ACS NANO

www.acsnano.org

ARTICLE

Ultralow Power In-Sensor Neuronal Computing with Oscillatory Retinal Neurons for Frequency-Multiplexed, Parallel Machine Vision

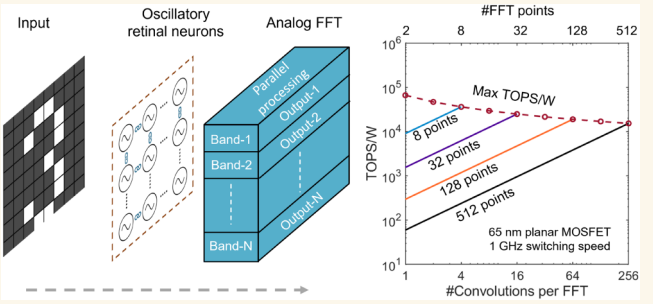
Ragib Ahsan, Hyun Uk Chae, Seyedeh Atiyeh Abbasi Jalal, Zezhi Wu, Jun Tao, Subrata Das, Hefei Liu, Jiang-Bin Wu, Stephen B. Cronin, Han Wang, Constantine Sideris, and Rehan Kapadia*

Cite This: *ACS Nano* 2024, 18, 23785–23796

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: **In-sensor and near-sensor computing architectures enable multiply accumulate operations to be carried out directly at the point of sensing. In-sensor architectures offer dramatic power and speed improvements over traditional von Neumann architectures by eliminating multiple analog-to-digital conversions, data storage, and data movement operations. Current in-sensor processing approaches rely on tunable sensors or additional weighting elements to perform linear functions such as multiply accumulate operations as the sensor acquires data. This work implements in-sensor computing with an oscillatory retinal neuron device that converts incident optical signals into voltage oscillations. A computing scheme is introduced based on the frequency shift of coupled oscillators that enables parallel, frequency multiplexed, nonlinear operations on the inputs. An experimentally implemented 3 × 3 focal plane array of coupled neurons shows that functions approximating edge detection, thresholding, and segmentation occur *in parallel*. An example of inference on handwritten digits from the MNIST database is also experimentally demonstrated with a 3 × 3 array of coupled neurons feeding into a single hidden layer neural network, approximating a liquid-state machine. Finally, the equivalent energy consumption to carry out image processing operations, including peripherals such as the Fourier transform circuits, is projected to be <20 fJ/OP, possibly reaching as low as 15 aJ/OP.**



KEYWORDS: *negative differential resistance, oscillator, oscillatory retinal neurons, parallel computing, ultralow power computing, in-sensor computing*

MAIN

In-sensor computing has emerged as a promising approach to improve computational speed and reduce energy consumption.[1−15] By eliminating the initial data conversion, storage, and transmission, in-sensor architectures offer dramatically higher speed and lower power consumption when compared to traditional von Neumann architectures. Local weighting devices or tunable responsivity sensors enable in-sensor architectures, where the input signal is multiplied by a weight at the point of sensing, resulting in local multiply accumulate (MAC) operations on the inputs which is at the heart of the convolutional neural networks (CNN). Optoelectronic synaptic sensors can realize different weights by tuning their responsivity and therefore enable neuromorphic convolution operations on an input image. Traditional silicon[16] and III−V materials,[17] ZnO,[18] more exotic 2D semiconductors such as WS_2,[19,20] WSe_2,[13,21] MoS_2,[22−24] MoSSe,[25] $MoS_{2-x}O_x$,[22] $MoTe_2$,[26] $PdSe_2$,[27] $ReSe_2$,[28] black phosphorus,[29] $PtSe_2$,[30] graphene,[22,31] Te nanoflake,[32] various organic semiconductors,[33−35] and double perovskites such as $Cs_2AgBiBr_6$[34] have been used as photoactive material in different implementations of these optoelectronic sensors. Traditional metal−insulator transition materials such as VO_2,[36−38] electrochemical migration based memristors such as Ag nanoparticles embedded in TiO_2,[39] lithium ions intercalated in Al_2O_3,[25]

Received: July 6, 2024
Revised: August 2, 2024
Accepted: August 6, 2024
Published: August 14, 2024

https://doi.org/10.1021/acsnano.4c09055
ACS Nano 2024, 18, 23785−23796

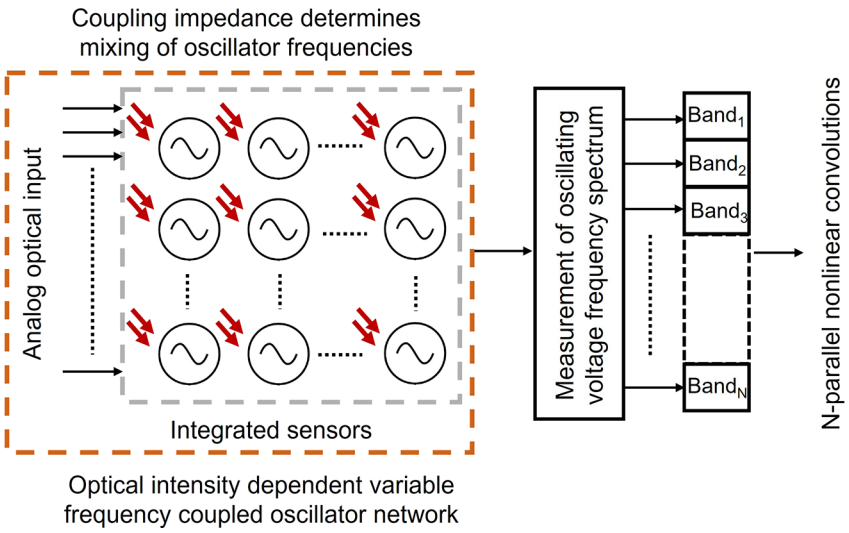


Figure 1. Block diagram of a frequency multiplexed parallel computing ORN network.

ferroelectrics such as PZT[40] and CIPS,[32] floating gate structures with Al_2O_3,[21,25,29] SiO_2,[18,31] HfO_2,[29] h-BN[21,31] etc. have been used to introduce nonvolatile responsivity to these sensors. These design efforts have been geared toward improving control on responsivity of the sensors, realizing signed responsivities to enable both positive and negative weights, and utilizing the photovoltaic regime of operation to minimize the energy consumption of convolution and MAC operations. In addition to optoelectronic vision[4,5,7,9−11,13,41] sensors, a wide variety of modalities, including auditory,[8,42−47] olfactory,[44] and tactile[48−50] sensors, benefit from the improved performance. However, these approaches generally execute a single MAC operation on the input data.[12−14,51−53] Furthermore, parallel operations require scaling the number of weighting devices connected to each sensor which can be costly from an area and power perspective.

In this work, we introduce an in-sensor computing approach where a coupled photosensor array carries out parallel computation on the input image. Each pixel in the array acts as an oscillator, generating an optical power-dependent frequency spectrum. When coupled, neighboring pixels also affect each pixel's frequency spectrum. The power in a frequency band then becomes a nonlinear function of the inputs. Separate frequency bands, therefore, encode separate nonlinear functions of the inputs in parallel. Here, each pixel is an oscillatory retinal neuron (ORN) that directly converts the input optical signal into voltage oscillations. We show through simulation and experiment that coupled ORN networks carry out approximations of both basic and advanced image processing functions, such as edge detection and image segmentation directly in the sensor, encoded by choice of frequency and bandwidth of the output filter. Figure 1 shows a block diagram of the system depicting how frequency multiplexed computing leads to parallel nonlinear convolution operations. It is important to note that this approach is distinct from the traditional oscillatory neural networks (ONN). There are two major approaches to performing image processing computations in traditional ONNs: (1) associative memory approach and (2) degree of match based convolution approach.[54−58] In the associative memory approach, input vector, such as vector of pixel intensities, is encoded in the initial phase of the oscillators in a coupled network that settle to a certain phase configuration. The coupling impedances of the network essentially "remembers" certain phase configurations, enabling recognition of different patterns. A variant of this associative memory approach is the Ising machine where the oscillators are forced to settle either in-phase or antiphase as directed by the coupling network. Computationally hard optimization problems that have an Ising Hamiltonian formulation, such as NP-complete max cut problem, can be mapped on to the impedances of a coupled oscillator network where the final phases of the oscillators encode the solution of the problem by minimizing the Ising Hamiltonian. On the other hand, the degree of match based convolution approach can be used to perform convolution like operations using a network of coupled oscillators. In this approach, the elementwise difference of input vector and the kernel weight vector is encoded as the frequency of individual oscillators. If the input vector is similar (dissimilar) to the kernel weight vector, all the oscillators oscillate with similar (dissimilar) frequency, leading to a synchronized large amplitude (unsynchronized small amplitude) oscillation. Amplitude of this resultant oscillation therefore denotes the degree of match between the input and kernel, analogous to a convolution operation. CMOS ring oscillators, relaxation oscillators made of insulator−metal−transition memristors, such as VO_2 oscillators have been the heart of such ONNs enabling pattern recognition, Ising machines,[59−67] and convolution operations.[57,58,62−64,68−71] Traditional ONNs are susceptible to different nonidealities such as phase distortion and propagation delays in high frequency circuits, and most importantly, extreme sensitivity to frequency variations. Our frequency multiplexed computing approach is distinct to conventional in-sensor computing and ONNs, as outlined by the following key differences: (1) ORNs incorporate oscillatory behavior into the sensor, transforming them into optoelectronic neuronal sensors; (2) while synaptic sensors require external programming for sequential responsivity adjustments to achieve various kernels, neuronal sensors encode responsivities in their oscillating voltage, enabling diverse responsivity configurations without external programming; (3) the ORN network does not rely on phase synchronization of oscillators, allowing them to operate at varying frequencies; (4) the use of frequency domain readout enables parallel computations across different

