Supplementary Information:

Unveiling ECRAM Switching Mechanism using Variable Temperature Hall Measurements for Accelerated AI Computation

Hyunjeong Kwak[1], Junyoung Choi[1], Eun Ho Kim[1], Chaeyoun Kim[2], Paul Solomon[3], Junyong Lee[1], Doyoon Kim[1], Byungha Shin[2], Donghwa Lee[1], Oki Gunawan[3*] and Seyoung Kim[1*]

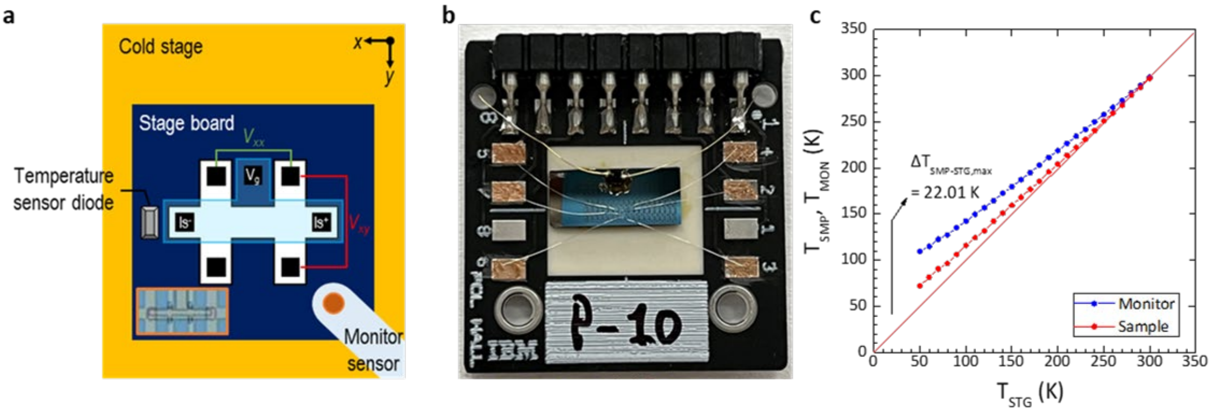
[1]Department of Materials Science and Engineering, Pohang University of Science and Technology, Pohang 37673, South Korea

[2]Department of Materials Science and Engineering, Korea Advanced Institute of Science and Technology, Daejeon 34141, South Korea

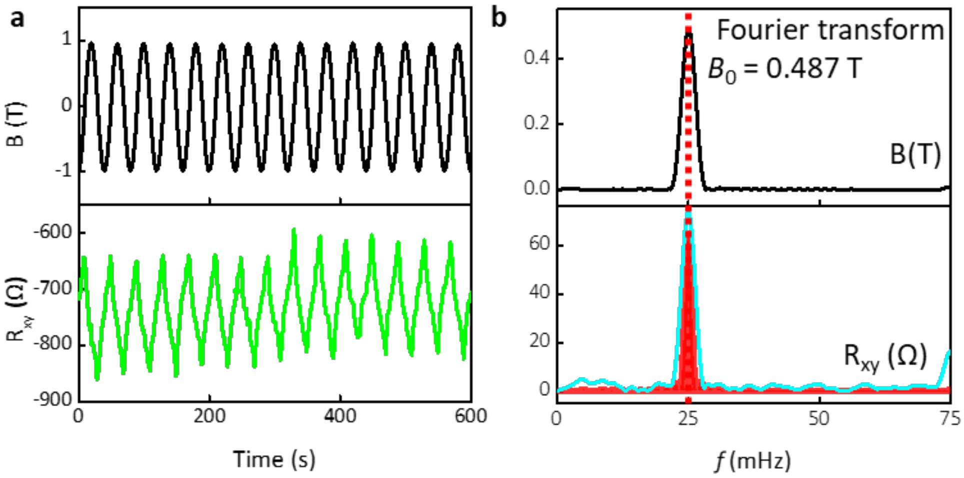
[3]IBM T.J. Watson Research Center, Yorktown Heights, New York 10598, USA

