

ARTICLE

<https://doi.org/10.1038/s41467-019-12035-6>

OPEN

Gaussian synapses for probabilistic neural networks

Amritanand Sebastian ¹, Andrew Pannone¹, Shiva Subbulakshmi Radhakrishnan ^{1,2} & Saptarshi Das ^{1,3,4}

The recent decline in energy, size and complexity scaling of traditional von Neumann architecture has resurrected considerable interest in brain-inspired computing. Artificial neural networks (ANNs) based on emerging devices, such as memristors, achieve brain-like computing but lack energy-efficiency. Furthermore, slow learning, incremental adaptation, and false convergence are unresolved challenges for ANNs. In this article we, therefore, introduce Gaussian synapses based on heterostructures of atomically thin two-dimensional (2D) layered materials, namely molybdenum disulfide and black phosphorus field effect transistors (FETs), as a class of analog and probabilistic computational primitives for hardware implementation of statistical neural networks. We also demonstrate complete tunability of amplitude, mean and standard deviation of the Gaussian synapse via threshold engineering in dual gated molybdenum disulfide and black phosphorus FETs. Finally, we show simulation results for classification of brainwaves using Gaussian synapse based probabilistic neural networks.

¹Department of Engineering Science and Mechanics, Pennsylvania State University, University Park, PA 16802, USA. ²Electrical and Electronics Engineering, Amrita Vishwa Vidyapeetham, Ettimadai, Coimbatore, Tamil Nadu 641112, India. ³Department of Materials Science and Engineering, Pennsylvania State University, University Park, PA 16802, USA. ⁴Materials Research Institute, Pennsylvania State University, University Park, PA 16802, USA. Correspondence and requests for materials should be addressed to S.D. (email: sud70@psu.edu)

The last five decades have witnessed an unprecedented and exponential growth in computational power, primarily driven by the success of the semiconductor industry. Relentless scaling¹ of complementary metal oxide semiconductor (CMOS) technology enabled by breakthroughs in material discovery², innovation in device physics³, transformation in micro and nanolithography techniques⁴, and the triumph of von Neumann architecture⁵ contributed to the computing revolution. Scaling has three characteristic aspects to it. Energy scaling to ensure practically constant computational power budget, size scaling to ensure faster and cheaper computing since more and more transistors can be packed into the same chip area, and complexity scaling to ensure incessant growth in computational power of single on-chip processor. The golden era of metal oxide semiconductor field effect transistor (MOSFET) scaling, also referred to as the Dennard scaling era⁶, has witnessed concurrent scaling of all three aspects for almost three decades. However, around 2005, the energy scaling ended owing to fundamental thermodynamic limitations at the device physics level, popularly known as the Boltzmann tyranny⁷. Size scaling continued for another decade albeit with new challenges⁸ and eventually ended in 2017 owing to limitations at the materials level imposed by quantum mechanics¹. Complexity scaling is also on decline owing to the non-scalability of traditional von Neumann computing architecture and the impending “Dark Silicon” era that presents a severe threat to multi-core processor technology⁹. In order to sustain the growth in computational power, it is imperative that all three aspects of scaling must be reinstated immediately through material rediscovery, device innovations, and advancement in higher complexity computing architectures.

The extraordinarily complex neurobiological architecture of the mammalian nervous system that seamlessly executes diverse and intricate functionalities such as adaption, perception, acquisition of sensory information, learning, memory formation, emotions, cognition, motor action, and many more has inspired computer scientists to think beyond the traditional von Neumann architecture in order to resurrect complexity scaling. The neural architecture deploys billions of information processing units, neurons, which are connected via trillions of synapses in order to accomplish massively parallel, synchronous, coherent, and concurrent computation. This is markedly different from the von Neumann architecture, where the logic and memory units are physically isolated and operate sequentially, i.e., instruction fetch and data operation cannot occur simultaneously. Furthermore, unlike the deterministic digital switches (logic transistors), neural architecture uses probabilistic and analog computational primitives in order to accomplish adaptive functionalities such as pattern recognition and pattern classification, which form the foundation for mammalian problem solving and decision making.

In the above context, IBM’s bioinspired CMOS chip, True North, is a remarkable breakthrough in neuromorphic computing, achieving the complexity of more than 1 million neurons or 256 million synapses while consuming a miniscule 70 mW of power¹⁰. Similarly, extensive work by Luca Benini et al. have recently shown that hardware digital neural networks consume comparable or less energy than the human brain for complex tasks, such as, image recognition¹¹. While these are impressive advancements, the inherent scaling challenges associated with the digital CMOS technology can ultimately limit the implementation of very-large-scale artificial neural networks (ANNs), invoking the critical and imminent need for energy efficient analog computing primitives for ANNs. Recent years have, therefore, witnessed innovation in analog devices such as the memristors^{12–14}, coupled oscillators¹⁵, and various targeted components^{16–18}, which can emulate neural spiking, neural transmission, and neural plasticity and hence can be used as computational

primitives in ANNs. While these devices do provide some energy benefit at the architectural level for specific applications, they fail to address the intrinsic energy and size scaling needs at the device level, which can ultimately lead to stagnation in complexity scaling. Further challenges associated with ANNs are often overlooked. For example, ANNs deployed for pattern classification problems require optimum training algorithms and learning rules to identify the class statistics with desired accuracy within a short training time. Unfortunately, the most popular and widely used heuristic backpropagation algorithm¹⁹, is inherently slow and remains vulnerable to local minima in spite of extensive modifications in recent years using methods such as conjugate gradient, quasi-Newton, and Levenberg–Marquardt (LM) to improve the convergence²⁰. In order to address the slow learning, incremental adaptation, and inherent unreliability of ANNs, novel classification techniques based on statistical principles must be embraced.

In this article, we experimentally demonstrate how a new class of analog devices, namely, Gaussian synapses, based on the heterostructure of novel atomically thin two dimensional (2D) layered semiconductors enables the hardware implementations of probabilistic neural networks (PNNs) and thereby reinstates all three aforementioned quintessential scaling aspects of computing. In short, 2D materials facilitate aggressive size scaling, analog Gaussian synapses offer energy scaling, and PNNs enable complexity scaling. Combined, these new developments can facilitate *Exascale Computing* and ultimately benefit scientific discovery, national security, energy security, economic security, infrastructure development, and advanced healthcare programs^{21,22}.

Results

Probabilistic neural network. Our overall approach is summarized in Fig. 1a. First, we reintroduce PNN that was proposed by Specht, D. F.²³ PNN is derived from Bayesian computing and Kernel density method²⁴. As shown in Fig. 1b, unlike ANNs, which necessitate multiple hidden layers, each with a large number of nodes, PNNs are comprised of a pattern layer and a summation layer and can map any input pattern to any number of output classifications. Furthermore, ANNs use activation functions such as sigmoid, rectified linear unit (ReLU), and their various derivatives for determining pattern statistics. This is often extremely difficult to accomplish with reasonable accuracy for non-linear decision boundaries. PNNs, on the contrary, use parent probability distribution functions (PDFs) approximated by Parzen window²⁵ and a non-parametric function to define the class probability, which in the case of Gaussian kernel is defined by Gaussian distribution as shown in Fig. 1c. PNNs, therefore, facilitate seamless and accurate classification of complex patterns with arbitrarily shaped decision boundaries. Furthermore, PNNs can be extended to map higher dimensional functions since multivariate Gaussian kernels are simply the product of univariate kernels as shown in Fig. 1d.

Unfortunately, in spite of the widespread applications and simplicity of PNNs, their hardware implementation is rather sparse. In fact, neither ANNs nor PNNs have been extensively realized using hardware components. While there is growing interest towards the development of devices for the hardware implementation of ANNs, the effort and investment towards the hardware implementation of PNNs are still very limited. One reason is that the hardware implementation of probability functions associated with the PNNs, such as the Gaussian, requires multicomponent digital CMOS circuits that leads to severe area and energy inefficiency. For example, the “Bump” circuit demonstrated by Delbruck, T. uses seven transistors²⁶. Similarly, the Gaussian synapse proposed by Choi, J. et al. consists