https://doi.org/10.1021/acsnano.4c09055
ACS Nano 2024, 18, 23785−23796

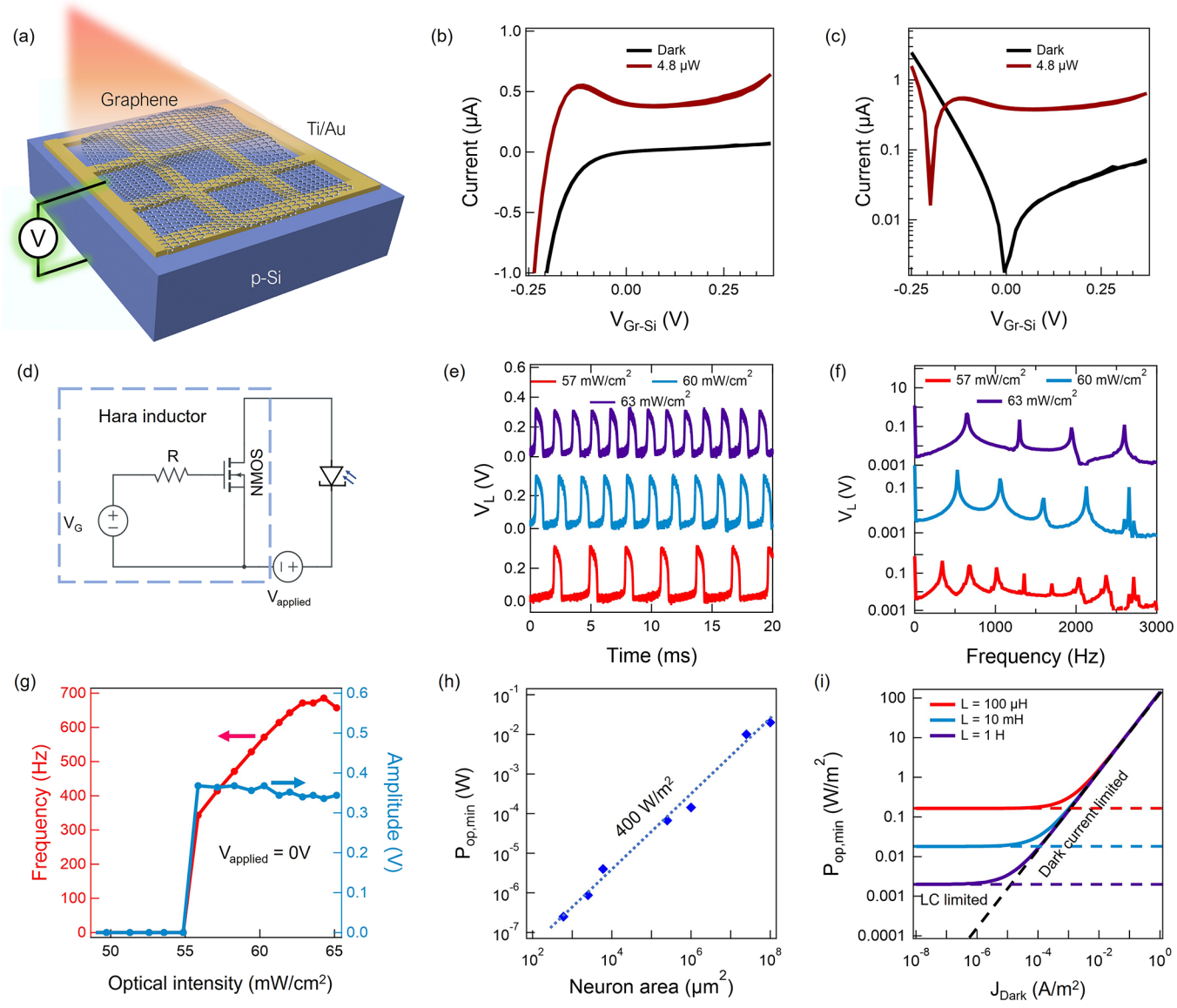


Figure 2. ORN enabled by SGM photodetector. (a) Schematic of the SGM photodetector device. (b) *I*–*V* curves measured at dark conditions and under uniform illumination (445 nm) in linear and (c) log scale. (d) Schematic of a single unit of ORN. (e) *V*–*t* curves measured at different optical intensities and (f) corresponding frequency spectra. (g) spiking frequency and amplitude as a function of optical intensity. (h) Experimental plot of minimum optical power required for oscillation with neuron area. (i) Calculation of dark current limited and LC limited $P_{op,min}$ for oscillation without external electrical power.

frequency bands. Notably, the ORNs do not require external electrical power, and when considering peripheral circuits such as buffers, selector circuits, and analog fast Fourier transform circuits, the *equivalent energy per operation can be smaller than 20 fJ/OP, possibly reaching as low as 15 aJ/OP with a signal-to-noise ratio (SNR) equivalent to that of a digital computation with 8 bit precision.*

RESULTS

The ORNs are composed of two elements, (i) a photodetector that exhibits voltage-controlled negative differential resistance (NDR) under illumination and (ii) an inductive element that can drive an electrical oscillation by taking advantage of the instability of the NDR behavior. A semiconductor–graphene–metal (SGM) photodetector, schematically shown in Figure 2a, exhibits NDR in the detector's power generation regime. The device comprises a p-type silicon substrate, a Ti/Au (5 nm/100 nm) metal grid, and a graphene layer. Linear scale *I*–*V* measurements of a 1 mm × 1 mm device under dark and uniform optical illumination are shown in Figure 2b. In the dark, the device exhibits Schottky-diode behavior, while exhibiting NDR under illumination. Figure 2c shows the log-scale *I*–*V* curves, highlighting that the NDR is only observed under illumination. Section S1 and Figures S1–S9 discuss the device-level behaviors and the measurement setup in detail. Connecting this device with an inductive element under appropriate bias conditions generates optical intensity dependent oscillations, as shown schematically in Figure 2d. An active inductive element, the Hara inductor, comprising a single MOSFET and a resistor, enables the scalability of the ORN. The observed oscillations are analogous to classical Van der Pol oscillators and the Fitzhugh–Nagumo model of neurons.[72−75] A prior work discusses the implementation and design of Hara active inductors in more detail.[76]

https://doi.org/10.1021/acsnano.4c09055
ACS Nano 2024, 18, 23785−23796

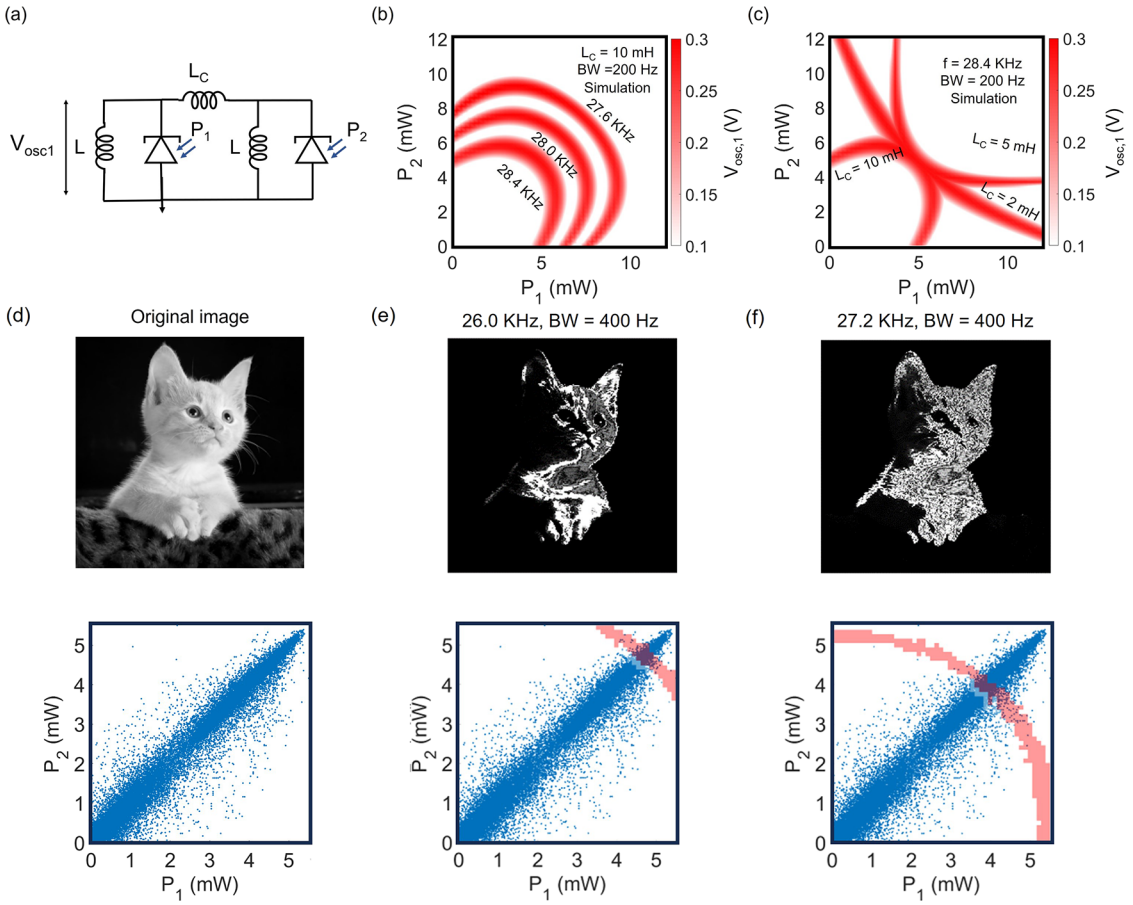


Figure 3. Frequency multiplexed computation with ORN. (a) Circuit schematic for two coupled ORNs. (b) ORN voltage colormap showing nonlinear peak surfaces and their shift at different center frequencies for L_C = 10 mH and BW = 200 Hz. (c) ORN voltage colormap showing different peak surface shapes for different L_C values. (d) Original image and the scatter plot showing all the (P_1,P_2) pairs for this image when input to a 1 × 2 convolutional kernel. (e,f) Experimentally measured image transformations when the two coupled ORNs (L_C = 10 mH) receive the (P_1,P_2) pairs as inputs similar to a convolution operation and the corresponding scatter plots. The overlap between red and blue scatter plots show how different subsets of inputs are thresholded by the network at different center frequencies (BW = 400 Hz). The original image has been used with permission from the original photographer, Tyler A. Swartz.

Other graphene-based photodetectors have exhibited NDR behavior, but all at a forward diode bias.[77−83] Photodetectors fabricated in other materials systems also show NDR behavior and have been used for designing optoelectronic spiking neuron circuits.[84,85] However, this device generates an open-circuit voltage and exhibits NDR at negative and zero applied voltages. This critical distinction allows oscillations at $V_{applied} \leq$ 0 V, which enables operation without external electrical power. Figure 2e shows experimental $V-t$ curves for a photodetector with an active area of 1 cm^2. Figure 2f shows the corresponding frequency spectra, illustrating the change as a function of the optical intensity. Figure 2g shows the oscillation frequency and amplitude as a function of incident optical intensity, where we observe that a minimum optical intensity is required to trigger oscillations in this ORN circuit. These measurements were all performed at $V_{applied}$ = 0 V.

