Abstract—Analog neural networks represent a massively parallel computing paradigm by mimicking the human brain. Two important functions that are not efficiently built by CMOS technology for their practical hardware implementations are weighting for synapse circuits and summing for neuron circuits. In this paper we propose the use of tunable analog resistances, such as multi-gate graphene devices, to efficiently enable these two functions. We design and demonstrate a complete analog neuromorphic circuitry enabled by such devices. Simulation results based on Verilog-A compact models for graphene devices confirm its functionality. We also provide experimental demonstration of our proposed graphene device along with projected circuit performance based on scaling targets. Our demonstration of our proposed graphene device along with confirm its functionality. We also provide experimental results based on Verilog-A compact models for graphene devices two functions. We design and demonstrate a complete analog neuromorphic computing enabled by multi-gate programmable resistive devices. neuromorphic circuits, several emerging technologies have been recently proposed for such purposes by exploiting their unique features [2]-[3]. In such implementations weighted signals are in current domain and various coupling mechanisms (e.g., thermal [2] and magnetic [3]) have been utilized to efficiently sum such current-mode signals and then generate a voltage-mode output signal. In this paper we propose to construct a fully-functional analog neural network using multi-gate programmable graphene devices. Different from the proposed designs in [2]-[3] this architecture represents couplings between neighboring neurons in voltage domain, and uses only electrical signals for both weighting and summing functions, thereby not requiring any special non-electrical properties for the enabling device.

I. INTRODUCTION

The human brain is a powerful computing system that performs information processing quite efficiently via its massively parallel neural mechanism. Neuromorphic circuits attempt to mimic this neural mechanism in the human brain via locally-coupled artificial neurons and synapses in order to overcome some bottlenecks of the traditional von Neumann machines, especially for computationally-intensive applications such as image processing and pattern recognition [1]-[3].

A conceptual neuromorphic computing architecture that can be used for both feed-forward and recurrent (feedback) networks is depicted in Fig. 1. In this model the outputs of the neighboring neurons are weighted with corresponding synaptic weights. Then these weighted signals are summed together in the neuron circuit to generate its own state. Finally, the neuron circuit takes this summation and generates its output based on an activation function that generally corresponds to a variant of the sigmoid function.

Two key challenges that require significant power and area resources when implemented with CMOS technology are to efficiently represent synaptic weights and sum neural signals coming from neighboring synapses. For example, the proposed CMOS-based designs in [4]-[5] attempt to use one digital-to-analog converter (DAC) and one variable gain multiplier for each artificial synapse, and one differential summing amplifier to represent each artificial neuron.

Although CMOS is inefficient for implementing analog

![Fig. 1. A generalized building block for analog neuromorphic circuitry.](image)

This work was supported by the Systems on Nanoscale Information fabrics (SONIC), one of six centers supported by the STARnet phase of the Focus Center Research Program (FCRP), a Semiconductor Research Corporation program sponsored by MARCO and DARPA.

II. MULTI-GATE TUNABLE RESISTIVE DEVICE EXAMPLE

Graphene is a two-dimensional material consisting of a single-atom thick layer of carbon that is arranged in a hexagonal lattice (see Fig. 2). With its 2D planar structure it has been shown to be compatible with traditional CMOS process [6]. Moreover, the charge carriers in graphene can move over great distances at a constant speed without scattering. This is similar to the behavior of photons travelling at the speed of light. In graphene the speed of charge carriers is slower than light by only a factor of 300 [7], thereby resulting in high saturation velocity. Graphene also has high carrier mobility and zero energy band-gap. Thus, it offers unique opportunities for future nanoelectronics such as high frequency applications [8]. In addition, the linear dispersion relation in graphene gives rise to relativistic behavior of charge carriers resulting in photon-like behavior and Klein tunneling [9].

Here we propose a novel graphene device configuration that could be used to represent artificial neurons and synapses. The number of charge carriers in a graphene ribbon can be controlled through local gating [10] that allows the gradual resistance change of the device. This is because the electric field between the gate and graphene ribbon due to applied gate voltage, attracts electrons in graphene, thereby altering its conductivity. By means of this feature and zero energy band-gap offered by this particular material, a resistor string can be built using multiple gates over a single continuous graphene ribbon, as illustrated in Fig. 3. Although the gates effectively dope the graphene creating multiple junctions, conductivity is maintained due to Klein tunneling at the junctions [9].

