

SCIENTIFIC REPORTS



OPEN

Graphene-based nonvolatile terahertz switch with asymmetric electrodes

Yan Li^{1,2}, Hui Yu¹, Xinyu Qiu¹, Tingge Dai¹, Jianfei Jiang¹, Gencheng Wang¹, Qiang Zhang¹, Yali Qin², Jianyi Yang¹ & Xiaoqing Jiang¹

We propose a nonvolatile terahertz (THz) switch which is able to perform the switching with transient stimulus. The device utilizes graphene as its floating-gate layer, which changes the transmissivity of THz signal by trapping the tunneling charges. The conventional top-down electrode configuration is replaced by a left-right electrode configuration, so THz signals could transmit through this device with the transmissivity being controlled by voltage pulses. The two electrodes are made of metals with different work functions. The resultant asymmetrical energy band structure ensures that both electrical programming and erasing are viable. With the aid of localized surface plasmon resonances in graphene ribbon arrays, the modulation depth is 89% provided that the Fermi level of graphene is tuned between 0 and 0.2 eV by proper voltage pulses.

The terahertz technology has gained increased research attention due to its industrial and academic applications^{1,2}. Various approaches, including those based on active metamaterials^{3–5} and two-dimensional materials^{6–8}, have been investigated over the past decade to manipulate the propagation of electromagnetic waves from RF to THz frequencies. However, most of them are volatile devices, as they return to original states once external stimulus are removed. Nonvolatile photonic devices, which can sustain the transformed properties with short triggering stimulus rather than persistent signals, hold an apparent advantage of saving the energy in the advent of photonic technologies. More importantly, the persistent photonic memory effect of these devices lend themselves to a variety of applications including protective optical/terahertz circuitry, adaptable transformation-optics devices, dynamically reconfigurable optical/THz networks or metamaterials, electrically controllable photonic memory^{9–15}, and so on. Furthermore, most planar lightwave circuits, especially those built on the silicon-on-insulator substrate that have high refractive index contrasts and strong optical confinements, are sensitive to fabrication tolerance. Therefore, many properties of practical photonic circuits such as the resonance wavelength would inevitably drift from their design values. Nonvolatile devices then can be incorporated into these photonics circuits as a post-trimming technique to compensate permanent errors.

Although a lot of nonvolatile devices, e.g., those based on floating gate structures^{14–16}, phase-change materials^{11–13}, organic materials¹⁷, ferromagnetic materials¹⁸, have been well developed for data storage, few of them turn out to be a feasible solution to implement the nonvolatile THz switching. The reason is that most of those devices contain functional materials or top-gate electrodes that are nontransparent at THz frequencies. For example, the floating-gate (FG) shown in Fig. 1(a) is so far the most successful configuration for the flash memory. It relies on charges stored in the FG layer to tune the threshold voltage of the transistor. In^{19,20} graphene is employed as the FG layer to store charges, which exhibits improved performance in terms of endurance, writing speed and operation voltage. However, THz signals cannot propagate through this configuration, since it demands a metallic control gate on the top, which is highly lossy at THz frequencies. Although metamaterial structures^{5,9–12}, for example, a periodic hole array patterned on an Au film as shown in Fig. 1(a), can be utilized to enhance the transmission through the metal layer, the limited bandwidth and the additional loss are tricky issues. Generally speaking, to design a technically viable and high-performance nonvolatile THz device still remains a challenge.

In this paper, we theoretically propose a nonvolatile THz switch by employing the graphene as the floating gate layer. The extraordinary optoelectronic property of graphene allows stored carriers to significantly tune the transmissivity of the THz signal. The conventional control-gate layer that covers the full device is replaced by source

¹College of Information Science and Electronics Engineering, Zhejiang University, Hangzhou, 310027, China.

²College of Information Engineering, Zhejiang University of Technology, Hangzhou, 310014, China. Correspondence and requests for materials should be addressed to H.Y. (email: huiyu@zju.edu.cn)

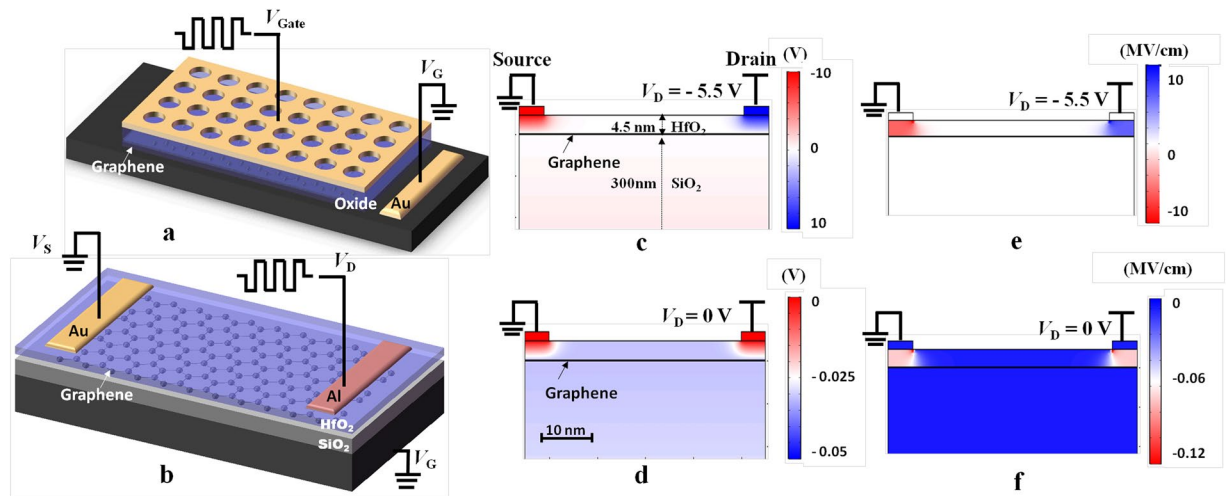


Figure 1. Device structures and numerical simulation results of electrostatic potential distributions. 3-D schematic diagrams of (a) the traditional FG memory, and (b) the graphene-based nonvolatile THz switch. The potential distributions in the nonvolatile THz switch at (c) the programming mode and (d) the retention mode, and the corresponding distributions of vertical electrical fields (e,f).

and drain electrodes made of different metals at the two sides as shown in Fig. 1(b). Hence the space between two electrodes acts as an open window for the THz transmission. With an appropriate thickness of tunneling layer (HfO_2), electrons from the drain electrode can tunnel through the HfO_2 layer and be stored in the FG (graphene) layer under a negative drain voltage. Meanwhile, since the tunneling barrier under the source electrode between graphene and HfO_2 is higher than that between the drain electrode and HfO_2 , electrons that tunnel out to the source electrode from the FG can be neglected. Therefore, electrons can accumulate in the FG. Similarly, under a positive drain voltage, the electrons stored in the FG can tunnel back to the drain, while electrons that tunnel into the FG from the source is negligible. Electrons therefore can be erased from the FG also due to the unequal tunneling barrier heights at the two electrode terminals. The nonvolatile switching is finally realized by taking advantage of different tunneling barriers at the two terminals.