To explore the scaling behavior of ORNs, photodetectors with areas between 600 μm^2 and 1 cm^2 have been fabricated and tested. The minimum optical power required for oscillation *without external electrical power* scales linearly with the device area, as shown in Figure 2h. Two parameters limit the oscillation dynamics of ORNs, the dark current and the capacitance. First, the dark current does not exhibit NDR and adds with the light current. Second, the photon flux should generate sufficient light current so that the valley of the NDR is greater than the dark current. There must also be sufficient photocurrent to charge and discharge the capacitance at time scales of the oscillation frequency. The addition of external power can mitigate this limitation. For a moderately doped (5 × 10^{15} cm^{-3}) p-Si substrate, the depletion capacitance at the graphene-silicon junction is ∼0.1 fF/μm^2. Figure 2i shows the minimum optical intensity for oscillation assuming a device capacitance of 0.1 fF/μm^2 as a function of device dark current density. We can see a crossover between two different regimes: (1) inductance−capacitance (LC) limited regime at smaller dark currents and (2) dark current limited regime at larger dark currents. For our photodetectors, the Schottky nature of the junction results in a larger dark current, limiting the threshold optical intensity to ∼400 W/m^2. At smaller dark current densities, it is possible to decrease this threshold to below 2 mW/m^2.

Next, we present a simple demonstration of how these coupled oscillators carry out computation. We use simulations of ORN circuits connected to bandpass filters to elucidate the behavior of coupled ORNs and how image processing occurs. We considered an ORN comprising a photodetector with an active area of 1 mm^2 connected to an external inductor (L = 10 mH) with $V_{applied}$ = 0 V. We simulated the $V-t$ curves of the

https://doi.org/10.1021/acsnano.4c09055
ACS Nano 2024, 18, 23785−23796

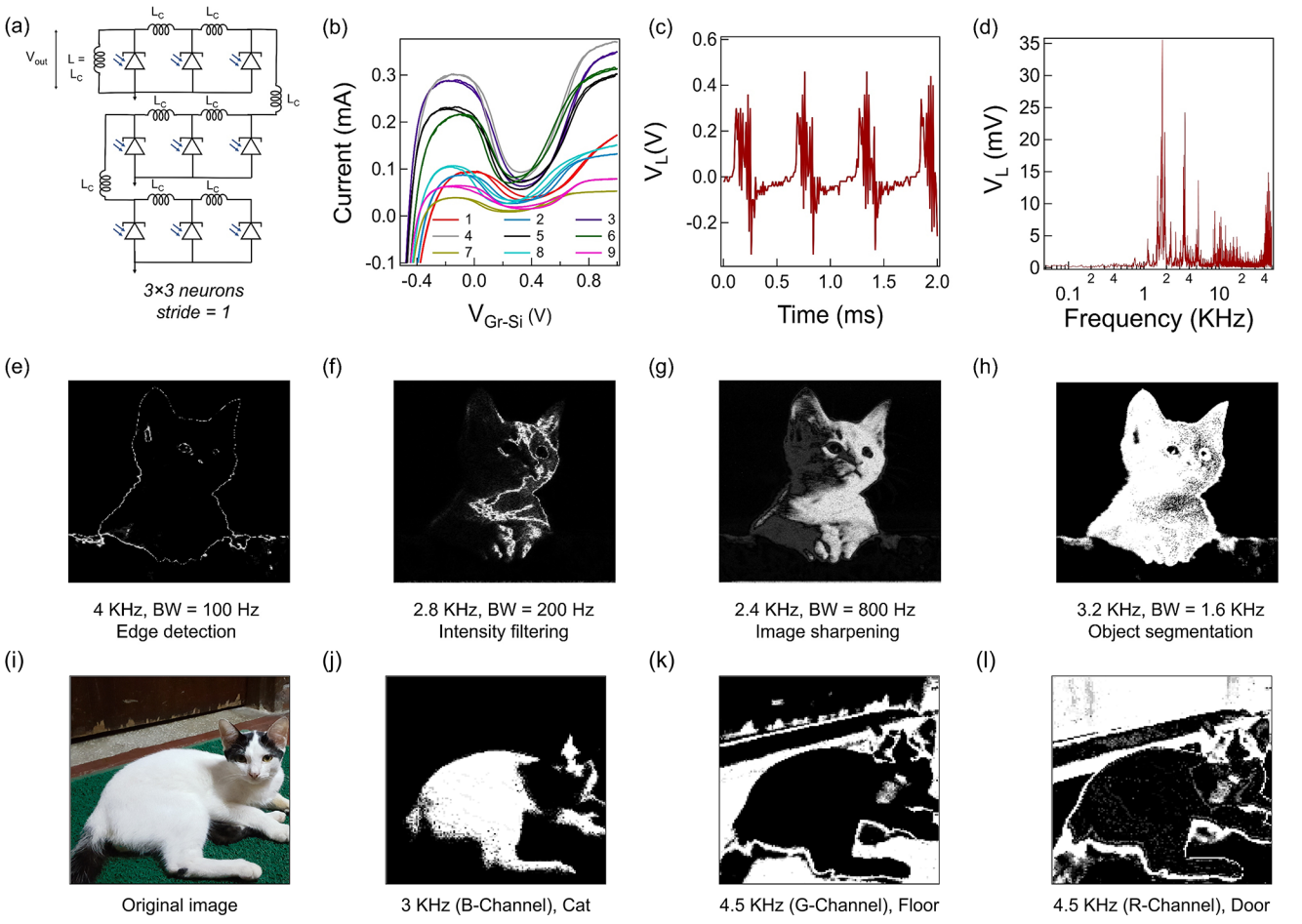


Figure 4. Image processing with coupled ORN network. (a) Circuit schematic for the ORN kernel. (b) $I-V$ curves of all 9 SGM detectors in the network under same optical illumination. (c) Oscillation $V-t$ and (d) FFT curves at the output node when all ORNs are under uniform illumination. (e) Frequency band filtered images showing edge detection, (f) intensity filtering, (g) image sharpening, (h) object segmentation. (i) Original color image and frequency domain images showing (j–l) image segmentation operation.

ORN using the experimental photodetector capacitance and $J-V$ values. The $V-t$ output of the simulation is filtered with varying center frequencies (f) and bandwidths (BW) representing different bandpass filters. Section S2 and Figure S10 show through simulation and analysis that each bandpass filtered output of a single ORN can be analytically approximated with Lorentzians. Figure 3a shows the schematic of two ORNs with inductive coupling, L_C = 10 mH. Figure 3b plots the bandpass filtered V_{osc1} magnitude as a function of P_1 and P_2 (incident optical power on the ORNs) for varying center frequencies f = 28.4, 28, and 27.6 kHz with BW = 200 Hz. The results show that two coupled oscillators define a curved subspace of the input. Figure 3c shows the simulation results for a fixed filter with f = 28.4 kHz and BW = 200 Hz and varying coupling impedance. This results in subspaces of varying shapes. While accurate solutions of the oscillator-coupled nonlinear differential equations require a numerical approach, we can analytically approximate the subspace by reducing the two oscillator problem to a single oscillator problem by introducing a new quantity $P_{12} = \sqrt{P_1^2 + P_2^2 + a(P_1 + P_2) + kP_1P_2 + b}$ which nonlinearly combines P_1 and P_2. The coupled oscillator result then becomes $V_{osc}(P_{12}, f, BW) = \frac{\gamma}{(P_{12} - P_{00})^2 + (\Delta P)^2}$, which can be fit to approximate the result from Figure 3c. Here, P_{00} is a function of the center frequency f and ΔP is a function of the filter bandwidth, BW.

To obtain a visual representation of how an image is processed in this scheme, we have fabricated a 1 × 2 focal plane array and set up the two coupled oscillator circuit (Figure 3a) with passive inductors. Treating this circuit as a 1 × 2 convolutional kernel of stride 1, we have first obtained all the possible input pixel intensity pairs for a grayscale image of a cat (Figure 3d, top panel) with 250 × 240 pixels. There are 60000 such pairs of pixels for this image (with zero padding). Using a digital projector and lens based optical setup, we have then projected two different optical powers on two photodetectors where a pixel intensity of 1 translated to an optical power of 5.5 mW and all pixel intensity values are linearly scaled with optical power. The bottom panel of Figure 3d shows all (P_1,P_2) optical power pairs as a scatter plot where each point corresponds to the pair of optical power inputs to the circuit, i.e., the 1 × 2 kernel. An oscilloscope measures the output $V-t$ signal from a single node of the array, $V_{osc,1}$ in this case. The output spectrum is then processed in software to obtain the FFT and filtered outputs. Some of the representative output $V-t$ signals and their corresponding frequency spectra are shown in Figure S11. The top panels in Figure 3e,f show the filtered output images for f = 26.0 and 27.2 kHz at a BW = 400 Hz. Clearly, the original image has

