p_B| \quad [9]$$

As an example, if $A = 01110110$ and $B = 011000100$ are two correlated stochastic streams representing $p_A = 5/8$ and $p_B = 3/8$, then $C = 00010010$ and $p_C = 2/8$. Note that conventional implementation of this function requires one *NOT* gate, one 2×1 *MUX*, and one finite state machine (FSM), increasing the area and energy overhead [18].

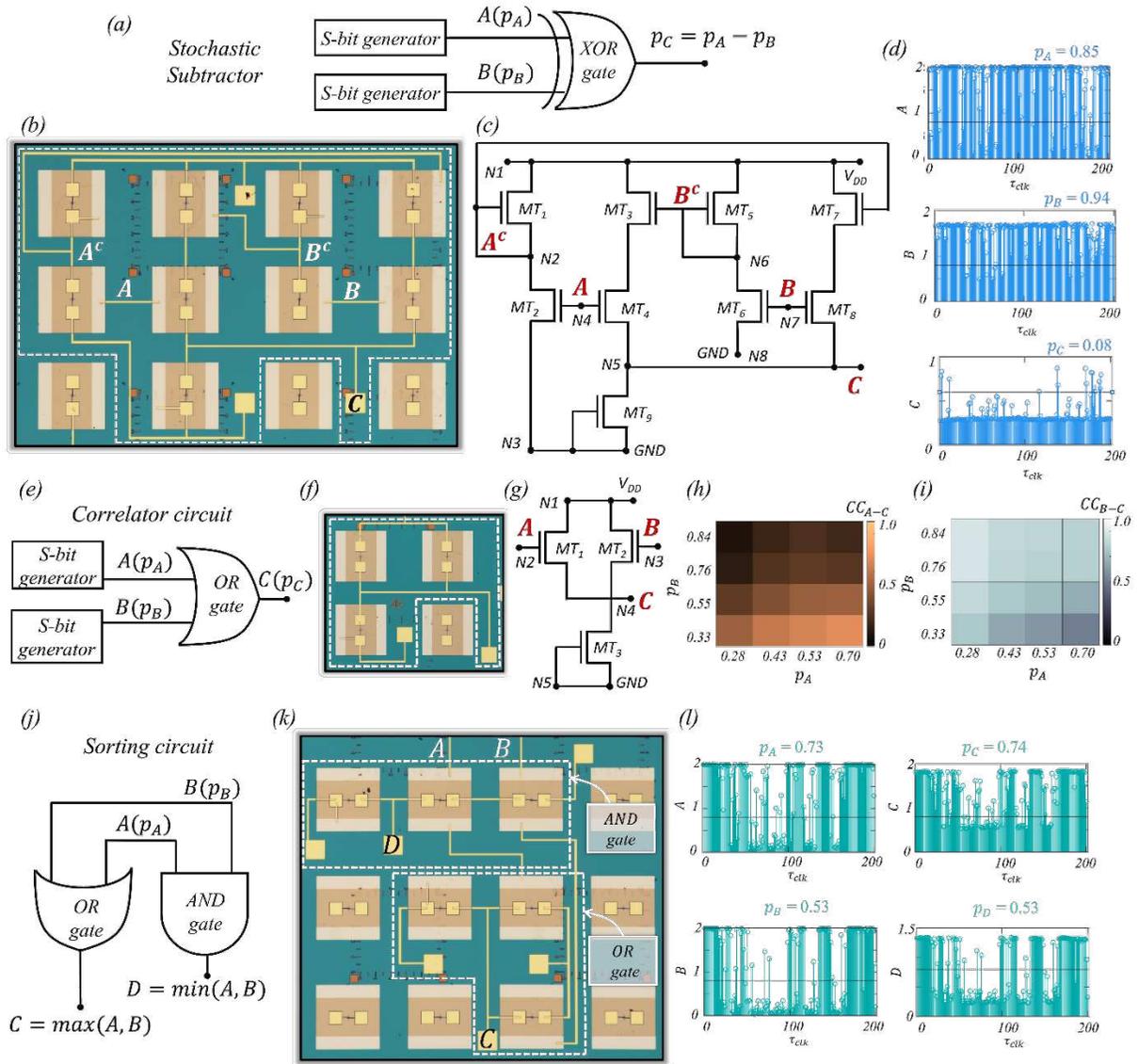


Figure 5. Stochastic subtraction and sorting using correlated s-bits. a) Schematic, b) optical image, and c) corresponding circuit configuration for stochastic subtraction using one XOR gate consisting of 9 memtransistors. d) Representative stochastic bit-streams for the random variables $A(p_A)$ and $B(p_B)$, which are highly correlated with $CC = 0.88$, and the corresponding output bit-stream for $C(p_C)$. e) Schematic, f) optical image, and g) corresponding circuit configuration for a correlator circuit based on OR gate consisting of 3 memtransistors. h) Colormaps of correlation coefficient between the output C and input A (CC_{A-C}) and input B (CC_{B-C}). i) Schematic, j) optical image, and k) corresponding circuit configuration of a sorting circuit comprising of one OR gate and one AND gate. l) Representative stochastic bit-streams for the correlated random variables A , B , and the sorted output C for maximum and D for minimum values, respectively.

Fig. 5b-c, respectively, show the optical image and corresponding circuit configuration of a XOR gate with a total of 9 memtransistors. Note that, memtransistor pairs, $MT1$ and $MT2$, and $MT5$ and $MT6$ are NOT gates used to invert A to A^c and B to B^c , respectively. A and B^c are applied to the local back-gates of $MT3$ and $MT4$, respectively, which are connected in series. Similarly, A^c and

B are applied to the local back-gates of $MT7$ and $MT8$, respectively, which are also connected in series. Finally the series connection of $MT3$ and $MT4$, and $MT7$ and $MT8$ are connected in parallel between node, $N1$ and $N5$. The drain terminal of $MT9$ is connected to $N5$, whereas, the source and gate terminals are shorted to the ground. The overall circuit accomplishes the XOR logic for the inputs, A and B at node $N5$. **Fig. 5d** shows the representative stochastic bit-streams for the correlated random variables $A(p_A = 0.85)$, $B(p_B = 0.93)$, and the output of the XOR gate, $C(p_C = 0.08)$, which is close to $|p_A - p_B|$. Clearly, our 9 memtransistor circuit is able to perform stochastic subtraction when the stochastic bit-streams are correlated. Note that the CC between A and B was intentionally made high, ~ 0.88 , by using a correlator circuit described below.

