Electronic Supplementary Material (ESI) for Materials Horizons.
This journal is © The Royal Society of Chemistry 2024

Supporting Information

Heterogeneous Density-based Clustering with Dual-functional Memristive Array

Dong Hoon Shin[1,†], Sunwoo Cheong[1,†], Soo Hyung Lee[1], Yoon Ho Jang[1], Taegyun Park[1], Janguk Han[1], Sung Keun Shim[1], Yeong Rok Kim[1], Joon-Kyu Han[1], In Kyung Baek[1], Néstor Ghenzi[1,2]*, and Cheol Seong Hwang[1]*

[†] These authors contributed equally to this work.

[*] Corresponding author (e-mail: *n.ghenzi@gmail.com, cheolsh@snu.ac.kr*)

[1] Department of Materials Science and Engineering and Inter-University Semiconductor Research Center, Seoul National University, Gwanak-ro 1, Gwanak-gu, Seoul 08826, Republic of Korea

[2] Universidad de Avellaneda UNDAV and Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)
Mario Bravo 1460, Avellaneda, Buenos Aires 1872, Argentina

<u>Table of contents</u>

- Supporting Note 1-4

- Supporting Figures S1-S26

- Supporting Tables S1-S4

- Supporting References

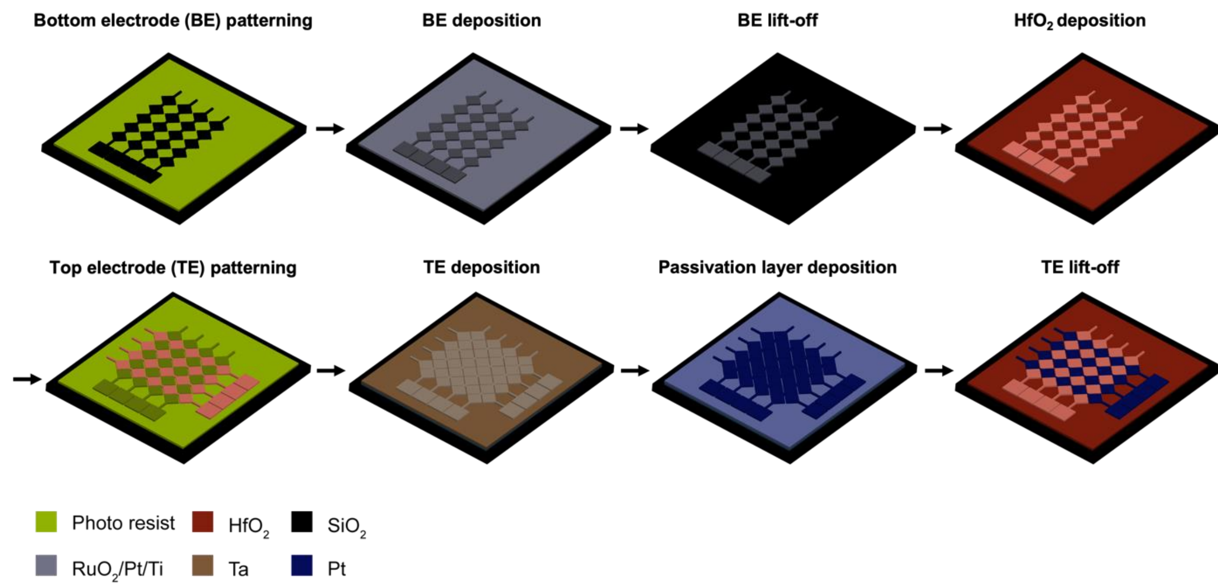

Figure S1. The fabrication process of the THR memristor crossbar array.

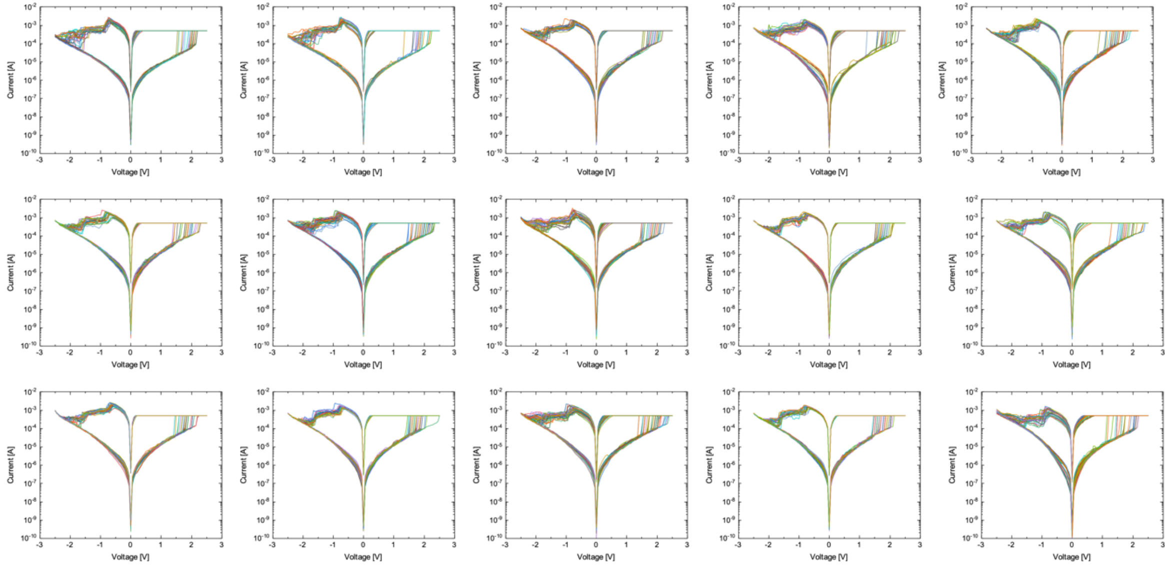

Figure S2. Collected I-V curves of 15 devices randomly selected from 9 by 9 THR CBA with 4 μm line width. Each plot contains 30 I-V curves, including the first cycles in the pristine state, showing the electroforming-free property and minimal device-to-device and cycle-to-cycle variations.

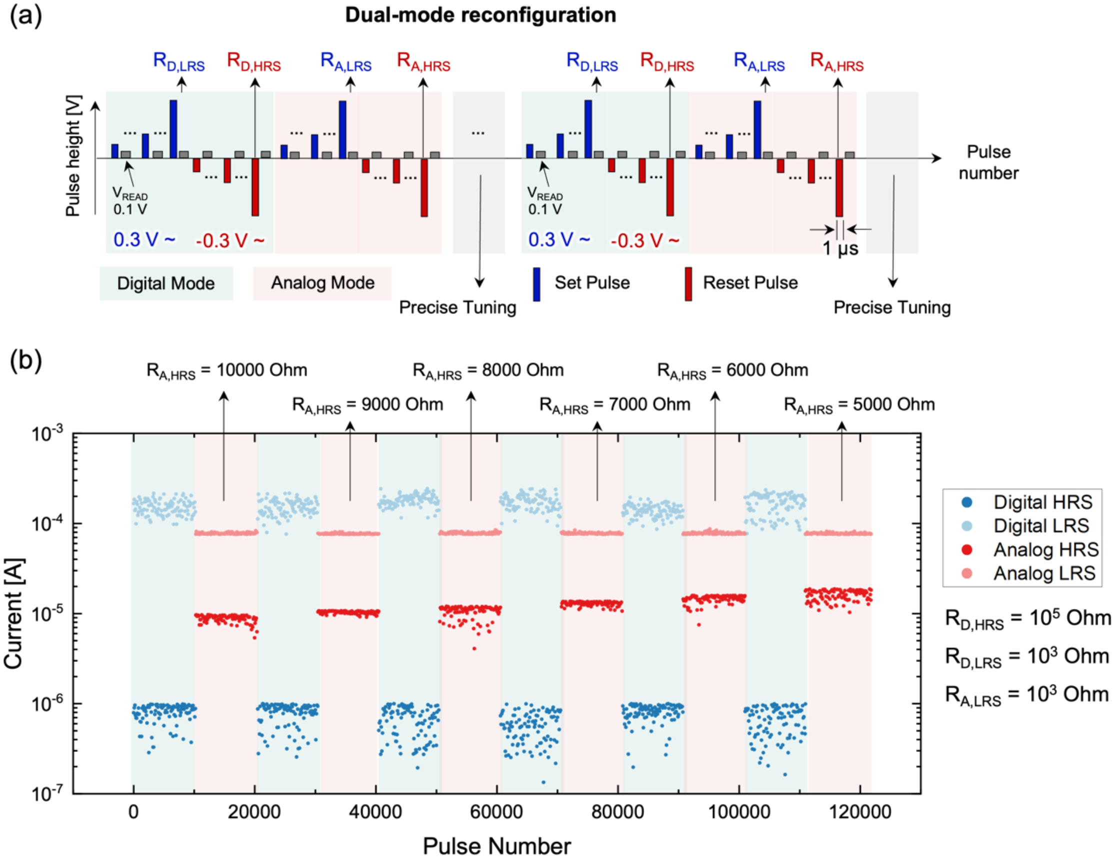

Figure S3. Pulsing application method and results of mode reconfigurations. (a) Pulsing scheme for dual-mode reconfiguration, where each digital mode and analog mode is repeated for 10^4 cycles. Here, the precise tuning scheme in gray color is employed to confirm the tunability of 5 states at 20 μS intervals between every mode reconfiguration cycles. In digital mode, incremental pulses are applied to cycle between $R_{D,LRS}$ (10^3 Ohm) and $R_{D,HRS}$ (10^5 Ohm), while in analog mode, pulses cycle between $R_{A,LRS}$ (10^3 Ohm) and $R_{A,HRS}$. To demonstrate the conductance tunability of the analog mode, the value of $R_{A,HRS}$ is varied from 10,000 Ohm to 5,000 Ohm with 1,000 Ohm interval after every 10^4 cycles. (b) The read current values (V_{READ} = 0.1 V) for HRS and LRS. In digital mode, due to the abrupt set behavior, reached $R_{D,LRS}$ is measured as a value smaller than the predetermined 10^3 Ohms, exhibiting a memory window greater than 10^2. Conversely, in analog mode, stable switching between the predetermined $R_{A,HRS}$ and $R_{A,LRS}$ values is observed, confirming the analog mode's conductance tunability.

