Automatic gain control of ultra-low leakage synaptic scaling homeostatic plasticity circuits

Ning Qiao
Giacomo Indiveri Institute of Neuroinformatics
University of Zurich and ETH Zurich
Zurich, Switzerland
Email: [qiaoning|giacomo]@ini.uzh.ch Chiara Bartolozzi
iCub Facility
Istituto Italiano di Tecnologia
Genova, Italy
Email: chiara.bartolozzi@iit.it

Abstract—**Homeostatic plasticity is a stabilizing mechanism that allows neural systems to maintain their activity around a functional operating point. This is an extremely useful mechanism for neuromorphic computing systems, as it can be used to compensate for chronic shifts, for example due to changes in the network structure. However, it is important that this plasticity mechanism operates on time scales that are much longer than conventional synaptic plasticity ones, in order to not interfere with the learning process. In this paper we present a novel ultra-low leakage cell and an automatic gain control scheme that can adapt the gain of analog log-domain synapse circuits over extremely long time scales. To validate the proposed scheme, we implemented the ultra-low leakage cell in a standard 180 nm Complementary Metal-Oxide-Semiconductor (CMOS) process, and integrated it in an array of dynamic synapses connected to an adaptive integrate and fire neuron. We describe the circuit and demonstrate how it can be configured to scale the gain of all synapses afferent to the silicon neuron in a way to keep the neuron's average firing rate constant around a set operating point. The circuit occupies a silicon area of $84\,\mu$m$\times 22\,\mu$m and consumes approximately 10.8 nW with a 1.8 V supply voltage. It exhibits time constants of up to 25 kilo-seconds, thanks to a controllable leakage current that can be scaled down to 1.2 atto-Amps (7.5 electrons/s).**

I. Introduction

One of the most remarkable properties of nervous systems is their ability of changing and adapting to the environment, in order to achieve and maintain robust computation. To this effect, a wide variety of *plasticity* mechanisms have been observed in neural circuits, optimized to attain specific goals: short-term plasticity over short temporal scales (e.g. of milliseconds) can mediate the selectivity to transient stimuli and contrast adaptation [1]; over longer time scales (e.g., tens to hundreds of milliseconds), spike-based synaptic plasticity mechanisms, such as Spike-Timing Dependent Plasticity (STDP), have been shown to mediate learning processes [2]; finally, over very long time scales (e.g., minutes, to hours, or more) it has been shown that intrinsic and homeostatic plasticity mechanisms are useful for adapting the system to long-lasting changes, maintaining the overall activity of neurons within functional boundaries [3]. Homeostatic plasticity comprises a variety of mechanisms acting at different levels, ranging from the tuning of the neuron excitability to modulating their activity by acting on the relative gain of the synapses connected to the neuron [4]. Synaptic scaling is a homeostatic mechanism that uniformly scales the efficacy of all the synapses impinging on the same neuron using a multiplicative effect which preserves the different ratios of synaptic weights among the synapses without disrupting the effect of activity dependent learning. This mechanism is crucial to adapt the activity of neural networks to compensate for changes in external conditions, such as increases in the input activity levels, or temperature drifts, or internal malfunctions of parts of the network. Homeostasis is therefore a particularly useful engineering strategy for the design of robust computational architectures in artificial neural networks that can automatically change their internal parameters to account for long-lasting changes in their operating conditions. However, despite being extremely important for the design of large scale neuromorphic computing platforms, only few works addressed the implementation of homeostasis in silicon neural networks, mainly because of the technical difficulty in obtaining the necessary extremely long time constants with the intrinsically fast CMOS circuital elements. Existing approaches focus on the use of floating gate transistors [5], [6], or propose to use off-chip methods, that require external memory and digital circuits [7]. Here we propose a novel auto-gain synaptic scaling circuit that exploits the features of an ultra-low leakage cell implemented using standard CMOS technology able to achieve extremely long time constant [8]–[10]. The design proposed here represents an improvement over a previous attempt [8] that could not achieve the same time scales, and that was more difficult to control. The synaptic scaling effect is obtained by making use of Differential Pair Integrator (DPI) synapse [11], [12] circuits, which have two independent parameters that can be used to set the global synaptic scaling term (via the homeostatic circuit), and the local individual synaptic weight terms (e.g., via spike-based learning circuits). In the next Sections we describe the circuits that implement the homeostatic automatic gain control loop and the ultra-low leakage CMOS cell, and
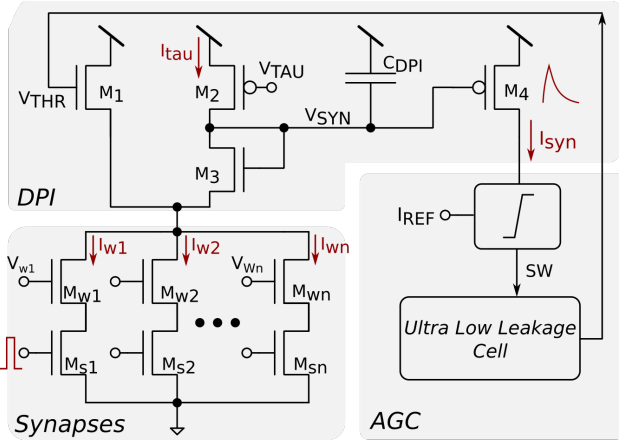
Fig. 1. Block diagram of proposed homeostatic AGC loop. The output current of the DPI block I_{syn} is scaled automatically over long time scales, by up- or down-regulating the V_{THR} control voltage.

