CMOS-compatible ferroelectric synapse technology for analog neural networks

Yanjie Shao, Elham R. Borujeny, Jorge Navarro, John C. Huang, Tyra E. Espedal, Dimitri A. Antoniadis, and Jesús A. del Alamo

Sponsorship: Intel through AI HW program; SRC GRC program

MIT AI Hardware Program Symposium

May 1, 2024

Modern AI burning too much energy…

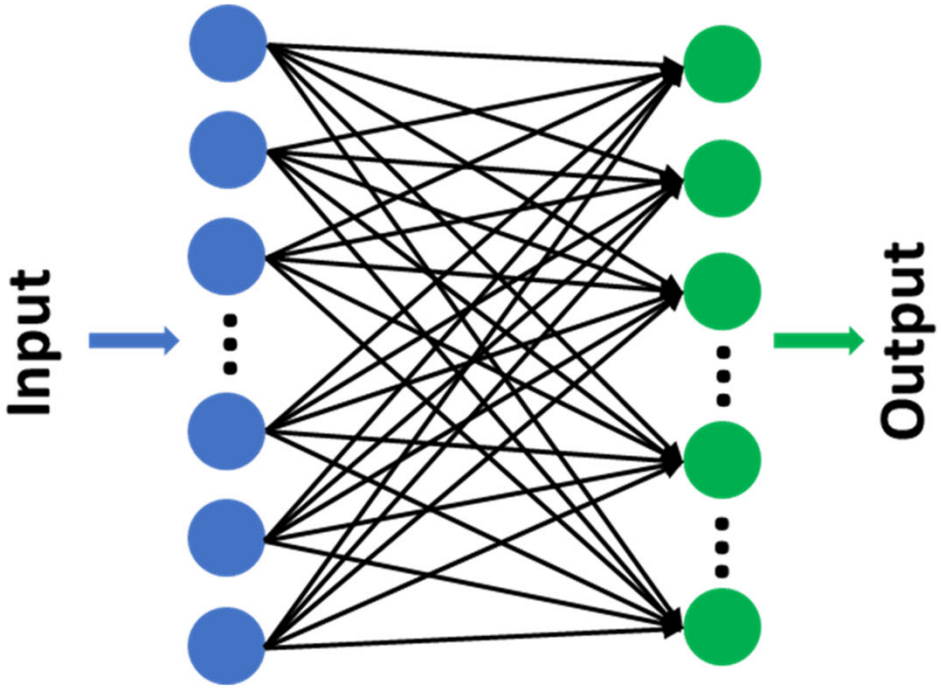
WIRED

WILL KNIGHT BUSINESS JAN 21, 2020 7:00 AM

AI Can Do Great Things—if It Doesn't Burn the Planet

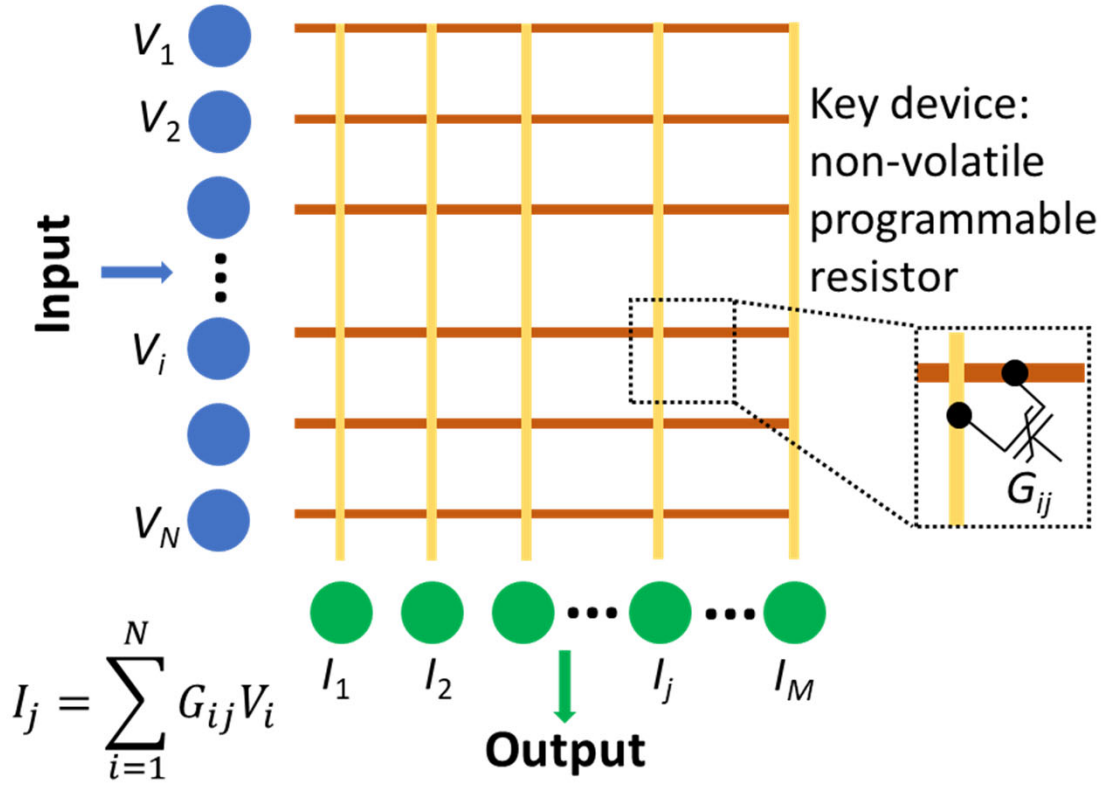
The computing power required for AI landmarks, such as recognizing images and defeating humans at Go, increased 300,000-fold from 2012 to 2018.

Analog hardware for AI



Artificial neural network (ANN)

Hardware implementation

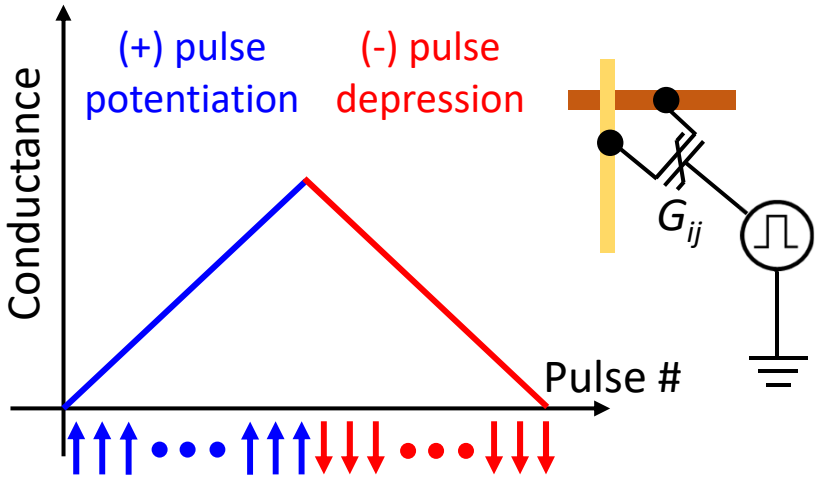


Crossbar array with analog synapse

- Multiply-Accumulate (MAC) operation at heart of Matrix Multiplication
- Analog vs. Digital NNs: Footprint ↓, Energy efficiency ↑↑, Latency ↓

