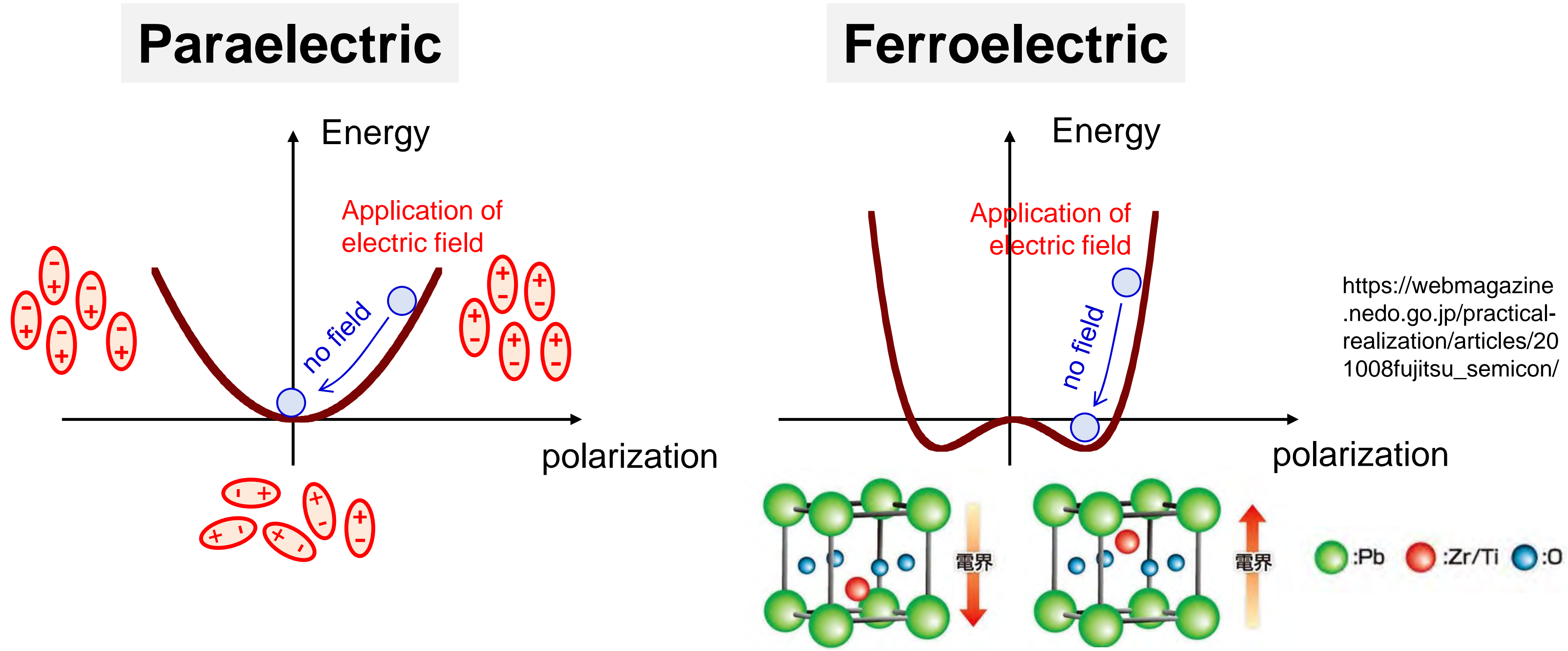
55th IEEE Semiconductor Interface Specialists Conference (SISC)
Tutorial, Dec 11, 2024
Catamaran Resort Hotel and Spa, San Diego, CA, USA
Hafnia-Based Ferroelectric FETs and Capacitors for Low-Power Memory and AI Applications: Physical Understanding of Device Operation and Reliability
Shinichi Takagi
The University of Tokyo, Tokyo, Japan
UTokyo

Outline

- Expectation of HfO_2-based ferroelectric materials and devices
- FeRAM ~ MFM capacitors
  - FeRAM operation
  - Reliability of MFM capacitor
- FeFET memory
  - FeFET memory operation
  - Understanding of coupling between polarization and carrier traps in FeFET
  - Issues related FeFET memory operation
  - FeFET reliability
- AI applications
  - Expectation of FeFETs for AI applications
  - Physical reservoir computing using FeFETs

Expectation of HfO_2-based ferroelectric materials and devices

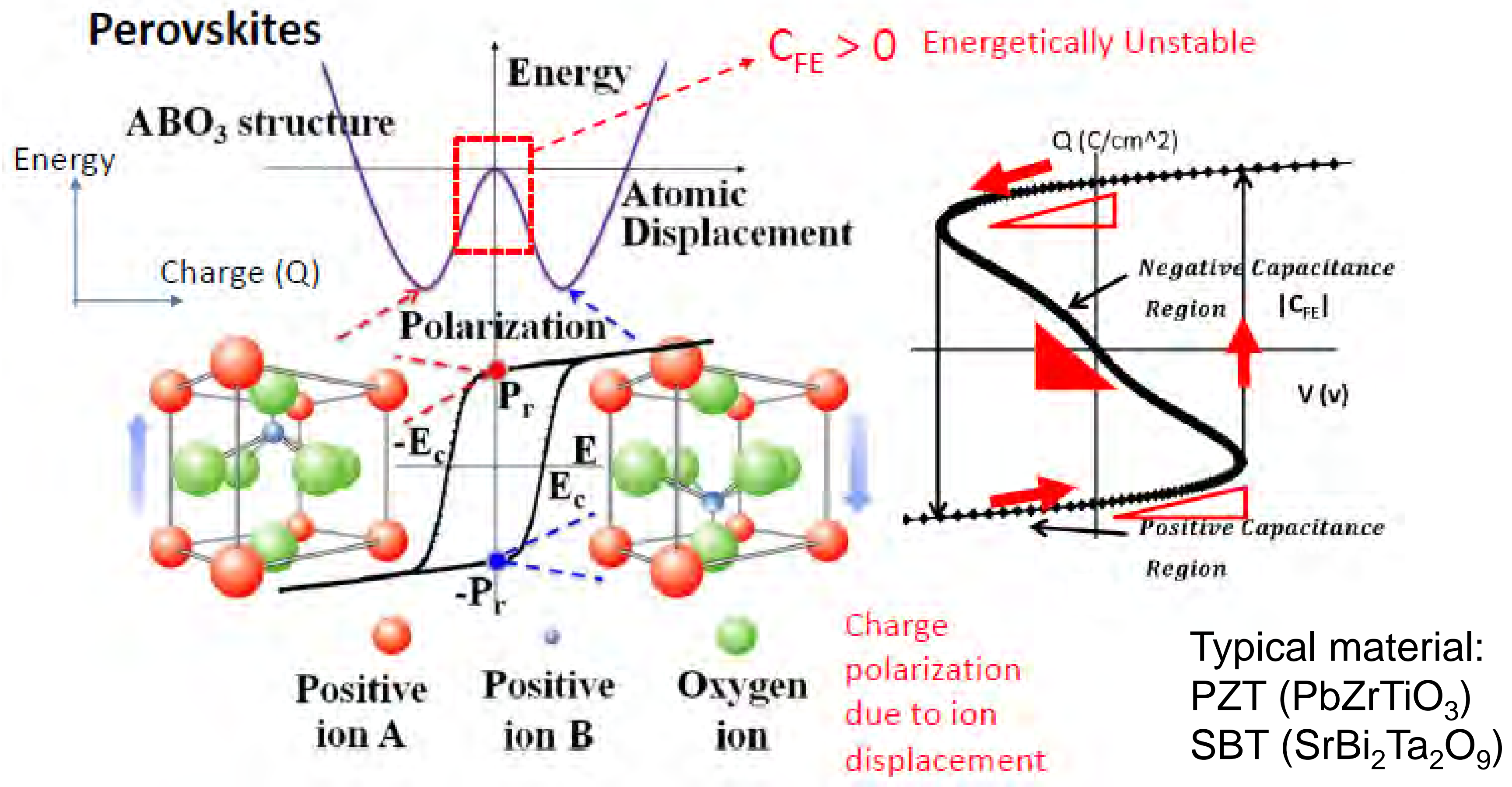
Ferroelectrics



Ferroelectric = material whose polarization is stable in the absence of an electric field (spontaneous polarization) and whose polarization can be controlled by an external electric field

Electrical characteristics of ferroelectric films

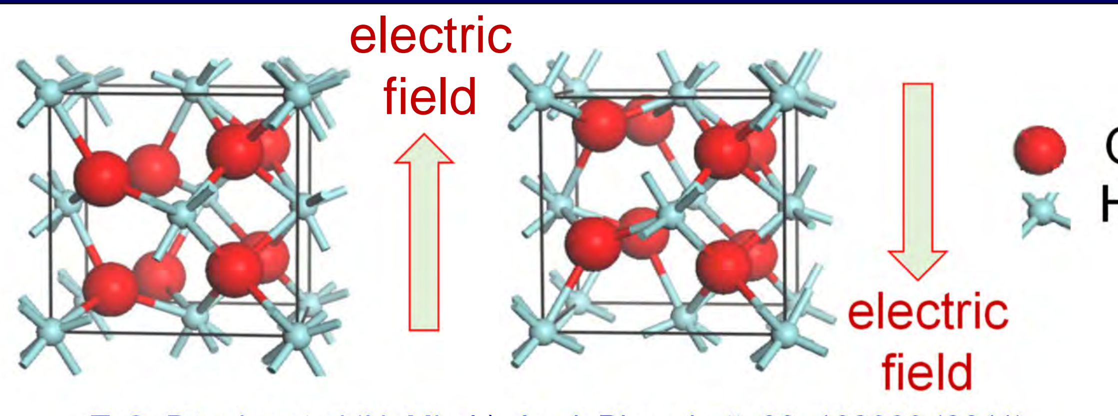
A. Thean, Option beyond FinFETs at 45 nm node, IEDM short course (2016)



Typical material:
PZT ($PbZrTiO_3$)
SBT ($SrBi_2Ta_2O_9$)

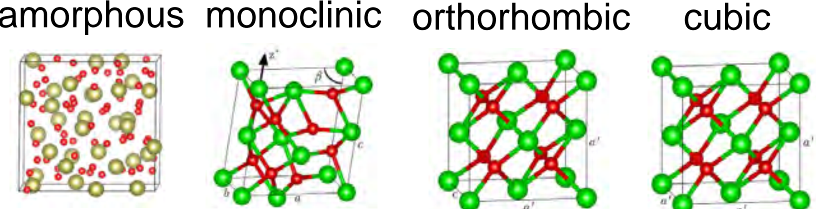
- The stability point of ions is determined by the direction of the electric field
- A change in the position of this stability point of ions causes a large polarization reversal

HfO_2-based ferroelectric film



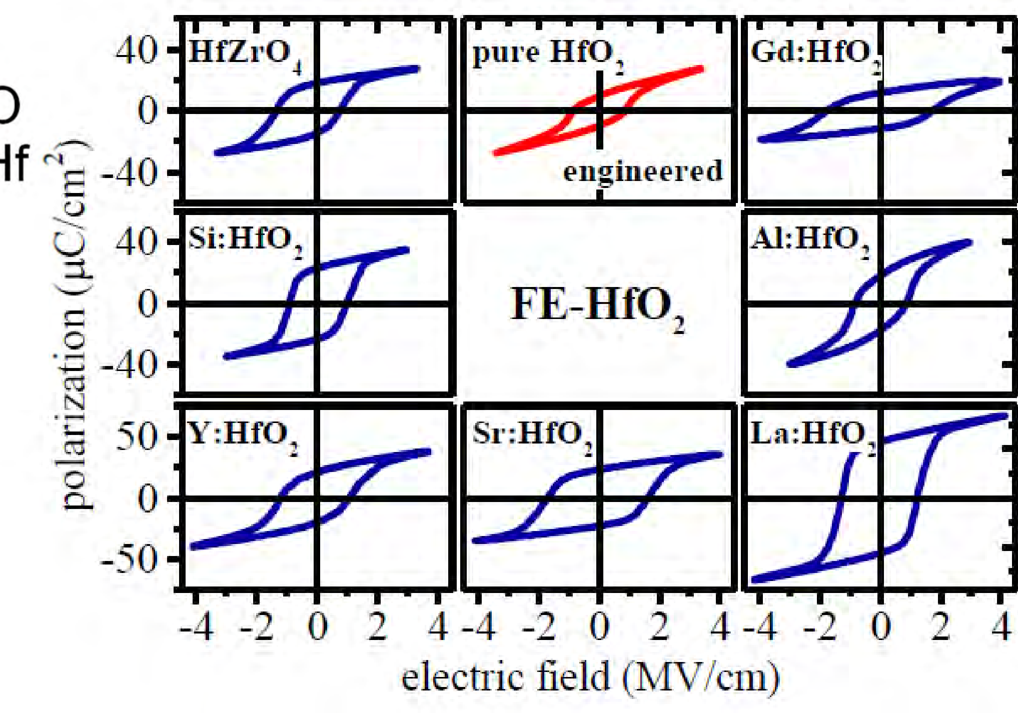
T. S. Böscke *et al* (NaMLab), Appl. Phys. Lett. **99**, 102903 (2011)

- Discovered in 2011
- Movement of oxygen atoms is the origin of spontaneous polarization



R. Materlik *et al*, J. Appl. Phys. **117**, 134109 (2015)
L. Zhao et al, Appl. Phys. Lett. **107**, 013504 (2015)

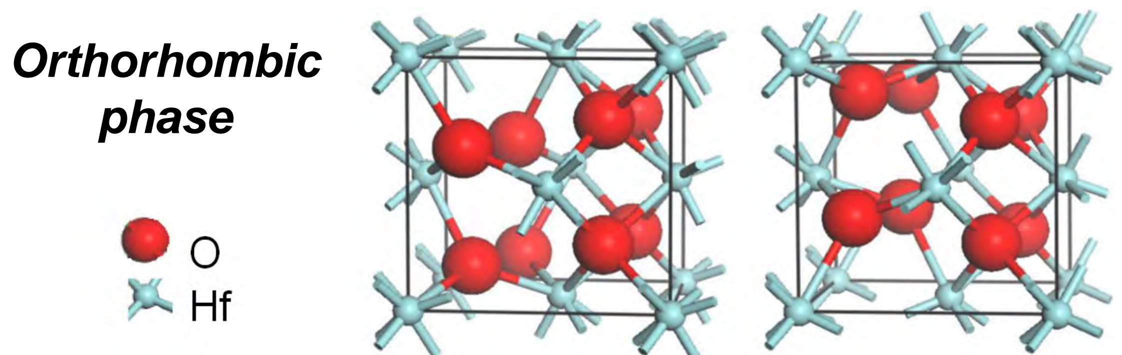
- Various crystal phases can exist
- Ferroelectricity originates in orthorhombic phase with an asymmetric crystal structure



J. Müller *et al*, IEDM (2013); IEDM Tutorial (2019)

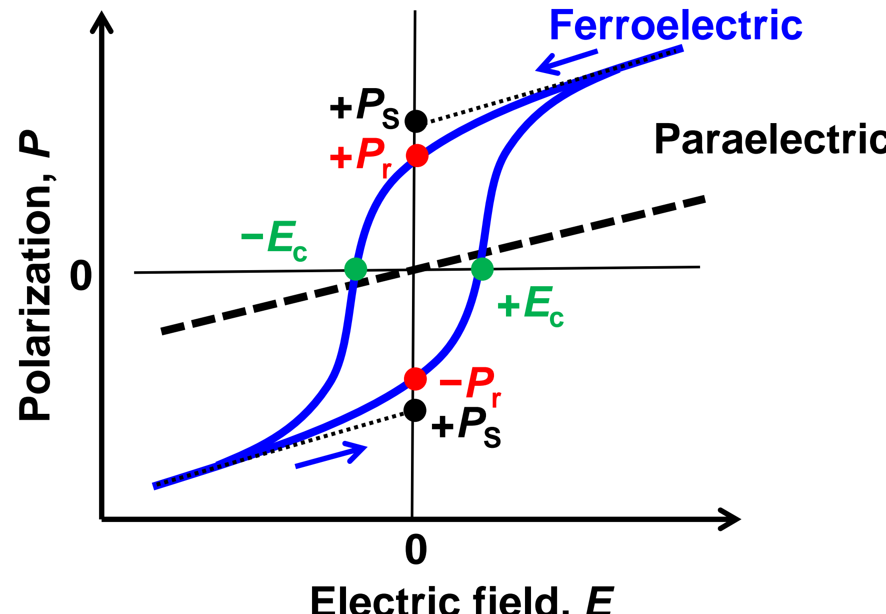
- Ferroelectric crystal phase (orthorhombic-phase) can be stabilized by doping various impurities
- Zr-doped HfO_2 (HZO) exhibits ferroelectricity in a wide range of Zr content

Comparison of ferroelectric characteristics



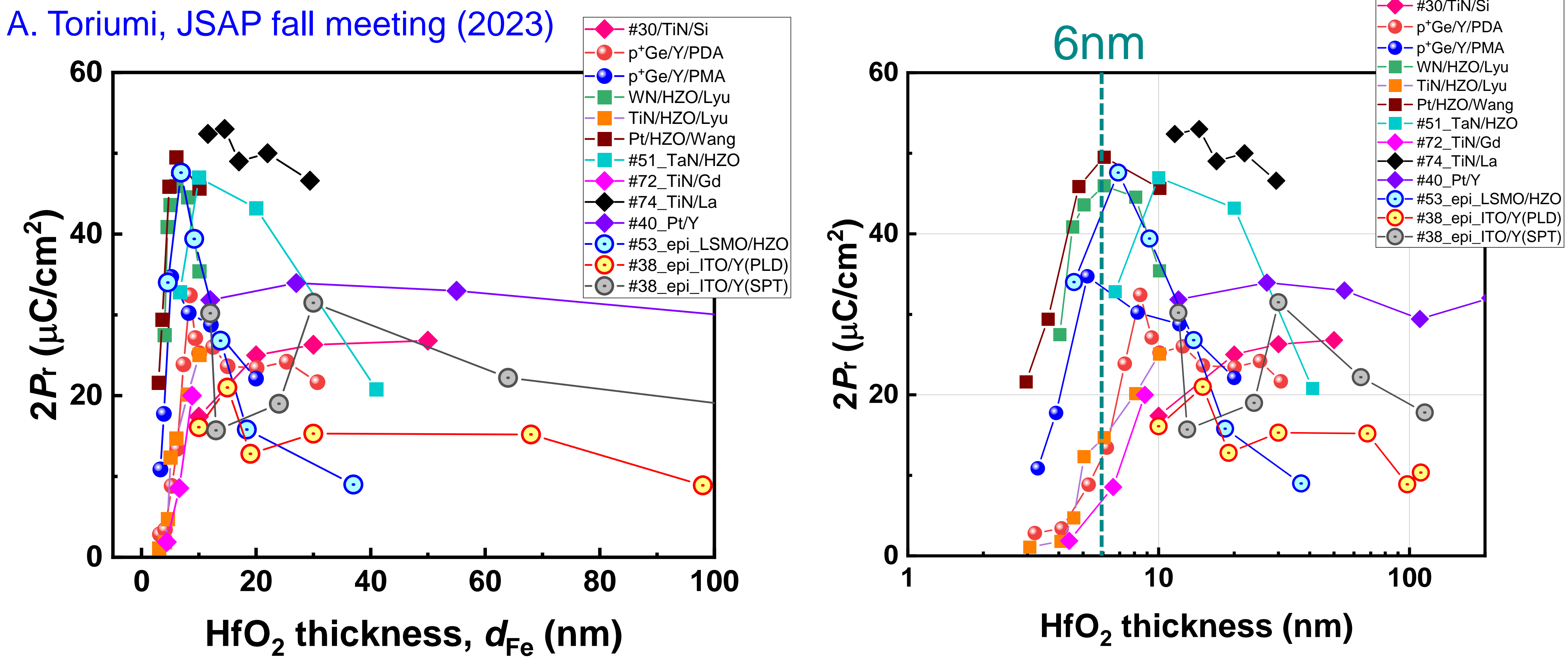
Discovery of doped HfO_2-based ferroelectric films in 2011 by NamLab group
(T.S. Böscke *et al. APL* **99** (2011))

	Pb(Zr, Ti)$_3$	$SrBi_2Ta_2O_9$	HfO_2-based
Film thickness (nm)	50-340	50-130	4-100
Dielectric constant	200	150	30
2Pr ($\mu C/cm^2$)	10-40	10-40	10-40
Ec (kV/cm)	31	90	1000



HfO_2-based materials:
- Higher E_c → (pros) better retention (cons) high operation voltage, worse reliability
- lower permittivity → (pros) higher ferroelectric field in MFIS

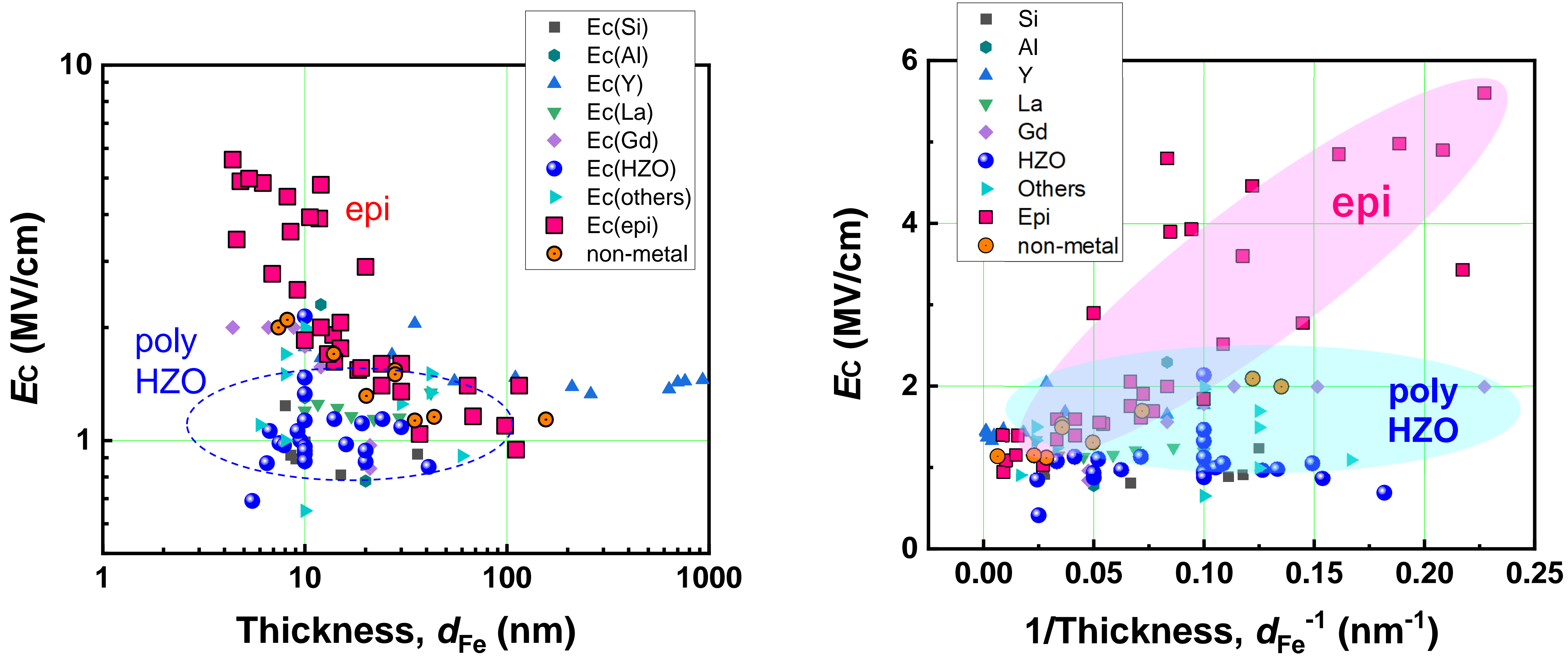
Impact of film thickness on HfO_2-based ferroelectric characteristics



- Sufficiently high $2P_r$ can be obtained in thickness less than 10 nm
- 2Pr increases with thinner films, has a peak around 6 nm and rapidly decreases down to 3 nm

Impact of film thickness on HfO_2-based ferroelectric characteristics

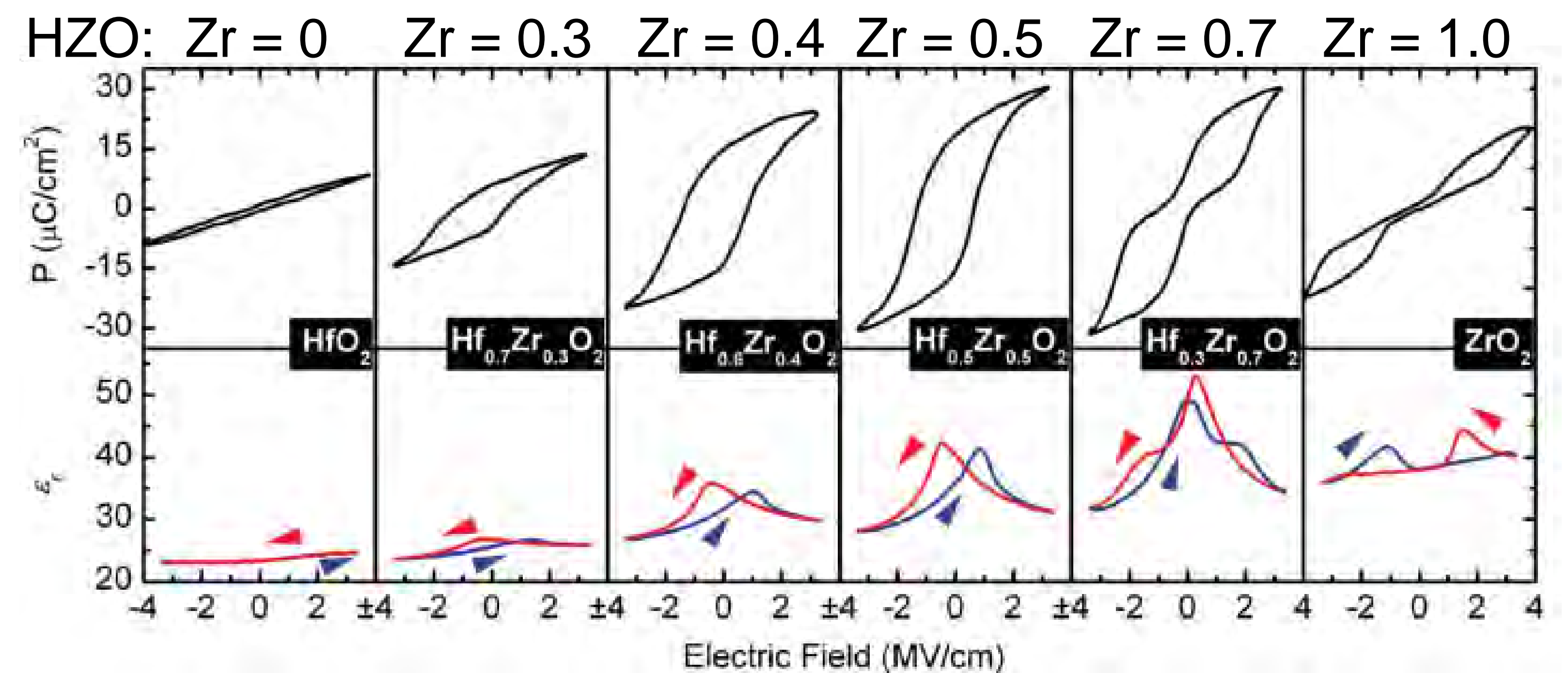
A. Toriumi, JSAP spring meeting (2024)



- E_c of poly HZO films (1-2 MV/cm) is almost constant, irrespective of thickness
- E_c of epi HZO films seem to be inversely proportional to thickness

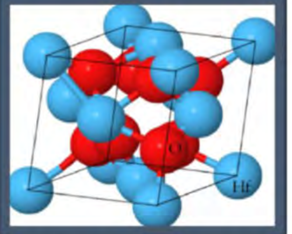
Control of ferroelectric/anti-ferroelectric properties by Zr content

HZO: Zr = 0 Zr = 0.3 Zr = 0.4 Zr = 0.5 Zr = 0.7 Zr = 1.0



Müller et al., Nano Lett., 12 (2012) 4318

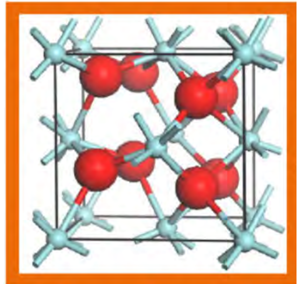
paraelectric



monoclinic

($P2_1/c$) $\varepsilon_r \sim 20$

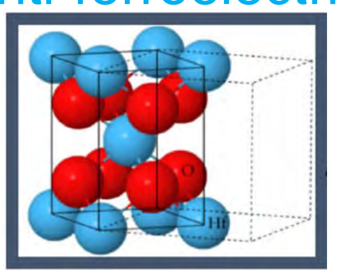
ferroelectric



orthorhombic

($Pbc2_1$) $\varepsilon_r \sim 25$

anti-ferroelectric

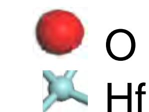


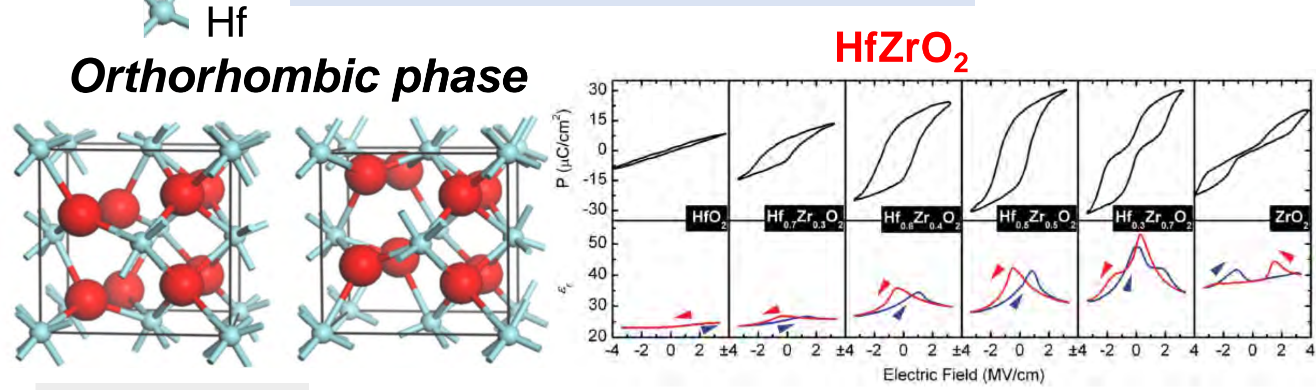
tetragonal

($P4_2/nmc$) $\varepsilon_r \sim 35$

- Ferroelectric/ anti-ferroelectric properties are well controlled by Zr content of $Hf_{1-x}Zr_xO_2$ (HZO) films
- low and mid Zr contents lead to ferroelectricity due to orthorhombic phase and high Zr contents lead to anti-ferroelectricity due to tetragonal phase