Fig. 2. Circuit implementation of the LLC used in the AGC loop.

present experimental results obtained from the measurements of a fabricated test circuit.

II. THE HOMEOSTATIC AUTOMATIC GAIN CONTROL LOOP

Typical neuromorphic computing architectures comprise arrays of silicon neurons each receiving input from a large number of input synapses [12], [13]. In these systems, it is possible to maintain neuron's overall spiking activity within given operating boundaries, without interfering with the network's signal processing and learning mechanisms, by adopting automatic gain control mechanisms with very long time constants for globally scaling the synaptic weights of the synapse circuits afferent to their corresponding neuron. A circuit that allows independent control of the synapse scaling gain and its synaptic weight is the DPI [11]. This circuit is a current-mode log domain integrator circuit. If all the synapses afferent to the neuron share the same temporal dynamics, it is possible to use one single integrator circuit per neuron and use the temporal superposition principle to combine the output of multiple synapses (see the multiple I_{wi} currents in Fig. 1). The circuit has the following transfer function (see [12] for a more thorough analysis using the translinear principle, and [11] for a time-domain linear system's analysis):

$$\tau_s \frac{d}{dt} I_{syn} + I_{syn} = \frac{I_w I_{gain}}{I_\tau} \quad (1)$$

where the term τ_s is $(C_{DPI}U_T)/(\kappa I_\tau)$, with U_T representing the thermal voltage, and κ the sub-threshold slope coefficient. At equilibrium, the steady-state value of I_{syn} is $I_{syn} = I_w I_{gain}/I_\tau$. The current I_τ is a bias current that needs to be tuned to properly set the integrator time constant. The current I_w corresponds to the sum of the individual synapse input currents $I_w = \sum_i I_{wi}$, set by their corresponding synaptic weight bias voltages V_{wi}. The current I_{gain} on the other hand represents an extra independent term that can be set by additional control circuits. This current is defined as

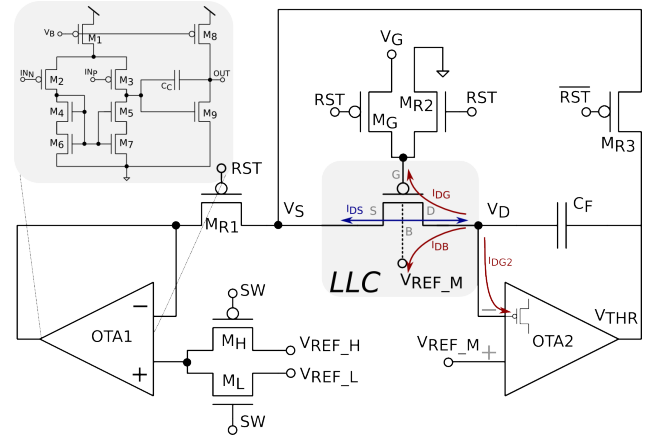$$I_{gain} = I_0 e^{\frac{\kappa(V_{THR}-V_{dd})}{U_T}} \quad (2)$$

