Dynamics of PSG-Based Nanosecond Protonic Programmable Resistors for Analog Deep Learning

M. Onen, J. Li, B. Yildiz, and J. A. del Alamo
Massachusetts Institute of Technology, Cambridge, MA 02139, USA, email: monen@mit.edu

2022 IEEE International Electron Devices Meeting (IEDM) | 978-1-6654-8959-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/IEDM45625.2022.10019365

***Abstract*—** We study the dynamics of a new class of three-terminal programmable resistors based on proton intercalation into a metal oxide channel using PSG as electrolyte for energy-efficient analog deep-learning hardware accelerators. These CMOS- and BEOL-compatible nanoscale devices exhibit excellent characteristics in terms of high operation speed (5 ns/pulse), high energy efficiency (~fJ/pulse), many (1000) nonvolatile conductance states across a large dynamic range (10×), linear and symmetric modulation, and high endurance [1]. This work presents the first real-time study of the dynamics of these devices evidencing impulse-like non-volatile conductance modulation with nanosecond-range pulses without any visible equilibration dynamics. Our study also reveals that the bottleneck to channel conductance modulation is proton transfer across the PSG/WO_3 interface and not proton transport across the PSG.

I. Introduction

Increasing workloads of deep learning problems and a slowdown in transistor scaling progress have fueled interest in energy-efficient analog hardware accelerators in recent years [2]. The unit element of an analog processor is a programmable nonvolatile resistor (also referred to as non-volatile memory, memristor, or crosspoint element). Given that computations in the analog domain rely on intrinsic physical properties, these devices need to satisfy strict requirements to avoid accuracy degradation [3]. Interest in engineering such devices for analog deep learning applications has skyrocketed in the last decade [4]. Many of these attempts have focused on repurposing mature memory technologies. Unfortunately, these efforts have come short as deep learning requires efficient processing of information much more importantly than its storage. Therefore, devices need to be primarily optimized for their state modulation properties rather than state preservation [1].

Recently, a novel device family relying on electrochemically controlled ion intercalation in transition-metal oxide channels (also referred to as ECRAM) has shown promising characteristics (**Fig. 1**) [5-12]. Following mature-but-CMOS-incompatible Li^+ ion devices [5, 6] and compatible-but-slow O^{2-} based alternatives [7-9], H^+ (proton)-based variants have come into focus. Protons are attractive because their small radius and light mass promise high speed and energy-efficient operation. However, the CMOS-incompatible nature of conventional proton electrolytes (e.g., liquid, organic, polymeric) has critically limited their application [10, 11]. Recently, we reported a new class of protonic devices using phosphosilicate glass (PSG) as the electrolyte, an eminently CMOS-compatible material [12]. PSG-based nanoscale devices have shown ultrafast modulation characteristics with outstanding energy efficiency under extreme electric fields (**Fig. 2**) [1].

This work investigates the real-time dynamics of conductance modulation of PSG-based nanoscale protonic programmable resistors. We demonstrate intriguing dynamics previously invisible to post-facto measurements (i.e., reading channel conductance much later than the programming pulse). We identify non-volatile intercalation into the channel in the nanosecond range in addition to a volatile proton-enhanced field effect. Our study suggests that proton crossing the PSG/WO_3 interface is the bottleneck to channel conductance modulation.

II. Methods

The self-aligned 3-terminal protonic programmable resistors investigated in this work consist of a Pd reservoir (also acts as the gate electrode), a PSG electrolyte, and a WO_3 channel [1, 12] (**Fig. 3**). Nanometer-scale devices were fabricated (**Fig. 4**) using a series of electron beam lithography steps [1]. The thickness of the PSG layer was varied in different samples (d_{PSG} =5, 10, and 20 nm), whereas the WO_3 layer was kept constant at 10 nm (**Fig. 5**).