https://doi.org/10.1021/acsnano.4c09055
ACS Nano 2024, 18, 23785−23796

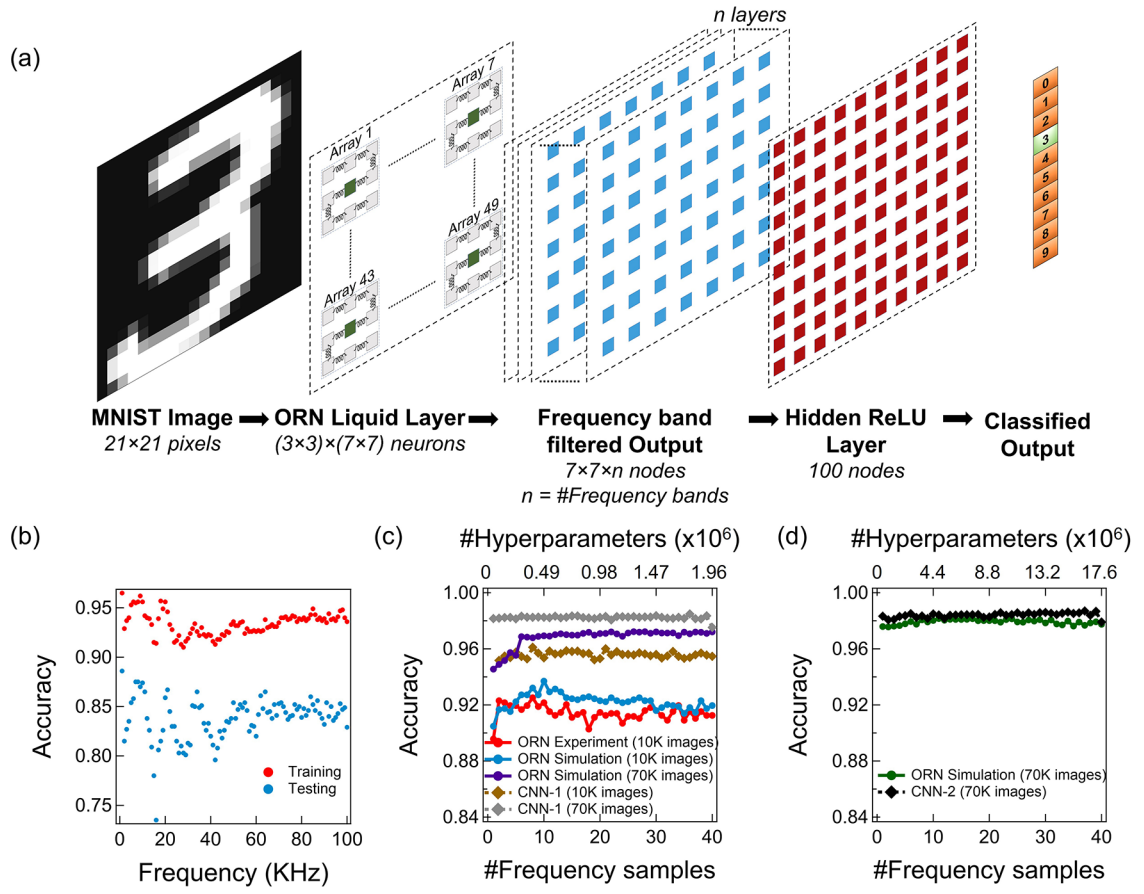


Figure 5. LSM implementation of ORN network for MNIST classification. (a) Image classification pipeline of the LSM structure showing an original input image, structure of the liquid layer, frequency sampled output images, and further processing at the readout layer by hidden ReLU units. (b) Training and testing accuracy of the readout layer for training data sets corresponding to different frequency samples. (c) Classification accuracy of the handwritten digits as a function of number of frequency samples for 7 × 7 pixels/sample and (d) for 21 × 21 pixels/sample. The image of handwritten digit "3" has been used under Creative Commons Attribution-Share Alike 3.0 license from MNIST data set.

been mapped to multiple processed images, indexed by the filter's center frequency. The bottom panels of Figure 3e,f show how the subspaces, defined by the ORN coupling, filter center frequency (f), and bandwidth (BW), overlap with the (P_1,P_2) pairs of the original image. The coupled ORNs map the subset of the pixel pairs that overlap with the defined subspace as high values and all other pixel pairs as low values. These results on a toy problem visually show how nonlinear computations are performed using coupled ORN oscillators.

To demonstrate how coupled ORNs carry out more useful and complex image processing functions, from edge detection to image sharpening, we have then performed similar experiments on the same image with a 3 × 3 ORN circuit with a cascaded connection, as shown in Figure 4a. We use this as a kernel that slides across an image in the same manner as a convolution operation in a convolutional neural network (CNN). A digital projector and external lens form the desired 3 × 3 segment of an image on the ORN focal plane array similar to the 1 × 2 kernel case. Output voltage oscillation is measured at the node V_{out} as shown in Figure 4a. Figure 4b shows the $I-V$ curves of all the SGM photodetectors in the experimental array under the same optical intensity (3 mW/mm^2). Figure 4c shows a representative $V-t$ curve obtained from the 3 × 3 array when all the pixels are illuminated with uniform intensity. Figure 4d shows the frequency spectrum of the $V-t$ curve of Figure 4c.

We then took the digital grayscale image of a cat (Figure 3d) and projected it on the 3 × 3 ORN focal plane array, using the array as a convolution kernel with a stride of one (pixel intensity of 1 refers to 5.5 mW incident optical power). Figure 4e–h show the images obtained at 4 kHz (BW = 100 Hz), 2.8 kHz (BW = 200 Hz), 2.4 kHz (BW = 800 Hz), and 3.2 kHz (BW = 1.6 kHz), respectively. These filtered images demonstrate edge detection, intensity filtering, image sharpening and object segmentation operations. As we increase the bandwidth from Figure 4e to Figure 4h, we observe a larger image region thresholded to bright pixels. In this way, smaller bandwidth filters enable lower-level feature extraction, such as edges, while high bandwidth filters lead to higher-level feature extraction, such as object segmentation. Bandwidth of these filters is directly related to the acquisition time of the oscillating output signal. A longer (shorter) acquisition time allows lower (higher) bandwidth and bins the frequency spectrum into more (fewer) bands available for computing. An image recognition task may require both higher and lower bandwidth filters while an image segmentation task may require only higher bandwidth filters. Therefore, the choice of bandwidth is specific to the application and is an important design parameter. Figure S12 shows the processed images

https://doi.org/10.1021/acsnano.4c09055
ACS Nano 2024, 18, 23785–23796

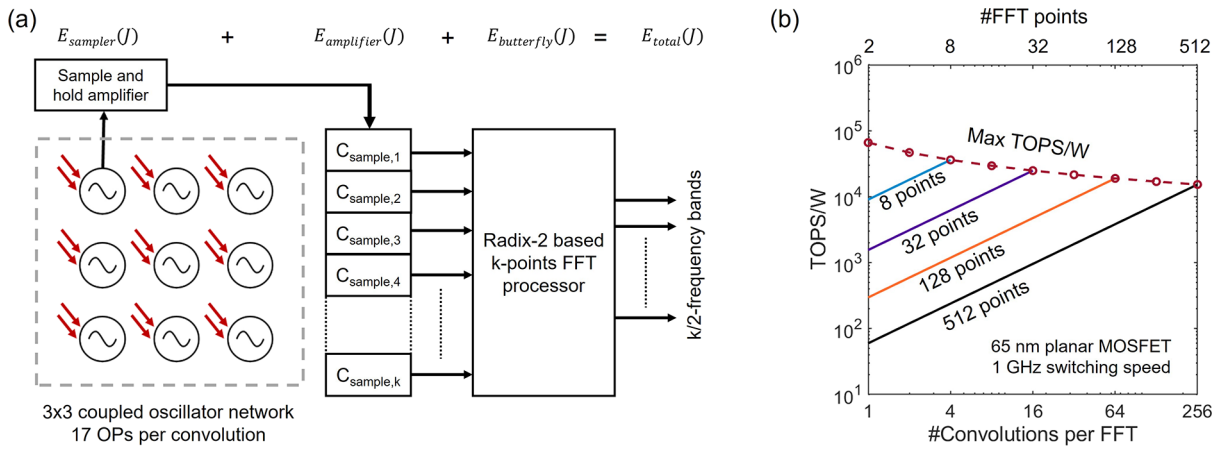


Figure 6. Energy consumption in ORN networks. (a) Block diagram of a 3 × 3 ORN network connected to an analog *k*-point FFT processor with *k*/2 frequency band outputs. (b) Performance of the ORN network as a function of useful convolution operations per FFT for different FFT implementations with 65 nm planar MOSFETs at a switching speed of 1 GHz.

when the image is projected at different optical power ranges. When the incident optical power range is lower, similar image processing can be obtained at higher center frequencies. This result shows that the choice of optical power range is not very critical if appropriate center frequencies are chosen. It is also important to note that the circuit configuration used here to couple the oscillators is not unique. Figure S13 shows image processing results obtained using a 10 × 10 coupled oscillator array kernel where the ORNs are connected to their nearest neighbors with a coupling inductance of 5H. Engineering the circuit configurations allows the implementation of a variety of image processing functions. Section S5 and Figures S14−S16 discuss the function implemented by the 3 × 3 kernel in more detail, specially how the edges are detected by the circuit.

Next, we have investigated whether the same 3 × 3 ORN focal plane array can perform image segmentation from an image with multiple objects. A color image of size 170 × 150 pixels (Figure 4i) that features a cat, a floor and a door was selected. The image is split into three different grayscale images according to the pixel intensities of the color channels (R-channel, G-channel, and B-channel). These three channels are then treated as independent images while being input to the ORN kernel, i.e., no coupling between different color channels were considered here. Therefore, each bandpass filter used had three different output images, one for each color channel. Figure 4j−l shows the images filtered at 3.5 kHz (B-channel), 4.5 kHz (G-channel), and 4.5 kHz (R-channel), respectively. The bandwidth used for each center frequency is 500 Hz. At 3.5 kHz (B-channel), the cat emerges as white and rest of the image is thresholded to black, effectively segmenting the cat. The images filtered at 4.5 kHz (G-channel) and 4.5 kHz (B-channel) segment the floor and the door, respectively. It is important to note that we have only used a single bandpass filter to segment an entire object in this case. Improved segmentation quality is expected when a linear combination of multiple frequencies is used. These results clearly illustrate how the ORN kernel can perform parallel, frequency multiplexed image processing and segmentation tasks.

These results show us two essential properties of this architecture: (1) the absence of any encoding or preprocessing for input, and (2) the ability to perform parallel computation at different frequencies. Since the projection of image and data acquisition are both performed in analog domain, inevitably noise is added to both the input and output of the system but can still obtain excellent results.

Inference is carried out by using a 3 × 3 pixel coupled oscillator network to act as a liquid layer to construct a liquid state machine (LSM). Images from the MNIST database cropped to 21 × 21 pixels were serially projected on the 3 × 3 array with a stride of 3, while output signals were acquired from a single pixel. This data acquisition mode converts 21 × 21 images into 7 × 7 × n data points where n is the number of frequency samples considered. Each frequency sample corresponds to a bandpass filtered output at a given center frequency and a bandwidth of 1 kHz. Ten thousand images from the MNIST database were projected on the array, and the output data was collected and fed into a readout layer consisting of a single hidden layer with 100 nodes followed by a 10-node output layer. The hidden layer used a ReLU activation function, and the output layer used a softmax activation function. We use backpropagation to train only the readout layer while keeping the liquid layer connections untouched. Section S6 summarizes the implementation of the readout layer. Figure 5a shows the LSM schematic. Figure 5b plots the accuracy obtained at the 50th epoch if only a single frequency from each coupled array is fed into the hidden layer.