It represents a virtual P-type subthreshold current that is not tied to any p-FET in the circuit of Fig. 1. By adjusting V_{THR}, I_{gain} can be tuned so as to increase or decrease I_{syn}, independent from changes of I_w (e.g., due to the regular learning process). A copy of the DPI output current I_{syn} is eventually injected into the neuron, which will then produce a firing rate proportional to its amplitude. Figure 1 shows how the AGC homeostatic control block is used to modulate the voltage V_{THR} in order to maintain the current I_{syn} around a set reference current I_{REF}: the I_{syn} current is fed into a high-gain voltage comparator that checks which of the voltages that set the I_{syn} and I_{ref} currents is greater than the other. Depending on the outcome of this comparison, the digital output voltage SW of this comparator is set to either ground or V_{DD}. This digital signal is then used to gate the control signals of a Low Leakage Cell (LLC) circuit which slowly adjusts V_{THR} to up-regulate or down-regulate I_{syn} accordingly.

III. THE ULTRA-LOW LEAKAGE CELL

To achieve long biological realistic time-scales, it is necessary to develop circuits with time constants that range from milliseconds to hours. Since these stringent specifications need to be met while keeping the circuit's capacitance to a minimum (e.g. in order to integrate many of these circuits on a single die), these circuits are required to produce extremely small currents. An example of such a circuit is the ultra-low leakage cell shown in Fig. 2. This circuit increases or decreases its output voltage V_{THR} by controlling the direction of a very small current across the channel an LLC p-FET to slowly charge or discharge the capacitor C_F. As in our LLC circuit implementation the capacitance C_F is set to 1 pF, the currents required to achieve kilo-second time constants have to be of the order of 1 fA, which is usually hundred times smaller than channel off-state leakage current of transistors, for the standard 180 nm CMOS process used. Ultra-low ranges of currents can be obtained by minimizing the leakage currents across LLC p-FET [9]. In particular, the leakage current I_{DB} can be minimized by biasing V_{DB} to be zero; this condition can be met by using a feedback Operational Transconductance Amplifier (OTA)

with large enough gain (see OTA2 in Fig.2). In order to get ultra-small leakage current from node D of the LLC p-FET, it is necessary to minimize the gate leakage currents I_{DG} and I_{DG2}: gate leakage current density normally is exponentially related to the thickness of gate oxide and strongly depends on gate bias [14]. For a standard 180 nm process with a gate oxide thickness of 4.6 nm, it is reasonable to assume the gate leakage current density with gate bias of 0.5 V to be smaller than $10^{-8}\,A/m^2$. To minimize these currents we designed the low leakage transistor with a W/L ratio of 0.5μm/1μm and the p-FET input transistor of the OTA2 with a W/L ratio of 8μm/1μm. Therefore the total gate leakage current is estimated to be smaller than 0.1 aA. While the OTA2 amplifier is used to implement a high-gain negative feedback loop to keep the potential of V_D as close as possible to V_{REF_M}, the OTA1 amplifier is used to clamp the voltage V_S of the LLC p-FET to one of the two V_{REF_L}, V_{REF_H} reference voltages. The detailed circuit schematic diagram of the OTA1 and OTA2 amplifiers are shown in the top-left inset of Fig. 2. To ensure high-gain and rail-to-rail output range, while minimizing power, we adopted a two stage pseudo-cascode split-transistor sub-threshold technique [15].

Given these small currents, at the beginning of an experiment it is necessary to initialize the AGC control loop to a proper initial condition, such as $V_{THR} = V_D = V_{REF\_M}$. This can be done by enabling the digital control signal RST to high, and resetting it to ground shortly after. At this point the direction of the current across the LLC p-FET of Fig. 2 will be set by the digital control signal SW, produced by the comparator of Fig. 1. If SW is high, then the V_{DS} of the LLC p-FET will correspond to $V_{REF\_M} - V_{REF\_L}$, otherwise it will correspond to $V_{REF\_M} - V_{REF\_H}$. By appropriately setting these reference voltages such that $V_{REF\_L} < V_{REF\_M} < V_{REF\_H}$, and by properly tuning the LLC p-FET's gate voltage V_G, it is possible to precisely control both direction and amplitude of the LLC p-FET I_{DS} current.