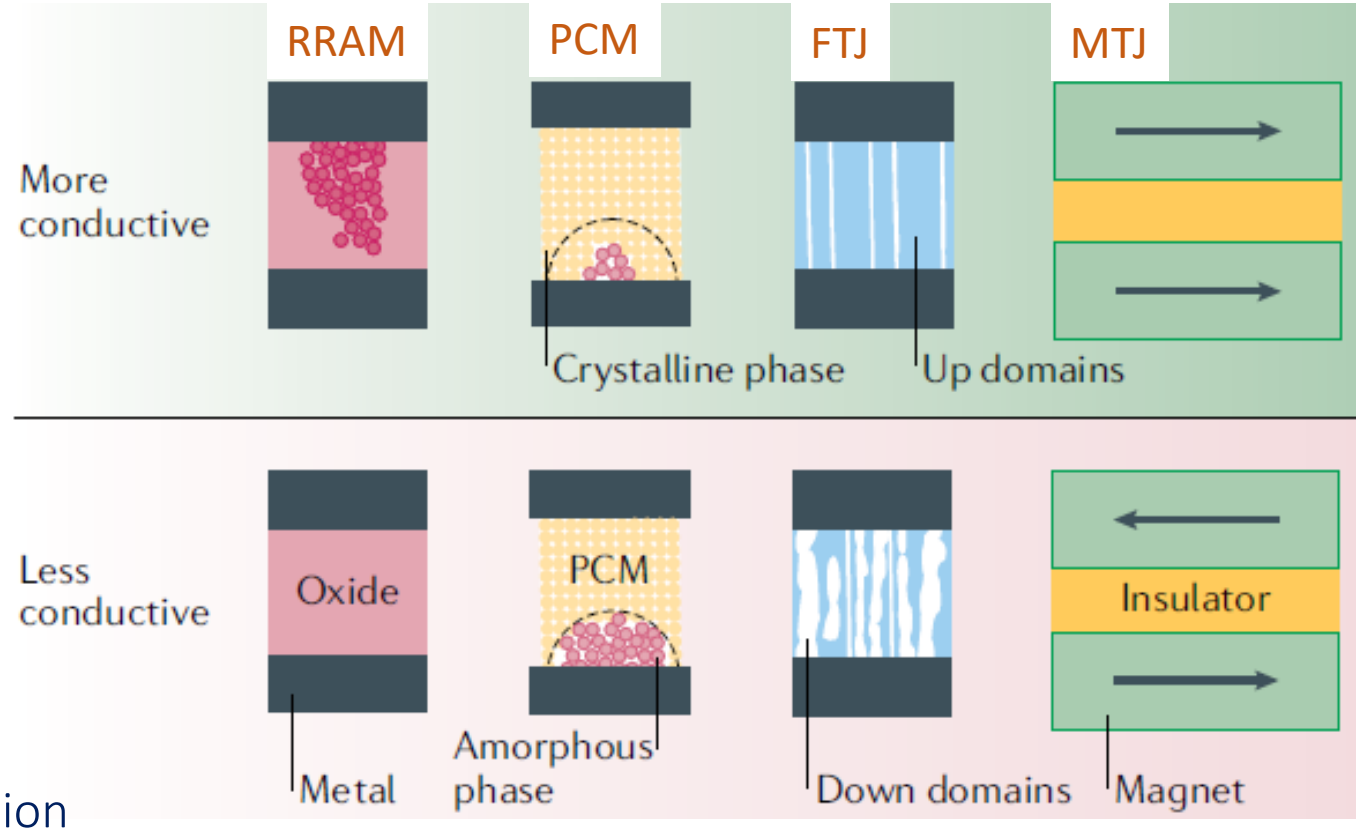
Requirement of programmable resistor

Ideal characteristics:



- G~0.1 μS
- High speed (ns)
- High efficiency (sub-pJ)
- Large dynamic range (~10x)
- Fine-tunable (~1000 levels)
- Low variability
- Symmetric modulation
- High endurance
- Good retention
- CMOS compatible

T. Gokmen and Y. Vlasov, *Frontiers in Neuroscience* 10, 333 (2016)



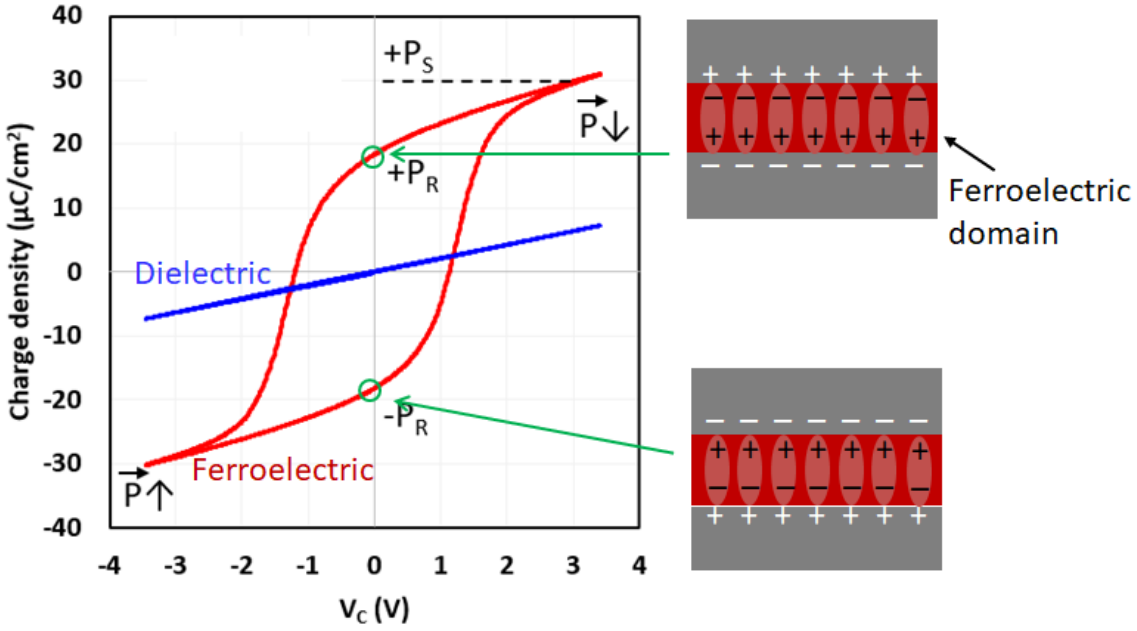
D. Marković et al., *Nature Reviews Physics* 2, 499–510 (2020)

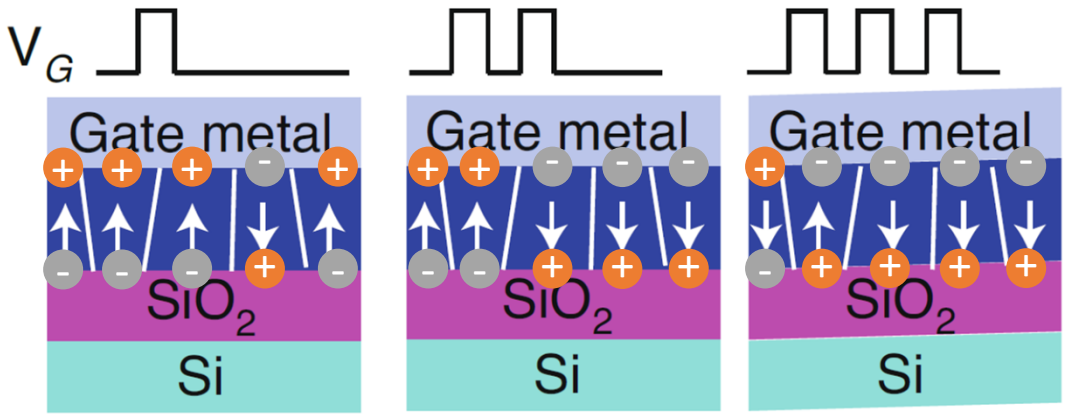
Device with desired characteristics does not exist today!

Programmable resistor based on ferroelectric effect

Ferroelectric (FE) $Hf_xZr_{1-x}O_2$ (HZO):

- stores polarization charge due to unique crystal structure
- CMOS compatibility





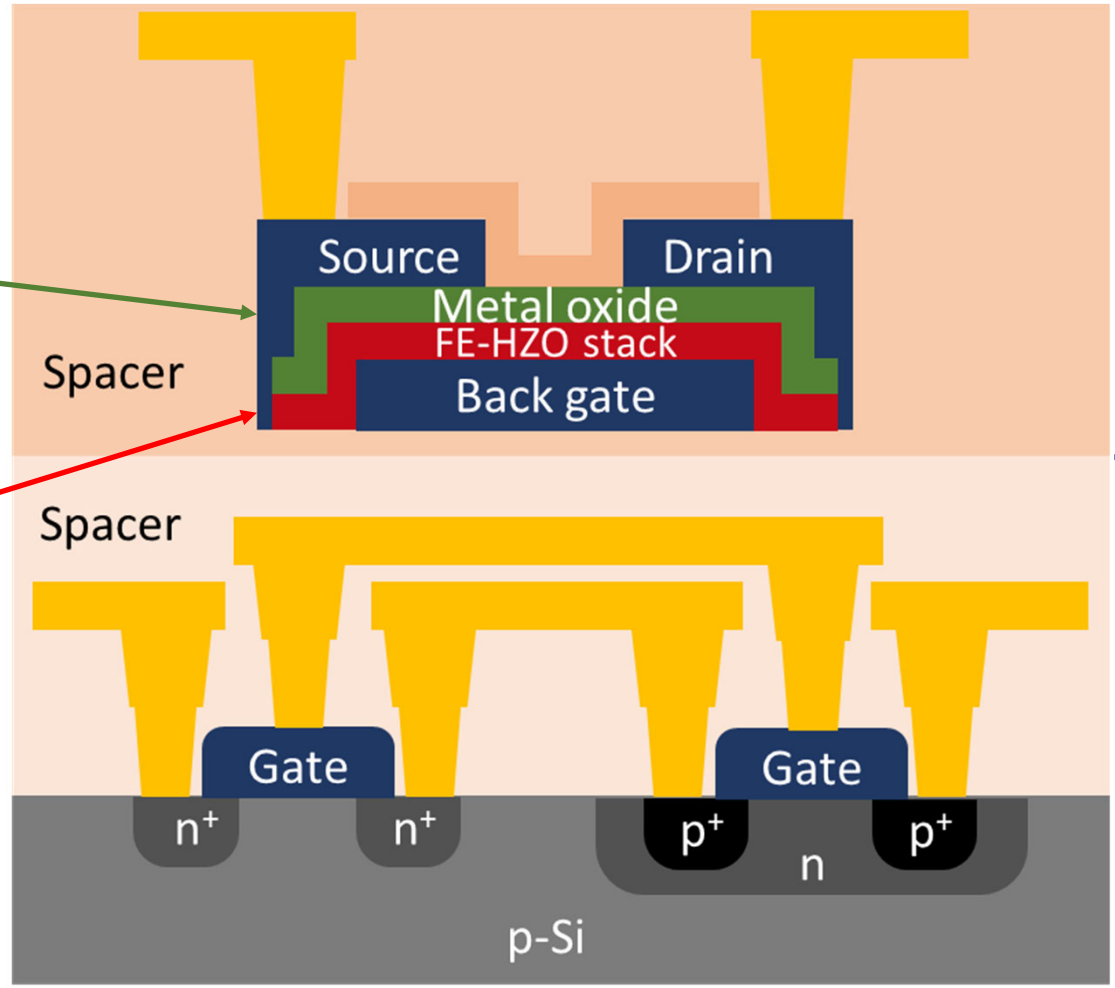
A. Khan *et al.*, *Nature Electronics* 3, 588-597 (2020).