As expected, the single-frequency results show that the resulting accuracy varies by filter frequency. Feeding multiple frequency samples per pixel to the hidden layer is expected to augment the accuracy of the network. Figure 5c shows how feeding multiple frequencies into the hidden network modifies the testing accuracies obtained at the 200th epoch. We have done this for both experimental and simulated ORN arrays. The experiments were carried out on 10000 images, limited by the speed of our data acquisition and projection setup. We observed a peak accuracy of 92.51% with 7 frequencies sampled per pixel. The data set was split into training and testing subsets with a 6:1 ratio. To evaluate the potential of this result if the full data set of 70000 images were used, a simulated version of the same 3 × 3 ORN focal plane array was also carried out. We see the resulting accuracy for the experimental and simulation cases with 10000 images are very similar. As the simulation uses the experimental device *I*−*V* curves, discrepancies between the simulation and experiment are attributed to the additional noise introduced by our image

https://doi.org/10.1021/acsnano.4c09055
ACS Nano 2024, 18, 23785−23796

Table 1. Comparison Between Different NPUs

NPU application	type	comment	bits	reported TOPS/W	normalized TOPS/W (8 bits)	normalized fJ/OP (8 bits)
analog to information conversion[14]	analog	in-sensor NN	8	43.5	43.5	23
VMM[15]	digital	SRAM	4	351	87.75	11.4
MAC macro[87]	analog	DRAM	4	217	54.25	18.4
VMM[88]	digital	DNN learning processor	8	146.52	146.52	6.8
arithmetic logic[89]	digital	superconducting logic devices	8	120	120	8.3
VMM[90]	analog	Si-CMOS/CAAC-IGZO based memory	6	210	118.13	8.5
VMM[91]	digital	stochastic NN accelerator	8	75	75	13.3
MAC macro[92]	digital	SRAM	8	63	63	15.9
MAC macro[93]	analog	SONOS memory	8	100	100	10
MAC macro[94]	digital	SRAM	1	20943	327.23	3.1
general purpose	digital	NVIDIA A100	8	4.992	4.992	200.3
general purpose	digital	Apple a16 Bionic	8	2.67	2.67	374.5
general purpose	digital	Qualcomm Snapdragon 865	8	4.5	4.5	222.2
nonlinear convolution (3 × 3 kernel) (this work)	analog	ORN	8	50−67000	50−67000	0.015−20

projection and data collection setup. The experimental data acquisition and simulation details are discussed in Section S7. In order to compare these results with a software defined neural network, we have constructed a CNN, named CNN-1 here, that has a convolutional layer (3 × 3 kernel with stride of 3, n channels) and ReLU activation that connects to the same fully connected layer (single hidden layer with 100 nodes). We vary the number of channels of the convolutional layer so that the number of trainable hyperparameters in this network remains the same as the ORN network. When trained with 10000 images, CNN-1 shows a significantly higher accuracy (∼96%) compared to ORN for both measured and simulated results. However, as we increase the number of training images to 60000 (with 10000 testing images), the ORN network shows an improved accuracy of 97.21%, just shy of the ∼98% accuracy achieved by CNN-1 for same number of training images. Critically, if the 3 × 3 array is used as a convolution kernel with a stride of 1, a peak accuracy of 98.16% is achieved for 11 frequency samples per pixel, as shown in Figure 5d. When we compare these results with CNN-2 (∼98.45%), employing a 3 × 3 convolutional kernel with a stride of 1, the difference in accuracy is even smaller. These results show that the parallel processing performed at different frequencies improves the network and that the coupling between pixels in the 3 × 3 array plays the similar role as a convolutional kernel in a CNN, allowing similar classification accuracies for ORN liquid based and fully software defined networks. In addition, the LSM architecture does not require the training of liquid layer interconnections, which significantly reduces the complexity and computational cost of the training.

While an ORN array does not require any external electrical power to drive the oscillations, the system requires peripheral circuitry to read the voltages and perform bandpass filtering operations. A charge domain on-chip FFT processor[86] can perform such operations with a low energy cost. Figure 6a shows a block diagram of a 3 × 3 ORN network showing how the oscillating output signal is filtered by an analog FFT processor into different frequency bands enabling parallel computation. As discussed in Section S8 and summarized in Figure 6b, an ORN array can perform convolution equivalent tasks with a performance as high as 67000 TOPS/W, which translates to an energy cost of 15 aJ/OP with a precision equivalent to 8-bit integer operations in digital systems. However, this energy cost depends on the implementation and speed of the FFT processor, and the number of frequency bands that perform useful convolution-like operations. For an example, a 32-points FFT processor can achieve a maximum performance of 25000 TOPS/W at 1 GHz frequency when power in all the 16 frequency bands is considered useful for the neural network. However, if only 1 frequency band returns useful convolution operation, the performance drops to ∼1550 TOPS/W, highlighting the importance of careful design optimizations that would multiplex many different useful convolution operations in different frequency bands to enable extreme energy efficiencies in the system. These projections clearly show that frequency multiplexed computing using coupled ORN array has the potential to completely replace the energy-expensive convolutional layers in CNN for deep learning applications.

Table 1 shows the performance comparison between different neural processing units (NPU) for deep learning.[14,15,87−94] Different NPUs operate at different bit resolutions and therefore an n-bit performance was scaled by a factor of $(\frac{n}{8})^2$ to get a normalized 8-bit performance. Such a scaling is reasonable[95,96] since number of transistors in digital logic typically scales as $\sim n^2$. Table S1 provides a more detailed list of NPUs and their performance as found in literature. Unfortunately, it is not very common for the works focusing on in-sensor computing to report their performance or energy costs. Most often, there is not sufficient information provided in these reports to calculate the energy costs. To avoid possible errors, Table 1 and Table S1 do not list the performance of in-sensor computing works unless directly reported. While NPUs performing a specific task may achieve a higher performance compared to a general-purpose GPU that performs different tasks, usefulness of these tasks is specific to the aimed applications. Therefore, such a comparison between NPUs serves as a guide to understanding the potential of different computing architectures, rather than a direct comparison of raw computing abilities.

CONCLUSION

In conclusion, we have introduced in-sensor neuronal computing as an alternative to in-sensor synaptic computing. We demonstrated that coupled ORNs enable highly parallel,

https://doi.org/10.1021/acsnano.4c09055
ACS Nano 2024, 18, 23785−23796

frequency multiplexed computation on input images without data conversion, storage, or transmission penalties. Experimental implementations using 3 × 3 array of coupled ORNs show parallel image processing on projected images. These include edge detection, intensity filtering, and object segmentation as examples of image processing tasks carried out at the detector array. We have also demonstrated that inference with these devices performs handwritten digit classification from the MNIST database with similar accuracy as software defined CNNs. While we have focused on image classification and image processing applications, we expect this computational approach to be general. Most importantly, ORN-based computation is extremely energy efficient considering the energy cost of the peripheral circuits, laying the framework for a general and ultralow power approach to oscillator-based computing.

METHODS

Semiconductor Substrate Preparation. Moderately boron doped (N_A = 5 × 10^{15} cm^{-3}) silicon (100) wafer was used as the semiconductor substrate. A 5 nm Ti/60 nm Au mesh is photolithographically defined and deposited by electron beam evaporation. A monolayer of CVD grown graphene is transferred on top of the metal mesh via wet transfer method.[97] A 100 nm aluminum film sputtered at the back side of the substrate acts as the contact to silicon.

Graphene Growth and Transfer. CVD graphene was grown on a Cu foil by using low pressure CVD. Cu foil was etched inside $FeCl_3$ copper etchant for 30 s before the graphene growth. Cu foil was annealed in a quartz tube furnace at 1000 °C for 30 min with 50 standard cubic centimeters per minute (sccm) hydrogen (H_2) flow rate. Graphene was synthesized under 7 sccm of methane (CH_4) and 50 sccm of hydrogen (H_2) for 40 min. For transfer, poly(methyl methacrylate) (PMMA A6495) was spin-coated on top of Cu foil at 2000 rpm for 60 s and baked for 5 min under 170 °C. PMMA spin-coated Cu foil was etched using $FeCl_3$ copper etchant graphene to remove the Cu while the remaining PMMA/graphene floats to the top. The stacked layer was cleaned with deionized (D.I) water and transferred to 10% hydrochloric acid solution to remove the remaining Cu etchants. After cleaning with D.I water once more, PMMA/graphene was transferred on top of the oxide/semiconductor substrate. The substrate was dried in the air overnight followed by 90 °C for 15 min, 150 °C for 30 min, and 90 °C for 15 min to ensure the adhesion between the graphene and the substrate. Finally, the substrate was immersed in acetone for 12 h to remove the PMMA.

Raman Spectroscopy for Graphene. CVD grown monolayer graphene transferred on the substrate was analyzed by Raman spectroscopy. Raman spectra were collected with Renishaw spectrometer with a 532 nm laser focused in a 0.5-μm spot through a Leica microscope with a 100× objective lens.

Wavelength Dependent Measurements. A supercontinuum laser with grating monochromator was used to illuminate the SGM photodetector with lights of different wavelengths between 400 and 1100 nm. Applied voltage was stepped while light and dark current measurements were performed. The difference between these two current measurements, i.e., the photocurrent was then used to measure the responsivity of the device.

ORN Measurements. A 5 × 5 array of SGM photodetectors was fabricated and individual devices were wire-bonded to a PCB. The devices were electrically connected to the inductors (all 10 mH) on a breadboard to form the ORN kernel. A digital projector was used to project the patterns on the device array (a 1 × 2 or 3 × 3 array from the 5 × 5 array) and an oscilloscope was used to record the oscillation waveforms. The whole process was automated using MATLAB environment.

ASSOCIATED CONTENT

Data Availability Statement

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

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsnano.4c09055.

Additional details about the mechanisms of NDR behavior of the photodetector, dynamics and energy consumption of coupled ORN circuits (PDF)

Accession Codes

The codes used to perform the simulations are available from the corresponding author upon reasonable request.

AUTHOR INFORMATION

Corresponding Author

Rehan Kapadia − *Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, California 90089, United States;* orcid.org/0000-0002-7611-0551; Email: rkapadia@usc.edu

Authors

Ragib Ahsan − *Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, California 90089, United States;* orcid.org/0000-0002-3833-7851

Hyun Uk Chae − *Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, California 90089, United States;* orcid.org/0000-0002-7655-2907

Seyedeh Atiyeh Abbasi Jalal − *Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, California 90089, United States*

Zezhi Wu − *Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, California 90089, United States*

Jun Tao − *Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, California 90089, United States*

Subrata Das − *Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, California 90089, United States*

Hefei Liu − *Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, California 90089, United States;* orcid.org/0000-0001-6533-7112

Jiang-Bin Wu − *Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, California 90089, United States;* orcid.org/0000-0002-8751-7082

https://doi.org/10.1021/acsnano.4c09055
ACS Nano 2024, 18, 23785−23796

Stephen B. Cronin − *Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, California 90089, United States;* orcid.org/0000-0001-9153-7687

Han Wang − *Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, California 90089, United States;* orcid.org/0000-0001-5121-3362

Constantine Sideris − *Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, California 90089, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsnano.4c09055

Notes

The authors declare no competing financial interest.

