

## Supporting Information

### Implementation of Bayesian Network and Bayesian Inference using $\text{Cu}_{0.1}\text{Te}_{0.9}/\text{HfO}_2/\text{Pt}$

#### Threshold Switching Memristor

*In Kyung Baek<sup>1,#</sup>, Soo Hyung Lee<sup>1,#</sup>, Yoon Ho Jang<sup>1</sup>, Hyungjun Park<sup>1</sup>, Jaehyun Kim<sup>1</sup>, Sunwoo Cheong<sup>1</sup>, Sung Keun Shim<sup>1</sup>, Janguk Han<sup>1</sup>, Joon-Kyu Han<sup>1</sup>, Gwang Sik Jeon<sup>1</sup>, Dong Hoon Shin<sup>1</sup> Kyung Seok Woo<sup>1,†</sup>, and Cheol Seong Hwang<sup>1,†</sup>*

<sup>1</sup>Department of Materials Science and Engineering and Inter-University Semiconductor Research Center, Seoul National University, Gwanak-ro 1, Daehag-dong, Gwanak-gu, Seoul, 08826, Republic of Korea.

#These two authors contributed equally.

†Correspondence: K.S.W. (e-mail: kevinwoo@snu.ac.kr), C.S.H. (e-mail: cheolsh@snu.ac.kr)

#### **Table of contents**

- **Supplementary Fig. S1-S11**
- **Supplementary Table S1**
- **Supplementary Note S1-S3**
- **Supplementary References**

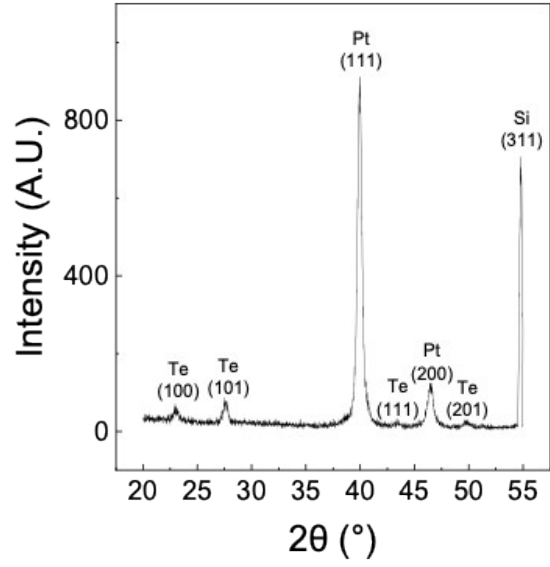
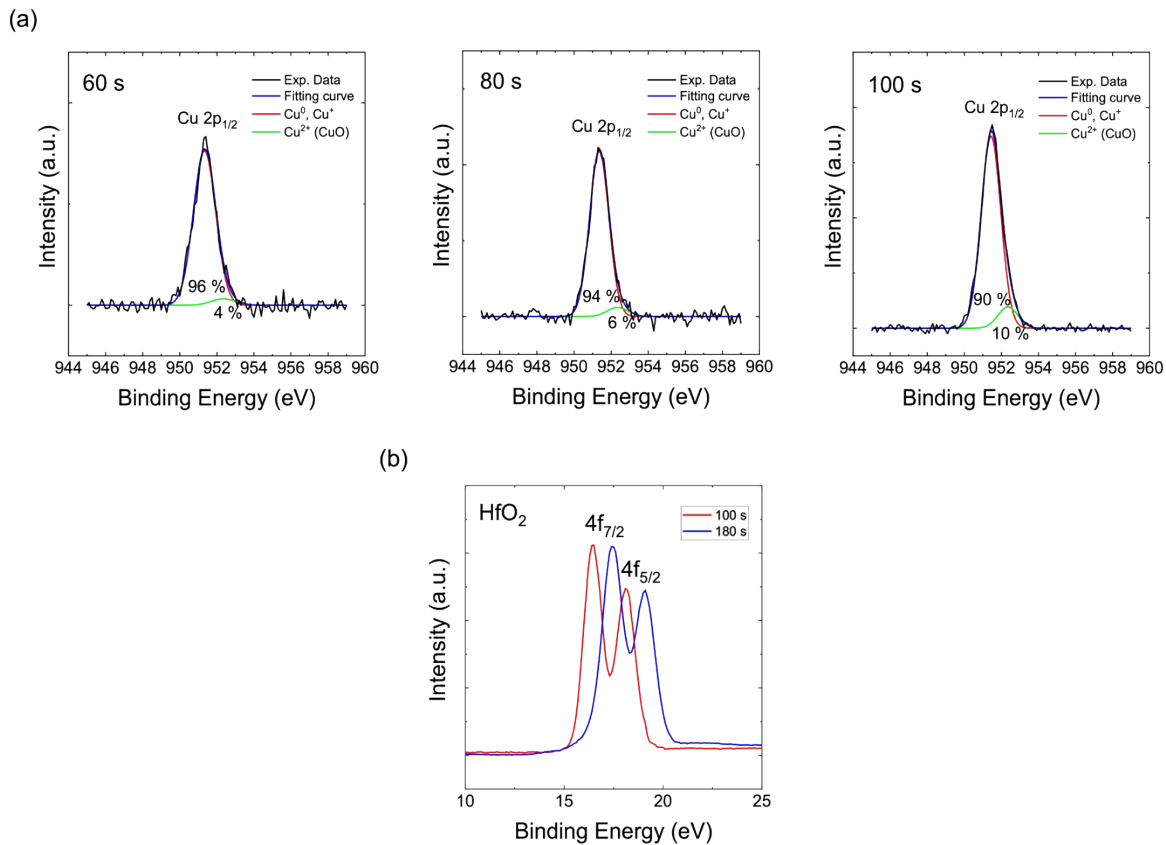
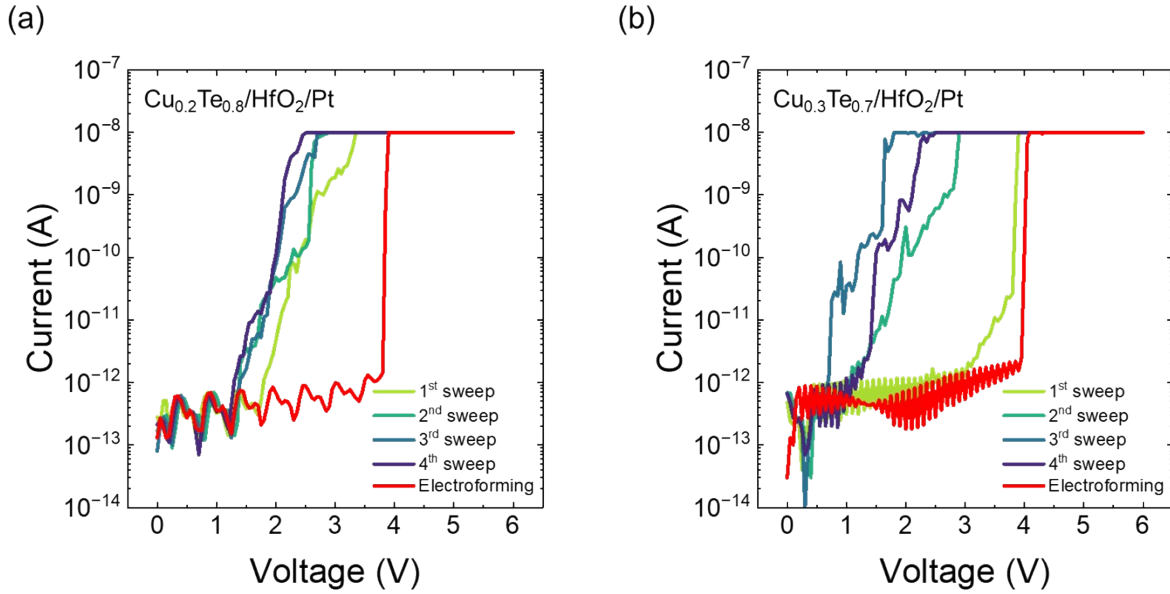