Multi-domain structure: partial polarization switching → analog behavior

Our approach: non-volatile ferroelectric TFT

InSnO channel
(room temperature)

FE-HZO gate stack
(T ~ 400 °C)

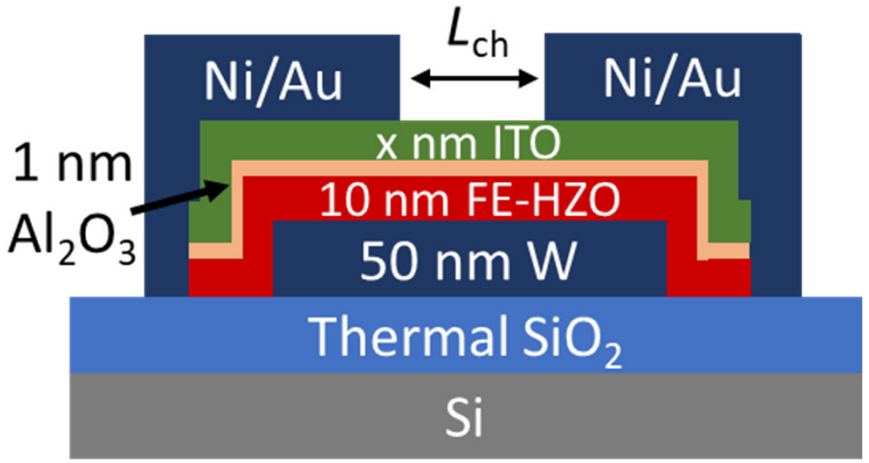


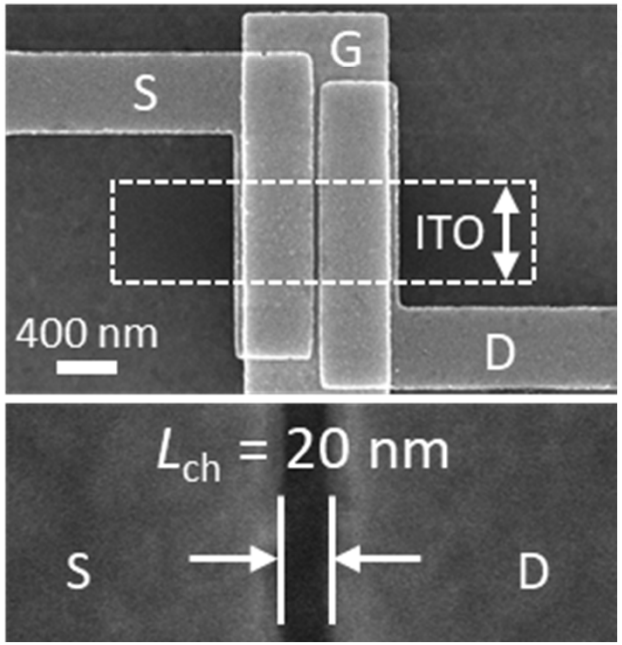
Back-end of the line (BEOL),
Analog FE-FET

Front-end of the line (FEOL),
Si-CMOS

Critical requirement: BEOL-compatibility → metal oxide + HZO thin-film transistor (TFT)!

BEOL-compatible FE-FET fabrication



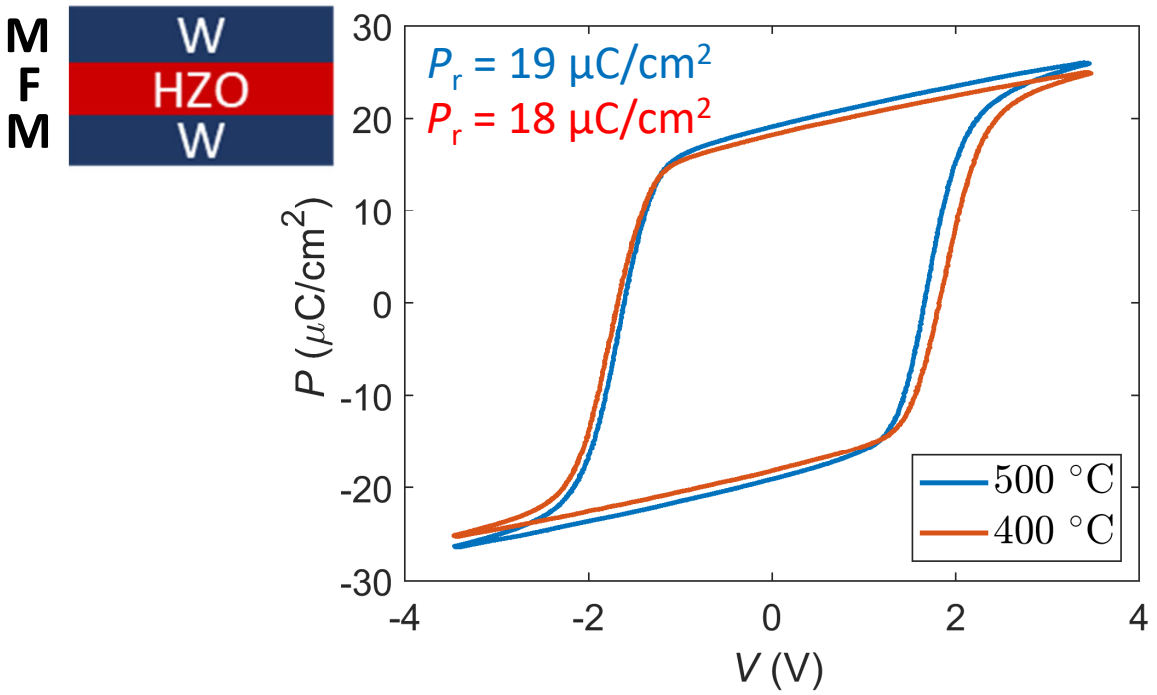


- 50 nm W sputtering
- W local back gate patterning
- 10nm HZO/1nm Al_2O_3 with PEALD
- RTA @ T = 400 °C, 1 min
- Gate via opening
- x nm ITO sputtering (x = 2.5, 5, 10 nm)
- Mesa patterning
- Ni/Au contact deposition
- Probe-pad fabrication