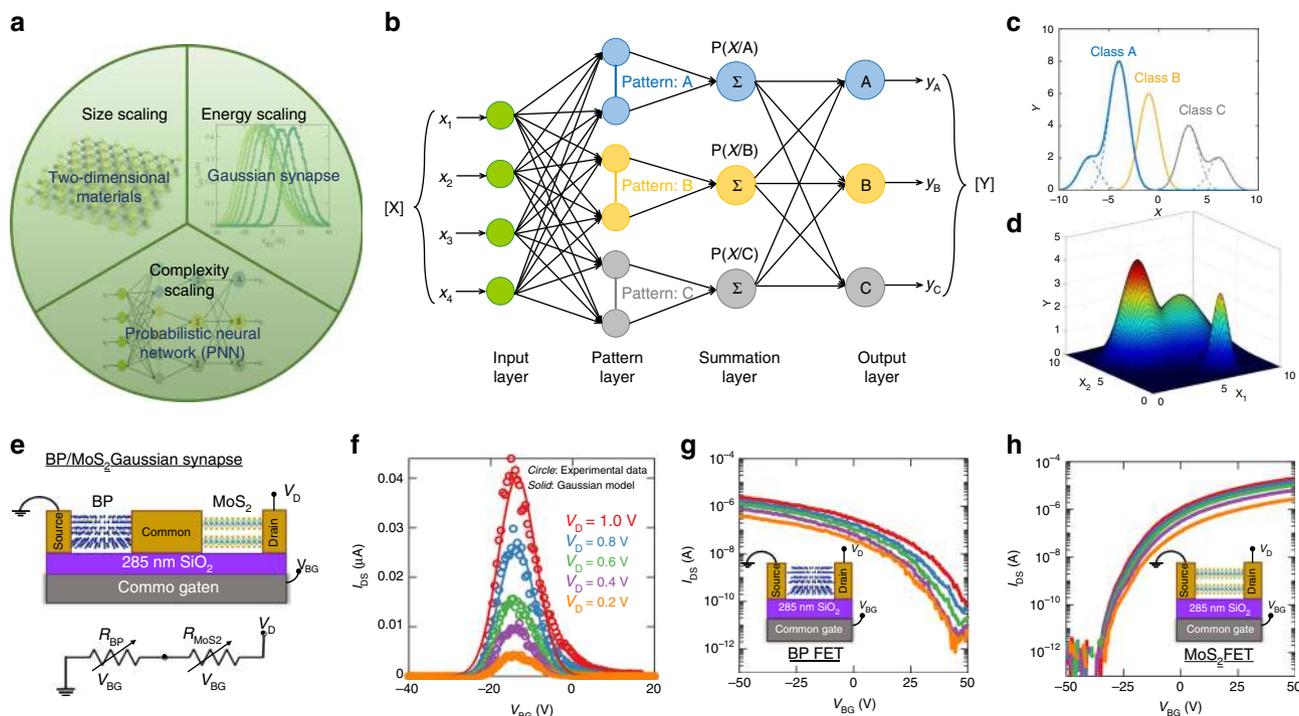


Fig. 1 Gaussian Synapse based probabilistic neural network (PNN). **a** Resurrection of three quintessential scaling aspects of computation i.e., complexity scaling through PNNs, size scaling through atomically thin 2D materials and energy scaling through analog Gaussian synapses. **b** Schematic representation of PNN that comprise of a pattern layer and a summation layer for mapping any input pattern to any number of output classifications. **c** Gaussian probability density functions (PDFs) facilitating seamless and accurate classification of complex patterns with arbitrarily shaped decision boundaries. **d** Multivariate Gaussian kernel for mapping higher dimensional functions. **e** Schematic of two transistor Gaussian synapse based on heterogeneous integration of n-type MoS₂ and p-type black phosphorus (BP) back-gated field effect transistors (FETs). The equivalent circuit diagram consists of two variable resistors connected in series. **f** Transfer characteristics i.e., the drain current (I_D) versus back-gate voltage (V_{BG}) of the Gaussian synapse for different drain voltages (V_D). Clearly, the experimental data (circles) can be modeled by Gaussian distributions (solid). **g** Transfer characteristics of p-type BP FET. **h** Transfer characteristics of n-type MoS₂ FET

of a pair of differential amplifiers and several arithmetic computational units²⁷. Madrenas, J. et al. introduced an alternate method to obtain Gaussian function by combining the exponential characteristics of MOSFETs in sub-threshold and square characteristics in inversion²⁸. The circuit was further improved for better symmetry and greater control and tunability by adding more transistors in a floating gate configuration^{29,30}. Another approach is to use a Gilbert Gaussian function, where two sigmoid curves are combined using a differential pair along with a current mirror^{31,32}. Without any extra circuitry to reduce asymmetry, the most compact circuit uses five transistors²⁸. As we will discuss in the following section, our experimental demonstration of Gaussian synapses uses only two transistors, which significantly improves the area and energy efficiency at the device level and provides cascading benefits at the circuit, architecture, and system levels. This will stimulate the much-needed interest in the hardware implementation of PNNs for a wide range of pattern classification problems.

Gaussian synapse. Figure 1e shows the schematic of our proposed two transistor Gaussian synapse based on heterogeneous integration of n-type molybdenum disulfide (MoS₂) and p-type black phosphorus (BP) back-gated field effect transistors (FETs). Fig. 1e also shows the equivalent circuit diagram for the Gaussian synapse, which simply consists of two variable resistors in series. The two variable resistors, i.e., R_{MoS_2} and R_{BP} correspond to the MoS₂ and BP FETs. Fig. 1f shows the experimentally measured transfer characteristics i.e., the drain current (I_D) versus back-gate voltage (V_{BG}) of the Gaussian synapse for different drain voltages

(V_D). The fabrication process flow and electrical measurement setup for Gaussian synapses are described in the experimental method section. Clearly, the transfer characteristics resemble a Gaussian distribution which can be modeled using the following equation.

$$I_D = \frac{A}{\sqrt{2\pi\sigma_V^2}} \exp\left[-\frac{(V_{BG} - \mu_V)^2}{2\sigma_V^2}\right]; A = \beta V_D \quad (1)$$

Where, A , μ_V and σ_V are, respectively, the amplitude, mean, and standard deviation of the Gaussian. For a specific MoS₂/BP pair, μ_V and σ_V are found to be constants, whereas, A varies linearly with V_D . The emergence of Gaussian transfer characteristics can be explained using the experimentally measured transfer characteristics of its constituents, i.e., the MoS₂ FET and the BP FET, as shown in Fig. 1g, h, respectively. MoS₂ FETs exhibit unipolar n-type characteristics, irrespective of the choice of contact metal, owing to the phenomenon of metal Fermi level pinning close to the conduction band that facilitates easier electron injection, whereas, BP FETs are predominantly p-type with large work function contact metals such as Ni³³⁻³⁶. Furthermore, unlike conventional enhancement mode Si FETs used in CMOS circuits, both MoS₂ and BP FETs are depletion mode, i.e., they are normally ON without applying any back-gate voltage. Remarkably, this simple difference results in the unique Gaussian transfer characteristics for the MoS₂/BP pair in spite of the device structure closely resembling a CMOS logic inverter. From the equivalent circuit diagram, the current (I_D) through the

Gaussian synapse can be written as:

$$I_D = \frac{V_D}{R_{\text{MoS}_2} + R_{\text{BP}}} \quad (2)$$

For extreme V_{BG} values, i.e., large negative (lesser than -30 V) and large positive (greater than 30 V), the MoS_2 FET and the BP FET are in their respective OFF states, making the corresponding resistances, i.e., R_{MoS_2} and R_{BP} very large (approximately Ω). This prevents any current conduction between the source and the drain terminal of the Gaussian synapse. However, as the MoS_2 FET switches from OFF state to ON state, current conduction begins and increases exponentially with V_{BG} following the subthreshold characteristics and reaches its peak magnitude determined by V_D . Beyond this peak, the current starts to decrease exponentially following the subthreshold characteristics of the BP FET. As a result, the series connection of the MoS_2 and BP FETs exhibits non-monotonic transfer characteristics with exponential tails that mimics a Gaussian distribution.

It must be noted that the Gaussian synapses do not utilize the ON state FET performance and, therefore, are minimally influenced by the carrier mobility values of the n-type and p-type FETs. Instead, the Gaussian synapse exploits the subthreshold FET characteristics, where the slope is independent of the carrier mobility of the semiconducting channel material. For symmetric Gaussian synapses, it is therefore more desirable to ensure similar sub-threshold slope (SS) values for the respective FETs than the carrier mobility. Ideally, the SS values for both FETs should be 60 mV decade $^{-1}$. However, presence of a nonzero interface trap capacitance worsens the SS. The SS can be improved by minimizing interface states at the 2D/gate-dielectric interface, as well as by scaling the thickness of the gate dielectric. It is also desirable to have FETs with Ohmic contacts for Gaussian synapses to ensure that the SS is determined by the thermionic emission of carriers in order to reach the minimum theoretical value of 60 mV decade $^{-1}$ at room temperature. For Schottky contact FETs, the SS can be severely degraded due to tunneling of carriers through the Schottky barrier.