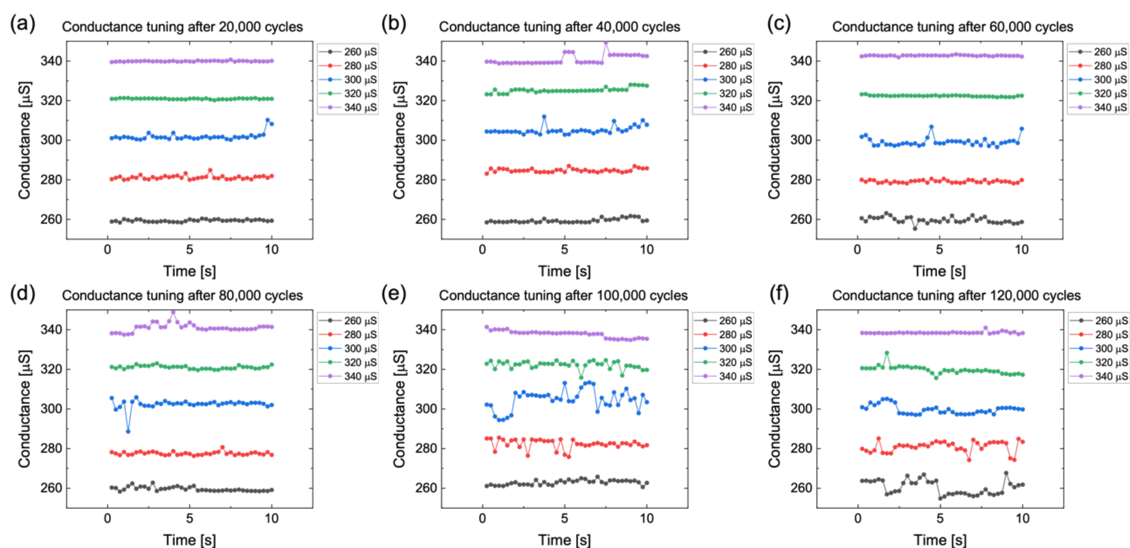

Figure S4. Conductance reading results (V_{READ} = 0.1 V) after each analog mode switching cycle after applying the precise tuning method (coarse-tuning, fine-tuning, and denoising). 5 states, spaced at 20 µS intervals (260 µS to 340 µS), are stably tuned. (a-f) 5 states, spaced at 20 µS intervals (260 µS to 340 µS), are stably tuned after switching cycles of (a) 20,000, (b) 40,000, (c) 60,000, (d) 80,000, (e) 100,000, (f) 120,000.

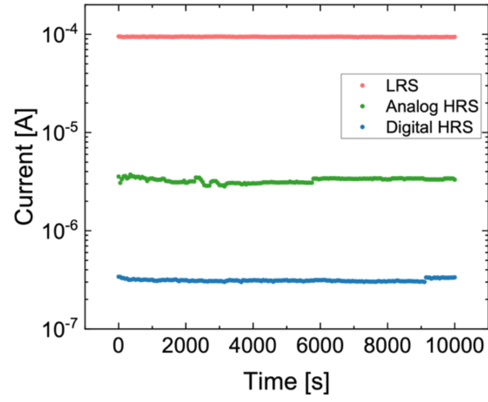

Figure S5. Retention characteristics of THR memristor at the temperature of 85 ^{o}C. HRS states for both modes, and LRS retain their states for more than 10^4 seconds without distinct degradation or state shift.

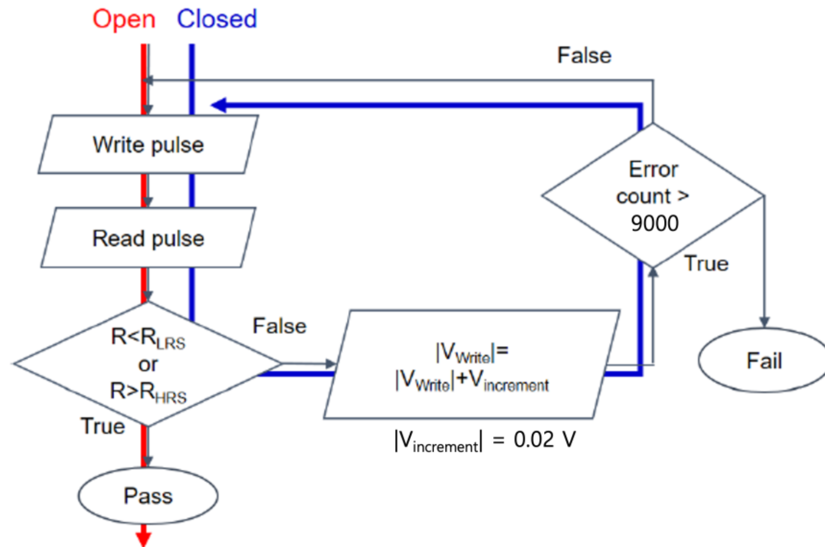

Figure S6. Flowchart of CLPS method[1–3] for the device switching endurance measurement. The custom-built field-programmable gate array board applies incremental pulses with predefined length and steps until the target resistance (R_{LRS} and R_{HRS}) is achieved. For the device set, positive incremental pulses are applied from 0.3 V step of 0.02 V until the device reaches the R_{LRS}, selected 1000 ~ 1200 Ohm for this work. When the device reaches the R_{LRS}, the board subsequently applies negative pulses with increasing magnitude in an absolute value starting from -0.3 V by -0.02 V step to reset the device until it reaches the target R_{HRS}, 10^5 Ohm for this work. After each program pulse, the resistance state is checked with a 0.1 V read pulse.

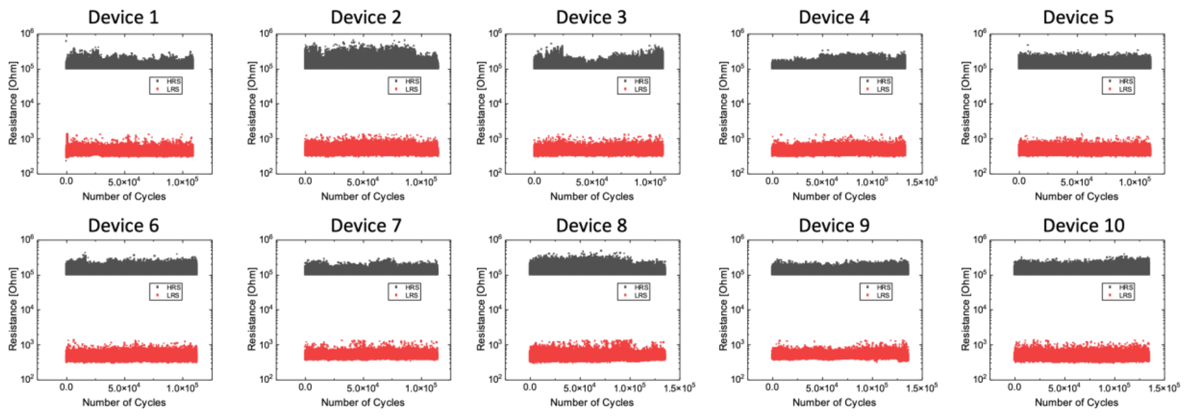

Figure S7. R_{LRS} and R_{HRS} during device switching endurance measurement of 10 THR memristors.

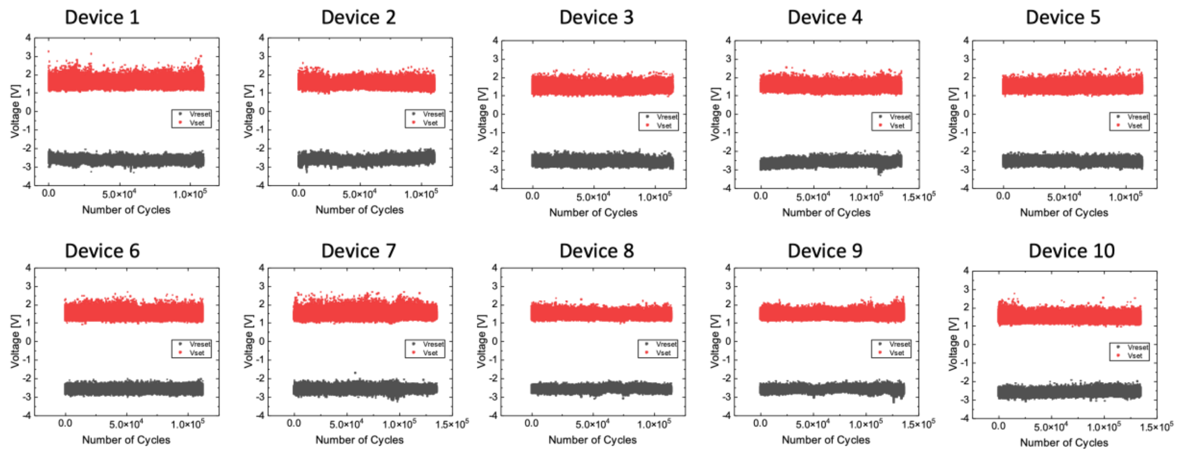

Figure S8. Switching voltages when the device reaches R_{LRS} and R_{HRS} of 10 THR memristors.

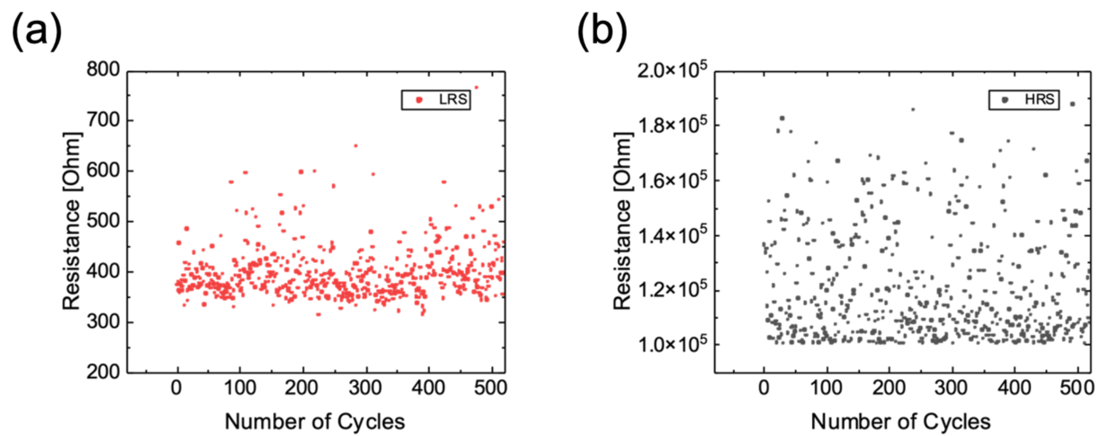

Figure S9. The magnified plot of first 500 cycles in the endurance measurement, R_{LRS} (a) and R_{HRS} (b).

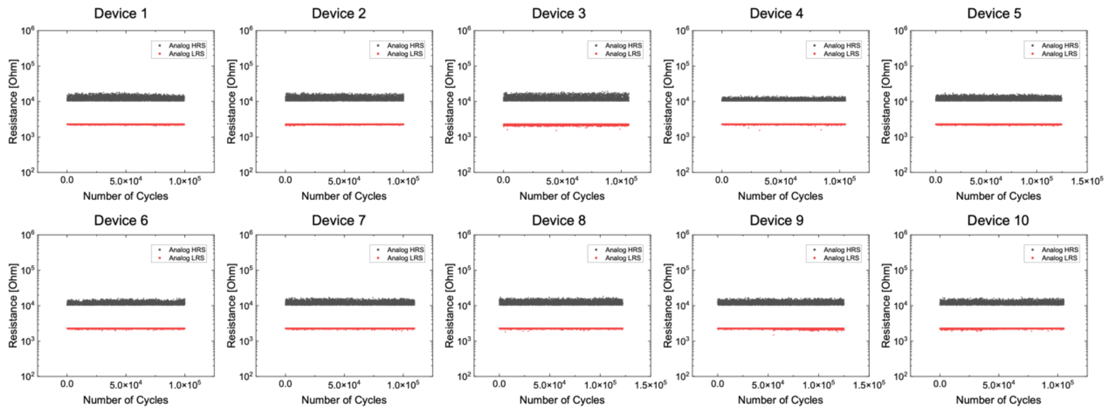

Figure S10. R_{LRS} and R_{HRS} during switching endurance of 10 different THR memristors in analog mode.