![Fig. 2. A graphene device [11].](image)
and corresponding gate length (i.e., contribution of each gate on the resistance level can be gate controls the conductivity of the corresponding area underneath it. Fig. 3. Cross-section drawing of the multi-gate graphene resistance. Each gate in both devices controls a binary-weighted resistor based on the corresponding area on graphene ribbon (e.g., \( l_1=2x \), \( l_2=2x \), and \( l_3=4x \), referring to Fig. 3). The number of control bits, \( i \), can be adjusted based on the requirements for the network function where the number of allowable resistance values is equal to \( 2^{i+1}-1 \), because each synapse consists of two components. Generally 15-31 different synaptic values are enough for a wide range of neural network applications [12]-[13]. This device configuration enables a compact D/A (digital-to-analog) conversion for digitally-controlled synaptic weights. The total resistance of two devices in each synapse component is always constant since binary bits are applied to them are complementary to each other (see Fig. 4). Hence the couplings between neuron output and synaptic weights are linear as given by:

\[
V_{\text{ex}_w} = \frac{V_{\text{out}} R_{\text{ex}_w}}{R_{\text{total}}} \quad \text{for excitatory synapse} \tag{1}
\]

\[
V_{\text{in}_w} = \frac{V_{\text{out}} R_{\text{in}_w}}{R_{\text{total}}} \quad \text{for inhibitory synapse} \tag{2}
\]

where \( V_{\text{out}} \) is the output of the corresponding neuron; \( R_{\text{ex}_w} \) and \( R_{\text{in}_w} \) denote excitatory and inhibitory synaptic weights, respectively; \( V_{\text{ex}_w} \) and \( V_{\text{in}_w} \) are weighted voltage signals corresponding to excitatory and inhibitory components, respectively; and

\[
R_{\text{total}} = R_{\text{ex}_w} + R_{\text{in}_w} = R_{\text{ex}_w} + R_{\text{in}_w} \tag{3}
\]

where \( R_{\text{ex}_w} \) and \( R_{\text{in}_w} \) are complementary versions of \( R_{\text{ex}_w} \) and \( R_{\text{in}_w} \), respectively.

The excitation component of each synapse circuit is activated by increasing \( R_{\text{ex}_w} \) via digital control bits (i.e., \( b_{\text{ex}_1}, b_{\text{ex}_2}, b_{\text{ex}_3} \)) in Fig. 4) when neighboring neurons are coupled to each other via a positive synaptic weight (i.e., two neighboring neurons, each representing an image pixel, that are more likely to be of the same color, either white or black). The inhibition component is activated in the same way as the excitation component (i.e., via \( b_{\text{in}_1}, b_{\text{in}_2}, b_{\text{in}_3} \)) in Fig. 4) when the relationship between neighboring neurons is represented by a negative synaptic weight (i.e., two neighboring neurons, each representing an image pixel, that are more likely to be of the opposite colors). A synaptic weight can be either positive or negative, so the corresponding synapse component is deactivated by setting \( b_{\text{ex}_1}, b_{\text{ex}_2}, b_{\text{ex}_3} \) or \( b_{\text{in}_1}, b_{\text{in}_2}, b_{\text{in}_3} \) to 0.

An interesting property of this proposed architecture is that couplings between neighboring neurons are voltage-mode signals instead of current-mode signals, in contrast to other proposed resistive networks [2]-[3]. This is dictated by the gates being controlled by voltage-mode signals. It is important to note that inhibitory synapses are connected to the gates of pull-up device in the neuron circuits, while excitatory synapses are connected to the gates of pull-down device, as shown in Fig. 4. This is because weighted signals coming from artificial synapses increase the resistance of the corresponding device. Therefore, when excitation becomes more dominant as compared to inhibition, the neuron circuit generates an output level that is closer to the supply voltage, \( V_{\text{supply}} \), in Fig. 4. In contrast, when inhibition is more dominant, the neuron circuit produces an output that is closer to 0V (ground).

IV. CIRCUIT SIMULATIONS

To evaluate the potential of our proposed neurocomputing architecture in Fig. 4 we have created a compact circuit simulation model in Verilog-A for the graphene device configurations based on device measurement data [10]. The relationship between graphene resistance and applied gate

Fig. 4. Proposed neuron and synapse circuits based on multi-gate programmable resistive devices. \( b_{\text{ex}_1}, b_{\text{ex}_2}, b_{\text{ex}_3} \) and \( b_{\text{in}_1}, b_{\text{in}_2}, b_{\text{in}_3} \) represent \( i \)-bits binary numbers for excitatory and inhibitory synaptic weights, respectively.
电压被模型为一个饱和函数，带有两个切线电压（即，$V_{cut\_min}$和$V_{cut\_max}$）。这个函数输出的最低电阻值（$R_{min}$）当应用栅电压小于$V_{cut\_min}$，最大电阻值（$R_{max}$）当应用栅电压高于$V_{cut\_max}$。当它介于这两电压之间时，设备的电阻线性增加与应用栅电压（从$R_{min}$到$R_{max}$）。