While our proof-of-concept demonstration of Gaussian synapses is based on exfoliated MoS_2 and BP flakes, it is well known that the micromechanical exfoliation is not a scalable manufacturing process for large-scale integrated circuits. Therefore, hardware implementation of PNNs using Gaussian synapses will necessitate large-area growth of MoS_2 and BP. Fortunately, recent years have seen tremendous progress in wafer-scale growth of high quality MoS_2 and BP using chemical vapor deposition (CVD) and metal organic chemical vapor deposition (MOCVD) techniques^{37–41}. Furthermore, while we have used two different 2D materials, MoS_2 and BP, for our demonstration of Gaussian synapses owing to their superior performance as n-type and p-type FETs, respectively, there are 2D materials, such as, WSe_2 , which offer ambipolar transport, i.e., the presence of both electron and hole conduction⁴² and can be grown over large area using CVD techniques⁴³. However, the performance of WSe_2 based n-type and p-type FETs are limited by the presence of large Schottky barriers at the metal/2D contact interfaces³⁶. By resolving the contact resistance related issues³⁶ and improving the quality of large-area synthesized WSe_2 , it is possible to implement Gaussian synapses based solely on WSe_2 as well. Moreover, in recent years several groups have demonstrated p-type MoS_2 and n-type BP, by implementing smart contact engineering and/or doping strategies^{44,45}. Therefore, very-large-scale integration of Gaussian synapses based on a CVD grown single 2D material will be possible in the near future for the hardware realization of PNNs. Since the focus of this article is to

introduce the novel Gaussian synapse and its benefit as a statistical computing primitive, we avoided material optimization.

Gaussian synapses are inherently low power since they exploit the subthreshold characteristics of the FET devices. In this context, we would like to remind the readers that the total power consumption (P_{total}) in digital CMOS circuit comprises, primarily, of dynamic switching power (P_{dynamic}) and static leakage power (P_{static}) and is given by the following equation:

$$P_{\text{total}} = P_{\text{dynamic}} + P_{\text{static}} = \eta C V_{\text{DD}}^2 f + I_{\text{static}} V_{\text{DD}} \quad (3)$$

Note that, η is the activity factor, C is the capacitance of the circuit, f is the switching frequency, and V_{DD} is the supply voltage. During the Dennard scaling era, the power consumption of the chip was dominated by P_{dynamic} , which was kept constant by scaling the threshold voltage (V_{TH}) and concurrently the supply voltage (V_{DD}) of the MOSFET. However, beyond 2005, the voltage scaling stalled since further reduction in V_{TH} resulted in an exponential increase in the static leakage current (I_{static}) and hence static power consumption. This is a direct consequence of the non-scalability of the subthreshold swing (SS) to below 60 mV decade $^{-1}$, as determined by the Boltzmann statistics. In fact, P_{total} in the present Dark-Si era is mostly dominated by P_{static} . Regardless of whether the dynamic or static power dominates, reinstating V_{DD} scaling is the only way to escape the Boltzmann tyranny. This is why in recent years, subthreshold logic circuits, which utilize V_{DD} that is close to or even less than V_{TH} , have received significant attention for ultra-low power applications^{46,47}. New subthreshold logic and memory design methodologies have already been developed and demonstrated on a fast Fourier transform (FFT) processor⁴⁸, as well as analog VLSI neural systems⁴⁹.

Note that our proposed Gaussian synapses naturally require operation in subthreshold regime in order to exploit the exponential feature in the transfer characteristics of the n-type and p-type transistors. Furthermore, as shown in Fig. 1f, Gaussian synapses maintain their characteristic features even when the supply voltage (V_{DS}) is scaled down to 200 mV or beyond. This allows the Gaussian synapses to be inherently low power. For the proof-of-concept demonstration of Gaussian synapses, we have used relatively thicker back-gate and top-gate oxides with respective thicknesses of 285 nm and 120 nm. This necessitates the use of rather large back-gate and top-gate voltages in the range of -50 V to 50 V and -35 V to 35 V, respectively. The power consumption by our proof-of-concept Gaussian synapses will still be high in spite of scaling the supply voltage and exploiting the subthreshold current conduction in the range of nano amperes through the MoS_2 and BP FETs. This is because power consumption by Gaussian synapse is simply the area under the I_D versus V_{BG} curve. By scaling the thicknesses of both the top and bottom gate dielectrics, it is possible to scale the operating gate voltages and thereby achieve desirable power benefits from the Gaussian synapses. Ultra-thin dielectric materials such as Al_2O_3 and HfO_2 , which offer much larger dielectric constants, ϵ_{ox} , of ≈ 9 and 25 , respectively, are now routinely used as gate oxides for highly scaled Si FinFETs⁵⁰. It must also be emphasized that the use of atomically thin 2D materials allows geometric miniaturization of Gaussian synapses without any loss of electrostatic integrity, which aids size scaling. We would like to remind the readers that the scalability of FETs is captured through a simple parameter called the screening length ($\lambda_{\text{SC}} = \sqrt{\frac{\epsilon_{\text{body}}}{\epsilon_{\text{ox}}} t_{\text{body}} t_{\text{ox}}}$), which determines the decay of the potential (band bending) at the source/drain contact interface into the semiconducting channel. In this expression, t_{body} and t_{ox} are the thicknesses and ϵ_{body} and ϵ_{ox} are the dielectric constants of

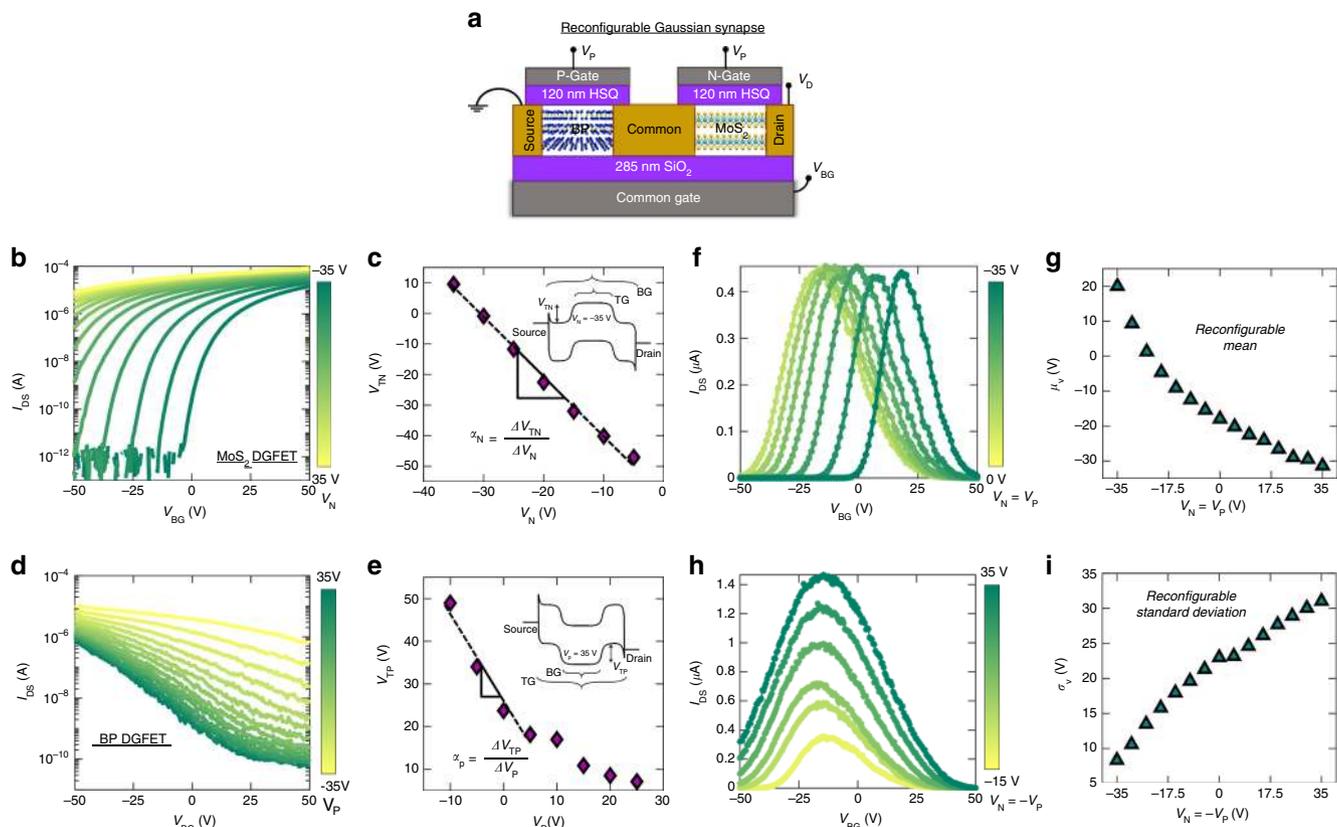


Fig. 2 Reconfigurable Gaussian synapse. **a** Schematic of a reconfigurable Gaussian synapse involving dual-gated n-type MoS₂ FET and p-type BP FET. The top-gate stack was fabricated using hydrogen silsesquioxane (HSQ) as the top-gate dielectric and nickel/gold (Ni/Au) as the top-gate electrode. **b** Back-gated transfer characteristics of the MoS₂ FET at V_D = 1 V, for different top-gate voltages (V_N). **c** Back-gate threshold voltage, V_{TN} of MoS₂ FET as function of V_N was extracted using the constant current method. Inset shows the band diagram elucidating how V_N controls V_{TN} by electrostatically adjusting the height of the thermal energy barrier for electron injection into the MoS₂ channel. **d** Back-gated transfer characteristics of the BP FET at V_D = 1 V, for different top-gate voltages (V_P). **e** Back-gate threshold voltage, V_{TP} of BP FET as function of V_P. Inset shows the band diagram elucidating how V_P controls V_{TP} by electrostatically adjusting the height of the thermal energy barrier for hole injection into the BP channel. As expected the slope (α_N = 1.91) of V_{TN} versus V_N and the slope (α_P = 2) of V_{TP} versus V_P are found to be similar and equal to the ratio of top-gate capacitance (C_{TG}) to the back-gate capacitance (C_{BG}), which in our case was ≈1.94. **f** Transfer characteristics of the Gaussian synapse for different values of V_N = V_P. This configuration allows us to shift the mean (μ_V) of the Gaussian synapse without changing the amplitude (A) and the standard deviation (σ_V). **g** μ_V as a function of V_N = V_P. **h** Transfer characteristics of the Gaussian synapse for different values of V_N = -V_P. This configuration allows us to configure σ_V while keeping μ_V constant. **i** σ_V as a function of V_N = -V_P. However, this configuration also results in an increase in the amplitude (A) of the Gaussian synapse as σ_V increases. This increase can be adjusted by changing the drain voltage (V_D) since A is linearly proportional to V_D. Nevertheless, by controlling V_N, V_P and V_D, it is possible to adjust the mean, standard deviation and amplitude of the Gaussian synapse