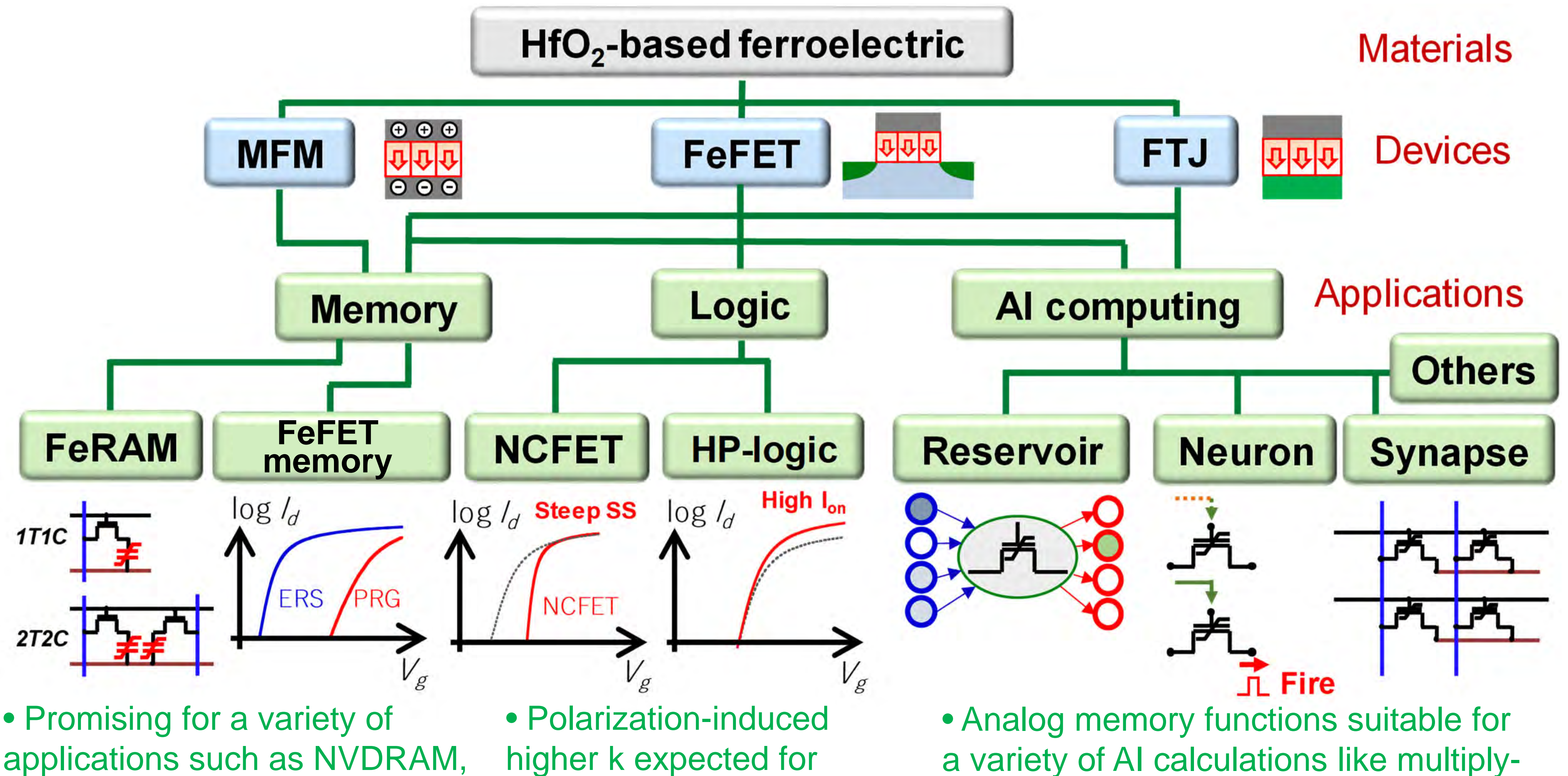
Comparison between HfO_2-based and Perovskite ferroelectric films





Perovskite ferroelectrics		HfO_2-based ferroelectrics
PZT, BTO, SBT, …		
Mass production over 20 years	Technology	First report in 2011, under development
10~40 $\mu C/cm^2$	Polarization (Pr)	10~40 $\mu C/cm^2$
50~200 nm	Film thickness	10 nm or less → high scalability
Noble metal (difficult to process)	Electrode	e.g. TiN (easy to process) } easy fabrication
Possible, but not easy	3D trench structure	Applicable (formed by ALD) } easy fabrication
Pb (special fabrication line needed)	Element	Hf, Zr, Si } Easy implementation into Si platform
Passivation layer needed for protection	hydrogen tolerance	Highly resistant } Easy implementation into Si platform

A variety of device applications using HfO_2-based ferroelectrics



- Promising for a variety of applications such as NVDRAM, SRAM, NAND flash etc.
- Polarization-induced higher k expected for logic applications
- Analog memory functions suitable for a variety of AI calculations like multiply-accumulate operation

Benchmark of ferroelectric memory

	Mainstream Charged based Memories				Emerging Non-volatile Memories					
	SRAM	DRAM	FLASH		PCM	RRAM	STT-MRAM	SOT-MRAM	FeRAM	FeFET
			NOR	NAND						
Cell area	>150F^2	6F^2	10F^2	<4F^2 (3D)	4~50F^2	4~50F^2	6~50F^2	12~100F^2	6~50F^2	6~50F^2
Multi-bit	1	1	2	3-4	2-3	2-3	1	1	1	2-3
Voltage	<1V	<1V	>10V	>10V	<3V	<3V	<1V	<1V	<2V	<3V
Read time	~1ns	~10ns	~50ns	~10µs	<10ns	<10ns	<10ns	~1ns	<100ns	<50ns
Write time	~1ns	~10ns	10µs-1ms	100µs-1ms	~50ns	<100ns	<20ns	<3ns	<100ns	<100ns
Retention	N/A	~64ms	>10y	>10y	>10y	>10y	>1y	>1y	>10y	>1y
Endurance	>1E16	>1E16	~1E5	1E3~1E4	1E6~1E9	1E3~1E9	1E6~1E14	~1E12	1E9~1E12	1E6~1E9
Write Energy (J/bit)	~fJ	~10fJ	100pJ	~10fJ	~10pJ	~pJ	~pJ	~pJ	~100fJ	~fJ

(HfO_2-based FE is also promising for SN in advanced NAND flash) S. Yu, IEDM short course 2 (2022)

Parameter/NVM	MRAM	FeRAM	FeFET	PCM	RRAM	NAND flash	NOR flash	DRAM
R/W speed	<10 ns	10 ns	100 ns	>100 ns	<10 ns	100 μs	1 μs to 1 s	~ 10 ns
Endurance	1×10^{15}	1×10^{15}	1×10^{5}	$>1 \times 10^{8}$	1×10^{9}	1×10^{4}	1×10^{6}	1×10^{16}
Retention	10 years	10 years	10 years	10 years	10 years	10 years	10 years	64 ms
Energy	2 pJ bit^{-1}	50 fJ bit^{-1}	<1 fJ bit^{-1}	3 pJ bit^{-1}	50 pJ bit^{-1}	1 nJ bit^{-1}	10 nJ bit^{-1}	pJ bit^{-1}
Cell size	$20F^2$	$6F^2$	$6–10F^2$	$5.5F^2$	$4F^2$	$4F^2$	$10F^2$	$6F^2$

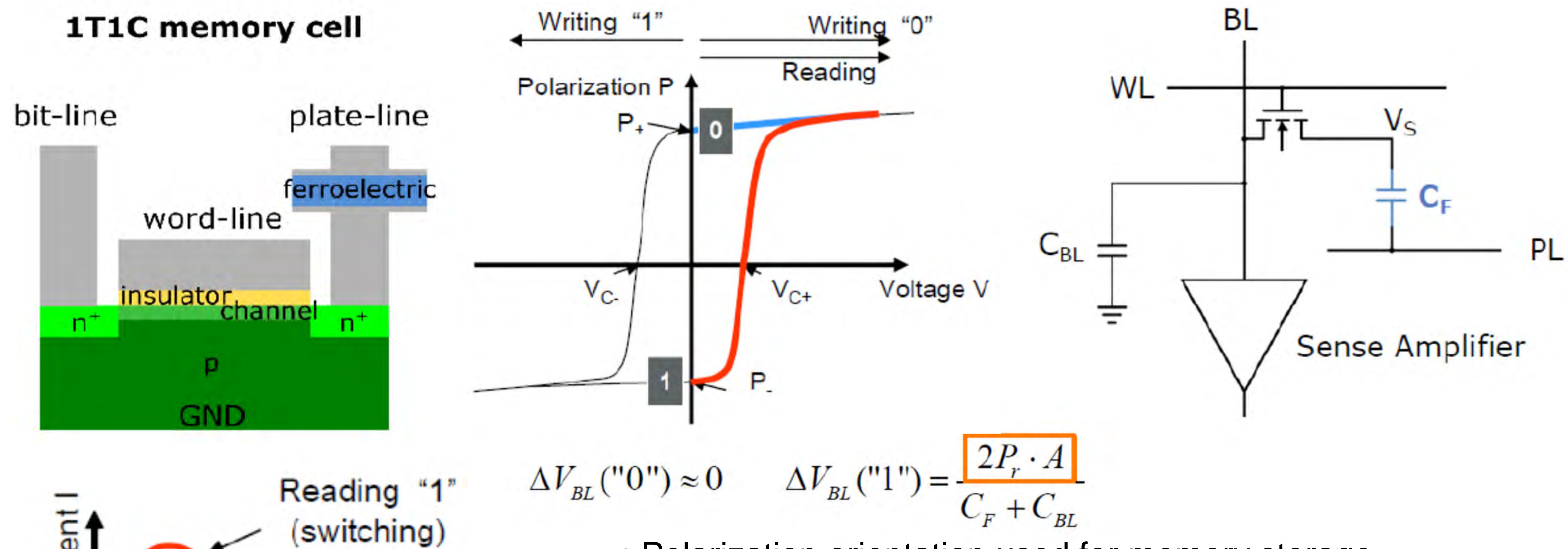
T. Schenk et al., *Rep. Prog. Phys.* **83**, 086501 (2020)

- Low energy consumption is an advantage for ferroelectric memory because of writing by voltage

FeRAM ~ MFM capacitors

FeRAM operation

Operation of FeRAM



$$\Delta V_{BL}("0") \approx 0 \qquad \Delta V_{BL}("1") = \frac{2P_r \cdot A}{C_F + C_{BL}}$$

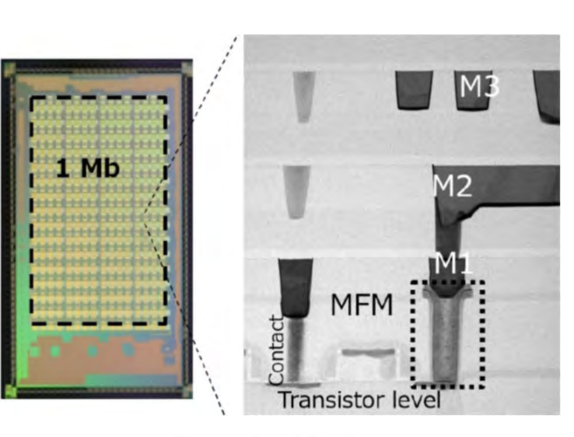
- Polarization orientation used for memory storage
- Destructive readout -> re-write necessary
- Memory window (sensing margin) in FRAM can be improved by increasing capacitor area A and remanent polarization Pr as well as w/ Pr/Ps ratio getting closer to unity

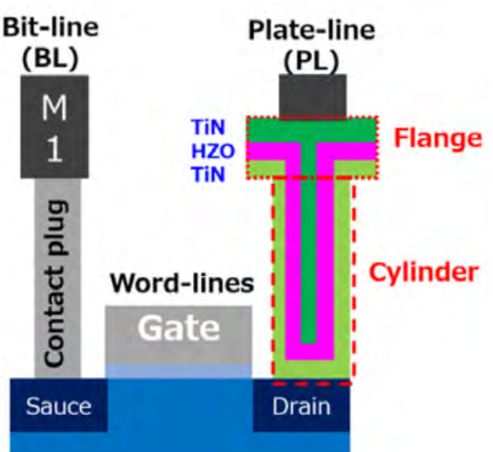
U. Schroeder, VLSI symp. short course (2020)

Recent Progress of HZO FeRAM – 1Mb 3D FeRAM, 32 Gb NVDRAM

1Mb 3D 1T/1C FeRAM

J. Okuno et al. (Sony), IEDM (2023) 11-7





Sense

Operating speed [ns]

Operating voltage [V]

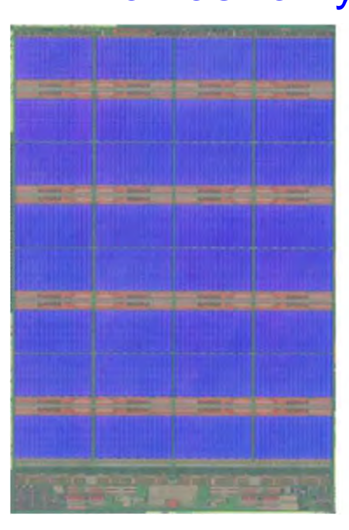
2.4 V, 26 ns

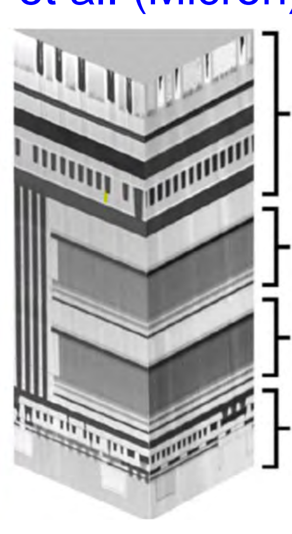
1.8 V, 40 ns

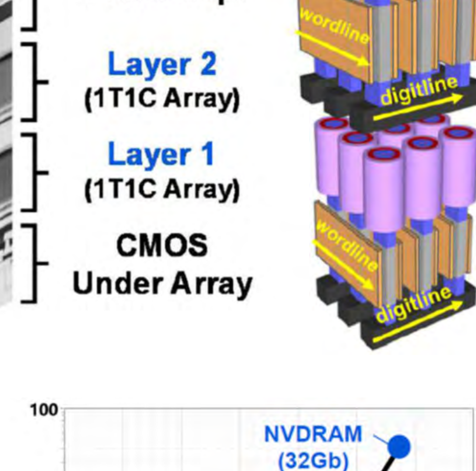
Device	IEDM'23 [This Work]	
Capacity	1 Mb	
MFM type	3D	
MFM size [μm²]	0.028	
Ferro. film [nm]	HZO 6 nm	
Write voltage [V]	1.8	2.4
Retention	85°C, >10^6 sec	85°C, > 10 year
Endurance *without refresh [cycles]	10^{11}	10^{12}

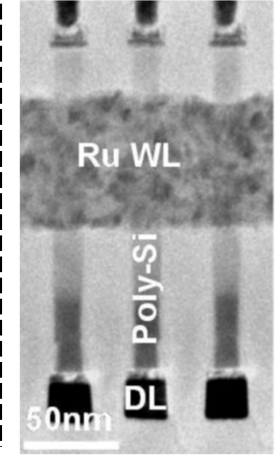
32Gb 3D Stacked Non-volatile DRAM

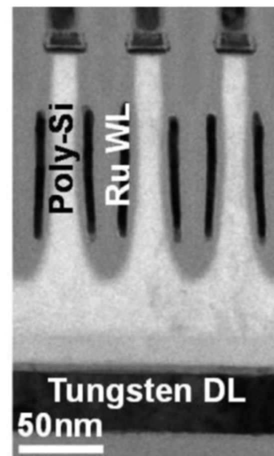
N. Ramaswamy et al. (Micron), IEDM (2023) 15-7

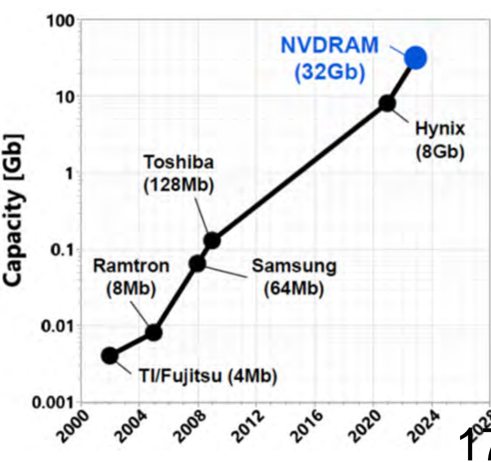








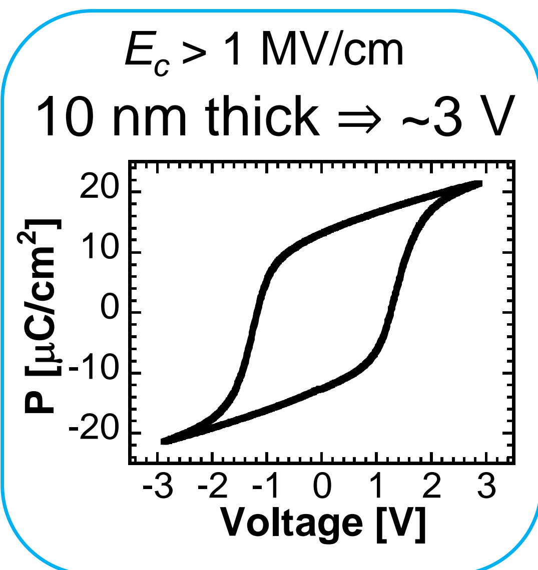




Critical issues of HZO MFM capacitors for FeRAM application

Challenge #1

Low operating voltage



High $2P_r$

10~50 μC/cm²

Low annealing temperature compatible BEOL

400~500ºC for $Hf_{0.5}Zr_{0.5}O_2$ (<400ºC is more desired)

Requirements for advanced FeRAM applications

High retention

Challenge #2

High endurance and reliability

PZT > 10^{13} to 10^{17} cycles

HfO_2 10^6~10^9 to 10^{11} cycles

- High polarization reversal voltage due to high E_c in HZO → difficulty in low voltage operation, needed for FeRAM embedded with advanced logic

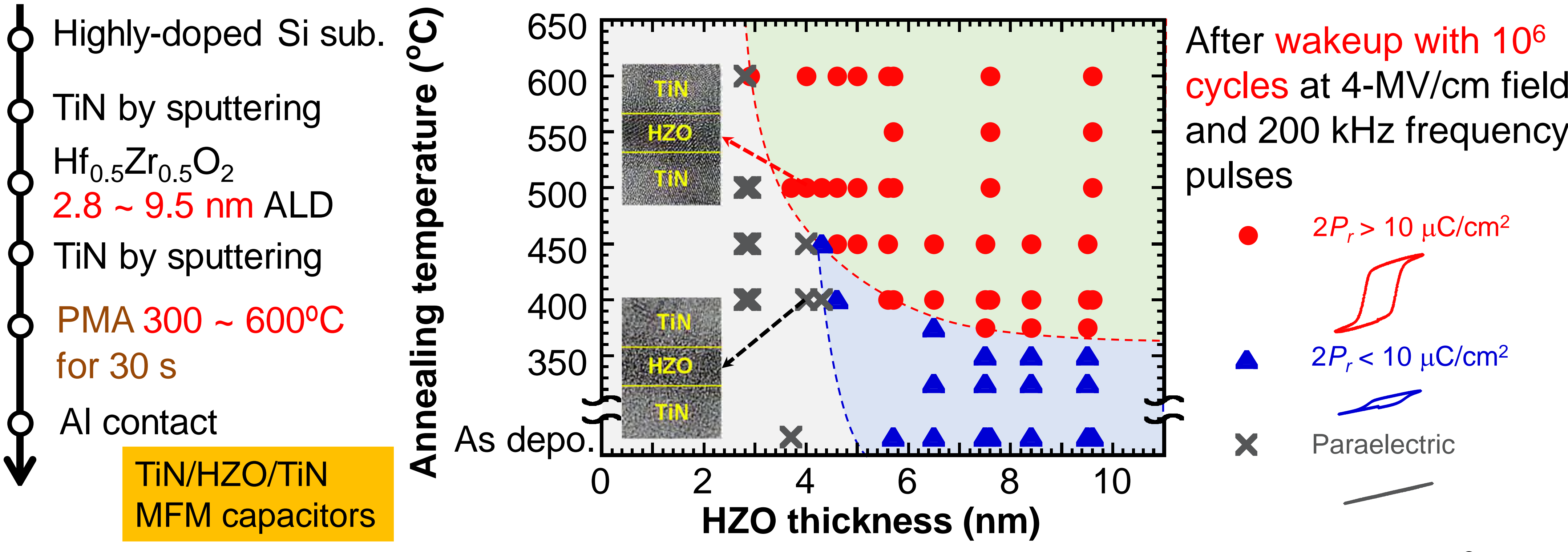
Impact is not too clear

- further improvement in endurance needed for HfO_2-based ferroelectrics

→ Expectation to HZO thickness scaling

Relationship between HZO thickness and crystallization temperature

K. Toprasertpong et al., ACS Appl. Mater. Interfaces **14**, 51137 (2022)

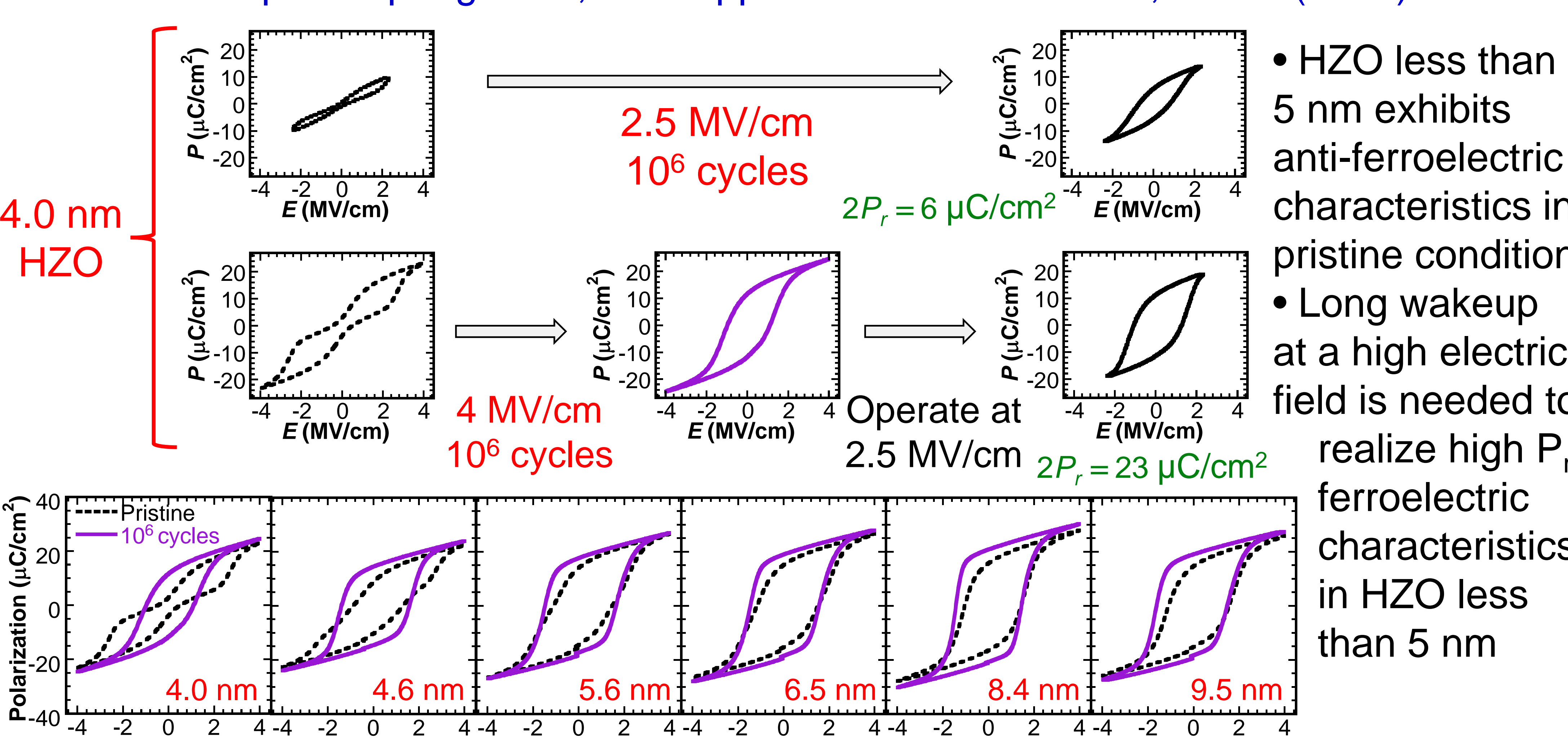


• An annealing temperature of 400 °C is sufficient to obtain 2Pr over 10 μC/cm² in HZO films thicker than 6 nm, whereas annealing temperatures of 450, 500, and 600 °C are required for 5, 4, and 3 nm films, respectively, even after 10^6 wakeup

• HZO thinning increases the crystallization temperature to ensure sufficient 2Pr, indicating there is a trade-off between HZO thickness and annealing temperature

Necessity of long wakeup for thin HZO capacitor

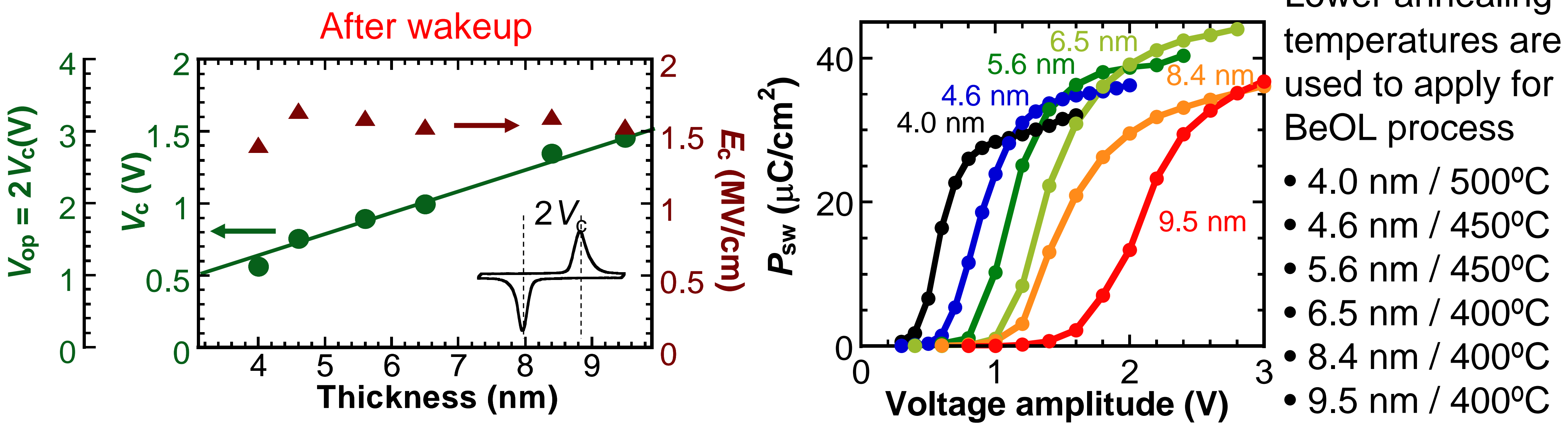
K. Toprasertpong et al., ACS Appl. Mater. Interfaces **14**, 51137 (2022)



- HZO less than 5 nm exhibits anti-ferroelectric characteristics in pristine condition
- Long wakeup at a high electric field is needed to realize high P_r ferroelectric characteristics in HZO less than 5 nm

Low voltage switching in scaled HZO MFM capacitors

K. Tahara et al., VLSI Symp. T7-3 (2021); K. Toprasertpong et al., ACS Appl. Mater. Interfaces **14**, 51137 (2022)

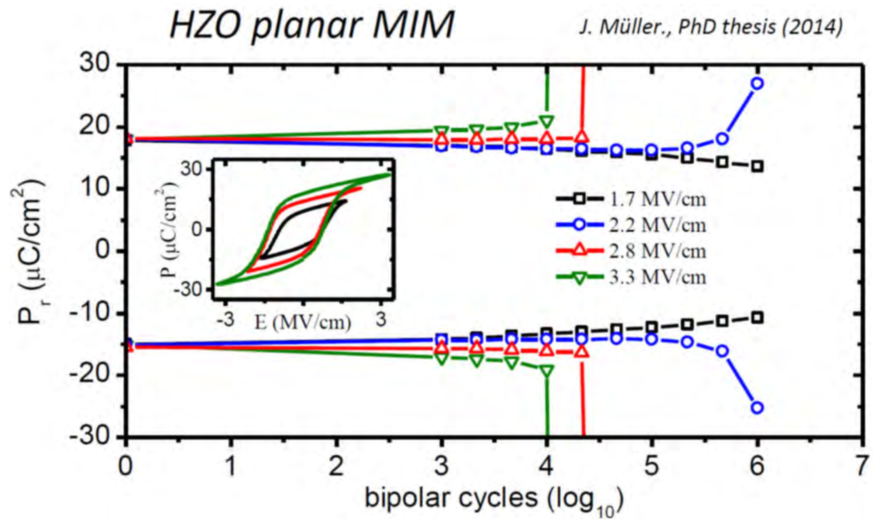


- E_c is found to have only a weak dependency on the thickness and is almost constant in the range of 4.0−9.5 nm
- Operation voltage can be significantly reduced down to less than 1 V (~0.8V) by thinning HZO down to 5-4 nm

Reliability of MFM capacitor

Reliability issues (1) in HZO MFM ~ breakdown, fatigue and wakeup

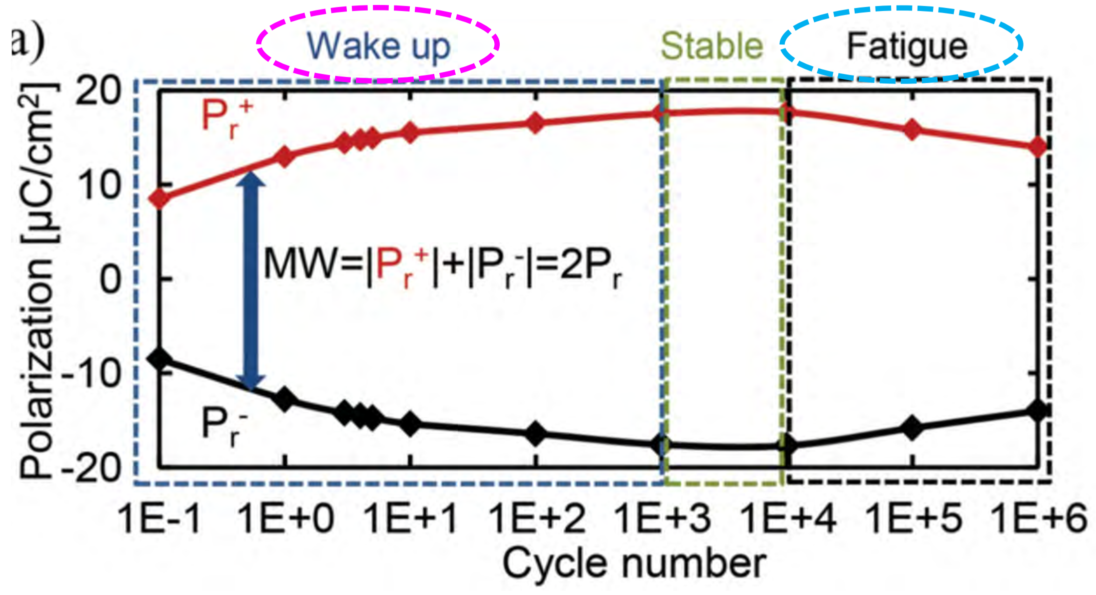
J. Muller, IEDM short course (2019)



	SBT (25 nm)	PZT (35 nm)	FE-HfO_2 (10 nm)
E_{SW}	~ 300 kV/cm	~ 375 kV/cm	~ 3 MV/cm
E_{BD}	~ 1.4 MV/cm	~ 1.2 MV/cm	~ 5 MV/cm
E_{BD} vs. E_{SW}	~ 21%	~ 31%	~ 60%