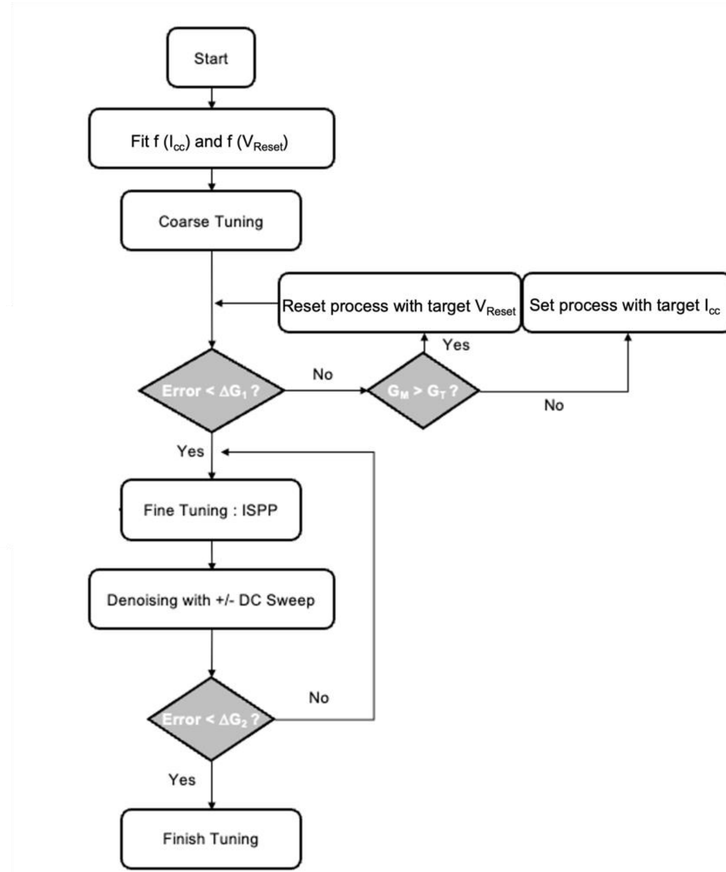

Figure S11. Algorithm of precise conductance tuning measurement. The algorithm can select the minimum and maximum conductance values (G_{MIN} and G_{MAX}) with the available memory window of the device. The three steps consist of coarse tuning, fine-tuning, and denoising steps, which follow the recent report by Rao et al..[4] First, the coarse tuning applies DC I-V sweeps to the memristor, with changing I_{cc} for set and V_{Reset} for reset to program the device with the target conductance margin (ΔG_1) of 50 µs. I_{cc} was (100 µA ~ 1000 µA), and V_{STOP} was (-0.5 V ~ -1.5 V). After the device conductance reaches the target conductance in a range of ΔG_1, the incremental step pulse programming (ISPP) algorithm was utilized to tune the conductance into a 5 µS margin finely (ΔG_2), applying 500 ns width and +0.3 V / -0.5 V height pulses with the +/- 25 mV step increment. The denoising process was optimized for state stabilization, applying repetitive I-V sweeps from 0 V to +0.3 V and 0 V to -0.3 V six times. Under these tuning processes, the THR memristor's state could be precisely modulated for the 8-bit precision (256 different states with 10 µs intervals), as in Figure 3e-f in the main text.

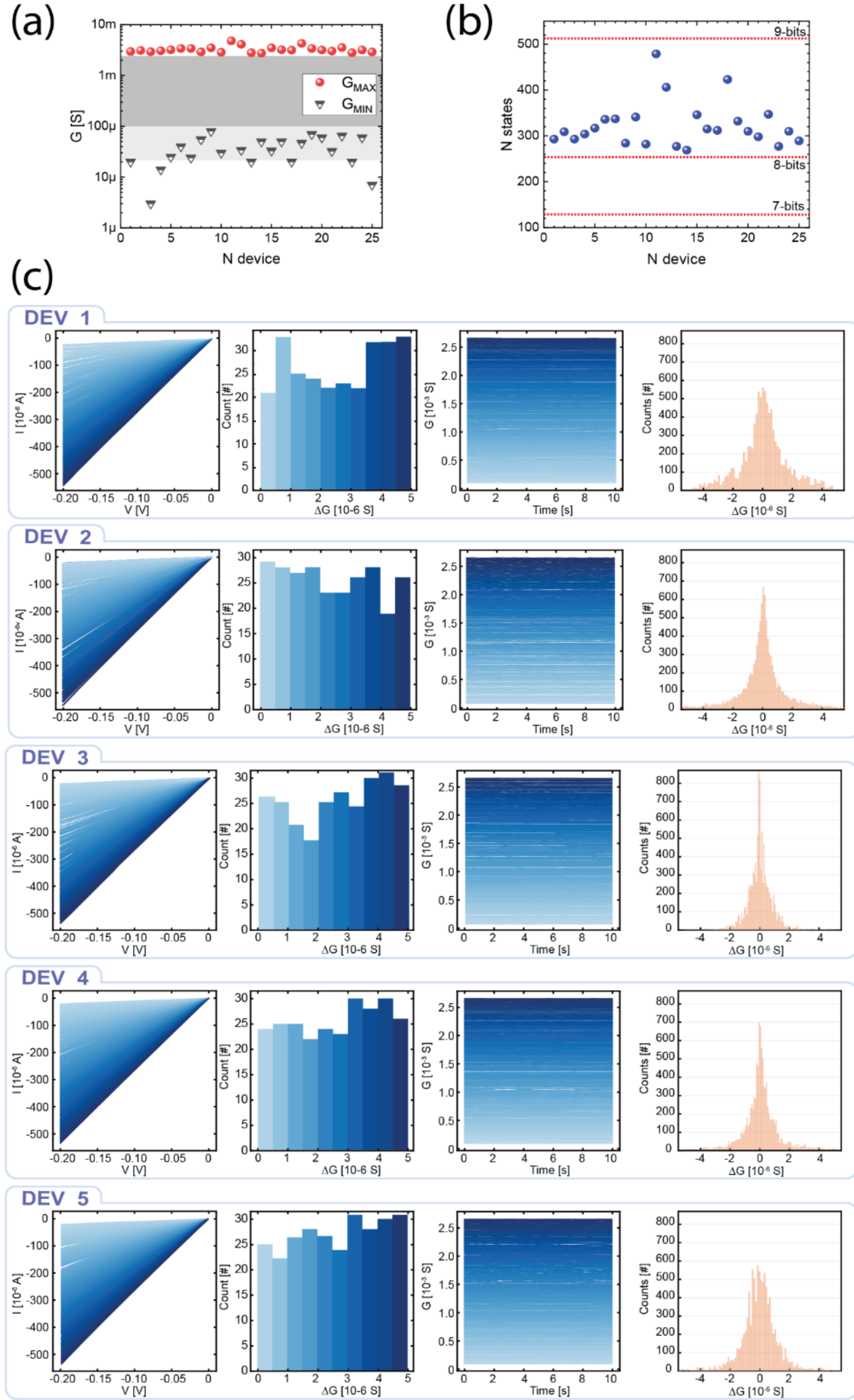

Figure S12. D2D variation analysis in the analog mode for the available conductance states analysis. Due to the reliable and repeatable tuning of the THR, minimum and maximum conductance range in analog mode can be secured for the 8-bit precision margin with a 10 µs state interval. (a) Maximum and minimum values of conductances. (b) Number of available states for 25 different THR devices. (c) Statistical analysis of the analog tuning operation for 5 representative devices. The first column shows the I-V response's linear characteristics for the 256 states. The second column shows the histogram of the error of the programmed conductance value calculated as the difference between the measured conductance in the linear

I-V response and the target value. The third column shows the retention characteristics for the 256 states. The fourth column shows the temporal deviation in the conductance value concerning the mean value for each state obtained in the retention measurements of the third column.

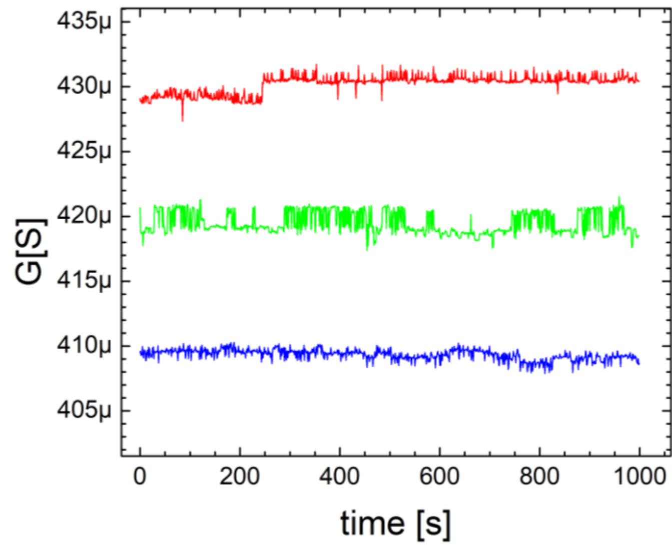

Figure S13. Retention characteristics of THR memristor after the precise conductance tuning process. The state interval was 10 μS, and the conductance state retained its mapped value without the mixing between states.

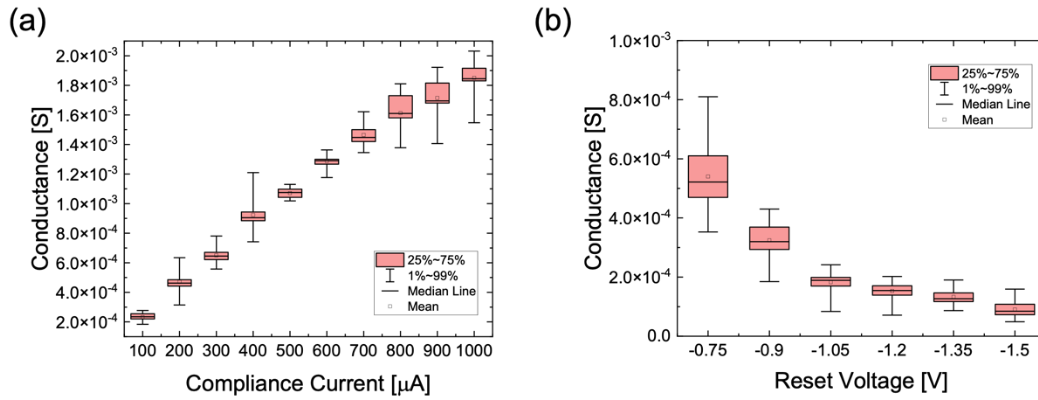

Figure S14. Effects of switching parameters on the coarse tuning process. (a) Achieved conductance during the analog mode set process as a function of compliance current (100 μA to 1000 μA). (b) Achieved conductance during the reset process as a function of reset voltage (-0.75 V to -1.5 V). Each box represents 100 repetitions. Each box represents the statistical results of 100 repetitions. Significant correlations between the switching parameters (compliance current and reset voltage) and the achieved conductance were observed during the coarse tuning process in both the set and reset directions.