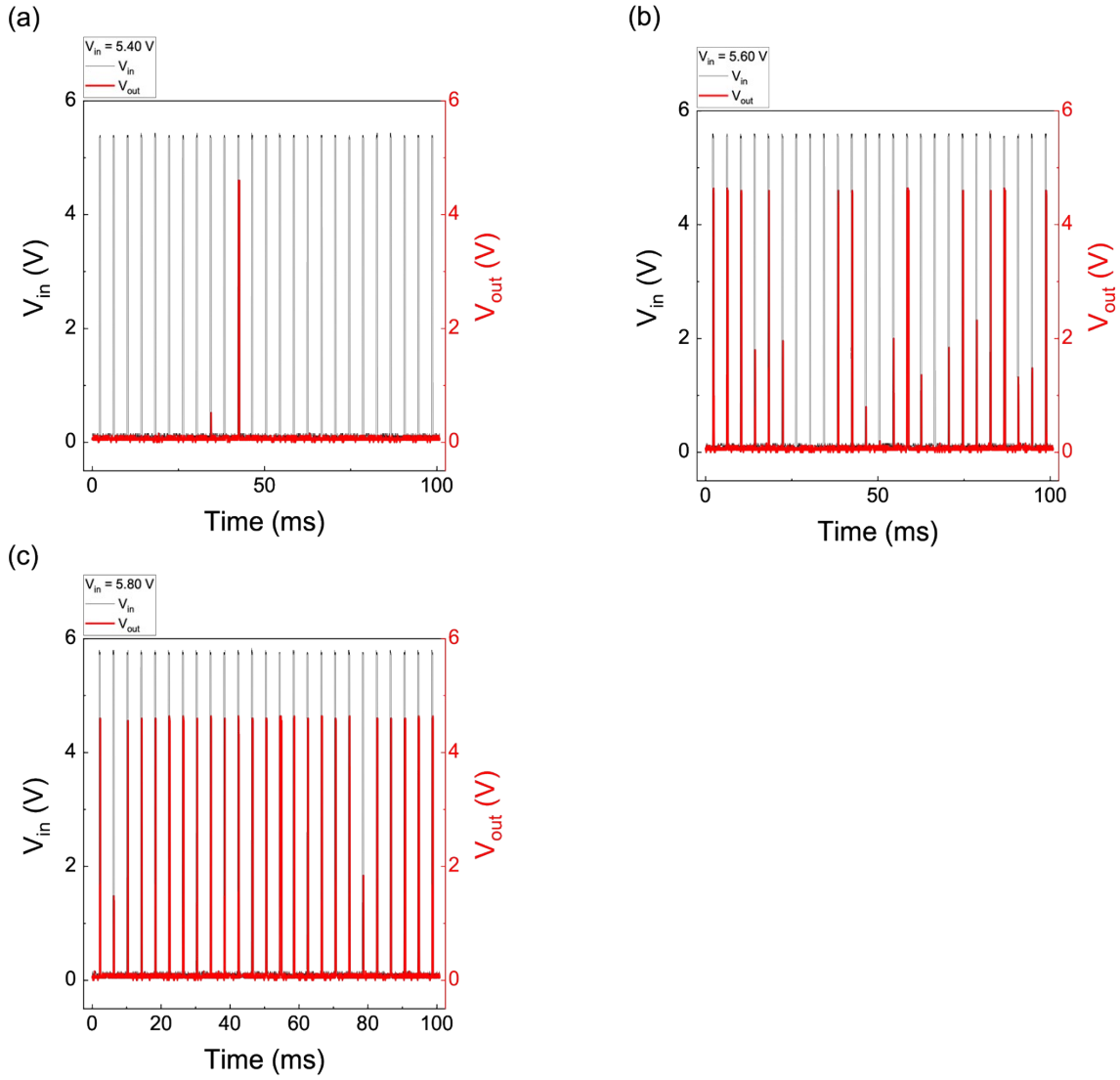
Fig. S1. Glancing angle X-ray diffraction (GAXRD) analysis results of the CTHP memristor.



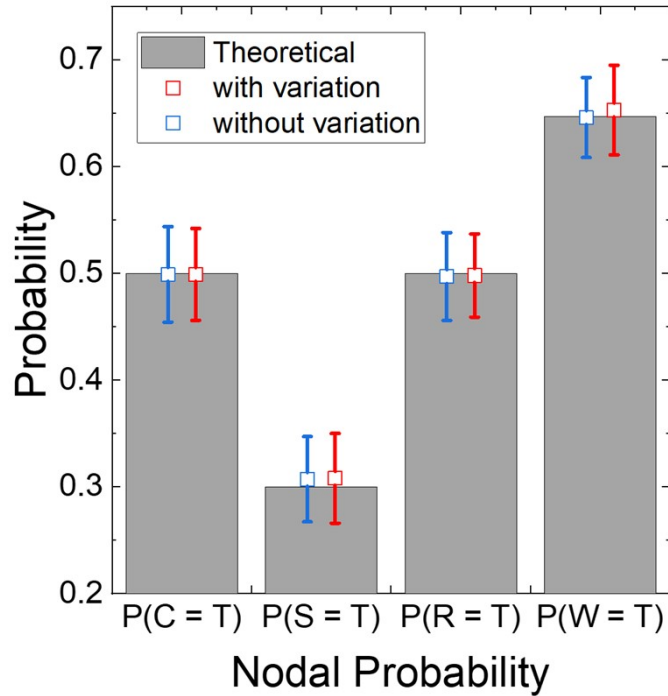
**Fig. S2. X-ray photoelectron spectroscopy (XPS) depth profiling analysis for Cu  $2p_{1/2}$  and Hf 4f peak.** (a) Cu,  $\text{Cu}_2\text{O}$  (951.33 eV) and  $\text{CuO } 2p_{1/2}$  peak (952.35 eV) of XPS analysis according to the etching time (60 s, 80 s, 100 s). The proportion of  $\text{CuO}$  peak increases as the etched surface becomes closer to the interfacial layer. (b) Hf 4f peak of XPS analysis according to the etching time (100 s, 180 s). 100 s and 180 s correspond to the interface and bulk of  $\text{HfO}_2$ , respectively.