While the s-bit generators produce uncorrelated bit-streams, correlated random variables can be created by using an OR gate as shown in **Fig. 5e**. The optical image and corresponding circuit configuration of the OR gate comprising of 3 memtransistors are shown in **Fig. 5f-g**, respectively. Two mutually independent or uncorrelated stochastics inputs, A and B , obtained from the s-bit generators are applied to the local back-gates of $MT1$ and $MT2$, which are connected in parallel among themselves and in series with $MT3$. As explained earlier, $MT3$ operates as a load resistor and the entire circuit serves as an OR gate. Interestingly, the output, C , obtained at node, $N4$ becomes correlated with either or both, A and B . **Fig. 5h-i** shows the correlation coefficient between C and A , i.e., CC_{A-C} and C and B , i.e., CC_{B-C} , respectively, for different values of p_A and p_B . Clearly, CC_{A-C} and CC_{B-C} values range from ~ 0 to ~ 1 . Also note that lower p_A values ensure higher correlation between C and B and vice versa. Nevertheless, the correlator circuit allows us to obtain correlated bit-stream with desirable correlation coefficients. The average energy expenditure for obtaining correlated bit stream is ~ 0.8 nJ.

Sorting: As we have shown earlier an *AND* gate functions as a stochastic multiplier for uncorrelated bit-streams. However, when the inputs become highly correlated, it gives the minimum of the two stochastic streams. As an example, if $A = 01101110$ and $B = 01100100$ are two correlated stochastic streams representing $p_A = 5/8$ and $p_B = 3/8$, then $C = 01100100$ and $p_C = 3/8$. Similarly, an *OR* gate, gives the maximum value of two stochastic streams, i.e. $C = 01101110$ and $p_C = 5/8$. In contrast conventional implementation [17] with uncorrelated inputs require FSM-based stochastic tanh function along with the three *MUX*s, which again increases area and energy overhead. **Fig. 5j-k**, respectively, show the schematic and optical image of a sorting circuit i.e. finding the minimum and maximum between two stochastic variables, A and B . The circuit consists of 6 memtransistors. **Fig. 5l** show the representative stochastic bit-streams for the correlated random variables A , B , and the sorted output C for maximum and D for minimum values, respectively.