Although the proof-of-principle design of the THz nonvolatile switch here is based on free-space optics with the THz wave normally incident on the graphene layer, this scheme is also well adapted to various waveguide-based switches operating at THz and optical wavelengths. A lot of graphene-based photonic modulators have been reported on either dielectric^{7,21,22} or plasmonic waveguides^{23–25}. We note that active graphene layers in many of these modulators can be readily replaced by the graphene-based floating-gate in Fig. 1(b) with the two asymmetrical electrodes locating at the two sides of the waveguides. Since waveguide modes are confined to propagate along the graphene plane, the interaction length between THz/optical waves and graphene is extended, and hence Performance metrics in terms of power consumption, footprint, and modulation depth can be improved significantly for nonvolatile switching.

Results

Graphene-based nonvolatile switching structure. A schematic diagram of the nonvolatile switch is shown in Fig. 1(b), which is built on the grounded silicon substrate. Monolayer graphene (work function: $\psi_g = 4.6 \text{ eV}$ ¹⁹) vertically stacking on a 300-nm-thick SiO_2 layer works as the floating-gate layer. Another 4.5-nm-thick HfO_2 layer ($\psi_{\text{HfO}_2} = 2.5 \text{ eV}$ ²⁶) acts as the tunneling/blocking layer. The drain and source electrodes on top are made of aluminum ($\psi_{\text{Al}} = 4.08 \text{ eV}$) and gold ($\psi_{\text{Au}} = 5.1 \text{ eV}$ ²⁷), respectively. To reveal the operation principle, the finite elements methods (Comsol Multiphysics) is utilized to model the electrostatic potential distribution in this device. The distance between the two electrodes is set to be 50 nm to better display the vertical configuration. It should be noted that the width of this transmission window for THz beams is on the scale of centimeters in practical devices. However, the electrostatic potential distribution in practical devices is similar to that presented in Fig. 1(c) and (d). On the other hand, although there are non-uniform fringe electrical fields at edges of electrodes as shown in Fig. 1(e) and (f), it is reasonable to assume fields between two electrodes and graphene are uniform when we calculate the tunnelling current. The reason is that widths of practical electrodes are on the scale of millimeter. For most areas below electrodes that are away from edges, electrical fields are uniform. In programming and erasing modes, large potential differences and accordingly high electrical fields are constructed between drain/source and graphene to enhance the tunneling of electrons; while low electrical fields are induced by charges stored in the FG layer for the retention mode. The extremely high conductivity of the graphene makes it an equipotential surface, which enables the electrons tunneling from the drain to spread throughout the entire graphene layer²⁰. Therefore, the potential difference from drain to graphene almost equals to that from graphene to source as shown in Fig. 1(c) and (d). Consequently, different metals can be utilized to induce imbalanced tunneling currents, which is crucial to implement necessary program/erase operations of a nonvolatile device as will be explained below.

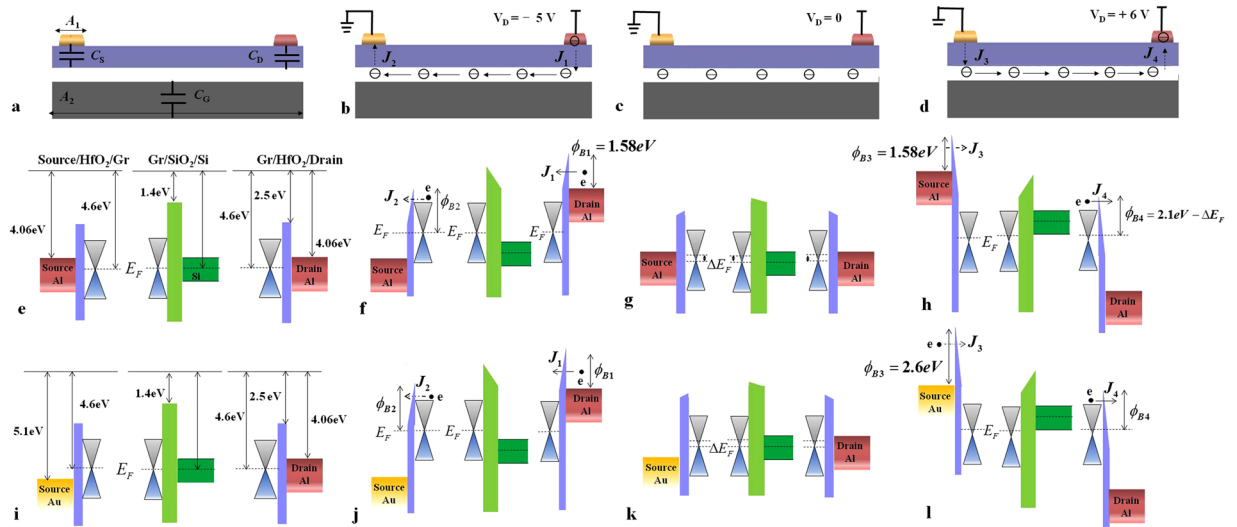


Figure 2. Different operation modes and corresponding vertically band diagrams of the nonvolatile THz device. (a) Initial mode (before metal/oxide/graphene contact) (b) programming mode, (c) retention mode and (d) erasing mode; the corresponding energy band diagrams for the symmetric electrodes (e–h) and asymmetric electrodes (i–l). The dashed line arrows indicate the tunneling directions of electrons between two electrodes and graphene; while the solid ones indicate the spreading of electrons in the graphene layer.