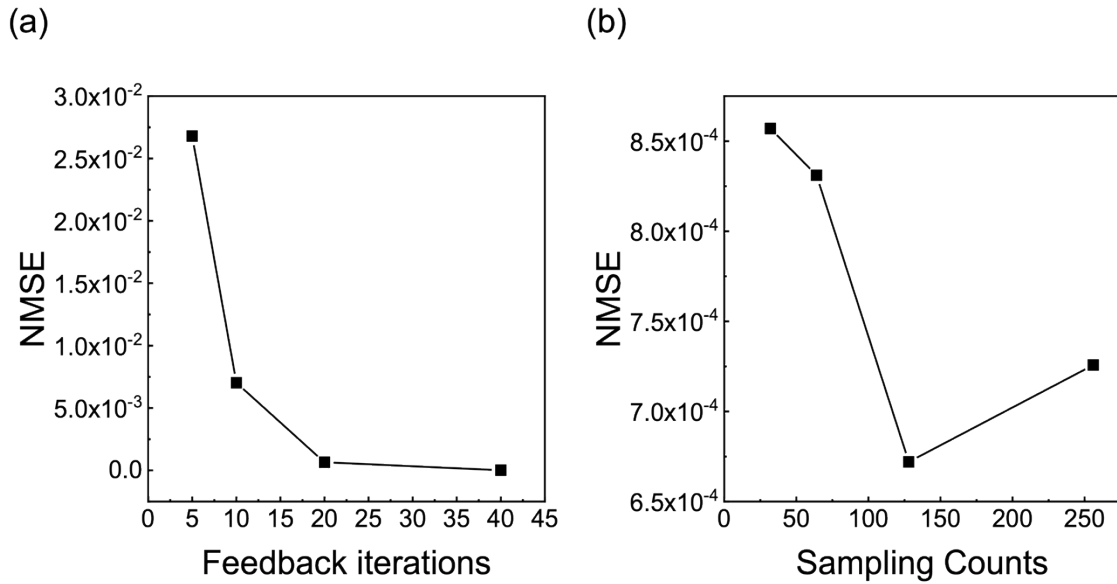
**Fig. S3. CTHP devices with different top electrode configurations.** (a) and (b) shows five DC sweep results for  $\text{Cu}_{0.2}\text{Te}_{0.8}/\text{HfO}_2/\text{Pt}$  and  $\text{Cu}_{0.3}\text{Te}_{0.7}/\text{HfO}_2/\text{Pt}$  devices, respectively. Both devices show an initial electroforming sweep colored red. However, as the Cu composition increases, CTHP exhibits a gradual shift in set voltage due to an excessive injection of Cu ions in the  $\text{HfO}_2$ . In contrast to  $\text{Cu}_{0.1}\text{Te}_{0.9}/\text{HfO}_2/\text{Pt}$  in Fig. 3a, gradual switching behavior after electroforming is observed for both devices due to the increased Cu clusters and thick filament formation.



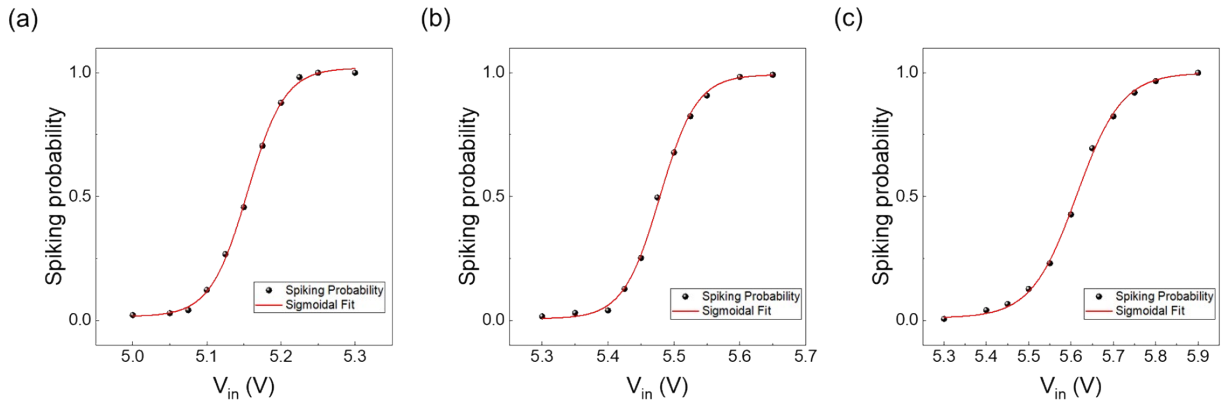
**Fig. S4. The response of CTHP-based p-bit neuron circuit by input pulse streams.** CTHP-based p-bit neuron outputs at (a),  $V_{in} = 5.40$  V, (b),  $V_{in} = 5.60$  V and (c),  $V_{in} = 5.80$  V. The figures show the partial responses of the neuron, resulting from 25 input pulses.



**Fig. S5. Impact of the cycle-to-cycle variation on probability inference accuracy.** Figure shows the inference results with and without the device's cycle-to-cycle (c-to-c) variation (Fig. 3) for the inference result in Table 1. The inferences of nodal probability with and without c-to-c variation are colored red and blue, respectively. As the inference process involves 128 sequential samplings of p-bit neurons, the inherent variation of CTHP is suppressed, resulting in two outcomes that exhibit no significant difference. The results show mean and standard deviation from 100 independent inferences.