Table 1 summarizes the SC architecture for different arithmetic operations.

Arithmetic operation	# of s-bit generators	Logic gates	# of memtransistors	Average energy expenditure
Multiplication	2	<i>AND</i>	15	~0.8 nJ
Addition	3	2×1 <i>MUX</i>	22	~1.2 nJ
Subtraction (correlated s-bits)	2	<i>XOR</i>	24 (including correlator circuit)	~0.8 nJ
Sorting (correlated s-bits)	2	<i>OR, AND</i>	21 (including correlator circuit)	~0.8 nJ

Conclusion

In conclusion, we have exploited the cycle-to-cycle variability in the programmed conductance of monolayer MoS₂ based 2D memtransistors and translated the same into s-bits with reconfigurable probability of obtaining ‘1’ in the bit-stream using a s-bit generator circuit comprising of 6 memtransistors and subsequently combined the s-bit generator with 2D memtransistor based logic gates to demonstrated a standalone SC architecture that can perform accurate arithmetic operations such as addition, subtraction, multiplication, and sorting. Our SC architecture consumes miniscule energy ~ 1 nano Joules to perform arithmetic operations and uses limited numbers of memtransistors with small active-area footprint. Our demonstration offers a way to accelerate SC on a non-von Neumann platform based on novel 2D materials and devices.

Methods

Fabrication of local back-gate islands: To define the back-gate island regions, the substrate 285 nm SiO₂ on p⁺⁺-Si was spin coated with bilayer photoresist consisting of Lift-Off-Resist (LOR 5A) and Series Photoresist (SPR 3012) baked at 185 °C and 95 °C, respectively. The bilayer photoresist was then exposed to Heidelberg Maskless Aligner (MLA 150) to define the island and developed using MF CD26 microposit, followed by a de-ionized (DI) water rinse. The back gate electrode of 20/50 nm TiN/Pt was deposited using reactive sputtering. The photoresist was removed using acetone and Photo Resist Stripper (PRS 3000) and cleaned using 2-propanol (IPA) and DI water. Atomic layer deposition (ALD) process was then implemented to grow 50 nm Al₂O₃ on the entire substrate including the island regions. To access the individual Pt back-gate electrodes etch patterns were defined using the same bilayer photoresist consisting of LOR 5A and SPR 3012. The bilayer photoresist was then exposed to MLA 150 and developed using MF CD26 microposit. 50 nm Al₂O₃ was subsequently dry etched using the BCl₃ chemistry at 5 °C for 20 seconds, which was repeated four times to minimize heating in the substrate. Next the photoresist was removed to give access to the individual Pt electrodes.

Large area monolayer MoS₂ film growth: Monolayer MoS₂ was deposited on epi-ready 2” c-sapphire substrate by metalorganic chemical vapor deposition (MOCVD). An inductively heated graphite susceptor equipped with wafer rotation in a cold-wall horizontal reactor was used to achieve uniform monolayer deposition as previously described [30]. Molybdenum hexacarbonyl (Mo(CO)₆) and hydrogen sulfide (H₂S) were used as precursors. Mo(CO)₆ maintained at 10°C and 650 Torr in a stainless-steel bubbler was used to deliver 1.1×10^{-3} sccm of the metal precursor for the growth, while 400 sccm of H₂S was used for the process. MoS₂ deposition was carried out at

1000°C and 50 Torr in H₂ ambient, where monolayer growth was achieved in 18 min. The substrate was first heated to 1000°C in H₂ and maintained for 10 min before the growth was initiated. After growth, the substrate was cooled in H₂S to 300°C to inhibit decomposition of the MoS₂ films. More details can be found in our earlier work [25, 31, 32].

