# In-Situ Hydrogenated CMOS-Compatible $WO_3$/Pd/PSG/$WO_3$ ECRAM

Dingyu Shen[1*], Paul M. Solomon[2], Guy Cohen[2], John Rozen[2] and Jesús A. del Alamo[1]

[1]*Microsystems Technology Laboratories, MIT, Cambridge, MA 02139, USA.*
[2]*IBM Research, T.J. Watson Research Center, Yorktown Heights, NY 10598, USA*
*Sponsorship: MIT-IBM Watson AI Lab.*

**Introduction:** To overcome the computational bottlenecks of large AI models, analog deep learning accelerators process information locally using special-purpose devices for matrix multiplication calculations and outer product updates in the analog domain. Among device candidates, Electrochemical Random-Access Memories (ECRAMs) modulate the resistance of a semiconductor channel through ionic exchange with a reservoir via an electrolyte. Ion intercalation is a non-volatile process with deterministic and reversible conductance potentiation and depression characteristics that makes ECRAMs most promising for neural network training with enhanced energy efficiency, non-volatility, and low latency.
Our previous work [1,2] has demonstrated nanoscale protonic ECRAM devices enabling channel conductance modulation over a 20× range with nano-second operation. Our recent research is a close collaboration between MIT and IBM Research focused on developing a CMOS-compatible ECRAM process based on in-situ hydrogenated $H_xWO_3$ channel, $PdH_y$ gate reservoir, and phosphosilicate glass (PSG) electrolyte (Fig. 1a,b), which enables control of baseline hydrogen level and thereby device conductance as part of its fabrication.
**Methods**: The fabrication process for the ECRAM devices investigated here is fully back-end-of-line (BEOL) CMOS compatible. The process begins with a 200-mm wafer with blanket 400°C annealed $WO_3$ layer prepared at IBM. All layers of the stack are defined by reactive ion etching (RIE). In-situ hydrogenation is achieved by low-power hydrogen plasma in a plasma-enhanced chemical vapor deposition (PECVD) reactor, immediately followed by deposition of PSG in the same chamber to prevent hydrogen loss. A self-aligned design is achieved by additional hydrogenation of gate and extrinsic channel regions performed after gate definition. The double-layer gate design (1 nm Pd + 15 nm $WO_3$) minimizes stress induced by Pd hydrogenation while maintaining a high-volume hydrogen reservoir with high interface diffusivity. $Si_3N_4$ encapsulation prevents H loss. Pulse characterization was performed in air on instrumentation with 80 MHz bandwidth.
**Results:** Our devices demonstrate fast, linear and symmetric channel conductance modulation with good endurance under sub-µs voltage pulses (Fig. 1c,d). Fig. 2a shows very low and reversible gate leakage over a wide conductance dynamic range. Conductance range and device scale can be further tailored based on the application. Fig. 2b shows exponential $V_{pulse}$ dependence and power-law $t_{pulse}$ dependence of conductance response, as expected theoretically [1]. Fig. 2c shows $I_s$ response to a trapezoidal pulse that includes contributions from displacement current, semiconductor field-effect current, and ion intercalation current [3]. Figs. 2d,e show $I_s$ after removing the displacement current as in [4], for protonation and deprotonation pulses with different $V_{pulse}$. Good retention is shown with a channel conductance drift smaller than 10% after 100 µs. Current contributions from field effect and ion intercalation are symmetric under positive and negative pulses with the intercalation effect being dominant (Fig. 2f). This indicates strong potential for analog deep learning training.

**References**
[1] M. Onen, et al. Science, 2022, 377, pp. 539-543. [2] M. Onen, et al. Nano Letters, 2021. 21(14): p. 6111-6116.
[3] M. Onen, et al. 2022 IEDM, pp. 2.6.1-2.6.4. [4] P.M. Solomon, et al. 2021 IRPS, pp. 1-7.