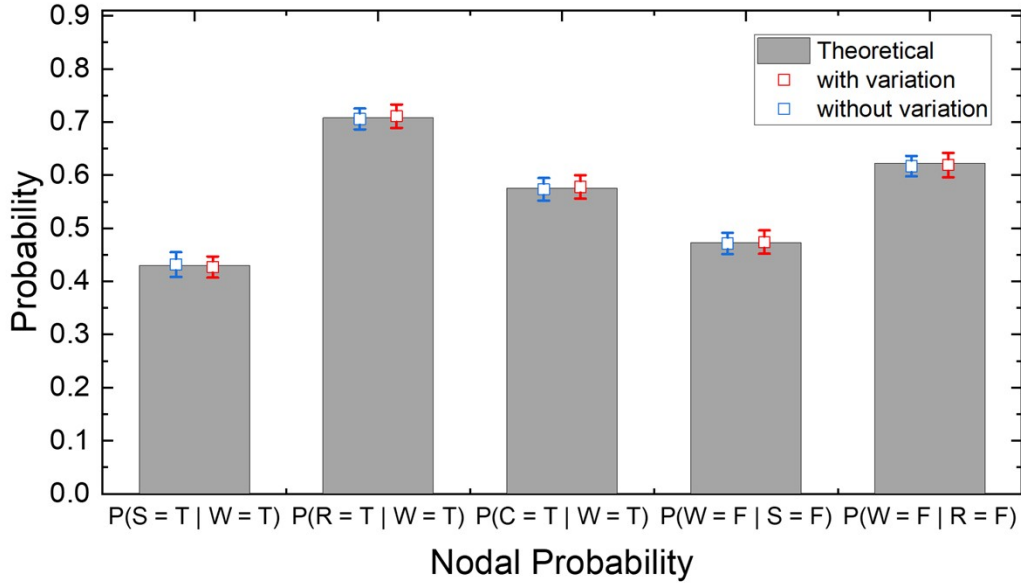


**Fig. S6. Normalized mean squared error (NMSE) for the inference in a simple Bayesian network in Fig. 1a.** (a) NMSE for the number of total feedback iterations (5, 10, 20, 40). The error decreases with increased feedback iterations but reaches a saturation value at 20 feedback iterations. (b) NMSE for the number of sampling counts (32, 64, 128, 256). The error decreases to 128 sampling counts but increases again to 256 sampling counts. This behavior is attributed to the inherent noise in the CTHP memristor. In each condition, the error was calculated for all the posterior probabilities of nodes being 'True.' The control variables were held constant at the values shown in Fig. 6. 20 feedback iterations and 128 sampling counts for inference provide the most efficient method to minimize the error between the inferred and theoretical values.



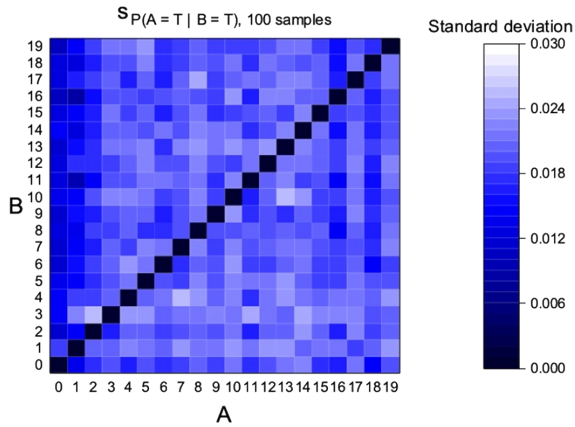
**Fig. S7. Device-to-device variation in CTHP devices and their sigmoidal curves.** (a)-(c) show three sigmoidal curves obtained from three different CTHP devices. Even though the devices exhibit a shift in sigmoidal curves due to the device-to-device (d-to-d) variation, it can be mitigated by applying different set voltages to each device. This process is performed by setting the DAC output voltages to the MUX (Fig. 4) according to the probability-voltage sigmoidal curve of each device.



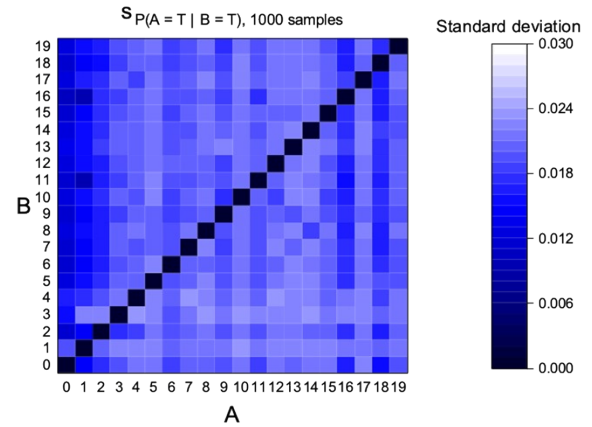


**Fig. S8. Impact of the cycle-to-cycle and device-to-device variation on probability inference accuracy with the division feedback logic.** The figure shows the inference results with and without the device's c-to-c and d-to-d variation for the inference result in Table 2. As the division feedback logic with an exponentially decreasing learning rate feasibly mitigates the effect of c-to-c variation, two inference results with and without variation show no significant difference. Moreover, according to the measurement data, d-to-d variation was controlled by applying a specific set voltages to each p-bit neuron.

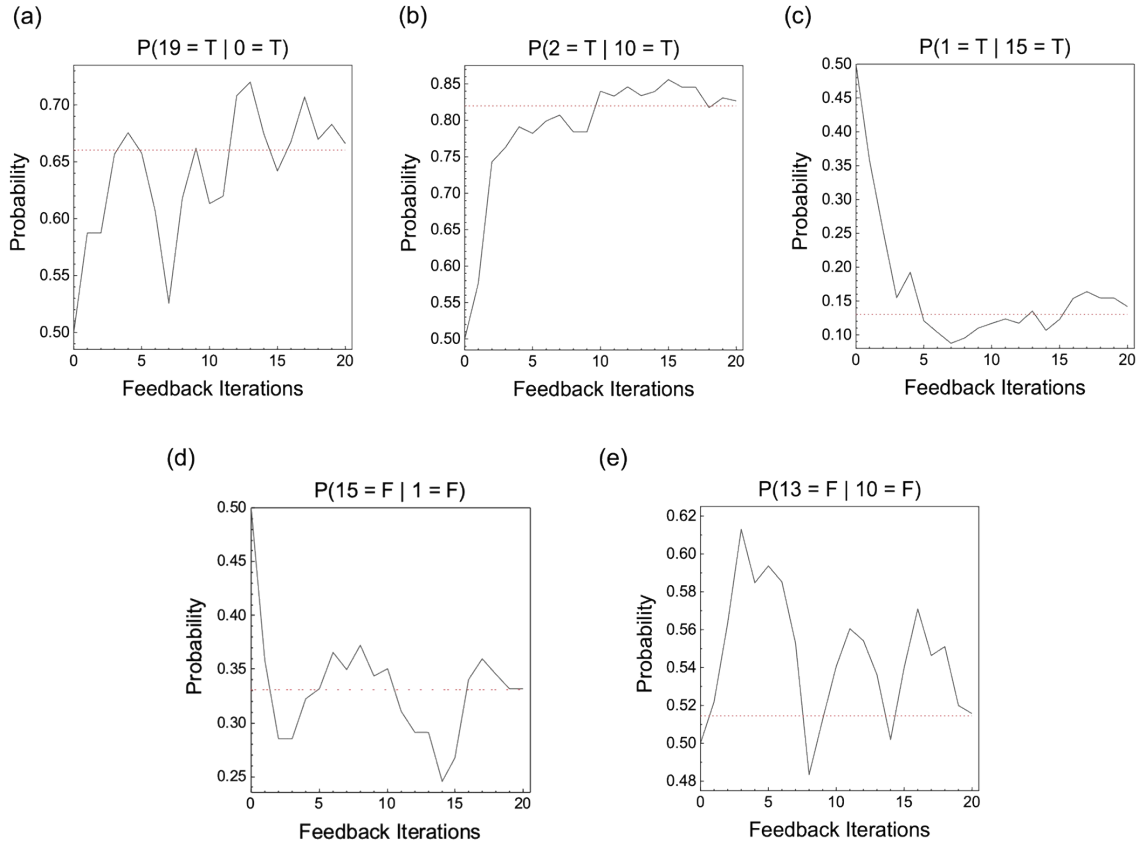
(a)