[*]kimseyoung@postech.ac.kr, ogunawa@us.ibm.com

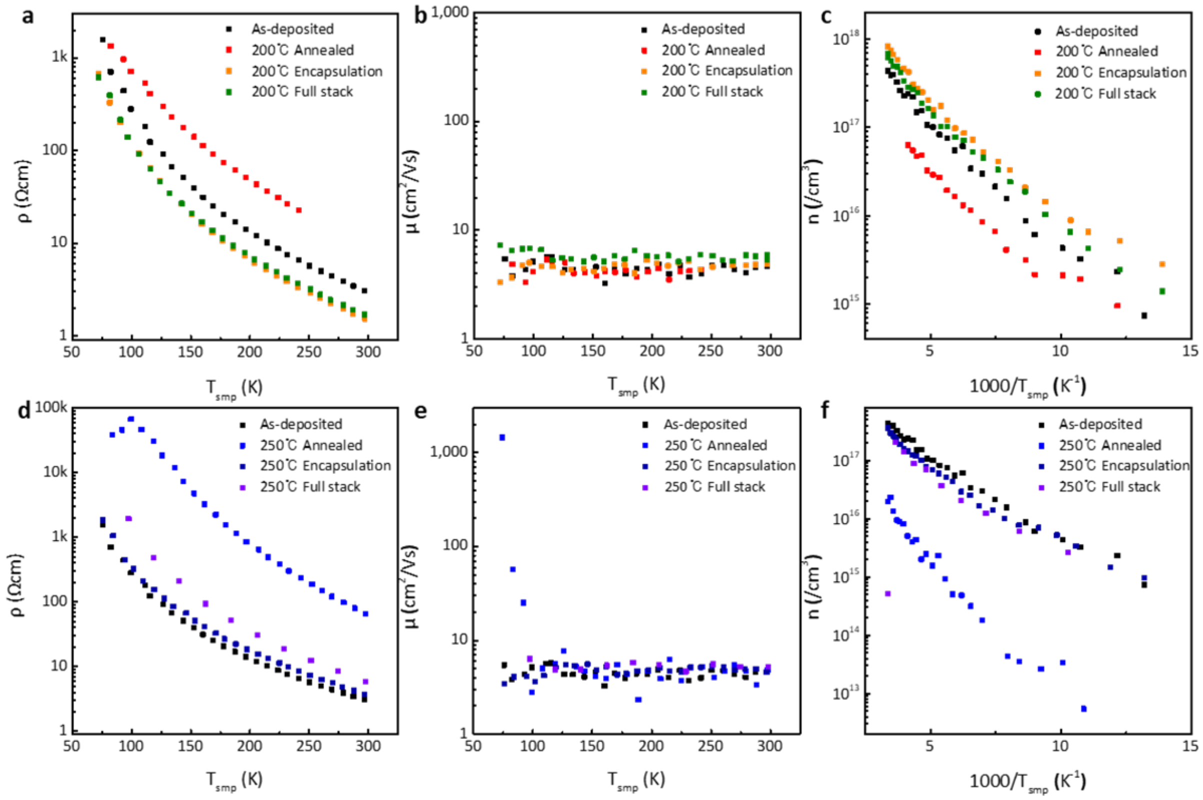
Supplementary figures



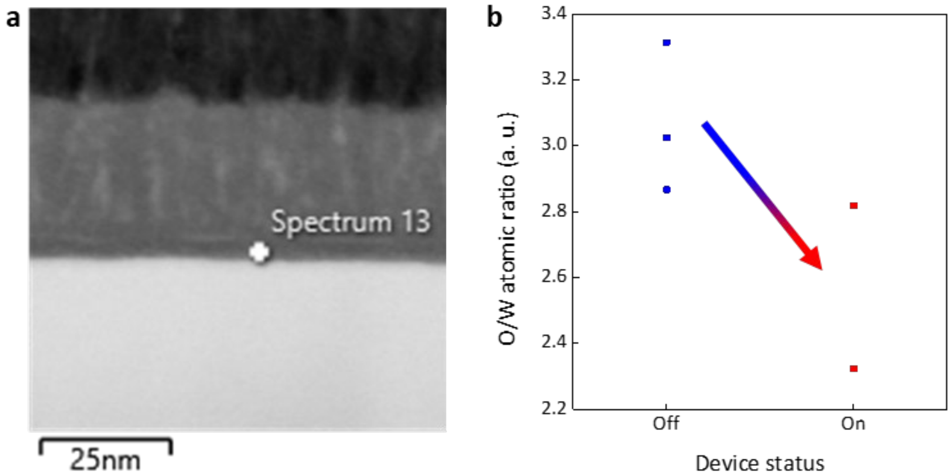
Supplementary Figure 1. Sample preparation and results for accurate temperature control and measurement. a Schematic of sample designed for variable-temperature PDL Hall measurement. A 1x1 cm^2 wafer is attached on a PCB substrate. The multi-terminal Hall bar structured ECRAM is deposited on silicon wafer, with a detail image resembling the inset optical image. A temperature diode sensor is attached to the silicon wafer substrate, and the PCB board is aligned with a monitor sensor. The stage board is fixed on the cold stage to control the temperature. **b** Photo of PCB board and prepared sample. Device pads on the wafer are wire bonded to outer pads, and measurements are conducted via SIP pins connected to the pads. **c** Temperature calibration graph. It depicts the applied temperature (Stage, STG), the temperature of device measured through the diode (Sample, SMP), and the temperature of the PCB board measured through the monitor sensor (Monitor, MON). Analyzing the gaps between these three temperatures enables accurate analysis of the Hall coefficient for the device. As the temperature decreases, the differences become more significant, necessitating more precise measurements.