During normal operation, if the total synaptic drive I_{syn} increases above the reference current I_{REF}, the comparator will set the digital signal SW to high. Since this enables the signal V_{REF_L} as input to the OTA1 amplifier, the current will slowly discharge C_F and cause an increase in V_{THR}. This will in turn downscale the value of I_{gain} of eq. (2), effectively reducing the synaptic current I_{syn} injected into neuron, and compensating for the initial change. Conversely, as I_{syn} decreases below I_{REF}, the comparator will enable V_{REF_H} as input to OTA1. This will cause the LLC p-FET current to slowly charge the C_F capacitor, thereby decreasing V_{THR} and increasing I_{gain}. This will counteract the source of the disturbance that caused the initial decrease of I_{syn}, and increase it back, until it reaches again the reference level I_{REF}.

IV. EXPERIMENTAL RESULTS

To characterize the response properties of the proposed circuits, we designed a prototype test chip in standard 180 nm CMOS process comprising a small array of neurons and synapses with embedded synaptic scaling circuits. Figure 3 shows the die-photo of the fabricated chip, with the synaptic scaling circuits highlighted in the Neuron/Synapse Array #1.
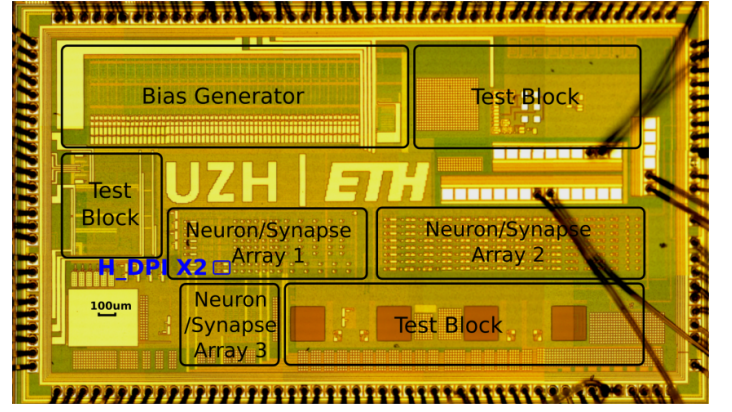

Fig. 3. Die photo of test chip implemented using a standard 180 nm CMOS process. The proposed DPI-based very long time scale automatic gain control synaptic scaling circuits are embedded in the Neuron/Synapse Array #1, and circled in blue. The whole chip occupies an area of 3.96 mm×2.29 mm, and the synaptic scaling circuits occupy an area of 84 μm×22 μm.
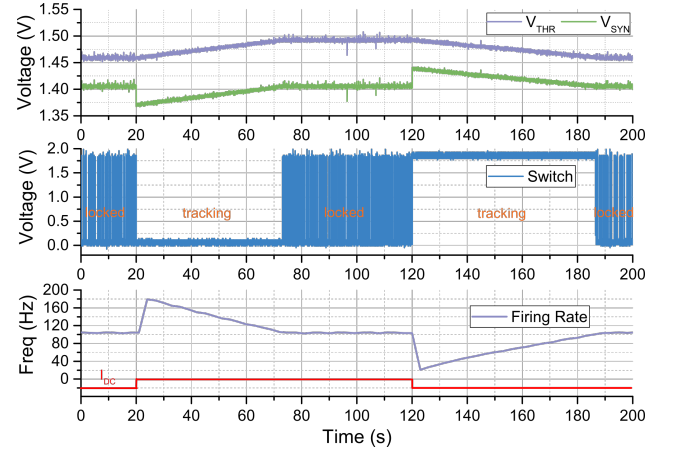

Fig. 4. Synaptic homeostasis measurements in response to step changes of the DPI input current. (Top): the voltage traces V_{THR} and V_{SYN}; (Middle): the comparator output digital signal SW; (Bottom): neuron's instantaneous firing rate and its input DC current.

In Fig. 4 we show the response of the circuit to a DC change in the input current I_{DC} applied as synaptic weight input current into the circuit's DPI block (see also Fig. 1).