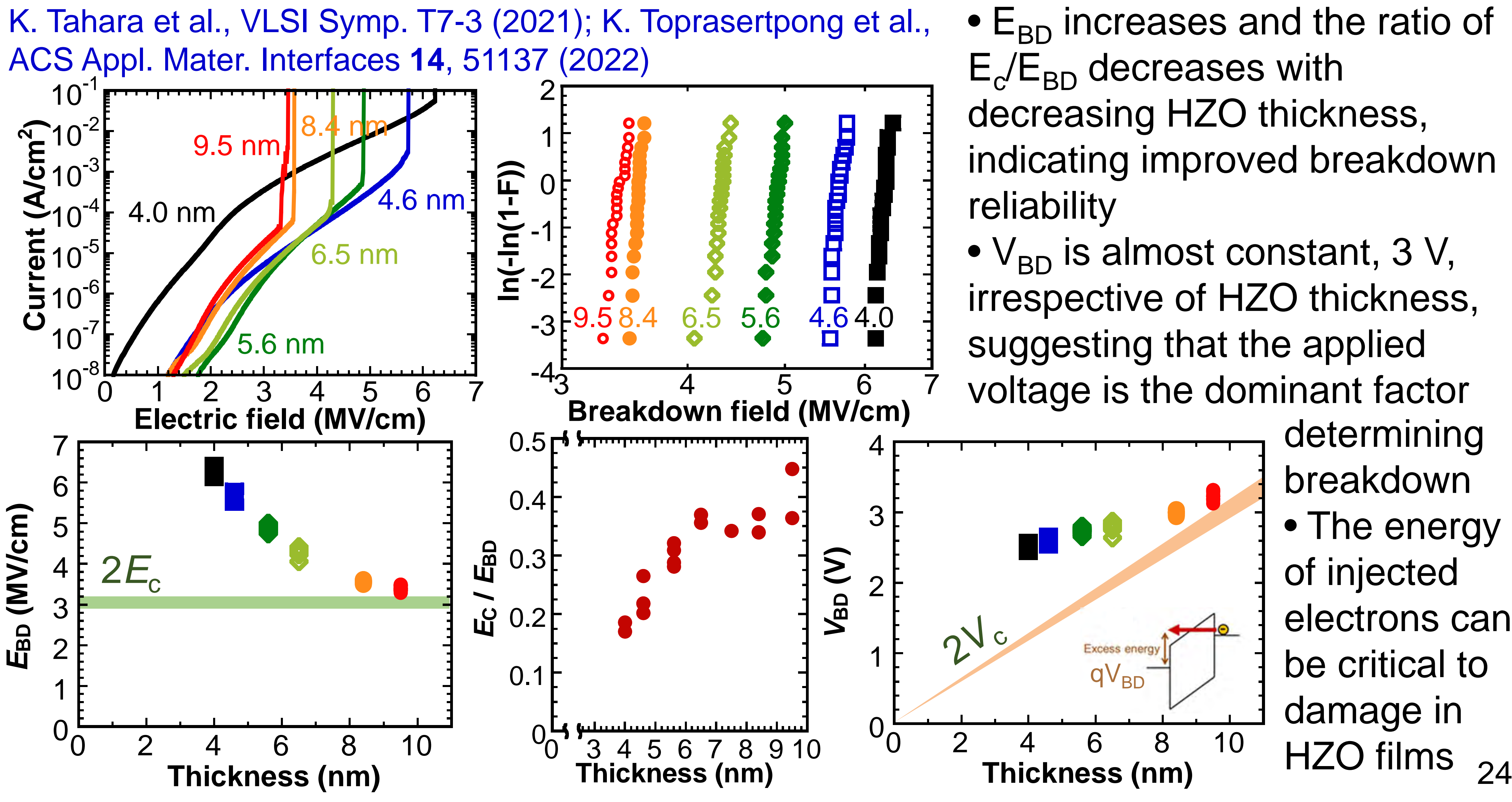
• Compared to other ferroelectrics the high E_c in FE-HfO_2 leads to an operation point ($E_{SW} \approx 3E_c$) which is close to E_{BD} (breakdown field)

M. Pesic et al., Adv. Funct. Mater. 26, 4601 (2016)



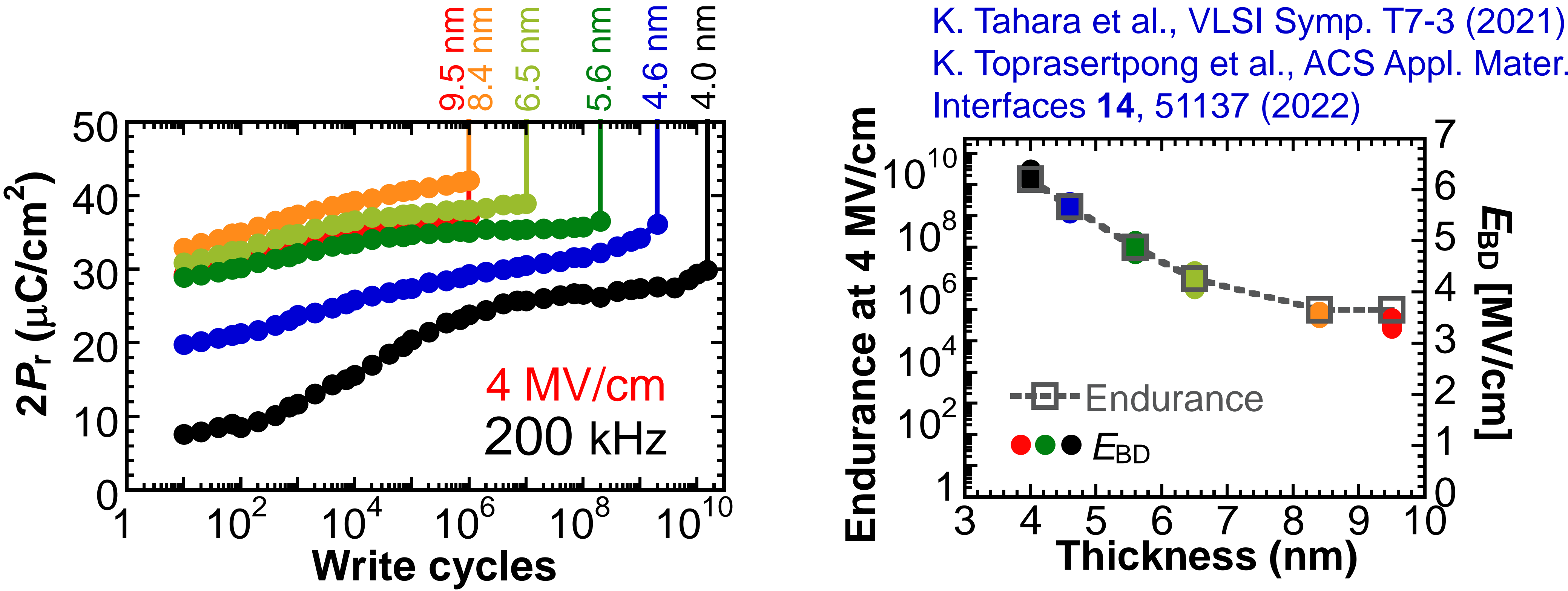
• Saturated operation endurance is limited mainly by hard dielectric breakdown induced by bipolar cycling stress

• In lower electric field operation, on the other hand, wakeup, where P_r increases with increasing cycle number, and fatigue, where P_r decreases with cycle number, are often observed



Impact of HZO thickness scaling on breakdown reliability
K. Tahara et al., VLSI Symp. T7-3 (2021); K. Toprasertpong et al., ACS Appl. Mater. Interfaces 14, 51137 (2022)
• E_{BD} increases and the ratio of E_c/E_{BD} decreases with decreasing HZO thickness, indicating improved breakdown reliability
• V_{BD} is almost constant, 3 V, irrespective of HZO thickness, suggesting that the applied voltage is the dominant factor determining breakdown
• The energy of injected electrons can be critical to damage in HZO films
Current (A/cm²)
Electric field (MV/cm)
9.5 nm
8.4 nm
4.0 nm
4.6 nm
6.5 nm
5.6 nm
ln(-ln(1-F))
Breakdown field (MV/cm)
9.5 8.4 6.5 5.6 4.6 4.0
E_{BD} (MV/cm)
$2E_c$
Thickness (nm)
E_c / E_{BD}
Thickness (nm)
V_{BD} (V)
$2V_c$
Excess energy
qV_{BD}
Thickness (nm)

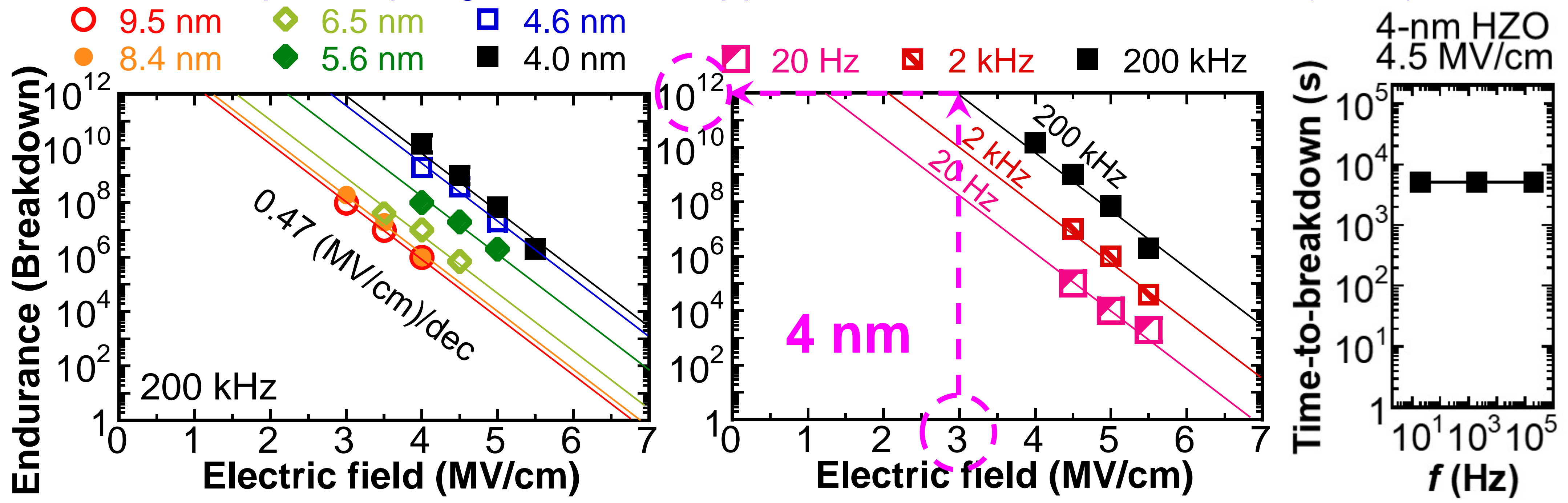
Write cycle endurance due to breakdown of MFM capacitors



- Significant improvement of endurance, limited by break down under 4 MV/cm cycling, is obtained by greater than 4 orders of magnitude by only thinning HZO films from 9.5 to 4.0 nm, because of the increase of E_{BD}
- Endurance cycle number limited by breakdown increases with a decrease in HZO thickness with the similar HZO thickness dependence of E_{BD}

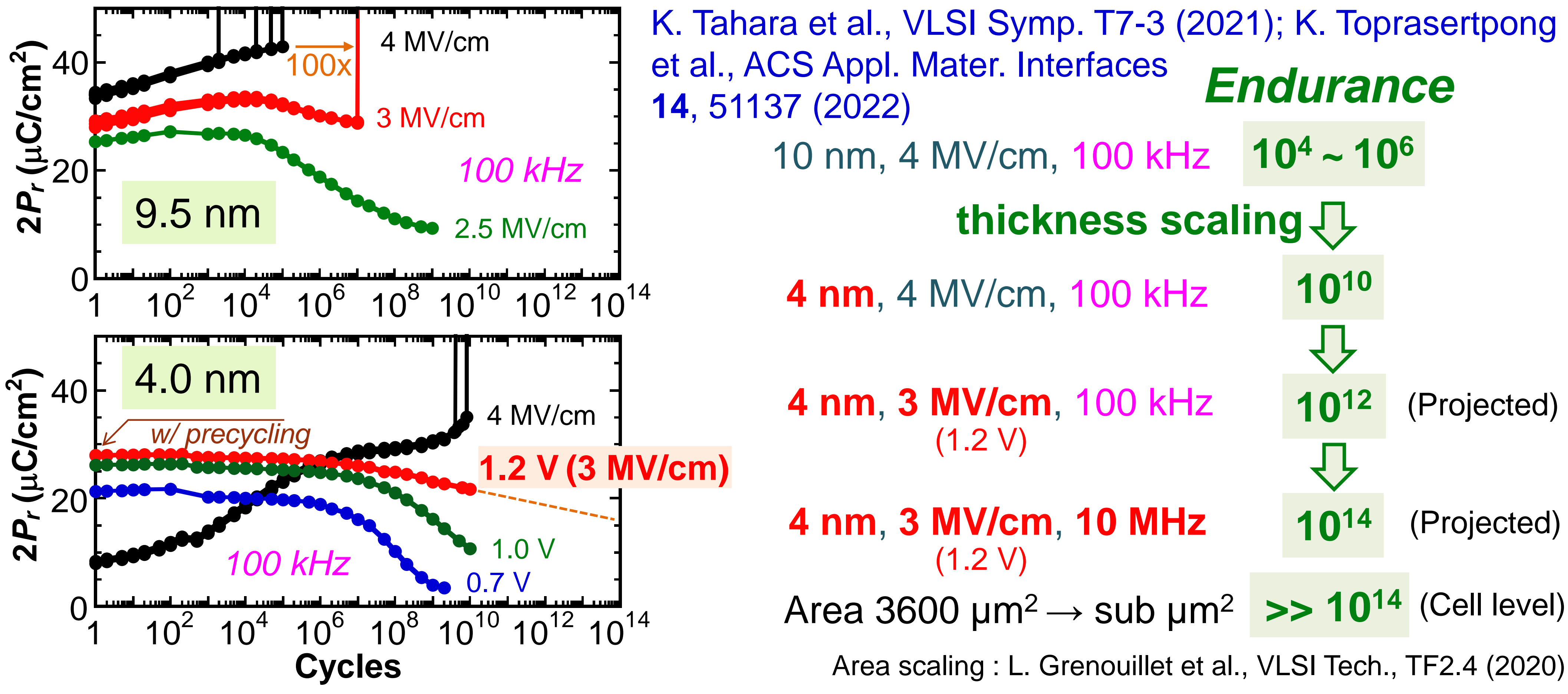
Breakdown endurance vs Electric field

K. Toprasertpong et al., ACS Appl. Mater. Interfaces **14**, 51137 (2022)



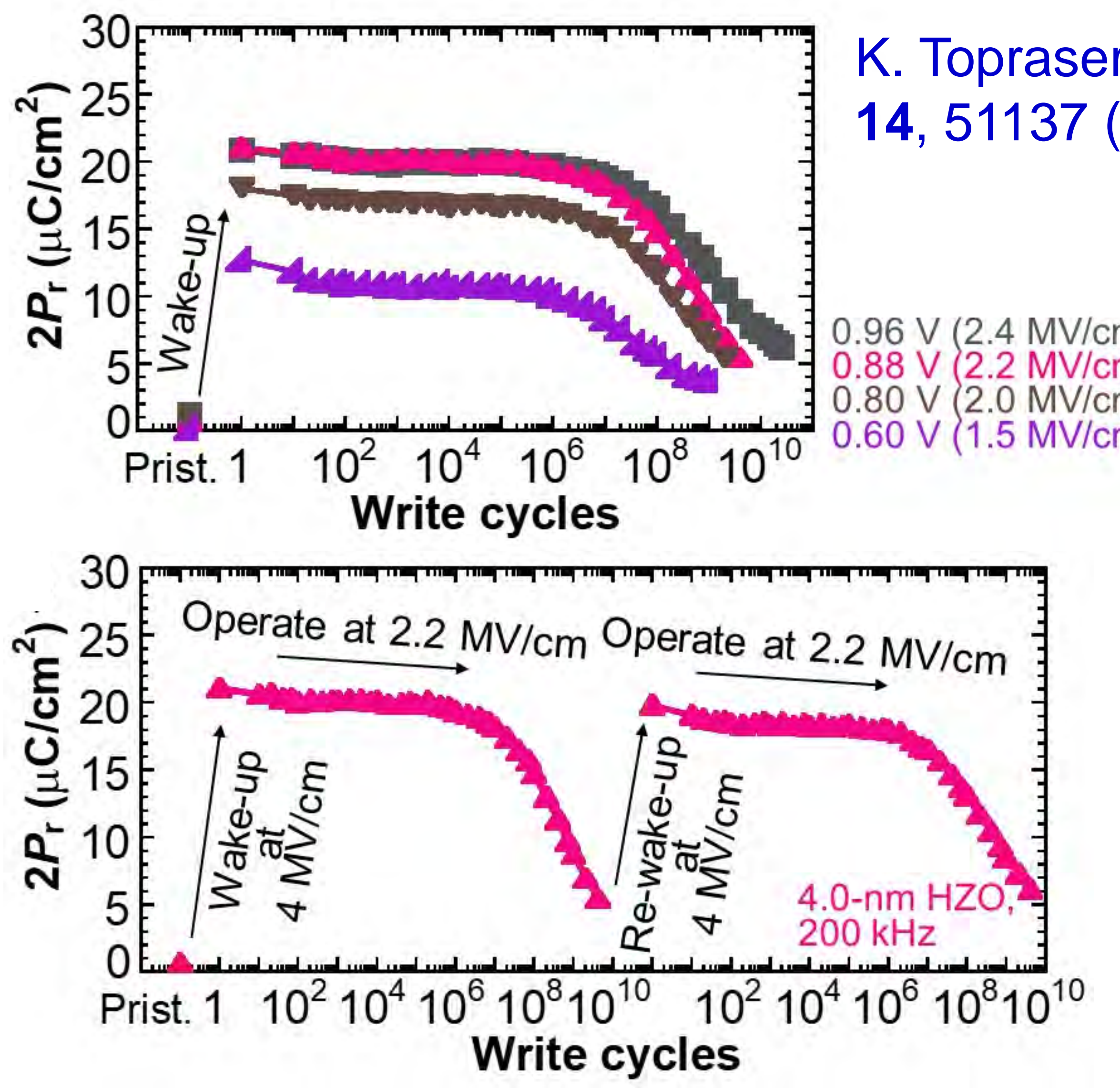
- Relationship between endurance and operating field is almost the same for all the HZO thicknesses
- The number of endurance cycles decreases by one order for every increasing field of 0.47 ± 0.08 MV/cm, independent of the HZO thickness, resulting in approximately 1×10^{12} cycles for breakdown-limited-endurance of the 4.0 nm HZO capacitor operating at 1.2 V (3 MV/cm)
- The stress frequency does not no change the field-acceleration factor
- The number of endurance cycles is proportional to the operating frequency; thus, the total time-to-breakdown is independent of the operating frequency

Improvement of endurance and reliability with HZO scaling



- The endurance cycle of 4-nm-thick HZO is around 10^{10} times at 4 MV/cm, and can be improved up to around 10^{12} times by lowering the electric field down to 3 MV/cm
- The endurance cycle of 10^{14} times or more can be expected at higher operation frequencies

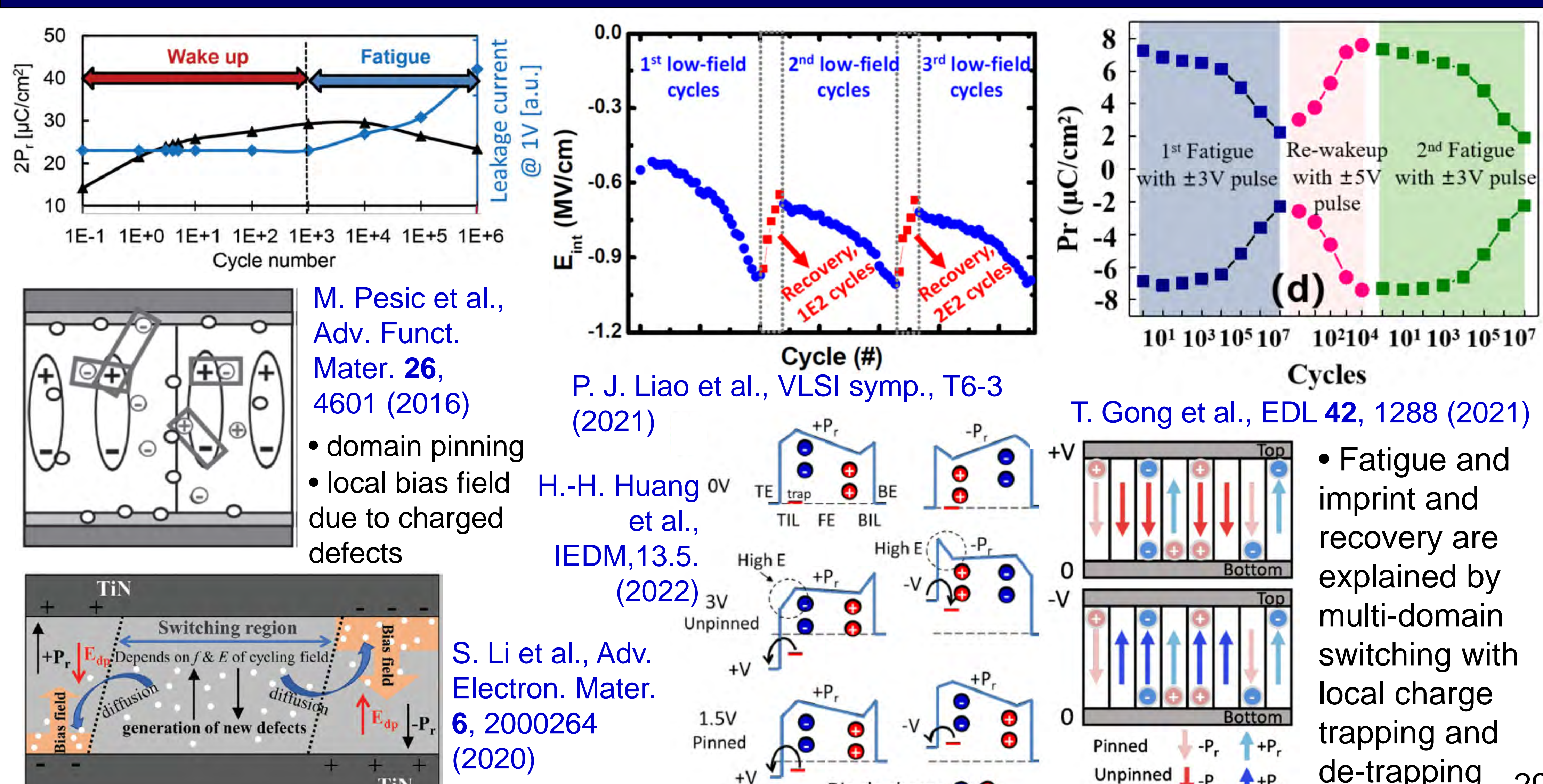
Fatigue and recovery of HZO MFM capacitors at low electric field



K. Toprasertpong et al., ACS Appl. Mater. Interfaces **14**, 51137 (2022)

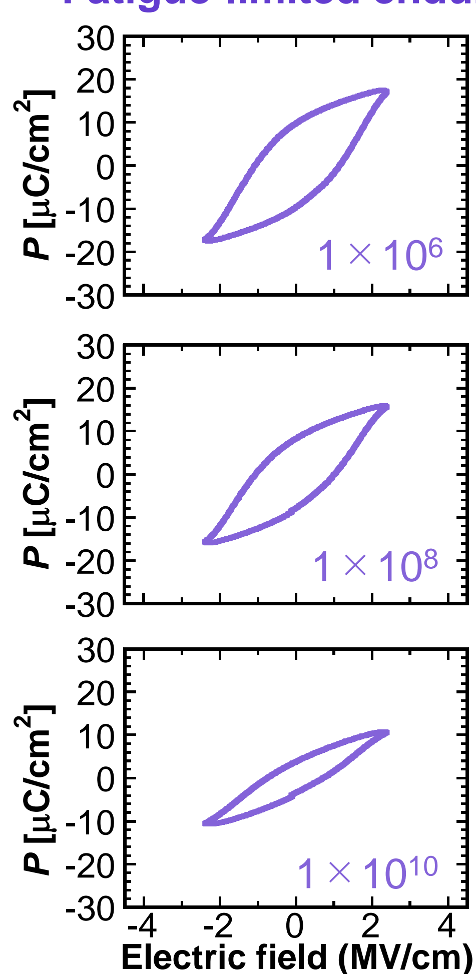
- Operating at a low field is found to cause another failure mode caused by fatigue
- At a low operating field, 2Pr values decrease with increasing the number of cycles
- This fatigue can be recovered by applying high operating voltage (electric field)
- This fatigue behavior is attributable to the influence of charge redistribution in the FE-HfO_2 films and eventually results in a read failure

Possible mechanism of fatigue and recovery of HZO MFM capacitors

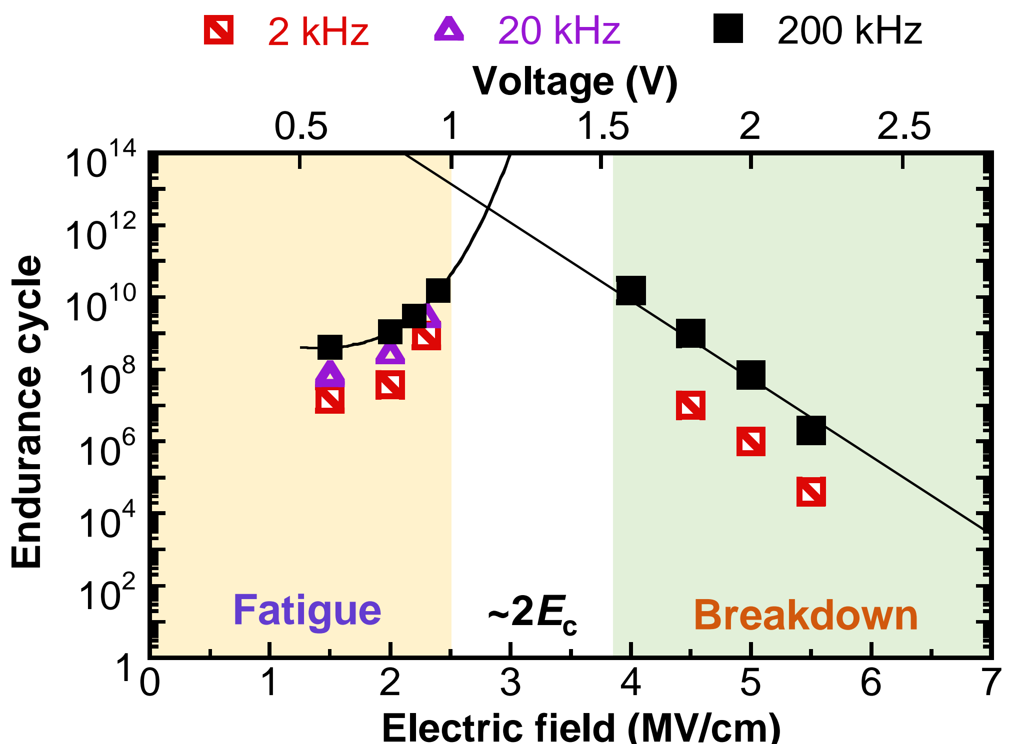


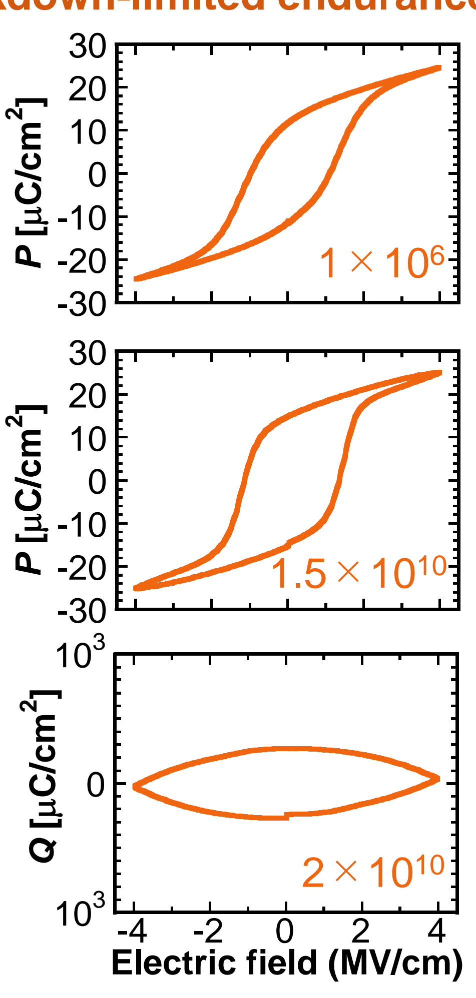
• Fatigue and imprint and recovery are explained by multi-domain switching with local charge trapping and de-trapping

Optimum electric field of HZO MFM capacitors for endurance



K. Toprasertpong et al., ACS Appl. Mater. Interfaces **14**, 51137 (2022)



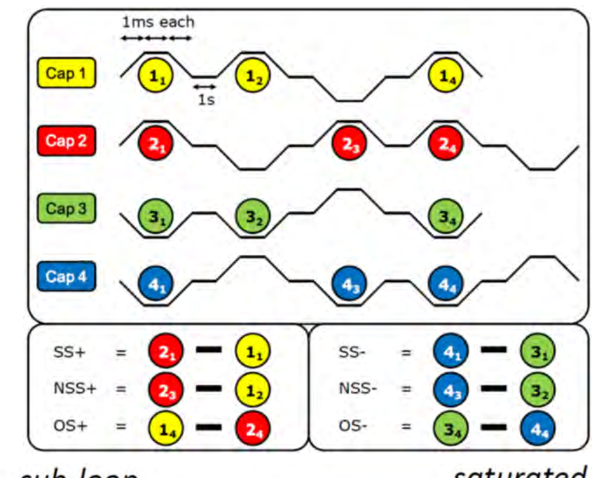


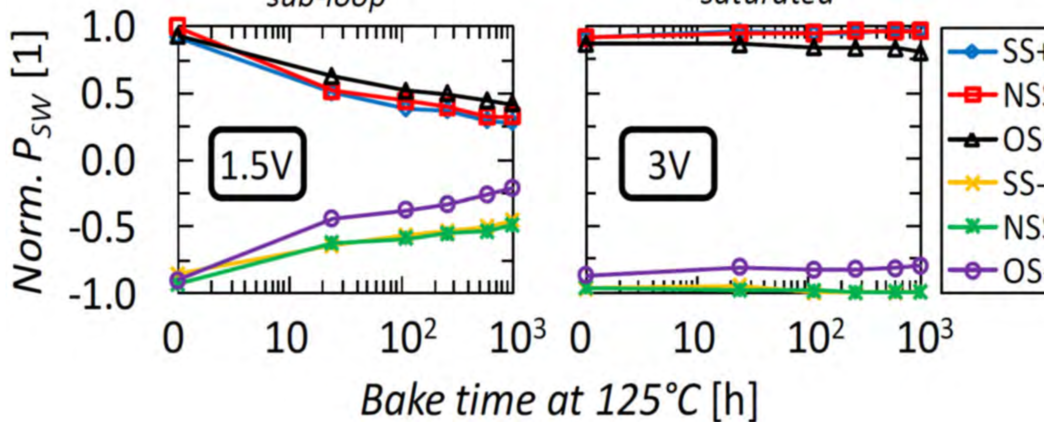
- There is an optimum voltage (field) regime for improving the endurance characteristics. The optimum field is around 3 MV/cm or around $2E_c$, indicating that the operating voltage of 1.2 V (3 MV/cm) for 4.0 nm HZO films

Reliability issues (2) in HZO MFM ~ retention and imprint

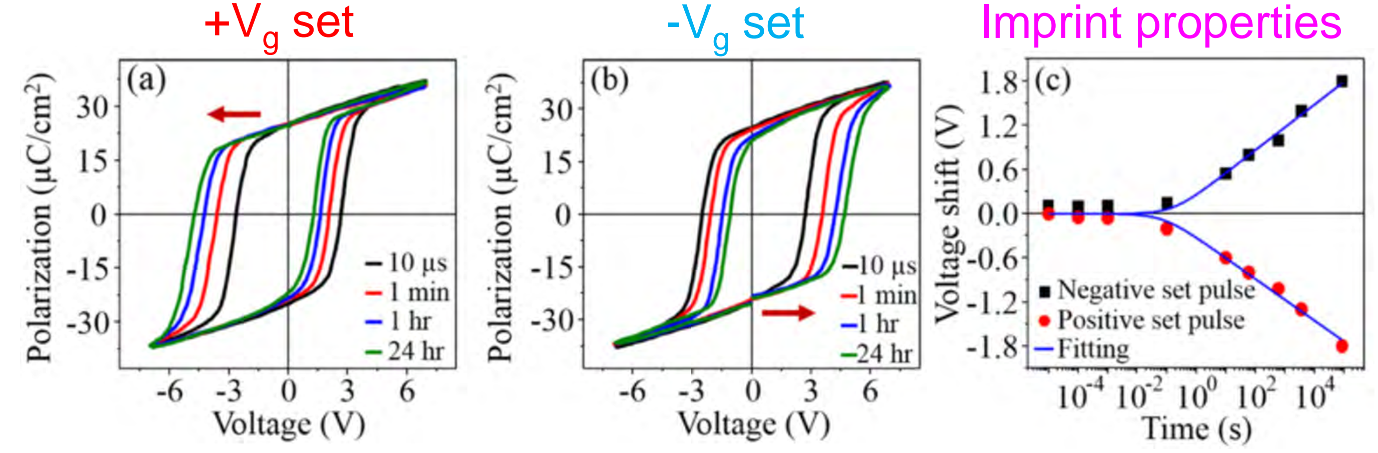
- Complex (FRAM standard) retention test to account for same (SS), new same state (NSS) and opposite state (OS) retention