Physical mechanism of the nonvolatile operation. To explain the necessity of using different metals for source and drain, we at first examine the case of both electrodes being made of the same metal, for example, Aluminum (Al). It is known that in a graphene MOS structure, charges will accumulate in graphene provided that work functions of graphene and metal are different. This localized metal-induced doping effect leads to internal built-in electrical fields inside the graphene, and hence bend its energy band. However, these fields only exist in narrow regions ($\sim 0.2 \mu\text{m}$) adjacent to drain/graphene and source/graphene interfaces²⁸, which is far less than the distance between the two terminals of our device. Therefore, the band bending is neglected to simplify the band diagrams. Resultant directions of tunneling currents and band diagrams at different operation states are shown in Fig. 2. Figure 2(e) depicts the initial energy band (before contact) without any external voltage signal applied on the device. In Fig. 2(b), a negative programming signal ($V_D < 0$) is applied on the drain. As already analyzed in Fig. 1, nearly one half of the voltage drops between drain and graphene, while the other half drops between graphene and HfO_2 ($\phi_{B1} = \psi_{\text{Al}} - \psi_{\text{HfO}_2} = 1.58 \text{ eV}$) is lower than that between graphene and HfO_2 ($\phi_{B2} = \psi_{\text{g}} - \psi_{\text{HfO}_2} = 2.1 \text{ eV}$). Since the tunneling probability strongly depends on the barrier height, the tunneling current cross the right barrier is greater than that cross the left barrier ($J_1 \gg J_2$), and then electrons accumulate in the FG layer. In the retention mode (Fig. 2c), the voltage applied on the drain is removed. Due to the aggregated electrons ($Q_{\text{FG}} \sim 3 \times 10^{12}/\text{cm}^2$) in the FG layer, potential and Fermi level of graphene rise by $\sim 0.08 \text{ V}$ and $\sim 0.2 \text{ eV}$, respectively. Tunneling currents from the FG layer to the source and drain is negligible as a result of the weak electrical field in the blocking layer. Therefore, electrons could be stored in the FG layer for a long period before completely leaking out. In the erasing mode of Fig. 2(d), a positive voltage is applied on the drain which is supposed to clear electrons in the FG layer. As the work function of charged graphene layer ($4.4 \pm 0.2 \text{ eV}$) still outnumbers that of the Al (4.06 eV), the tunneling current from source to the graphene is greater than that from the graphene to the drain, i.e., $J_3 \gg J_4$ as shown in Fig. 2(h). There is a net electron tunneling current to increase the electron density in the FG layer. Therefore, the device is unable to accomplish the electrical erasing. In fact, as long as the device employs the same metal for source and drain, either electrical programming or erasing is forbidden.

To solve this issue, the Al source is replaced by a Gold (Au) source. The different work functions of Al and Au lead to an asymmetrical energy band structure as shown in Fig. 2(i). For the programming and retention modes presented in Figs 2(j) and 3(k), accumulation and leaking of electrons are analogous to those depicted in Fig. 2(f) and (g). However, since the work function of charged graphene ($4.6 \pm 0.2 \text{ eV}$) is less than gold (5.1 eV), the barrier height on the left side is higher than that on the right side during the erasing process ($\phi_{B3} > \phi_{B4}$) as shown in Fig. 2(l). The tunneling current J_3 is accordingly insignificant as compared with J_4 . Therefore, the stored electrons could be erased by a positive voltage due to the asymmetric-electrode configuration.

A quantitative calculation is provided below. Considering the two-dimensional character and band structure peculiarities of graphene, a rigorous treatment should resort to the Bardeen transfer Hamiltonian approach, as it has been utilized to explain tunneling currents in relevant two-dimensional systems^{29,30}. However, a strict calculation based on this approach is much more cumbersome than that based on the Fowler-Nordheim (FN) theory³¹. The FN theory is initially built on the free-electron theory of metals and leads to a very simple relationship between the density of emission current and the applied electric field. Strictly, the FN model applies only to bulk systems, e.g., metal/insulator/crystalline solid. However, FN equations are also frequently used as an approximation to describe low dimensional systems, including nanocrystals^{32,33} and two-dimension materials^{34,35}. For

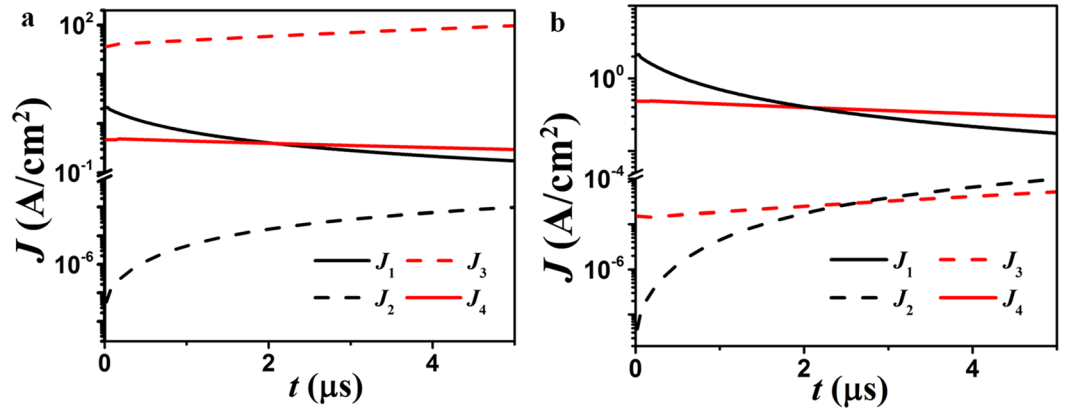


Figure 3. The time responses of the tunneling current density for different electrodes. (a) Symmetric Aluminum/Aluminum and (b) asymmetric Aluminum/Gold electrodes as drain/source.

example, the FN model captures measured tunneling currents in a graphene/insulator/metal system quite well with proper modifying factors in ref.³⁵. Based on these considerations, we employ the FN theory without considering the linear energy-momentum relation of graphene. This greatly simplifies the calculation with acceptable accuracy. Regarding the applicability of the FN model to the graphene/insulator/metal interfaces in our device, we discuss how the density of states of graphene affects the tunneling current density in the Supplementary Information. However, a rigorous treatment of the tunneling current is still suggested for our future work.

