nature electronics

Article

https://doi.org/10.1038/s41928-023-01042-7

Reconfigurable mixed-kernel heterojunction transistors for personalized support vector machine classification

Received: 23 August 2022

Accepted: 12 September 2023

Published online: 12 October 2023

Check for updates

Xiaodong Yan[1,7], Justin H. Qian[1,7], Jiahui Ma[2,7], Aoyang Zhang[2], Stephanie E. Liu[1], Matthew P. Bland[1], Kevin J. Liu[1], Xuechun Wang[3], Vinod K. Sangwan[1] ✉, Han Wang[2,4] ✉ & Mark C. Hersam[1,5,6] ✉

Advances in algorithms and low-power computing hardware imply that machine learning is of potential use in off-grid medical data classification and diagnosis applications such as electrocardiogram interpretation. However, although support vector machine algorithms for electrocardiogram classification show high classification accuracy, hardware implementations for edge applications are impractical due to the complexity and substantial power consumption needed for kernel optimization when using conventional complementary metal–oxide–semiconductor circuits. Here we report reconfigurable mixed-kernel transistors based on dual-gated van der Waals heterojunctions that can generate fully tunable individual and mixed Gaussian and sigmoid functions for analogue support vector machine kernel applications. We show that the heterojunction-generated kernels can be used for arrhythmia detection from electrocardiogram signals with high classification accuracy compared with standard radial basis function kernels. The reconfigurable nature of mixed-kernel heterojunction transistors also allows for personalized detection using Bayesian optimization. A single mixed-kernel heterojunction device can generate the equivalent transfer function of a complementary metal–oxide–semiconductor circuit comprising dozens of transistors and thus provides a low-power approach for support vector machine classification applications.

The support vector machine (SVM) is a supervised machine learning algorithm based on quadratic programming and statistical learning theory[1]. It is an efficient classification tool for many applications, including channel estimation and voice detection[2]. Compared with neural network classifiers, SVM algorithms are less computationally demanding, making them more suitable for hardware implementation in low-power applications including real-time off-grid health monitoring[3–6]. For linearly separable problems, an SVM classifier identifies the

[1]Department of Materials Science and Engineering, Northwestern University, Evanston, IL, USA. [2]Ming Hsieh Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, CA, USA. [3]Biomedical Engineering Department, University of Southern California, Los Angeles, CA, USA. [4]Mork Family Department of Chemical Engineering and Materials Science, University of Southern California, Los Angeles, CA, USA. [5]Department of Electrical and Computer Engineering, Northwestern University, Evanston, IL, USA. [6]Department of Chemistry, Northwestern University, Evanston, IL, USA. [7]These authors contributed equally: Xiaodong Yan, Justin H. Qian, Jiahui Ma. ✉e-mail: vinod.sangwan@northwestern.edu; han.wang.4@usc.edu; m-hersam@northwestern.edu

hyperplane that maximizes the margin distance between two classes in the feature space. However, for practical nonlinear problems, a kernel function is typically employed to map low-dimensional inputs into a higher-dimensional space, thus making linear non-separable inputs easier to classify. Although a variety of functions can, in principle, be employed for this purpose, kernels are most typically derived from linear, polynomial, Gaussian and sigmoid functions.

Although the Gaussian function (also known as the radial basis function) is the most commonly used kernel, additional kernels are often better suited for typical classification tasks[2]. In particular, the Gaussian kernel has interpolation ability and is effective at identifying local properties. As a result, SVMs with Gaussian kernels have a strong local learning ability (Supplementary Fig. 1a–d). In contrast, sigmoid kernels are better suited for identifying global characteristics but have a relatively weak interpolation ability (Supplementary Fig. 1e–h). Therefore, mixed kernels that combine the advantages of Gaussian and sigmoid kernels often have the best classification performance for practical applications. Complementary metal–oxide–semiconductor (CMOS) implementations of SVM algorithms are highly effective for classification; the energy consumption can be improved by using alternative (non-CMOS) platforms for real-time applications such as continuous monitoring. Consequently, SVM hardware accelerators have been widely explored including the hardware implementations of dynamically reconfigurable kernel functions using both digital[7–9] and analogue[10–12] circuits. Analogue circuits are theoretically more effective than digital architectures due to their lower power consumption and areal footprint, but even the simplest implementations of analogue Gaussian functions require a large number of circuit elements[13], which is worsened when implementing tunable and mixed-kernel functions[14].

In this Article, we report dual-gated mixed-kernel heterojunction (MKH) transistors using monolayer molybdenum disulfide (MoS_2) (grown using chemical vapour deposition (CVD)) as an n-type material and solution-processed semiconducting carbon nanotubes (CNTs) as the p-type material. Precise control over the electric-field screening in MKH transistors enables the generation of a complete set of fine-grained Gaussian, sigmoid and mixed-kernel functions using only a single device. In conjunction with Bayesian optimization (BO), the MKH transistors provide effective and efficient hyperparameter searching, which further enhances the classification performance. By taking into account user diversity through personalized hyperparameter optimization, precise arrhythmia detection can be derived from electrocardiogram (ECG) data. We also illustrate the scalability advantage of our MKH transistors in SVM hardware implementations that employ an $n \times n$-kernel matrix. Compared with equivalent kernel generation using conventional CMOS circuits, our MKH transistors reduce the number of circuit elements for mixed-kernel SVM by approximately two orders of magnitude, thereby providing a high classification accuracy in a scalable and energy-efficient manner.

MKH transistor structure and tunable mixed kernels

Our MKH transistors are designed to realize rich and distinct functionalities that have not been realized in previous anti-ambipolar devices. First, the semi-vertical geometry of our device design allows for the dual gating of both overlapping and non-overlapping regions of MoS_2 and CNTs in the heterojunction (Fig. 1a–c). The overlap region of the MoS_2/CNT heterostructure forms a p–n junction diode (Fig. 1d) with nanomaterial-enabled partial electric-field screening[15] in the overlap region. The overlap region in combination with the MoS_2 and CNT transistors in series in the non-overlapping regions enables highly tunable anti-ambipolar transfer characteristics[16]. Figure 1e–g shows the charge transport measurements under different dual-gating conditions that yield Gaussian kernel functions with tunable mean (μ) (Fig. 1e), amplitude (A) (Fig. 1f) and standard deviation (σ) (Fig. 1g). Compared with previous literature[17,18], the Gaussian behaviour in our MKH transistors is symmetric and shows an unprecedented independent tunability of each hyperparameter including standard deviation (Supplementary Fig. 2), which is enabled by the weak screening in the overlap region. Solution-processed CNTs are advantageous for this purpose since their network density can be tuned over a wide range, thus allowing precise control over the degree of screening and mobility. In the MKH transistors, a linear density of ~7 CNTs per micrometre is used, which avoids the n-type arm in the CNT ambipolar response compared with higher CNT densities (Supplementary Figs. 3 and 4) while concurrently providing optimal top-gate screening and mobility[19].