the semiconducting channel and the insulating oxide, respectively. As discussed by Frank et al.⁵¹, to avoid short channel effects the channel length of an FET (L_{CH}) has to be at least three times higher than the screening length (λ_{SC}). For atomically thin semiconducting monolayers of 2D materials, t_{body} is 0.6 nm, which corresponds to λ_{SC} of 1.3 nm, whereas, for the most advanced FinFET technology the thickness of Si fins can be scaled down to only 5 nm without severely increasing the bandgap due to quantum confinement effects and reducing the mobility due to enhanced surface roughness scattering. Nevertheless, the above discussions, clearly articulate how BP/MoS₂ 2D heterostructure based Gaussian synapses can facilitate effortless hardware realization of PNNs and thereby aid complexity scaling without compromising energy and size scaling.

Reconfigurable Gaussian synapse. For the hardware implementation of Gaussian synapses, it is highly desirable to demonstrate complete tunability of the device transfer function, i.e., A, μ_V, and σ_V of the Gaussian distribution. This could be

achieved, seamlessly, in our device structure *via* threshold engineering through additional gating of either or both MoS₂ and BP FETs. Fig. 2a shows the schematic representation of a reconfigurable Gaussian synapse, where, both MoS₂ and BP FETs are dual-gated (DG). The top-gate stack was fabricated using hydrogen silsesquioxane (HSQ)^{52,53} as the top-gate dielectric with nickel/gold (Ni/Au) as the top-gate electrode. The fabrication process flow is described in the experimental method section. Fig. 2b shows the experimentally measured back-gated transfer characteristics of the MoS₂ FET at V_D = 1 V for different top-gate voltages (V_N). Clearly, V_N controls the back-gate threshold voltage, V_{TN} of the MoS₂ FET as shown in Fig. 2c. The energy band diagram shown in the inset of Fig. 2c can be used to explain the concept of threshold voltage engineering using gate electrostatics. The top-gate voltage determines the height of the potential barrier for electron injection inside the MoS₂ channel, which must be overcome by applying a back-gate voltage to enable current conduction from the source to the drain terminal. Negative top-gate voltages increase the potential barrier for electron injection and hence necessitate larger positive back-gate voltages to achieve

similar level of current conduction. As such, V_{TN} becomes more positive (less negative) for large negative V_N values. Note that the slope (α_N) of V_{TN} versus V_N in Fig. 2c must be proportional to the ratio of top-gate capacitance (C_{TG}) to the back-gate capacitance (C_{BG}). This follows directly from the principle of charge balance, which ensures that the channel charge induced by the top-gate voltage must be compensated by the back-gate voltage at threshold. We extracted the value for α_N to be 1.91. This is consistent with the theoretical prediction of approximately 1.94, given that the top-gate and back-gate dielectric thicknesses are 120 nm and 285 nm, respectively and the top-gate insulator, HSQ, has a slightly lower dielectric constant (≈ 3.2) than the back-gate insulator, SiO₂ (3.9). Fig. 2d shows the experimentally measured back-gated transfer characteristics of the BP FET at $V_D = 1$ V, for different top-gate voltages (V_p). As expected, V_p controls the back-gate threshold voltage, V_{TP} of the BP FET as shown in Fig. 2e. Here, the top-gate voltage influences the height of the potential barrier for hole injection, which is overcome by applying a back-gate voltage, enabling current conduction from the drain to the source terminal. The corresponding energy band diagram is shown in the inset of Fig. 2e. Positive top-gate voltages increase the potential barrier for hole injection and hence necessitate larger negative back-gate voltages to achieve similar level of current conduction. As such, V_{TP} becomes more negative (less positive) for large positive V_p values. We also extracted the slope (α_p) of V_{TP} versus V_p in Fig. 2e and, as expected, found a similar value of ≈ 2 .

The dual-gated MoS₂ and BP FETs allow complete control of the shape of the Gaussian synapse. Fig. 2f shows the experimentally measured transfer characteristics of the Gaussian synapse for different values of $V_N = V_p$. This configuration allows us to shift the mean (μ_V) of the Gaussian synapse without changing the amplitude (A) or the standard deviation (σ_V) Fig. 2g shows μ_V plotted as a function of $V_N = V_p$. We are able to do this since the back-gate threshold voltages for both MoS₂ and BP FETs shift in the same direction in this configuration. Similarly, Fig. 2h shows the experimentally measured transfer characteristics of the Gaussian synapse for different values of $V_N = -V_p$. Under this configuration, the back-gate threshold voltages for MoS₂ and BP FETs shift in opposite directions. As such the μ_V of the Gaussian distribution remains constant, whereas, σ_V keeps increasing. Fig. 2i shows σ_V plotted as a function of $V_N = -V_p$. However, this configuration also results in an increase in the amplitude (A) of the Gaussian synapse as σ_V increases. This increase can be adjusted by changing the drain voltage (V_D) since A is linearly proportional to V_D . Nevertheless, by controlling V_N , V_p , and V_D , it is possible to adjust the mean, standard deviation, and amplitude of the Gaussian synapse.

Scaled Gaussian synapses. In order to project the performance of scaled Gaussian synapses, we used the Virtual Source (VS) model that was originally developed by Khakifirooz, A. et al. for short channel Si MOSFETs⁵⁴, for short channel Si MOSFETs. The Gaussian transfer characteristics (I_D versus V_G for different V_D) were simulated in the following Eqs. 4, 5, and 6. In the VS model, both the subthreshold and the above threshold behavior is captured through a single semi-empirical and phenomenological relationship that describes the transition in channel charge density from weak to strong inversion (Eq. 5).

$$I_D = \frac{V_D}{R_N + R_p}; R_N = \frac{L_N}{W_N \mu_N Q_N}; R_p = \frac{L_p}{W_p \mu_p Q_p}; \quad (4)$$

$$Q_N = C_{BG} m \frac{k_B T}{q} \log \left[1 + \exp \left(\frac{V_G - V_{TN}}{m k_B T / q} \right) \right]; \quad (5)$$

$$Q_p = C_{BG} m \frac{k_B T}{q} \log \left[1 + \exp \left(- \frac{V_G - V_{TP}}{m k_B T / q} \right) \right];$$

$$V_{TN} = \alpha V_N; V_{TP} = \alpha V_p; \alpha = C_{TG} / C_{BG}; \quad (6)$$

Here, R_N and R_p are the resistances, L_N and L_p are the lengths, W_N and W_p are the widths, μ_N and μ_p are the carrier mobility values, and Q_N and Q_p are the inversion charges corresponding to the n-type and the p-type 2D-FETs, respectively. The band movement factor m can be assumed to be unity for a fully depleted and ultra-thin body 2D-FET with negligible interface trap capacitance. Finally, V_{TN} and V_{TP} are the threshold voltages of the n-type and p-type 2D FETs determined by their respective top-gate voltages V_N and V_p . Note, that in the subthreshold regime, the inversion charges i.e., Q_N and Q_p increase exponentially with V_G , whereas above threshold, the inversion charge is a linear function of V_G , which is seamlessly captured through the VS model. Fig. 3a, b show the simulated transfer characteristics of the individual n-type and p-type 2D FETs, respectively, and Fig. 3c shows the transfer characteristics of the Gaussian synapse based on their heterostructure for different combinations of the top-gate voltages, following the VS model, as described above. Furthermore, Fig. 3d–f, respectively, demonstrate the tunability of A , μ_V , and σ_V of the Gaussian synapse via top-gate voltages V_N and V_p . More details on the design of Gaussian synapses can be found in the supplementary information section.