S. Müller et al., TDMR 99 (2013)





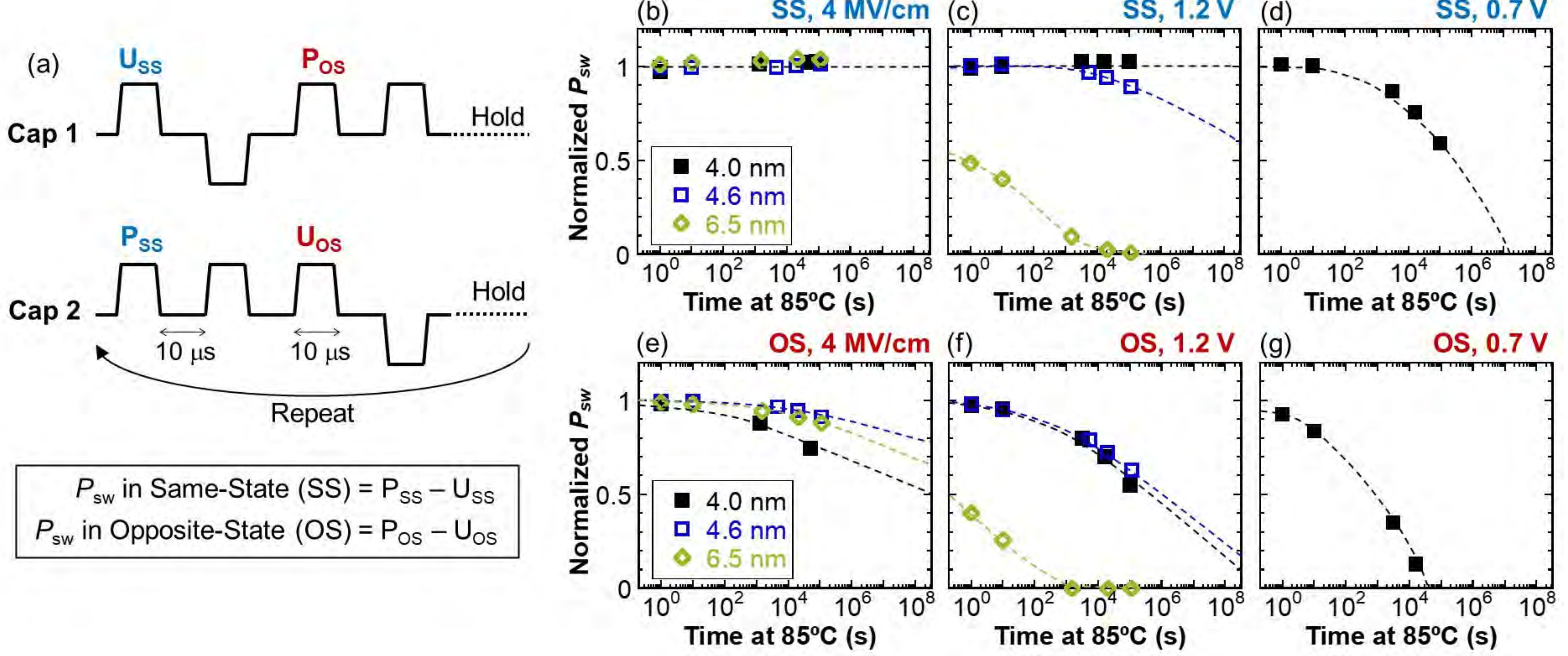
P. Buragohain et al., ACS Applied Mat. & Int. 11, 35115 (2019)



- Good high-T data retention for same and opposite state is commonly observed for FE HfO_2 MFM capacitors
- Sub-loop operation makes retention unstable
- Worse retention for opposite state is a problem, which is attributable to imprint properties of HfO_2-based MFM
- Comparatively large imprint and fully recovery are also commonly observed for HfO_2-based MFM, which can be one possible critical reliability issue

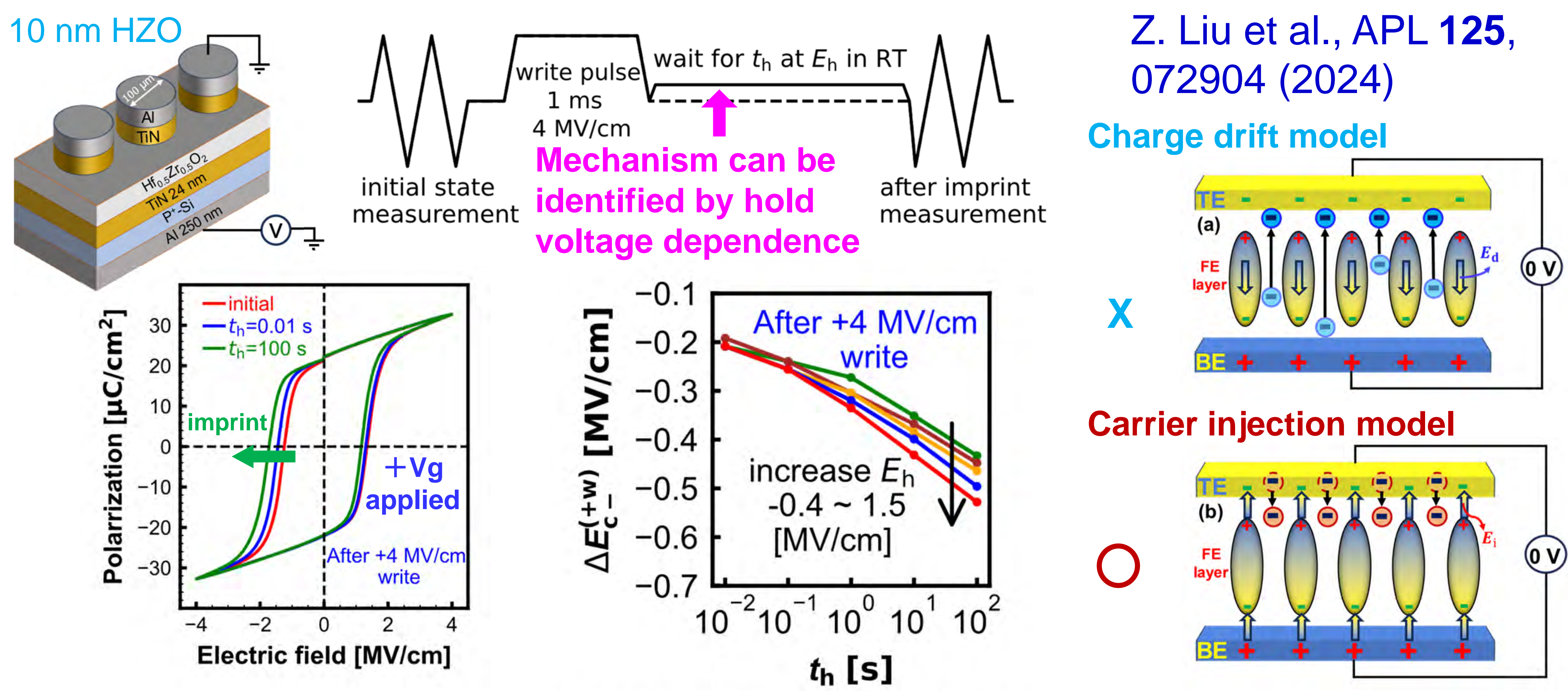
Retention properties of thin HZO MFM capacitors

K. Toprasertpong et al., ACS Appl. Mater. Interfaces **14**, 51137 (2022)



- 4-nm-thick HZO MFM exhibits excellent same state (SS) retention properties even at 1.2 V (3 MV/cm)
- On the other hand, the opposite state (OS) retention is weaker than the SS retention, attribute to imprint properties of HZO films, which need to be further improved

Physical mechanism of imprint in HZO

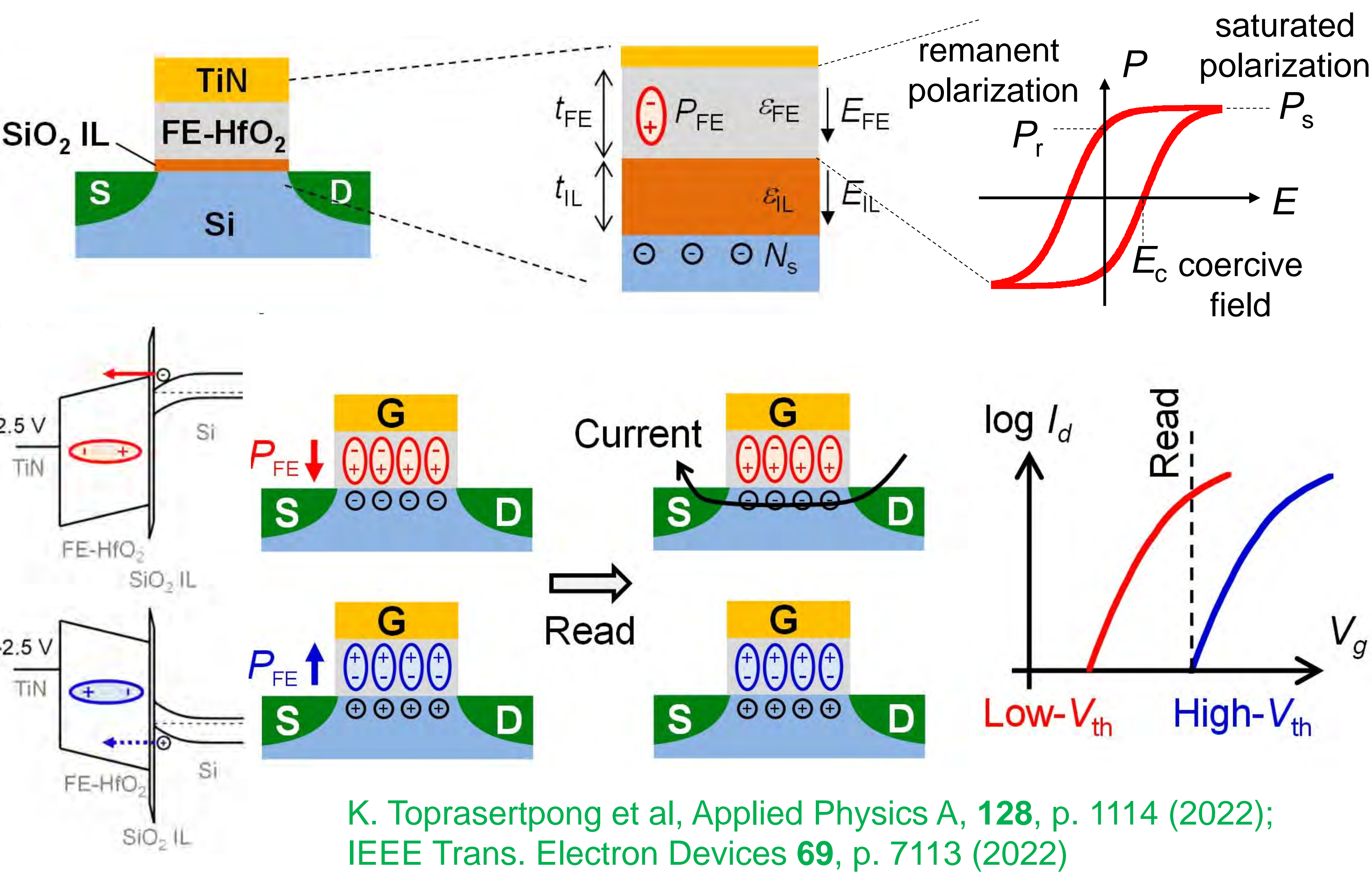


• Imprint in HZO films can be caused by charge generation due to injection (or emission) of carriers from (or into) electrodes (carrier injection model)
(This model has also been supported by P. Vishnumurthy, IRPS 7A.1 (2024))

FeFET memory

FeFET memory operation

Ferroelectric gate insulator Field-Effect Transistors (FeFET)

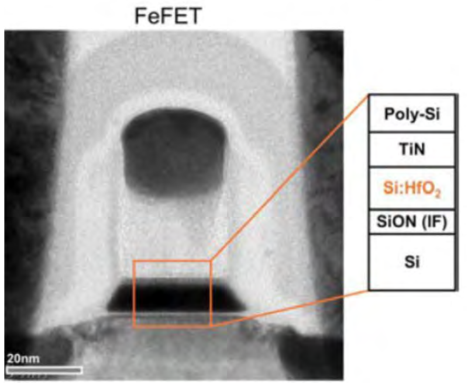


- Polarization condition of FE can control the electric field at MOS interface and resulting V_{th} of I_d-V_g characteristics of MOSFETs
- Nonvolatile memory function is obtained by reading out V_{th} condition with low applied V_g
- Non-destructive readout expected

K. Toprasertpong et al, Applied Physics A, **128**, p. 1114 (2022); IEEE Trans. Electron Devices **69**, p. 7113 (2022)

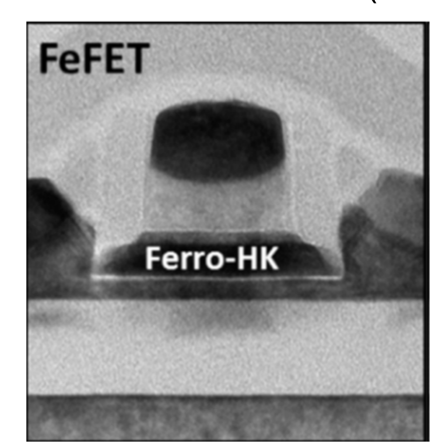
Recent Progress of HZO devices – scaled FeFETs and advanced structures

28nm node FeFET (GF)



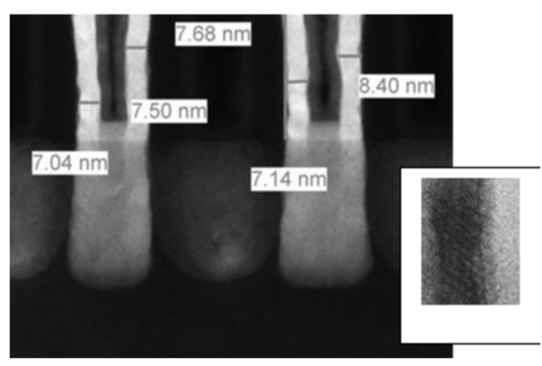
M. Trentzsch, IEDM (2016)

22nm node FDSOI (GF)



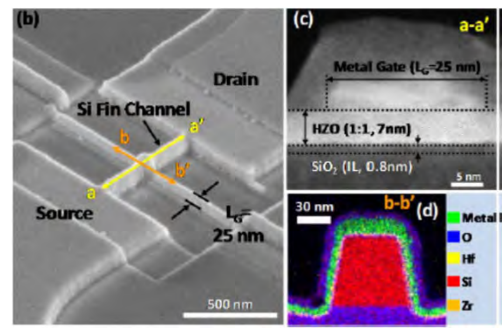
S. Dünkel, IEDM (2017)

14nm node FinFET (GF)



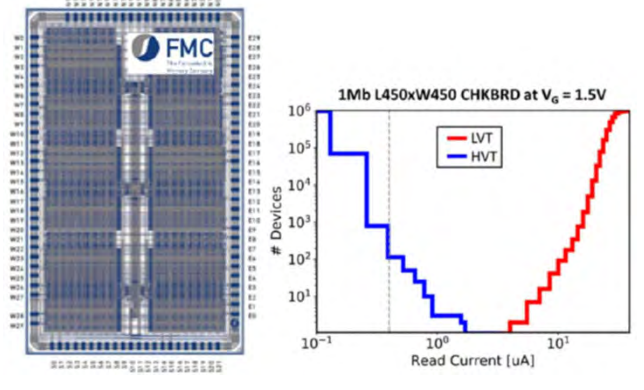
Z. Krivokapic, IEDM (2017)

25nm FinFET (Samsung)



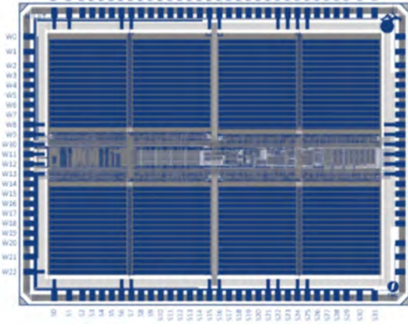
H. Bae, IEDM (2020)

10Mb cell array (GF, FMG)



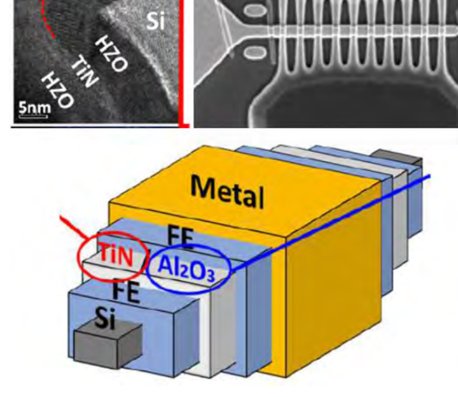
S. Beyer, IMW (2020)

32Mb produce macro (FMG, GF)



S. Mueller, VL (2021)

GAA FeFET (NTNU)



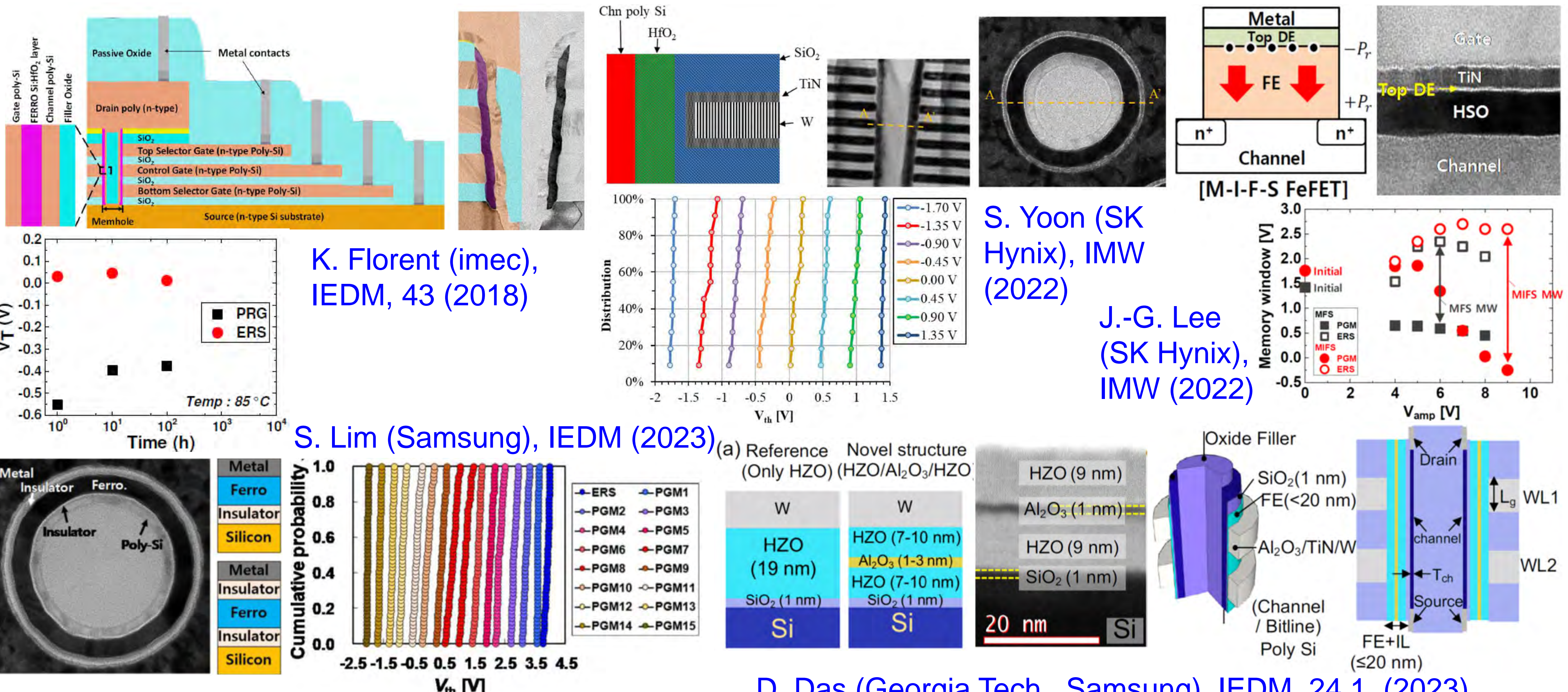
C.-Y. Liao, VL (2022)

FeFET on IGZO (Peking U.)

Source
Top gate
L_{ch} = 40 nm
Drain
200 nm
Gate
Pt/Au
FE
S
D
IGZO
Buffer layer
Si substrate

M. Zeng, IEDM (2023)

Recent Progress of HZO devices – application to FLASH memory cell

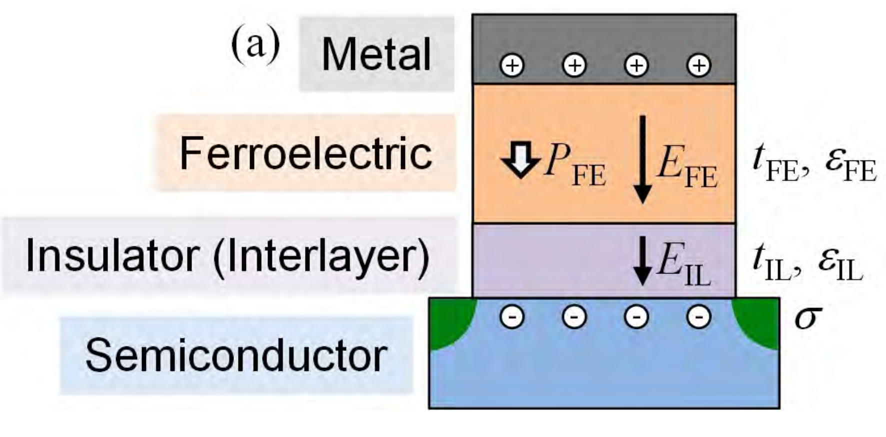


- Application of the FeFET structure to NAND flash memory cell by replacing charge-trap layers by ferroelectric films is actively being studied → one of the most promising and near-term applications

Idealistic value of memory window in FeFET

K. Toprasertpong et al, IEEE TED **69**, 7113 (2022)

- When sufficient electric field is applied across FE, below P_r and E_c match those in major loop P-E curve



(b)

$P = P_{FE} + \varepsilon_{FE}\varepsilon_0 E_{FE}$

P_r

$-\sigma_0$

E_{FE}

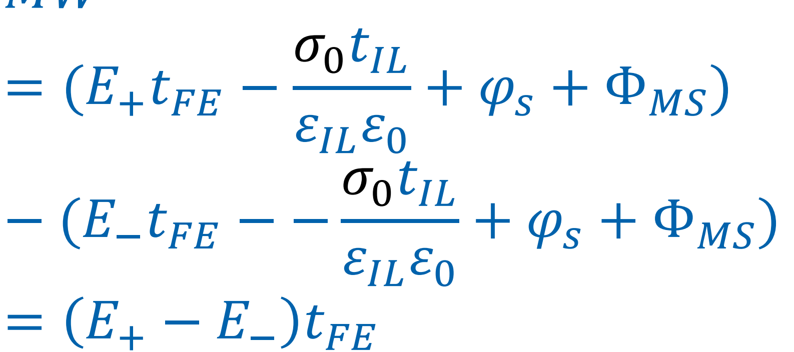
E_- E_+

Field determining MW

η : form factor of P-E curve

$$P_{FE\pm}(E_\pm) + \varepsilon_{FE}\varepsilon_0 E_\pm = P_s \tanh\left(\frac{\eta(E_\pm \mp E_c)}{E_c}\right) + \varepsilon_{FE}\varepsilon_0 E_\pm = \sigma_0 \approx 0$$

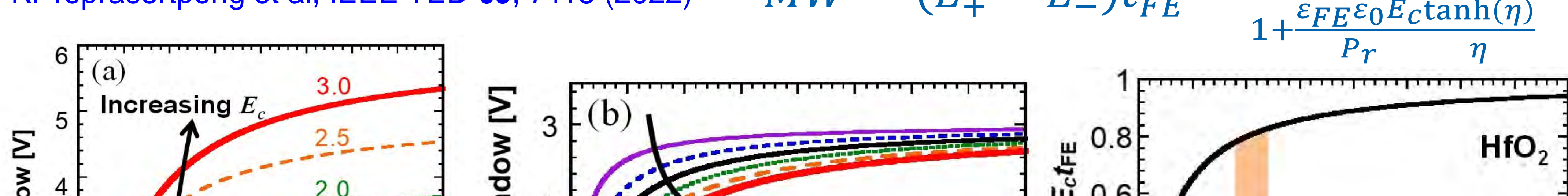
$$E_\pm \approx \pm \frac{E_c}{1 + \frac{\varepsilon_{FE}\varepsilon_0 E_c}{P_r}\frac{\tanh(\eta)}{\eta}}$$



$$MW = (E_+ - E_-)t_{FE} = \frac{2E_c t_{FE}}{1+\frac{\varepsilon_{FE}\varepsilon_0 E_c}{P_r}\frac{\tanh(\eta)}{\eta}}$$

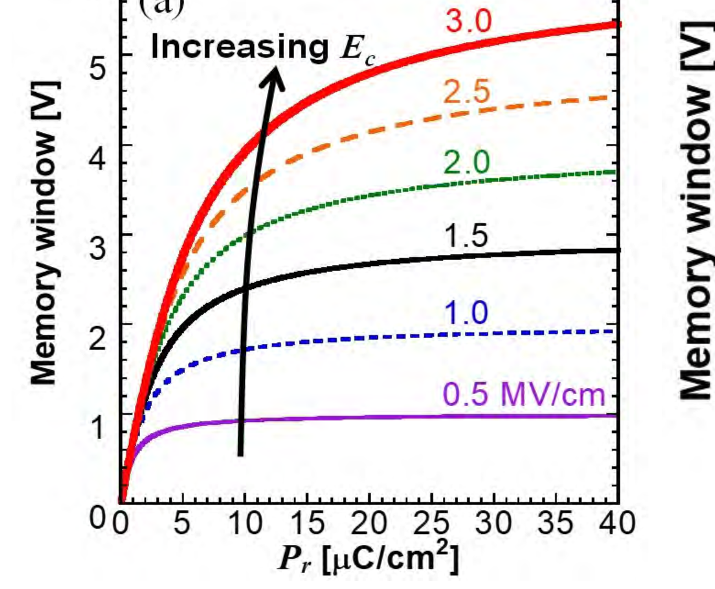
- Under the idealistic condition, maximum memory window is given by $2E_c t_{FE}$

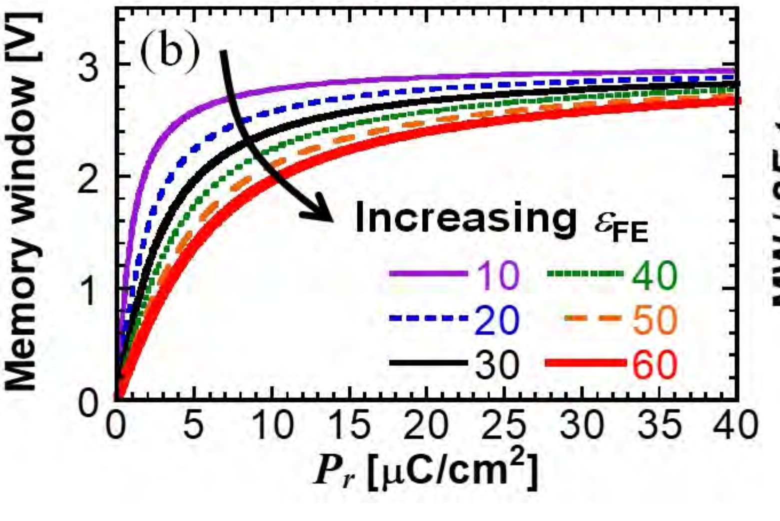
For increasing memory window of FeFET

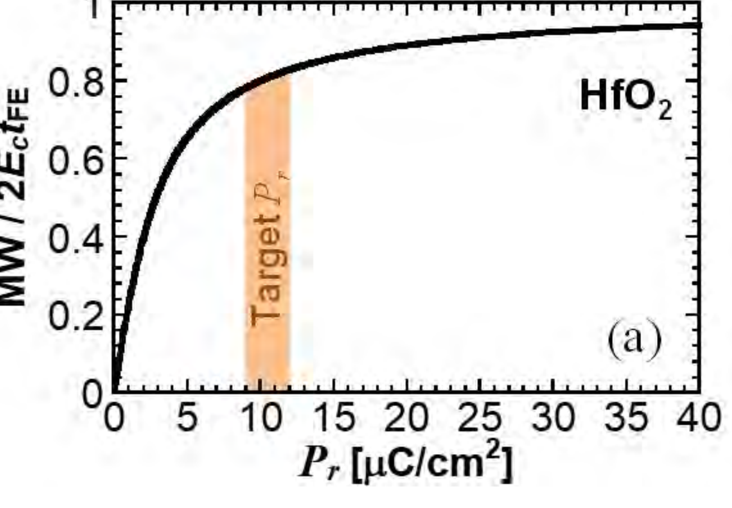


K. Toprasertpong et al, IEEE TED **69**, 7113 (2022)

$$MW = (E_+ - E_-)t_{FE} = \frac{2E_c t_{FE}}{1+\frac{\varepsilon_{FE}\varepsilon_0 E_c}{P_r}\frac{\tanh(\eta)}{\eta}}$$







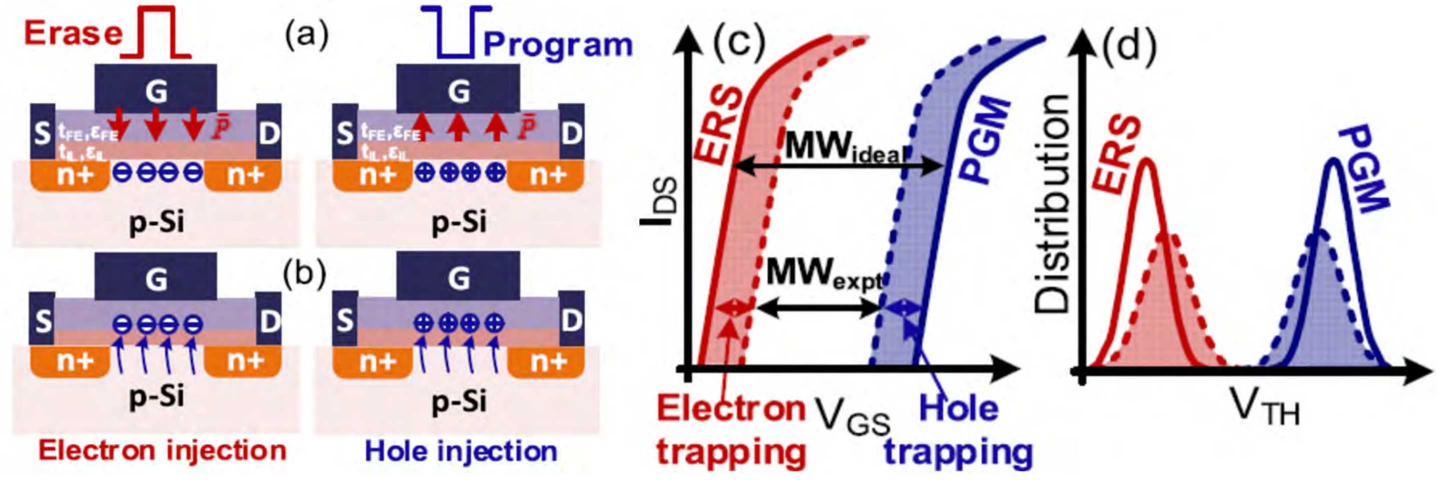
- higher E_c
- lower ε_{FE}
- $P_r > 3\,\varepsilon_{FE}\varepsilon_0 E_c$ ~ 10 μC/cm²