Unless indicated, devices were tested under a low vacuum, following their exposure to forming gas (FG, 3% H_2 in N_2) for protonation of the stack. Waveforms were acquired at 80 MHz bandwidth except for **Fig. 7** which was recorded at 170 MHz. To maximize the bandwidth of the signals, devices with a high W/L ratio (≥10) for the active region were selected in all experiments.

III. Results

Fig. 6 illustrates incremental conductance modulation of a nanoscale protonic device (d_{PSG}=10 nm) after individual 100 ns programming pulses with V_{GS}=10 V and V_{DS}=0.5 V. Each colored segment represents source current, I_S, recorded separately with ~10 s in between. This indicates that the channel conductance perfectly retained its state between successive acquisitions. Averaging of I_S values (black line on left) shows well-defined discrete steps of linear channel conductance modulation.

It is important to establish that not only do the devices respond to short pulses, but also complete the modulation without any extended equilibration period following the application of the pulse. **Fig. 7A** shows the time evolution of I_S during the application of a negative pulse train of 5 ns each at V_{GS}=−12 V. Note that a negative channel bias of V_{DS} =−0.5 V is used for all negative pulses to limit $|V_{GD}|$. These results evidence that conductance modulation occurs in an impulse-like fashion (within 170 MHz instrument bandwidth) which is highly desirable for analog deep learning applications. In addition, the modulation of channel conductance was perfectly retained over 10^4× the programming pulse length (**Fig. 7B**).

We have carried out a detailed study of the time evolution of the channel conductance during programming (**Fig. 8**). For this, we used trapezoidal pulses of 1 μs duration with 1 μs rise and fall times

(**Fig. 8, right**). In general, we expect a channel terminal current response that consists of four components: 1) direct gate leakage current, 2) displacement current associated with the rising and falling edges of the pulse, 3) volatile field-effect channel current since WO_3 is a semiconductor, and 4) non-volatile channel current due to proton insertion. We show below that the gate leakage is negligible. To isolate the remaining three components, we first recorded the channel current I_S for a device *before* exposure to FG (red trace). In this case, in response to a V_{GS}=3 V pulse, the waveform shows negligible I_S during the pulse but prominent current during the transients which can be attributed to displacement current. The negligible field-effect current reflects an unprotonated E-mode device. Without mobile protons within the device, intercalation is absent, as expected.

Application of an identical pulse on the same device *after* FG treatment (blue trace) still does not result in proton intercalation in the channel (no residual increase in I_S after the pulse). On the other hand, there is a clear additional channel current (blue trace) riding on top of the displacement currents observed in air (red trace). Importantly, this current grows during the rising edge, remains constant during the pulse plateau, and gradually disappears during the falling edge, leaving no permanent change behind (i.e., volatile). This extra current can then be associated with an enhancement of the electron concentration and therefore channel current due to the field-effect action of protons migrating within the PSG layer closer to the WO_3 interface under the influence of an electric field. We label this as a field-effect channel current due to a proton-induced negative threshold voltage shift.

Non-volatile proton intercalation is ultimately seen when a higher voltage pulse (4.5 V, green trace) is applied to the FG-treated device, with a distinct current rise during the pulse plateau as well as a permanent current enhancement after the pulse ends. Similar behavior for negative voltage pulses is shown in **Fig. 9**.

The absence of direct gate leakage current and the separation between gate stack displacement current and channel current can be confirmed by performing experiments on a protonated device at V_{DS}=0.5 V and V_{DS}=0 (**Fig. 10**). At V_{DS}=0, the static source current (about half of $-I_G=(I_S+I_D)$) is negligible thanks to the electronic insulation properties of PSG [1,12].

We have carried out similar experiments with pulses of different voltage amplitudes before (**Fig. 11A**) and after FG exposure (**Fig. 11B**). For FG-treated devices, the field-effect current is found to be linear with the gate voltage (**Fig. 11C**). This is more clearly seen in **Fig. 12** which shows the time evolution of *the difference* of I_S produced by FG exposure (i.e., inclusion of free protons in the device stack) for the same device, vs. $V_{GS}(t)$. Using the net current ($\Delta I_S(t) = I_S^{\text{After FG}}(t) - I_S^{\text{Before FG}}(t)$) for analysis effectively eliminates the displacement and parasitic currents. For low pulse voltages, the $\Delta I_S(t)$ is purely the field-effect current, whereas increased voltages lead to non-volatile ion intercalation, which manifests itself as hysteretic incomplete loops.