Brainwave classification. Next, we show simulation results suggesting that PNNs based on Gaussian synapses can be used for the classification of various neural oscillations, also known as the brainwaves that are fundamental to human awareness, cognition, emotions, and actions. These rhythmic and repetitive oscillations that originate from synchronous and complex firing of neural ensembles are observed throughout the central nervous system and are essential in controlling the neuro-physiological health of any individual. As shown in Fig. 4a, the brainwaves are divided into five frequency bands based on the present understanding and interpretation of their functions. Lower frequency (0.5–3.5 Hz) and higher amplitude delta waves (δ) are generated during deepest meditation and dreamless sleep, suspending all external awareness and facilitating healing and regeneration. Disruptions in δ -wave activity can lead to neurological disorders such as dementia, schizophrenia, parasomnia, epilepsy, and Parkinson's disease. Low frequency (4–8 Hz) and high amplitude theta waves (θ) originate during sleep or deep meditation with senses withdrawn from the external world and focused within. Normal firing of θ -waves enables learning, memory, intuition, and introspection, while excessive activity can lead to attention-deficit/hyperactivity disorder (ADHD). Mid-frequency (8–14 Hz) and mid-amplitude alpha waves (α) represent the resting state of the brain and facilitate mind/body coordination, mental peace, alertness, and learning. High frequency (16–32 Hz) and low amplitude beta waves (β) dominate the wakeful state of consciousness, direct our concentration towards cognitive tasks such as problem solving and decision-making, and at the same time consume a tremendous amount of energy. In clinical context, β -waves can be used as biomarkers as they indicate the release of gamma aminobutyric acid, the principal inhibitory neurotransmitter in the mammalian nervous system. Finally, gamma waves (γ) are the fastest (32–64 Hz) and quietest brain waves. These waves were dismissed as neural

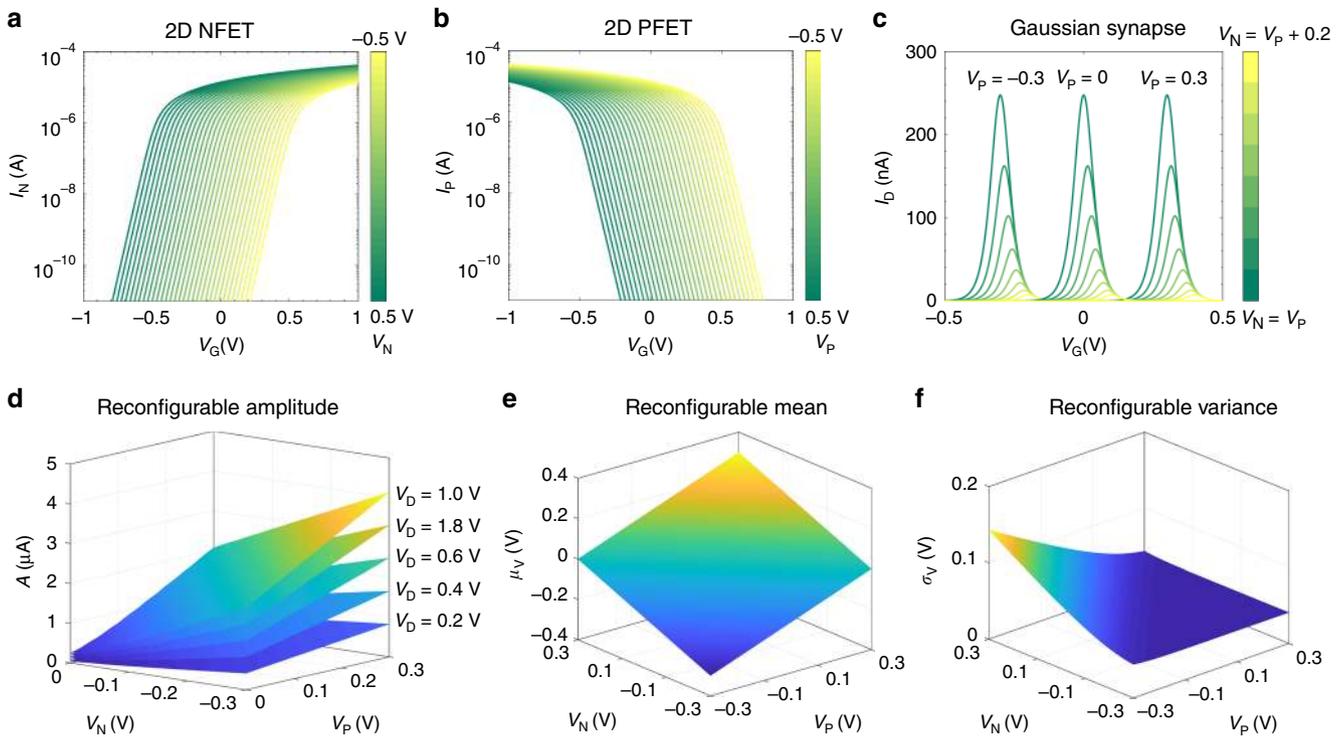


Fig. 3 Scaled Gaussian Synapses. Simulated back-gated transfer characteristics of **(a)** n-type and **(b)** p-type 2D FET for corresponding different top-gate voltages V_N and V_P , respectively, using the Virtual Source (VS) model. In the VS model, both the subthreshold and the above threshold behavior are captured through a single semi-empirical and phenomenological relationship that describes the transition in channel charge density from weak to strong inversion. **(c)** Transfer characteristics, **(d)** amplitude, **(e)** mean, and **(f)** standard deviation of the Gaussian synapse obtained via heterogeneous integration of the n-type and the p-type 2D FETs for different combinations of V_N and V_P . The following parameters were used for the simulation. $L_N = L_P = 1 \mu\text{m}$; $W_N = W_P = 2 \mu\text{m}$; $\mu_N = \mu_P = 20 \text{ cm}^2(\text{Vs})^{-1}$; $C_{BG} = 7 \times 10^{-3} \text{ F m}^{-2}$; $M = 1$; $\alpha = 1$; $V_D = 1 \text{ V}$

noise until recently, when researchers discovered the connection to greater consciousness and spiritual activity culminating in the state of universal love and altruism⁵⁵. The above discussion clearly shows the immense importance of brainwaves in regulating our daily experience. Instabilities in brain rhythm can be catastrophic, leading to insomnia, narcolepsy, panic attacks, obsessive-compulsive disorder, agitated depression, hyper-vigilance, and impulsive behaviors. Early diagnosis of abnormal brainwave activity through neural networks can help prevent chronic neuro diseases and mental and emotional disorders.

Figure 4b, c, and 4d show the frequency pattern of the normalized power spectral density (PSD) for each type of brainwaves, extracted from the Fast Fourier Transform (FFT) of the time domain Electroencephalography (EEG) data (sequential montage) with increasing sampling times that correspond to sample sizes of $N = 512, 2560,$ and $6400,$ respectively. Clearly, as the training set becomes more and more exhaustive, the discrete frequency responses corresponding to each type of brainwave evolve into continuous spectrums that show complex patterns. Furthermore, the system is highly nonlinear with functional dependence of the PSDs on frequency being rather complicated for each type of brainwave. As such, classification of brainwave patterns using conventional ANNs, can be challenging^{56–58}. In addition, ANNs require optimum training algorithms and extensive feature extraction and preprocessing of the training sample in order to achieve reasonable accuracy. In contrast, as demonstrated in Fig. 4d, the PNN adopts single pass learning by defining the class PDF for each of the brainwave pattern in the frequency domain using Gaussian mixture model (GMM)^{59,60}. GMM is represented as the weighted sum of a finite number of scaled (different variance) and shifted (different mean) normal

distributions as described by Eq. 2.

$$p(x) = \sum_{i=1}^K \psi_i N[x/\mu_i, \sigma_i^2]; N[x/\mu_i, \sigma_i^2] = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left[-\frac{(x-\mu_i)^2}{2\sigma_i^2}\right]; \sum_{i=1}^K \psi_i = 1 \tag{7}$$

A GMM with K components is parameterized by two types of values, the component weights (ψ_i) and the component means (μ_i) and variances (σ_i^2) with the constraint that $\sum_{i=1}^K \psi_i = 1$, so that the total probability distribution normalizes to unity. For each type of brainwave pattern, the GMM parameters for the K component were estimated from the training data corresponding to $N = 25,600,$ using the non-linear least square method. Figure 4e shows root mean square errors (RMSEs) calculated as a function of $K,$ i.e., the number of Gaussian curves used in the corresponding GMMs. Clearly, a very limited number of Gaussian functions are necessary to capture the non-linear decision boundary for each of the brainwaves. This enormously reduces the energy and size constraint for the PNNs based on Gaussian synapses.