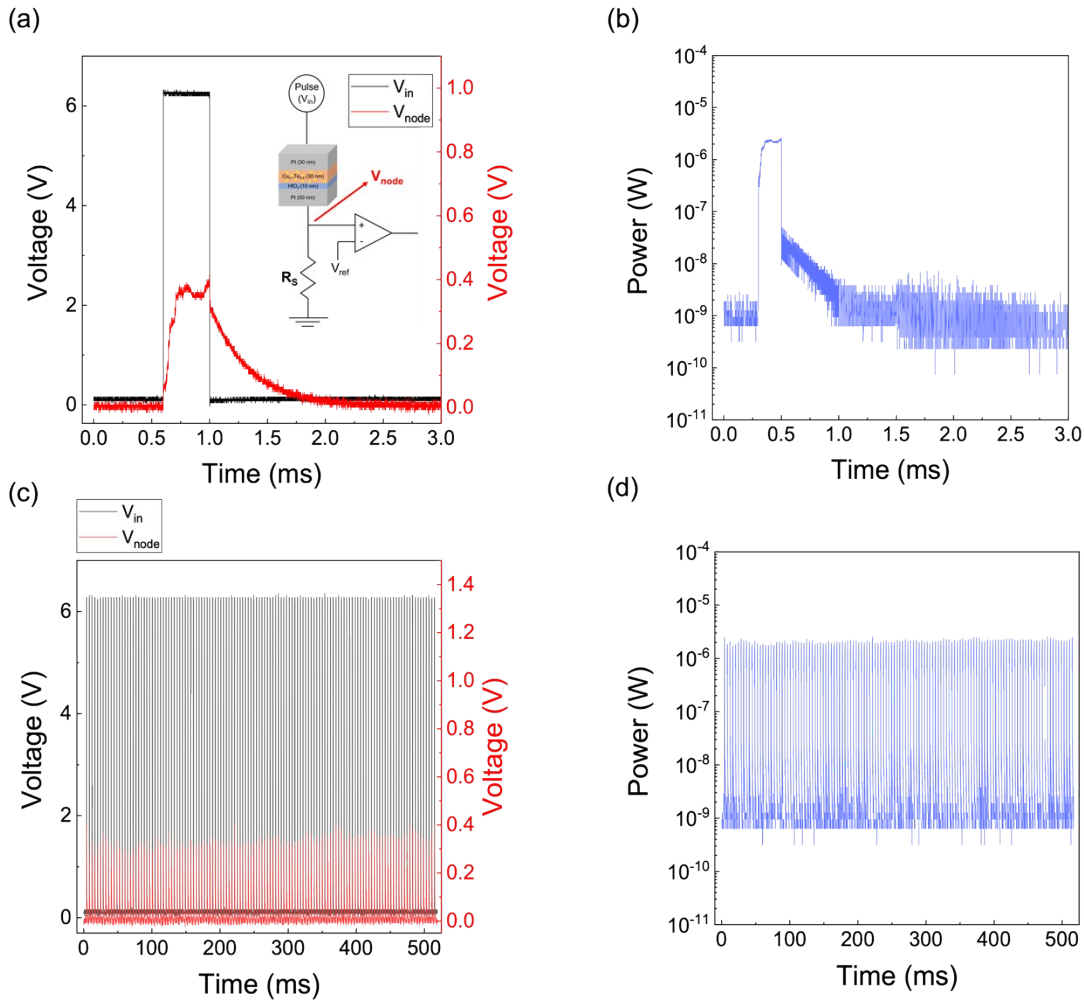
(b)



**Fig. S9. Standard deviation values of the Bayesian inference.** (a)-(b) Matrix color mapping for the standard deviation of inference values for the network, shown in Fig. 7a of the main text. Inference results show all the conditional probabilities,  $P(A = T | B = T)$  through the number of samples. Generally, the result for 1000 samples shows less deviation than for 100 samples.



**Fig. S10. Posterior probability inferences of a complex network.** (a)-(e) The inference results of five posterior probabilities in a complex Bayesian network, shown in Fig. 7a of the main text, as a function of feedback iterations. The black lines are the inference values, and the red dashed lines are the theoretical values.



**Fig. S11. The power consumption of the CTHP memristor and a serial resistor of the p-bit neuron.** (a) The pulse measurement of the CTHP memristor and a serial resistor of the p-bit neuron. The input voltage of 6.3 V, marked as black, is applied to the top electrode of the memristor. The  $V_{node}$ , marked as red, is the divided voltage between the memristor and the resistor. The inset shows a circuit configuration of the measurement. (b) Total power of the CTHP memristor and the serial resistor ( $P = V_{in} \times V_{node} / R_s$ ). The average power of one period is 186 nW. (c) The spiking behavior in response to 128 input pulses with the identical measurement as a. A width and a period of the pulses are 400  $\mu$ s and 4 ms, respectively. The most outputs correspond to spiking pulses since the  $V_{node} > 0.3$  V at the most input pulses. (d) Total power of the CTHP memristor and the serial resistor. The average power is 156 nW despite approximately 100 % spiking probability for 128 input pulses.

**Table S1. Energy consumption breakdown of different works for hardware implementation of the Bayesian network.** The energy for the representation of the single node in a Bayesian network is shown for each work. A clock frequency of 1.2 GHz is assumed, and a detailed explanation for the energy calculation is provided in Note S3.

	<b>Component</b>	<b>Energy (fJ)</b>	<b>Counts</b>
<b>CTHP (This work)</b>	CTHP device + resistor ( $R_s$ )	1.284	1
	MUX	40	1
	Comparator	400	1
	1-bit register (D flip-flop)	0.076	1
	<b>Total</b>	<b>441.4</b>	
<b>SiO<sub>x</sub> nanorod<sup>1</sup></b>	SiO <sub>x</sub> nanorod device	3.383	1
	Comparator	400	1
	Resistors ( $R_p, R_L$ )	1363.125	1
	<b>Total</b>	<b>1767</b>	
<b>MTJ<sup>2</sup></b>	MTJ device	132.4	1
	MUX	40	4
	Comparator	400	1
	Transistor	-	1
	<b>Total</b>	<b>692.4</b>	
<b>CMOS<sup>3,4</sup></b>	LFSR	$275.6 \times 10^9$	1
	Inference circuit	-	-
	<b>Total</b>	<b><math>&gt; 275.6 \times 10^9</math></b>	