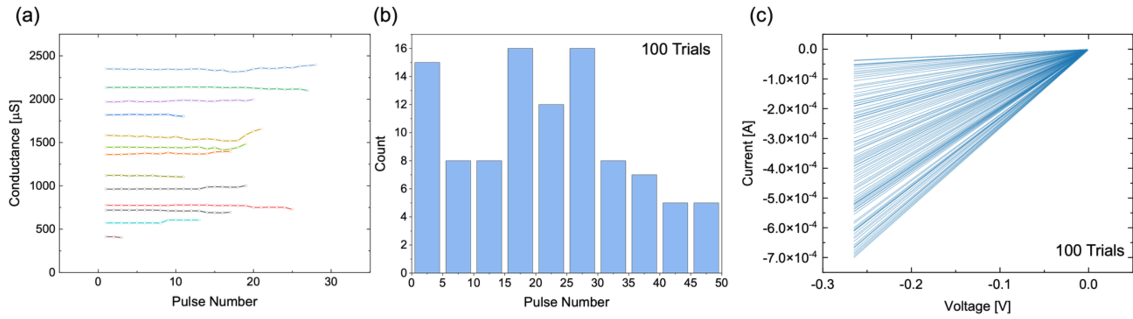

Figure S15. Results of the fine tuning process and the number of pulses necessary to iteratively write different conductance states. (a) The process of analog mapping by the precise tuning method in Figure S11. The figure shows the conductance ($V_{READ} = 0.1$ V) as a function of pulse number during the fine-tuning stage, where incremental pulses (starting from +0.3 V and -0.5 V, increasing by ±25 mV steps) are applied to reach the target conductance. (b) A histogram of the number of pulses required to reach the target conductance. The histogram shows that in 100 target conductance tuning trials, fewer than 50 pulses were required. (c) The analog conductance read results (up to -0.26 V) from these 100 trials demonstrate well-separated states and a linear resistance relationship.

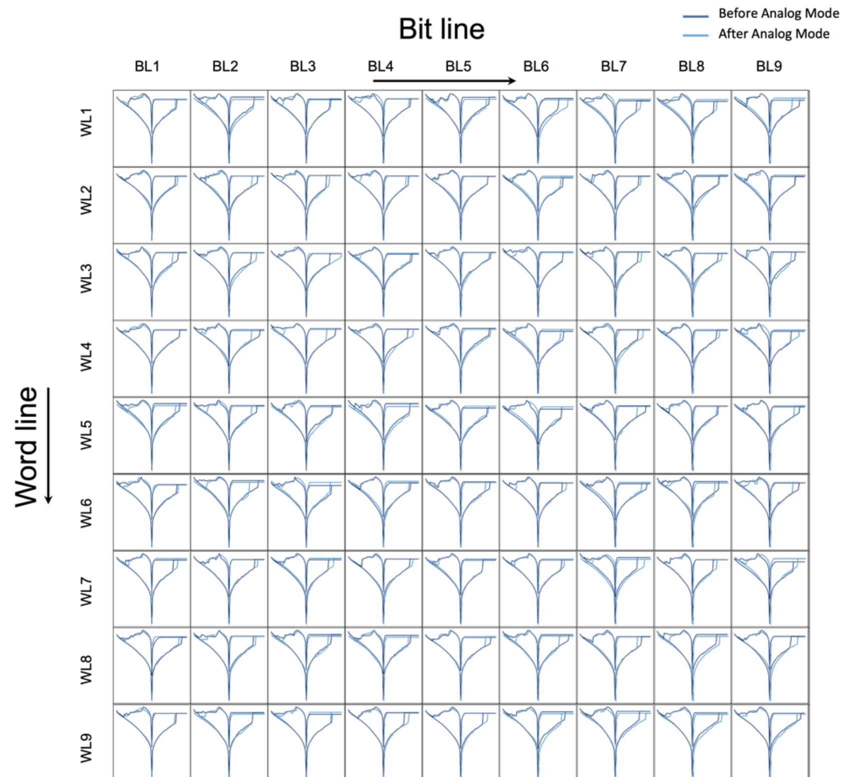

Figure S16. Digital mode I-V characteristics of 81 cells in 9 by 9 THR CBA. The dark blue lines indicate the I-V curves before the analog mode operation, while the light blue lines represent the curves after the analog mode operation. (Figure S17)

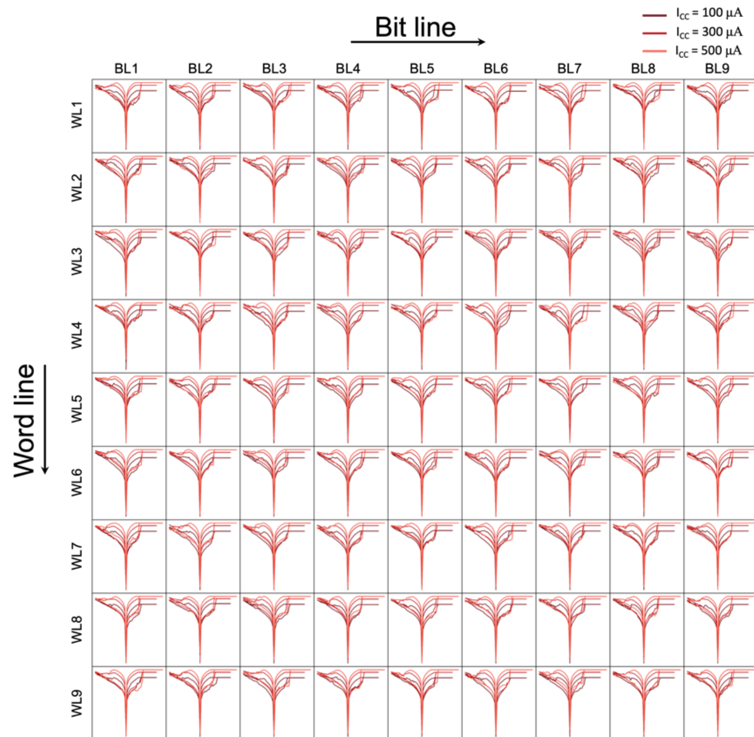

Figure S17. Analog mode I-V characteristics of 81 cells in 9 by 9 THR CBA. Each cell under compliance currents of 100 μA, 300 μA, and 500 μA are shown in different colors.

Supporting Note 1. Input Mapping and ED calculation in the analog array.

Notations :

a-CBA : analog crossbar arrray

C: Coordinate Matrix

S: Squared Vector

W: Weight Vector

The ED between two points *i* and *j* can be calculated as following Equation (S1):

$$ED = ||\boldsymbol{U_i} - \boldsymbol{U_j}|| = \sqrt{\boldsymbol{U_i^2} - 2\boldsymbol{U_i} \cdot \boldsymbol{U_j} + \boldsymbol{U_j^2}} \qquad (S1)$$

, where the single subscripted $\boldsymbol{U_i}$ and $\boldsymbol{U_j}$ are coordinate vectors of points i and j, respectively. Equation (S1) is modified to Equation (S2) to calculate and compare ED directly without any further post-processing in a-CBA:

$$-\frac{ED^2}{2} = \boldsymbol{U_i} \cdot \boldsymbol{U_j} - \frac{\boldsymbol{U_j^2}}{2} - \frac{\boldsymbol{U_i^2}}{2} \qquad (S2)$$

, where the dot product term can directly be calculated as a current output of the a-CBA, and other terms can be calculated with appropriately mapped conductances and voltages. Accordingly, the a-CBA was divided into three sub-matrices, C, S, and W, which are used to calculate the $\boldsymbol{U_i} \cdot \boldsymbol{U_j}$, $-\frac{\boldsymbol{U_j^2}}{2}$, and $-\frac{\boldsymbol{U_i^2}}{2}$ terms, respectively.

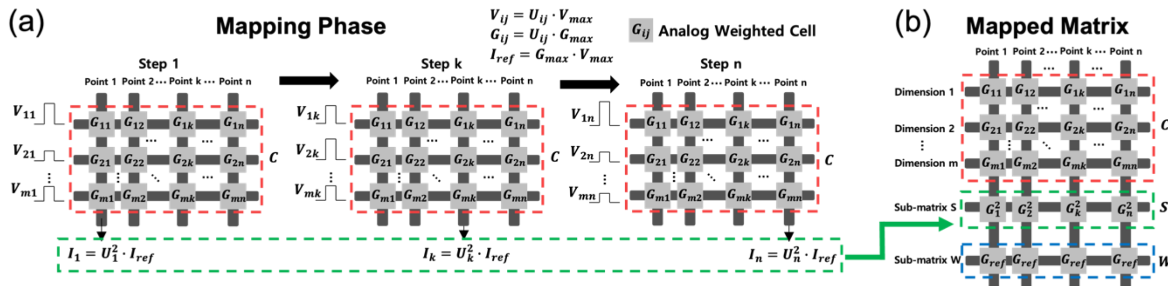

Figure S18. Analog array operations for Euclidean distance and neighbor relationship calculations (a) Mapping phase of a-CBA to calculate the squared term of coordinate U based on sub-matrix C. The m rows of the array represent the dataset dimension, and the n columns present n different data points. Total n steps are performed with appropriate mapping of voltage and conductance to output the elements of sub-matrix S. The kth step operation is depicted to show equation level derivation. (b) Mapped a-CBA with sub-matrices C, S, and W. These matrices are used to calculate separate terms of the vector-matrix multiplication (VMM) in Euclidean distance.

During the mapping phase, the spatial coordinates of each data point are appropriately translated into memristors, representing the conductance values within the C sub-matrix. These conductance values, denoted as scalar G_{ij} values in Figure S18a, utilize a two-subscript notation (i for dimension and j for data point number). Specifically, within the C sub-matrix, each n data point occupies a column, with the m-dimensional coordinate information of each point distributed across the corresponding rows within that column. This mapping process relies on THR memristors' precise analog conductance mapping capabilities. Following the mapping of the C sub-matrix, additional sub-matrices S and W undergo mapping through array-level operations, as depicted in Figure S18b. Once these sub-matrices are configured, the Euclidean distance between point i and point j is computed by analyzing the current flow through the relevant BL, following Ohm's and Kirchhoff's laws. Throughout this calculation phase, the coordinates of points $\boldsymbol{U_i}$ and $\boldsymbol{U_j}$ are translated into corresponding voltage pulses or state conductances.

Implementing VMM at the hardware level through mapping encounters two primary constraints: the permissible ranges of memristor cell conductance and reading voltage. The physical properties of conductance preclude the representation of negative values, necessitating the assignment of maximum and minimum conductance values (G_{max}, G_{min}). Additionally, memristors exhibit a linear I-V relationship solely within a specific voltage range, mandating that device operations adhere to this range to uphold Ohm's law for VMM in a-CBA. The methodology for mapping data coordinates to memristor conductances and determining the input voltage range while considering these constraints is elucidated in Supporting Note 2. In mem-DBSCAN utilizing the THR device, the coordinates of data points (U_{ij}) undergo quantization and normalization into 256 states to facilitate the straightforward mapping of coordinate values to conductance and voltage within the given limitations. Notably, U_{ij} can be converted to conductance (G_{ij}) as $G_{ij} = U_{ij} \cdot G_{max}$ and can also be transformed into voltage by $V_{ij} = U_{ij} \cdot V_{max}$ for input purposes when required.