To better capture the voltage dependence and time evolution of intercalation during programming, we used pulses with a longer duration (100 µs). **Fig. 13** shows real-time proton intercalation into the channel via the rising channel current during the pulse. Unlike the field-effect (which remains constant over extended pulse duration once the protonic interface capacitance is charged at the programming voltage), the intercalation effect linearly increases with time as more protons cross the PSG/WO_3 interface. The linearity of the modulation indicates a diluted proton concentration in the WO_3 channel material [11, 12].

Repeating this experiment with different pulse voltage shows a highly nonlinear voltage dependence for intercalation (**Fig. 14**). In fact, the voltage dependence of the nonvolatile channel conductance change is exponential (**Fig. 15**). This is consistent with Butler-Volmer charge transfer kinetics, where ion insertion occurs by field-induced lowering of a relatively high energy barrier ($\gg$ kT) [13]. To test this hypothesis, we repeated the same experiments in devices with different PSG thickness showing an exponential dependence of non-volatile conductance modulation on the electric field across the gate stack for all devices (**Fig. 16**).

IV. Discussion

Our research reveals that the field-effect channel current linearly follows the gate voltage in time indicating fast proton motion though the PSG and charging/discharging of the protonic interface capacitance. This conclusion also holds at low voltages where the field acceleration of proton hopping should diminish. We conclude from this that proton conductivity of the PSG electrolyte does not represent a bottleneck for proton intercalation.

In contrast, protons can only cross the PSG-WO_3 interface under the presence of a high field, as evidenced by a clear threshold behavior **(Fig. 15)**. This high nonlinearity also provides a ratchet effect, ensuring protons stay in the channel after the programming pulse comes to an end. Suitable PSG-WO_3 interface engineering should speed up proton transfer and lower the programming voltage. This might come with a retention trade-off which is less critical for analog deep learning applications.

V. Conclusions

In this work we analyzed the dynamics of channel conductance modulation of PSG-based nanoscale protonic programmable resistors. Results not only definitively validate previous post-facto measurements of nanosecond operation, but also reveal three distinct components of channel current during the application of a pulse: displacement, field effect, and ion intercalation. Our study provides guidance on how to reduce operating voltage.

Acknowledgment

Work was supported by MIT-IBM Watson AI Lab and MIT Quest for Intelligence. We thank Baoming Wang and Frances M. Ross at MIT for TEM imaging.

References

[1] M. Onen, et. al., *Science*, In Press. [2] T. Gokmen & Y. Vlasov, *Front. Neurosci*. **10**, (2016). [3] M. Onen et. al., *Front. Artif. Intell.,* **5**, (2022). [4] Z. Wang, et. al., *Nat. Nano.*, **15**, (2020). [5] E. J. Fuller, *Adv. Mater.*, **29**, (2017). [6] J. Tang et. al., IEDM (2018). [7] S. Kim, et. al., IEDM (2019). [8] Y. Li, et. al., *Adv. Mater.* **32** (2020). [9] P. Solomon, et. al., IRPS, (2021). [10] Y. van de Burgt et.al., *Nat. Mater.,* **16**, (2017). [11] X. Yao, et. al., *Nat. Comm.*, **11**, (2020). [12] M. Onen, et. al., *Nano Lett.*, **21**, (2021). [13] A. Röthel, et. al., *Z. Phys. Chem.,* **224**, (2010).