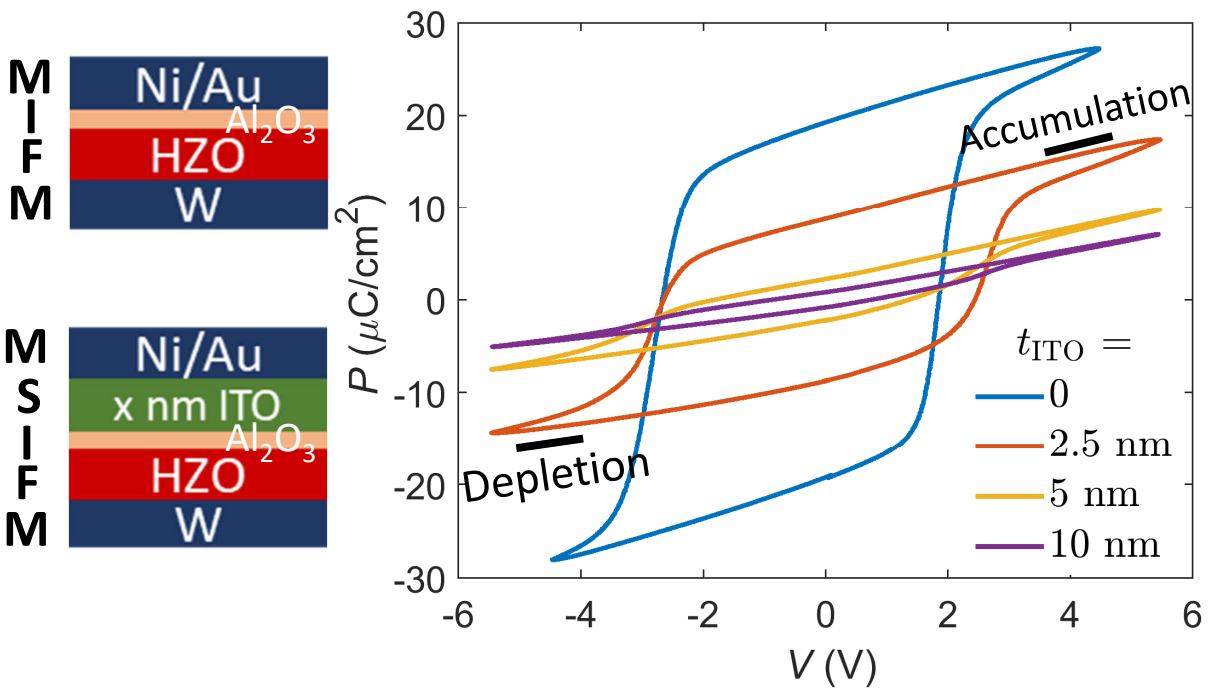
- CMOS-compatible thermal budget
- Highly-scaled device geometry

Scaling channel thickness for improved performance (I)

MFM

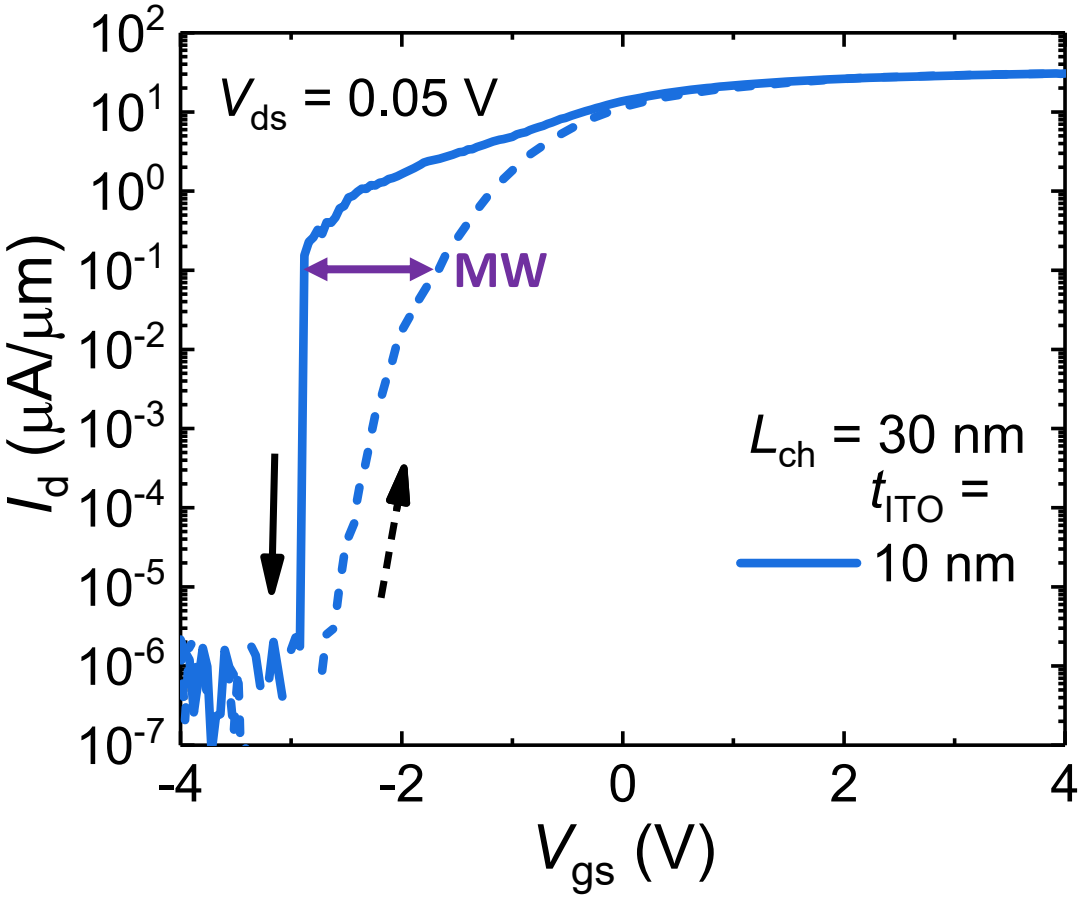


MFIM + MFISM



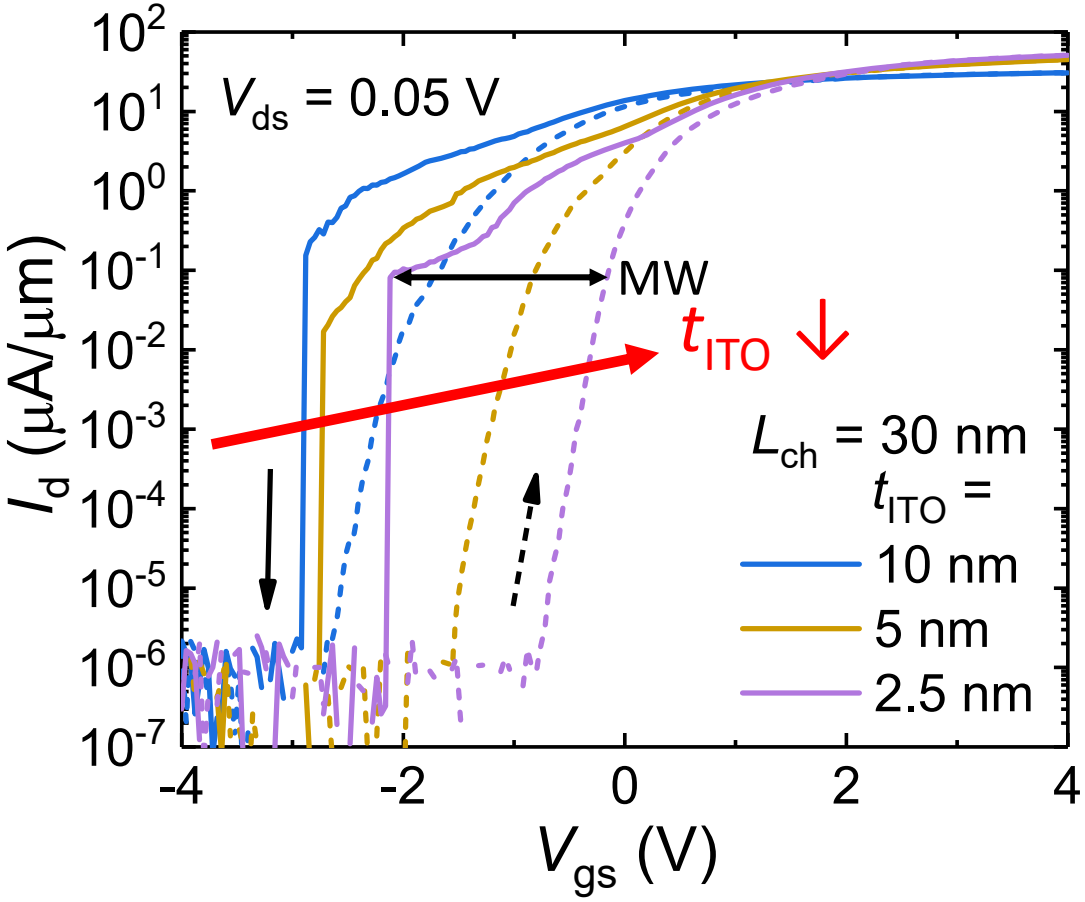
- Nice FE property with ≤ 400 °C thermal budget
- P_r ↑ as t_{ITO} ↓, influence of depletion layer