The conventional FN formula of tunneling current describe the field-assisted emission of electrons under a very high electrical field ($E_D > q\phi_{B1}/d$). It can be used to calculate tunneling currents in programming/erasing modes, but is not suitable for the retention mode ($V_D \approx 0$) of very weak electrical field. Generalized formulas are derived by Simmons which enable the calculation of tunneling currents for several different voltage ranges³⁶. According to Simmons' formulas, tunneling current densities J at programming/erasing and retention modes can be calculated as

$$\begin{aligned} J &= A_1 \phi_B^{1/2} \exp(-C_1 \phi_B^{1/2}), \quad (V_D \approx 0) \\ J &= A_2 E^2 \exp(-C_2/E). \quad (E_D > q\phi_{B1}/d) \end{aligned} \quad (1)$$

definitions of coefficients A and C are: $A_1 = q^2(2m^*)^{1/2}/(4d\pi^2\hbar^2)$ and $C_1 = 2d(2m^*)^{1/2}/\hbar$, $A_2 = q^3/(16\pi^2\hbar\phi_B)$ and $C_2 = 4(2m^*)^{1/2}(\phi_B)^{3/2}/3\hbar q$, here m^* denotes the effective mass of the electron, \hbar is the reduced Plank constant, ϕ_B is the barrier height and q is the electron charge. The electrical field between source and FG (E_D) and that between drain and FG (E_S) are determined by the potential of the floating gate layer V_{FG} , which depends on the amount of stored electrons Q_{FG}

$$V_{FG} = \frac{Q_{FG}}{C_T} + \frac{C_S}{C_T} V_S + \frac{C_D}{C_T} V_D + \frac{C_G}{C_T} V_G, \quad (2)$$

where V_S , V_D , and V_G are the potentials of source, drain and ground, respectively. As shown in Fig. 2(a), C_S , C_D , and C_G represent the capacitances (per unit area) between graphene/source, graphene/drain, and graphene/ground, respectively; and C_T is their sum ($C_T = C_S + C_D + C_G$)³⁷.

According to the MOS theory, C_S , C_D , and C_G cannot be simply described by their corresponding oxide capacitances C_{ox} alone due to the finite density of states (DOS) of graphene. A part of the applied voltage drops in the graphene layer itself, and this effect can be captured in the quantum capacitance C_q of graphene, which can be calculated as³⁸

$$C_q = \frac{2q^2 k_B T}{\pi(\hbar v_F)^2} \ln \left[2 \left(1 + \cosh \frac{E_F}{k_B T} \right) \right] \quad (3)$$

When the Fermi level of graphene locates at the Dirac point, i.e., $E_F = 0$, the quantum capacitance reaches its minimum value of $C_q(E_F = 0) = \frac{2q^2 k_B T}{\pi(\hbar v_F)^2} \ln(4) \approx 0.8 \mu\text{F}/\text{cm}^2$. In contrast, the capacitance of the 300-nm-thick bottom silica layer is $C_{Gox} = 11.5 \text{ nF}/\text{cm}^2$, which is far less than the quantum capacitance. Hence, the influence of the quantum capacitance on C_G can be neglected. On the other hand, as the thickness of the HfO_2 layer is very thin, its capacitance is comparable with the quantum capacitance. Therefore, both C_S and C_D should be considered as series connections of their respective oxide capacitances and quantum capacitances: $1/C_S = 1/C_{Sox} + 1/C_{Sq}$ and $1/C_D = 1/C_{Dox} + 1/C_{Dq}$.

The stored-charge density in the FG layer is related to the tunneling currents across the two barriers. It can be calculated as

$$\begin{aligned}\frac{dQ_{FG}}{dt} &= \frac{W_1}{W_2}(J_1 - J_2) \text{ (Program mode),} \\ \frac{dQ_{FG}}{dt} &= \frac{W_1}{W_2}(J_4 - J_3) \text{ (Erase mode),}\end{aligned}\quad (4)$$

W_1/W_2 is the width ratio of the electrode and the FG layer as shown in Fig. 2. The effect of V_G can be neglected due to the fact that C_G is far less than C_S and C_D .

The discussion above has not considered the influences of different work functions between metal and graphene, and the resultant inevitable doping of graphene^{39–41}. If these effects need to be reflected, V_D should be replaced by corresponding effective voltages: $V_{D,eff} = V_D - V_{D0}$. Here V_{D0} is drain voltage at the Dirac point without external voltages, whose role is very similar to that of the flat-band voltages in conventional MOS transistors. Hence electric fields E_D and E_S can be approximately written as

$$\begin{aligned}E_D &= \frac{V_{D,eff}(C_T - C_D)}{C_T d} - \frac{Q_{FG}}{C_T}, \\ E_D &= \frac{V_{D,eff}C_D}{C_T d} + \frac{Q_{FG}}{C_T}.\end{aligned}\quad (5)$$

here, ε_0 , ε_r are the vacuum permittivity and relative permittivity of HfO_2 , respectively, d is the thickness of the tunneling layer.

With the aid of equations (1–5), we can calculate time response of current density in programming (J_1 , J_2) and erasing (J_3 , J_4) modes. Parameters in the calculation are: $\varepsilon_r = 17^{42}$, $W_1 = 1$ mm, $W_2 = 1$ cm, $d = 4.5$ nm, $V_D = -5.5$ V and $+6$ V in programming and erasing modes. In Fig. 3(a), both source and drain are made of aluminum. As already explained, electrons can tunnel to graphene with negligible leakage to source ($J_1 \gg J_2$) in the programming mode. However, in the erasing mode the electron density in the FG layer keeps on growing as $J_4 \ll J_3$, and therefore they cannot be erased. In Fig. 3(b), Al and Au are used as drain and source electrodes, respectively. Because of their different work functions, the electron transportation between source and graphene is inappreciable in both programming and erasing modes. By reversing the polarity of the drain voltage, we can either increase or decrease the amount of electrons stored in FG layer for different modes.

A critical parameter which decides the device performance is the thickness d of the HfO_2 layer. We calculate time responses of tunneling current densities in programming mode, electrical field strength between drain and graphene, density of stored charge, i.e., J_1 , E_D , and Q_{FG} , in Fig. 4(a) and (b) with d as a variable. The drain voltage in the calculation is -5.5 V. It can be seen that reducing the thickness d could enhance the electrical E_D and consequently the amplitude of tunneling current density J_1 , which implies a reduced programming time for storing the same amount of electrons in graphene. However, reducing the thickness d would deteriorate the retention time, and even the programming process forbidden. As shown in Fig. 4(a), J_2 becomes equal to J_1 at $t = 0.5$ μs for $d = 3.5$ nm. Therefore, a very thin tunneling layer is not advisable. Based on the trade-off between the programming speed and the retention time, a 4.5-nm-thick HfO_2 is chosen as the tunneling layer.