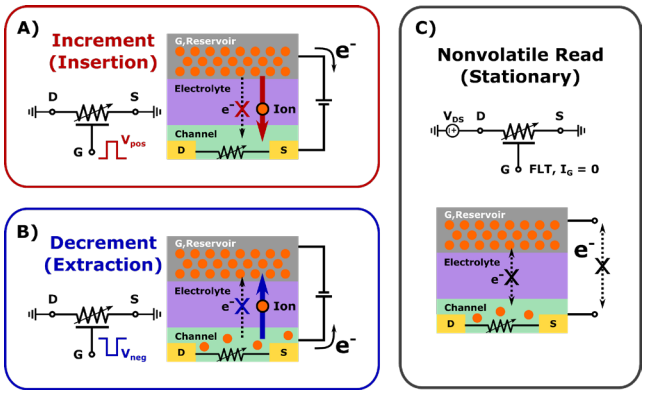

Fig. 1. Operation of programmable resistor based on ion intercalation: A) Positive voltage pulses shuttle protons into the channel material to increase its conductivity. B) Negative pulses reverse this. C) In the absence of a programming voltage the channel retains its conductance state.

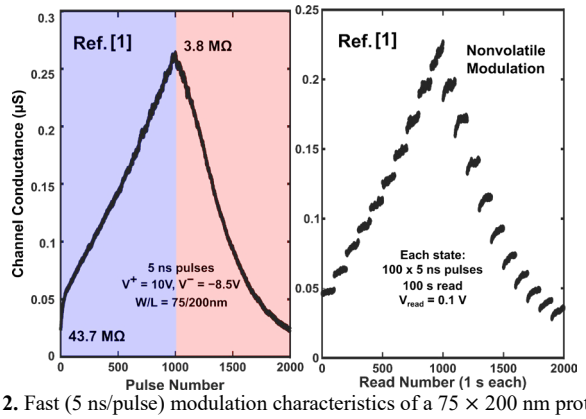

Fig. 2. Fast (5 ns/pulse) modulation characteristics of a 75 × 200 nm protonic programmable resistor with d_{PSG}= 10 nm showing nearly linear, and symmetric characteristics with good retention (100 s) over a wide dynamic range (~10×) [1].

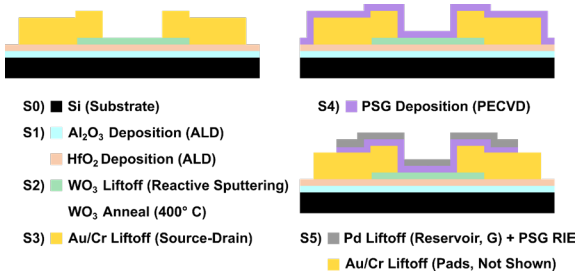

Fig. 3. Process flow of protonic programmable resistors fabricated using a series of self-aligned electron-beam lithography steps.

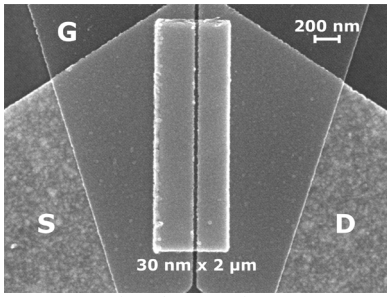

Fig. 4. Scanning electron microscope (SEM) image of a 30 × 2000 nm device.

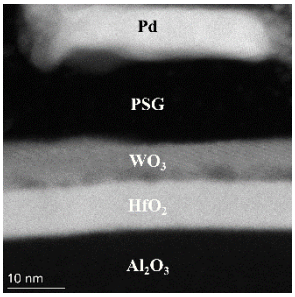

Fig. 5. Surface transmission electron microscope image of device cross section.

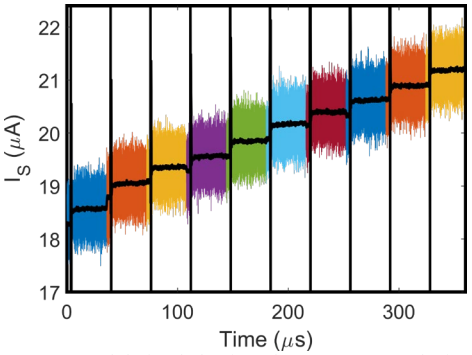

Fig. 6. Modulation behavior of a 75 × 750 nm device with d_{PSG}= 10 nm for 10 V, 100 ns gate pulses under V_{DS}=0.5 V.