A version of this manuscript was deposited on a preprint server: Ragib Ahsan, Hyun Uk Chae, Seyedeh Atiyeh Abbasi Jalal, Zezhi Wu, Jun Tao, Subrata Das, Hefei Liu, Jiang-Bin Wu, Stephen Cronin, Han Wang, Constantine Sideris, and Rehan Kapadia. "Ultra-low power in-sensor neuronal computing with oscillatory retinal neurons for frequency-multiplexed, parallel machine vision," 2024, 2935296, Research Square, 10.21203/rs.3.rs-2935296/v2 (accessed 30 January, 2024).

ACKNOWLEDGMENTS

This work was supported by Department of Energy Grant No. DE-SC0022248, National Science Foundation Award No. 2004791, Office of Naval Research Grant No. N00014-21-1-2634, Air Force Office of Scientific Research Grant No. FA9550-21-1-0305. R.A. and J.T. acknowledge USC Provost Graduate Fellowships. Z.W. and S.D. acknowledge USC Graduate School Fellowships.

REFERENCES

(1) Amir, M.; Kim, D.; Kung, J.; Lie, D.; Yalamanchili, S.; Mukhopadhyay, S. NeuroSensor: A 3D image sensor with integrated neural accelerator. *2016 IEEE SOI-3D-Subthreshold Microelectronics Technology Unified Conference (S3S)*; IEEE, 2016, 12.

(2) Amir, M. F.; Ko, J. H.; Na, T.; Kim, D.; Mukhopadhyay, S. 3-D stacked image sensor with deep neural network computation. *IEEE Sens. J.* **2018**, *18* (10), 4187−4199.

(3) Choo, K. D.; Xu, L.; Kim, Y.; Seol, J.-H.; Wu, X.; Sylvester, D.; Blaauw, D. Energy-efficient motion-triggered IoT CMOS image sensor with capacitor array-assisted charge-injection SAR ADC. *IEEE J. Solid-State Circuits* **2019**, *54* (11), 2921−2931.

(4) Du, Z.; Fasthuber, R.; Chen, T.; Ienne, P.; Li, L.; Luo, T.; Feng, X.; Chen, Y.; Temam, O. ShiDianNao: Shifting vision processing closer to the sensor, *Proceedings of the 42nd Annual International Symposium on Computer Architecture*; IEEE, 2015, 92104.

(5) Finateu, T.; Niwa, A.; Matolin, D.; Tsuchimoto, K.; Mascheroni, A.; Reynaud, E.; Mostafalu, P.; Brady, F.; Chotard, L.; LeGoff, F. 5.10 A 1280× 720 back-illuminated stacked temporal contrast event-based vision sensor with 4.86 μm pixels, 1.066 GEPS readout, programmable event-rate controller and compressive data-formatting pipeline. *2020 IEEE International Solid-State Circuits Conference-(ISSCC)*; IEEE, 2020, 112114.

(6) Hsu, T.-H.; Chiu, Y.-C.; Wei, W.-C.; Lo, Y.-C.; Lo, C.-C.; Liu, R.-S.; Tang, K.-T.; Chang, M.-F.; Hsieh, C.-C. AI edge devices using computing-in-memory and processing-in-sensor: From system to device. *2019 IEEE International Electron Devices Meeting (IEDM)*; IEEE, 2019, 2225.

(7) Hsu, T.-H.; Chen, Y.-K.; Chiu, M.-Y.; Chen, G.-C.; Liu, R.-S.; Lo, C.-C.; Tang, K.-T.; Chang, M.-F.; Hsieh, C.-C. A 0.8 V multimode vision sensor for motion and saliency detection with ping-pong PWM pixel. *IEEE J. Solid-State Circuits* **2021**, *56* (8), 2516−2524.

(8) Jiménez-Fernández, A.; Cerezuela-Escudero, E.; Miró-Amarante, L.; Dominguez-Moralse, M. J.; de Asís Gómez-Rodríguez, F.; Linares-Barranco, A.; Jiménez-Moreno, G. A binaural neuromorphic auditory sensor for FPGA: A spike signal processing approach. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *28* (4), 804−818.

(9) Lichtsteiner, P.; Posch, C.; Delbruck, T. A 128$\times $128 120 dB 15$\mu $ s latency asynchronous temporal contrast vision sensor. *IEEE J. Solid-State Circuits* **2008**, *43* (2), 566−576.

(10) LiKamwa, R.; Hou, Y.; Gao, J.; Polansky, M.; Zhong, L. Redeye: Analog convnet image sensor architecture for continuous mobile vision. *ACM SIGARCH Comput. Archit. News* **2016**, *44* (3), 255−266.

(11) Wang, C.-Y.; Liang, S.-J.; Wang, S.; Wang, P.; Li, Z. A.; Wang, Z.; Gao, A.; Pan, C.; Liu, C.; Liu, J.; et al. Gate-tunable van der Waals heterostructure for reconfigurable neural network vision sensor. *Sci. Adv.* **2020**, *6* (26), No. eaba6173.

(12) Zhou, F.; Chai, Y. Near-sensor and in-sensor computing. *Nat. Electron.* **2020**, *3* (11), 664−671.

(13) Mennel, L.; Symonowicz, J.; Wachter, S.; Polyushkin, D. K.; Molina-Mendoza, A. J.; Mueller, T. Ultrafast machine vision with 2D material neural network image sensors. *Nature* **2020**, *579* (7797), 62−66.

(14) Sadasivuni, S.; Bhanushali, S. P.; Banerjee, I.; Sanyal, A. In-sensor neural network for high energy efficiency analog-to-information conversion. *Sci. Rep.* **2022**, *12* (1), 18253.

(15) Dong, Q.; Sinangil, M. E.; Erbagci, B.; Sun, D.; Khwa, W.-S.; Liao, H.-J.; Wang, Y.; Chang, J. 15.3 A 351TOPS/W and 372.4 GOPS compute-in-memory SRAM macro in 7nm FinFET CMOS for machine-learning applications. *2020 IEEE International Solid-State Circuits Conference-(ISSCC)*; IEEE, 2020, 242244.

(16) Jang, H.; Hinton, H.; Jung, W.-B.; Lee, M.-H.; Kim, C.; Park, M.; Lee, S.-K.; Park, S.; Ham, D. In-sensor optoelectronic computing using electrostatically doped silicon. *Nat. Electron.* **2022**, *5* (8), 519−525.

(17) Choi, C.; Kim, H.; Kang, J.-H.; Song, M.-K.; Yeon, H.; Chang, C. S.; Suh, J. M.; Shin, J.; Lu, K.; Park, B.-I.; et al. Reconfigurable heterogeneous integration using stackable chips with embedded artificial intelligence. *Nat. Electron.* **2022**, *5* (6), 386−393.

(18) Dang, B.; Liu, K.; Wu, X.; Yang, Z.; Xu, L.; Yang, Y.; Huang, R. One-Phototransistor−One-Memristor Array with High-Linearity Light-Tunable Weight for Optic Neuromorphic Computing. *Adv. Mater.* **2022**, *35*, 2204844.

(19) Sun, Y.; Xu, S.; Xu, Z.; Tian, J.; Bai, M.; Qi, Z.; Niu, Y.; Aung, H. H.; Xiong, X.; Han, J.; et al. Mesoscopic sliding ferroelectricity enabled photovoltaic random access memory for material-level artificial vision system. *Nat. Commun.* **2022**, *13* (1), 5391.

(20) Li, T.; Miao, J.; Fu, X.; Song, B.; Cai, B.; Ge, X.; Zhou, X.; Zhou, P.; Wang, X.; Jariwala, D. Reconfigurable, non-volatile neuromorphic photovoltaics. *Nat. Nanotechnol.* **2023**, *18*, 1303−1310.

(21) Wang, S.; Pan, X.; Lyu, L.; Wang, C.-Y.; Wang, P.; Pan, C.; Yang, Y.; Wang, C.; Shi, J.; Cheng, B. Nonvolatile van der Waals heterostructure phototransistor for encrypted optoelectronic logic circuit. *ACS Nano* **2022**, *16* (3), 4528−4535.

(22) Fu, X.; Li, T.; Cai, B.; Miao, J.; Panin, G. N.; Ma, X.; Wang, J.; Jiang, X.; Li, Q.; Dong, Y.; et al. Graphene/MoS2− xOx/graphene photomemristor with tunable non-volatile responsivities for neuromorphic vision processing. *Light: Sci. Appl.* **2023**, *12* (1), 39.

(23) Kwak, D.; Polyushkin, D. K.; Mueller, T. In-sensor computing using a MoS2 photodetector with programmable spectral responsivity. *Nat. Commun.* **2023**, *14* (1), 4264.

(24) Dodda, A.; Jayachandran, D.; Pannone, A.; Trainor, N.; Stepanoff, S. P.; Steves, M. A.; Radhakrishnan, S. S.; Bachu, S.; Ordonez, C. W.; Shallenberger, J. R.; et al. Active pixel sensor matrix based on monolayer MoS2 phototransistor array. *Nat. Mater.* **2022**, *21* (12), 1379−1387.