**Note S1.** Calculations of the nodal probabilities using the Bayes' theorem.

$$P_{node}(S = T)$$

$$\begin{aligned} &= P(S = T \cap C = T) + P(S = T \cap C = F) \\ &= P(S = T | C = T)P(C = T) + P(S = T | C = F)P(C = F) \\ &= 0.1 \times 0.5 + 0.5 \times 0.5 = 0.3 \end{aligned}$$

$$P_{node}(R = T)$$

$$\begin{aligned} &= P(R = T \cap C = T) + P(R = T \cap C = F) \\ &= P(R = T | C = T)P(C = T) + P(R = T | C = F)P(C = F) \\ &= 0.8 \times 0.5 + 0.2 \times 0.5 = 0.5 \end{aligned}$$

$$P_{node}(W = T)$$

$$\begin{aligned} &= P(W = T \cap S = T \cap R = T) + P(W = T \cap S = T \cap R = F) \\ &+ P(W = T \cap S = F \cap R = T) + P(W = T \cap S = F \cap R = F) \\ &= P(W = T | S = T \cap R = T)P(S = T \cap R = T) + P(W = T | S = T \cap R = F)P(S = T \cap R = F) \\ &+ P(W = T | S = F \cap R = T)P(S = F \cap R = T) + P(W = T | S = F \cap R = F)P(S = F \cap R = F) \end{aligned}$$

$$\begin{aligned} &= P(W = T | S = T \cap R = T)\{P(S = T \cap R = T \cap C = T) + P(S = T \cap R = T \cap C = F)\} \\ &+ P(W = T | S = T \cap R = F)\{P(S = T \cap R = F \cap C = T) + P(S = T \cap R = F \cap C = F)\} \\ &+ P(W = T | S = F \cap R = T)\{P(S = F \cap R = T \cap C = T) + P(S = F \cap R = T \cap C = F)\} \\ &+ P(W = T | S = F \cap R = F)\{P(S = F \cap R = F \cap C = T) + P(S = F \cap R = F \cap C = F)\} \end{aligned}$$

$$\begin{aligned} &= P(W = T | S = T \cap R = T)\{P(S = T \cap R = T | C = T)P(C = T) + P(S = T \cap R = T | C = F)P(C = F)\} \\ &+ P(W = T | S = T \cap R = F)\{P(S = T \cap R = F | C = T)P(C = T) + P(S = T \cap R = F | C = F)P(C = F)\} \\ &+ P(W = T | S = F \cap R = T)\{P(S = F \cap R = T | C = T)P(C = T) + P(S = F \cap R = T | C = F)P(C = F)\} \\ &+ P(W = T | S = F \cap R = F)\{P(S = F \cap R = F | C = T)P(C = T) + P(S = F \cap R = F | C = F)P(C = F)\} \end{aligned}$$

$$\begin{aligned} &= P(W = T | S = T \cap R = T)\{P(S = T | C = T)P(R = T | C = T)P(C = T) + P(S = T | C = F)P(R = T | C = F)P(C = F)\} \\ &+ P(W = T | S = T \cap R = F)\{P(S = T | C = T)P(R = F | C = T)P(C = T) + P(S = T | C = F)P(R = F | C = F)P(C = F)\} \\ &+ P(W = T | S = F \cap R = T)\{P(S = F | C = T)P(R = T | C = T)P(C = T) + P(S = F | C = F)P(R = T | C = F)P(C = F)\} \\ &+ P(W = T | S = F \cap R = F)\{P(S = F | C = T)P(R = F | C = T)P(C = T) + P(S = F | C = F)P(R = F | C = F)P(C = F)\} \end{aligned}$$

$$= 0.99 \times (0.1 \times 0.8 \times 0.5 + 0.5 \times 0.2 \times 0.5) + 0.9 \times (0.1 \times 0.2 \times 0.5 + 0.5 \times 0.8 \times 0.5)$$

$$+ 0.9 \times (0.9 \times 0.8 \times 0.5 + 0.5 \times 0.2 \times 0.5) + 0 \times (0.9 \times 0.2 \times 0.5 + 0.5 \times 0.8 \times 0.5)$$