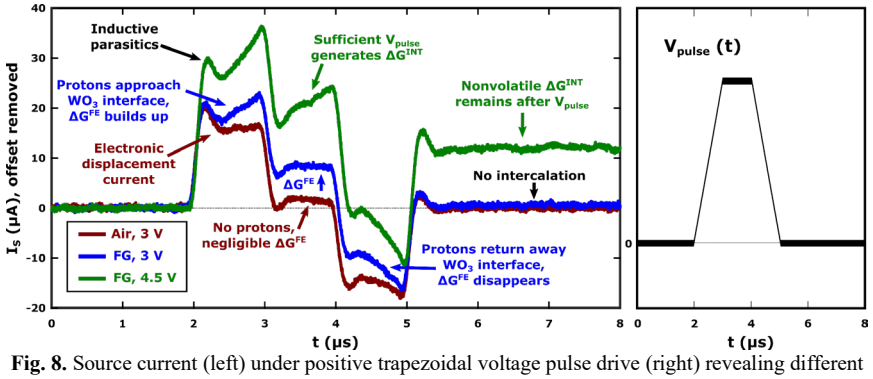

Fig. 8. Source current (left) under positive trapezoidal voltage pulse drive (right) revealing different components of conductance modulation for a device with d_{PSG}=5 nm under V_{DS}=0.5 V. Negligible ΔG^{FE} in the absence of protons (i.e., *before* FG treatment). *After* FG treatment (blue, green), a volatile ΔG^{FE} is observed. When V_{pulse} is sufficiently high (green), a nonvolatile ΔG^{INT} is generated.

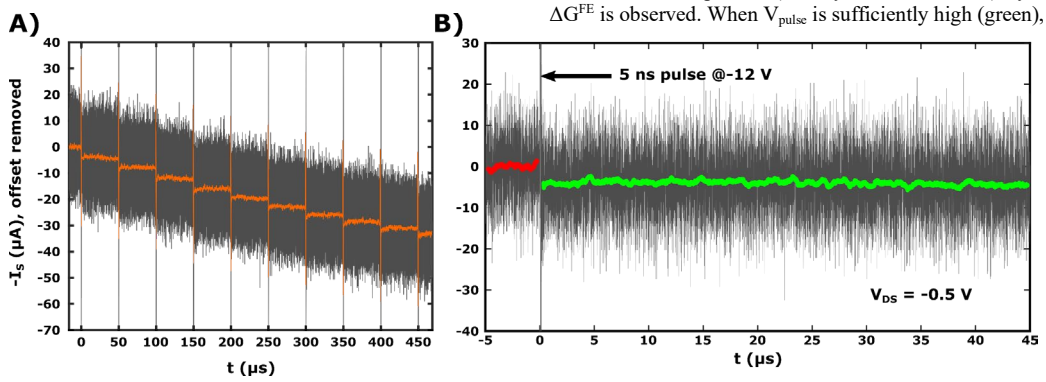

Fig. 7. Modulation behavior of a 30 × 5000 nm device with d_{PSG}=10 nm for -12 V, 5 ns gate pulses under V_{DS}=-0.5 V captured at 170 MHz. (A) Linear, gradual deprotonation of the channel over 10 individual pulses. (B) I_S(t) before, during, and after the first pulse on a finer time scale, clearly displaying impulse-like fast modulation characteristics.