- HfO_2-based ferroelectrics with higher E_c and lower ε_{FE} are suitable materials than conventional Perovskite ferroelectrics such as PZT and SBT

Importance extrinsic factors to reduce memory window

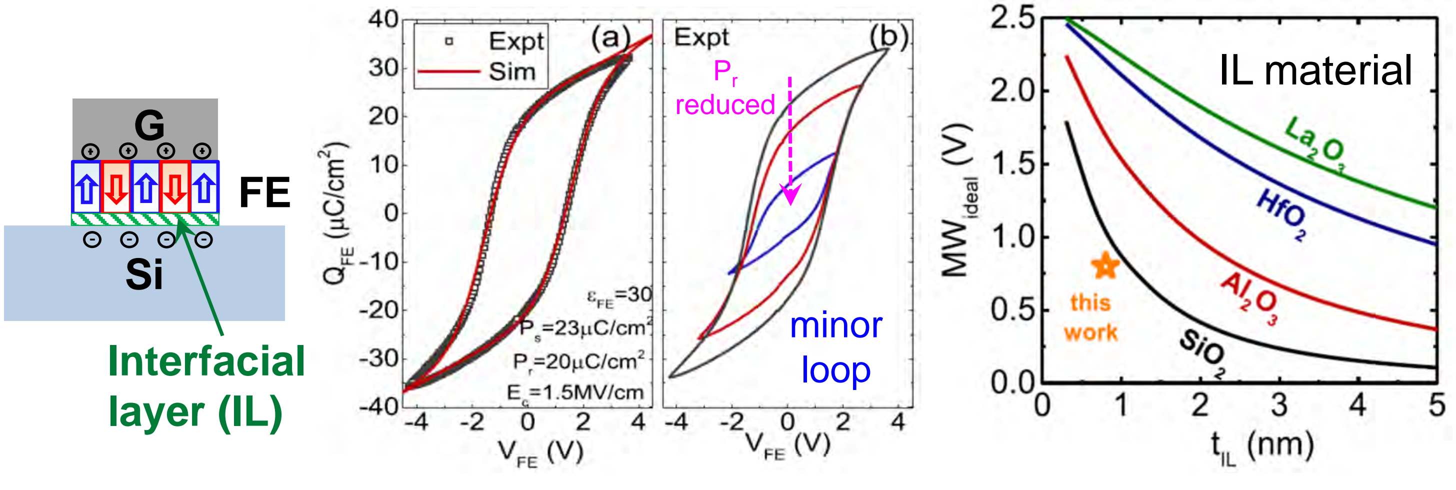
(1) Influence of trapped electron and holes

K. Ni et al., TED **65** (2018) 2461



- Electrons and holes trapped during erase and program operation reduce memory window

(2) Reduction in E_{FE} and P_r due to interfacial layer (IL)



- Electric field and voltage across FE) and IL are determined by capacitance balance between FE and IL, when amounts of trapped charges are small
- Thicker IL and lower permittivity reduce the memory window

Understanding of coupling between polarization and carrier traps in FeFET

Importance of MFIS interfaces for FeFET operation

Coupling between polarization, traps and inversion charges

electric field E across FE
polarization P
trap charge Q_{trap}
inversion layer charge Q_s

Q_s is induced by P
⇒ a part of Q_s is trapped (Q_{trap})
⇒ Q_{trap} changes $E(x)$ across FE
⇒ P changes by $E(x)$ across FE
※ **complicated characteristics**

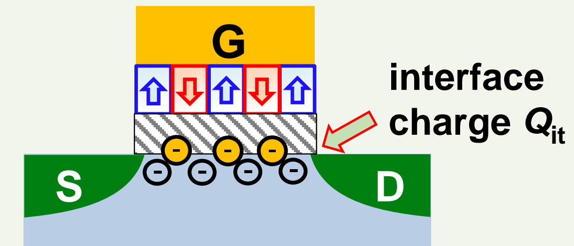
MFIS interface (MOS interface)

M : Metal
F : Ferroelectric
I : Insulator
S : Semiconductor

Basic question:
- typical P ~ 10~20 μC/cm²
→ $N_s = \sigma_s / q > 10^{14}$ cm^{-2}
→ too high N_s (typical N_s of MOSFET = 10^{11}~10^{13} cm^{-2})
→ $E_{IL} = \sigma_s / \varepsilon_{IL} > 30$ MV/cm
→ too high E_{IL}

K. Toprasertpong et al., Applied Physics A, 128 (2022) 1114

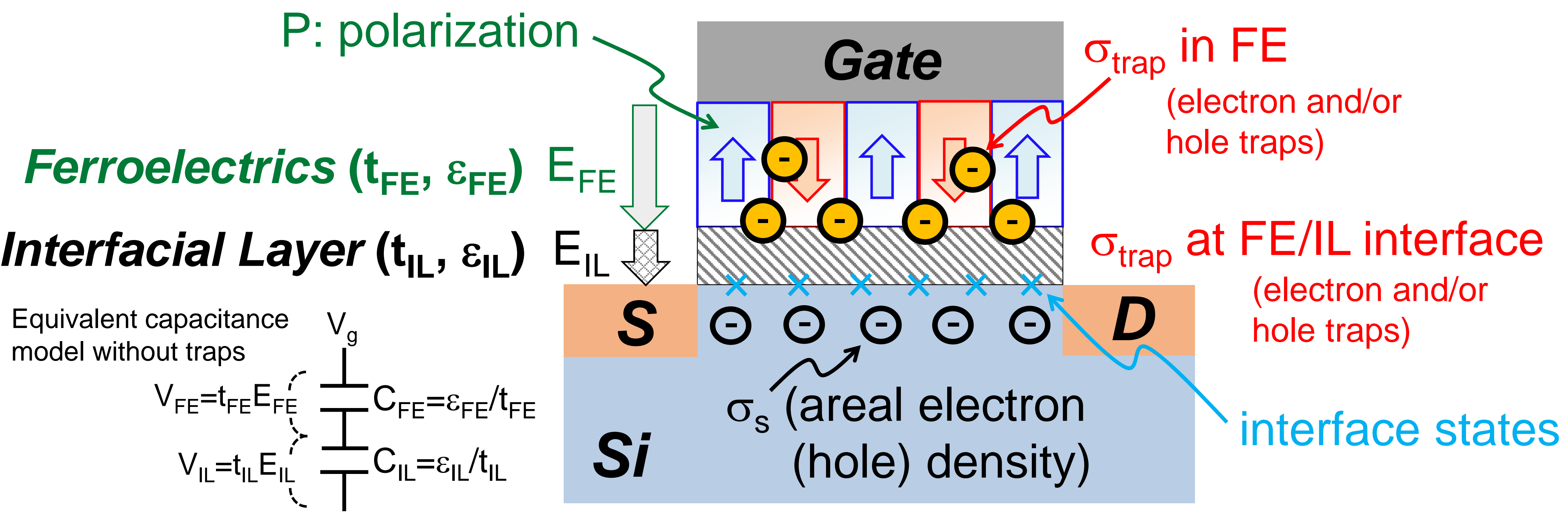
Interface state D_{it}



☐ affect SS and mobility
※ **degrade FET performance**
(However, they do not cause complicated coupling with polarization, because $E(x)$ across IL does not change at same surface potential with different P)

Understanding MFIS interface and carrier trapping properties is critical to FeFET operation and characteristics

Issues of ferroelectric gate stacks and the impact on FeFET

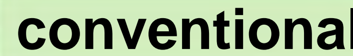


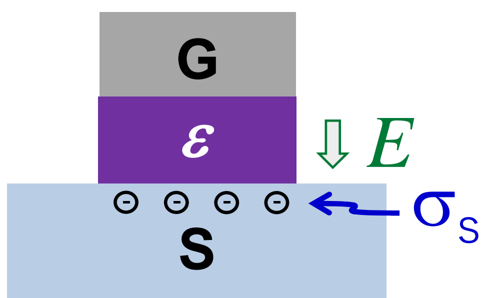
- When traps can be ignored, electric field across ferroelectrics (E_{FE}) and electric field across interfacial layers (E_{IL}) are determined by the capacitance ratio of FE and interfacial layers (IL)
- Here, thinner thickness and higher permittivity of IL lead to higher E_{FE}
- When a large amount of traps are included, we need to take into account the charge balance between polarization, inversion-layer charges, trap charges and interface state charges, which can strongly affect E_{FE} and E_{IL}

Significant influence of traps on FeFET operation

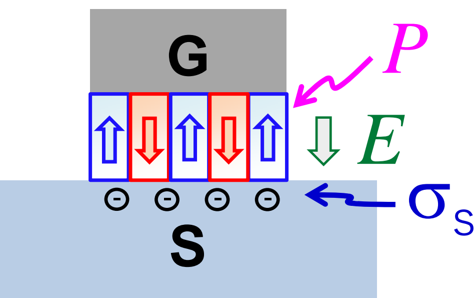
Under inversion condition (MOSFET operation condition)

K. Toprasertpong et al., IEDM (2019) 570

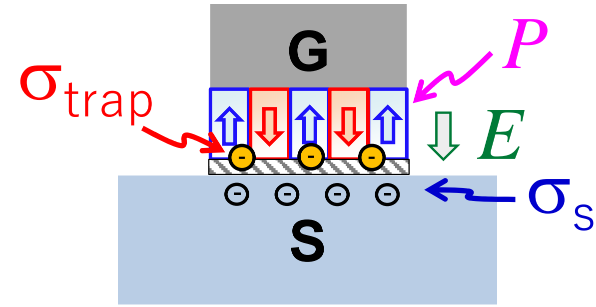




Ferro gate (ideal)



Ferro gate (with traps)



$\varepsilon E = \sigma_s$

$\varepsilon E + P = \sigma_s$

$\varepsilon E + P = \sigma_s + \sigma_{trap}$

P: polarization

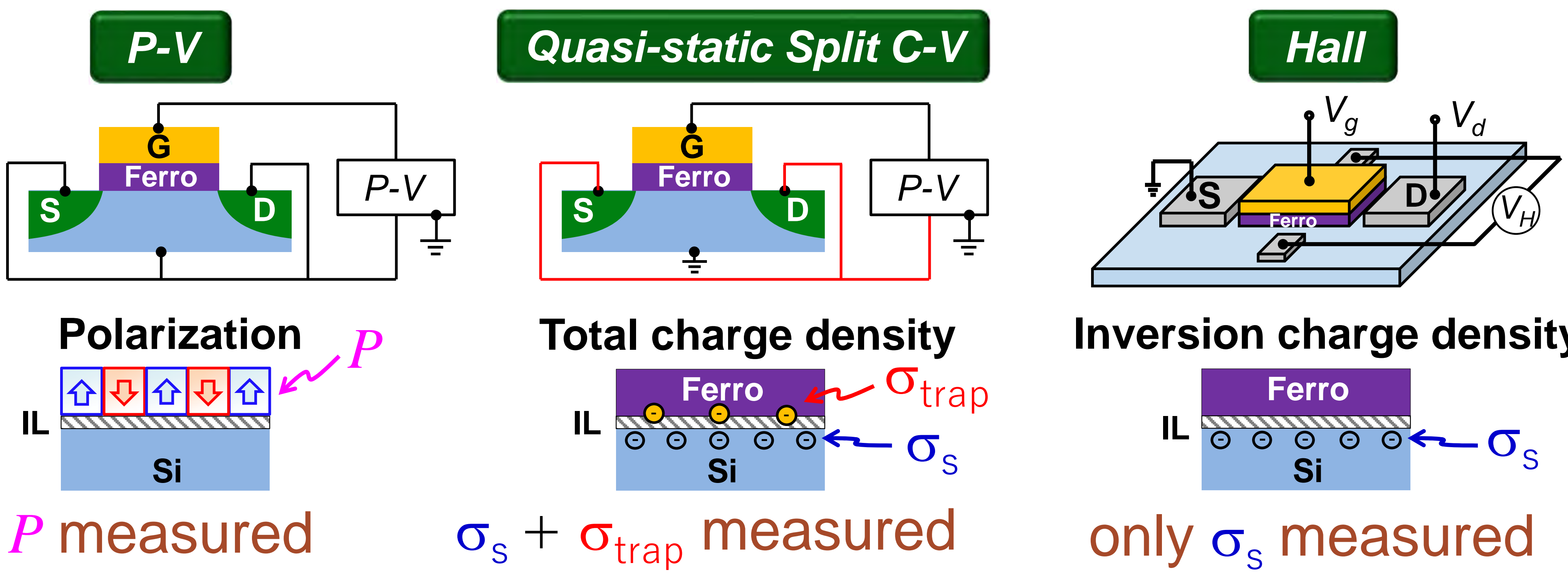
σ_s : areal density of inversion charges

σ_{trap} : areal density of trapped charges

Quantitative discrimination of polarization, inversion charges and trapped charges is quite important to understand FeFET operation

Evaluation method to quantify each component

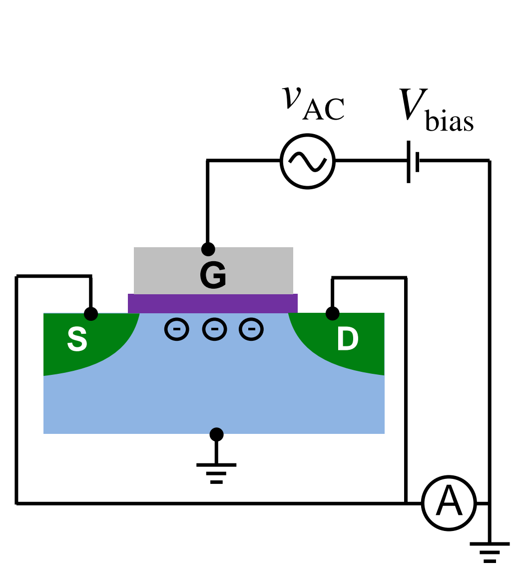
K. Toprasertpong et al., IEDM (2019) 570, VLSI Symp. (2020)TF1.5

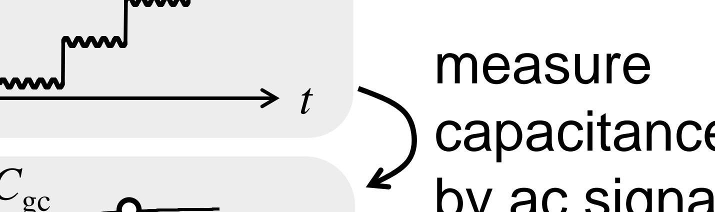


P measured $\quad$ $\sigma_S + \sigma_{trap}$ measured $\quad$ only σ_S measured

- We can independently evaluate *P*, σ_S and σ_{trap} by using (1) P-V, (2) quasi-static (QS) split C-V, (3) Hall measurements
- The contribution of traps (σ_{trap}) on FeFET operation can be quantified

Evaluation of areal inversion charge density by conventional split C-V

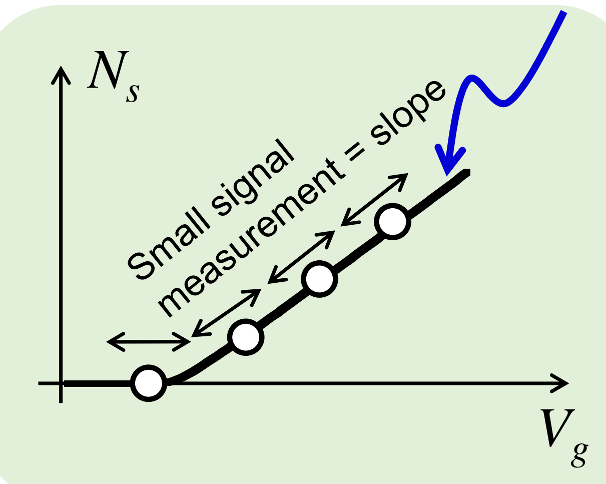




measure capacitance by ac signal

integration of capacitance ⇒ charge density

Relationship to be determined



C_{gc}: gate-channel capacitance

$$\sigma_S = qN_s$$

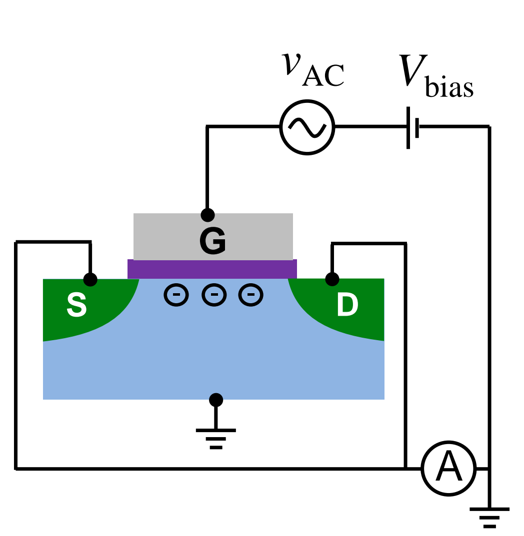
• When traps and interface states can be ignored, N_s corresponds to areal density of free (mobile) inversion carriers

Real procedure for determining relationship between N_s and V_g

(1) Evaluate slope of N_s-V_g characteristics by small signal ac measurement

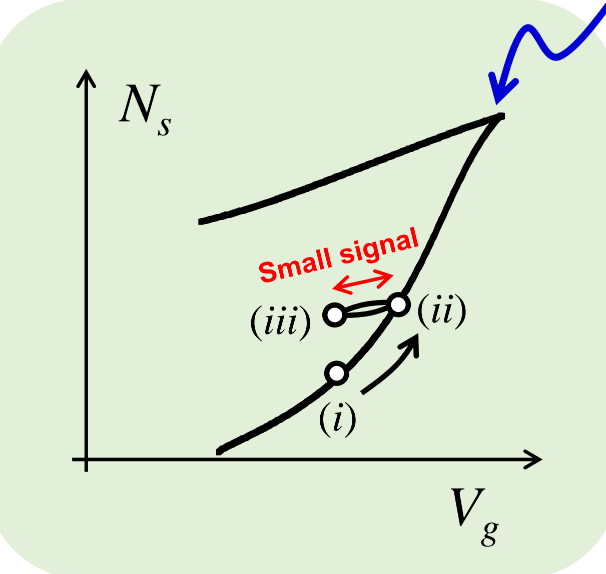
(2) Represent N_s-V_g by connecting the slope at each V_g point

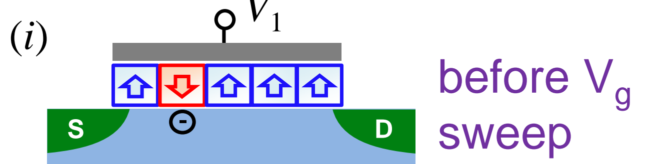
Problem of applying conventional split C-V to FeFET



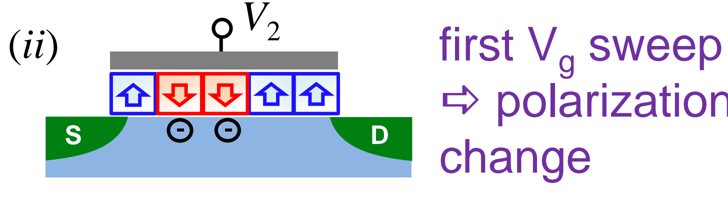
FeFET

Relationship to be determined



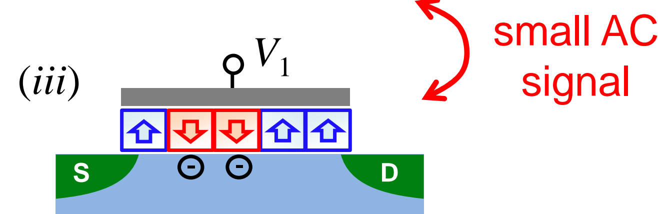


before V_g sweep



first V_g sweep
⇨ polarization change

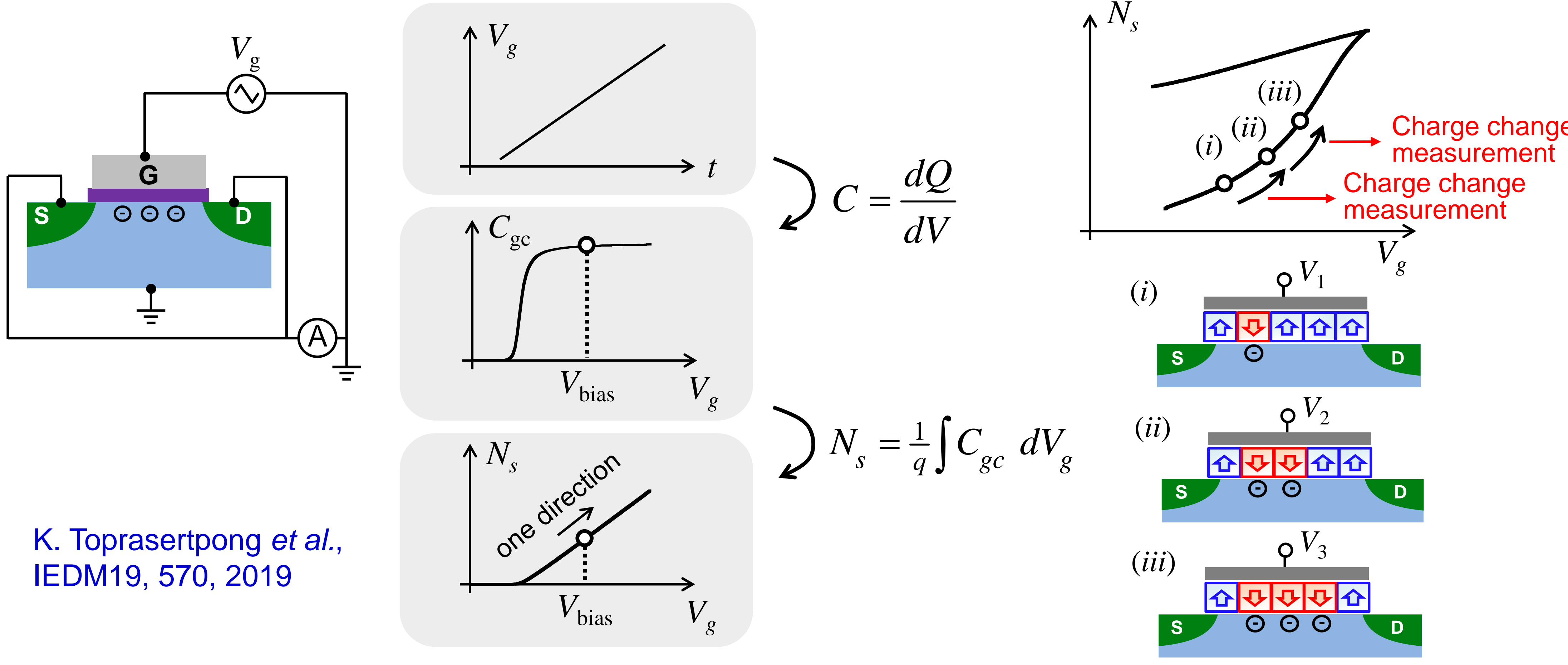
small AC signal



polarization retained
⇨ response of minor-loop
⇨ AC response ≠ slope of N_s -V_g

Conventional split C-V is not applicable to determination of N_s in FeFET

Application of quasi-static split C-V to FeFET



- When a large-signal (one-directional and one-time sweep) C-V measurement is used instead of the small-signal one, the "change in N_s due to polarization reversal" can be correctly evaluated
- Note that this N_s includes channel charges, interface charges and injected/trapped charges

P-V and quasi-static split C-V measurements

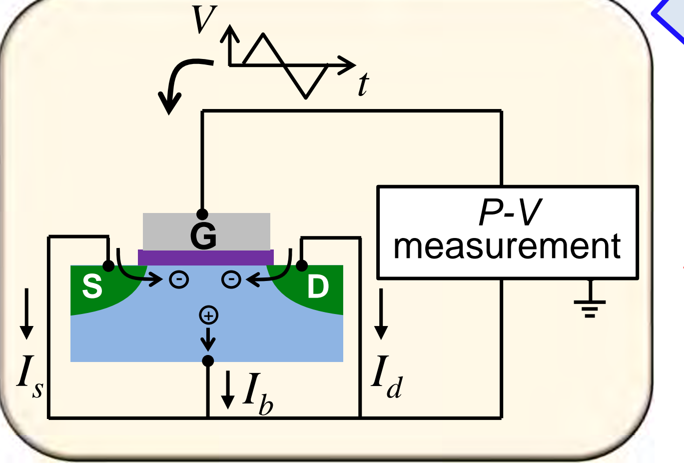
P-V measurement of MFM capacitor

= charge movement during voltage sweep (not P but ΔP measured)

V t *P-V* measurement

$Q = \int I \, dt$

P-V measurement of FeFET

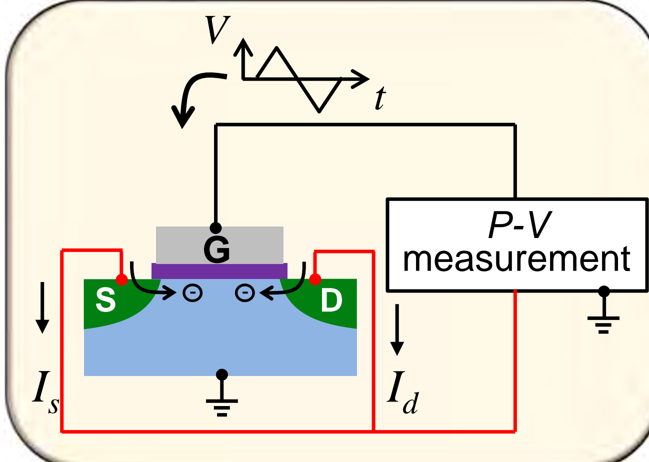


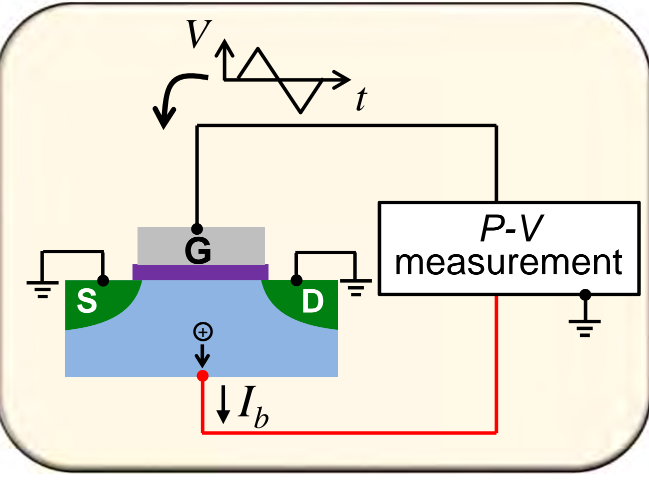
K. Toprasertpong *et al*, Appl. Phys. Lett. **116**, 242903, 2020

S/D current

Split

substrate current



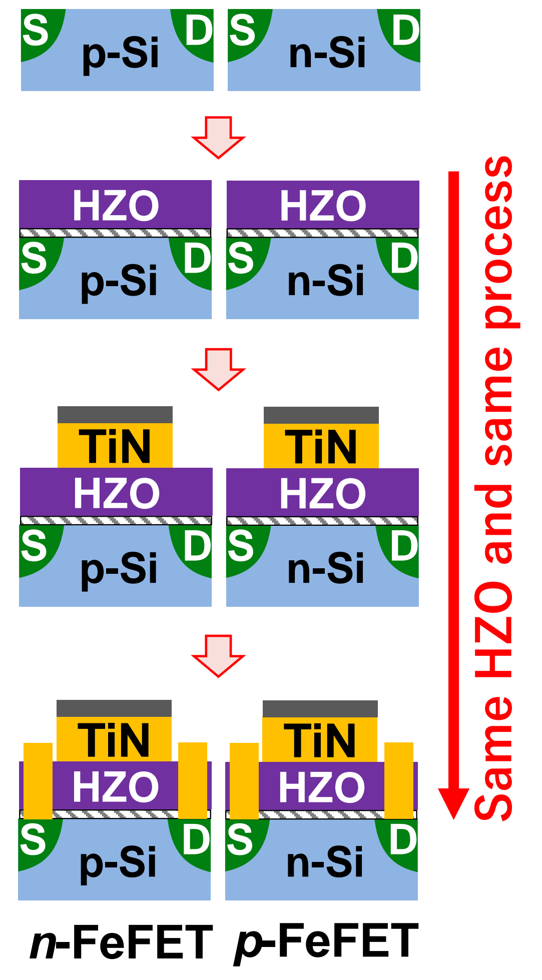


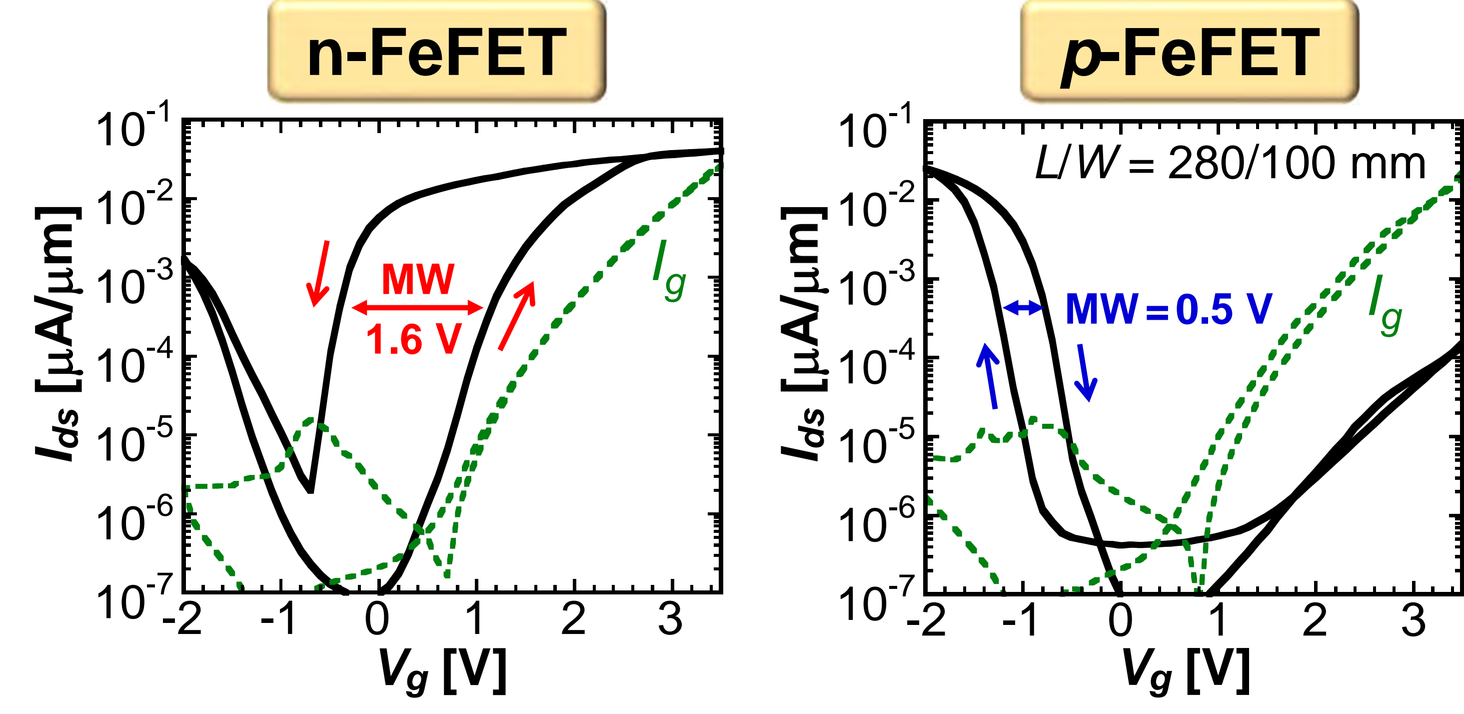
Measurable by standard *P-V* evaluation machines