MoS₂ film transfer to local back-gate islands: To fabricate the 2D memtransistors, MOCVD grown monolayer MoS₂ film was transferred from the sapphire to SiO₂/p⁺⁺-Si substrate with local back-gate islands using PMMA (polymethyl-methacrylate) assisted wet transfer process. First, MoS₂ on sapphire substrate was spin coated with PMMA and then baked at 180 °C for 90 s. The corners of the spin-coated film were scratched using a razor blade and immersed inside 1 M NaOH solution kept at 90 °C. Capillary action causes the NaOH to be drawn into the substrate/film interface, separating the PMMA/ MoS₂ film from the sapphire substrate. The separated film was rinsed multiple times inside a water bath and finally transferred onto the SiO₂/p⁺⁺-Si substrate with local back-gate islands and then baked at 50 °C and 70 °C for 10 min each to remove moisture and residual PMMA, ensuring a pristine interface.

Fabrication of 2D memtransistors: To define the channel regions for the memtransistors, the substrate was spin-coated with PMMA and baked at 180 °C for 90 s. The resist was then exposed to electron beam (e-beam) and developed using 1:1 mixture of 4-methyl-2-pentanone (MIBK) and 2 propanol (IPA). The monolayer MoS₂ film was subsequently etched using sulfur hexafluoride (SF₆) at 5 °C for 30 s. Next, the sample was rinsed in acetone and IPA to remove the e-beam resist. To define the source and drain contacts, sample is then spin coated with methyl methacrylate (MMA) followed by A3 PMMA. Then using e-beam lithography source and drain contacts are

patterned and developed by using 1:1 mixture of MIBK and IPA for 60s. 40 nm of Nickel (Ni) and 30 nm of Gold (Au) are deposited using e-beam evaporation. Finally, lift-off process is performed to remove the evaporated Ni/Au except from the source/drain patterns by immersing the sample in acetone for 30 min followed by IPA for another 30 mins. Each island contains one memtransistor to allow for individual gate control.

Monolithic Integration: To define the connections between the respective memtransistors the substrate was spin coated with MMA and PMMA, followed by the e-beam lithography and developing using 1:1 mixture of MIBK and IPA, and e-beam evaporation of 60 nm Au. Finally, the e-beam resist was rinsed away by lift-off process using acetone and IPA.

Electrical Characterization: Electrical characterization of the fabricated devices are performed using Lake Shore CRX-VF probe station under atmospheric condition using a Keysight B1500A parameter analyzer.

Data Availability: The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Code Availability: The codes used for plotting the data are available from the corresponding authors on reasonable request.

AUTHOR INFORMATION

Corresponding Author

sud70@psu.edu, das.sapt@gmail.com

Author Contributions

S.D conceived the idea and designed the experiments. H.R., Y.Z., and T. F. S. fabricated the memtransistors. H.R., Y.Z., and S.D performed the measurements, analyzed the data, discussed the results, and agreed on their implications. N. T. grew MOCVD MoS₂. All authors contributed to the preparation of the manuscript.

Competing Interest

The authors declare no competing interests

Acknowledgement

The work was supported by Army Research Office (ARO) through Contract Number W911NF1920338. Authors also acknowledge the materials support from the National Science Foundation (NSF) through the Pennsylvania State University 2D Crystal Consortium–Materials Innovation Platform (2DCCMIP) under NSF cooperative agreement DMR-1539916.