While the conductance mapping method in the C sub-matrix is relatively straightforward, the techniques employed in S and W require further elaboration, as outlined below. The sub-matrix S consists of a single 1 x n row, where the conductance value of each component is determined as $-\frac{\boldsymbol{U_j^2}}{2}$. Following a specific procedure, these conductance values can be calculated physically using the pre-mapped C sub-matrix described earlier. For instance, considering the kth column's BL current as an example, the coordinates of point k in Figure S18a are initially mapped in the C sub-matrix ($G_{lk}, l = 1,2,\cdots,m$). By then applying these coordinates of point

k to the m rows in a voltage format ($V_{lk}, l = 1,2, \cdots, m$), the BL current corresponds to $\boldsymbol{U_k^2}$. This process is easily understood through the following equations. Initially, the BL current corresponds to Equation (S3):

$$I_k = \sum_{l=1}^{m} G_{lk} \cdot V_{lk} = \sum_{l=1}^{m} (U_{lk} \cdot G_{max}) \cdot (U_{lk} \cdot V_{max}) = G_{max} \cdot V_{max} \cdot \sum_{l=1}^{m} U_{lk} \cdot U_{lk} \quad (S3)$$

, when the input voltage vector is V_{lk}. For the matrix multiplication, Equation (S4) is used,

$$\sum_{l=1}^{m} U_{lk} \cdot U_{lk} = U_{1k}^2 + U_{2k}^2 + \cdots + U_{mk}^2 = \boldsymbol{U_k^2}. \quad (S4)$$

Therefore, for the output current, Equation (S5) is derived:

$$I_k = \boldsymbol{U_k^2} \cdot G_{max} \cdot V_{max} \quad (S5)$$

Defining the I_{ref} as Equation (S6):

$$I_{ref} = G_{max} \cdot V_{max} \quad (S6),$$

, and by obtaining the output current through VMM, the squared term of each data point's coordinates can be computed as $\boldsymbol{U_k^2}$, which is represented as the kth component of the S sub-matrix in the form of memristor conductances ($G_k^2 = \frac{U_k^2 \cdot G_{max}}{2}$). The division by two ensures the mapping range, a concept detailed in Supporting Note 2. It is important to note that all BLs except for the kth one will generate BL currents for the given V_{lk} input, but they are not relevant to the calculation of the squared term. With n points in consideration, this process iterates n times to compute all of the squared terms, thereby completing the preparation of the sub-matrix S.

The sub-matrix W is also composed of a single 1 x n row, but the conductance value of each component has a common value of $G_{ref}\ (= G_{max})$, and the reason for assigning this specific value will be understood during the subsequent explanation of the calculation phase.

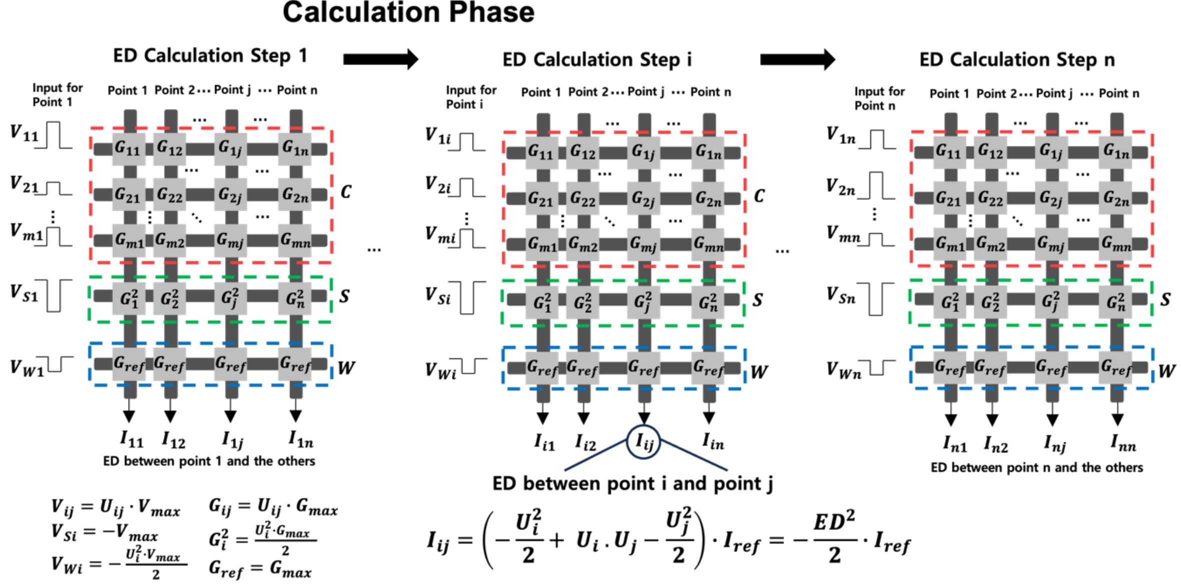

Figure S19. Schematic of parallel Euclidean distance calculations in the conductance-mapped a-CBA. n steps are computed with a-CBA VMM to calculate ED between all data points, and the ith step calculation with the jth column to calculate the distance between point i and point j is shown.

In the calculation phase, ED between point i and all other points (1, 2, ..., j, ..., n) is computed through n iterations. This process is explained using the ith calculation step in the following. First, V_{li} $(l = 1,2,\cdots,m)$ vector is applied to the C sub-matrix, as shown in Figure S19b. Then, the jth BL current of the C sub-matrix is given as Equation (S7).

$$I_{ij,C} = \sum_{l=1}^{m} G_{lj} \cdot V_{li} = \sum_{l=1}^{m} (U_{lj} \cdot G_{max}) \cdot (U_{li} \cdot V_{max}) = I_{ref} \cdot \sum_{l=1}^{m} U_{li} \cdot U_{lj} = \boldsymbol{U_i} \cdot \boldsymbol{U_j} \cdot I_{ref} \quad (S7)$$

This value corresponds to the first term of Equation (S2).

In the row corresponding to the S sub-matrix, a constant voltage $V_{Si} = -V_{max}$ is applied, and the corresponding output current is given as in Equation (S8).

$$I_{ij,S} = G_j^2 \cdot (-V_{max}) = -\boldsymbol{U_j^2} \cdot \frac{G_{max}}{2} \cdot V_{max} = -\frac{\boldsymbol{U_j^2}}{2} \cdot I_{ref} \quad (S8)$$

Lastly, in the row corresponding to the W sub-matrix, $V_{Wi} = V_i^2 = -\frac{\boldsymbol{U_i^2}}{2} \cdot V_{max}$ is applied, and all the BL currents are given as in Equation (S9) for the given conductance of G_{max}.

$$I_{ij,W} = G_{ref} \cdot V_i^2 = G_{max} \cdot (-\frac{\boldsymbol{U_i^2}}{2} \cdot V_{max}) = -\frac{\boldsymbol{U_i^2}}{2} \cdot I_{ref} \quad (S9)$$

Summing up these three currents, Equation (S10) comes out,

$$I_{ij} = I_{ij,C} + I_{ij,S} + I_{ij,W} = (\boldsymbol{U_i} \cdot \boldsymbol{U_j} - \frac{\boldsymbol{U_j^2}}{2} - \frac{\boldsymbol{U_i^2}}{2}) \cdot I_{ref} \quad (S10)$$

showing the ED between points i and j is achieved through the assessment of the current. It becomes apparent that applying a single voltage vector to the C, S, and W sub-matrices yields the EDs between the selected data point (associated with the provided input voltage vector) and all other data points, showcasing the remarkable parallel computational capability of the a-CBA. Subsequently, this process is reiterated for all n data points, reducing the number of calculation steps from $\frac{n \cdot (n-1)}{2}$ to n. The proposed approach is an efficient, precise, and highly parallel method for ED calculation utilizing a memristor-based CBA, thereby streamlining the time-intensive task during the clustering process.

Supporting Note 2. Mapping method of coordinate to the conductance and voltage.

Supporting Note 1 delves into the process of translating mathematical values, such as coordinates, into conductance and voltage for hardware processing and discusses two constraints pertinent to hardware mapping. In certain datasets, coordinate values may include negative values, which pose a challenge for direct representation in physical mappings such as those involving memristors. Several mapping methodologies have been proposed to mitigate this challenge. In this study, the dataset coordinates are initially normalized from 0 to 1 to serve as physical parameters, encompassing analog conductance state and input voltage for a-CBA. The normalization process is determined individually for each dimension of the dataset, taking into account the maximum value ($U_{i,max}$), minimum value ($U_{i,min}$) of each dimension i, and the actual values for all n points ($U_{ij,real}, i = 1,2,\cdots,m, j = 1,2,\cdots,n$) according to the following procedure:

$$U_{ij} = \frac{U_{ij,real} - U_{i,min}}{U_{i,max} - U_{i,min}} \quad (S11)$$

and these U_{ij} values should be quantized into the number of available conductance states, N (256 states for this work). These values become in the range of [0,1].

Further, for the conductance and voltage mapping of the coordinates based on the below equations:

$$G_{ij} = U_{ij} \cdot G_{max} \quad (S12)$$

$$V_{ij} = U_{ij} \cdot V_{max} \quad (S13)$$

for normalized and quantized U_{ij}.

Based on these mapping, the a-CBA firstly applies the kth input voltage vector (V_{lk}) to the kth column, which is holding G_{lk} values to calculate $\boldsymbol{U_k^2}$, and write these values into the S sub-matrix in conductance form ($G_k^2 = \frac{U_k^2 \cdot G_{max}}{2}$) based on equation S14.

$$\sum_{l=1}^{m} U_{lk} \cdot U_{lk} = U_{1k}^2 + U_{2k}^2 + \cdots + U_{mk}^2 = \boldsymbol{U_k^2}. \quad (S14)$$

Here, each component $U_{1k,} U_{2k,} \cdots, U_{mk}$ is transformed into normalized and quantized values within the range of [0,1]. Consequently, the squared term of each element $U_{1k}^2,\ U_{2k}^2, \cdots, U_{mk}^2$ also falls within the range of [0,1], and the summation of the terms $\boldsymbol{U_k^2}$ lies within the range of

[0, m]. These $\boldsymbol{U_k^2}$ terms are further mapped to the conductance form ($G_k^2 = \frac{U_k^2 \cdot G_{max}}{2}$), expanding the range of G_k^2 to [0, $\frac{m*G_{max}}{2}$], which surpasses the physical limit of conductances (G_{max}).

On the contrary, concerning voltage mapping, the input voltages for the W sub-matrix in the calculation phase, denoted as $V_{Wi} = V_i^2 = -\frac{U_i^2}{2} \cdot V_{max}$, also involve squared $\boldsymbol{U_i^2}$ terms within a range of [0,m], leading to a voltage range of [-$\frac{m*V_{max}}{2}$,0], exceeding the voltage constraint ($\pm V_{max}$ in a symmetric voltage range case). To address these issues, normalization of G_{ij} and V_{ij} considering the dataset dimension m becomes necessary, albeit at the expense of sacrificing some coordinate precision to accommodate a smaller size (m+2 by n) of a-CBA.