K. Toprasertpong *et al.*, IEDM19, 570, 2019

I_d-V_g characteristics of n- and p-FeFET with same gate stacks

K. Toprasertpong et al., IEDM (2019) 570, VLSI Symp. (2020)TF1.5





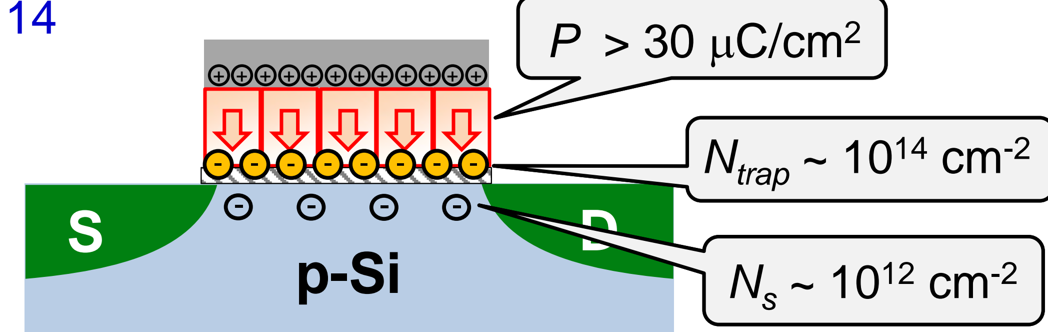
- Under the same HZO and gate stack, memory window of n-FeFET is much higher than that of p-FeFET
- These devices are used for characterization of polarization, inversion carriers and trapped carriers

Asymmetric carrier trapping in positive and negative gate voltage

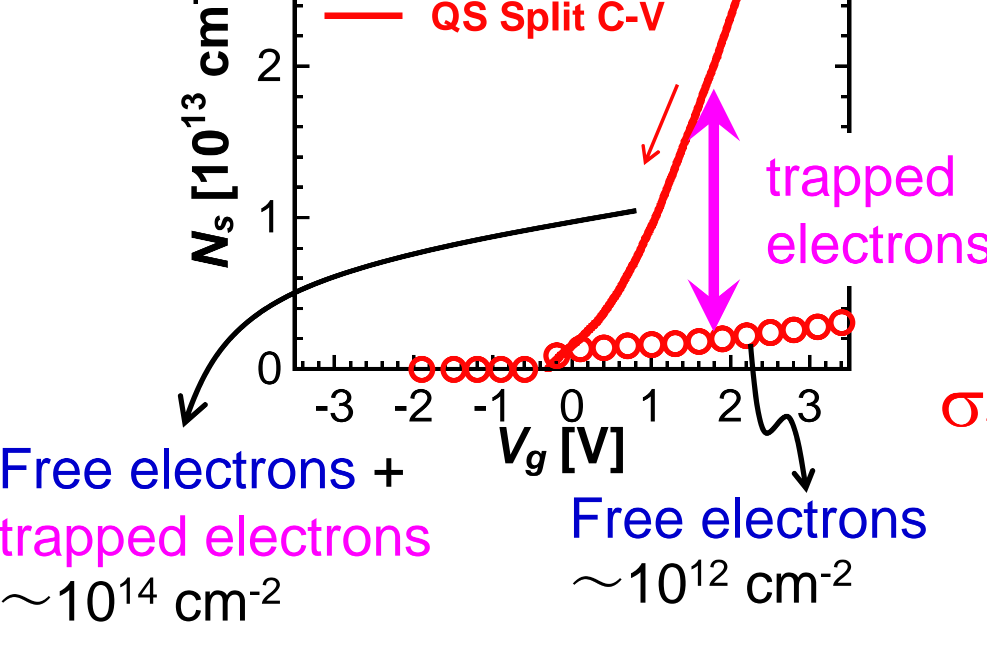
Electron distribution under positive V_g

K. Toprasertpong *et al.*, IEDM19 (2019) 570; VLSI Symp. (2020) TF1.5, Appl. Phys. A, 128 (2022) 1114

Similar large trap density at FE/IL interfaces has been reported independently by R. Ichihara et al. (Kioxia), VLSI symp. (2020)



Electrons in n-FeFET



- Although ferroelectric polarization induces an areal electron density of around 10^{14} cm^{-2}, the induced electrons are mostly trapped in HZO (probably around HZO/IL interface)

Electric field across FE is shielded by large amounts of electron traps

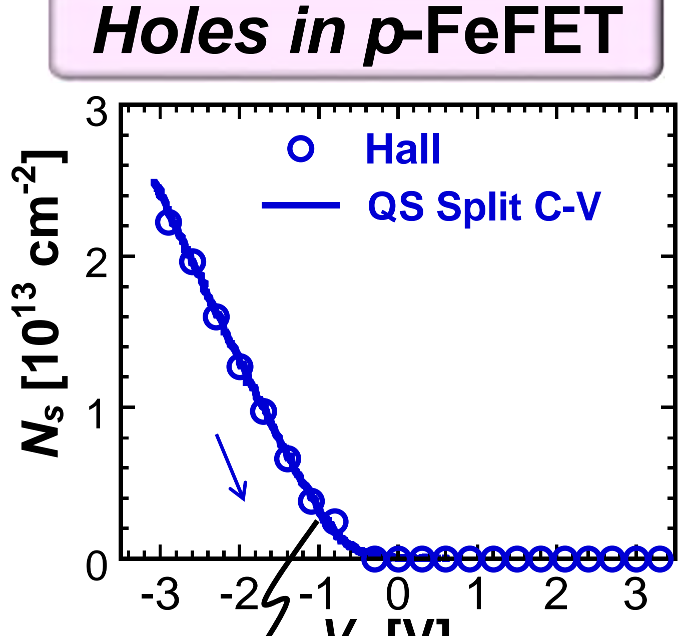
$$\varepsilon E + P = \sigma_s + \sigma_{trap}$$

Asymmetric carrier trapping in positive and negative gate voltage

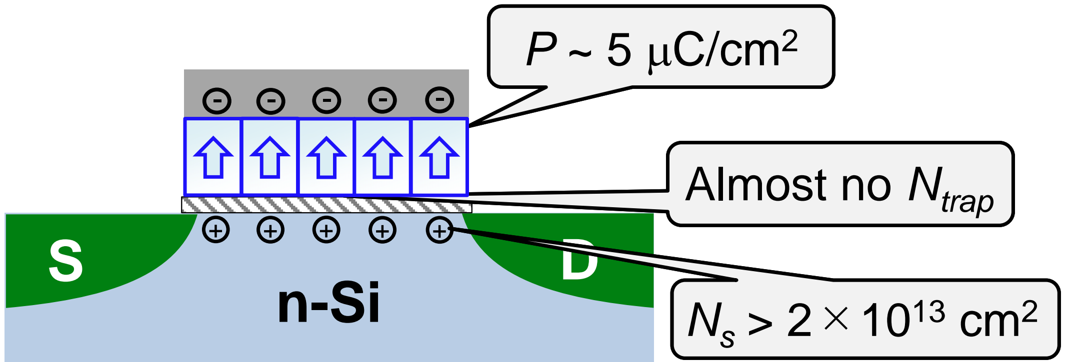
Hole distribution under negative V_g

K. Toprasertpong *et al.*, IEDM19 (2019) 570; VLSI Symp. (2020) TF1.5, Appl. Phys. A, 128 (2022) 1114

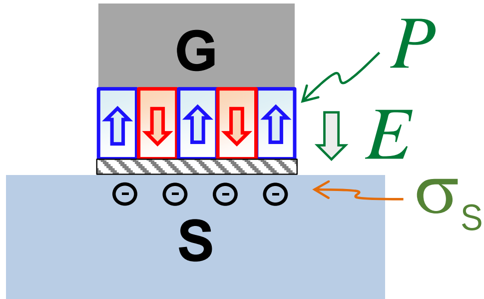
Holes in *p*-FeFET



Free holes + trapped holes ~ Free holes → no significant trapping



- Polarization-induced inversion-layer holes are rarely trapped ⇨ Polarization P = hole areal density N_s, which is much higher than electron one
- P in p-FeFET is smaller than that in n-FeFET

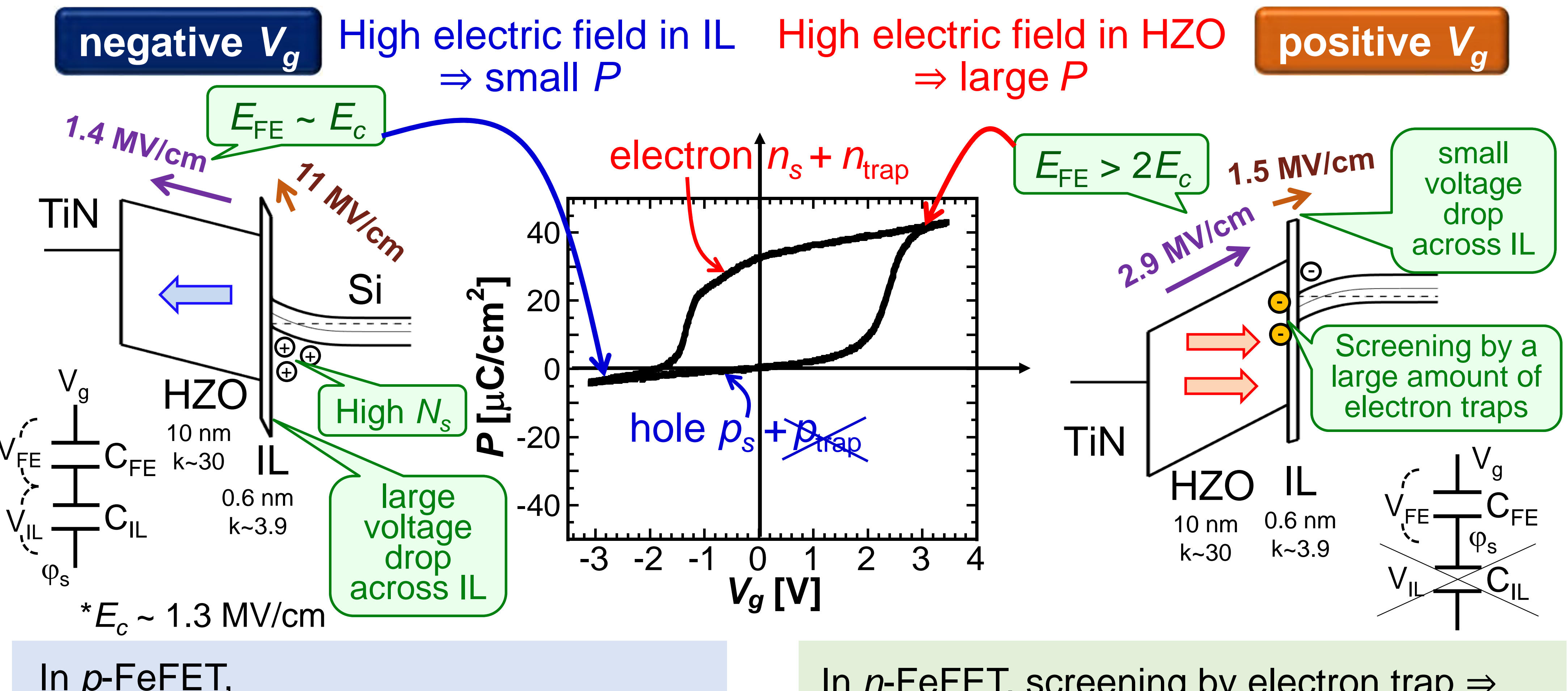


Electric fields across ferroelectric and IL layers are determined by capacitance ratio

$$\varepsilon E + P = \sigma_S$$

Electron-trap-enhanced polarization in n-FeFET

K. Toprasertpong et al., IEDM (2019) 570, VLSI Symp. (2020)TF1.5; Appl. Phys. A, 128 (2022) 1114



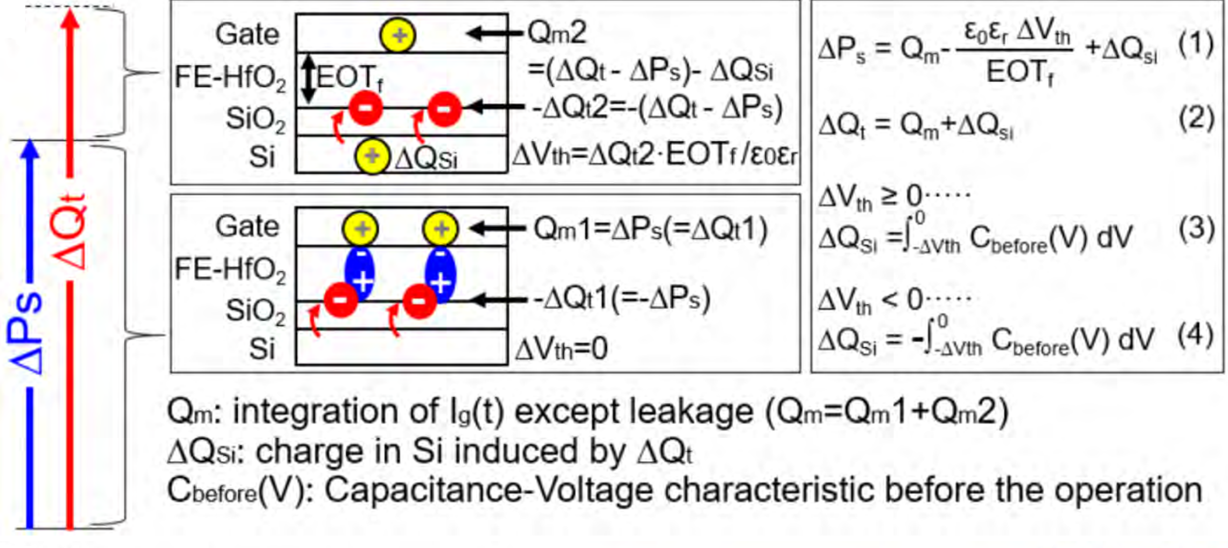
In p-FeFET,
high N_s ⇒ high E_{IL} ⇒ low E_{FE} ⇒ small P

In n-FeFET, screening by electron trap ⇒
high E_{FE} ⇒ large P ⇒ large memory window

Another report on trap-assisted polarization in FeFET

FE-HfO_2/SiO_2/Si FeFET

R. Ichiharaet al., VLSI Symp., (2020), IEDM, 130 (2021)



$$\Delta P_s = Q_m - \frac{\varepsilon_0 \varepsilon_r \Delta V_{th}}{EOT_f} + \Delta Q_{Si} \quad (1)$$

$$\Delta Q_t = Q_m + \Delta Q_{Si} \quad (2)$$

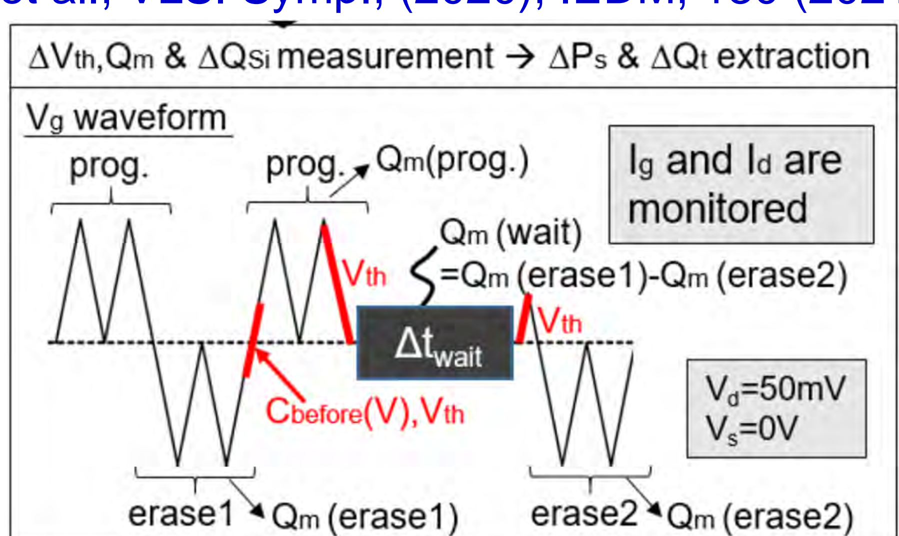
$\Delta V_{th} \geq 0 \cdots\cdots$

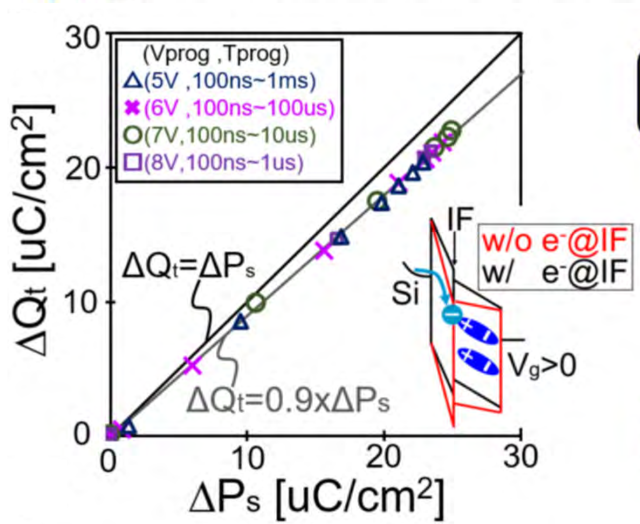
$$\Delta Q_{Si} = \int_{-\Delta V_{th}}^{0} C_{before}(V)\, dV \quad (3)$$

$\Delta V_{th} < 0 \cdots\cdots$

$$\Delta Q_{Si} = -\int_{-\Delta V_{th}}^{0} C_{before}(V)\, dV \quad (4)$$

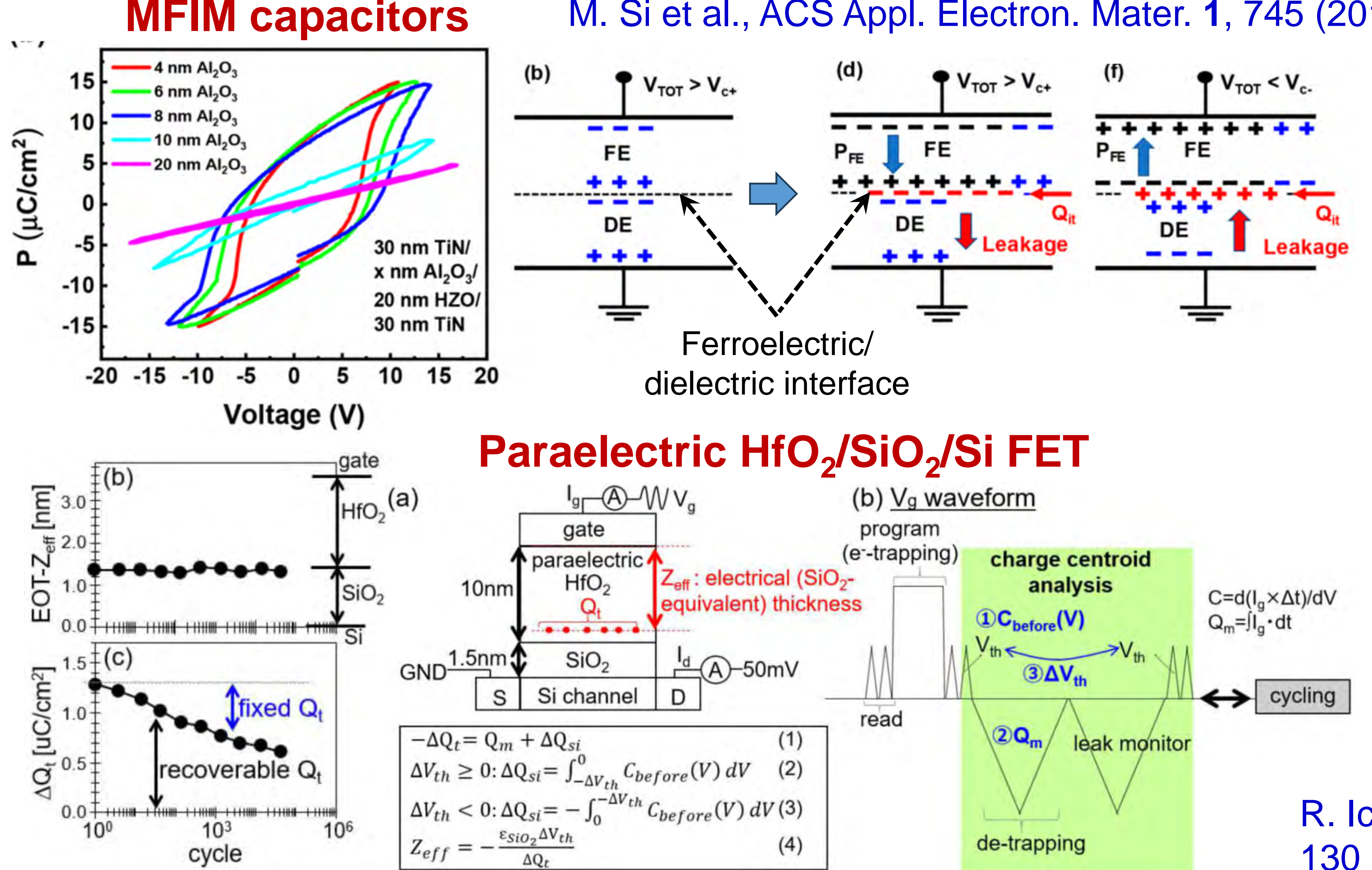
Q_m: integration of $I_g(t)$ except leakage ($Q_m = Q_{m1} + Q_{m2}$)
ΔQ_{Si}: charge in Si induced by ΔQ_t
$C_{before}(V)$: Capacitance-Voltage characteristic before the operation





- Polarization and trapped charges are estimated by monitoring I_g and I_d separately
- Trapped charges near FE/IL interfaces amount to 90 % of polarization

Carrier traps at ferroelectric/IL interface



$$-\Delta Q_t = Q_m + \Delta Q_{si} \quad (1)$$
$$\Delta V_{th} \geq 0: \Delta Q_{si} = \int_{-\Delta V_{th}}^{0} C_{before}(V)\,dV \quad (2)$$
$$\Delta V_{th} < 0: \Delta Q_{si} = -\int_{0}^{-\Delta V_{th}} C_{before}(V)\,dV \quad (3)$$
$$Z_{eff} = -\frac{\varepsilon_{SiO_2}\Delta V_{th}}{\Delta Q_t} \quad (4)$$

- Full polarization switching of an FE layer in an MFIS capacitor is obtained by leakage-current-induced charges at the FE/ultrathin-IL interface

- The trap position estimated from Ig and Id for similar paraelectric $HfO_2/SiO_2/Si$ FETs is almost at the HfO_2/SiO_2 interface

R. Ichihara et al., IEDM, 130 (2021)

Pros and cons of large amounts of electron traps in FeFETs

Impact of large amounts of electron traps around FE/IL interfaces

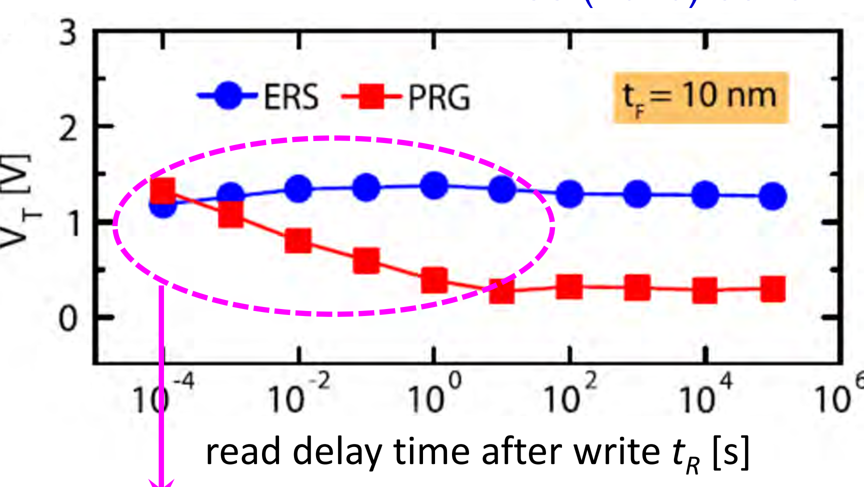
- Cons

- reduction in memory window due to existing traps
- memory window narrowing due to generated traps
- necessity of long read delay time after write

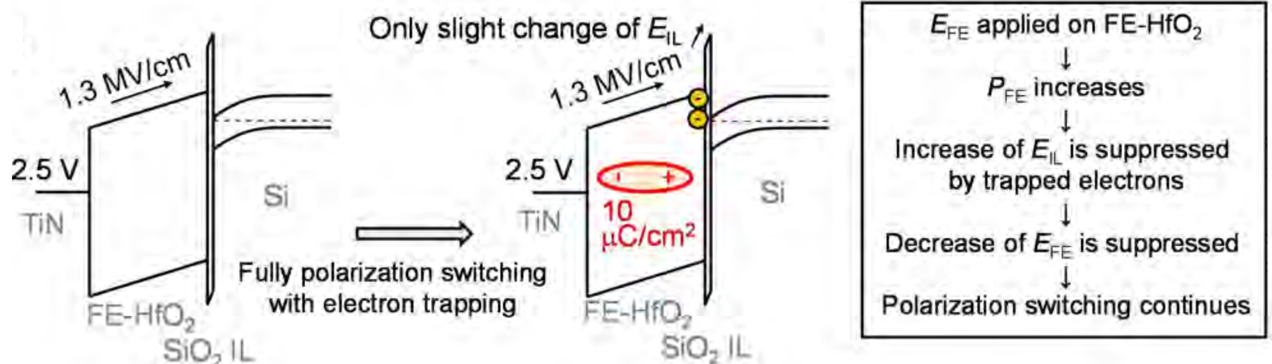
- Pros

- increased electric field across FE (trap-assisted polarization switching)
- mitigation of depolarization field

Mulaosmanovic et al, TED 66 (2019) 3828



lower PRG V_{th} after positive V_g pulse is obtained after sufficient read delay time to induce **electron de-trapping** (explained later)

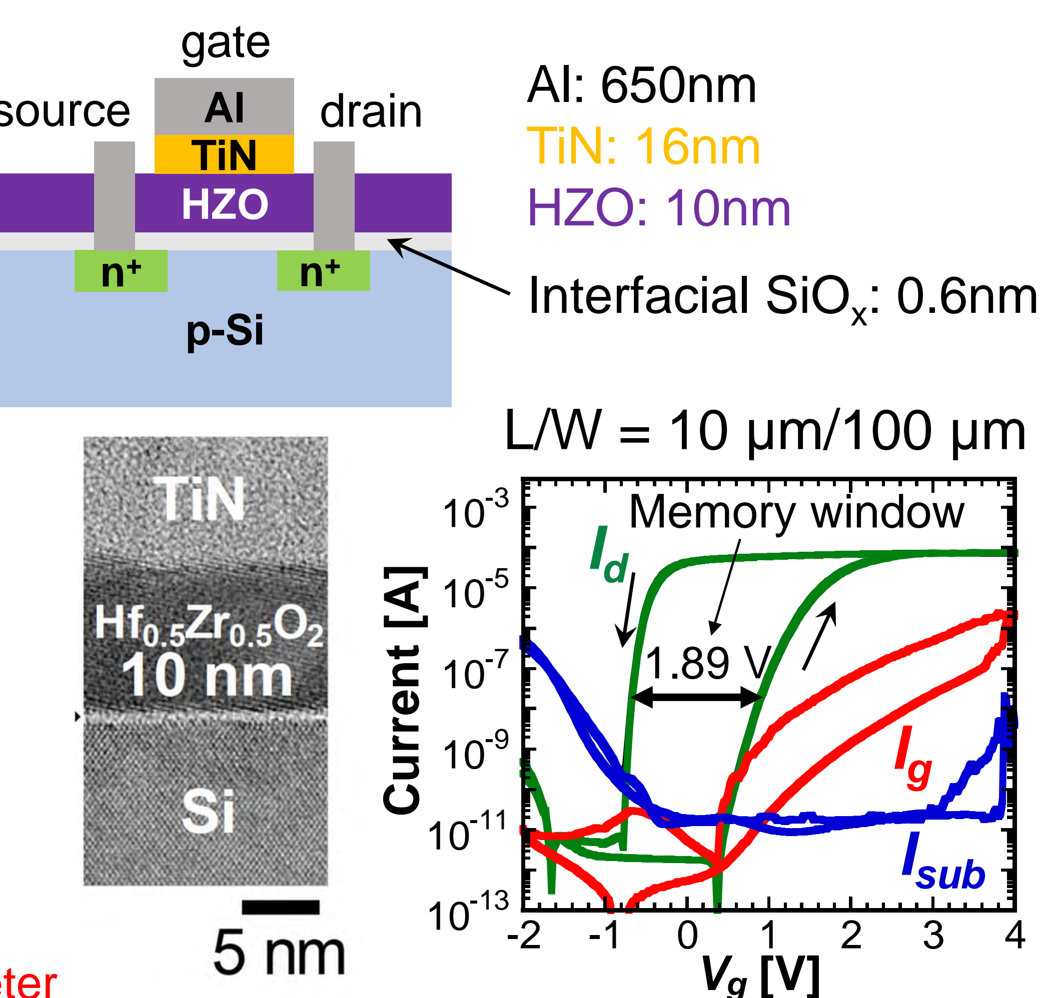


K. Toprasertpong et al, Appl. Phys. A 128, 1114 (2022)

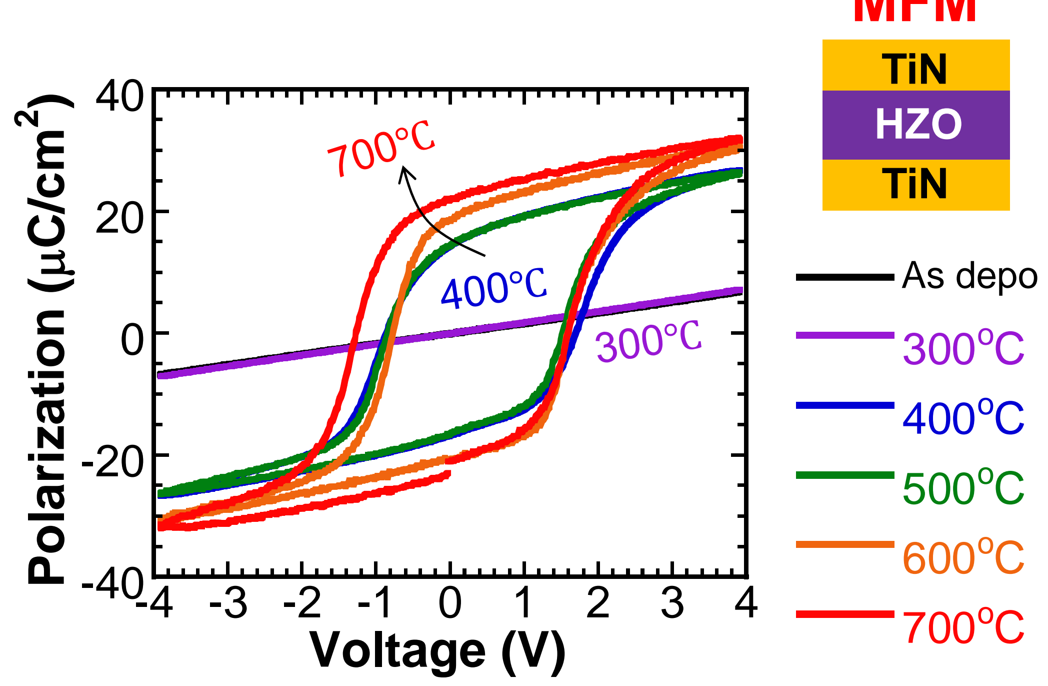
Impact of annealing temperature on structure and FeFET characteristics