By varying the area of the overlapping region of the MoS_2/CNT heterojunction, the MKH transistors can also produce sigmoid kernel functions with tunable slopes by sweeping the top gate (V_{TG}) and keeping the back gate (V_{BG}) fixed at 5 V (Fig. 1h). This additional sigmoid functionality has not been realized in previous anti-ambipolar devices[18,20–24]. In the MKH transistors, a lateral overlap length of 10 μm (Supplementary Fig. 10) in the MoS_2/CNT heterojunction yields optimal sigmoid functions compared with smaller overlap lengths (Supplementary Fig. 4) because it increases the effective channel width of MoS_2 that is screened from the top gate. When a constant V_{BG} bias of 5 V is applied, the current in the screened MoS_2 channel will saturate at a level that produces sigmoid functions with minimal distortion. At smaller overlap lengths, the ideal saturation level of the MoS_2 is too low, leading to a partial Gaussian behaviour before saturation (Supplementary Fig. 4). Supplementary Note 1 provides further details.

Additionally, by tuning V_{BG}, a series of transfer characteristics can be generated that contain both sigmoid and Gaussian characteristics (Fig. 1i), where curve fitting confirms Gaussian and sigmoid mixed-kernel functionality (Supplementary Fig. 6). This single-device mixed-kernel generation is also unique to our MKH transistors, where the dual-gated architecture coupled with optimized electric-field screening allows for tailored control over the carrier concentrations in the MoS_2, CNT and overlapping MoS_2/CNT heterojunction regions (Supplementary Figs. 3 and 4).

In single-gated anti-ambipolar devices[21,23,24], the lack of hyperparameter tunability of the Gaussian curves limits their utility for kernel generation since hardware implementations would require additional overhead circuitry. Lateral anti-ambipolar devices also suffer from higher power consumption, which limits their scalability. Therefore, the reconfigurable MKH transistors reported here have multiple advantages over incumbent anti-ambipolar devices. First, multiple types of kernel function including Gaussian, sigmoid and mixed Gaussian and sigmoid curves can be generated in a single MKH device. Second, dynamic reconfigurability of the kernel functions is realized by simply modulating the gate fields in the device channel by utilizing the weak screening of atomically thin semiconductors. Third, even without scaling the devices down to the deep submicrometre limit, the power consumption for MKH transistors to generate Gaussian curves is already comparable with one of the lowest power consumption that has been experimentally reported in scaled CMOS Gaussian generation circuits[14]. Moreover, the latter CMOS circuit requires more device overhead and complicated designs, which ultimately preclude its use for edge applications. These multifold advantages of MHK transistors for generating mixed kernels enable efficient and effective SVM classification, as demonstrated below for the specific case of personalized arrhythmia detection from ECG data.

Mixed-kernel SVM for arrhythmia detection

To illustrate the utility and effectiveness of our MKH transistors in practical applications, we explored mixed-kernel SVM classification for personalized arrhythmia detection from the ECG data. Our proposed system for real-time arrhythmia detection includes data acquisition, mixed-kernel SVM circuitry for classification and a user interface (Fig. 2a). Typical ECG input signals are shown in Fig. 2b for a variety of arrhythmia types[25,26]. As shown in Fig. 2c, six different

https://doi.org/10.1038/s41928-023-01042-7

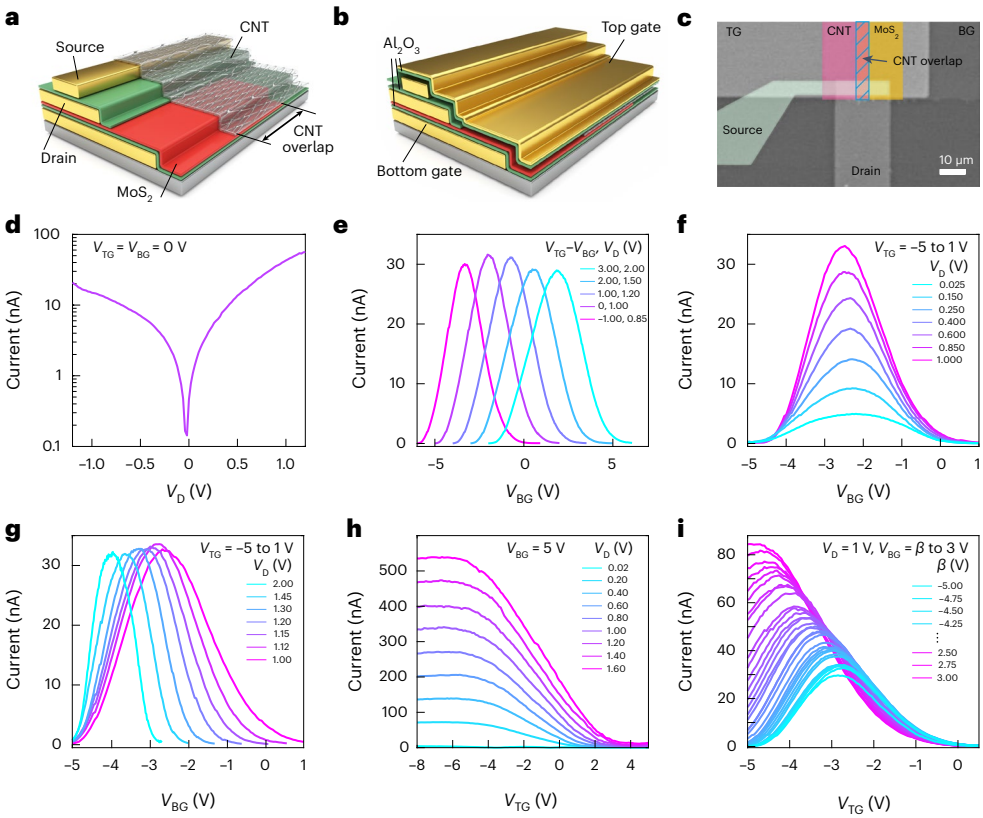

Fig. 1 | MKH transistor schematic, structure and performance. a, Partial schematic of the device showing the MoS_2/CNT channels and overlap. **b**, Full schematic of a completed device. **c**, Electron microscopy image of the fabricated device with the device regions labelled. **d**, I_D–V_D curve of the device with both top and bottom gates grounded. **e**, I_D–V_{BG} curves of the device in the Gaussian operation mode showing tunable μ. **f**, I_D–V_{BG} curves of the device in the Gaussian operation mode showing tunable amplitude. **g**, I_D–V_{BG} curves of the device in the Gaussian operation mode showing tunable σ. **h**, I_D–V_{TG} curves of the device in the tunable sigmoid operation mode. **i**, I_D–V_{TG} curves of the device showing tunable mixed-kernel generation. Supplementary Note 2 provides details of the biasing conditions.