In Fig. 4(c) and (d), we plot time responses of the device for different amplitudes of programming/erasing signals. Increasing the amplitude of the drain voltage can substantially shorten the pulse duration that is required to store or remove a certain amount of electrons. However, the maximum voltage is limited by the breakdown electrical field of HfO_2 (17 MV/cm in⁴²) and the stability of device. The drain voltages in the programming and erasing modes are finally set as $V_D = -5.5$ V and $+6$ V, respectively. The absolute amplitude of V_D in the erasing mode is higher and operating time is longer than those in programming mode. This is due to the fact that tunneling electrons experience a higher barrier height in the erasing mode as shown in Fig. 2(l). In order to equalize the programming and erasing times, the pulse amplitudes of the two operation modes should be adjusted accordingly. According to Eq. (1), the leakage current density from graphene to drain is on the order of $\sim 10^{-10}$ A/cm² provided that the charge density in graphene layer is $\sim 3 \times 10^{12}$ /cm². This order of magnitude is in accordance with the calculations in refs^{43,44}, and implies a very long retention time in theory. However, since insulating abilities of practical thin dielectric layers cannot be ideal, leakage currents in realistic nonvolatile devices are usually much larger than this theoretical estimation. Here RC time constant is not considered in calculation, since it is usually on the order of picoseconds (ps)^{5,7}.

Nonvolatile switching of THz signal. In this part, we study the switching characteristic and the dynamic response of this nonvolatile device. The charges stored in graphene (Q_{FG}) change the transmission of THz signal by varying its optical conductivity ($\sigma_g(\omega)$). Since charges can be stored for a long period before an opposite erasing voltage comes, the device hence acts as a nonvolatile THz switch. Optical absorption in graphene includes interband and intraband transitions. However, in THz range the intraband transition dominates, so its optical conductivity can be described by a Drude-like dispersion formula⁴⁵ $\sigma_g(\omega) = iqE_F/\pi\hbar^2(1 + \omega^2\tau^2)$, here τ is the carrier momentum scattering time. The Fermi level of graphene is related to the density of stored charges: $|E_F| = \hbar v_F\sqrt{\pi|Q_{FG}|}$, here $v_F = 10^6$ m/s is the Fermi velocity of Dirac fermions. The transmissivity of THz signal can be calculated as⁴⁶

$$T(\omega, n_s) = \frac{\sqrt{\varepsilon_2}}{\sqrt{\varepsilon_1}} \frac{4(\varepsilon_1\varepsilon_0)^2}{|(\sqrt{\varepsilon_1\varepsilon_2} + \varepsilon_1)\varepsilon_0 + \sqrt{\varepsilon_1}\text{Re}(\sigma_g(\omega))/c|^2}\quad (6)$$

here, ε_1 and ε_2 are the permittivities of covering layer and substrate, respectively.

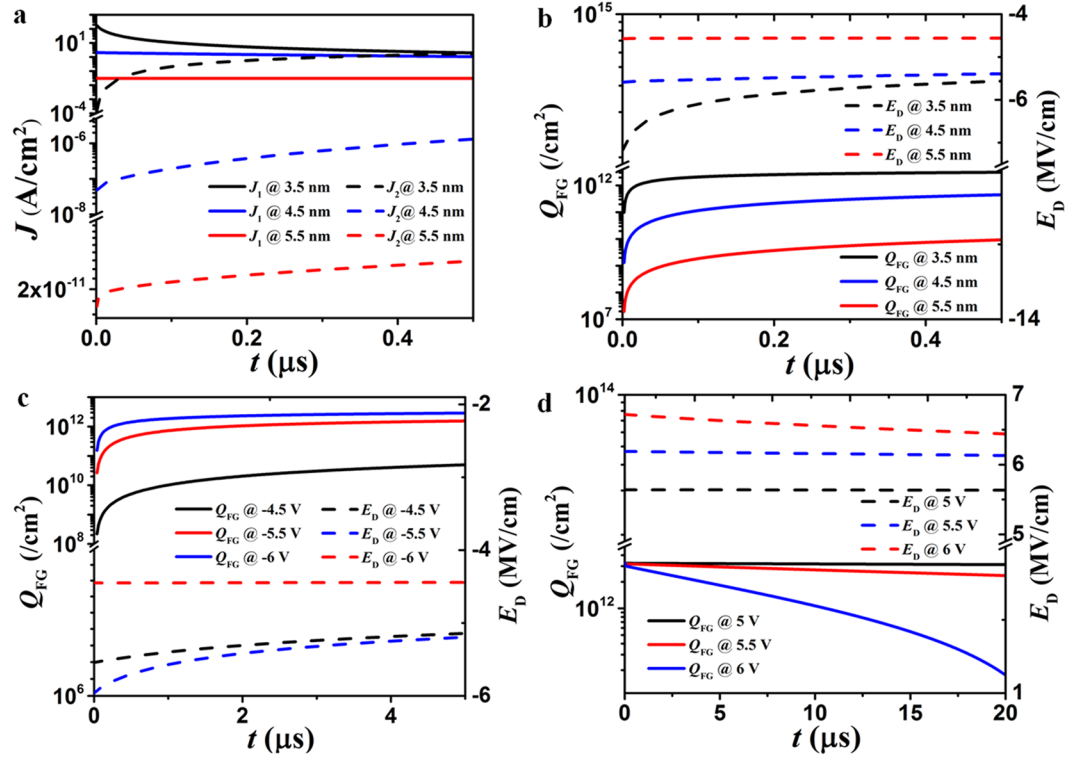


Figure 4. The processes of programming and erasing charges. The time responses of (a) tunneling current densities J_1 and J_2 , (b) electrical field strength between drain and graphene E_D , density of stored charge Q_{FG} for different thicknesses of tunneling layer d ; the time responses of Q_{FG} and E_D with different drain voltages V_D for programming mode (c) and erasing mode (d).