在我们的电路模拟中，我们使用$V_{supply}=1V$, $V_{cut\_min}=0.1V$, $V_{cut\_max}=0.9V$，且$R_{max}/R_{min}=100$，这在我们技术[10]中是可实现的。这个R-V特性可以由我们电路中一个适当选择的参考电压对应的_embedded_footnotes_text into content_output_text_text_output_text_2015 Design, Automation & Test in Europe Conference & Exhibition (DATE)

这就给我们创造了一个5-神经元的示例，它成功地回忆起存储的模式。即使有40%的输入模式失真，它仍然能够识别到模式[1 0 1 1 0]。图5展示了模式识别的示例。

电路中使用的神经元电路有9个插入层（即，$l_1=l_2=\ldots=l_9$）所以每个神经元有9个本地连接。用于在我们的电路中模拟我们所需的电压（例如，约-3V）或一个精心调整的背栅/衬底电压，或者两者。我们连接了一个1pF电容在每个神经元电路到模型设备的寄生。

所使用的设备在神经元电路中有9个插入层，这些层是插入层电路中的二进制权重的，也就是说，$l_1=2l_2=4l_3$。这些权重允许15种不同的突触权重，可以在模型中定义（考虑了兴奋性和抑制性突触）。设备电阻在突触电路中被设计为100倍于在神经电路中，以防止显著的电压在输出电压中出现的，由于扇出。

我们首先展示了一个5-神经元的联想记忆基于该架构。根据图4。使用一个模式的示例[1 0 1 1 0]和[1 0 1 0 1]作为在我们的电路中模拟我们所需的电压（例如，约-3V）或一个精心调整的背栅/衬底电压，或者两者。我们连接了一个1pF电容在每个神经元电路到模型设备的寄生。

这些设备使用的神经元电路有9个插入层（即，$l_1=l_2=\ldots=l_9$）所以每个神经元有9个本地连接。用于在我们的电路中模拟我们所需的电压（例如，约-3V），或者一个精心调整的背栅/衬底电压，或者两者。我们连接了一个1pF电容在每个神经元电路到模型设备的寄生。

我们还构建了一个20-像素灰度样本，用于更大神经网络电路，具有局部互连接通的人工神经元（例如，对于每个神经元有9个连接）基于的架构。在图4中，我们展示了我们所关心的相同的设备模拟模型来评估一个包含20个神经元的联想记忆。图6中的模式所示。模式在图6中显示。

图5展示了模式识别的示例。联想记忆充分识别模式[1 0 1 1 0]，尽管有40%的失真在原始输入模式中。

图6展示了20-神经元网络的模式记忆示例。联想记忆成功地回忆起存储的模式。这个设备被设计为三种顶栅，它们的长度变化在二进制方式，与一个背栅一起。这种顶栅电路中使用的电压（用以控制设备的电阻范围）是由组合应用顶栅和背栅电压。背栅电压电压被用于控制设备的电阻范围，而且顶栅是用于控制电荷载荷在背栅下。由此有效地调节电阻值的范围。

测量结果提供在图10。尽管这种原型在用于表征取一个更大的设备（1μmx10μm）时，我们不会期望使用在我们的电路设计中，它展示了预期的行为。这种一式多样的设备提供了一个简单电压作为功能的，突触电压，并展示了单个的性能，可以用于灵活的多栅可编程电阻。在紧凑和有效的方式。通过进一步的发展和规模，可以使用这种技术的高度灵活的神经网络电路组件可以被制造和集成在一起，从而构建一个完全功能的系统。
It should be noted that this controlled resistance based on gating of a zero band-gap graphene ribbon would not be possible with a MOSFET device due to the inherent threshold voltage of the latter.

With proper selection of back gate voltage, the minimum device dimensions possible with today’s lithography could be used for the proposed graphene devices while still attaining the required high-to-low resistance ratio (e.g., 100 used in our circuit simulations). A sheet resistance of approximately 1 kΩ can be achieved for a 0.1 nm thick graphene ribbon using today’s technology [16]. A minimum size device can be used in the neuron circuits, and larger devices in the synapse circuits (e.g., 100x larger). Our simulations indicate that with the device width of 10 nm and varying lengths of 10 nm to 1 μm for the minimum-size devices used in the circuitry the proposed implementation provides more than three orders of magnitude improvement in both power and area when compared to a CMOS-based design fabricated in 65 nm.

VI. CONCLUSION

In this paper we propose and design a complete analog neuromorphic circuitry based on the multi-gate programmable resistances that provide a natural summing and compact D/A conversion for the implementation of artificial neurons and synapses. The operation of the proposed architecture has been confirmed using simple gray-scale pattern recognition and image processing examples. The fabrication process and measurement results for enabling device have been provided to anticipate scaling targets required for a highly-efficient network realization. This research is intended to provide insights into the development of new devices with intriguing capabilities that could enable feasible implementations of analog neurocomputing circuits.

REFERENCES