References

- [1] D. A. Reed and J. Dongarra, "Exascale computing and big data," *Communications of the ACM*, vol. 58, pp. 56-68, 2015.
- [2] X.-k. Liao, K. Lu, C.-q. Yang, J.-w. Li, Y. Yuan, M.-c. Lai, *et al.*, "Moving from exascale to zettascale computing: challenges and techniques," *Frontiers of Information Technology & Electronic Engineering*, vol. 19, pp. 1236-1244, October 01 2018.
- [3] M. Yamaoka, T. Okuyama, S. Tanaka, M. Hayashi, and C. Yoshimura, "New computing paradigm for analyzing increasingly complex social infrastructure systems," *Hitachi Review*, vol. 64, pp. 525-531, 2015.
- [4] G. Indiveri and S.-C. Liu, "Memory and information processing in neuromorphic systems," *Proceedings of the IEEE*, vol. 103, pp. 1379-1397, 2015.
- [5] B. R. Gaines, "Stochastic computing," in *Proceedings of the April 18-20, 1967, spring joint computer conference*, 1967, pp. 149-156.
- [6] W. Poppelbaum, C. Afuso, and J. Esch, "Stochastic computing elements and systems," in *Proceedings of the November 14-16, 1967, fall joint computer conference*, 1967, pp. 635-644.
- [7] S. C. Smithson, N. Onizawa, B. H. Meyer, W. J. Gross, and T. Hanyu, "Efficient CMOS invertible logic using stochastic computing," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 66, pp. 2263-2274, 2019.
- [8] A. Ardakani, F. Leduc-Primeau, N. Onizawa, T. Hanyu, and W. J. Gross, "VLSI implementation of deep neural network using integral stochastic computing," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 25, pp. 2688-2699, 2017.
- [9] P. Knag, W. Lu, and Z. Zhang, "A native stochastic computing architecture enabled by memristors," *IEEE Transactions on Nanotechnology*, vol. 13, pp. 283-293, 2014.
- [10] S. Gaba, P. Knag, Z. Zhang, and W. Lu, "Memristive devices for stochastic computing," in *2014 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2014, pp. 2592-2595.
- [11] S. Gaba, P. Sheridan, J. Zhou, S. Choi, and W. Lu, "Stochastic memristive devices for computing and neuromorphic applications," *Nanoscale*, vol. 5, pp. 5872-5878, 2013.
- [12] R. Venkatesan, S. Venkataramani, X. Fong, K. Roy, and A. Raghunathan, "Spintastic: Spin-based stochastic logic for energy-efficient computing," in *2015 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2015, pp. 1575-1578.
- [13] G. Finocchio, M. Di Ventra, K. Y. Camsari, K. Everschor-Sitte, P. K. Amiri, and Z. Zeng, "The promise of spintronics for unconventional computing," *arXiv preprint arXiv:1910.07176*, 2019.
- [14] J. Hu, B. Li, C. Ma, D. Lilja, and S. J. Koester, "Spin-hall-effect-based stochastic number generator for parallel stochastic computing," *IEEE Transactions on Electron Devices*, vol. 66, pp. 3620-3627, 2019.
- [15] W. A. Borders, A. Z. Pervaiz, S. Fukami, K. Y. Camsari, H. Ohno, and S. Datta, "Integer factorization using stochastic magnetic tunnel junctions," *Nature*, vol. 573, pp. 390-393, 2019.
- [16] A. Alaghi and J. P. Hayes, "Survey of stochastic computing," *ACM Transactions on Embedded computing systems (TECS)*, vol. 12, pp. 1-19, 2013.