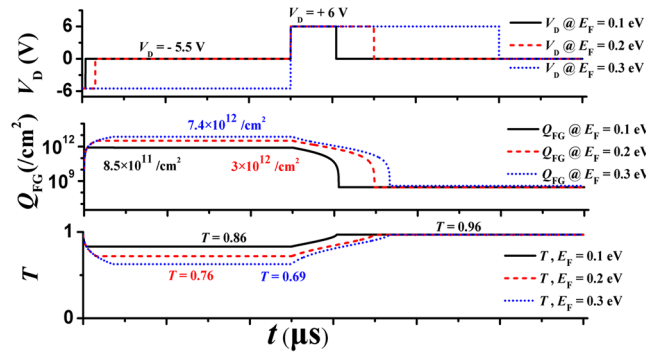


Figure 5. Waveforms of driving signals and the corresponding time responses of the device. E_F is tuned from 0 eV to 0.1 eV (black, solid), 0 eV to 0.2 eV (red, dash), and 0 eV to 0.3 eV (blue dot) by adding drain voltage pulses of different durations.

Discussion

According to the analysis above, the amplitudes of the programming/erasing (negative/positive) pulses are chosen to be -5.5 V and $+6$ V. Waveforms of the programming/erasing pulses applied on the drain and the corresponding responses of the device are shown in Fig. 5. With a 0.4 - μ s-long programming pulse, electrons in the FG layer finally reach a density of $\sim 8.5 \times 10^{11}$ /cm². E_F of graphene then rises by 0.1 eV. The modulation depth (MD), defined as $|T(n_s) - T(n_0)|/T(n_0)$, is calculated to be 9.5% at 5 THz according to equation (6). The corresponding erasing time is 4.5 μ s. If the programming and erasing times are extended to 2.4 μ s and 9.6 μ s, respectively, more electrons ($\sim 3 \times 10^{12}$ /cm²) are stored, and thus the MD is enhanced to 20% . By further prolonging the programming and erasing time (7.65 μ s and 12 μ s, respectively), Q_{FG} can reach to ($\sim 7.4 \times 10^{12}$ /cm²), meanwhile, its MD is increased to 27% .

In accordance with Figs 4 and 5, the MD (or stored charges Q_{FG}) relies on the duration time of the programming/erasing pulses as long as amplitudes of driving pulses are fixed. Enhancing the MD by using long pulses is in contradiction with improving the operation speed. In order to overcome this tradeoff, we can take

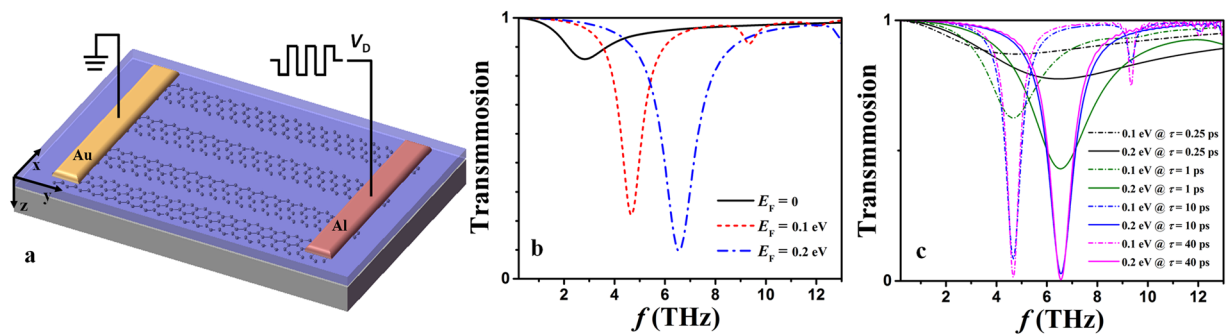


Figure 6. (a) Schematic diagram of graphene-ribbon-array based nonvolatile switch. (b) Transmission for $E_F = 0$ eV, 0.1 eV and 0.2 eV as a function of frequency. (c) Transmission for $E_F = 0.1$ eV and 0.2 eV as a function of τ .

advantage of graphene-based plasmons. As collective oscillations of charge carriers, plasmons can occur in patterned graphene⁴⁷. For example, the graphene-ribbon array shown in Fig. 6(a) supports highly confined plasmons which can resonantly couple with the incident light. The response is similar to the monolayer graphene sheet while the incident light polarized along the ribbon, therefore, a normally incident light polarized perpendicular to the ribbons is used to excite the surface Plasmon mode^{48–52}. The simulation result shows that this structure helps to enhance the MD remarkably. The modeled resonance curves are presented in Fig. 6(b) by using a three-dimensional finite-difference time domain method (Lumerical FDTD Solutions). The simulation condition is given in Methods. Figure 6(b) indicates that the MDs reach 80% (4.6 THz) and 89% (6.5 THz) by tuning E_F from 0 eV to 0.1 eV and to 0.2 eV, respectively. The power consumption is proportional to the total stored charges and hence the area of the floating-gate layer, the corresponding power consumptions per unit area are 0.0068 J/cm² and 0.0085 J/cm² to tune the Femi level from 0 to 0.1 eV (programming mode) and from 0.1 eV back to 0 eV (erasing mode), respectively. These values become 0.026 J/cm² and 0.033 J/cm² for tuning Femi level from 0 to 0.2 eV and from 0.2 to 0 eV, respectively. The simulation in Fig. 6(b) is based on a scattering time of $\tau = 3$ ps for the graphene on SiO₂ substrate according to data in⁵². It is observed that both impurities and phonons affect scattering rates of practical graphene layers^{53–55}, so the value of scattering time is not fixed but depends on the particular processing condition. To investigate the impact of the scattering time, we also carry out the simulation in Fig. 6(c) with scattering times of 0.25 ps, 1 ps, 10 ps and 40 ps according to the data in^{8,51}. It can be seen that a high-quality graphene layer of long scattering time is favorable to realize a narrow resonance curve with high resonance depth, since the conductivity of graphene is proportional to $i\tau^{-2}$.

In conclusion, we propose a graphene-based nonvolatile THz intensity switch with asymmetric electrodes. If a uniform graphene layer is used as the floating gate layer, the modulation depths are 9.5%/20% by nonvolatile tuning E_F of graphene between 0 and 0.1/0.2 eV with voltage pulses of appropriate amplitude and duration. Moreover, the MD can be further improved to 89% by patterning the graphene layer into a ribbon array. The performance of this configuration can be further improved by different manners. For example, since charged impurities in oxides introduce disorders of graphene, h-BN can be used to replace HfO₂ so as to increase the quality of graphene⁵⁶; using multi-layer graphene instead of graphene monolayer is able to enhance the effective refraction index variation¹⁹, and so on.