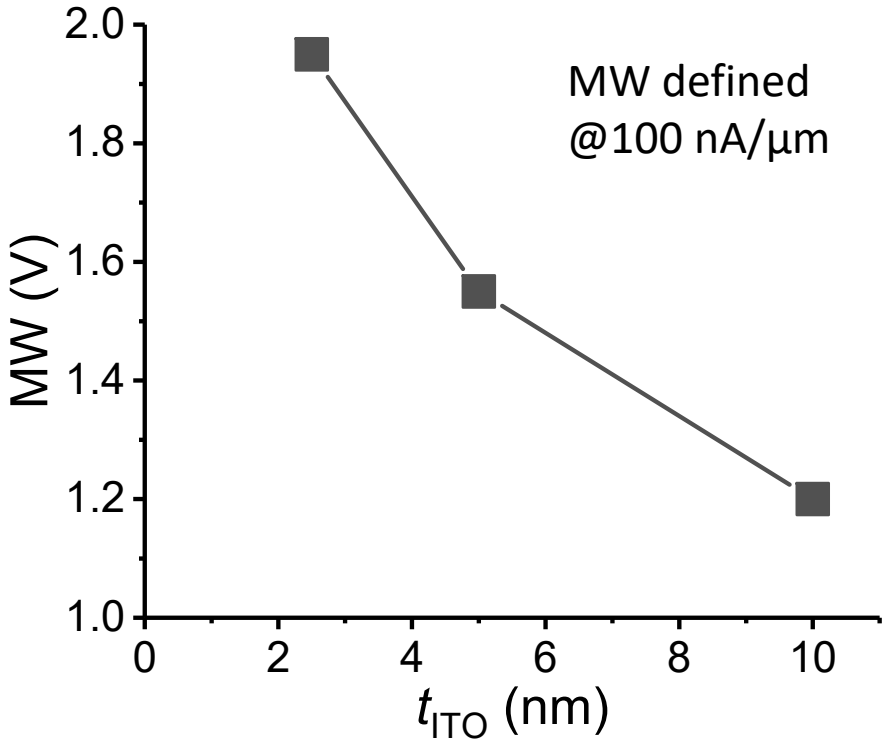
Scaling channel thickness for improved performance (II)



- Counter-clockwise hysteresis loop due to FE switching
- Memory window as a key metric

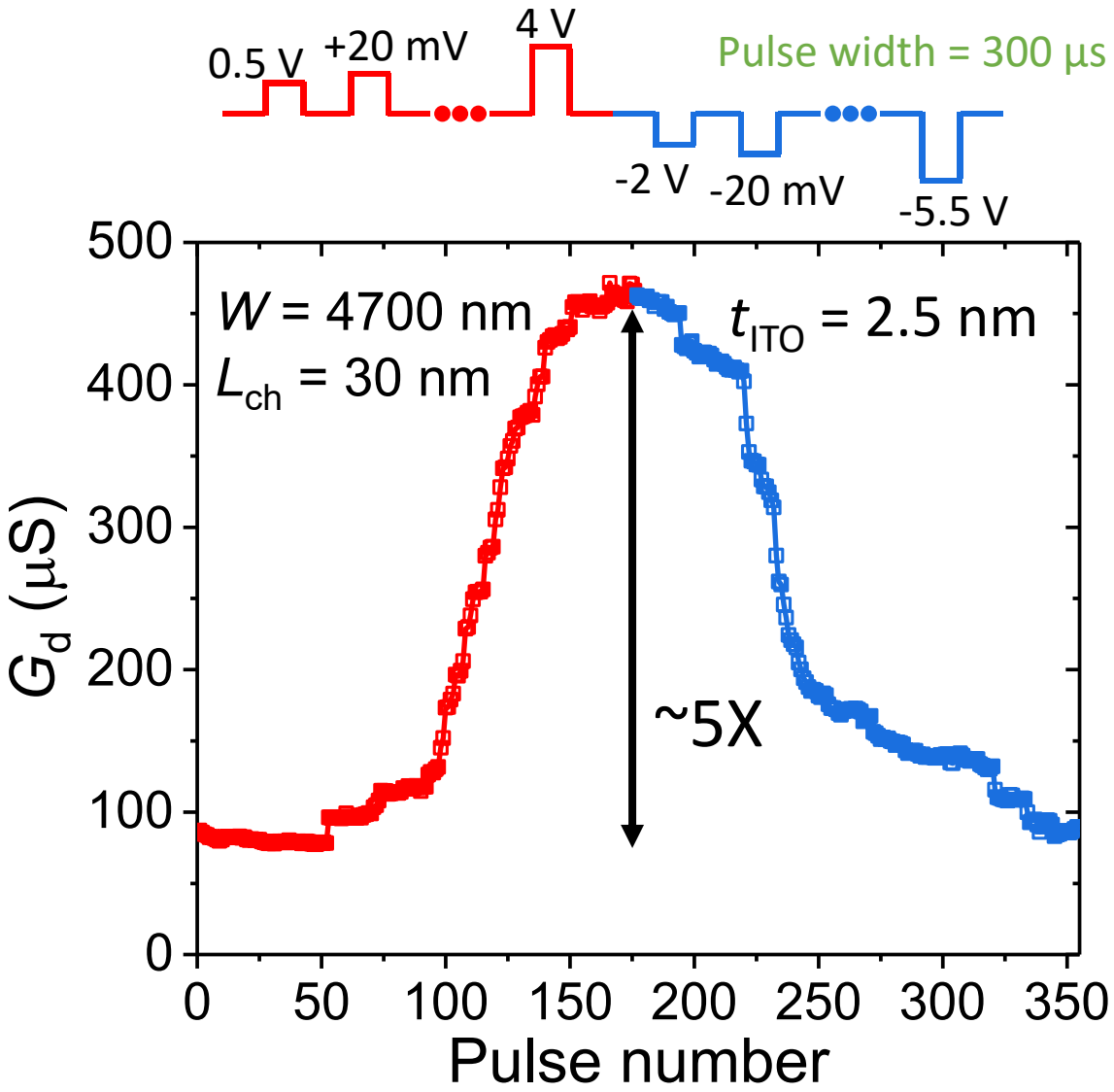
Scaling channel thickness for improved performance (III)





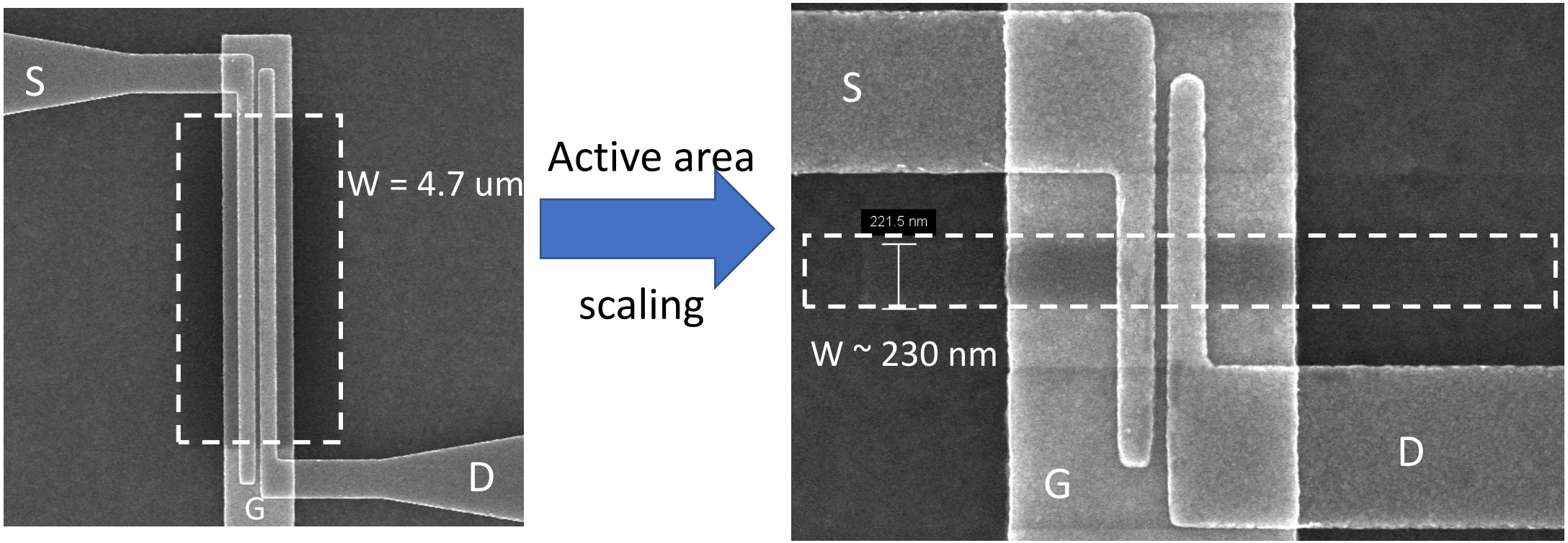
- Memory window (MW) ↑ and $\Delta V_t > 0$ as t_{ITO} ↓
- More possible states with larger MW
- D-mode → larger conductance dynamic range with more positive V_t