types of arrhythmia are considered here: (1) normal beat (N), (2) atrial premature beat (A), (3) premature ventricular contraction (V), (4) paced beat (/), (5) left bundle branch block beat (L) and (6) right bundle branch block beat (R). Ambulatory ECG recordings are collected from biosensors, amplified and preprocessed by analogue-to-digital converters (Fig. 2d, channel 1 input). Two approaches are considered for kernel function generation (Fig. 2d (channel 2 input) and Supplementary Fig. 6). The first approach uses only one MKH transistor to internally generate tunable Gaussian, sigmoid and mixed kernels. The second approach uses two MKH transistors that are separately optimized for tunable Gaussian kernels and tunable sigmoid kernels, which are then externally mixed to produce a more complete set of mixed kernels compared with the first approach. The mixing ratio is dynamically tuned by using an additive modulator based on the optimization results. A mixed-kernel SVM module then receives channel 1 and channel 2 inputs to perform arrhythmia detection. A final user interface monitors the two channel inputs and displays arrhythmia-type classification results.

Different combinations of Gaussian/sigmoid mixed kernels were tested for ECG classification. In particular, five different mixed kernels were generated using one MKH transistor, which correspond to mixed-kernel functions having β values of −5, −3, −1, 1 and 3 V (Fig. 1i). Figure 2e shows the classification accuracies for correctly identifying each arrhythmia type (that is, N, A, V, /, L and R) out of 10,000 input ECG pulse waveform samples acquired from the MIT-BIH arrhythmia database[27,28]. The mixed kernel with a β value of 1 V yields the highest average arrhythmia detection accuracy with all the arrhythmia types being detected with an accuracy at or above the ~90% level, and is an improvement compared with conventional purely Gaussian (β = −5 V) or purely sigmoid (β = 3 V) kernels. These experimental results confirm that different mixed-kernel ratios affect the SVM classification accuracy, thus highlighting the importance of dynamic tunability in mixed-kernel hardware implementations.

Bayesian-optimized mixed kernels for personalized detection

For SVM classification, choosing the optimal kernel is critical for a high classification accuracy. Since the optimal selection of hyperparameters can greatly vary for different applications and scenarios, brute-force combinatorial optimization is typically impractical, especially for personalized, real-time applications. Therefore, a dynamically reconfigurable mixed-kernel hardware solution needs to be used in conjunction with an efficient hyperparameter optimization strategy. In this regard, BO is a promising option since it belongs to a class of sequential model-based optimization algorithms that have been shown to outperform random research or grid search in terms of convergence performance[29]. In our case, the hyperparameters for the mixed kernel are iteratively optimized using BO by maximizing the marginal likelihood of arrhythmia detection using a Gaussian process. A Gaussian process is a generalized Gaussian distribution that is specified by mean and covariance functions, which acts as a prior probability model. The BO process initiates from a random sample in the hyperparameter space. After each BO iteration, the expected improvement serves as an acquisition function to determine the next search point in the hyperparameter space. Figure 3a–c shows the BO results after 5, 15 and 25 iterations, respectively. After 25 iterations, the search points