Methods

The numerical simulation of electrostatic potential distribution and the corresponding electrical distribution as shown in Fig. 1 is performed by the electrostatic module of commercial software (Comsol Multiphysics) based on finite element method (FEM). The free-triangular mesh is employed and the electrical properties of graphene are derived from refs^{8,9}. The width of electrodes (drain and source) is $W_1 = 1$ mm; and that of the floating-gate layer is $W_2 = 1$ cm. Since W_1 and W_2 are several orders of magnitude larger than the thickness of graphene, they are scaled down to 5 nm and 50 nm in the simulation for a better eye-capture.

The tunneling processes of electrons in programing/erasing modes are calculated by Matlab based on the model in refs^{31,34–41}; the results are shown in Figs 3 and 4. The effective drain voltage $V_{D,eff}$ is calculated via the equations in refs^{39–41}. The voltage at the Dirac point is normally related to work function difference between metal and graphene as well as interface trap states. In this study we assume the graphene is ideal, so only the effect of different work functions is considered. The transmissivities through the monolayer graphene in Fig. 5 are also calculated by Matlab based on equation (6). Three chemical potentials (Femi levels) are considered, i.e., 0, 0.1, 0.2 eV, in Fig. 5.

The numerical simulation of the THz transmissivity through graphene ribbons is performed with the 3-D finite-difference time domain method (Lumerical FDTD Solutions). The result is shown in Fig. 6. The graphene ribbons are oriented along the y direction. A plane-wave source is normally incident (z -direction) on graphene. The periodic boundary condition is used along the y direction, while the anti-symmetric boundary condition is used along the x direction. The perfectly matched layer (PML) is set at the z -direction. The graphene material type (temperature of 300 K) together with the 2D rectangle geometry described in modeling methodology are employed to model the graphene layer following the surface conductivity approach. The grid size for graphene layer is $\Delta x = \Delta y = 2.5$ nm and $\Delta z = 0.25$ nm.