Conductance switching in FE-FETs: wide channel



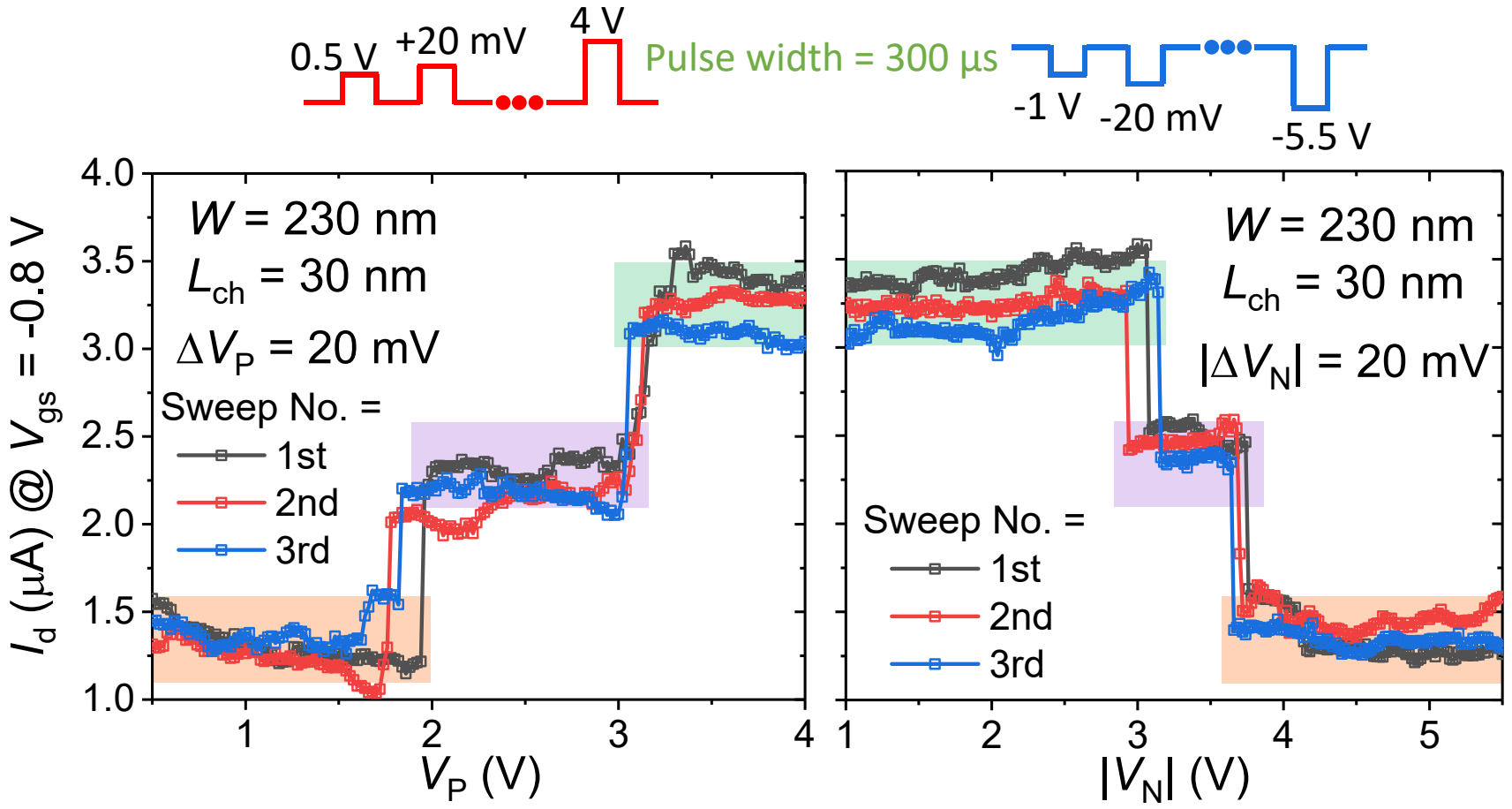
- Gradual conductance switching
- Conductance dynamic range ~5x
- Overall symmetric modulation
- “Increasing pulse amplitude” scheme

Achieving denser arrays with active area scaling



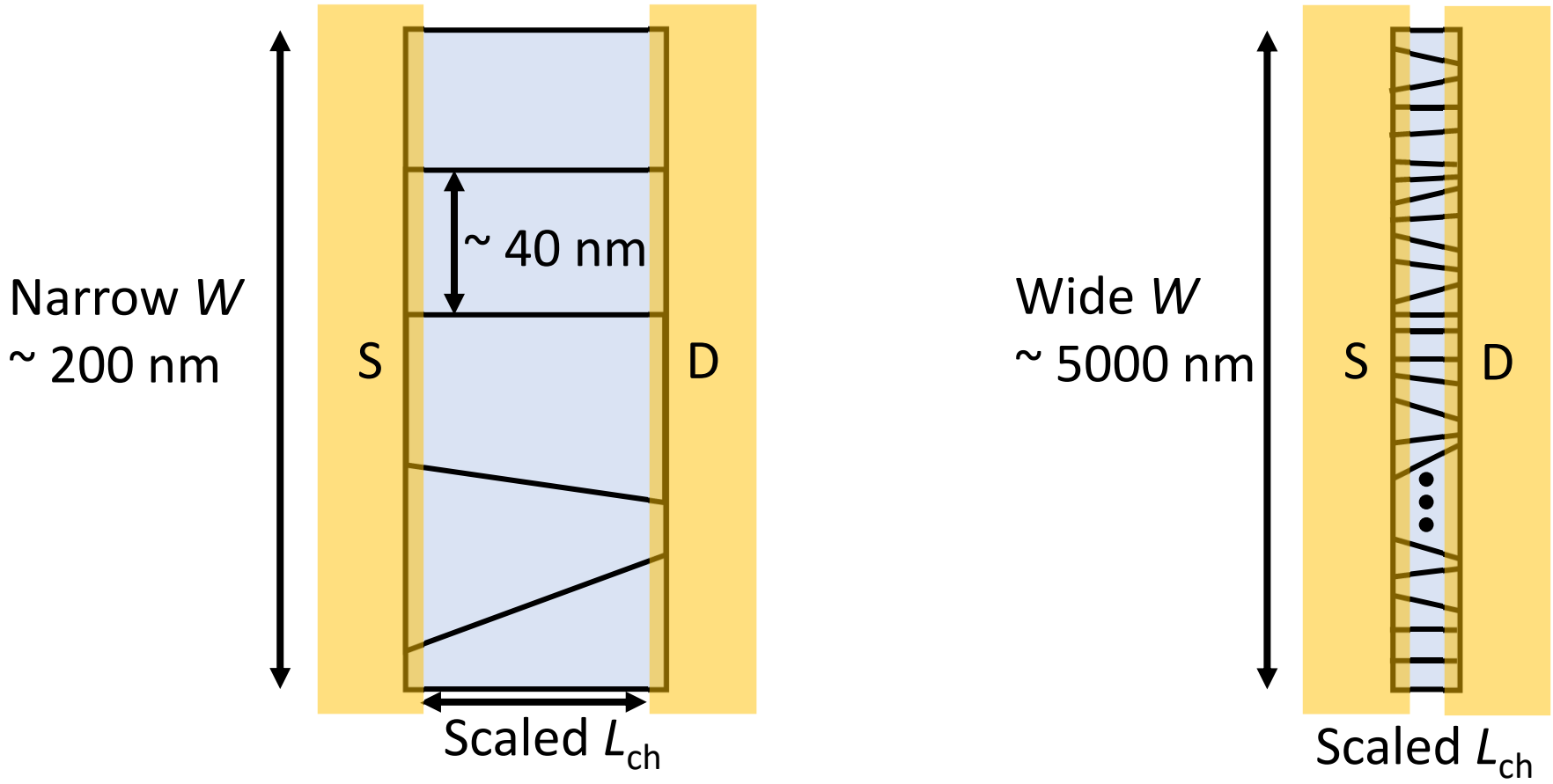
- Small active area desired for dense crossbar arrays
- Need to examine the device operation with narrow width

Conductance switching in FE-FETs: narrow channel



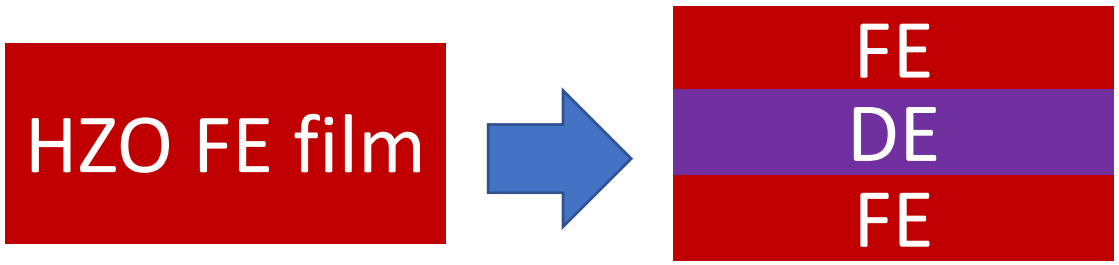
- Discrete instead of gradual conductance switching
- Limited number of active FE domains