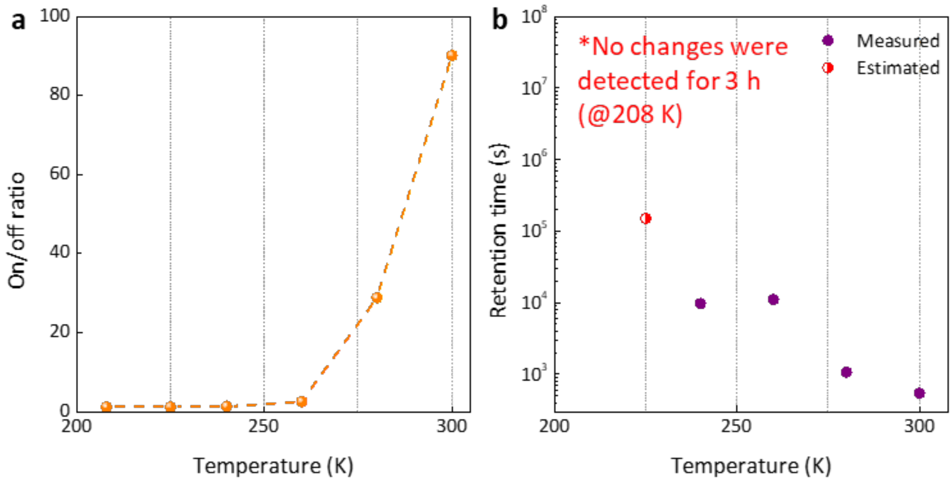
Supplementary Figure 2. Hall signal analysis with lock-in detection. a Applied magnetic field and magnetoresistance sweep (R_{xy}). **b** Fourier transform and lock-in detection of the Hall signal. The R_{xy} signal shows a clear and robust Fourier peak at *f*(red dotted line), which produces a steady lock-in output R_H.



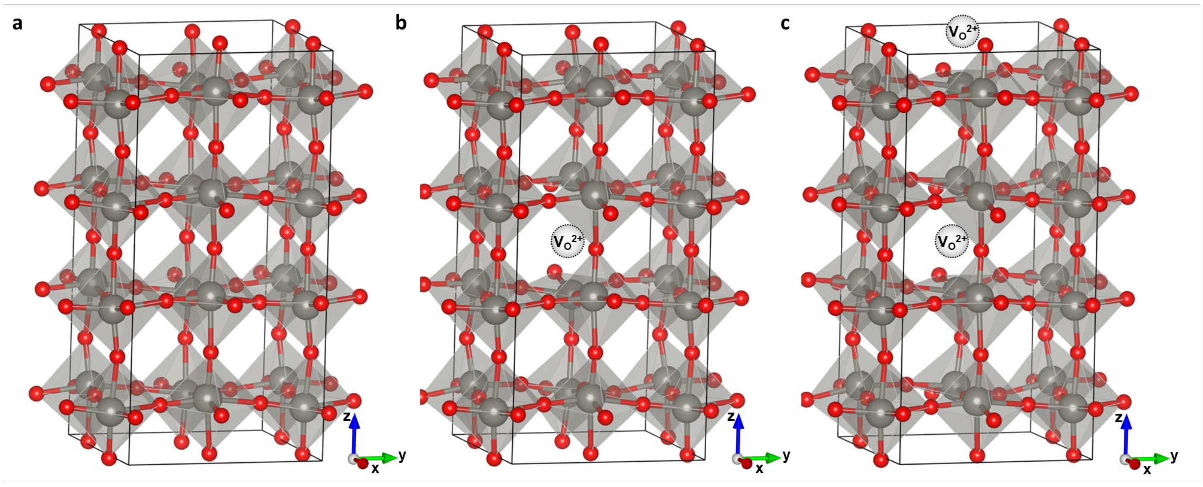
Supplementary Figure 3. ECRAM Hall measurement results at various conditions. Results of Hall measurements performed according to the fabrication process: [1]Samples with only depositing using a WO_3 target. [2]Sample with channel deposition followed by annealing at 200 °C (or 250 °C) for 75 minutes to observe the difference due to annealing. [3]Sample with HfO_2 electrolyte deposition on the channel-deposited sample. [4]Sample with deposition up to tungsten oxide reservoir and W gate electrode. **a-c** includes annealing at 200 °C, while **d-f** include annealing at 250 °C. Clear differences in resistivity and carrier density due to annealing are evident, but upon deposition of the full stack, values tend to recover to levels similar to the initial ones. **a, d** Resistivity. **b, e** Carrier mobility. **c, f** Carrier density and activation energy.