In this experiment we set I_{DC} to start at 0.3 nA, the reference current I_{REF} to be 20 nA, and the parameters of the silicon neuron (e.g. gain, time constant and refectory period) in a way to obtain a firing rate of approximately 100 Hz. By properly setting the V_G bias voltage of Fig. 2, we tuned the time constant of the homeostatic control circuit to be around 60 seconds. In these conditions, the AGC loop of Fig. 1 clamps V_{THR} to a value around 1.46 V, and V_{SYN} around 1.4 V, thus maintaining the neuron's firing rate stable at its initial value. After 20 seconds we made I_{DC} change from 0.3 to 0.6 nA. As expected, this increased the DPI output current I_{syn}, decreased the V_{SYN} voltage accordingly, and increased the neuron's firing rate from 100 to about 180 Hz. The synaptic scaling homeostatic circuits now start having an effect and slowly scale down the total
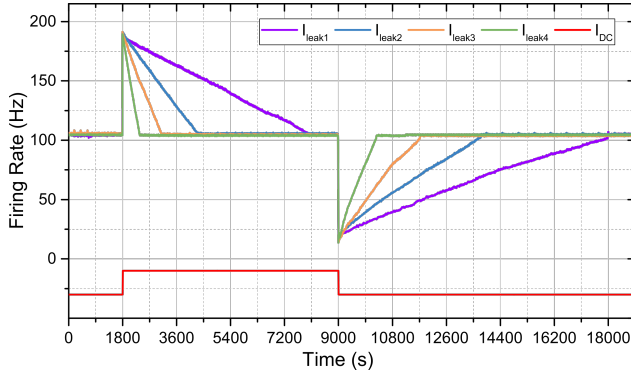
Fig. 5. Neuron's firing rate modulated by the homeostatic mechanism, tuned to respond with different time scales. The bottom red curve represents the DPI's input current I_{DC}.

TABLE I. Homeostatic plasticity circuit key figures.

Process Technology	AMS 180 nm 1P6M CMOS
Silicon Area of DPI	$84\,\mu$m $\times$ $22\,\mu$m
Size of LLC (W/L)	$0.5\,\mu$m / $1\,\mu$m
Power Consumption	10.8 nW
Temporal Constant	25 k seconds
Leakage Slope (1pF)	$1.2\,\mu$V/s
Controllable Leakage Current	1.2 aA (7.5 Electrons/sec)

synaptic current I_{syn} being injected in the neuron, which in turn starts to slowly decrease its output firing rate. This is done by slowly increasing the V_{THR} signal, which is shared by all input synapses afferent to the same neuron, and which modulated the I_{gain} current. After approximately 60 seconds I_{syn} and the firing rate of the neuron are both restored to their initial values. At around $t = 120\,$s we change the I_{DC} current back from 0.6 to 0.3 nA. In this case, the neuron's firing rate drops below its original value and the AGC loop is activated such that after about 60 seconds, the neuron's firing rate is restored back to its original value. Due to the *bang-bang* nature of the the AGC control loop, when the neuron's firing rate is close to the reference (see "locked" regions in Fig. 4) the homeostatic circuits keep on alternating the SW signal from high to low, in order to keep the I_{syn} current around the I_{REF} reference current.

In Fig. 5 we show how we can tune the homeostatic circuits to work with different leakage rates. These can be achieved by changing the V_G bias voltage of LLC p-FET, which sets the amplitude of the I_{DS} current on Fig. 2, and by modulating the difference between V_{REF_M}, and V_{REF_L}/V_{REF_H}, which control the voltage drop across the LLC p-FET channel.

In each condition the AGC succeeds in restoring the neuron's firing rate to its original 100 Hz rate. Although the longest time scale we measured in this experiment is around 9k seconds, we verified, with further tests that the same experimental setup could achieve time scales of about than 25k seconds. In these test the voltage on the 1 pF C_F capacitor of Fig. 2 changed of approximately 30 mV (similar to what happens in the top trace of Fig. 4) with a slope of $1.2\,\mu$V/s. Therefore the leakage current used to discharge the capacitor is approximately 1.2 aA (7.5 Electrons/second).

A summary of the key figures of the homeostatic synaptic scaling circuit is shown in Table I.

V. Conclusion

We presented a compact low-power ultra-low leakage synaptic scaling circuit for implementing homeostatic plasticity mechanism with biologically realistic time constants in standard CMOS processes. We showed how the DPI-based automatic gain control circuit can properly tune the gain of DPI for appropriately scaling the circuit's output current. We designed, fabricated and tested an ultra-low leakage cell that allowed us to obtain extremely long time constants in a controllable way. We measured the low leakage currents obtained from well-biased signal p-FET device and demonstrated how, with a 1 pF capacitor, it is possible to reach time scales as large as 25 k seconds, and leakage currents as small as 1.2 aA. The proposed circuits occupies an area of $84\,\mu$m$\times22\,\mu$m in a standard 180 nm process, and consumes 10.8 nW with 1.8 V supply power during normal operation. In comparison to previously proposed designs, this circuit does not require additional floating gate devices or off-chip methods. This makes it suitable for being easily integrated with other low-power neuromorphic circuits on the same device.