Possible domain distribution in FE-FETs

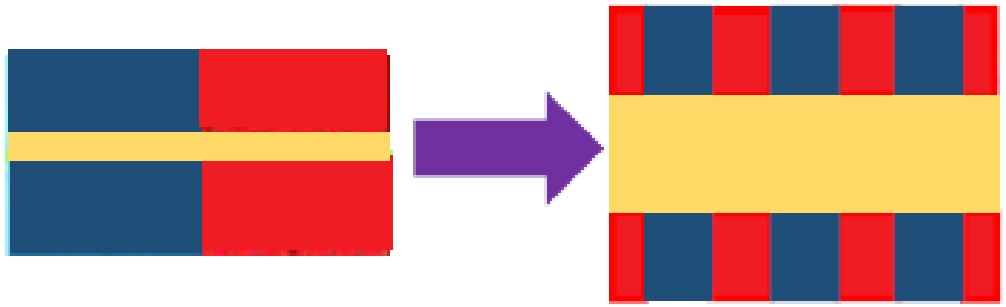


- Number of active domains ↑ as W ↑

Achieving higher domain density…



Superlattice design of FE/DE stacks



Aabrar et al. *IEEE IEDM*, 19.6.1-19.6.4 (2021);
Aabrar et al. *IEEE TED*, 69, 2094-2100 (2022).

- Next step: FE stack engineering

Conclusions and future work

Conclusions:

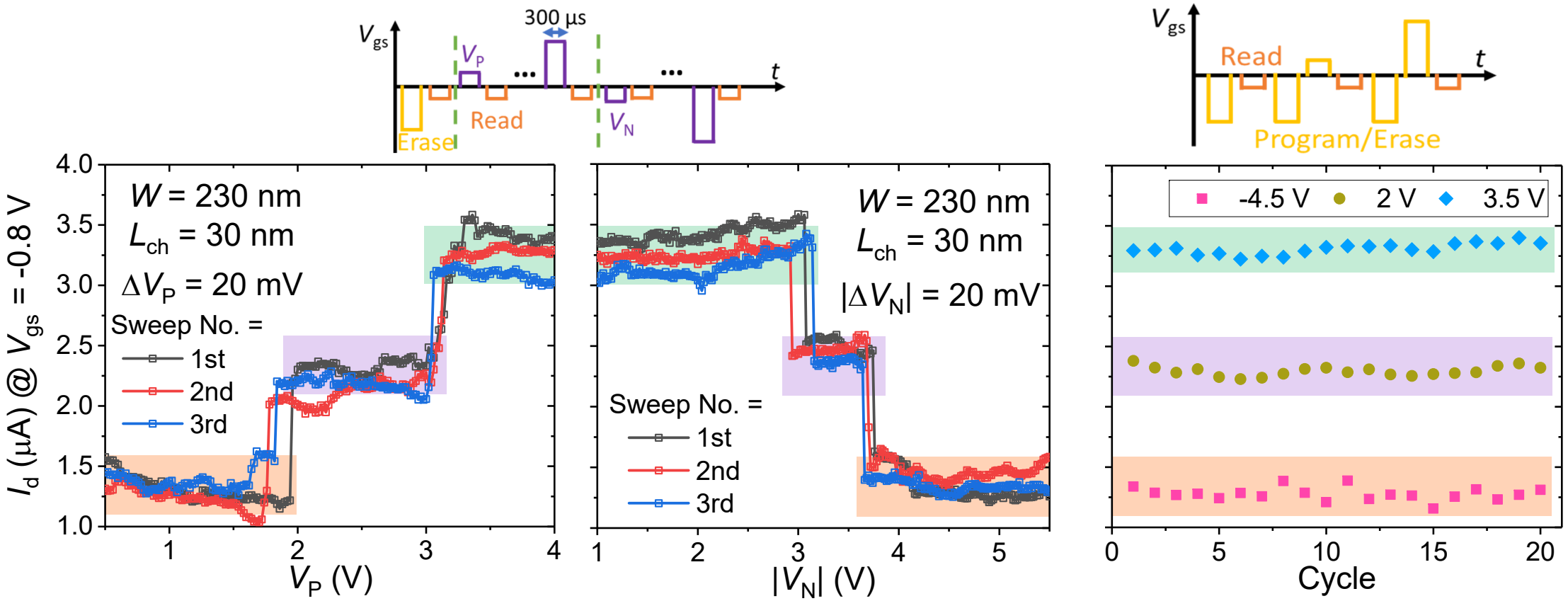
- Scaled channel thickness for improved FE-FET performance
- Gradual conductance switching in wide-channel FE-FETs
- Discrete FE domain switching in narrow-channel FE-FETs

Future work:

- Achieving gradual conductance switching with identical pulse amplitude
- Mapping FE domain size with material engineering
- Study of endurance and retention