- [17] P. Li, D. J. Lilja, W. Qian, K. Bazargan, and M. D. Riedel, "Computation on stochastic bit streams digital image processing case studies," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 22, pp. 449-462, 2013.
- [18] P. Li and D. J. Lilja, "Using stochastic computing to implement digital image processing algorithms," in *2011 IEEE 29th International Conference on Computer Design (ICCD)*, 2011, pp. 154-161.
- [19] K. Yang, D. Fick, M. B. Henry, Y. Lee, D. Blaauw, and D. Sylvester, "16.3 A 23Mb/s 23pJ/b fully synthesized true-random-number generator in 28nm and 65nm CMOS," in *2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, 2014, pp. 280-281.
- [20] W. Sun, B. Gao, M. Chi, Q. Xia, J. J. Yang, H. Qian, *et al.*, "Understanding memristive switching via in situ characterization and device modeling," *Nature communications*, vol. 10, pp. 1-13, 2019.
- [21] A. Jaiswal, X. Fong, and K. Roy, "Comprehensive scaling analysis of current induced switching in magnetic memories based on in-plane and perpendicular anisotropies," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 6, pp. 120-133, 2016.
- [22] A. Sengupta, P. Panda, P. Wijesinghe, Y. Kim, and K. Roy, "Magnetic tunnel junction mimics stochastic cortical spiking neurons," *Scientific reports*, vol. 6, pp. 1-8, 2016.
- [23] Z. Fu, Y. Tang, X. Zhao, K. Lu, Y. Dong, A. Shukla, *et al.*, "An Overview of Spintronic True Random Number Generator," *Frontiers in Physics*, vol. 9, p. 172, 2021.
- [24] 2DCC. *2d-crystal-consortium*. Available: <https://www.mri.psu.edu/2d-crystal-consortium/user-facilities/thin-films/list-thin-film-samples-available>
- [25] A. Sebastian, R. Pendurthi, T. H. Choudhury, J. M. Redwing, and S. Das, "Benchmarking monolayer MoS2 and WS2 field-effect transistors," *Nature Communications*, vol. 12, p. 693, 2021/01/29 2021.
- [26] S. Das, H.-Y. Chen, A. V. Penumatcha, and J. Appenzeller, "High performance multilayer MoS2 transistors with scandium contacts," *Nano letters*, vol. 13, pp. 100-105, 2013.
- [27] D. S. Schulman, A. J. Arnold, and S. Das, "Contact engineering for 2D materials and devices," *Chemical Society Reviews*, vol. 47, pp. 3037-3058, 2018.
- [28] S. Chuang, C. Battaglia, A. Azcatl, S. McDonnell, J. S. Kang, X. Yin, *et al.*, "MoS2 p-type transistors and diodes enabled by high work function MoO_x contacts," *Nano letters*, vol. 14, pp. 1337-1342, 2014.
- [29] A. Alaghi and J. P. Hayes, "Exploiting correlation in stochastic circuit design," in *2013 IEEE 31st International Conference on Computer Design (ICCD)*, 2013, pp. 39-46.
- [30] Y. Xuan, A. Jain, S. Zafar, R. Lotfi, N. Nayir, Y. Wang, *et al.*, "Multi-scale modeling of gas-phase reactions in metal-organic chemical vapor deposition growth of WSe₂," *Journal of Crystal Growth*, vol. 527, 2019.
- [31] D. Jayachandran, A. Oberoi, A. Sebastian, T. H. Choudhury, B. Shankar, J. M. Redwing, *et al.*, "A low-power biomimetic collision detector based on an in-memory molybdenum disulfide photodetector," *Nature Electronics*, vol. 3, pp. 646-655, 2020/10/01 2020.
- [32] A. Dodda, A. Oberoi, A. Sebastian, T. H. Choudhury, J. M. Redwing, and S. Das, "Stochastic resonance in MoS2 photodetector," *Nature Communications*, vol. 11, p. 4406, 2020/09/02 2020.

Figure Caption

Figure 1. Fabrication and characterization of 2D memtransistors for acceleration of stochastic computing (SC). a) Optical image of a representative 2D memtransistors based medium scale integrated circuit for the hardware acceleration of SC. b) Optical image and corresponding 3D schematic of a representative 2D memtransistor based on monolayer MoS₂, which are locally back-gated using a stack comprising of atomic layer deposition (ALD) grown 50 nm Al₂O₃ on sputter deposited 40/30 nm Pt/TiN. All back-gate islands were fabricated on SiO₂/p⁺⁺-Si substrate. c) Transfer characteristics, i.e. source to drain current (I_{DS}) versus local back-gate voltage (V_{BG}) measured using source to drain bias, $V_{DS} = 1$ V for a representative MoS₂ memtransistor with channel length, $L = 1$ μm , and channel width, $W = 5$ μm in linear and logarithmic scale. d) Output characteristics, i.e. I_{DS} versus V_{DS} for different V_{BG} for the same MoS₂ memtransistor. e) Device-to-device variation in the transfer characteristics and f) corresponding histogram of extracted field effect mobility (μ_{FE}) distribution across 50 memtransistors. Analog g) programming and h) erase capability of 2D memtransistor when subjected to negative “Write” (V_P) and positive “Erase” (V_E) voltage pulses of different amplitudes ranging from 6 V to 13 V applied to the local back-gate electrode, each for a duration of $\tau_{P/E} = 100$ ms. i) Non-volatile retention for 4 representative programmed and erased states for 100 seconds.