For this method, Equation (S12) and (S13) are exchanged to (S15) and (S16)

$$G_{ij}' = U_{ij} \cdot G_{max} \cdot \sqrt{\frac{2}{m}} \quad (S15)$$

$$V_{ij}' = U_{ij} \cdot V_{max} \cdot \sqrt{\frac{2}{m}} \quad (S16)$$

In this way, the mapping of $\boldsymbol{U_i^2}$ in conductance and voltage changes to the below

$$G'^2_i = \frac{2}{m} \cdot \boldsymbol{U_i^2} \cdot G_{max} \quad (S17)$$

$$V'^2_i = \frac{2}{m} \cdot \boldsymbol{U_i^2} \cdot V_{max} \quad (S18)$$

, which can be converted to $G'^2_i = \frac{2}{m} \cdot \boldsymbol{U_i^2} \cdot G_{max}$ in a range of [0, G_{max}] and $V_{Wi} = V'^2_i = \frac{2}{m} \cdot \boldsymbol{U_i^2} \cdot V_{max}$ in a range of [-V_{max}, 0], which can be mapped in conductance with no conflict to conductance range constriction. (The THR devices show linear I-V relations in -0.4V to 0.4V, symmetric voltage range)

In this way, for the ED calculation phase,

(1) sub-matrix C calculates

$$I_{ij,C} = \sum_{l=1}^{m} G_{lj}' \cdot V_{li}' = \sum_{l=1}^{m} \left(U_{lj} \cdot G_{max} \cdot \sqrt{\frac{2}{m}}\right) \cdot \left(U_{li} \cdot V_{max} \cdot \sqrt{\frac{2}{m}}\right) = I_{ref} \cdot \frac{2}{m} \cdot \sum_{l=1}^{m} U_{li} \cdot U_{lj}$$

$$= \boldsymbol{U_i} \cdot \boldsymbol{U_j} \cdot I_{ref} \cdot \frac{2}{m} \quad (S19)$$

(2) sub-matrix S calculates

$$I_{ij,S} = G_j'^2 \cdot (-V_{max}) = -\frac{2}{m} \cdot \boldsymbol{U_j^2} \cdot \frac{G_{max}}{2} \cdot V_{max} = -\frac{\boldsymbol{U_j^2}}{m} \cdot I_{ref} \quad (S20)$$

(3) sub-matrix W calculates

$$I_{ij,W} = G_{ref} \cdot V_i'^2 = G_{max} \cdot (-\frac{2}{m} \cdot \frac{\boldsymbol{U_i^2}}{2} \cdot V_{max}) = -\frac{\boldsymbol{U_i^2}}{m} \cdot I_{ref} \quad (S21)$$

and the total current comes out as

$$I_{ij} = I_{ij,C} + I_{ij,S} + I_{ij,W} = (\,\boldsymbol{U_i} \cdot \boldsymbol{U_j} - \frac{\boldsymbol{U_j^2}}{2} - \frac{\boldsymbol{U_i^2}}{2}) \cdot I_{ref} \cdot \frac{2}{m} = -\frac{ED^2}{m} \cdot I_{ref} \quad (S22)$$

showing the ED between the points i and j with the factor of $\frac{2}{m}$, which can be compared with the threshold value of $I_{th,\varepsilon}{}' = -\frac{\varepsilon^2}{m} \cdot I_{ref}$ to construct a neighbor matrix.

Supporting Note 3. Simulated hardware of THR memristors.

To verify the operation of the THR devices on a large array scale and its application in HDC algorithm, a simulated hardware model was built to mimic the behavior and operational variability of THR memristors. In order to account the switching parameters such as switching voltage and resistance variation, a diode-like model was employed for the observed electrical response.[5] This approach assumes that the current flows through a filamentary-type structure can be approximated by a potential barrier with height E_B as given by the expression[5]:

$$I = eA\,\{exp[-\gamma(E_b - \mu_L)] - exp[-\gamma(E_b - \mu_R)]\} \tag{S23}$$

, where μ_L and μ_R represent the electrochemical potentials at the left and right reservoirs of the structure, respectively, A and γ are constants related to the particular characteristics of the potential barrier, and e denotes the elementary charge of an electron. Under the application of a positive bias voltage V > 0, $\mu_L - \mu_R = eV$ so that Eq. (S23) becomes:

$$I(V) = I_0[exp(\beta V) - 1] \tag{S24}$$

which is formally analogous to the ideal diode equation, I_0 represents the diode amplitude (which may exhibit voltage dependence through E_b), and $\beta \approx dln(I)/dV$ is the logarithmic conductance of the diode. For negative bias voltage V < 0, a similar equation, with the appropriate signs, holds.

For a comprehensive model, series (R_S) and parallel resistances (R_{P1}, R_{P2}) are incorporated. Physically, R_S may represent a residual local potential barrier. In contrast, R_{P1} and R_{P2} may represent localized and area-distributed parallel leakage current pathways. The overall current-voltage relationship can thus be approximated by the expression[5]:

$$I(V) = sgn(V)\{(\beta R_S)^{-1}\, W[\beta I_0 R_S d_1 \exp(\beta d_1 \exp\,(\beta d_1(|V| + I_0 R_S))] + d_1(|V|G_{P1} - I_0) - |V|G_{P2}\} \tag{S25}$$

where $G_{P1} = 1/R_{P1}, G_{P2} = 1/R_{P2}, d_1 = (1 + G_{P1}R_S)^{-1}$, *sgn* is the sign function, and W is the Lambert function. For the Lambert function, we considered the following approximation, which is based on a Hermite-Padé expansion[6]:

$$W(x) = \ln(1 + x)\ \left(1 - \frac{\ln\,(1+\ln\,(1+x))}{2+\ln\,(1+x)}\right) \tag{S26}$$

for real x > 0 with a relative error less than 10^{-2}.

Within the context of this work, the term I_0 in Eq. (S2) is modeled by employing sigmoidal-type threshold functions. This selection is motivated by the ability of these functions

to generate smooth transitions between the two resistance states during the application of a voltage signal:

$$I_0 = I_{0Min} + (I_{0Max} - I_{0Min})\, \lambda(V) \tag{S27}$$

$$\text{with } \lambda^{on/off}(V) = \lambda_{Max}\, tangh\left[-\eta\left(V - \frac{V_{th}^{\frac{on}{off}}}{\beta_r}\right)\right] \tag{S28}$$

β_r regulates the threshold at which switching occurs between the two resistance states, and it is directly linked to the current compliance. $V_{th}^{on/off}$ is the threshold voltage to switch from the HRS to the LRS, and back to the HRS. The λ function plays the role of the state equation of the memristor.

Eqs. (S23)-(S28) show that the proposed model possesses essential characteristics for the compact modeling of memristors as it is analytic, continuous, and differentiable. The tractable mathematical functions accurately reproduce diverse electrical responses, including linear/non-linear and pinched/non-pinched I-V responses. η determines the transition rates for both switching processes.

Additionally, to incorporate the resistance state and voltage distribution into the model, I_0 values and $V_{th}^{on/off}$ values are determined from switching parameters of digital and analog mode characteristics. Figure S20 shows the fitting results for the I-V curves of digital and analog modes, confirming that the model accurately simulates the dispersion of switching voltage, switching window, and state dispersion. Supporting Table S2 and S3 shows the fitting parameters representing THR devices' switching characteristics, including the switching variations.

The effect of line resistance on VMM calculation for validation in CBA was also analyzed. Figure S21 shows the result of analog mode VMM operation with a 1 by 9 CBA. For the measurement, 9 different conductance states were mapped for 9 memristors using the proposed precise tuning algorithm, and various voltages under 0.4 V were applied to measure the output current. A maximum of 0.4 V was selected to ensure that the mapped conductance was read within a linear region without state change. VMM through simulated hardware was calculated via Spice simulation, considering line resistance. For the precise mapping of conductances in analog mode, conductance fluctuations of different analog states shown in Figure S12 were implemented to Spice simulation to incorporate fluctuations result from

nonlinear I-V relationship and read noises of analog states to calculate the ED, KD, and RD values in Figure 4 in the main text.

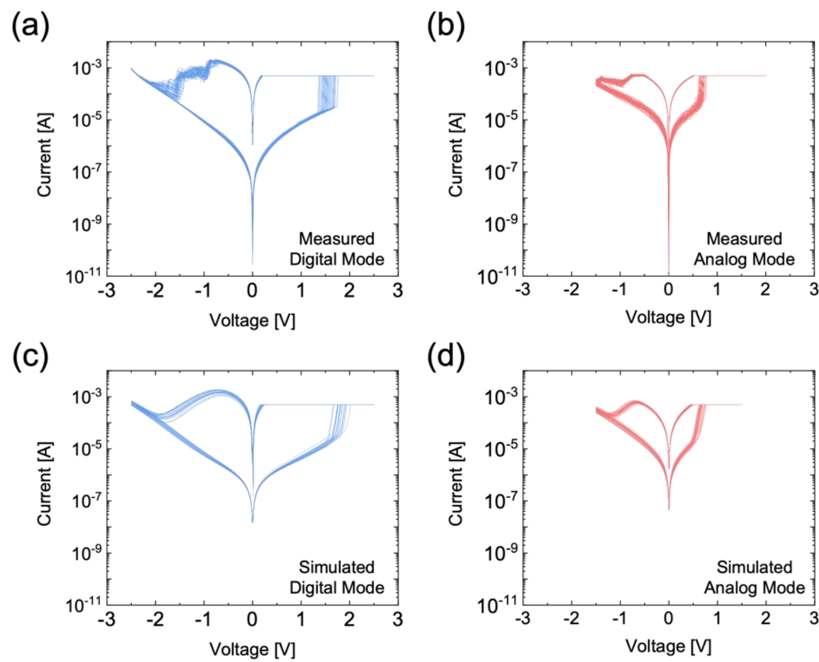

Figure S20. Simulated hardware characteristics. (a) and (b) are measured I-V characteristics of digital and analog modes, 50 cycles each respectively. (c) and (d) are I-V characteristics of digital and analog modes via constructed simulated hardware, respectively.

.

Supporting Table S1. Fitting parameters for the digital mode.

Parameter	SET	RESET
η	30	7.5
α	3	3
$I_{O,min}$	1 mA for LRS / 100 nA for HRS	
$I_{O,max}$	3 mA for LRS / 200 nA for HRS	
Rs	200 ohm	
Rp	10^4 ohm	
$V_{th}^{on/off}$	$N(2, 0.1^2)$	$N(0.6, 0.05^2)$

Supporting Table S2. Fitting parameters for the analog mode.