$$= 0.0891 + 0.189 + 0.369 + 0 = 0.6471$$

**Note S2.** Calculations of the posterior probability using the Bayes' theorem.

$$\begin{aligned}
 P_{post}(\mathbf{R} = \mathbf{T} \mid \mathbf{W} = \mathbf{T}) &= \frac{P(\mathbf{R} = \mathbf{T} \cap \mathbf{W} = \mathbf{T})}{P(\mathbf{W} = \mathbf{T})} \\
 &= \frac{P(\mathbf{R} = \mathbf{T} \cap \mathbf{W} = \mathbf{T} \cap \mathbf{S} = \mathbf{T}) + P(\mathbf{R} = \mathbf{T} \cap \mathbf{W} = \mathbf{T} \cap \mathbf{S} = \mathbf{F})}{P(\mathbf{W} = \mathbf{T} \cap \mathbf{S} = \mathbf{T} \cap \mathbf{R} = \mathbf{T}) + P(\mathbf{W} = \mathbf{T} \cap \mathbf{S} = \mathbf{T} \cap \mathbf{R} = \mathbf{F}) + P(\mathbf{W} = \mathbf{T} \cap \mathbf{S} = \mathbf{F} \cap \mathbf{R} = \mathbf{T}) + P(\mathbf{W} = \mathbf{T} \cap \mathbf{S} = \mathbf{F} \cap \mathbf{R} = \mathbf{F})} \\
 &= \frac{P(\mathbf{W} = \mathbf{T} \mid \mathbf{R} = \mathbf{T} \cap \mathbf{S} = \mathbf{T})P(\mathbf{R} = \mathbf{T} \cap \mathbf{S} = \mathbf{T}) + P(\mathbf{W} = \mathbf{T} \mid \mathbf{R} = \mathbf{T} \cap \mathbf{S} = \mathbf{F})P(\mathbf{R} = \mathbf{T} \cap \mathbf{S} = \mathbf{F})}{P(\mathbf{W} = \mathbf{T} \mid \mathbf{S} = \mathbf{T} \cap \mathbf{R} = \mathbf{T})P(\mathbf{S} = \mathbf{T} \cap \mathbf{R} = \mathbf{T}) + P(\mathbf{W} = \mathbf{T} \mid \mathbf{S} = \mathbf{T} \cap \mathbf{R} = \mathbf{F})P(\mathbf{S} = \mathbf{T} \cap \mathbf{R} = \mathbf{F}) + P(\mathbf{W} = \mathbf{T} \mid \mathbf{S} = \mathbf{F} \cap \mathbf{R} = \mathbf{T})P(\mathbf{S} = \mathbf{F} \cap \mathbf{R} = \mathbf{T}) + P(\mathbf{W} = \mathbf{T} \mid \mathbf{S} = \mathbf{F} \cap \mathbf{R} = \mathbf{F})P(\mathbf{S} = \mathbf{F} \cap \mathbf{R} = \mathbf{F})} \\
 &= \frac{P(\mathbf{W} = \mathbf{T} \mid \mathbf{R} = \mathbf{T} \cap \mathbf{S} = \mathbf{T})\{P(\mathbf{R} = \mathbf{T} \cap \mathbf{S} = \mathbf{T} \cap \mathbf{C} = \mathbf{T}) + P(\mathbf{R} = \mathbf{T} \cap \mathbf{S} = \mathbf{T} \cap \mathbf{C} = \mathbf{F})\} + P(\mathbf{W} = \mathbf{T} \mid \mathbf{R} = \mathbf{T} \cap \mathbf{S} = \mathbf{F})\{P(\mathbf{R} = \mathbf{T} \cap \mathbf{S} = \mathbf{F} \cap \mathbf{C} = \mathbf{T}) + P(\mathbf{R} = \mathbf{T} \cap \mathbf{S} = \mathbf{F} \cap \mathbf{C} = \mathbf{F})\}}{[P(\mathbf{W} = \mathbf{T} \mid \mathbf{S} = \mathbf{T} \cap \mathbf{R} = \mathbf{T})\{P(\mathbf{S} = \mathbf{T} \cap \mathbf{R} = \mathbf{T} \cap \mathbf{C} = \mathbf{T}) + P(\mathbf{S} = \mathbf{T} \cap \mathbf{R} = \mathbf{T} \cap \mathbf{C} = \mathbf{F})\} + P(\mathbf{W} = \mathbf{T} \mid \mathbf{S} = \mathbf{T} \cap \mathbf{R} = \mathbf{F})\{P(\mathbf{S} = \mathbf{T} \cap \mathbf{R} = \mathbf{F} \cap \mathbf{C} = \mathbf{T}) + P(\mathbf{S} = \mathbf{T} \cap \mathbf{R} = \mathbf{F} \cap \mathbf{C} = \mathbf{F})\} + P(\mathbf{W} = \mathbf{T} \mid \mathbf{S} = \mathbf{F} \cap \mathbf{R} = \mathbf{T})\{P(\mathbf{S} = \mathbf{F} \cap \mathbf{R} = \mathbf{T} \cap \mathbf{C} = \mathbf{T}) + P(\mathbf{S} = \mathbf{F} \cap \mathbf{R} = \mathbf{T} \cap \mathbf{C} = \mathbf{F})\} + P(\mathbf{W} = \mathbf{T} \mid \mathbf{S} = \mathbf{F} \cap \mathbf{R} = \mathbf{F})\{P(\mathbf{S} = \mathbf{F} \cap \mathbf{R} = \mathbf{F} \cap \mathbf{C} = \mathbf{T}) + P(\mathbf{S} = \mathbf{F} \cap \mathbf{R} = \mathbf{F} \cap \mathbf{C} = \mathbf{F})\}]} \\
 &= \frac{P(\mathbf{W} = \mathbf{T} \mid \mathbf{R} = \mathbf{T} \cap \mathbf{S} = \mathbf{T})\{P(\mathbf{R} = \mathbf{T} \cap \mathbf{S} = \mathbf{T} \mid \mathbf{C} = \mathbf{T})P(\mathbf{C} = \mathbf{T}) + P(\mathbf{R} = \mathbf{T} \cap \mathbf{S} = \mathbf{T} \mid \mathbf{C} = \mathbf{F})P(\mathbf{C} = \mathbf{F})\} + P(\mathbf{W} = \mathbf{T} \mid \mathbf{R} = \mathbf{T} \cap \mathbf{S} = \mathbf{F})\{P(\mathbf{R} = \mathbf{T} \cap \mathbf{S} = \mathbf{F} \mid \mathbf{C} = \mathbf{T})P(\mathbf{C} = \mathbf{T}) + P(\mathbf{R} = \mathbf{T} \cap \mathbf{S} = \mathbf{F} \mid \mathbf{C} = \mathbf{F})P(\mathbf{C} = \mathbf{F})\}}{[P(\mathbf{W} = \mathbf{T} \mid \mathbf{S} = \mathbf{T} \cap \mathbf{R} = \mathbf{T})\{P(\mathbf{S} = \mathbf{T} \cap \mathbf{R} = \mathbf{T} \mid \mathbf{C} = \mathbf{T})P(\mathbf{C} = \mathbf{T}) + P(\mathbf{S} = \mathbf{T} \cap \mathbf{R} = \mathbf{T} \mid \mathbf{C} = \mathbf{F})P(\mathbf{C} = \mathbf{F})\} + P(\mathbf{W} = \mathbf{T} \mid \mathbf{S} = \mathbf{T} \cap \mathbf{R} = \mathbf{F})\{P(\mathbf{S} = \mathbf{T} \cap \mathbf{R} = \mathbf{F} \mid \mathbf{C} = \mathbf{T})P(\mathbf{C} = \mathbf{T}) + P(\mathbf{S} = \mathbf{T} \cap \mathbf{R} = \mathbf{F} \mid \mathbf{C} = \mathbf{F})P(\mathbf{C} = \mathbf{F})\} + P(\mathbf{W} = \mathbf{T} \mid \mathbf{S} = \mathbf{F} \cap \mathbf{R} = \mathbf{T})\{P(\mathbf{S} = \mathbf{F} \cap \mathbf{R} = \mathbf{T} \mid \mathbf{C} = \mathbf{T})P(\mathbf{C} = \mathbf{T}) + P(\mathbf{S} = \mathbf{F} \cap \mathbf{R} = \mathbf{T} \mid \mathbf{C} = \mathbf{F})P(\mathbf{C} = \mathbf{F})\} + P(\mathbf{W} = \mathbf{T} \mid \mathbf{S} = \mathbf{F} \cap \mathbf{R} = \mathbf{F})\{P(\mathbf{S} = \mathbf{F} \cap \mathbf{R} = \mathbf{F} \mid \mathbf{C} = \mathbf{T})P(\mathbf{C} = \mathbf{T}) + P(\mathbf{S} = \mathbf{F} \cap \mathbf{R} = \mathbf{F} \mid \mathbf{C} = \mathbf{F})P(\mathbf{C} = \mathbf{F})\}]} \\
 &= \frac{[P(\mathbf{W} = \mathbf{T} \mid \mathbf{R} = \mathbf{T} \cap \mathbf{S} = \mathbf{T})\{P(\mathbf{R} = \mathbf{T} \mid \mathbf{C} = \mathbf{T})P(\mathbf{S} = \mathbf{T} \mid \mathbf{C} = \mathbf{T})P(\mathbf{C} = \mathbf{T}) + P(\mathbf{R} = \mathbf{T} \mid \mathbf{C} = \mathbf{F})P(\mathbf{S} = \mathbf{T} \mid \mathbf{C} = \mathbf{F})P(\mathbf{C} = \mathbf{F})\} + P(\mathbf{W} = \mathbf{T} \mid \mathbf{R} = \mathbf{T} \cap \mathbf{S} = \mathbf{F})\{P(\mathbf{R} = \mathbf{T} \mid \mathbf{C} = \mathbf{T})P(\mathbf{S} = \mathbf{F} \mid \mathbf{C} = \mathbf{T})P(\mathbf{C} = \mathbf{T}) + P(\mathbf{R} = \mathbf{T} \mid \mathbf{C} = \mathbf{F})P(\mathbf{S} = \mathbf{F} \mid \mathbf{C} = \mathbf{F})P(\mathbf{C} = \mathbf{F})\}]}{[P(\mathbf{W} = \mathbf{T} \mid \mathbf{S} = \mathbf{T} \cap \mathbf{R} = \mathbf{T})\{P(\mathbf{S} = \mathbf{T} \mid \mathbf{C} = \mathbf{T})P(\mathbf{R} = \mathbf{T} \mid \mathbf{C} = \mathbf{T})P(\mathbf{C} = \mathbf{T}) + P(\mathbf{S} = \mathbf{T} \mid \mathbf{C} = \mathbf{F})P(\mathbf{R} = \mathbf{T} \mid \mathbf{C} = \mathbf{F})P(\mathbf{C} = \mathbf{F})\} + P(\mathbf{W} = \mathbf{T} \mid \mathbf{S} = \mathbf{T} \cap \mathbf{R} = \mathbf{F})\{P(\mathbf{S} = \mathbf{T} \mid \mathbf{C} = \mathbf{T})P(\mathbf{R} = \mathbf{F} \mid \mathbf{C} = \mathbf{T})P(\mathbf{C} = \mathbf{T}) + P(\mathbf{S} = \mathbf{T} \mid \mathbf{C} = \mathbf{F})P(\mathbf{R} = \mathbf{F} \mid \mathbf{C} = \mathbf{F})P(\mathbf{C} = \mathbf{F})\} + P(\mathbf{W} = \mathbf{T} \mid \mathbf{S} = \mathbf{F} \cap \mathbf{R} = \mathbf{T})\{P(\mathbf{S} = \mathbf{F} \mid \mathbf{C} = \mathbf{T})P(\mathbf{R} = \mathbf{T} \mid \mathbf{C} = \mathbf{T})P(\mathbf{C} = \mathbf{T}) + P(\mathbf{S} = \mathbf{F} \mid \mathbf{C} = \mathbf{F})P(\mathbf{R} = \mathbf{T} \mid \mathbf{C} = \mathbf{F})P(\mathbf{C} = \mathbf{F})\} + P(\mathbf{W} = \mathbf{T} \mid \mathbf{S} = \mathbf{F} \cap \mathbf{R} = \mathbf{F})\{P(\mathbf{S} = \mathbf{F} \mid \mathbf{C} = \mathbf{T})P(\mathbf{R} = \mathbf{F} \mid \mathbf{C} = \mathbf{T})P(\mathbf{C} = \mathbf{T}) + P(\mathbf{S} = \mathbf{F} \mid \mathbf{C} = \mathbf{F})P(\mathbf{R} = \mathbf{F} \mid \mathbf{C} = \mathbf{F})P(\mathbf{C} = \mathbf{F})\}]} \\
 &= \frac{0.99 \times (0.8 \times 0.1 \times 0.5 + 0.2 \times 0.5 \times 0.5) + 0.9 \times (0.8 \times 0.9 \times 0.5 + 0.2 \times 0.5 \times 0.5)}{0.99 \times (0.1 \times 0.8 \times 0.5 + 0.5 \times 0.2 \times 0.5) + 0.9 \times (0.1 \times 0.2 \times 0.5 + 0.5 \times 0.8 \times 0.5) + 0.9 \times (0.9 \times 0.8 \times 0.5 + 0.5 \times 0.2 \times 0.5) + 0 \times (0.9 \times 0.2 \times 0.5 + 0.5 \times 0.8 \times 0.5)} \\
 &= \frac{0.99 \times (0.04 + 0.05) + 0.9 \times (0.36 + 0.05)}{0.99 \times (0.04 + 0.05) + 0.9 \times (0.01 + 0.20) + 0.9 \times (0.36 + 0.05) + 0 \times (0.09 + 0.20)} \\
 &= \frac{0.0891 + 0.369}{0.0891 + 0.189 + 0.369 + 0} \\
 &= \frac{0.4581}{0.6471} \approx 0.70793
 \end{aligned}$$