https://doi.org/10.1038/s41928-023-01042-7

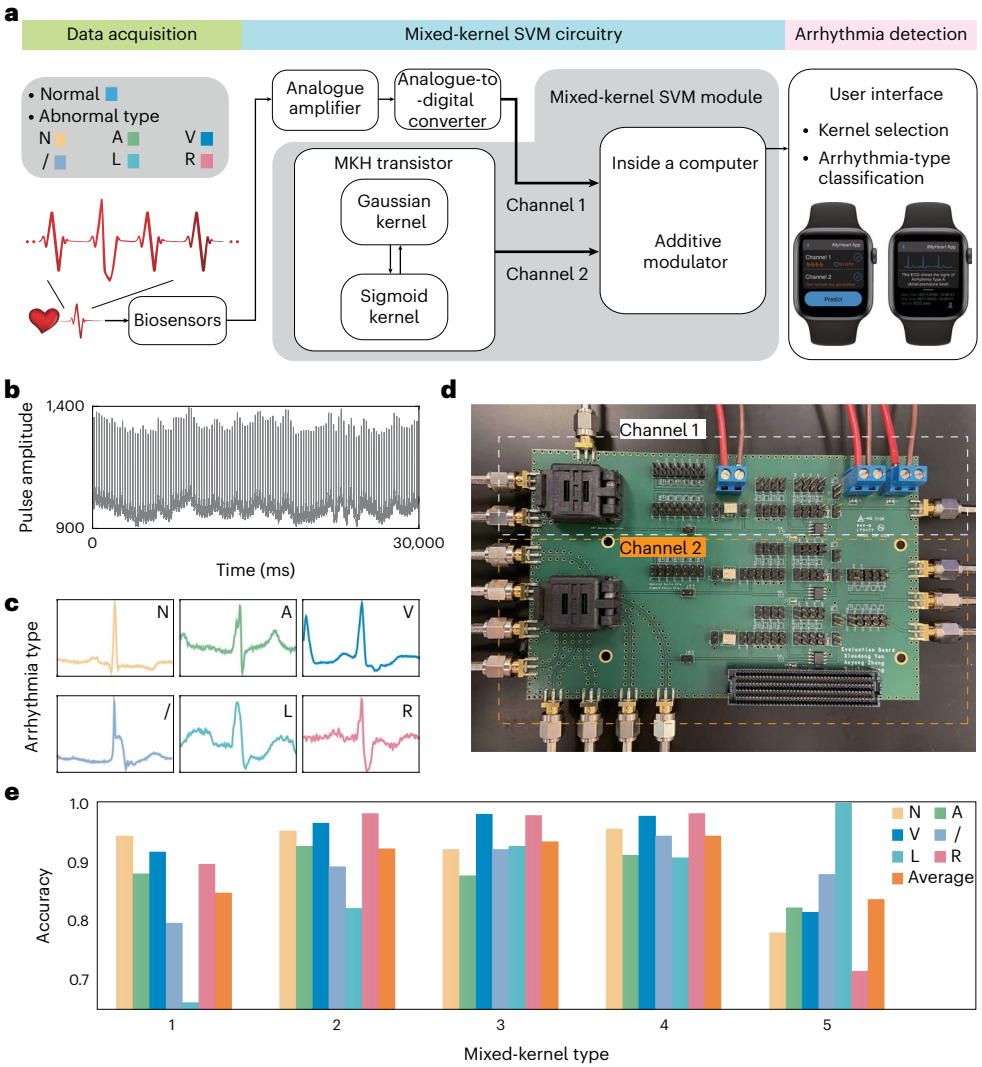

Fig. 2 | Arrhythmia detection using single-device-generated mixed kernels. **a**, Schematic of the mixed-kernel SVM circuitry and user interface design for arrhythmia detection. Ambulatory ECG recordings are collected from biosensors, and an optimal mixed kernel is used to identify six different types of arrhythmia. **b**, Typical heartbeat ECG dataset. A binning window size of 320 ms was used to discretize the data. **c**, Six arrhythmia types: normal beat (N), atrial premature beat (A), premature ventricular contraction (V), paced beat (/), left bundle branch block beat (L) and right bundle branch block beat (R). **d**, Photograph of the circuit board used to implement the mixed-kernel SVM. **e**, Classification accuracy for different mixed kernels corresponding to β values of −5, −3, −1, 1 and 3 V (Fig. 1i). Mixed-kernel type 1 is purely Gaussian, mixed-kernel type 5 is purely sigmoid and mixed-kernel types 2–4 are 75%/25%, 50%/50% and 25%/75% Gaussian/sigmoid kernels, respectively.

have converged to the optimal hyperparameter combination, where the highest classification accuracy is achieved (Fig. 3c, red region).

Since the BO process enables highly efficient optimization, it is suitable for determining optimal mixed-kernel hyperparameters for personalized or group-based classification. In addition, because BO can accommodate relatively small and noisy datasets, it is appropriate for mobile-friendly arrhythmia detection and related edge-computing use cases. To quantify the effectiveness of this approach, a BO-optimized mixed-kernel SVM was used for personalized arrhythmia detection (Fig. 3d). This assessment was carried out by randomizing 100 arrhythmia records from the publicly available MIT-BIH arrhythmia database[27,28]. These records were used as the input dataset to the mixed-kernel SVM system, after which BO was employed for the mixed-kernel hyperparameter selection for each case. The average classification accuracy over the six arrhythmia types was then calculated, where the results of ten specific records are provided (Fig. 3e). Compared with the classification results achieved using only Gaussian or only sigmoid kernels, the personalized mixed kernels were better suited for diverse patient datasets, resulting in consistently higher arrhythmia detection accuracy (Fig. 3f).

Comparing MKH classification with CMOS implementations

In this section, we compare our MKH devices with previous experimentally demonstrated Gaussian/sigmoid mixed-kernel hardware implementations based on conventional CMOS circuits. Early CMOS demonstrations of Gaussian function generation utilized bump circuits[30], after which CMOS Gaussian function circuits were studied for hardware implementations of machine learning algorithms[31], smart sensors[32] and neuromorphic computing systems[33]. Although the bump circuit is still the most robust and low-power CMOS method for generating fixed Gaussian curves, scalability issues limit the use of bump circuits for tunable Gaussian function generation. Since a bump circuit generates only one type of Gaussian curve, the generation of tunable Gaussian functions requires cascading multiple bump circuits with each circuit containing CMOS transistors of different channel widths and