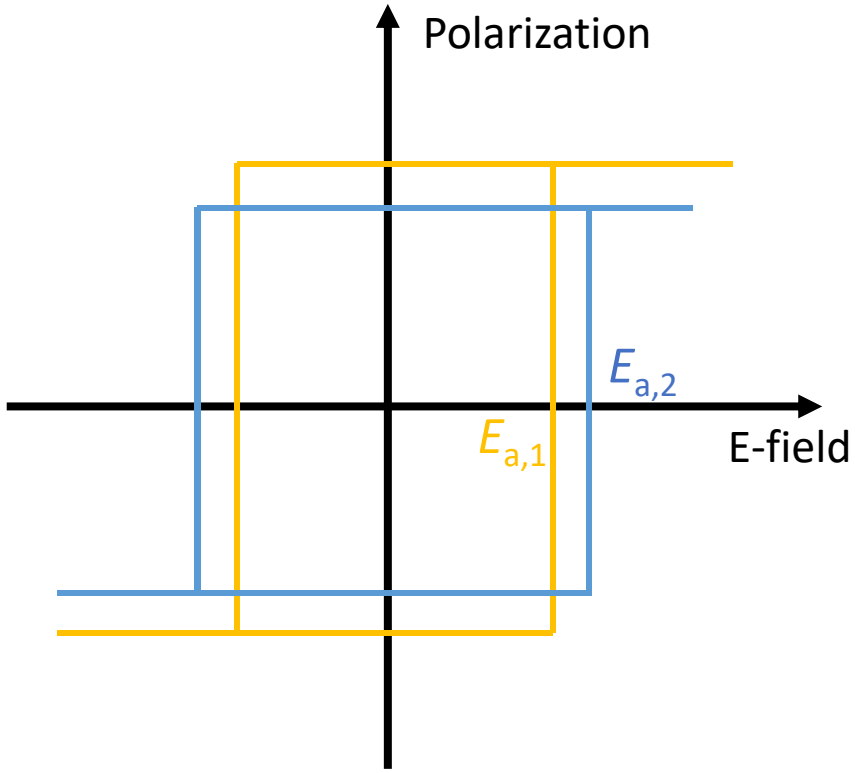
Appendices

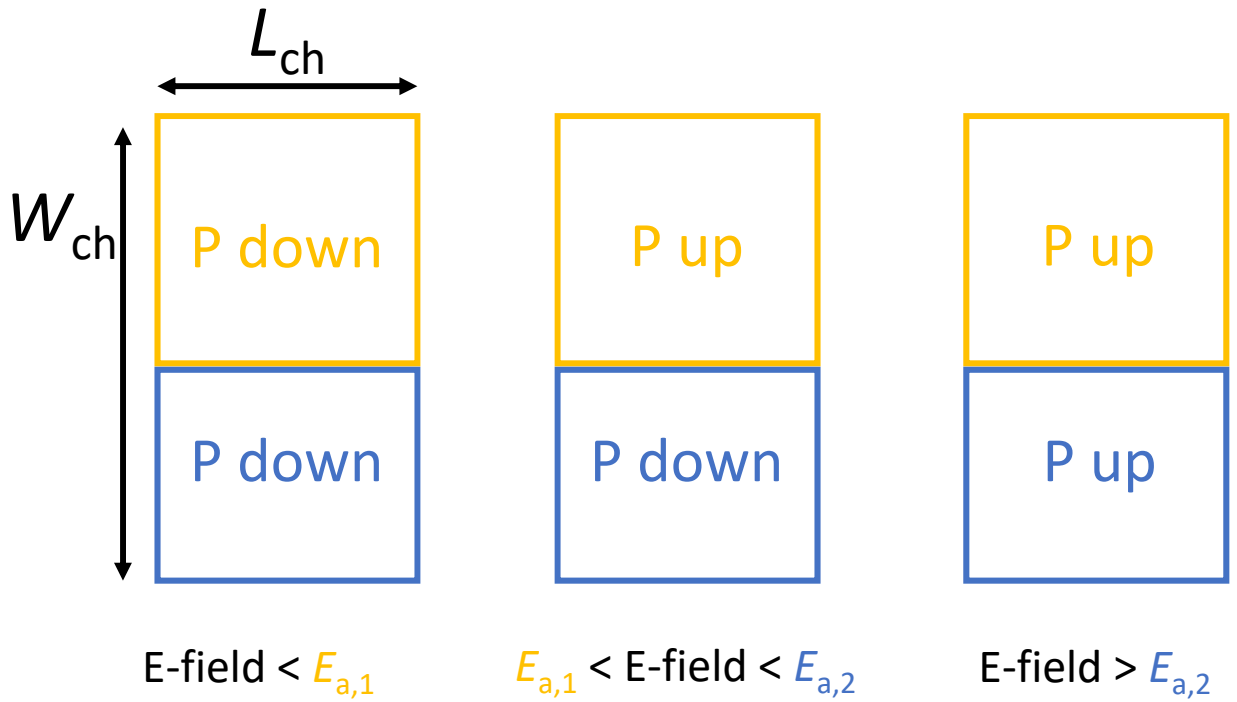
Discrete conduction states



- Clear three quantized states with repeatable measurements
- Three-level memory in agreement with “increasing pulse scheme” results
- Discrete domain switching in highly-scaled devices (both W_{ch} and L_{ch})

Discrete domain switching

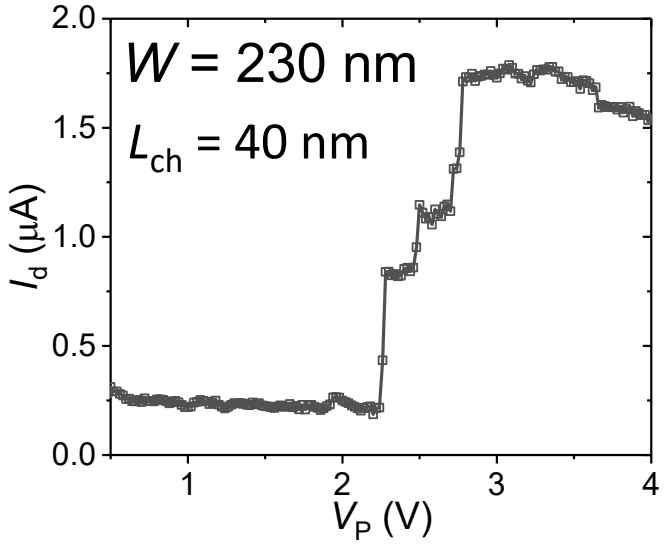


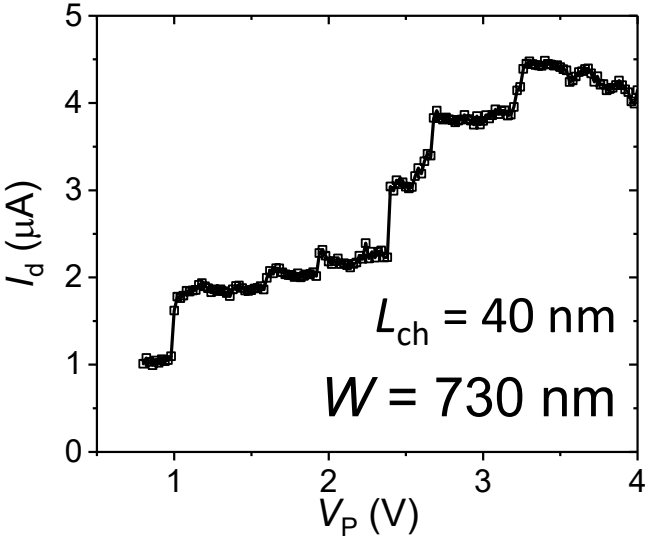


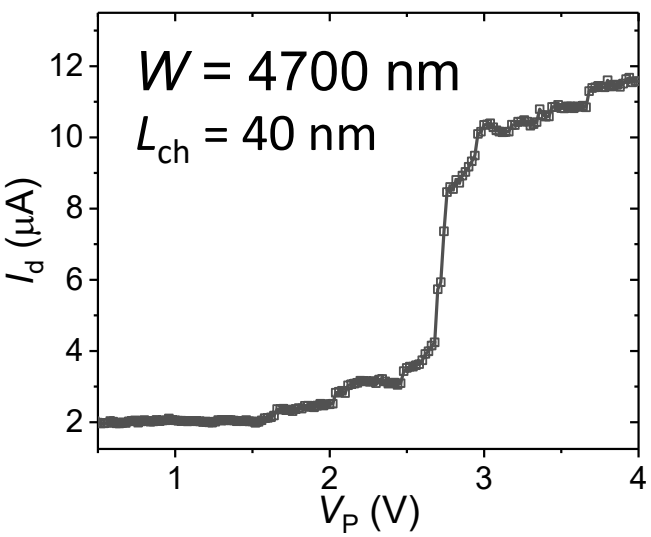
Top-down view of our FE-FETs

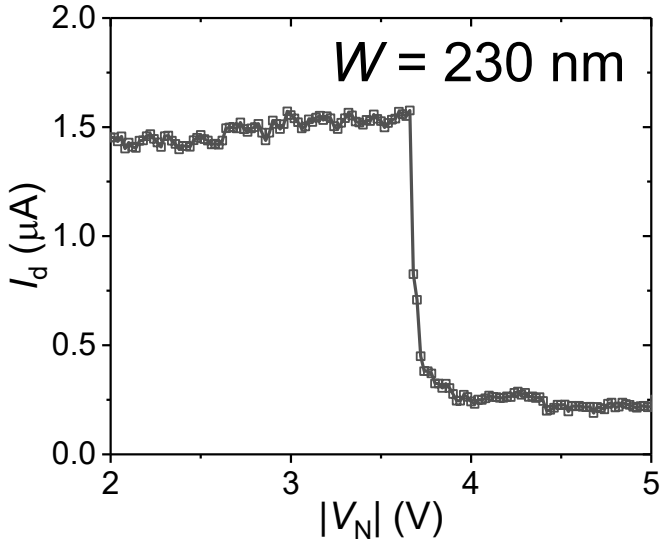
- Discrete local activation fields (E_a) for different domains
- Wider devices → smoother distribution of local activation field

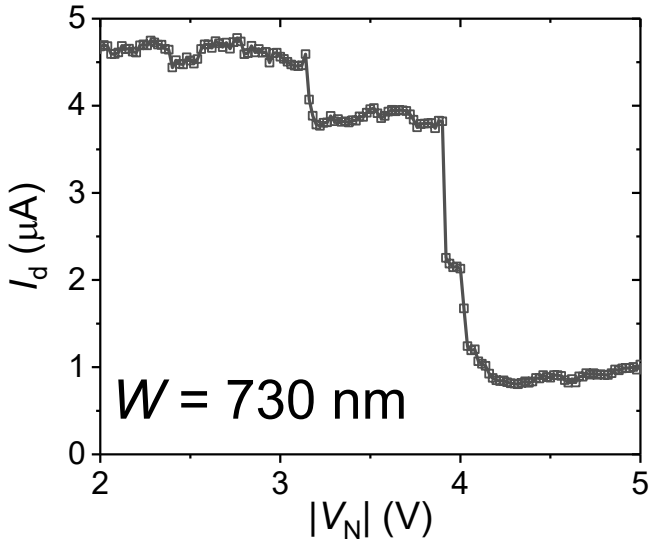
Conductance switching in FE-FETs: channel width scaling

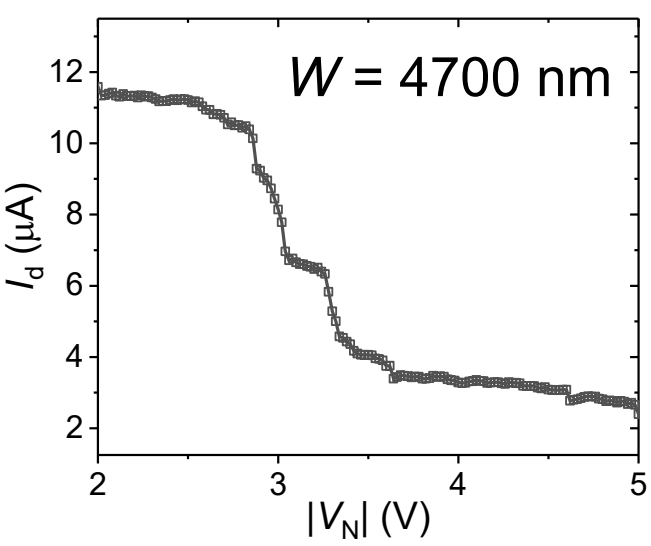












- Number of states ↑ & more gradual switching as W ↑