Finally, Fig. 5a shows simulation results evaluating the PNN architecture for the detection of new brainwave patterns. The PNN consists of 4 layers: input, pattern, summation, and output. The amplitude of the new FFT data is relayed from the input layer to the pattern layer as the drain voltage (V_D) of the Gaussian synapses, whereas, the frequency range (0–64 Hz) is mapped to the back-gate voltage (V_G) range. The summation layer integrates the current over the full swing of V_G from the individual pattern blocks and communicates with the winner-takes-it-all (WTA)

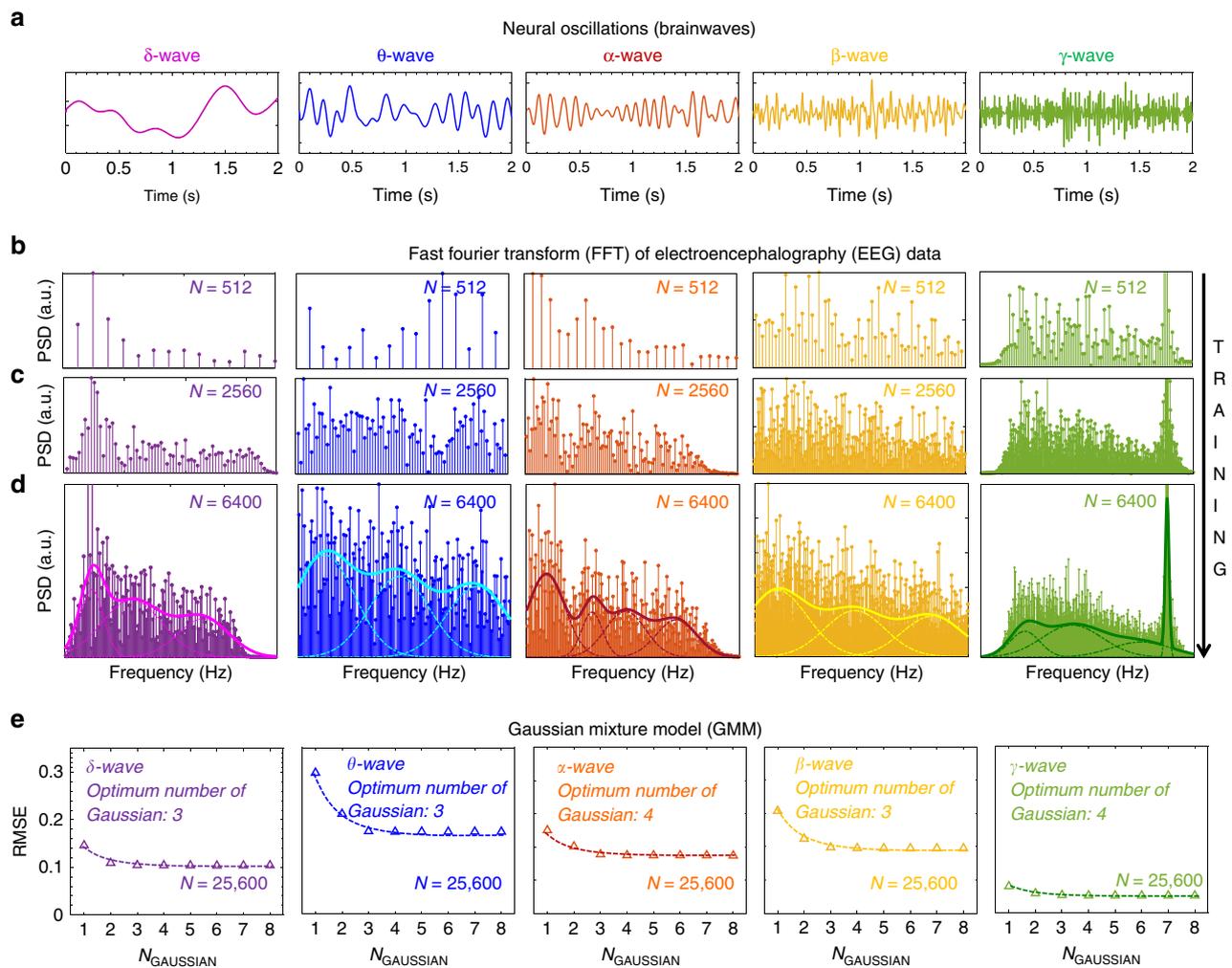


Fig. 4 Recognition of Brainwaves using Gaussian Mixture Model (GMM). **a** Brainwaves are divided into five frequency bands. 0.5–3.5 Hz: Delta waves (δ), 4–8 Hz: Theta waves (θ), 8–14 Hz: Alpha waves (α), 16–32 Hz: Beta waves (β) and 32–64 Hz: Gamma waves (γ). Normalized power spectral density (PSD) as a function of frequency for each type of brainwaves, extracted using the Fast Fourier Transform (FFT) of the time domain Electroencephalography (EEG) data (sequential montage) with increasing sampling times that correspond to sample sizes of **(b)** $N = 512$, **(c)** $N = 2560$, and **(d)** $N = 6400$. As the training set becomes more and more exhaustive, the discrete frequency responses evolve into continuous spectrums. The highly nonlinear functional dependence of the PSDs on frequency for each type of brainwave can be classified using GMM, which is represented as the weighted sum of a finite number of scaled (different variance) and shifted (different mean) normal distributions. **e** Root mean square error (RMSE) as a function of K , i.e., the number of Gaussian used in the GMMs for each type of brainwave. Clearly, a very limited number of Gaussian synapses are necessary to capture the non-linear decision boundaries

circuit that allows the output layer to recognize the brainwave patterns. We implemented our PNN architecture on 10 whole-night polysomnographic recordings, obtained from 10 healthy subjects in a sleep laboratory using a digital 32-channel polygraph (details can be found in the method section). The percentage of different brainwave components as recognized by the PNN are shown as a color map in Fig. 5b. As expected, the PNN recognizes the dominant presence of delta and theta waves in the sleep samples. Furthermore, Fig. 5c shows the total power consumption (details can be found in the supplementary information section) by the PNN as a function of the supply voltage (V_{DD}) and sample volume (N). As expected, the power dissipation scales with N and V_{DD} . Interestingly, even for a large sample volume of $N = 2 \times 10^5$, corresponding to 8 h of EEG data, the power consumption by the proposed PNN architecture was found to be as frugal as $3 \mu\text{W}$ for $V_{DD} = 0.1 \text{ V}$, which increases to only $350 \mu\text{W}$ for $V_{DD} = 1.0 \text{ V}$. A direct comparison of power dissipation with digital CMOS will be premature at this time, especially since the peripheral circuits required for the proposed PNN architecture

will add power dissipation overhead. Nevertheless, these preliminary results show that the PNN architectures based on Gaussian synapses can offer extreme energy efficiency.

Discussion

In conclusion, we have demonstrated reconfigurable Gaussian synapses based on the heterostructure of atomically thin 2D layered semiconductors as a new class of analog and probabilistic computational primitives that can reinstate both energy and size scaling aspects of computation. Furthermore, we elucidated how Gaussian synapses enable direct hardware realization of PNNs, which offer simple and effective solutions to a wide range of pattern classification problems and thereby resurrect complexity scaling. Finally, we show simulation results suggesting that PNN architecture based on Gaussian synapses is capable of recognizing complex neural oscillations or brainwave patterns from large volumes of EEG data with extreme energy efficiency. We believe that these findings will foster the much-needed interest in