https://doi.org/10.1038/s41928-023-01042-7

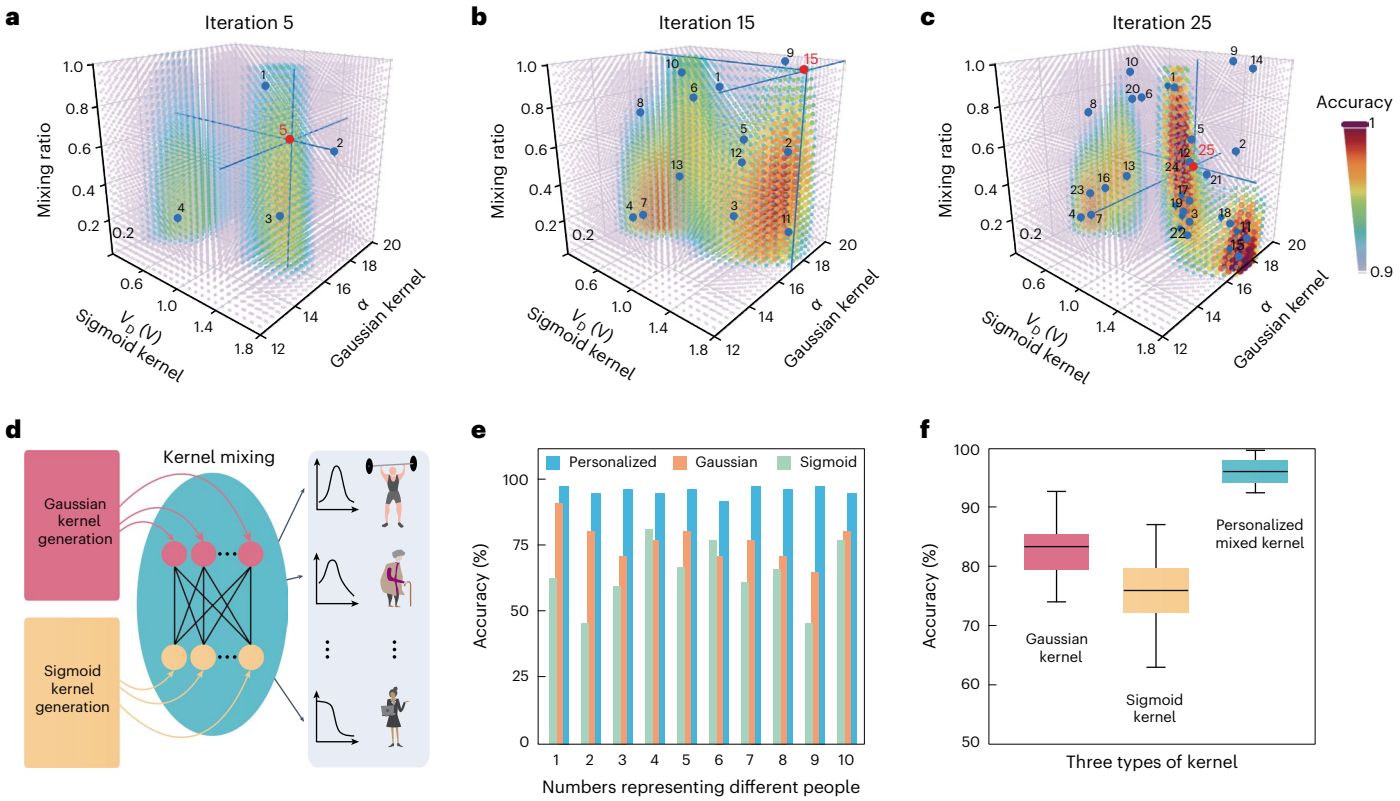

Fig. 3 | BO of personalized mixed kernels. a–**c**, BO of the Gaussian kernel hyperparameter, sigmoid kernel hyperparameter and mixing ratio after 5 (**a**), 15 (**b**) and 25 (**c**) iterations. **d**, Schematic of personalized SVM classification enabled by using two MKH transistors to generate a complete set of mixed kernels. **e**, Arrhythmia detection accuracy comparison for SVMs using only a Gaussian kernel, only a sigmoid kernel or a personalized mixed kernel for ten different patient datasets. **f**, Box plots derived from different patient datasets ($n = 100$), showing that the personalized mixed kernel has higher arrhythmia detection accuracy than SVMs using only a Gaussian kernel or only a sigmoid kernel. Box plots show median, 25th and 75th percentiles and whiskers that extend to the maximum and minimum classification accuracy.

lengths[14]. Therefore, the number of CMOS transistors quickly increases for fine-grained tunable Gaussian kernels. Alternatively, the addition of extra components to the bump circuit, such as operational transconductance amplifiers[34] and digital-to-analogue converters[10], can provide tunable Gaussian curves but also introduce additional scalability issues. For example, the operational transconductance amplifier approach requires matching of the channel length modulation parameters of short-channel transistors, whereas the digital-to-analogue converter approach has high power consumption (>100 μW) and large footprint (>0.02 mm^2) due to the peripheral supporting circuitry. Beyond bump circuits, alternative CMOS circuits that employ absoluters, squarers and exponentiators also require high power consumption and large footprints[35]. The generation of tunable sigmoid functions with conventional CMOS circuits face similar scalability issues as CMOS-tunable Gaussian circuits[36]. Although full-tunable CMOS mixed kernels can, in principle, be realized by combining the tunable Gaussian and tunable sigmoid CMOS circuits, the aforementioned impractical scaling issues have precluded any experimental demonstrations to date.

In comparison, MKH transistors feature a major advantage over CMOS technology for realizing a complete set of tunable SVM mixed kernels. For example, only two MKH transistors are required for a complete set of tunable mixed-kernel generation in contrast to the many building blocks used in CMOS implementations (Fig. 4a). Conservatively, CMOS requires more than 100 devices to generate a similarly complete set of tunable mixed kernels as can be achieved with only two MKH transistors (Fig. 4b). Consequently, the MKH approach provides a clear footprint advantage compared with CMOS for tunable mixed-kernel generation[37].

The scaling advantage of MKH transistors becomes even more evident when implementing an $n \times n$-kernel matrix. Mathematically, finding the primal SVM optimization is equivalent to solving a Lagrangian dual problem as defined in equation (1) by training n-input vectors $\mathbf{x}_i$:

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j \kappa(\mathbf{x}_i, \mathbf{x}_j)$$
$$\text{s.t.}\ \alpha_i \geq 0 \ \& \ \sum_{i=1}^{n} \alpha_i y_j = 0, i = 1, 2, \ldots, n \quad (1)$$

where α_i is the Lagrange multiplier corresponding to the inequality constraints in equation (2):

$$y_i (w^{\mathrm{T}} x_i + b) \geq 1, i = 1, 2, \ldots, n \quad (2)$$

Here y_i is the true classification label corresponding to $\mathbf{x}_i$, and $\kappa(\mathbf{x}_i, \mathbf{x}_j)$ is the kernel function operation on $\mathbf{x}_i$ and $\mathbf{x}_j$ after substituting the Karush–Kuhn–Tucker conditions into the primal Lagrangian. Correspondingly, the SVM hardware needs to implement an $n \times n$-kernel matrix to calculate the kernel function operation $\kappa(\mathbf{x}_i, \mathbf{x}_j)$ for each pair of input vectors (Fig. 4c). Each kernel cell in this matrix needs to generate a complete set of mixed kernels by having at least 100 CMOS devices or equivalently only two MKH transistors, leading to the plot shown in Fig. 4d that compares the total number of devices required for SVM hardware with personalized kernel functionality. In applications where the hardware footprint needs to be minimized (for example, wearable electronics), the MKH approach has clear advantages over conventional CMOS.