(25) Meng, J.; Wang, T.; Zhu, H.; Ji, L.; Bao, W.; Zhou, P.; Chen, L.; Sun, Q.-Q.; Zhang, D. W. Integrated in-sensor computing

optoelectronic device for environment-adaptable artificial retina perception application. *Nano Lett.* **2022**, *22* (1), 81−89.
(26) Wang, F.; Hu, F.; Dai, M.; Zhu, S.; Sun, F.; Duan, R.; Wang, C.; Han, J.; Deng, W.; Chen, W. A two-dimensional mid-infrared optoelectronic retina enabling simultaneous perception and encoding. *Nat. Commun.* **2023**, *14* (1), 1938.
(27) Pi, L.; Wang, P.; Liang, S.-J.; Luo, P.; Wang, H.; Li, D.; Li, Z.; Chen, P.; Zhou, X.; Miao, F. Broadband convolutional processing using band-alignment-tunable heterostructures. *Nat. Electron.* **2022**, *5* (4), 248−254.
(28) Seo, S.; Lee, J. J.; Lee, R. G.; Kim, T. H.; Park, S.; Jung, S.; Lee, H. K.; Andreev, M.; Lee, K. B.; Jung, K. S.; et al. An Optogenetics-Inspired Flexible van der Waals Optoelectronic Synapse and its Application to a Convolutional Neural Network. *Adv. Mater.* **2021**, *33* (40), 2102980.
(29) Lee, S.; Peng, R.; Wu, C.; Li, M. Programmable black phosphorus image sensor for broadband optoelectronic edge computing. *Nat. Commun.* **2022**, *13* (1), 1485.
(30) Islam, M. M.; Krishnaprasad, A.; Dev, D.; Martinez-Martinez, R.; Okonkwo, V.; Wu, B.; Han, S. S.; Bae, T.-S.; Chung, H.-S.; Touma, J.; Jung, Y. Multiwavelength optoelectronic synapse with 2D materials for mixed-color pattern recognition. *ACS Nano* **2022**, *16* (7), 10188−10198.
(31) Yang, Q.; Luo, Z. D.; Zhang, D.; Zhang, M.; Gan, X.; Seidel, J.; Liu, Y.; Hao, Y.; Han, G. Controlled optoelectronic response in van der waals heterostructures for In-sensor computing. *Adv. Funct. Mater.* **2022**, *32* (45), 202207290.
(32) Zha, J.; Shi, S.; Chaturvedi, A.; Huang, H.; Yang, P.; Yao, Y.; Li, S.; Xia, Y.; Zhang, Z.; Wang, W.; et al. Electronic/Optoelectronic Memory Device Enabled by Tellurium-based 2D van der Waals Heterostructure for in-Sensor Reservoir Computing at the Optical Communication Band. *Adv. Mater.* **2023**, *35*, 2211598.
(33) Wu, X.; Wang, S.; Huang, W.; Dong, Y.; Wang, Z.; Huang, W. Wearable in-sensor reservoir computing using optoelectronic polymers with through-space charge-transport characteristics for multi-task learning. *Nat. Commun.* **2023**, *14* (1), 468.
(34) Lao, J.; Yan, M.; Tian, B.; Jiang, C.; Luo, C.; Xie, Z.; Zhu, Q.; Bao, Z.; Zhong, N.; Tang, X. Ultralow-Power Machine Vision with Self-Powered Sensor Reservoir. *Adv. Sci.* **2022**, *9* (15), 2106092.
(35) Seung, H.; Choi, C.; Kim, D. C.; Kim, J. S.; Kim, J. H.; Kim, J.; Park, S. I.; Lim, J. A.; Yang, J.; Choi, M. K.; et al. Integration of synaptic phototransistors and quantum dot light-emitting diodes for visualization and recognition of UV patterns. *Sci. Adv.* **2022**, *8* (41), No. eabq3101.
(36) Li, G.; Xie, D.; Zhong, H.; Zhang, Z.; Fu, X.; Zhou, Q.; Li, Q.; Ni, H.; Wang, J.; Guo, E.-J.; et al. Photo-induced non-volatile VO2 phase transition for neuromorphic ultraviolet sensors. *Nat. Commun.* **2022**, *13* (1), 1729.
(37) Yuan, R.; Duan, Q.; Tiw, P. J.; Li, G.; Xiao, Z.; Jing, Z.; Yang, K.; Liu, C.; Ge, C.; Huang, R.; et al. A calibratable sensory neuron based on epitaxial VO2 for spike-based neuromorphic multisensory system. *Nat. Commun.* **2022**, *13* (1), 3973.
(38) Kumar, M.; Lim, S.; Kim, J.; Seo, H. Picoampere Dark Current and Electro-Opto-Coupled Sub-to-Super-linear Response from Mott-Transition Enabled Infrared Photodetector for Near-Sensor Vision Processing. *Adv. Mater.* **2023**, *35*, 2210907.
(39) Shan, X.; Zhao, C.; Wang, X.; Wang, Z.; Fu, S.; Lin, Y.; Zeng, T.; Zhao, X.; Xu, H.; Zhang, X. Plasmonic Optoelectronic Memristor Enabling Fully Light-Modulated Synaptic Plasticity for Neuromorphic Vision. *Adv. Sci.* **2022**, *9* (6), 2104632.
(40) Cui, B.; Fan, Z.; Li, W.; Chen, Y.; Dong, S.; Tan, Z.; Cheng, S.; Tian, B.; Tao, R.; Tian, G.; et al. Ferroelectric photosensor network: An advanced hardware solution to real-time machine vision. *Nat. Commun* **2022**, *13* (1), 1707.
(41) Zhou, F.; Zhou, Z.; Chen, J.; Choy, T. H.; Wang, J.; Zhang, N.; Lin, Z.; Yu, S.; Kang, J.; Wong, H.-S. P.; et al. Optoelectronic resistive random access memory for neuromorphic vision sensors. *Nat. Nanotechnol.* **2019**, *14* (8), 776−782.
(42) Lyon, R. F.; Mead, C. An analog electronic cochlea. *IEEE Trans. Acoust., Speech, Signal Process.* **1988**, *36* (7), 1119−1134.
(43) Hasler, P.; Smith, P. D.; Graham, D.; Ellis, R.; Anderson, D. V. Analog floating-gate, on-chip auditory sensing system interfaces. *IEEE Sens. J.* **2005**, *5* (5), 1027−1034.
(44) Hsieh, H.-Y.; Tang, K.-T. VLSI implementation of a bio-inspired olfactory spiking neural network. *IEEE Trans. Neural Netw. Learn. Syst.* **2012**, *23* (7), 1065−1073.
(45) Wen, B.; Boahen, K. A 360-channel speech preprocessor that emulates the cochlear amplifier. *2006 IEEE International Solid State Circuits Conference-Digest of Technical Papers*; IEEE, 2006, 22682277.
(46) Ellis, R.; Yoo, H.; Graham, D. W.; Hasler, P.; Anderson, D. V. A continuous-time speech enhancement front-end for microphone inputs. *2002 IEEE International Symposium on Circuits and Systems. Proceedings (Catal. No. 02CH37353)*; IEEE, 2002.
(47) Schrauwen, B.; D'Haene, M.; Verstraeten, D.; Van Campenhout, J. Compact hardware for real-time speech recognition using a liquid state machine. In *2007 international joint conference on neural networks*, IEEE, 2007, pp 10971102.
(48) Tan, H.; Tao, Q.; Pande, I.; Majumdar, S.; Liu, F.; Zhou, Y.; Persson, P. O. Å.; Rosen, J.; van Dijken, S. Tactile sensory coding and learning with bio-inspired optoelectronic spiking afferent nerves. *Nat. Commun.* **2020**, *11* (1), 1369.
(49) Kim, Y.; Chortos, A.; Xu, W.; Liu, Y.; Oh, J. Y.; Son, D.; Kang, J.; Foudeh, A. M.; Zhu, C.; Lee, Y.; et al. A bioinspired flexible organic artificial afferent nerve. *Science* **2018**, *360* (6392), 998−1003.
(50) Zhang, X.; Zhuo, Y.; Luo, Q.; Wu, Z.; Midya, R.; Wang, Z.; Song, W.; Wang, R.; Upadhyay, N. K.; Fang, Y. An artificial spiking afferent nerve based on Mott memristors for neurorobotics. *Nat. Commun.* **2020**, *11* (1), 51.
(51) Zhang, X.; Basu, A. A 915−1220 TOPS/W Hybrid In-Memory Computing based Image Restoration and Region Proposal Integrated Circuit for Neuromorphic Vision Sensors in 65nm CMOS. *2022 IEEE Custom Integrated Circuits Conference (CICC)*; IEEE, 2022, 12.
(52) Bose, S. K.; Basu, A. A 389TOPS/W, 1262fps at 1Meps Region Proposal Integrated Circuit for Neuromorphic Vision Sensors in 65nm CMOS. *2021 IEEE Asian Solid-State Circuits Conference (A-SSCC)*; IEEE, 2021, 13.
(53) Chai, Y. *In-sensor computing for machine vision*; Nature Publishing Group UK: London, 2020.
(54) Koo, M.; Pufall, M.; Shim, Y.; Kos, A. B.; Csaba, G.; Porod, W.; Rippard, W.; Roy, K. Distance computation based on coupled spin-torque oscillators: Application to image processing. *Phys. Rev. Appl.* **2020**, *14* (3), 034001.
(55) Csaba, G.; Porod, W. Coupled oscillators for computing: A review and perspective. *Appl. Phys. Rev.* **2020**, *7* (1), 011302.
(56) Csaba, G.; Raychowdhury, A.; Datta, S.; Porod, W. Computing with coupled oscillators: Theory, devices, and applications. *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*; IEEE, 2018, 15.
(57) Raychowdhury, A.; Parihar, A.; Smith, G. H.; Narayanan, V.; Csaba, G.; Jerry, M.; Porod, W.; Datta, S. Computing with networks of oscillatory dynamical systems. *Proc. IEEE* **2019**, *107* (1), 73−89.
(58) Nikonov, D. E.; Csaba, G.; Porod, W.; Shibata, T.; Voils, D.; Hammerstrom, D.; Young, I. A.; Bourianoff, G. I. Coupled-oscillator associative memory array operation for pattern recognition. *IEEE J. Explor. Solid-State Comput. Devices Circuits* **2015**, *1*, 85−93.
(59) Chou, J.; Bramhavar, S.; Ghosh, S.; Herzog, W. Analog coupled oscillator based weighted Ising machine. *Sci. Rep.* **2019**, *9* (1), 14786.
(60) Dutta, S.; Khanna, A.; Assoa, A.; Paik, H.; Schlom, D. G.; Toroczkai, Z.; Raychowdhury, A.; Datta, S. An Ising Hamiltonian solver based on coupled stochastic phase-transition nano-oscillators. *Nat. Electron.* **2021**, *4* (7), 502−512.
(61) Wang, T.; Roychowdhury, J. OIM: Oscillator-based Ising machines for solving combinatorial optimization problems. In *Unconventional Computation and Natural Computation: 18th International Conference, UCNC 2019*, Springer, 2019, pp 232256.
(62) Nikonov, D. E.; Kurahashi, P.; Ayers, J. S.; Li, H.; Kamgaing, T.; Dogiamis, G. C.; Lee, H.-J.; Fan, Y.; Young, I. Convolution