References

[1] F. Chance, S. Nelson, and L. Abbott, "Synaptic depression and the temporal response characteristics of V1 cells," *The Journal of Neuroscience*, vol. 18, no. 12, pp. 4785–99, 1998.

[2] L. Abbott and S. Nelson, "Synaptic plasticity: taming the beast," *Nature Neuroscience*, vol. 3, pp. 1178–1183, November 2000.

[3] G. Turrigiano and S. Nelson, "Homeostatic plasticity in the developing nervous system," *Nature Reviews Neuroscience*, vol. 5, pp. 97–107, February 2004.

[4] G. Turrigiano, "Homeostatic plasticity in neural networks: the more things change, the more they stay the same," *Trends in Neuroscience*, vol. 22, no. 5, pp. 221–227, 1999.

[5] S.-C. Liu and B. Minch, "Homeostasis in a silicon integrate-and-fire neuron," in *Advances in Neural Information Processing Systems (NIPS)*, T. Leen, T. Dietterich, and V. Tresp, Eds., vol. 13. MIT Press, 2001, pp. 727–733.

[6] S. Nease and E. Chicca, "Floating-gate-based intrinsic plasticity with low-voltage rate control," in *International Symposium on Circuits and Systems ISCAS 2014*. IEEE, 2016, pp. 2507–2510.

[7] C. Bartolozzi, O. Nikolayeva, and G. Indiveri, "Implementing homeostatic plasticity in VLSI networks of spiking neurons," in *International Conference on Electronics, Circuits, and Systems, ICECS 2008*. IEEE, 2008, pp. 682–685.

[8] G. Rovere, N. Qiao, C. Bartolozzi, and G. Indiveri, "Ultra low leakage synaptic scaling circuits for implementing homeostatic plasticity in neuromorphic architectures," in *International Symposium on Circuits and Systems, (ISCAS), 2014*. IEEE, 2014, pp. 2073–2076.

[9] K. Roy, S. Mukhopadhyay, and H. Mahmoodi-Meimand, "Leakage current mechanisms and leakage reduction techniques in deep-submicrometer CMOS circuits," *Proceedings of the IEEE*, vol. 91, no. 2, pp. 305–327, 2003.

[10] M. O'Halloran and R. Sarpeshkar, "A 10-nw 12-bit accurate analog storage cell with 10-aa leakage," *Solid-State Circuits, IEEE Journal of*, vol. 39, no. 11, pp. 1985–1996, 2004.

[11] C. Bartolozzi and G. Indiveri, "Synaptic dynamics in analog VLSI," *Neural Computation*, vol. 19, no. 10, pp. 2581–2603, Oct 2007.

[12] E. Chicca, F. Stefanini, C. Bartolozzi, and G. Indiveri, "Neuromorphic electronic circuits for building autonomous cognitive systems," *Proceedings of the IEEE*, vol. 102, no. 9, pp. 1367–1388, Sep 2014.

[13] N. Qiao, H. Mostafa, F. Corradi, M. Osswald, F. Stefanini, D. Sumislawska, and G. Indiveri, "A re-configurable on-line learning spiking neuromorphic processor comprising 256 neurons and 128k synapses," *Frontiers in Neuroscience*, vol. 9, no. 141, 2015.

[14] W. K. Henson, N. Yang, S. Kubicek, E. M. Vogel, J. J. Wortman, K. De Meyer, and A. Naem, "Analysis of leakage currents and impact on off-state power consumption for CMOS technology in the 100-nm regime," *Electron Devices, IEEE Transactions on*, vol. 47, no. 7, pp. 1393–1400, 2000.

[15] M. Yang, S.-C. Liu, and T. Delbruck, "Subthreshold DC-gain enhancement by exploiting small size effects of MOSFETs," *Electronics Letters*, vol. 50, no. 11, pp. 835–837, 2014.