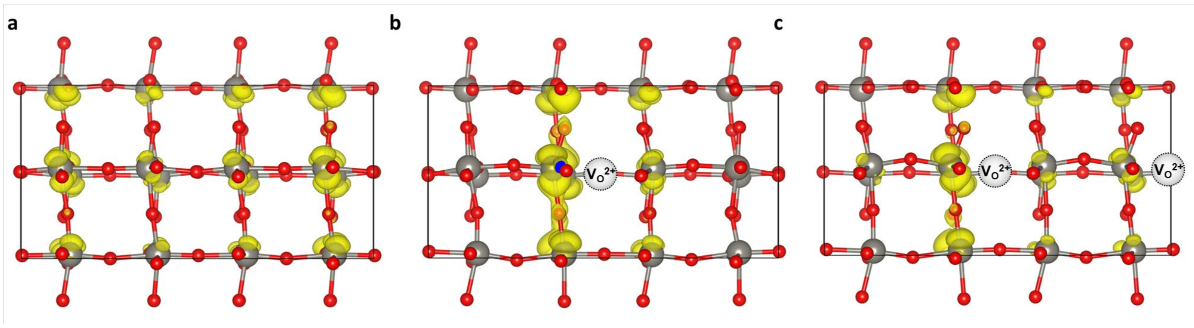
Supplementary Figure 4. EDS analysis for tracking changes of oxygen amount between on and off states. a After switching the device to the on state, a cross-sectional sampling was performed using focused ion beam (FIB), followed by transmission electron microscope image. **b** Energy dispersive X-ray spectrometer (EDS) analysis of the device in both on (high conductivity) and off (low conductivity) states. The spectra analysis in the channel region allowed for the calculation of the atomic percentages of W and O. It was observed that the oxygen ratio decreased when transitioning from the off to on state.



Supplementary Figure 5. Detailed analysis of switching and retention behaviors depending on temperature. a Temperature-dependent switching characteristics. The on/off ratio sharply decreases with decreasing temperature, with minimal change observed below 240 K, where it remains slightly above 1. **b** Temperature-dependent retention characteristics. Retention significantly increases with decreasing temperature, with no observable change at 208 K. Retention values were experimentally confirmed for temperatures ranging from 300 K to 240 K, measured over a period of 3 hours. Below 220 K, retention values were extrapolated using an exponential decay function based on the 3-hour measurements.



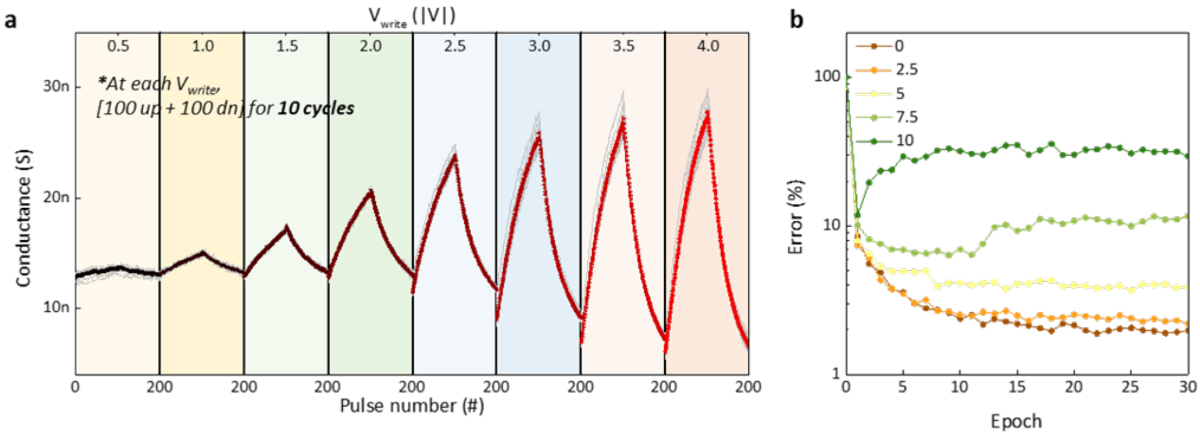
Supplementary Figure 6. Geometry-optimized structure of a WO_3. **b** $WO_{2.938}$. **c** $WO_{2.875}$. Oxygen vacancy sites are marked as V_O^{2+}.