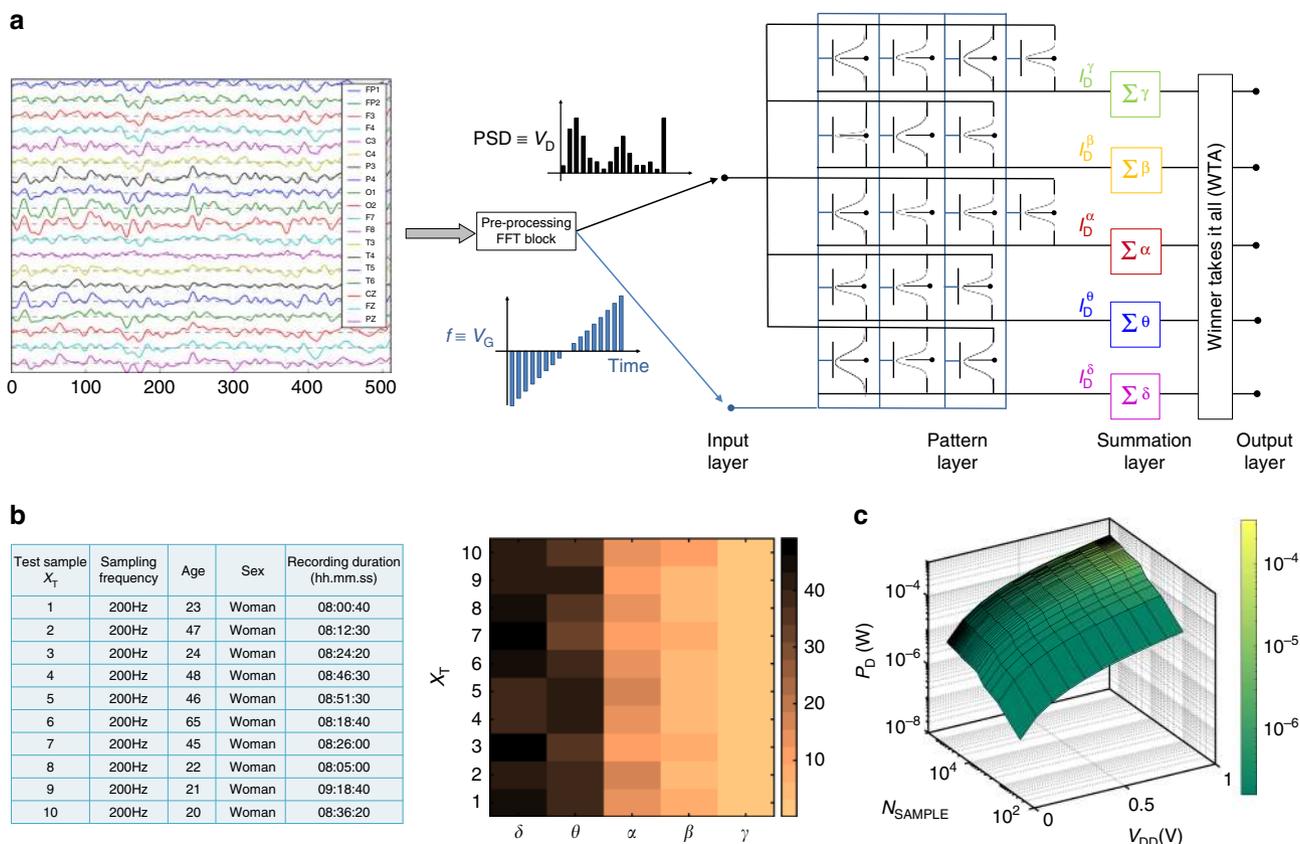


Fig. 5 PNN Architecture for Brainwave Recognition. **a** The PNN consists of 4 layers: input, pattern, summation, and output. The amplitude of the FFT data is relayed from the input layer to the pattern layer as drain voltage (V_D) of the Gaussian synapses, whereas, the frequency range (0–64 Hz) is mapped to the back-gate voltage (V_G) range. The summation layer integrates the current over the full swing of V_G from the individual pattern blocks and communicates with the winner-takes-it-all (WTA) circuit that allows the output layer to recognize the brainwave patterns. **b** Implementation of PNN Architecture: 10 whole-night polysomnographic recordings and the corresponding outcome of the PNN architecture shown using a color map. The PNN recognizes the dominant presence of delta and theta waves in all the sleep samples. **c** Power Consumption by PNN Architecture: The total power consumption by the PNN as a function of the supply voltage (V_{DD}) and sample volume (N_{SAMPLE}). As expected, the power dissipation scales with N_{SAMPLE} and V_{DD}

hardware implementation of PNNs and ultimately aid high performance and low power computing infrastructure.

Methods

Device fabrication and measurements. MoS₂ and BP flakes were micro-mechanically exfoliated on 285 nm thermally grown SiO₂ substrates with highly doped Si as the back-gate electrode. The thicknesses of the MoS₂ and BP flakes were in the range of 3–20 nm. MoS₂ is a 2D layered material with the lattice parameters $a = 3.15 \text{ \AA}$, $b = 3.15 \text{ \AA}$, $c = 12.3 \text{ \AA}$, $\alpha = 90^\circ$, $\beta = 90^\circ$, and $\gamma = 120^\circ$. The layered nature due to van der Waals (vdW) bonding results in a higher value for c . This enables mechanical exfoliation of the material to obtain ultra-thin layers of MoS₂. BP exhibits a puckered honeycomb lattice structure. It has phosphorous atoms existing on two parallel planes. The lattice parameters are given by $a = 3.31 \text{ \AA}$, $b = 10.47 \text{ \AA}$, $c = 4.37 \text{ \AA}$, $\alpha = 90^\circ$, $\beta = 90^\circ$, and $\gamma = 90^\circ$. The Source/Drain contacts were defined using electron-beam lithography (Vistec EBPG5200). Ni (40 nm) followed by Au (30 nm) was deposited using electron-beam (e-beam) evaporation for the contacts. Both devices were fabricated with a channel length of 1 μm . The width of the MoS₂ and BP devices were 0.78 μm and 2 μm , respectively. The top-gated devices were fabricated with hydrogen silsesquioxane (HSQ) as the top-gate dielectric. The top-gate dielectric was deposited by spin coating 6% HSQ in methyl isobutyl ketone (MIBK) (Dow Corning XR-1541-006) at 4000 rpm for 45 s and baked at 80 $^\circ\text{C}$ for 4 min. The HSQ was patterned using an e-beam dose of 2000 $\mu\text{C cm}^{-2}$ and developed at room temperature using 25% tetramethylammonium hydroxide (TMAH) for 30 s following a 90 s rinse in deionized water (DI). Next, the HSQ was cured in air at 180 $^\circ\text{C}$ and then 250 $^\circ\text{C}$ for 2 min and 3 min, respectively. The thickness of the HSQ layer, used as the top-gate dielectric was 120 nm. Top-gate electrodes with Ni (40 nm) followed by Au (30 nm) were patterned with the same procedure as the source and drain contacts. Given the instability of BP, we took special care to ensure minimal exposure time to the air

while fabricating BP devices by storing the material in vacuum chambers between different fabrication steps. In addition, all the three lithography steps involved in the device fabrication were done within a period of 3 days. The electrical characterizations were obtained at room temperature in high vacuum ($\approx 10^{-6}$ Torr) a Lake Shore CRX-VF probe station and using a Keysight B1500A parameter analyzer.

EEG data. The data used for our study were obtained from the DREAMS project, which were acquired in a sleep laboratory of a Belgium hospital using a digital 32-channel polygraph (BrainnetTM System of MEDATEC, Brussels, Belgium). They consist of whole-night polysomnographic recordings, coming from healthy subjects. At least two EOG channels (P8–A1, P18–A1), three EEG channels (CZ–A1 or C3–A1, FP1–A1, and O1–A1), and one submental EMG channel were recorded. The standard European Data Format (EDF) was used for storing. The sampling frequency was 200 Hz. These recordings have been specifically selected for their clarity (i.e., that they contain few artifacts) and come from persons, free of any medication, volunteers in other research projects, conducted in the sleep lab.

Data availability

The data that support the plots within this paper and other findings of this study are available from the corresponding authors upon reasonable request.

Code availability

The codes used for data analysis are available from the corresponding authors upon request.