https://doi.org/10.1021/acsnano.4c09055
ACS Nano 2024, 18, 23785−23796

inference via synchronization of a coupled CMOS oscillator array. *IEEE J. Explor. Solid-State Comput. Devices Circuits* **2020**, *6* (2), 170−176.
(63) Delacour, C.; Carapezzi, S.; Abernot, M.; Todri-Sanial, A. Energy-Performance Assessment of Oscillatory Neural Networks Based on VO2 Devices for Future Edge AI Computing. *IEEE Trans. Neural Netw. Learn. Syst.* **2024**, *35* (7), 10045−10058.
(64) Núñez, J.; Avedillo, M. J.; Jiménez, M.; Quintana, J. M.; Todri-Sanial, A.; Corti, E.; Karg, S.; Linares-Barranco, B. Oscillatory Neural Networks Using VO2 Based Phase Encoded Logic. *Front. Neurosci.* **2021**, *15*, 655823.
(65) Corti, E.; Cornejo Jimenez, J. A.; Niang, K. M.; Robertson, J.; Moselund, K. E.; Gotsmann, B.; Ionescu, A. M.; Karg, S. Coupled VO2 oscillators circuit as analog first layer filter in convolutional neural networks. *Front. Neurosci.* **2021**, *15*, 628254.
(66) Ji, J.; Ren, X.; Gomez, J.; Bashar, M. K.; Shukla, N.; Datta, S.; Zorlutuna, P. Large-Scale Cardiac Muscle Cell-Based Coupled Oscillator Network for Vertex Coloring Problem. *Adv. Intell. Syst.* **2023**, *5* (5), 2200356.
(67) Vaidya, J.; Surya Kanthi, R.; Shukla, N. Creating electronic oscillator-based Ising machines without external injection locking. *Sci. Rep.* **2022**, *12* (1), 981.
(68) Zahedinejad, M.; Awad, A. A.; Muralidhar, S.; Khymyn, R.; Fulara, H.; Mazraati, H.; Dvornik, M.; Åkerman, J. Two-dimensional mutually synchronized spin Hall nano-oscillator arrays for neuromorphic computing. *Nat. Nanotechnol.* **2020**, *15* (1), 47−52.
(69) Nikonov, D. E.; Young, I. A.; Bourianoff, G. I. *Convolutional networks for image processing by coupled oscillator arrays*. arXiv, 2014.
(70) Jackson, T. C.; Sharma, A. A.; Bain, J. A.; Weldon, J. A.; Pileggi, L. Oscillatory neural networks based on TMO nano-oscillators and multi-level RRAM cells. *IEEE J. Emerg. Sel. Top. Circuits Syst.* **2015**, *5* (2), 230−241.
(71) Romera, M.; Talatchian, P.; Tsunegi, S.; Abreu Araujo, F.; Cros, V.; Bortolotti, P.; Trastoy, J.; Yakushiji, K.; Fukushima, A.; Kubota, H.; et al. Vowel recognition with four coupled spin-torque nano-oscillators. *Nature* **2018**, *563* (7730), 230−234.
(72) Nagumo, J.; Arimoto, S.; Yoshizawa, S. An active pulse transmission line simulating nerve axon. *Proc. IRE* **1962**, *50* (10), 2061−2070.
(73) Smith-Saville, R. Variable frequency tunnel diode relaxation oscillator. *Nucl. Instrum. Methods* **1967**, *55*, 120−124.
(74) Keener, J. P. Analog circuitry for the van der Pol and FitzHugh-Nagumo equations. *IEEE Trans. Syst. Man. Cybern.* **1983**, *5*, 1010−1014.
(75) Izhikevich, E. M.; FitzHugh, R. Fitzhugh-nagumo model. *Scholarpedia* **2006**, *1* (9), 1349.
(76) Ahsan, R.; Wu, Z.; Jalal, S. A. A.; Kapadia, R. Ultralow Power Electronic Analog of a Biological Fitzhugh−Nagumo Neuron. *ACS Omega* **2024**, *9* (16), 18062−18071.
(77) Ho, C.-L.; Wu, M.-C.; Ho, W.-J.; Liaw, J.-W. Light-induced negative differential resistance in planar InP/InGaAs/InP double-heterojunction p-i-n photodiode. *Appl. Phys. Lett.* **1999**, *74* (26), 4008−4010.
(78) Liu, W.; Guo, H.; Li, W.; Wan, X.; Bodepudi, S. C.; Shehzad, K.; Xu, Y. Light-induced negative differential resistance in gate-controlled graphene-silicon photodiode. *Appl. Phys. Lett.* **2018**, *112* (20), 201109.
(79) Qin, S.; Wang, F.; Liu, Y.; Wan, Q.; Wang, X.; Xu, Y.; Shi, Y.; Wang, X.; Zhang, R. A light-stimulated synaptic device based on graphene hybrid phototransistor. *2D Mater.* **2017**, *4* (3), 035022.
(80) Wang, X.; Wang, Y.; Feng, M.; Wang, K.; Bai, P.; Tian, Y. Light-induced negative differential resistance effect in a resistive switching memory device. *Curr. Appl. Phys.* **2020**, *20* (3), 371−378.
(81) Antonova, I. V.; Shojaei, S.; Sattari-Esfahlan, S.; Kurkina, I. I. Negative differential resistance in partially fluorinated graphene films. *Appl. Phys. Lett.* **2017**, *111* (4), 043108.
(82) Zhang, Q.; Chen, S.; Zhang, S.; Shang, W.; Liu, L.; Wang, M.; Yu, H.; Deng, L.; Qi, G.; Wang, L. Negative differential resistance and hysteresis in graphene-based organic light-emitting devices. *J. Mater. Chem. C* **2018**, *6* (8), 1926−1932.
(83) Lee, K. W.; Jang, C. W.; Shin, D. H.; Kim, J. M.; Kang, S. S.; Lee, D. H.; Kim, S.; Choi, S.-H.; Hwang, E. Light-induced negative differential resistance in graphene/Si-quantum-dot tunneling diodes. *Sci. Rep.* **2016**, *6* (1), 30669.
(84) Nath, S. K.; Das, S. K.; Nandi, S. K.; Xi, C.; Marquez, C. V.; Rúa, A.; Uenuma, M.; Wang, Z.; Zhang, S.; Zhu, R. J.; et al. Optically Tunable Electrical Oscillations in Oxide-Based Memristors for Neuromorphic Computing. *Adv. Mater.* **2024**, *36*, 2400904.
(85) Al-Taai, Q. R. A.; Hejda, M.; Zhang, W.; Romeira, B.; Figueiredo, J. M.; Wasige, E.; Hurtado, A. Optically-triggered deterministic spiking regimes in nanostructure resonant tunnelling diode-photodetectors. *Neuromorphic Comput. Eng.* **2023**, *3* (3), 034012.
(86) Sadhu, B.; Sturm, M.; Sadler, B. M.; Harjani, R. Analysis and design of a 5 GS/s analog charge-domain FFT for an SDR front-end in 65 nm CMOS. *IEEE J. Solid-State Circuits* **2013**, *48* (5), 1199−1211.
(87) Chen, Z.; Chen, X.; Gu, J. 15.3 a 65nm 3T dynamic analog RAM-based computing-in-memory macro and CNN accelerator with retention enhancement, adaptive analog sparsity and 44TOPS/W system energy efficiency. *2021 IEEE International Solid-State Circuits Conference (ISSCC)*; IEEE, 2021, 240242.
(88) Kim, S.; Lee, J.; Kang, S.; Lee, J.; Yoo, H.-J. A 146.52 TOPS/W deep-neural-network learning processor with stochastic coarse-fine pruning and adaptive input/output/weight skipping. *2020 IEEE Symposium on VLSI Circuits*; IEEE, 2020, 12.
(89) Nagaoka, I.; Tanaka, M.; Sano, K.; Yamashita, T.; Fujimaki, A.; Inoue, K. Demonstration of an energy-efficient, gate-level-pipelined 100 tops/w arithmetic logic unit based on low-voltage rapid single-flux-quantum logic. *2019 IEEE International Superconductive Electronics Conference (ISEC)*; IEEE, 2019, 13.
(90) Chen, M.-C.; Ohshita, S.; Amano, S.; Kurokawa, Y.; Watanabe, S.; Imoto, Y.; Ando, Y.; Hsieh, W.-H.; Chang, C.-H.; Wu, C.-C. A> 64 Multiple States and> 210 TOPS/W High Efficient Computing by Monolithic Si/CAAC-IGZO+ Super-Lattice ZrO 2/Al 2 O 3/ZrO 2 for Ultra-Low Power Edge AI Application. *2022 International Electron Devices Meeting (IEDM)*; IEEE, 2022, 18.2.1−18.2.4.
(91) Romaszkan, W.; Li, T.; Garg, R.; Yang, J.; Pamarti, S.; Gupta, P. A 4.4−75-TOPS/W 14-nm Programmable, Performance-and Precision-Tunable All-Digital Stochastic Computing Neural Network Inference Accelerator. *IEEE Solid-State Circuits Lett.* **2022**, *5*, 206−209.
(92) Fujiwara, H.; Mori, H.; Zhao, W.-C.; Chuang, M.-C.; Naous, R.; Chuang, C.-K.; Hashizume, T.; Sun, D.; Lee, C.-F.; Akarvardar, K. A 5-nm 254-TOPS/W 221-TOPS/mm 2 fully-digital computing-in-memory macro supporting wide-range dynamic-voltage-frequency scaling and simultaneous MAC and write operations*2022 IEEE International Solid-State Circuits Conference (ISSCC)*IEEE, 2022, 13.
(93) Agrawal, V.; Prabhakar, V.; Ramkumar, K.; Hinh, L.; Saha, S.; Samanta, S.; Kapre, R. In-memory computing array using 40nm multibit SONOS achieving 100 TOPS/W energy efficiency for deep neural network edge inference accelerators. *2020 IEEE International Memory Workshop (IMW)*; IEEE, 2020, 14.
(94) Lin, C.-S.; Tsai, F.-C.; Su, J.-W.; Li, S.-H.; Chang, T.-S.; Sheu, S.-S.; Lo, W.-C.; Chang, S.-C.; Wu, C.-I.; Hou, T.-H. A 48 TOPS and 20943 TOPS/W 512kb Computation-in-SRAM Macro for Highly Reconfigurable Ternary CNN Acceleration. *2021 IEEE Asian Solid-State Circuits Conference (A-SSCC)* **2021**, 1−3.
(95) Yang, Q.; Li, H. BitSystolic: A 26.7 TOPS/W 2b∼ 8b NPU with configurable data flows for edge devices. *IEEE Trans. Circuits Syst. I: Regul. Pap.* **2021**, *68* (3), 1134−1145.
(96) Horowitz, M. Computing's energy problem (and what we can do about it). In *2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, IEEE, 2014.
(97) Rezaeifar, F.; Ahsan, R.; Lin, Q.; Chae, H. U.; Kapadia, R. Hot-electron emission processes in waveguide-integrated graphene. *Nat. Photonics* **2019**, *13*, 843−848.

https://doi.org/10.1021/acsnano.4c09055
ACS Nano 2024, 18, 23785−23796