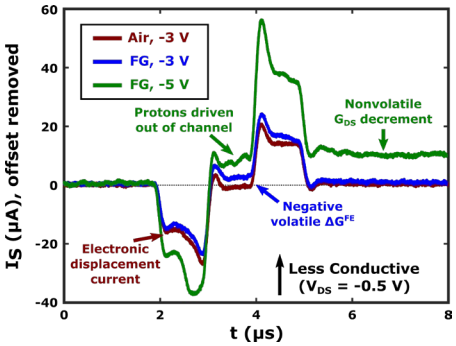

Fig. 9. Components of conductance modulation of the device in Fig. 8 during the application of negative voltage pulses for V_{DS} = -0.5 V. Color coding as in Fig. 8.

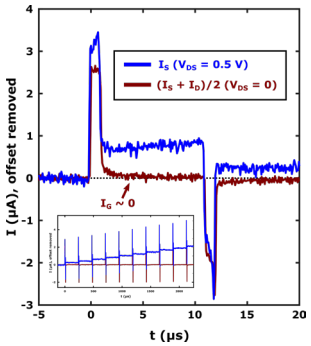

Fig. 10. Current waveforms of device in Fig. 8 (FG-treated) under V_{pulse}=5 V for V_{DS}=0 and V_{DS}= 0.5V. Inset shows 10 pulses.

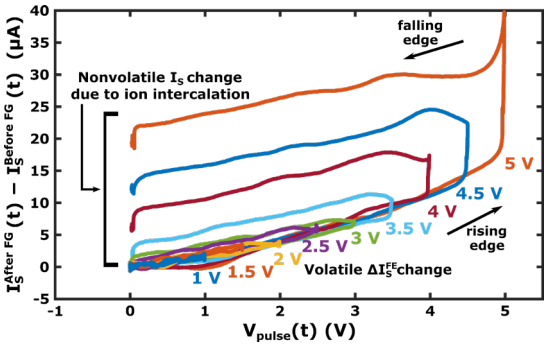

Fig. 12. Difference in I_S for identical pulses due to FG exposure for the device in Fig. 8 as vs. $V_{pulse}(t)$. Higher voltage pulses (>3 V) yield intercalation while others show volatile field effect only. For V_{pulse}>3V, $I_S^{Before\ FG}(t)$ is extrapolated from Fig. 11A.

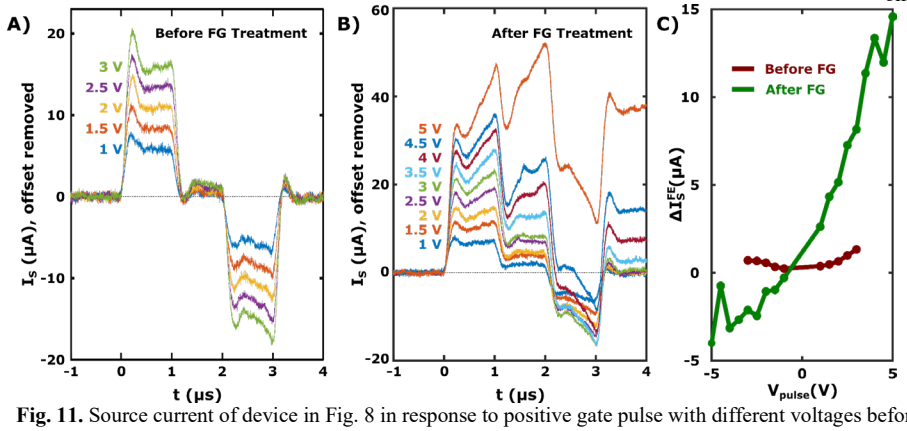

Fig. 11. Source current of device in Fig. 8 in response to positive gate pulse with different voltages before (A) and after (B) FG treatment. For FG-treated devices V_{pulse} >3 V result in intercalation while lower pulses cause volatile field effect only. High voltages prior to FG treatment were avoided as they were found to damage the stack. (C) Voltage dependence of volatile field-effect current, appearing after FG-treatment.