To calculate the posterior probability, the equations in the denominator and numerator should be converted to those in CPT. CPT expresses the relationship between a node and its parent node. Consequently, the probability between the node and the parent node should be marginalized. For example, "P(R = T ∩ W = T ∩ S = T)" can be expressed as "P(W = T | R = T ∩ S = T)P(R = T ∩ S = T)" according to Bayes' rule. R and S are parents of W. "P(R = T ∩ S = T ∩ C = T)" can be expressed as "P(R = T ∩ S = T | C = T)P(C = T)" according to Bayes' rule. C is a common

parent of R and S. " $P(R = T \cap S = T \mid C = T)P(C = T)$ " can be represented as " $P(R = T \mid C = T)P(S = T \mid C = T)P(C = T)$ ", due to the conditional independence between the child nodes R and S when the parent node C is determined as true.

This calculation process involves marginalization in the Bayesian network consisting of 4 nodes and 3 layers, shown in Fig. 1a. Meanwhile, in a complex Bayesian network such as the Bayesian network consisting of 20 nodes and 7 layers, shown in Fig. 7a, the marginalization process becomes exponentially intricate. In particular, exponentially more marginalization and multiplication processes are required in the Bayesian inference as the two nodes are farther from the common ancestor. Consequently, the Bayesian inference with conventional hardware requires a substantial amount of calculations attributed to floating-point calculations.



**Note S3.** Calculations of the power and energy consumption.

First, the power consumption of the devices (LFSR, MTJ, and SiO<sub>x</sub> nanorod) was calculated based on the values provided in the previous works.<sup>1-3</sup> As the measured energy values are given in the previous works, the device operating speed was considered to calculate the power. For the MTJ device, the power consumption was calculated most conservatively, where the bitstream number of 256 is assumed.<sup>2</sup> Furthermore, it is important to highlight that, to ensure a fair comparison, the technological node of the CMOS was consistently assumed to be 180 nm. As a result, the power consumption of the LFSR in the previous work was scaled regarding this assumption.<sup>4</sup>

Then, the energy for representing a single node in a Bayesian network was calculated. For the energy consumption of the peripheral CMOS circuit, such as MUX, register, and comparator, the work of Yi et al. was referred to.<sup>5</sup> As in the previous work, a 1.2 GHz clock frequency was assumed for all the devices to take parasitic elements in the measurement system. Meanwhile, the energy consumption of a 1-bit register (D flip-flop) was calculated by dividing the energy consumption of the I/O register by its number of bits.

## References

- 1 S. Choi, G. S. Kim, J. Yang, H. Cho, C. Y. Kang and G. Wang, *Advanced Materials*, DOI:10.1002/adma.202104598.
- 2 X. Jia, J. Yang, Z. Wang, Y. Chen, H. H. Li and W. Zhao, in *Proceedings of the Asia and South Pacific Design Automation Conference, ASP-DAC*, 2018.
- 3 J. Kaiser and S. Datta, *Appl Phys Lett*, 2021.
- 4 W. A. Borders, A. Z. Pervaiz, S. Fukami, K. Y. Camsari, H. Ohno and S. Datta, *Nature*, DOI:10.1038/s41586-019-1557-9.
- 5 S. in Yi, J. D. Kendall, R. S. Williams and S. Kumar, *Nat Electron*, DOI:10.1038/s41928-022-00869-w.