Received: 6 February 2019 Accepted: 13 August 2019

Published online: 13 September 2019

References

- Frank, D. J. et al. Device scaling limits of Si MOSFETs and their application dependencies. *Proc. IEEE* **89**, 259–288 (2001).
- Thompson, S. E. & Parthasarathy, S. Moore's law: the future of Si microelectronics. *Mater. Today* **9**, 20–25 (2006).
- Sze, S. M. & Sze, S. *Modern Semiconductor Device Physics*. (Wiley, New York, 1998).
- Ma, X. & Arce, G. R. *Computational Lithography*. Vol. 77 (John Wiley & Sons, 2011).
- Myers, G. J. *Advances in Computer Architecture*. (John Wiley & Sons, Inc., 1982).
- Dennard, R. H., Gaensslen, F. H., Rideout, V. L., Bassous, E. & LeBlanc, A. R. Design of ion-implanted MOSFET's with very small physical dimensions. *IEEE J. Solid-St Circ.* **9**, 256–268 (1974).
- Meindl, J. D., Chen, Q. & Davis, J. A. Limits on silicon nanoelectronics for terascale integration. *Science* **293**, 2044–2049 (2001).
- Yu, B. et al. in *Electron Devices Meeting. IEDM'02 International*. 251–254 (IEEE, 2002).
- Esmailzadeh, H., Blem, E., Amant, R. S., Sankaralingam, K. & Burger, D. in *Computer Architecture (ISCA) 38th Annual International Symposium on ISCA*. 365–376 (IEEE, 2011).
- Merolla, P. A. et al. A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science* **345**, 668–673 (2014).
- Benini, L. & Micheli, G. System-level power optimization: techniques and tools. *ACM Trans. Des. Autom. Electron. Syst.* **5**, 115–192 (2000).
- Pickett, M. D., Medeiros-Ribeiro, G. & Williams, R. S. A scalable neuristor built with Mott memristors. *Nat. Mater.* **12**, 114 (2013).
- Prezioso, M. et al. Training and operation of an integrated neuromorphic network based on metal-oxide memristors. *Nature* **521**, 61 (2015).
- Wang, Z. et al. Memristors with diffusive dynamics as synaptic emulators for neuromorphic computing. *Nat. Mater.* **16**, 101–108 (2017).
- Baldi, P. & Meir, R. Computing with arrays of coupled oscillators: an application to preattentive texture discrimination. *Neural Comput.* **2**, 458–471 (1990).
- Arnold, A. J. et al. Mimicking neurotransmitter release in chemical synapses via hysteresis engineering in MoS₂ Transistors. *ACS Nano* **11**, 3110–3118 (2017).
- Tian, H. et al. Anisotropic black phosphorus synaptic device for neuromorphic applications. *Adv. Mater.* **28**, 4991–4997 (2016).
- Jiang, J. et al. 2D MoS₂ neuromorphic devices for brain-like computational systems. *Small* **13**, 1700933 (2017).
- Dreiseitl, S. & Ohno-Machado, L. Logistic regression and artificial neural network classification models: a methodology review. *J. Biomed. Inform.* **35**, 352–359 (2002).
- Hagan, M. T. & Menhaj, M. B. Training feedforward networks with the Marquardt algorithm. *IEEE Trans. Neural Netw.* **5**, 989–993 (1994).
- Bergman, K. et al. Exascale computing study: technology challenges in achieving exascale systems. *Defense Advanced Research Projects Agency Information Processing Techniques Office (DARPA IPTO), Technical Report 15* (2008).
- Reed, D. A. & Dongarra, J. Exascale computing and big data. *Commun. ACM* **58**, 56–68 (2015).
- Specht, D. F. Probabilistic neural networks. *Neural Netw.* **3**, 109–118 (1990).
- Besag, J., Green, P., Higdon, D. & Mengersen, K. Bayesian computation and stochastic systems. *Stat. Sci.* **10**, 3–41 (1995).
- Schlöer, H. & Hartmann, U. Mapping neural network derived from the Parzen window estimator. *Neural Netw.* **5**, 903–909 (1992).
- Delbruck, T. Bump circuits for computing similarity and dissimilarity of analog voltages. *IEEE IJCNN*, A475–A479 (1991).
- Choi, J., Sheu, B. J. & Chang, J. C. F. A Gaussian synapse circuit for analog vlsi neural networks. *1994 IEEE Int. Symp. Circuits Syst.* **6**, F483–F486 (1994).
- Madrenas, J., Verleysen, M., Thissen, P. & Voz, J. L. A CMOS analog circuit for Gaussian functions. *IEEE Trans. Circuits-I* **43**, 70–74 (1996).
- Lin, S. Y., Huang, R. J. & Chiueh, T. D. A tunable Gaussian/square function computation circuit for analog neural networks. *IEEE Trans. Circuits Syst. I-Analog Digit. Signal Process.* **45**, 441–446 (1998).
- Srivastava, R., Singh, U. & Gupta, M. Analog circuits for Gaussian function with improved performance. In *Analog circuits for Gaussian function with improved performance* 934–938 (IEEE, 2011).
- Kang, K. & Shibata, T. An on-chip-trainable Gaussian-Kernel analog support vector machine. *IEEE Trans. Circuits-I* **57**, 1513–1524 (2010).
- Wu, N. et al. A real-time and energy-efficient implementation of difference-of-Gaussian with flexible thin-film transistors. *IEEE Comp. Soc. Ann.* 455–460 <https://doi.org/10.1109/ISVLSI.2016.87> (2016).
- Das, S., Chen, H. Y., Penumatcha, A. V. & Appenzeller, J. High performance multilayer MoS₂ transistors with scandium contacts. *Nano Lett.* **13**, 100–105 (2013).
- Das, S., Demarteau, M. & Roelofs, A. Ambipolar phosphorene field effect transistor. *ACS nano* **8**, 11730–11738 (2014).
- Das, S., Robinson, J. A., Dubey, M., Terrones, H. & Terrones, M. Beyond graphene: progress in novel two-dimensional materials and van der Waals Solids. *Annu Rev. Mater. Res.* **45**, 1–27 (2015).
- Schulman, D. S., Arnold, A. J. & Das, S. Contact engineering for 2D materials and devices. *Chem. Soc. Rev.* 3037–3058 (2018).
- Andrzejewski, D. et al. Improved luminescence properties of MoS₂ monolayers grown via MOCVD: role of pre-treatment and growth parameters. *Nanotechnology* **29**, 295704 (2018).
- Nguyen, T. K. et al. High photoresponse in conformally grown monolayer MoS₂ on a rugged substrate. *ACS Appl. Mater. Interfaces* **10**, 40824–40830 (2018).
- Smithe, K. K. H., Suryavanshi, S. V., Munoz Rojo, M., Tedjarati, A. D. & Pop, E. Low variability in synthetic monolayer MoS₂ devices. *ACS Nano* **11**, 8456–8463 (2017).
- Zhang, J. et al. Scalable growth of high-quality polycrystalline MoS(2) monolayers on SiO(2) with tunable grain sizes. *ACS Nano* **8**, 6024–6030 (2014).
- Smith, J. B., Hagaman, D. & Ji, H. F. Growth of 2D black phosphorus film from chemical vapor deposition. *Nanotechnology* **27**, 215602 (2016).
- Das, S. & Appenzeller, J. WSe₂ field effect transistors with enhanced ambipolar characteristics. *Appl. Phys. Lett.* **103**, 103501 (2013).
- Zhang, X. et al. Defect-controlled nucleation and orientation of WSe₂ on hBN: a route to single-crystal epitaxial monolayers. *ACS Nano*. <https://doi.org/10.1021/acsnano.8b09230> (2019).
- Chuang, S. et al. MoS₂ p-type transistors and diodes enabled by high work function MoO_x contacts. *Nano Lett.* **14**, 1337–1342 (2014).
- Perello, D. J., Chae, S. H., Song, S. & Lee, Y. H. High-performance n-type black phosphorus transistors with type control via thickness and contact-metal engineering. *Nat. Commun.* **6**, 7809 (2015).
- Wang, A., Calhoun, B. H. & Chandrakasan, A. P. in *Sub-threshold Design for Ultra Low-Power Systems (Series on Integrated Circuits and Systems)*. (Springer-Verlag, 2006).
- Hanson, S., Seok, M., Sylvester, D. & Blaauw, D. Nanometer device scaling in subthreshold logic and SRAM. *IEEE Trans. Electron Devices* **55**, 175–185 (2008).
- Wang, A. & Chandrakasan, A. A 180-mV subthreshold FFT processor using a minimum energy design methodology. *IEEE J. Solid-St. Circ.* **40**, 310–319 (2005).
- Andreou, A. G. et al. Current-mode subthreshold MOS circuits for analog VLSI neural systems. *IEEE Trans. Neural Netw.* **2**, 205–213 (1991).
- Wallace, R. M. & Wilk, G. D. High-κ dielectric materials for microelectronics. *Crit. Rev. Solid State Mater. Sci.* **28**, 231–285 (2003).
- Frank, D. J., Taur, Y. & Wong, H.-S. P. Generalized scale length for two-dimensional effects in MOSFETs. *Electron Device Lett. IEEE* **19**, 385–387 (1998).
- Nasr, J. R. & Das, S. Seamless fabrication and threshold engineering in monolayer MoS₂ dual-gated transistors via hydrogen silsesquioxane. *Adv. Electron. Mater.* **5**, 1800888 (2019).
- Nasr, J. R., Schulman, D. S., Sebastian, A., Horn, M. W. & Das, S. Mobility deception in nanoscale transistors: an untold contact story. *Adv. Mater.* **31**, 1806020 (2019).
- Khakifirooz, A., Nayfeh, O. M. & Antoniadis, D. A simple semiempirical short-channel MOSFET current-voltage model continuous across all regions of operation and employing only physical parameters. *IEEE Trans. Electron Devices* **56**, 1674–1680 (2009).
- Rubik, B. Neurofeedback-enhanced gamma brainwaves from the prefrontal cortical region of meditators and non-meditators and associated subjective experiences. *J. Altern. Complement. Med.* **17**, 109–115 (2011).
- Nakayama, K., Kaneda, Y. & Hirano, A. A Brain Computer Interface Based on FFT and multilayer neural network - feature extraction and generalization. *International Symposium on ISPACS'07* 826–829 (IEEE, 2007).
- Nakayama, K. & Inagaki, K. A Brain Computer Interface Based on Neural Network with Efficient Pre-Processing. *International Symposium on ISPACS'06* 673–676 (IEEE, 2006).
- Subasi, A. & Erceleb, E. Classification of EEG signals using neural network and logistic regression. *Comput. Methods Prog. Biomed.* **78**, 87–99 (2005).
- Zivkovic, Z. An Improved Moving Object Detection Algorithm Based on Gaussian Mixture Models. *Proceedings of the 17th International Conference on Pattern Recognition (ICPR)* 28–31 (IEEE, 2004).
- Reynolds, D. Gaussian mixture models. *Ency. Biometrics* 827–832 (2015).

Acknowledgements

The authors would like to acknowledge the contribution of Joseph R Nasr, Harikrishnan Ravichandran, Harikrishnan Jayachandrakurup, and Sarbashis Das for their help in device fabrication. This work was partially supported through Grant Number FA9550-17-1-0018 from Air Force Office of Scientific Research (AFOSR) through the Young Investigator Program.

Author contributions

S.D. conceived the idea, designed the experiments and wrote the paper. A.S., S.S.R., and A.P. performed the experiments. All authors analyzed the data, discussed the results, agreed on their implications, and contributed to the preparation of the paper.

Additional information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-019-12035-6>.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Peer Review Information Nature Communications thanks Qing Wan, Su-Ting Han and other, anonymous, reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019