https://doi.org/10.1038/s41928-023-01042-7

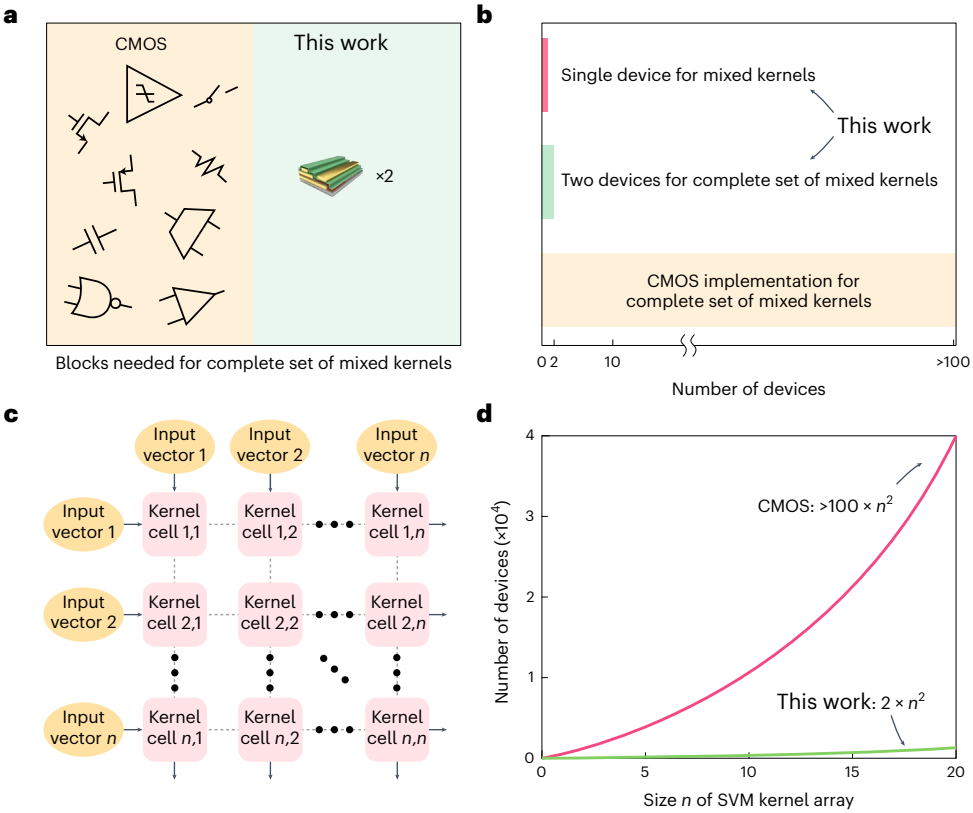

Fig. 4 | Mixed-kernel circuit complexity comparison with conventional CMOS. a, Comparison of the circuit blocks needed to generate a complete set of mixed kernels for CMOS and MKH implementations. CMOS circuits require many building blocks compared with only two MKH transistors. **b**, Comparison of the number of devices needed to generate a complete set of mixed kernels for CMOS and MKH implementations. CMOS circuits require more than 100 devices compared with only two MKH transistors. **c**, Schematic showing an $n \times n$-kernel matrix that is required to perform kernel computation for two n-input vectors in a hardware-level SVM implementation. **d**, Comparison of the number of devices needed to implement an $n \times n$-kernel matrix for CMOS and MKH implementations.

Finally, the MKH approach has reduced power consumption compared with CMOS for mixed-kernel SVM hardware. The tunable Gaussian, sigmoid and mixed kernels generated by MKH transistors (Fig. 1e–i) only require tens of nanowatts of power (for example, the estimated power consumption shown in Fig. 1e is 30 nA × 2 V = 60 nW). In contrast, tunable mixed kernels generated using conventional CMOS circuits require tens of microwatts to milliwatts depending on the specific architecture[35]. This power consumption difference is particularly important when implementing an $n \times n$-kernel matrix.

Conclusions

By tailoring the degree of electric-field screening through control over CNT density and overlap area, dual-gated MoS_2/CNT heterojunctions can enable the design of MKH transistors with tunable Gaussian, sigmoid and mixed-kernel functionality. The self-aligned, semi-vertical device geometry implies that a complete set of mixed Gaussian/sigmoid kernels can be simply achieved by varying the biases to the top and bottom gates. Using MKH transistors, we developed a mixed-kernel SVM platform for arrhythmia detection, where optimal Gaussian/sigmoid hyperparameters and mixed-kernel ratios were determined by BO to achieve exceptionally high arrhythmia detection accuracies that are superior to what can be obtained using purely Gaussian kernels. MKH transistors are also amenable to personalized kernels that provide arrhythmia detection accuracies approaching 95% for diverse patient profiles. Our MKH approach offers advantages over conventional CMOS implementations including simpler circuit designs, smaller footprints and lower power consumption. MKH transistors could, thus, be of use in a range of SVM classification applications in wearable and edge settings.

Methods

MKH transistor fabrication

Photolithography was performed on a Heidelberg MLA150 Maskless Aligner with an exposure wavelength of 375 nm and an exposure dosage of 750 mJ cm^{-2}, and on a Suss MABA6 Mask Aligner with an exposure wavelength of 365 nm and an exposure intensity of 10 mW cm^{-2}. A negative resist (NR9-1000PY, Futurrex) was used on undoped Si substrates, and baked for 1 min at 150 °C pre-exposure and 1 min at 100 °C post-exposure. Resist development was performed in 1:1 diluted RD6 (Futurrex), and liftoff was performed in *N*-methyl-2-pyrrolidone at 70 °C. Metal deposition was performed using an AJA electron-beam evaporator, and atomic layer deposition was performed using a Cambridge Nanotech ALD S100 device. Continuous monolayer MoS_2 was synthesized using CVD. Sulfur (S) powder (MilliporeSigma, 99.98%) and molybdenum trioxide (MoO_3) powder (MilliporeSigma, 99.97%) were used as chemical precursors, and single-crystal sapphire (MTI Corporation, <0001>) was used as the substrate. After growth, monolayer MoS_2 was transferred to the undoped Si substrate through a polycarbonate-assisted transfer process. The MoS_2 monolayer was patterned using a positive-resist bilayer of MicroChem PMGI baked at 170 °C and MicroChem S1813 baked at 115 °C. MoS_2 was etched by reactive ion etching with a Samco RIE-10NR using 50 s.c.c.m. Ar at 13.3 Pa and 50 W for 20 s, where select regions were encapsulated with 4 nm Al_2O_3 to protect from future etching steps. Semiconducting CNTs (IsoNanotubes-S 99% purity, NanoIntegris) were vacuum filtered onto a cellulose membrane (VMWP, 0.05 µm pore size, MilliporeSigma) and acetone bath transferred overnight onto the sample. The CNTs were patterned using S1813 resist, and etched with reactive ion etching using

20 s.c.c.m. O_2 at 26.5 Pa and 100 W for 10 s. The thickness of the metal electrodes is ~50 nm, and the thickness of the atomic-layer-deposited Al_2O_3 dielectric layer is 35 nm.

Materials characterization and electrical measurements

The thicknesses of the different device layers were characterized by atomic force microscopy in ambient conditions using an Asylum Cypher atomic force microscope. All the electrical measurements were performed in ambient conditions on a Cascade MicroTech semi-automated probe system using a Keithley 4200 semiconductor analyser.