Figure 2. Programming stochasticity in 2D memtransistor and s-bit generation. a) Transfer characteristics of a representative 2D memtransistor, measured each time after the application of $V_P = -10$ V and $V_E = 10$ V each for $\tau_s = 100$ ms, for a total of 100 cycles. b) Distribution of post-programmed and post-erased conductance (G_{MT}) measured using $V_{BG} = 0$ V follow Gaussian random distributions, which forms the basis for high-quality randomness for s-bit generation. c)

optical image and d) corresponding circuit diagram for the proposed s-bit generator consisting of six memtransistors ($MT1, MT2, MT3, MT1, MT2, MT3$). e) Voltage waveforms applied to the nodes, $N1, N2$, i.e., V_{N1}, V_{N2} . During each clock cycle (τ_{clk}), V_{N1} toggles between 0 V, 0 V, and $V_{DD} = 2$ V and V_{N2} toggles between $V_P = -7$ V, $V_E = 10$ V, and $V_R = 1$ V. Voltages applied to nodes, $N3$, and $N4$, i.e., V_{N3} , and V_{N4} are held constant at 1V and 0 V, respectively. f) Voltage readout at node, $N5$, i.e., V_{N5} . Since the memtransistors, $MT1$ and $MT2$ are connected in series and G_{MT1} fluctuates due to programming and reset every τ_{clk} , so does V_{N5} . g) Distribution of V_{N5} over 200 τ_{clk} follows a random Gaussian distribution with mean, $\mu_{VN5} = 0.36$ V and standard deviation, $\sigma_{VN5} = 0.05$ V. h) Output, V_{N6} , of an inverting amplifier constructed using $MT3$ and $MT4$ as a function of the input, V_{N5} with a gain of ~ 7 . i) V_{N6} corresponding to V_{N5} shown in (f). j) Distribution of V_{N6} which follows a random Gaussian distribution with mean, $\mu_{VN6} = 0.77$ V and an increased standard deviation of $\sigma_{VN6} = 0.33$ V. k) Output, V_{N7} , of a thresholding inverter constructed using $MT5$ and $MT6$ as a function of the input, V_{N6} for different inversion threshold, V_{IT} . l) V_{N7} corresponding to V_{N6} shown in (i) for different V_{IT} . m) Probability of obtaining '1' in the bit stream (p_s) as a function of V_{IT} . This clearly shows the ability of the proposed circuit to transform the cycle-to-cycle conductance fluctuations in 2D memtransistor into s-bits with reconfigurable p_s that lie between [0,1].

Figure 3. Stochastic multiplier. a) Schematic, b) optical image, and c) corresponding circuit configuration of a stochastic multiplier consisting of 2 s-bit generator and one *AND* gate with a total of 15 memtransistors. d) Representative stochastic bit-streams for the random variables, $A(p_A)$ and $B(p_B)$ obtained from their respective s-bit generators and the corresponding output bit-stream for $C(p_C)$. Colormaps of e) percentage errors (ϵ) for multiplication and f) corresponding

correlation coefficient (CC) for different combinations of p_A and p_B . Lower values of ϵ is a direct consequence of near ideal CC values close to zero indicating mutual independence of A and B , which is critical for accurate multiplication. We have used bit-streams of length 200-bit to evaluate the probability values associated with the random variable.

Figure 4. Stochastic adder. a) Schematic, b) optical image, and c) corresponding circuit configuration of a stochastic adder consisting of 3 s-bit generator and one 2×1 MUX gate with a total of 22 memtransistors. d) Representative stochastic bit-streams for the random variables $S(p_S)$, $A(p_A)$, and $B(p_B)$ obtained from their respective s-bit generation modules at nodes $N7$, $N12$, and $N13$ and the corresponding output bit-stream for $C(p_C)$. e) Colormaps of percentage errors (ϵ) for scaled addition for different combinations of p_A , p_B , for $p_S \approx 0.5$.

Figure 5. Stochastic subtraction and sorting using correlated s-bits. a) Schematic, b) optical image, and c) corresponding circuit configuration for stochastic subtraction using one XOR gate consisting of 9 memtransistors. d) Representative stochastic bit-streams for the random variables $A(p_A)$ and $B(p_B)$, which are highly correlated with $CC = 0.88$, and the corresponding output bit-stream for $C(p_C)$. e) Schematic, f) optical image, and g) corresponding circuit configuration for a correlator circuit based on OR gate consisting of 3 memtransistors. h) Colormaps of correlation coefficient between the output C and input A (CC_{A-C}) and input B (CC_{B-C}). i) Schematic, j) optical image, and k) corresponding circuit configuration of a sorting circuit comprising of one OR gate and one AND gate. l) Representative stochastic bit-streams for the correlated random variables A , B , and the sorted output C for maximum and D for minimum values, respectively.