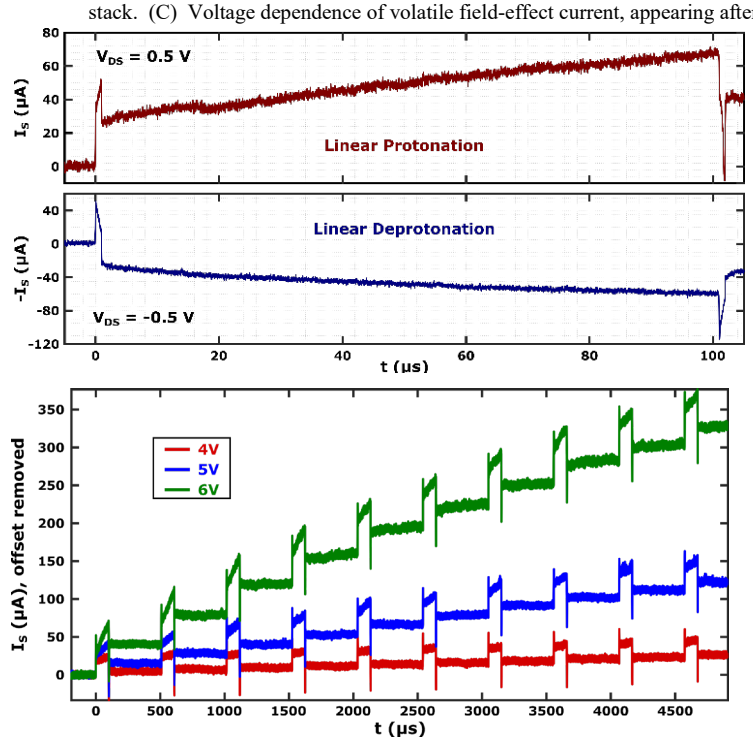

Fig. 13. Time evolution of protonation and deprotonation of a 75 × 750 nm device with d_{PSG} = 5 nm over the course of a long (100 μs) 4 V pulse. The waveform shows linear protonation and deprotonation characteristics throughout.

Fig. 14. Channel current of a 75 × 750 nm device with d_{PSG}=10 nm under different pulse voltages for 10 consecutive pulses (V_{DS}=0.5 V) showing nonlinear dependence between intercalation and pulse amplitude.

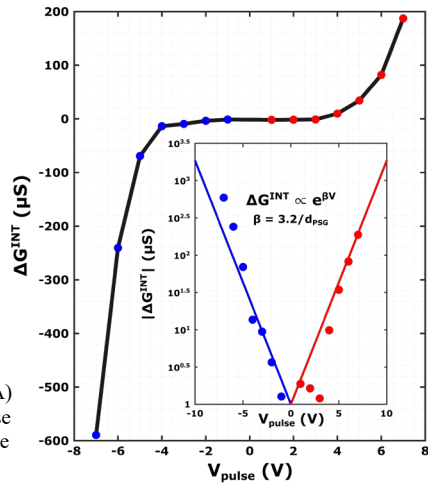

Fig. 15. Voltage dependence of nonvolatile channel conductance modulation (ΔG^{INT}) for device in Fig. 13 under 100 μs pulses.

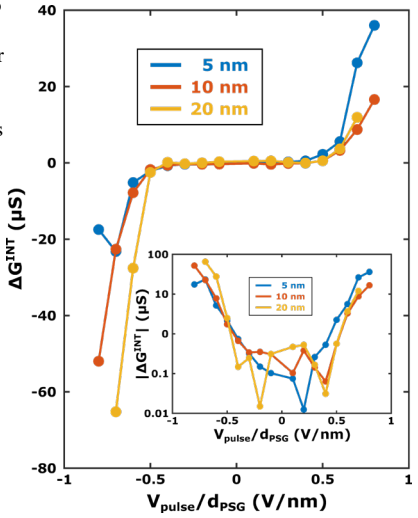

Fig. 16. Electric field dependence of nonvolatile channel conductance modulation (ΔG^{INT}) for 75 × 750 nm devices with different PSG thicknesses (5, 10, 20 nm) under 100 μs pulses.