Parameter	SET	RESET
η	30	10
α	3	3
$I_{O,min}$	0.2 mA for LRS / 3 μA for HRS	
$I_{O,max}$	0.3 mA for LRS / 6 μA for HRS	
Rs	200 ohm	
Rp	10^4 ohm	
$V_{th}^{on/off}$	N(0.7, 0.05^2)	N(0.7, 0.03^2)

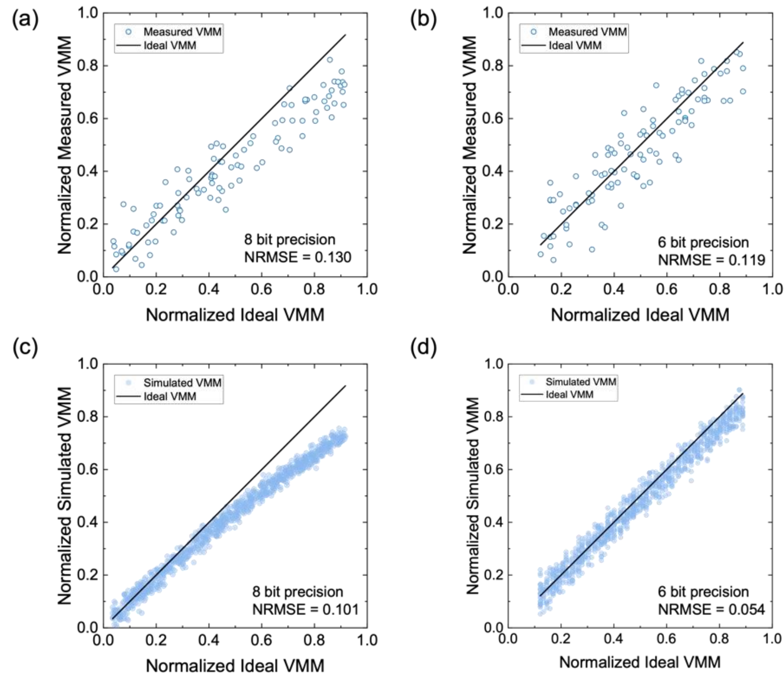

Figure S21. Analog mode VMM result for 1 by 9 array with different bit precision. (a) Measured VMM currents of 1 by 9 analog array, with 8 bit state precision. (100 μS to 2660 μS, interval of 10 μS) (b) Measured VMM currents of 1 by 9 analog array, with 6 bit state precision. (100 μS to 740 μS, interval of 10 μS) (c) Simulated VMM currents of 1 by 9 analog array, with 8 bit state precision and the line resistance between cells of 3 Ohm. (d) Simulated VMM currents of 1 by 9 analog array, with 6 bit state precision and the line resistance between cells of 3 Ohm.

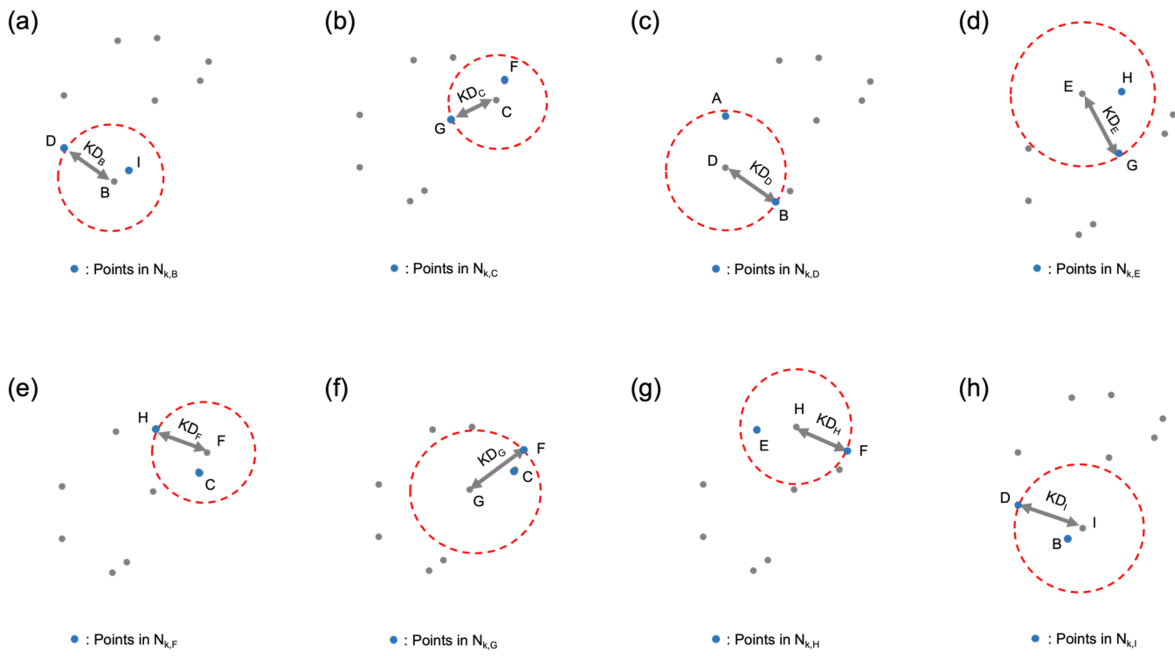

Figure S22. All other KNN and KD relationships (K = 2) of the 9-point dataset, explained in the main text.

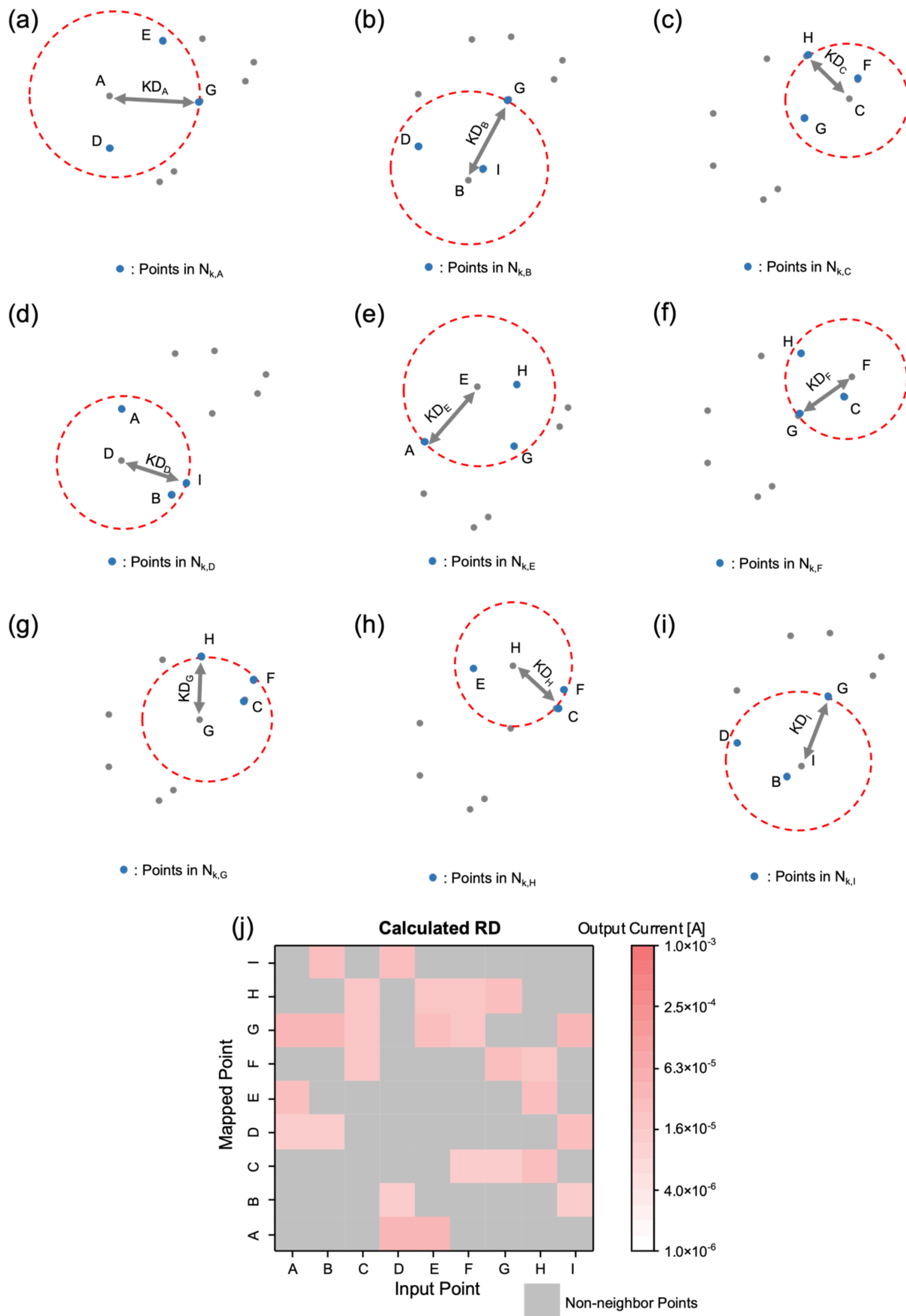

Figure S23. KNN, KD, and RD relationship of the 9-point dataset with K = 3.

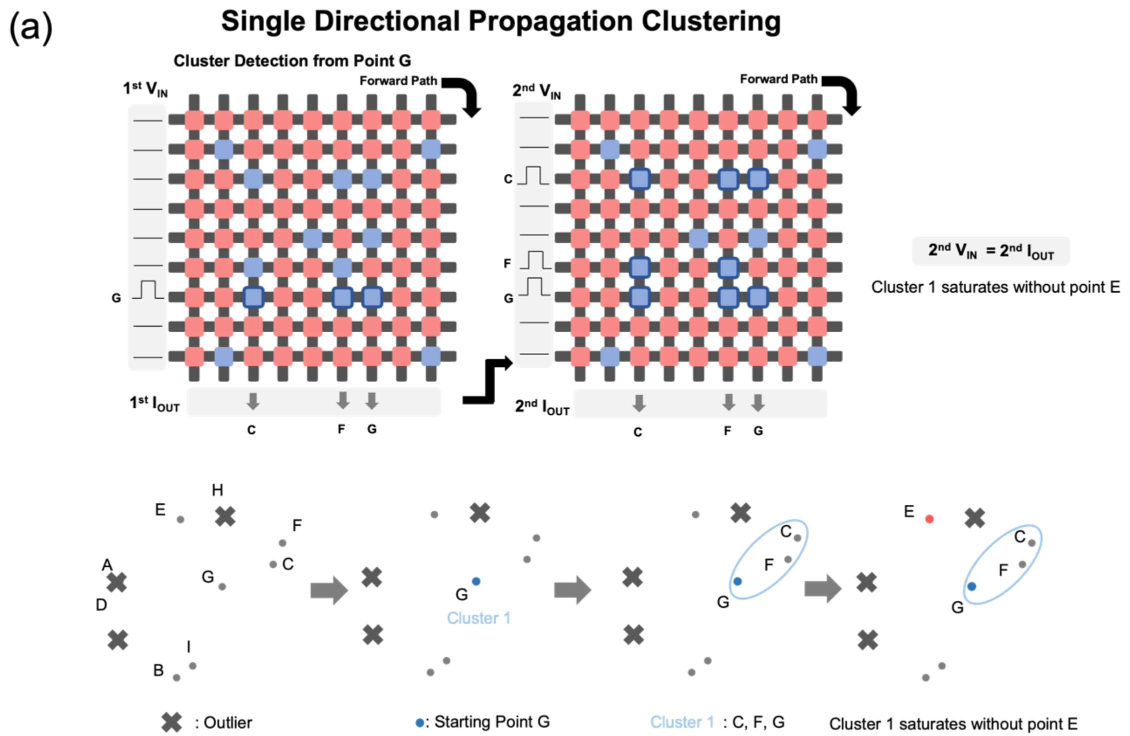

Figure S24. Digital array operation and graphical representation within the single directional propagation clustering, starting from point G. By one step of VMM, cluster 1 saturates with the points C, F, and G, which differs from the result from the starting point E. Due to the asymmetric mapping of the digital matrix, unidirectional propagation clustering results in the starting point-dependent cluster results, which is not favorable for the proposed HDC algorithm.