References

- Siegel, P. H. Terahertz technology. *IEEE Trans. Microw. Theory Tech.* **50**, 910–928 (2002).
- Zhang, X. & Ferguson, B. Materials for terahertz science and technology. *Nat. Mater.* **1**, 26–33 (2002).
- Fan, Y. *et al.* An electromagnetic modulator based on electrically controllable metamaterial analogue to electromagnetically induced transparency. *Sci. Rep.* **7**, 40441 (2017).
- Islam, M., Rao, S., Kumar, G., Pal, B. & Chowdhury, D. Role of Resonance Modes on Terahertz Metamaterials based Thin Film Sensors. *Sci. Rep.* **7**, 1–8 (2017).
- Zide, J. M. O. *et al.* Active terahertz metamaterial devices. *Nature* **444**, 597–600 (2006).
- Sensale-Rodriguez, B., Yan, R., Liu, L., Jena, D. & Xing, H. G. Graphene for reconfigurable terahertz optoelectronics. *Proc. IEEE* **101**, 1705–1716 (2013).
- Liu, M., Yin, X. & Zhang, X. Double-layer graphene optical modulator. *Nano Lett.* **12**, 1482–1485 (2012).
- Otsuji, T., Popov, V. & Ryzhii, V. Active graphene plasmonics for terahertz device applications. *J. Phys. D: Appl. Phys.* **47** (2014).
- Lee, S. H. *et al.* Switching terahertz waves with gate-controlled active graphene metamaterials. *Nat. Mater.* **11**, 936–941 (2012).
- Ou, J., Plum, E., Zhang, J. & Zheludev, N. I. An electromechanically reconfigurable plasmonic metamaterial operating in the near-infrared. *Nat. Nanotech.* **8**, 252–255 (2013).
- Wang, Q. *et al.* Optically reconfigurable metasurfaces and photonic devices based on phase change materials. *Nat. Photon.* **10** (2016).
- Driscoll, T. *et al.* Memory Metamaterials. *Science* **325**, 1518–1521 (2009).
- Pernice, W. H. P. & Bhaskaran, H. Photonic non-volatile memories using phase change materials. *Appl. Phys. Lett.* **101**, 171101 (2012).
- Song, J. *et al.* Integrated photonics with programmable non-volatile memory. *Sci. Rep.* **6**, 22616 (2016).
- Song, J. *et al.* Silicon photonic integrated circuits with electrically programmable non-volatile memory functions. *Opt. Express* **24**, 21744–21751 (2016).
- Lee, J. S. Review paper: Nano-floating gate memory devices. *Electron. Mater. Lett.* **7**, 175–183 (2011).
- Ji, Y. *et al.* Organic nonvolatile memory devices with charge trapping multilayer graphene film. *Nat. Nanotechnol.* **23**, 105202 (2012).
- Zheng, Y. *et al.* Gate-controlled nonvolatile graphene-ferroelectric memory. *Appl. Phys. Lett.* **94**, 163505–163505-3 (2009).
- Hong, A. J. *et al.* Graphene Flash Memory. *ACS Nano* **5**, 7812–7817 (2011).
- Vu, Q. *et al.* Two-terminal floating-gate memory with van der Waals heterostructures for ultrahigh on/off ratio. *Nat. Commun.* **7**, 12725 (2016).
- Liu, M. *et al.* A graphene-based broadband optical modulator. *Nature* **474**, 64–67 (2011).
- Midrio, M. *et al.* Graphene-based optical phase modulation of waveguide transverse electric modes. *Photonics Res.* **2**, A34–A40 (2014).
- Lao, J. *et al.* Tunable graphene-based plasmonic waveguides: nano modulators and nano attenuators. *Laser & Photonics Rev.* **8**, 569–574 (2014).
- Ansell, D. *et al.* Hybrid graphene plasmonic waveguide modulators. *Nat. Commun.* **6**, 8846 (2015).
- Svintsov, D., Vyurkov, V., Ryzhii, V. & Otsuji, T. Voltage-controlled surface plasmon-polaritons in double graphene layer structures. *Journal of Appl. Phys.* **113**, 338–344 (2013).
- Wilk, G. D., Wallace, R. M. & Anthony, J. M. High-k gate dielectrics: Current status and materials properties considerations. *J. Appl. Phys.* **89**, 5243–5275 (2001).
- Haynes, W. M. *Handbook of Chemistry and Physics*. 94th Ed, (CRC Press, Boca Raton, 2013).
- Mueller, T., Xia, F. & Avouris, P. Graphene photodetectors for high-speed optical communications. *Nat. photon.* **4**, 297–301 (2010).
- Feenstra, R. M., Jena, D. & Gu, G. Single-particle tunneling in doped graphene-insulator-graphene junctions. *Journal of Appl. Phys.* **111** (2012).
- Tersoff, J. & Hamann, D. R. Theory of the scanning tunneling microscope. *Phys. Rev. B: Condens. Matter* **31**, 805 (1985).
- Fowler, R. H. & Nordheim, L. Electron Emission in Intense Electric Fields. Proceedings of the Royal Society of London **119**, 173–181 (1928).
- Seike, K. *et al.* Floating-gated memory based on carbon nanotube field-effect transistors with Si floating dots. *Japanese Journal of Appl. Phys.* **53** (2014).
- Liu, Z. *et al.* Metal nanocrystal memories - Part I: Device design and fabrication. *IEEE Trans. Electron Devices* **49**, 1606–1613 (2002).
- An, Y., Shekhawat, A., Behnam, A., Pop, E. & Ural, A. Gate tunneling current and quantum capacitance in metal-oxide-semiconductor devices with graphene gate electrodes. *Appl. Phys. Lett.* **109** (2016).
- Mao, L. F. *et al.* The Gate Leakage Current in Graphene Field-Effect Transistor. *IEEE Electr. Device Lett.* **29**, 1047–1049 (2008).
- Simmons, J. G. Generalized formula for the electric tunnel effect between similar electrodes separated by a thin insulating film. *J. Appl. Phys.* **34**, 1793–1803 (1963).
- Lenzlinger, M. & Snow, E. H. Fowler-Nordheim tunneling into thermally grown SiO₂. *J. Appl. Phys.* **40**, 278–283 (1969).
- Fang, T., Jena, D., Konar, A. & Xing, H. *et al.* Carrier Statistics and Quantum Capacitance of Graphene Sheets and Ribbons. *Appl. Phys. Lett.* **91**, 092109–092109-3 (2007).
- Xu, H. *et al.* Top-gated graphene field-effect transistors with high normalized transconductance and designable Dirac point voltage. *ACS Nano* **5**, 5031–5037 (2011).
- Schaefer, J., Thiele, S. & Schwierz, F. Modeling of graphene metal-oxide-semiconductor field-effect transistors with gapless large-area graphene channels. *J. Appl. Phys.* **107**, 094505 (2010).
- Burghard, M., Balasubramanian, K., Weitz, R. T., Lee, E. J. H. & Kern, K. Contact and edge effects in graphene devices. *Nat. Nanotech.* **3**, 486–490 (2008).
- Sire, C., Blonkowski, S., Gordon, M. & Baron, T. Statistics of electrical breakdown field in HfO₂ and SiO₂ films from millimeter to nanometer length scales. *Appl. Phys. Lett.* **91**, 242905 (2007).
- Lee, J. J., Wang, X., Bai, W., Lu, N. & Kwong, D. Theoretical and experimental investigation of Si nanocrystal memory device with HfO₂ high-k tunneling dielectric. *IEEE Trans. Electron Devices* **50**, 2067–2072 (2003).
- Bertolazzi, S., Krasnozhan, D. & Kis, A. Nonvolatile memory cells based on MoS₂/graphene heterostructures. *ACS Nano* **7**, 3246–3252 (2013).
- Gao, W. *et al.* High-contrast terahertz wave modulation by gated graphene enhanced by extraordinary transmission through ring apertures. *Nano Lett.* **14**, 1242–1248 (2014).
- Dong, H. M., Xu, W., Zhang, J. & Yuan, Y. Z. Terahertz absorption window and high transmission in graphene bilayer. *Opt. Commun.* **283**, 3695–3697 (2010).
- Christensen, J. *et al.* Graphene Plasmon Wave guiding and Hybridization in Individual and Paired Nanoribbons. *ACS Nano* **6**, 431–440 (2012).
- Fan, Y. *et al.* Tunable terahertz coherent perfect absorption in a monolayer graphene. *Opt. Lett.* **39**, 6269–6272 (2014).
- Fan, Y. *et al.* Tunable mid-infrared coherent perfect absorption in a graphene meta-surface. *Sci. Rep.* **5**, 13956 (2015).
- Fan, Y. *et al.* Electrically Tunable Goos-Hänchen Effect with Graphene in the Terahertz Regime. *Adv. Opt. Mater.* **4**, 1824–1828 (2016).
- Ju, L. *et al.* Graphene plasmonics for tunable terahertz metamaterials. *Nat. Nanotech.* **6**, 630–634 (2011).
- Nikitin, A. Y. *et al.* Surface plasmon enhanced absorption and suppressed transmission in periodic arrays of graphene ribbons. *Phys. Rev. B Condens. Matter Mater. Phys.* **85** (2012).

53. Vasko, F.T. & Ryzhii, V. Voltage and temperature dependencies of conductivity in gated graphene. *Phys. Rev. B: Condens. Matter* **76** (2007).
54. Bolotin, K. I. *et al.* Temperature-dependent transport in suspended graphene. *Phys. Rev. Lett.* **101** (2008).
55. Adam, S. *et al.* Theory of charged impurity scattering in two-dimensional graphene. *Solid State Commun.* **149**, 1072–1079 (2009).
56. Lee, G. H. *et al.* Electron tunneling through atomically flat and ultrathin hexagonal boron nitride. *Appl. Phys. Lett.* **99**, 243114 (2011).

Acknowledgements

This work is supported by the Nature Basic Research Program of China (2013CB632105), the Zhejiang Provincial Natural Science Foundation for Distinguished Young Scholar (LR15F050002), and the National Natural Science Foundation of China (61675184).

Author Contributions

L.Y. designed the structure and wrote the manuscript. H.Y. and X.Y.Q. discussed about the mechanism principle. T.G.D. and Q.Z. discussed about the theoretical analyses and calculations. J.F.J. and G.C.W. provided the excitation of graphene-ribbon resonance in the simulation. Y.L.Q., J.Y.Y. and X.Q.J. helped to plot the figures. H.Y. supervised the project. All authors discussed the results and commented on the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-20047-3>.

Competing Interests: The authors declare that they have no competing interests.

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



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

© The Author(s) 2018