SVM processing

An SVM algorithm finds an optimal separating hyperplane by maximizing the possible margins between the points that belong to different classes. It can be mathematically formulated as a linearly constrained quadratic programming problem, as shown in equation (3):

$$\underset{w,b}{\text{minimize}} f(w,b) = \frac{1}{2}||w||^2 \tag{3}$$

subject to the condition in equation (2), where $w \in \mathbb{R}^n$ gives the normal direction of the hyperplanes and b is a scalar. A dual problem can be equivalently derived with regard to the primal quadratic programming problem, which simplifies the computation. This dual Lagrangian formulation is described in equation (1).

SVM can also be generalized to linearly non-separable applications with the introduction of kernel functions. The kernel functions transform the input data into a higher-dimensional Hilbert feature space before performing linear separation. A kernel is essentially a symmetric function $K(x_i, x_j)$ under a necessary and sufficient condition given by Mercer's theorem:

$$\sum_{i=1}^{n}\sum_{j=0}^{n} \lambda_i \lambda_j K(x_i, x_j) \geq 0 \tag{4}$$

for any datasets $\mathbf{x}_{1:n} = \{x_1,\ldots,x_n\}$ and any real numbers $\lambda_{1:n} = \{\lambda_1,\ldots,\lambda_n\}$. In this study, we specifically utilize a Gaussian kernel, also known as the radial basis function kernel: $K(x_i, x_j) = e^{-||x_i - x_j||/2\sigma^2}$; a sigmoid kernel: $K(x_i, x_j) = \tanh(\gamma(x_i, x_j))$; and mixed kernels with tunable mixing ratios.

BO

In this study, we use a BO algorithm for efficient hyperparameter searching. BO is a method that uses prior observations of predefined loss f to determine the next search point. For an input dataset $\mathbf{x}_{1:n} = \{x_1,\ldots,x_n\}$, the loss function $f_{1:n} = \{f(\mathbf{x}_1),\ldots,f(\mathbf{x}_n)\}$ can be described by a Gaussian process in equation (5) for computing the prior distribution:

$$f_{1:n} \approx \mathcal{N}(m(\mathbf{x}_{1:n}), \mathbf{K}) \tag{5}$$

where the mean function is $m(\mathbf{x}_{1:n}) = [m(\mathbf{x}_1),\ldots, m(\mathbf{x}_n)]^T$, and the $n \times n$ covariance kernel matrix $\mathbf{K}$ is defined by the component in equation (6):

$$[K]_{ij} = k(\mathbf{x}_i, \mathbf{x}_j) \tag{6}$$

An acquisition function is a function of the posterior distribution over the loss function $f_{1:n}$. The next search points are determined by maximizing the expected improvement of the acquisition function, as defined in equation (7):

$$EI(\mathbf{x}) = \mathbb{E}[\max\{0, f(\mathbf{x}) - f(\hat{\mathbf{x}})\}], \tag{7}$$

where $\hat{\mathbf{x}}$ is the current optimal hyperparameter set, and the data points to sample in the next iteration are calculated using equation (8):

$$\mathbf{x}_{\text{new}} = \arg\max EI(\mathbf{x}). \tag{8}$$

Data availability

The data for all the figures in this Article are available via the Harvard Dataverse repository at https://doi.org/10.7910/DVN/0AAPVD.

Code availability

The code used in this study is available via GitHub at https://github.com/JennyMa0517/mixed-kernel-heterojunction.

References

1. Noble, W. S. What is a support vector machine? *Nat. Biotechnol.* **24**, 1565–1567 (2006).
2. Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L. & Lopez, A. A comprehensive survey on support vector machine classification: applications, challenges and trends. *Neurocomputing* **408**, 189–215 (2020).
3. Genov, R. & Cauwenberghs, G. Kerneltron: support vector 'machine' in silicon. *IEEE Trans. Neural Netw.* **14**, 1426–1434 (2003).
4. Devikanniga, D., Ramu, A. & Haldorai, A. Efficient diagnosis of liver disease using support vector machine optimized with crows search algorithm. *EAI Endorsed Trans. Energy Web* **7**, e10 (2020).
5. Wang, H., Zheng, B., Yoon, S. W. & Ko, H. S. A support vector machine-based ensemble algorithm for breast cancer diagnosis. *Eur. J. Oper. Res.* **267**, 687–699 (2018).
6. Ahmad, I., Basheri, M., Iqbal, M. J. & Rahim, A. Performance comparison of support vector machine, random forest, and extreme learning machine for intrusion detection. *IEEE Access* **6**, 33789–33795 (2018).
7. Afifi, S. M., GholamHosseini, H. & Sinha, R. Hardware implementations of SVM on FPGA: a state-of-the-art review of current practice. *Int. J. Innov. Sci. Eng. Technol.* **2**, 733–752 (2015).
8. Shoeb, A. H. & Guttag, J. V. Application of machine learning to epileptic seizure detection. In *Proc. 27th International Conference on Machine Learning (ICML-10)* 975–982 (PMLR, 2010).
9. Bin Altaf, M. A. & Yoo, J. A 1.83 µJ/classification, 8-channel, patient-specific epileptic seizure classification SoC using a non-linear support vector machine. *IEEE Trans. Biomed. Circuits Syst.* **10**, 49–60 (2016).
10. Kang, K. & Shibata, T. An on-chip-trainable Gaussian-kernel analog support vector machine. *IEEE Trans. Circuits Syst. I, Reg. Papers* **57**, 1513–1524 (2010).
11. Zhang, R. & Shibata, T. Fully parallel self-learning analog support vector machine employing compact Gaussian generation circuits. *Jpn. J. Appl. Phys.* **51**, 04DE10 (2012).
12. Alimisis, V., Gourdouparis, M., Dimas, C. & Sotiriadis, P. P. A 0.6 V, 3.3 nW, adjustable Gaussian crcuit for tunable kernel functions. In *2021 34th SBC/SBMicro/IEEE/ACM Symposium on Integrated Circuits and Systems Design (SBCCI)* 1–6 (IEEE, 2021).
13. Vrtaric, D., Ceperic, V. & Baric, A. Area-efficient differential Gaussian circuit for dedicated hardware implementations of Gaussian function based machine learning algorithms. *Neurocomputing* **118**, 329–333 (2013).
14. Reda Mohamed, A., Qi, L., Li, Y. & Wang, G. A generic nano-watt power fully tunable 1-D Gaussian kernel circuit for artificial neural network. *IEEE Trans. Circuits Syst., II, Exp. Briefs* **67**, 1529–1533 (2020).
15. Sangwan, V. K. & Hersam, M. C. Neuromorphic nanoelectronic materials. *Nat. Nanotechnol.* **15**, 517–528 (2020).
16. Sangwan, V. K. et al. Self-aligned van der Waals heterojunction diodes and transistors. *Nano Lett.* **18**, 1421–1427 (2018).
17. Sebastian, A., Pannone, A., Subbulakshmi Radhakrishnan, S. & Das, S. Gaussian synapses for probabilistic neural networks. *Nat. Commun.* **10**, 4199 (2019).
18. Beck, M. E. et al. Spiking neurons from tunable Gaussian heterojunction transistors. *Nat. Commun.* **11**, 1565 (2020).