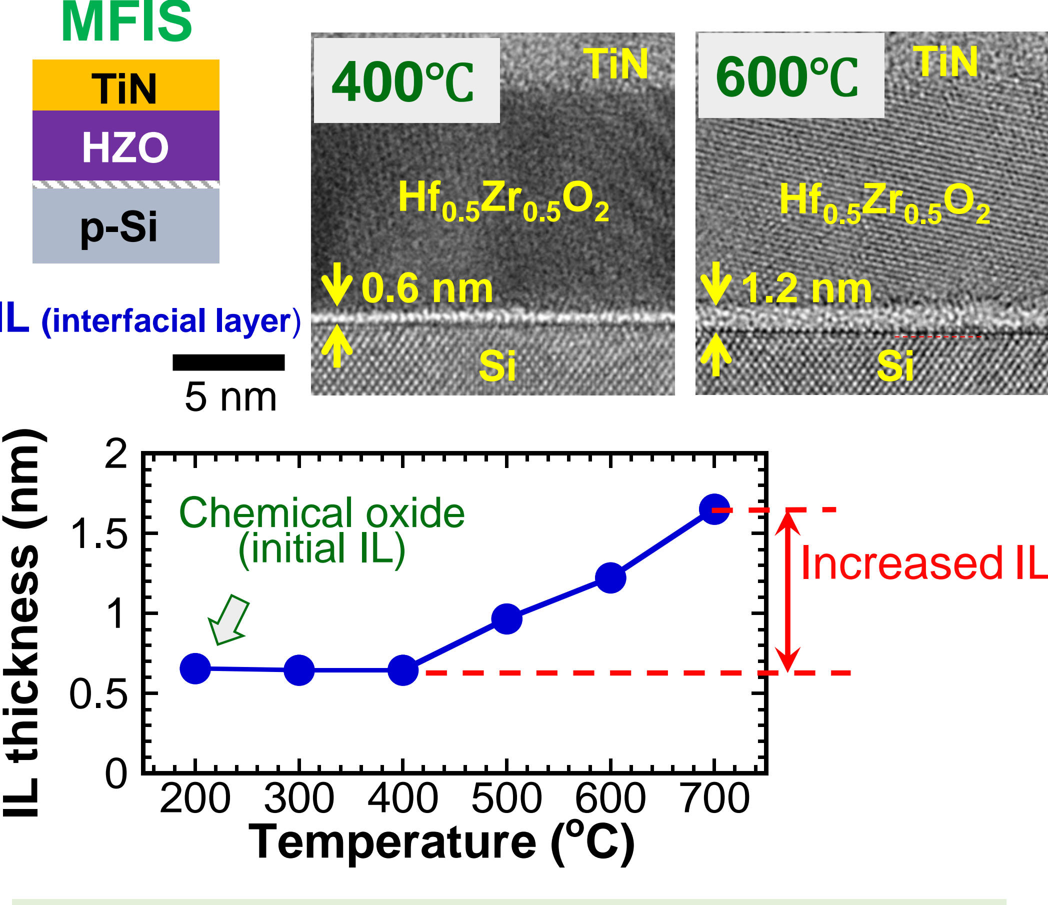
- Si sub cleaning
- SiO_2 oxidation 208nm
- Marking 400nm
- Active area definition by BHF
- SiO_2 re-oxidation 25nm
- S/D patterning by photoresist
- Ion implantation
  - ● P 30keV 2×10^{14} cm^{-2}
- Thermal activation @ 1000 ℃
- SiO_x 0.6 nm with SC2
- ALD $Hf_{0.5}Zr_{0.5}O_2$ 10nm
- TiN 16 nm by EB evaporation
- Al 650 nm
- Gate patterning
- Contact etching BHF
- Al:Si(2%) 350nm for S/D Contact
- Al 500 nm for back contact
- PMA 300 ~ 700 ℃ 30s → annealing for crystallization of HZO → critical process parameter

K. Toprasertpong et al., IEDM (2019) 570, VLSI Symp. (2020)TF1.5, APL 116 (2020) 242903, EDL 41 (2020) 1588



Impact of annealing condition on ferroelectric and MFIS interface properties





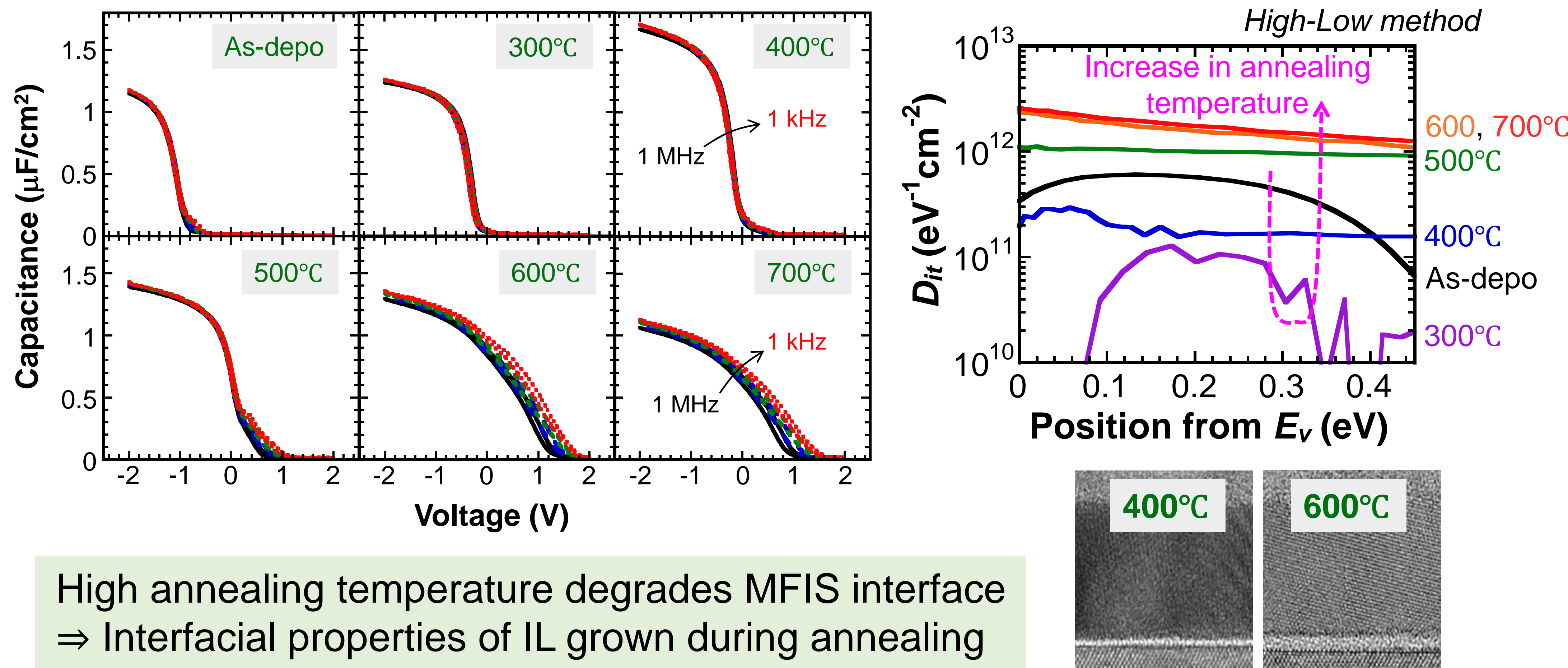
- Ferroelectric characteristics appear at annealing temperatures higher than 400 °C
- Gradual increase in polarization with an increase in annealing temperature

- Annealing temperature higher than 400 °C → Increase in IL thickness

K. Toprasertpong et al., EDL 41 (2020) 1588

Evaluation of MFIS interface properties by C-V characteristics

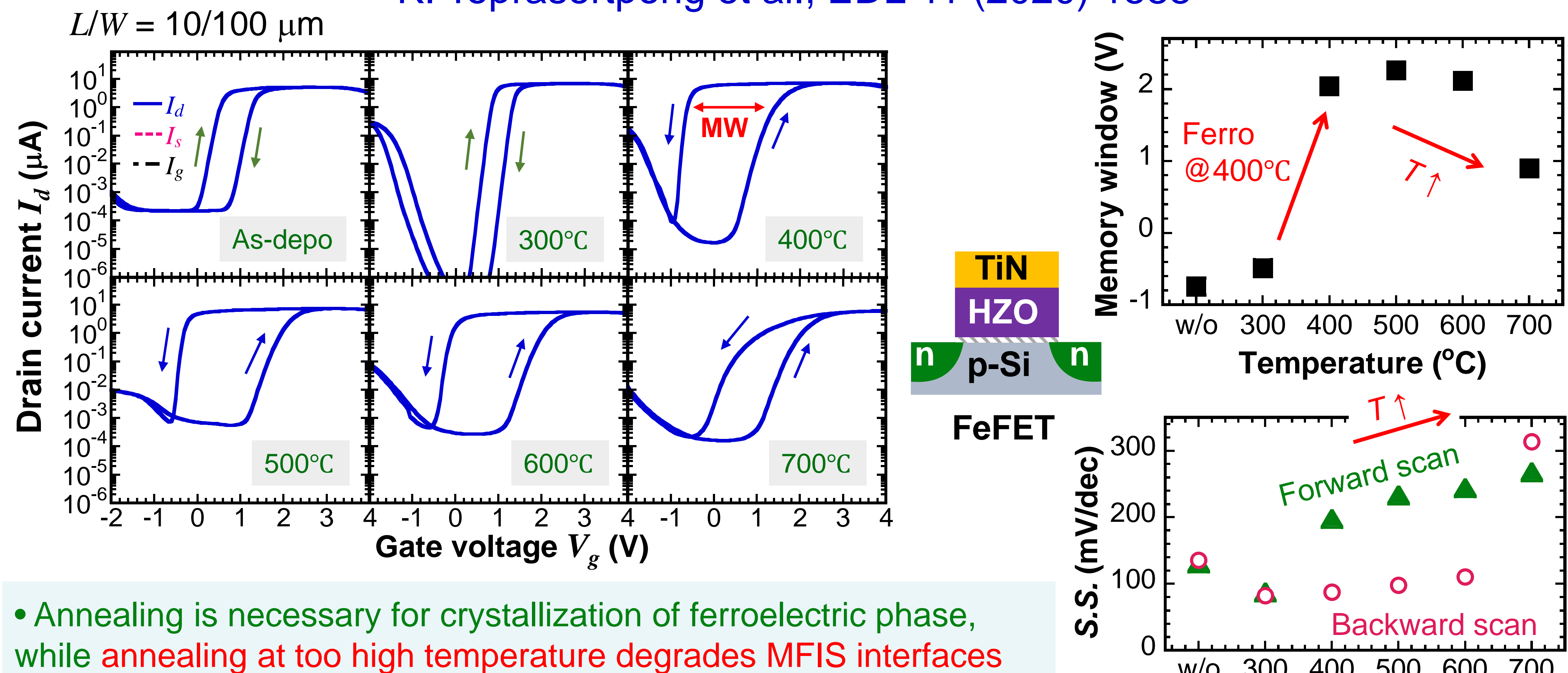
K. Toprasertpong *et al*, IEEE Electron. Dev. Lett. **41**, 1588 (2020)



High annealing temperature degrades MFIS interface
⇒ Interfacial properties of IL grown during annealing may not be good

Impact of annealing condition on FeFET characteristics

K. Toprasertpong et al., EDL 41 (2020) 1588

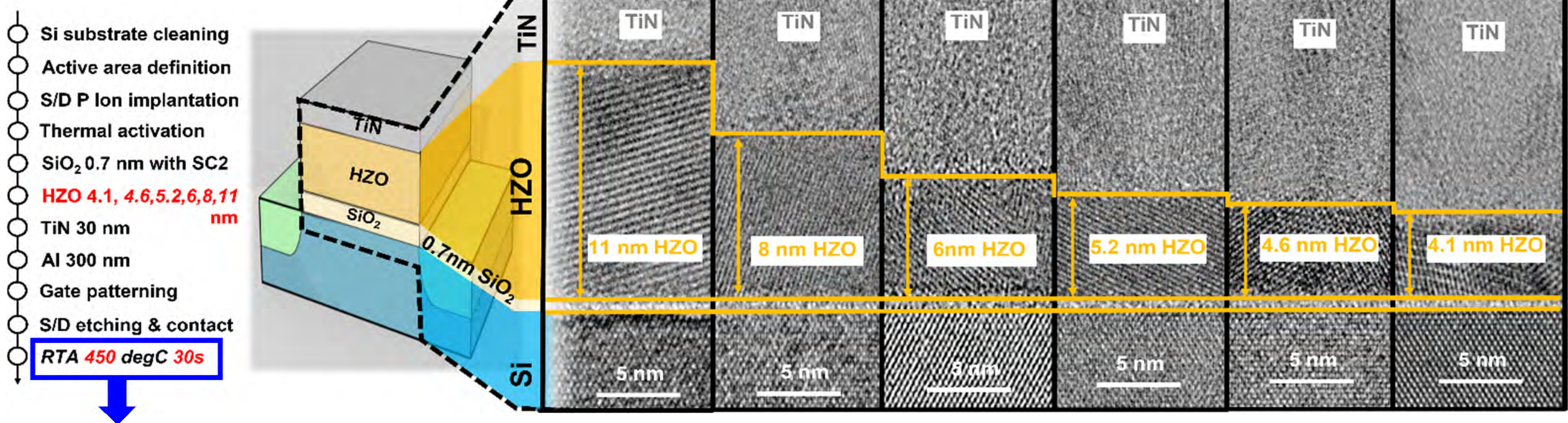


- Annealing is necessary for crystallization of ferroelectric phase, while annealing at too high temperature degrades MFIS interfaces (high D_{it}, high S.S., high I_{off} and narrow memory window)

⇨ Optimum annealing temperature exists (400℃ in this study)

HZO FeFETs with different HZO thickness

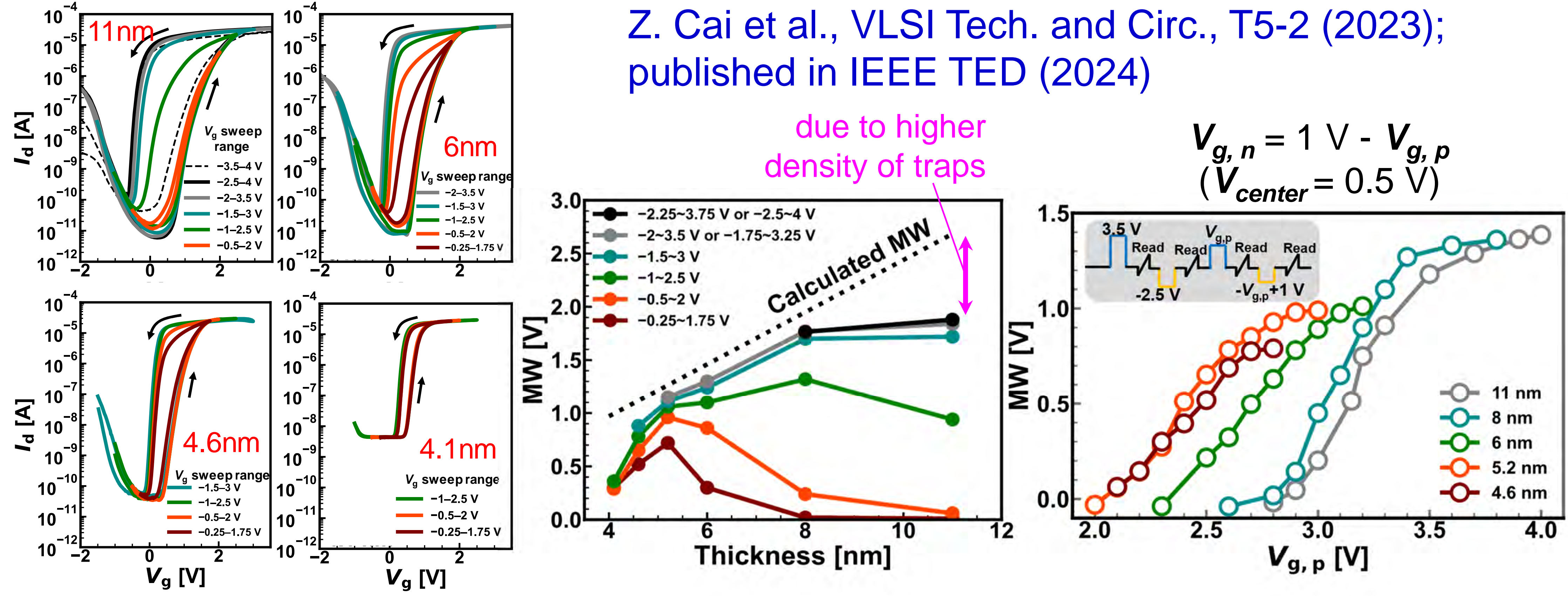
Z. Cai et al., VLSI Tech. and Circ., T5-2 (2023); published in IEEE TED (2024)



✓ Enough ferro-electric property and good interface quality

- ❑ Impact of HZO film thickness scaling on FeFET characteristics and reliability is also systematically examined for low voltage operation and better reliability in FeFETs
- ❑ ALD $Hf_{0.5}Zr_{0.5}O_2$ films with 0.7-nm-thick SiO_2 interfacial layers show good crystallinity from 11 nm to 4.1 nm after 450ºC annealing

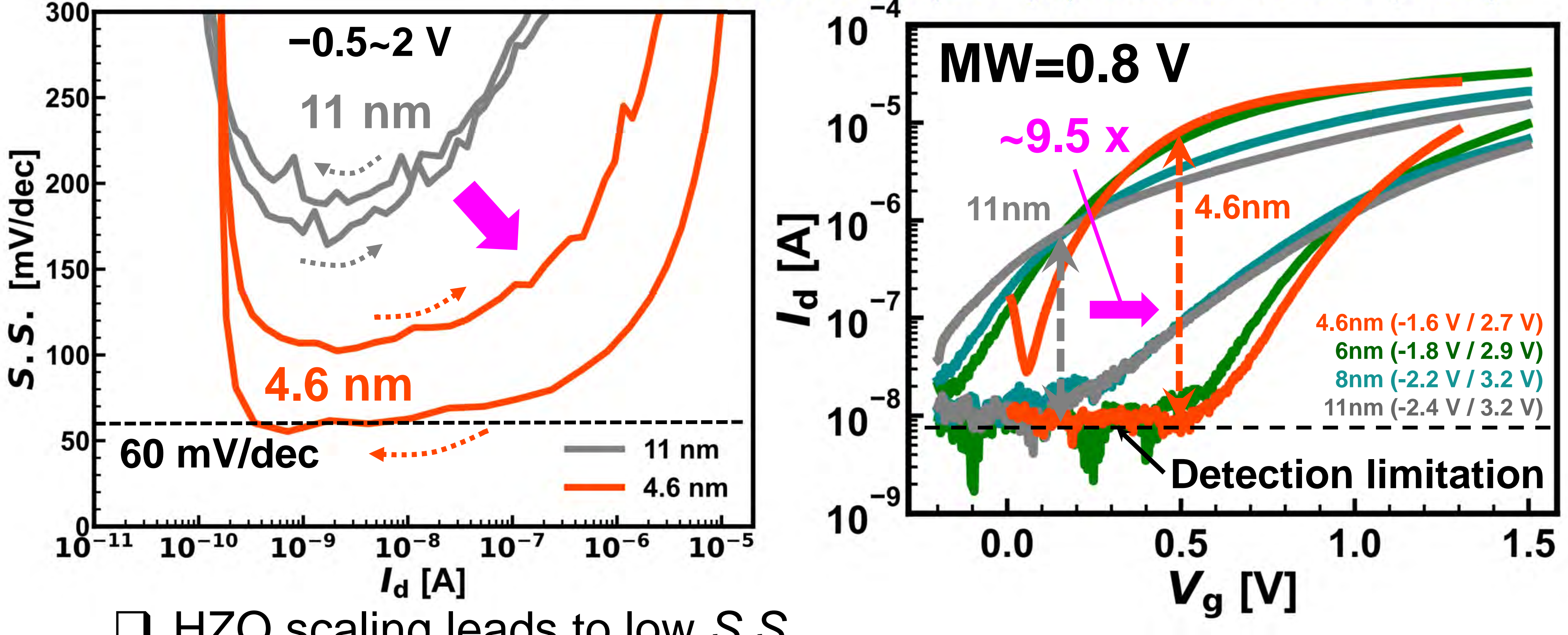
Memory window evaluated by DC and pulse measurements



- Saturated memory window (MW) is higher with thicker HZO (saturated MW ~ $2E_c t_{FE}$ → decrease in saturated MW is a drawback of thinner HZO)
- Under low-voltage operation, larger MW is obtained in an optimum thin HZO
- Sufficient MW in low voltages can be obtained by appropriate HZO scaling

Impact of HZO scaling on sub-threshold swing and I_{on}/I_{off} ratio

Z. Cai et al., VLSI Technology and Circuits, T5-2 (2023), published in TED (2024)

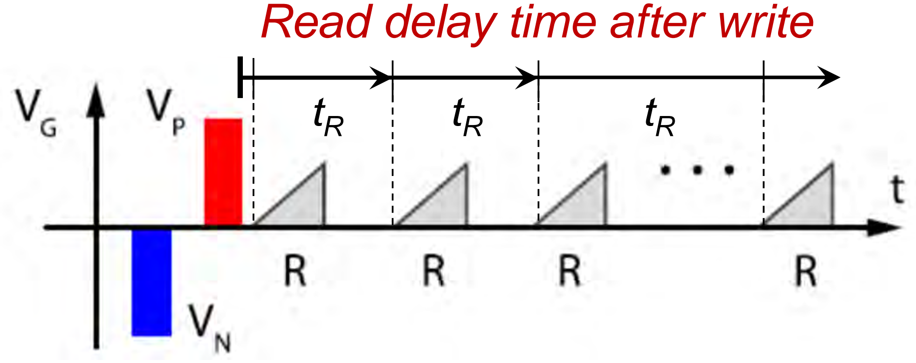


- HZO scaling leads to low *S.S.*
- I_{on}/I_{off} ratio at a given V_g (index in current readout operation) is improved by HZO thickness scaling

Issues related FeFET memory operation

Necessity of long read-after-write time

Mulaosmanovic et al, TED 66 (2019) 3828



- Memory window does not open at short read-after-write time
- Sufficient read delay time after positive V_g pulse to induce **electron de-trapping** is necessary for the readout of low V_{th} (program: PGM) condition

Continuous de-trapping of electrons after writing lowers PRG V_{th} with longer waiting time

ERS PRG t_F = 10 nm

V_T [V]

3 2 1 0

10^{-4} 10^{-2} 10^{0} 10^{2} 10^{4} 10^{6}

time [s]

During writing

Linear polarization

Hold condition
Just after write

Hold condition
waiting for ~sec

TiN S D Si

TiN S D Si

TiN S D Si

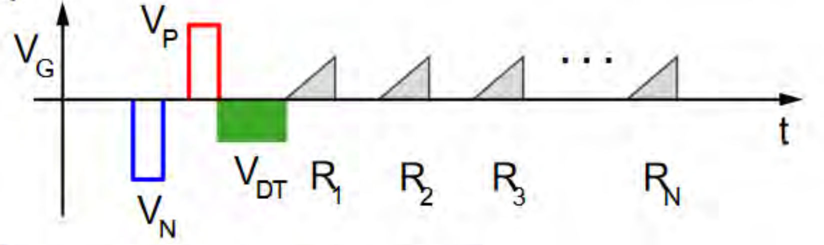
Negative MW

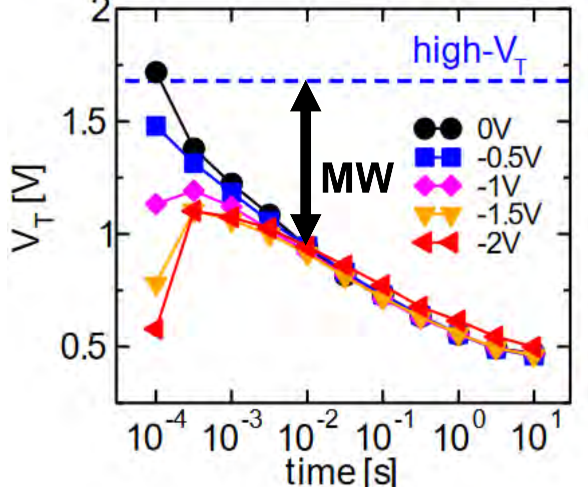
Positive MW

Partial relaxation of polarization

de-trapping of electrons

For improving long read-after-write time issue

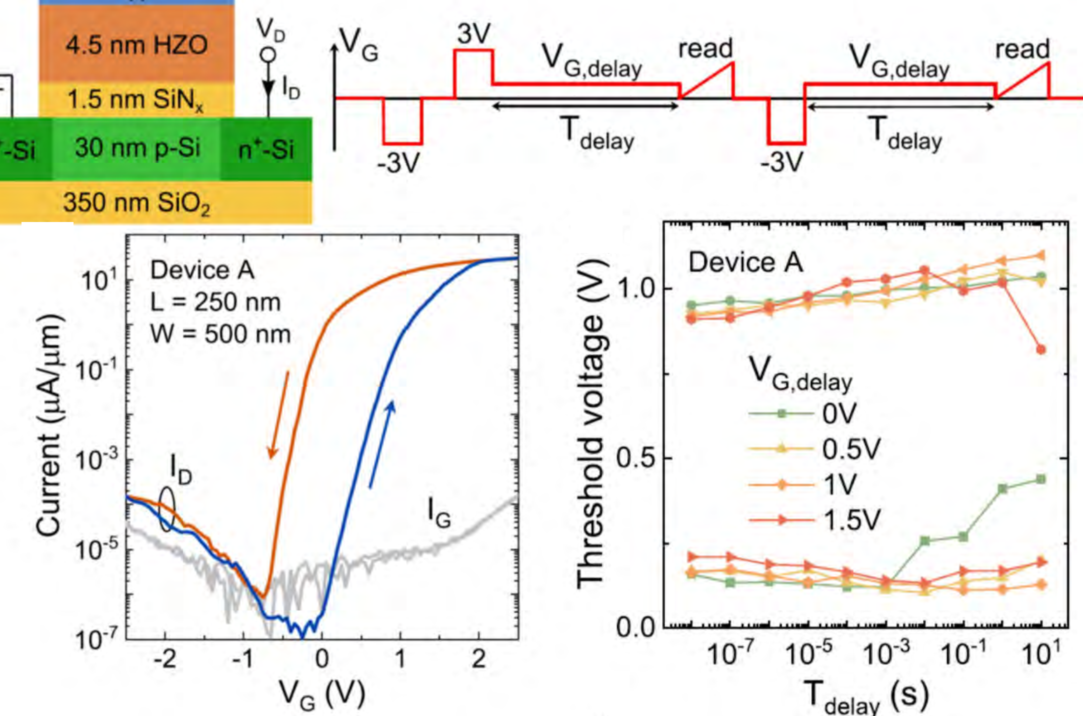




- Short negative pulse just after write operation can enhance de-trapping of trapped electrons, which can increase memory window in short term

H. Mulaosmanovic et al, EDTM, 7C-4 (2020)

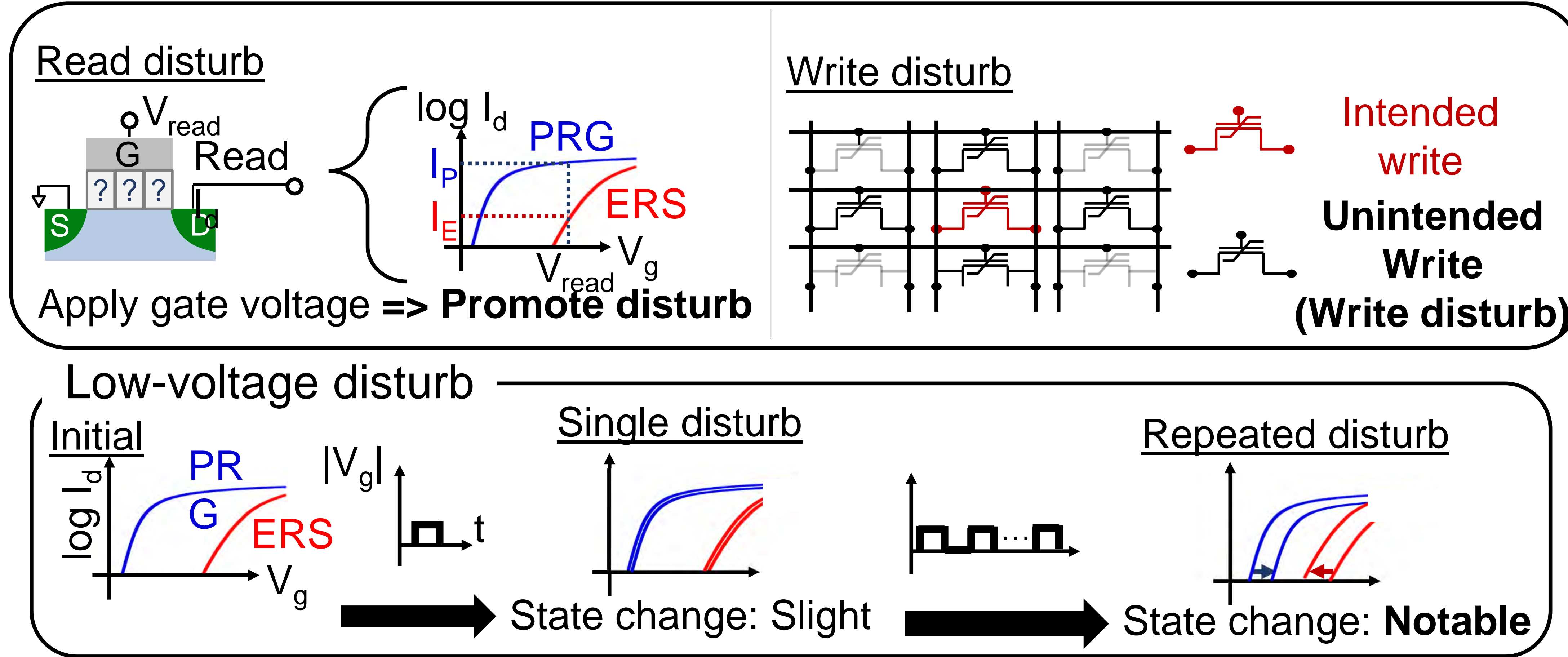
H. Zhou et al, IEDM, 395 (2020)
N. Tasneem et al, IEDM, 122 (2021)
Z. Wang et al, IEDM, 430 (2021)



- IL engineering, allowing fast de-trapping of electrons or reducing electron traps, is expected to realize fast read-after-write time

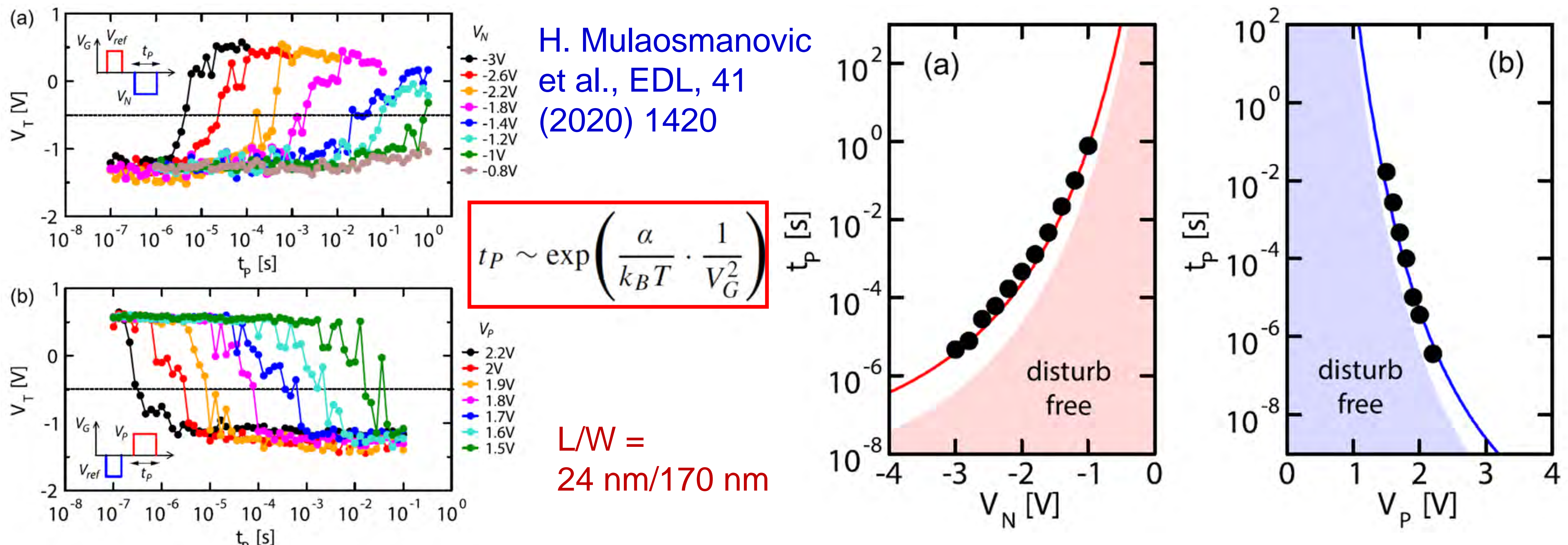
M. Hoffmann *et al.*, *IEEE EDL* **43**, 717 (2022)