Supplementary Figure 7. Spatial distribution of the CBM states in a WO_3. **b** $WO_{2.938}$. **c** $WO_{2.875}$. The iso-surface level is set 0.0002 e/Å^3.

Computational details

The Vienna Ab initio Simulation Package (VASP) served as the computational framework for all density functional theory (DFT) calculations conducted in this study[1]. Electron-ion interactions were modeled using the plane-wave projector augmented wave (PAW) method[2]. Exchange-correlation effects were treated within the framework of the Perdew–Burke–Ernzerhof for solids (PBEsol) functional[3]. The plane wave cutoff energy was consistently set to 520 eV across all calculations. The geometrically optimized lattice parameters of the monoclinic WO_3 are determined to be a = 7.46 Å, b = 7.60 Å, c = 7.73 Å, β = 90.2°, closely match experimental parameters, exhibiting a deviation of less than 2%[4]. The Brillouin zone was sampled using a 4 × 4 × 4 Monkhorst-pack k-point grid for unit cell calculations and 4 × 4 × 2 Monkhorst-pack k-point grid for (1 × 1 × 2) supercell calculations. To ensure accurate results, structural relaxation was performed until both energy and forces met convergence criteria, set at 10^{-6} eV for energy and 0.01 eV/Å for the forces acting on the atoms. For HSE06 calculations, atomic coordinates are fixed to the geometrically optimized positions obtained from PBEsol. We opted for a 28% intermixing of Hartree-Fock exact exchange, slightly deviating from the default value of 25%, because it matches well with the experimental value of WO_3 band gap.



Supplementary Figure 8. Cycle to cycle variation and its impact on neural network training. a The cycle-to-cycle variation of a three-terminal ECRAM device, identical to those measured in Hall measurements. Switching characteristics were measured from ± 0.5 V to ± 4.0 V in 0.5 V increments. At each write voltage, 100 pulses of potentiation and 100 pulses of depression with pulse width of 0.5s were repeated in sets of 10 cycles. As the write voltage increased, the cycle-to-cycle variation improved. Furthermore, the data was obtained through continuous measurements, validating an endurance of 16,000. **b** Neural network training error as a function of cycle-to-cycle variation. It was observed that when the cycle-to-cycle variation exceeds 5 %, the training accuracy sharply decreases.

Simulation details

To assess the impact of cycle-to-cycle variation on neural network training, we employed linear regression using a stochastic gradient descent (SGD) algorithm to simulate pattern recognition accuracy with the MNIST dataset. Supplementary Figure 8b displays the results of evaluating recognition accuracy (error rate). The neural network used in the simulation is a multilayer perceptron, comprising input, first hidden, second hidden, and output layers with 784, 256, 128, and 10 neurons, respectively. The 784 input neurons represent 28x28

MNIST image data, while the 10 output neurons represent digits from 0 to 9. Weight updates were executed with a learning rate of 0.01, a batch size of 1, and 30 epochs.

The evaluation of recognition accuracy revealed ideal synaptic characteristics, achieving 98.02% accuracy in the ideal scenario (0% variation). However, as cycle-to-cycle variation increased, accuracy decreased slightly to 97.8% and 96.13% at 2.5% and 5% variation, respectively. Accuracy dropped significantly to 88.31% at 7.5% and 70.63% at 10% variation. Notably, ECRAM demonstrates minimal cycle-to-cycle variation between 0% and 2.5%, making it a promising candidate for AI accelerators with high neural network training accuracy.

References

1. Kresse G, Furthmüller J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Physical review B* **54**, 11169 (1996).
2. Blöchl PE. Projector augmented-wave method. *Physical review B* **50**, 17953 (1994).
3. Perdew JP*, et al.* Restoring the density-gradient expansion for exchange in solids and surfaces. *Physical review letters* **100**, 136406 (2008).
4. Loopstra B, Rietveld H. Further refinement of the structure of WO3. *Acta Crystallographica Section B: Structural Crystallography and Crystal Chemistry* **25**, 1420-1421 (1969).