* Corresponding author: email: deanshen@mit.edu

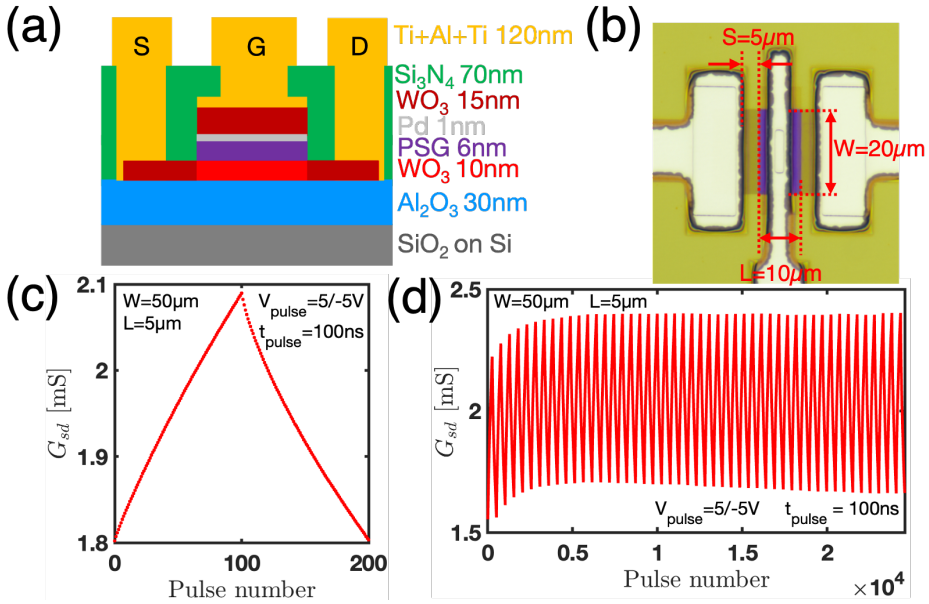


**Fig. 1.** (a) $WO_3$/Pd/PSG/$WO_3$ protonic ECRAM structure. Dark (light) red regions indicate high (low) hydrogen doping level in $WO_3$ through hydrogen plasma treatment. (b) Microscopic photo of a fabricated ECRAM device. W, width; L, length. (c) Voltage pulse modulation performance, showing fast (100 ns, 5 V), linear and symmetric characteristics. $WO_3$ base conductance can be tuned in the range of 5 μS to 2.5 mS by $H_2$ plasma treatment. (d) Endurance characterization of the same device with pulse parameters as in (c), displaying stable modulation characteristics over $2\times10^4$ pulses after shifting to the symmetry point. Each cycle has 256 up and 256 down pulses.

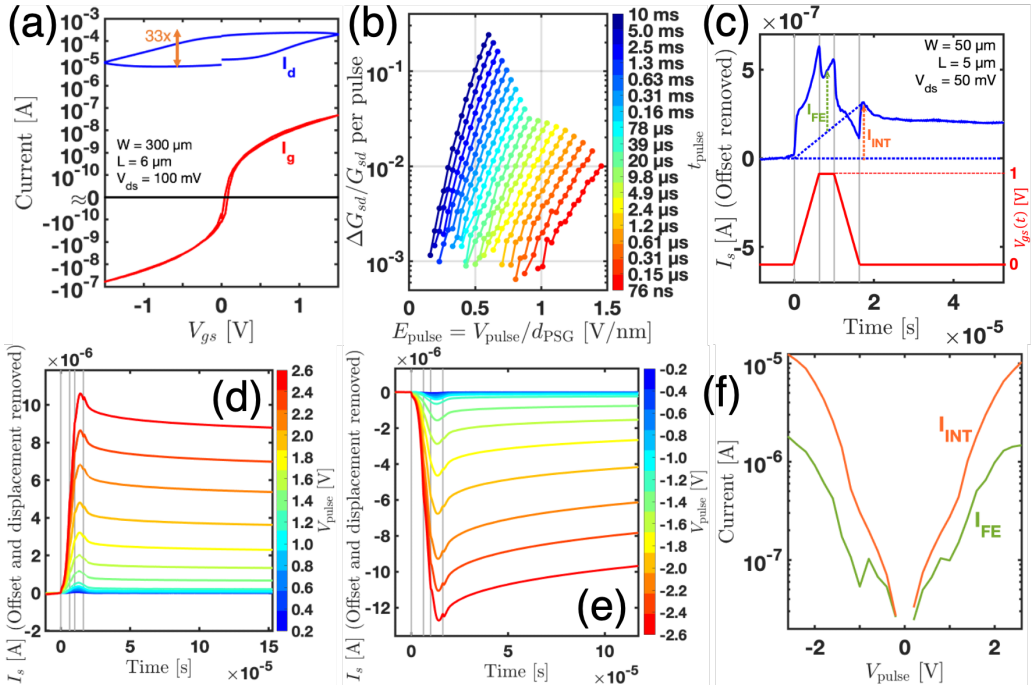


**Fig. 2.** (a) $I_d$, $I_g$–$V_{gs}$ double sweep showing dynamic range (~33×) and gate leakage. Source is grounded and $V_{ds}$ = 100 mV. (b) Relative change in channel conductance per pulse for different electric field ($V_{pulse}/d_{PSG}$) and pulse width. (c to f) are measured from a 50-μm-by-5-μm device by averaging on 1M alternate direction pulses with transimpedance amplifier cancelling constant offset current of 50 μA. (c) Source current (left axis) under positive trapezoidal voltage pulse drive (right axis). After removing displacement current, the $I_s$ dependence on different pulse voltages are shown in (d) (positive) and (e) (negative). (f) Field-effect current ($I_{FE}$) and intercalation current ($I_{INT}$) defined in (c) extracted from (d and e), showing exponential dependence on pulse voltage.