19. Duong, D. L., Lee, S. M. & Lee, Y. H. Origin of unipolarity in carbon nanotube field effect transistors. *J. Mater. Chem.* **22**, 1994–1997 (2012).
20. Wang, K.-C. et al. Atomic-level charge transport mechanism in gate-tunable anti-ambipolar van der Waals heterojunctions. *Appl. Phys. Lett.* **118**, 083103 (2021).
21. Kim, C. H., Hayakawa, R. & Wakayama, Y. Fundamentals of organic anti-ambipolar ternary inverters. *Adv. Electron. Mater.* **6**, 1901200 (2020).
22. Wu, E. et al. Photoinduced doping to enable tunable and high-performance anti-ambipolar $MoTe_2/MoS_2$ heterotransistors. *ACS Nano* **13**, 5430–5438 (2019).
23. Kobashi, K., Hayakawa, R., Chikyow, T. & Wakayama, Y. Multi-valued logic circuits based on organic anti-ambipolar transistors. *Nano Lett.* **18**, 4355–4359 (2018).
24. Li, Y. et al. Anti-ambipolar field-effect transistors based on few-layer 2D transition metal dichalcogenides. *ACS Appl. Mater. Interfaces* **8**, 15574–15581 (2016).
25. Song, M. H., Lee, J., Cho, S. P., Lee, K. J. & Yoo, S. K. Support vector machine based arrhythmia classification using reduced features. *Int. J. Control Autom. Syst.* **3**, 571–579 (2005).
26. Asl, B. M., Setarehdan, S. K. & Mohebbi, M. Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal. *Artif. Intell. Med.* **44**, 51–64 (2008).
27. Moody, G. B. & Mark, R. G. The impact of the MIT-BIH arrhythmia database. *IEEE Eng. Med. Biol. Mag.* **20**, 45–50 (2001).
28. Goldberger, A. L. et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex phiysologic signals. *Circulation* **101**, e215–e220 (2000).
29. Shahriari, B., Swersky, K., Wang, Z., Adams, R. P. & De Freitas, N. Taking the human out of the loop: a review of Bayesian optimization. *Proc. IEEE* **104**, 148–175 (2015).
30. Delbruck, T. 'Bump' circuits for computing similarity and dissimilarity of analog voltages. In *IJCNN-91-Seattle International Joint Conference on Neural Networks* 475–479 (IEEE, 1993).
31. Verleysen, M., Thissen, P., Voz, J.-L. & Madrenas, J. An analog processor architecture for a neural network classifier. *IEEE Micro* **14**, 16–28 (1994).
32. Nam, M. & Cho, K. Implementation of real-time image edge detector based on a bump circuit and active pixels in a CMOS image sensor. *Integration* **60**, 56–62 (2018).
33. Payvand, M. & Indiveri, G. Spike-based plasticity circuits for always-on on-line learning in neuromorphic systems. In *2019 IEEE International Symposium on Circuits and Systems (ISCAS)* 1–5 (IEEE, 2019).
34. Lu, J., Yang, T., Jahan, M. & Holleman, J. Nano-power tunable bump circuit using wide-input-range pseudo-differential transconductor. *Electron. Lett.* **50**, 921–923 (2014).
35. Alimisis, V., Gourdouparis, M., Gennis, G., Dimas, C. & Sotiriadis, P. P. Analog Gaussian function circuit: architectures, operating principles and applications. *Electronics* **10**, 2530 (2021).
36. Youssefi, B., Leigh, A. J., Mirhassani, M. & Wu, J. Tunable neuron with PWL approximation based on the minimum operator. *IEEE Trans. Circuits Syst., II, Exp. Briefs* **66**, 387–391 (2018).
37. Yan, X., Qian, J. H., Sangwan, V. K. & Hersam, M. C. Progress and challenges for memtransistors in neuromorphic circuits and systems. *Adv. Mater.* **34**, 2108025 (2022).

Acknowledgements

Device fabrication was supported by the DOE ASCR and BES Microelectronics Threadwork Program, which is funded by the US Department of Energy (DOE), Office of Science, under contract no. DE-AC02-06CH11357. Device testing was supported by the National Science Foundation Materials Research Science and Engineering Center at Northwestern University under contract no. DMR-1720139 and by the National Science Foundation Neuroplane Program under contract no. CCF-2106964. SVM analysis and BO were supported by the Army Research Office under contract no. W911NF-21-2-0128. Materials growth and characterization were supported by the Laboratory Directed Research and Development Program at Sandia National Laboratories (SNL). SNL is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the US DOE National Nuclear Security Administration under contract DE-NA0003525. This work describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the United States Government. This work made use of the Northwestern University Micro/Nano Fabrication Facility (NUFAB), which has received support from the Soft and Hybrid Nanotechnology Experimental (SHyNE) Resource (NSF ECCS-1542205), the Materials Research Science and Engineering Center (NSF DMR-1720139) and the State of Illinois.

Author contributions

M.C.H., H.W. and V.K.S. conceived the idea. X.Y. and J.H.Q. designed the experiments and performed the device fabrication, measurements and analysis. J.M. performed the SVM analysis and BO. A.Z. and X.W. assisted with the circuit design, analysis and comparison. S.E.L., M.P.B. and K.J.L. synthesized and characterized the MoS_2 films. All authors discussed the results and contributed to the writing of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41928-023-01042-7.

Correspondence and requests for materials should be addressed to Vinod K. Sangwan, Han Wang or Mark C. Hersam.

Peer review information *Nature Electronics* thanks Kyusang Lee and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.