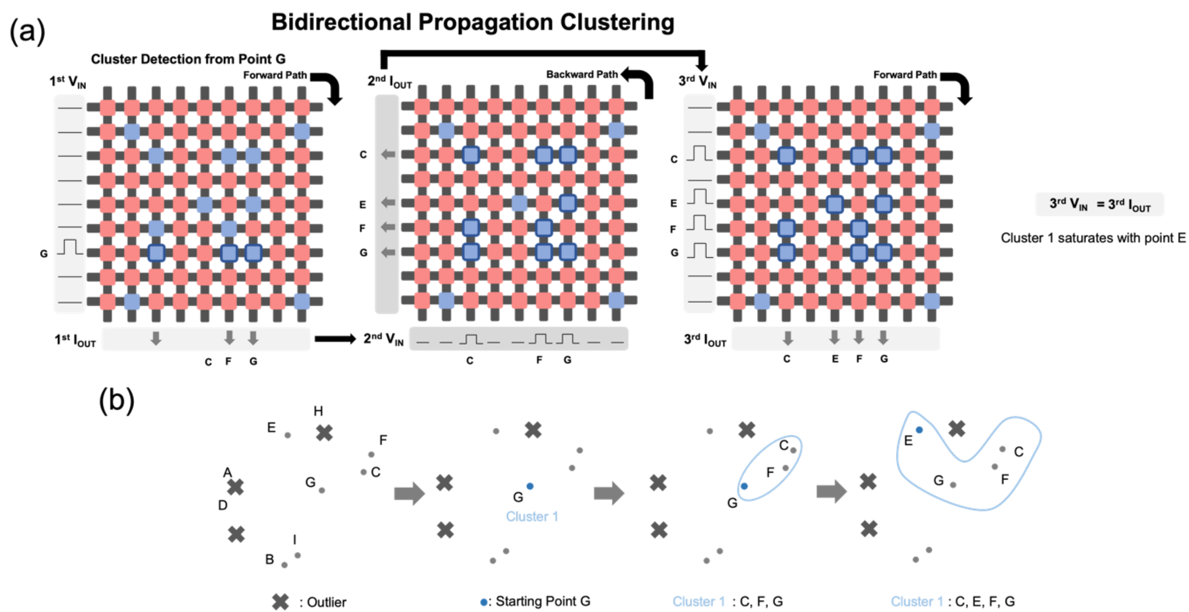

Figure S25. Digital array operation and graphical representation within the bidirectional propagation clustering, starting from point G.

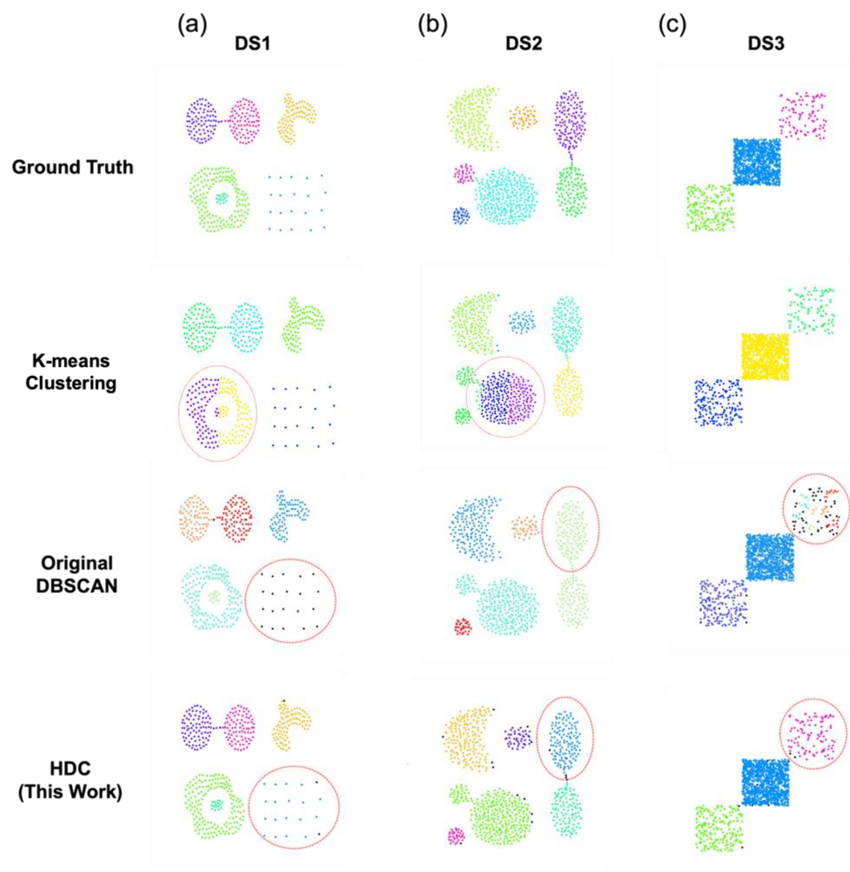

Figure S26. Clustering results for three different datasets (DS1-DS3) with various clustering methods.

Supporting Table S3. Comparison of ARI between three different clustering methods

	DS1	DS2	DS3
K-menas	0.783	0.793	1.0
DBSCAN	0.930	0.883	0.929
HDC (This work)	0.994	0.894	0.985

Supporting Note 4. Energy and latency calculation of HDC algorithm.

The energy consumption of K-means clustering for a CPU-based digital computing system was estimated as 45 W (thermal design power of CPU, Ryzen 5 4600H) × 7.85 % (CPU utilization during data clustering) × 199.5 ms (computational time for data clustering) = 704.7 mJ. In the same way, the energy consumption and latency of DBSCAN was estimated as 226.1 mJ and 114.7 ms, respectively.

The energy consumption of the analog mode array was the sum of switching energy ($E_{switching}$) for the mapping and static energy (E_{static}) for the VMM operations. Thus, the total energy was estimated using the following equations:

$$E_{switching} = N_{ISPP,mean} \cdot \int V(t). I(t). dt \ . (N_{dimension} + 2). N_{dataset} \quad \text{(S29)}$$

$$E_{static} = I_{analog} \cdot t_{read} \cdot V_{read,mean} \cdot (N_{dimension} + 2). N_{dataset} \ . (N_{dataset} + 1 + N_K) \quad \text{(S30)}$$

, where $E_{switching}$ is the energy for the analog conduction mapping process, $V(t)$ is the voltage till the switching process occurs, $I(t)$ is the current that flows through the device during the switching event, and $N_{ISPP,mean}$ is the average of the required number of pulses for analog conductance mapping as shown in Figure S15. The read voltage ($V_{read,\ mean}$) was 0.2 V on average, and its output current (I_{analog}) was 2 mA on average. The width of the read pulse, t_{read}, was 100 ns. $N_{dimension}$, $N_{dataset}$, and N_K depend on the dataset.

Similarly, the energy consumption of the digital mode CBA was the sum of switching energy ($E_{switching}$) for the mapping and static energy (E_{static}) for the VMM operations. Thus, the total energy was estimated using the following equations:

$$E_{switching} = \int V(t). I(t). dt \ . N_{dataset} \cdot N_K \quad \text{(S31)}$$

$$E_{static} = I_{digital} \cdot t_{read} \cdot V_{read} \cdot N_{dataset} \cdot N_{dataset} \cdot N_{iteration} \quad \text{(S32)}$$

The read voltage (V_{read}) was 0.1 V, and its output current $I_{digital}$ depends on N_K.

The energy consumption of the peripheral circuit on the both array operations were estimated as $P_{peri} \cdot t. N_{element}$ based on 130-nm Si CMOS technology. The DAC, ADC, and comparator power were about 7.2 mW, 2.3 mW, and 0.2 mW.[7–10] In addition, peripheral circuits, such as DAC, ADC, and voltage drivers, were estimated based on an 8-bit I/O system. Finally, this work's step-by-step results of energy calculations regarding latency and energy consumption are described in Supporting Table S4.

Supporting Table S4. Summarized step-by-step energy consumption and latency

	This work	
	Energy (mJ)	**Latency (ms)**
Analog CBA mapping	3.84×10^{-2}	4
Analog CBA VMM	4.02×10^{-2}	0.3
Digital CBA mapping	3×10^{-3}	0.1
Digital CBA VMM	2.5×10^{-3}	5×10^{-3}
Peripheral Circuits	1.47	3.82×10^{-2}
Total	1.56	4.44

Supporting References

1 T. H. Park, H. J. Kim, W. Y. Park, S. G. Kim, B. J. Choi and C. S. Hwang, *Nanoscale*, 2017, **9**, 6010–6019.

2 T. H. Park, Y. J. Kwon, H. J. Kim, H. C. Woo, G. S. Kim, C. H. An, Y. Kim, D. E. Kwon and C. S. Hwang, *ACS Appl Mater Interfaces*, 2018, **10**, 21445–21450.

3 G. S. Kim, H. Song, Y. K. Lee, J. H. Kim, W. Kim, T. H. Park, H. J. Kim, K. Min Kim and C. S. Hwang, *ACS Appl Mater Interfaces*, 2019, **11**, 47063–47072.

4 M. Rao, H. Tang, J. Wu, W. Song, M. Zhang, W. Yin, Y. Zhuo, F. Kiani, B. Chen, X. Jiang, H. Liu, H. Y. Chen, R. Midya, F. Ye, H. Jiang, Z. Wang, M. Wu, M. Hu, H. Wang, Q. Xia, N. Ge, J. Li and J. J. Yang, *Nature*, 2023, **615**, 823–829.

5 J. Blasco, N. Ghenzi, J. Suae, P. Levy and E. Miranda, *IEEE Electron Device Letters*, , DOI:10.1109/LED.2014.2297992.

6 S. Winitzki, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, DOI:10.1007/3-540-44839-x_82.

7 F. Solis, Á. Fernández Bocco, A. C. Galetto, L. Passetti, M. R. Hueda and B. T. Reyes, *International Journal of Circuit Theory and Applications*, DOI:10.1002/cta.3029.

8 A. K. Adepu and K. K. Kolupuri, *International Journal on Cybernetics & Informatics*, DOI:10.5121/ijci.2016.5434.

9 F. Moradi, G. Panagopoulos, G. Karakonstantis, H. Farkhani, D. T. Wisland, J. K. Madsen, H. Mahmoodi and K. Roy, *Microelectronics J.*, DOI:10.1016/j.mejo.2013.09.009.

10 B. B. A. Fouzy, M. B. I. Reaz, M. A. S. Bhuiyan, M. T. I. Badal and F. H. Hashim, in *2016 International Conference on Advances in Electrical, Electronic and Systems Engineering, ICAEES 2016*, 2016.