Read/write disturb characteristics of FeFET memory



- One advantage in FeFET memory is non-destructive readout
- Read disturb characteristics can determine non-destructive readout time
- Write disturb characteristics are critical for memory array operation

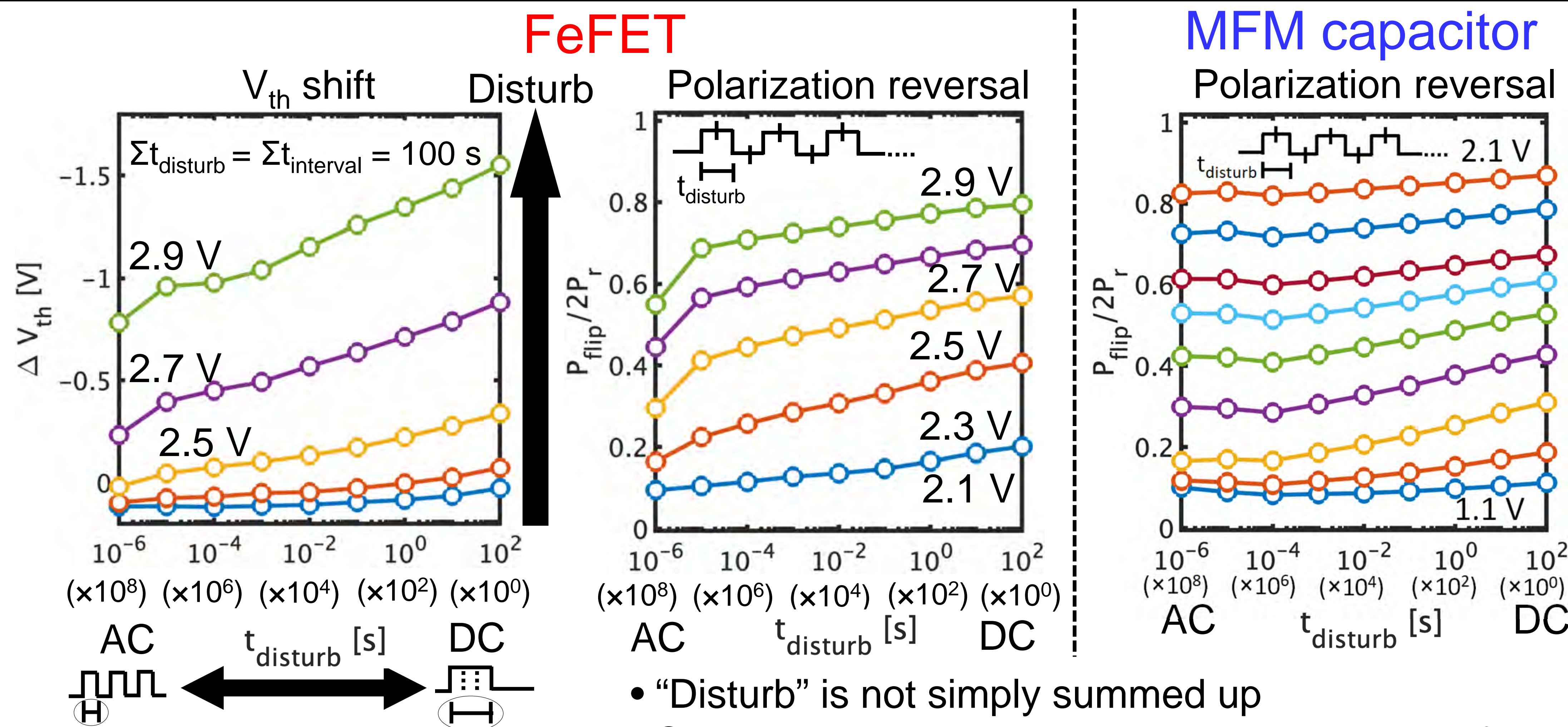
Dependence of polarization switching on ulse voltage and time



• Ferroelectric switching is strongly dependent on voltage (electric field) and time during read operation, indicating that the readout voltage and time are critical to non-destructive readout

• The disturb time can be well represented as a function of the gate voltage, which can provide guidelines for a disturb-free operation of FeFETs

Experimental disturb characteristics of FeFET and MFM capacitor

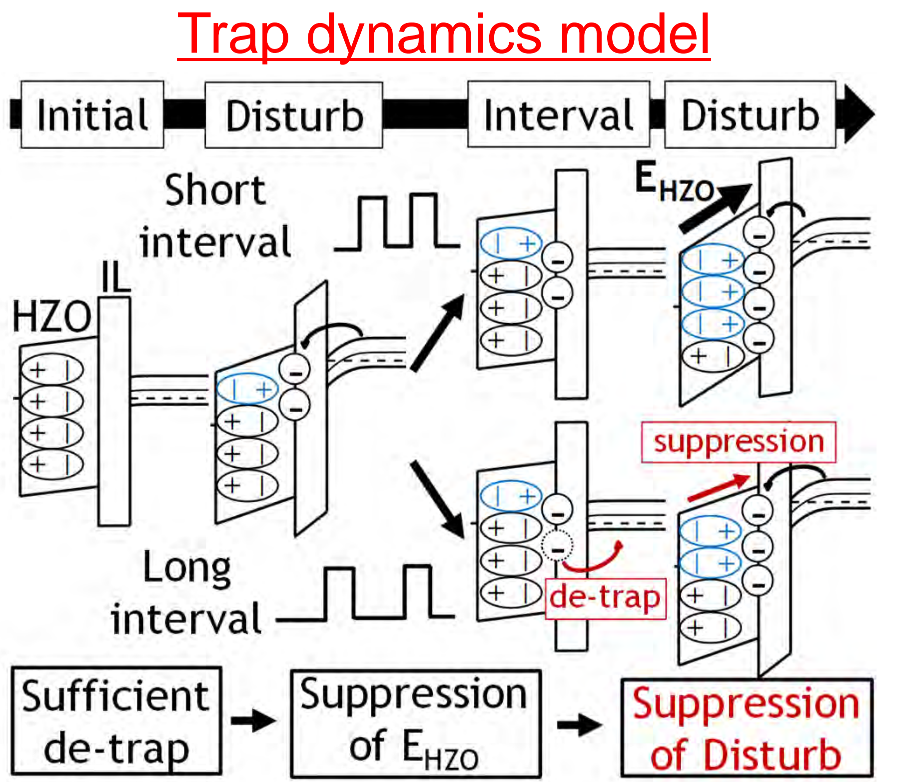


- “Disturb” is not simply summed up
- Shorter disturb time leads to more relaxation of both threshold and polarization, resulting in less disturb

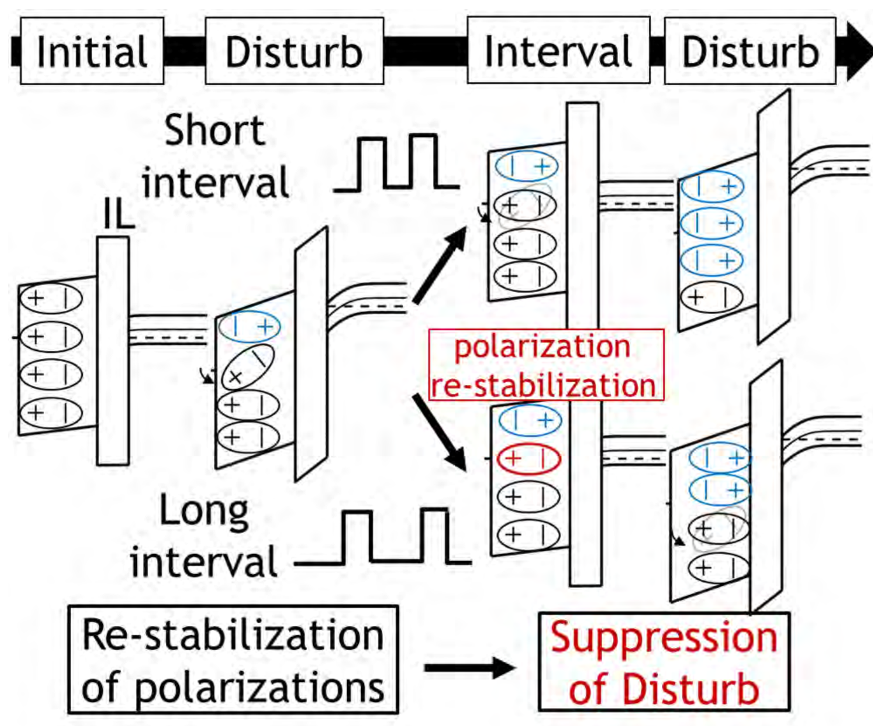
M. Otomo et al., VLSI symp, 2024

Physical model of non-accumulative characteristics of FeFET

Trap dynamics model



Polarization dynamics model



- The non-accumulation effect that can improve the disturbance characteristics can be explained by two mechanisms: the electrons are detrapped during the hold, weakening the electric field for the next disturbance, and the polarization that is nearly reversed by the disturbance is restabilized during the hold.

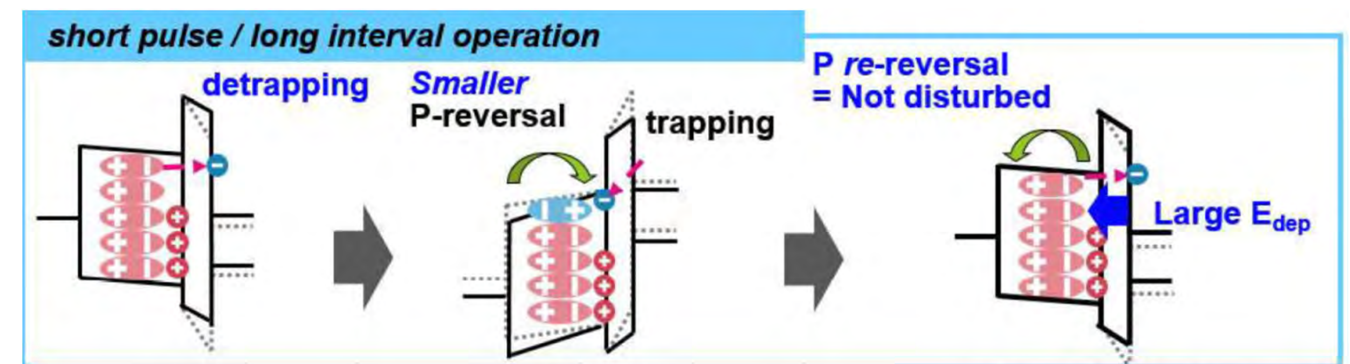
M. Otomo et al., VLSI symp, 2024

Importance of trap-property-aware device operation for FeFETs

- Amounts of trapped/de-trapped electrons can be modulated by bias conditions → **electric field across FE can be intentionally controlled** → **write/read disturb can be mitigated by trap control**

(1) long interval/short V_g pulse operation

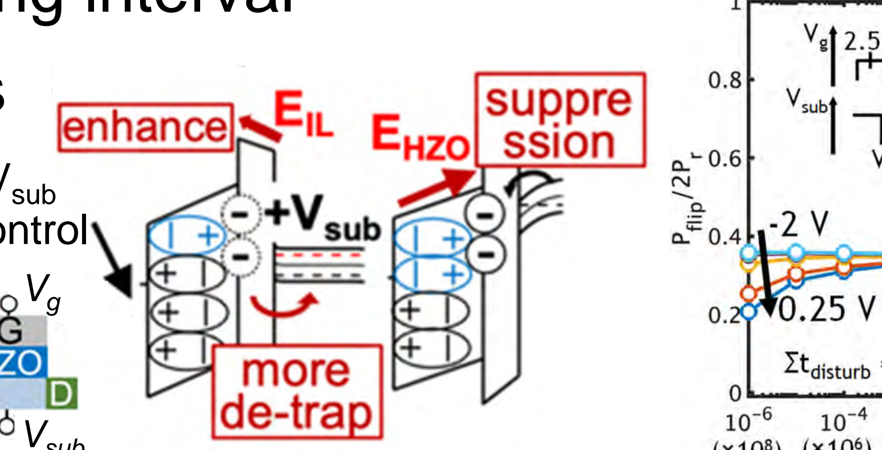
T. Hamai (Kioxia) et al., IRPS 6A.1-2 (2023)



(2) short/low negative V_g pulse during interval

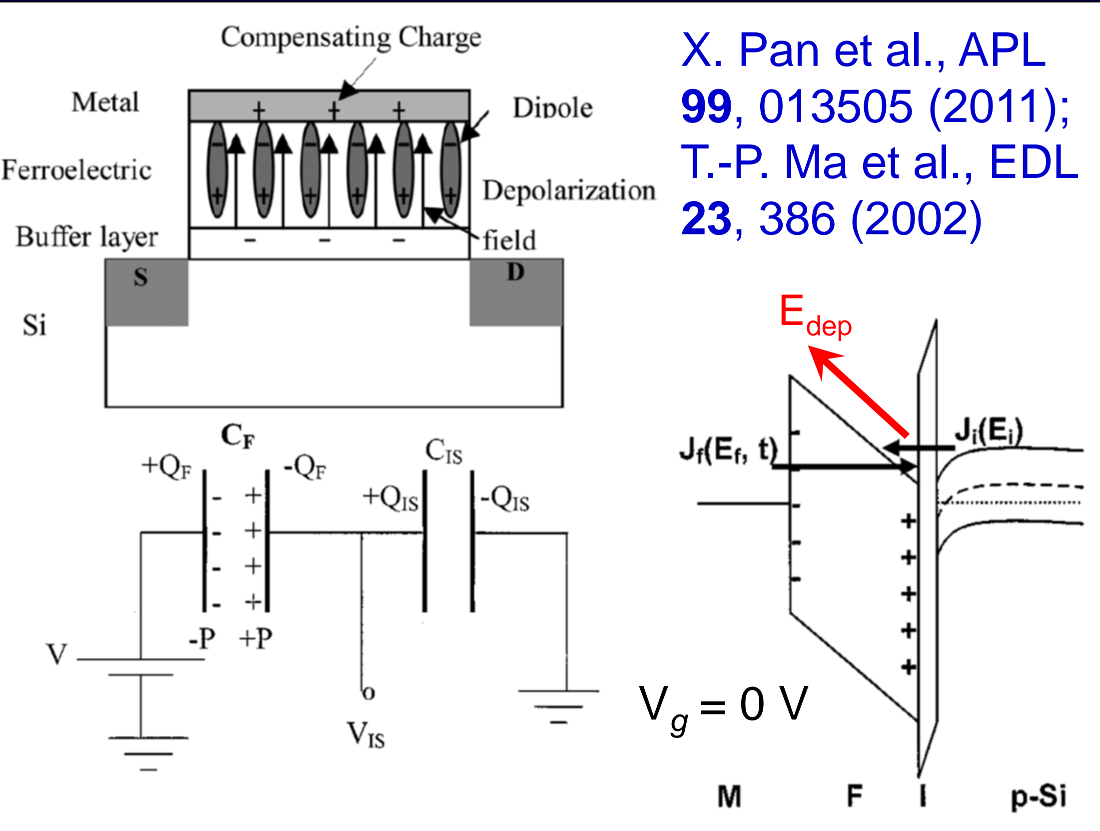
(3) short/low positive substrate bias during interval

M. Otomo et al., VLSI symp, 2024



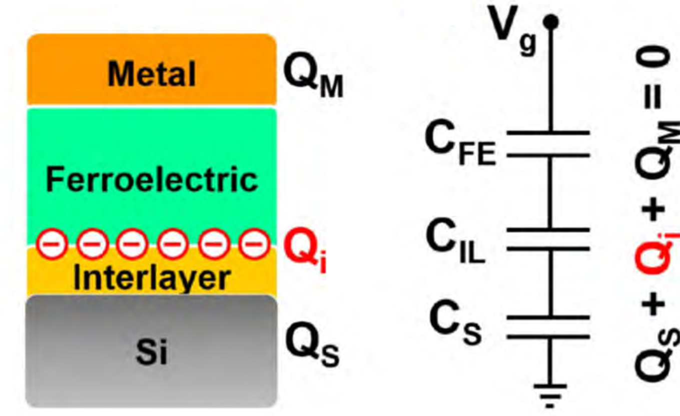
FeFET reliability

Depolarization field in FeFET gate stack



X. Pan et al., APL **99**, 013505 (2011);
T.-P. Ma et al., EDL **23**, 386 (2002)

X. Wang et al., TED **67**, 4500 (2020)

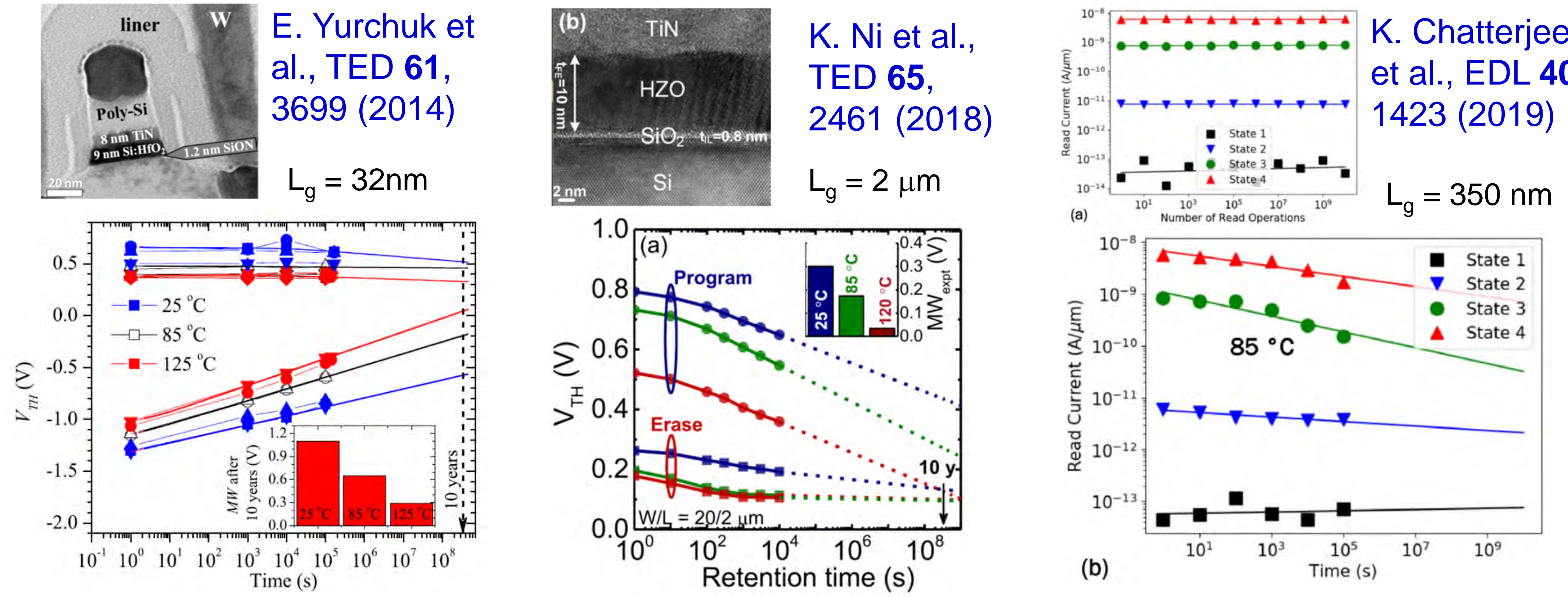


$$E_{dep} = \frac{V_{FE}}{t_{FE}}\bigg|_{V_g=0} = -\frac{P_0(Q_i) + Q_i}{(C_{FE} + C_{IS})t_{FE}}$$

- Q_i: electron or hole trap density. The sign is opposite to that of P_0

- Interfacial layers of FeFETs introduce depolarization field across FE in the direction of reducing the polarization at V_g = 0 V, tending to reduce the memory retention time
- Trapped charges at FE/IL interfaces can reduce E_{dep}, which is expected to improve retention characteristics

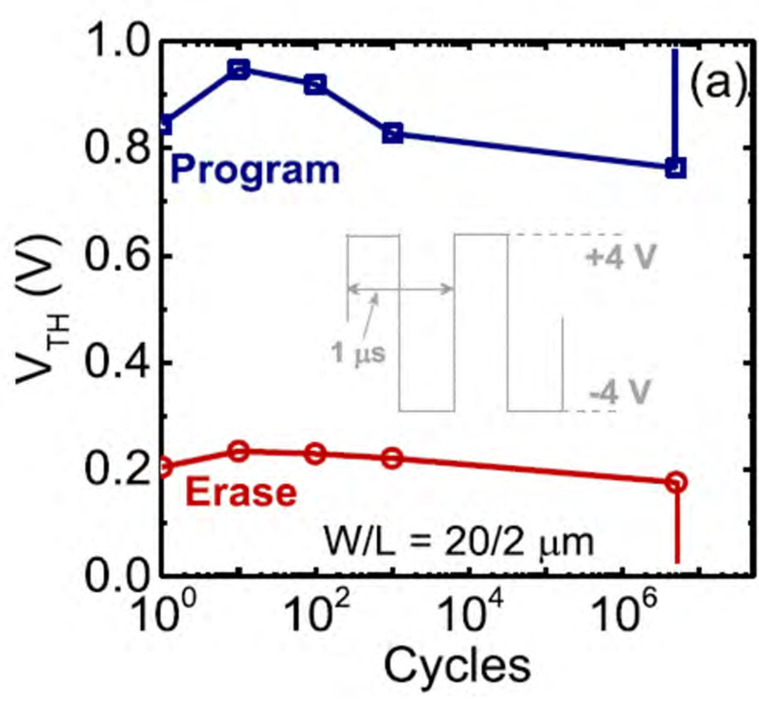
Typical retention characteristics of FeFET



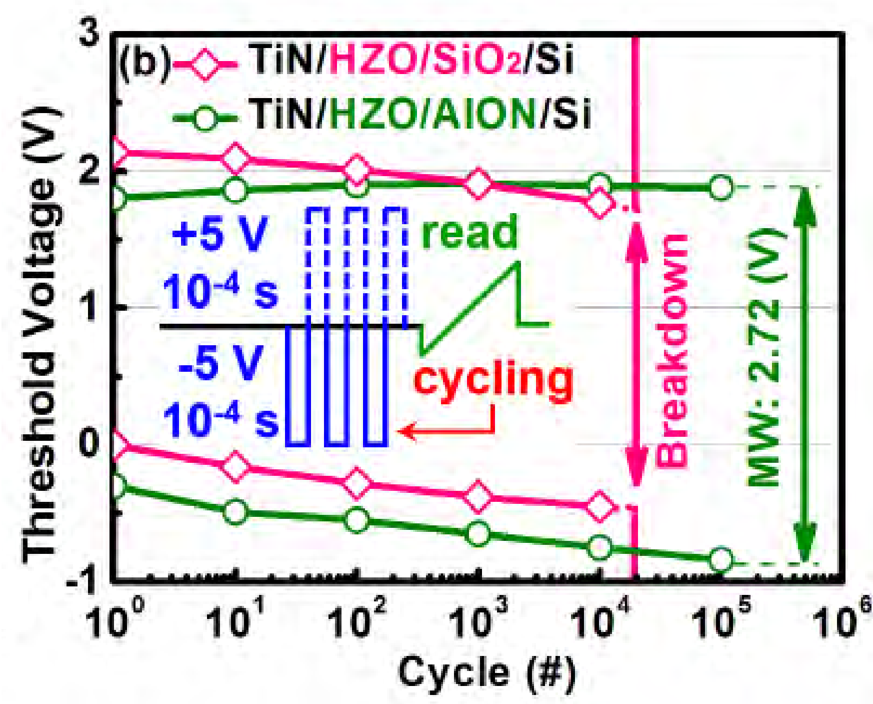
• Since early stage of studies on FeFETs, good retention characteristics over 10 year have been often reported

• Superior retention characteristics of HfO2-based FeFETs to perovskite FeFETs are attributable to higher E_c of HfO2-based FE and higher carrier trapping

Endurance of FeFET ~ gate stack breakdown



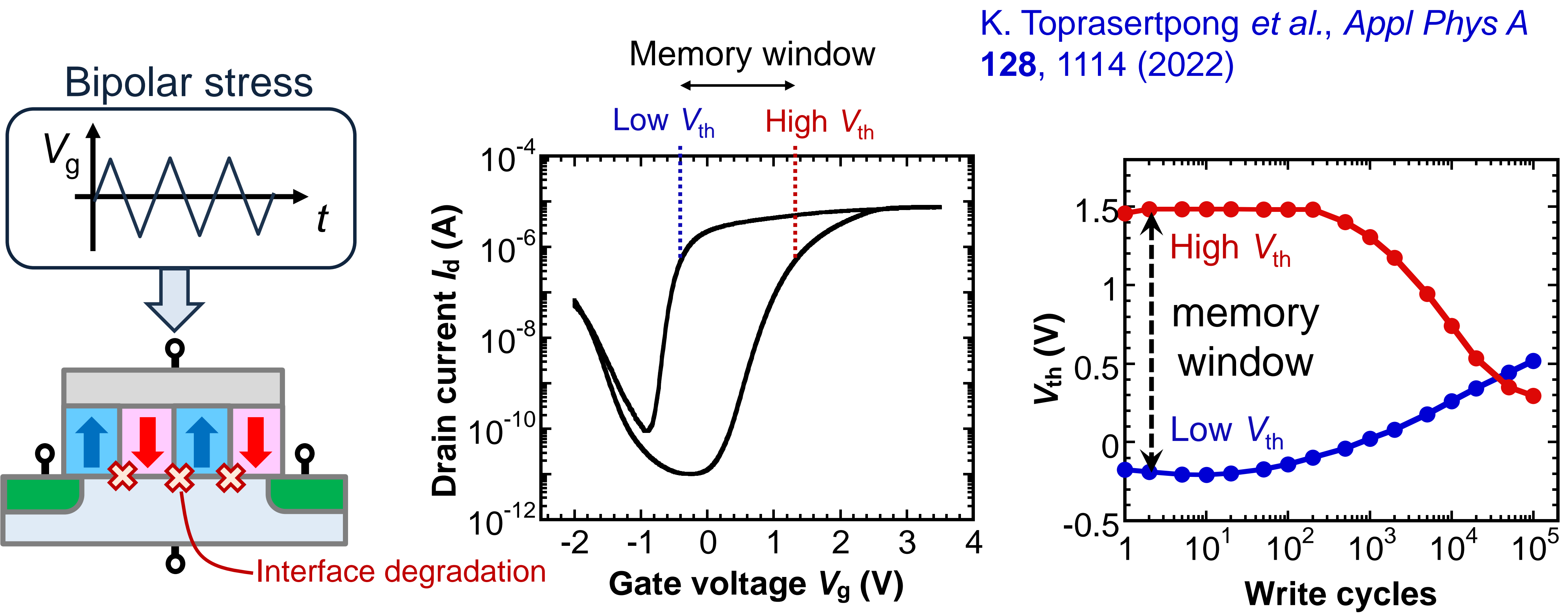
K. Ni et al, TED 65, 2461 (2018)



C.-Y. Chan et al, VLSI, TF2.5 (2020)

- Some cases have been reported in which endurance properties are determined by dielectric breakdown of gate stacks, attributable to too high operation voltage

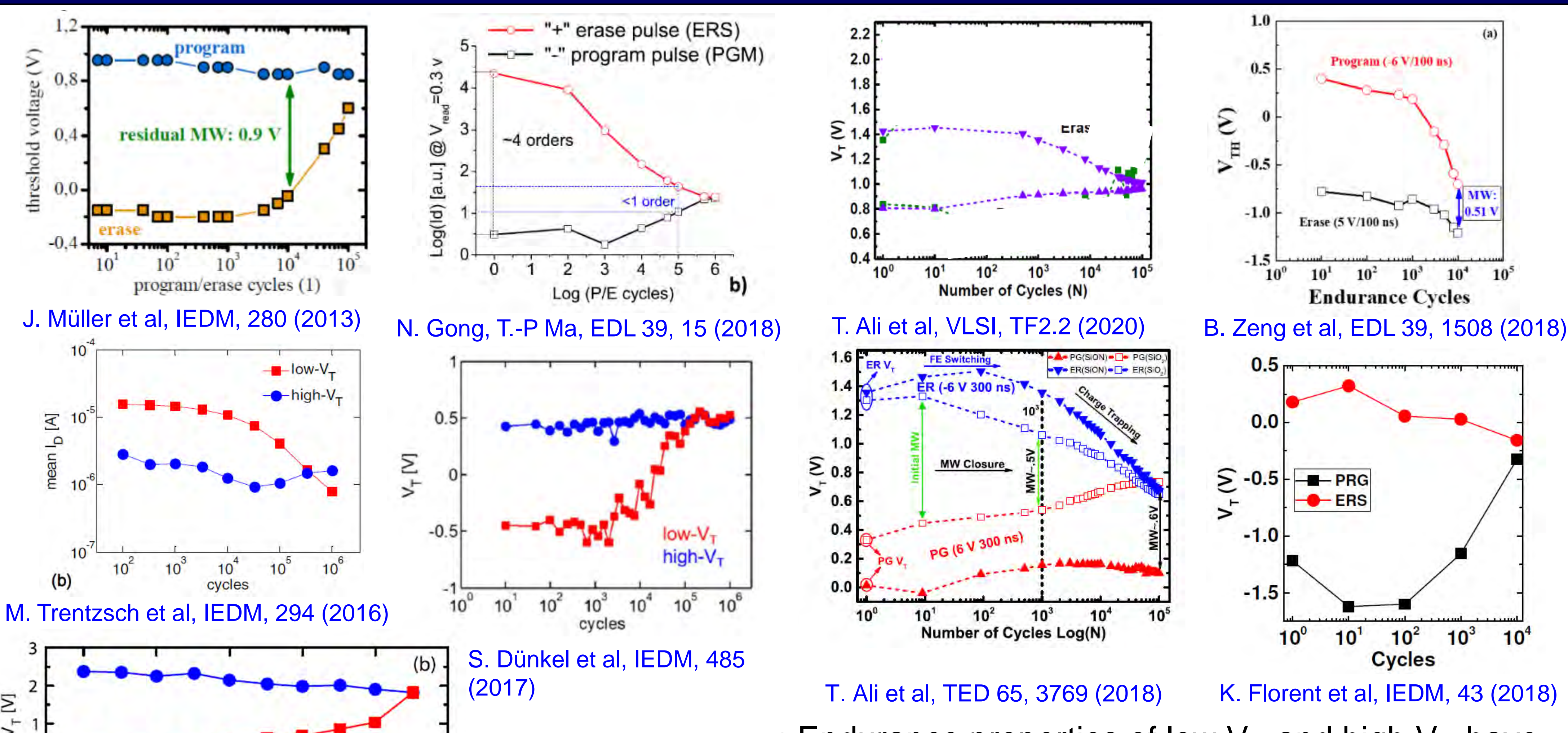
Endurance of FeFET ~ window narrowing



In FeFET memory, bipolar stress results in failure of memory operation

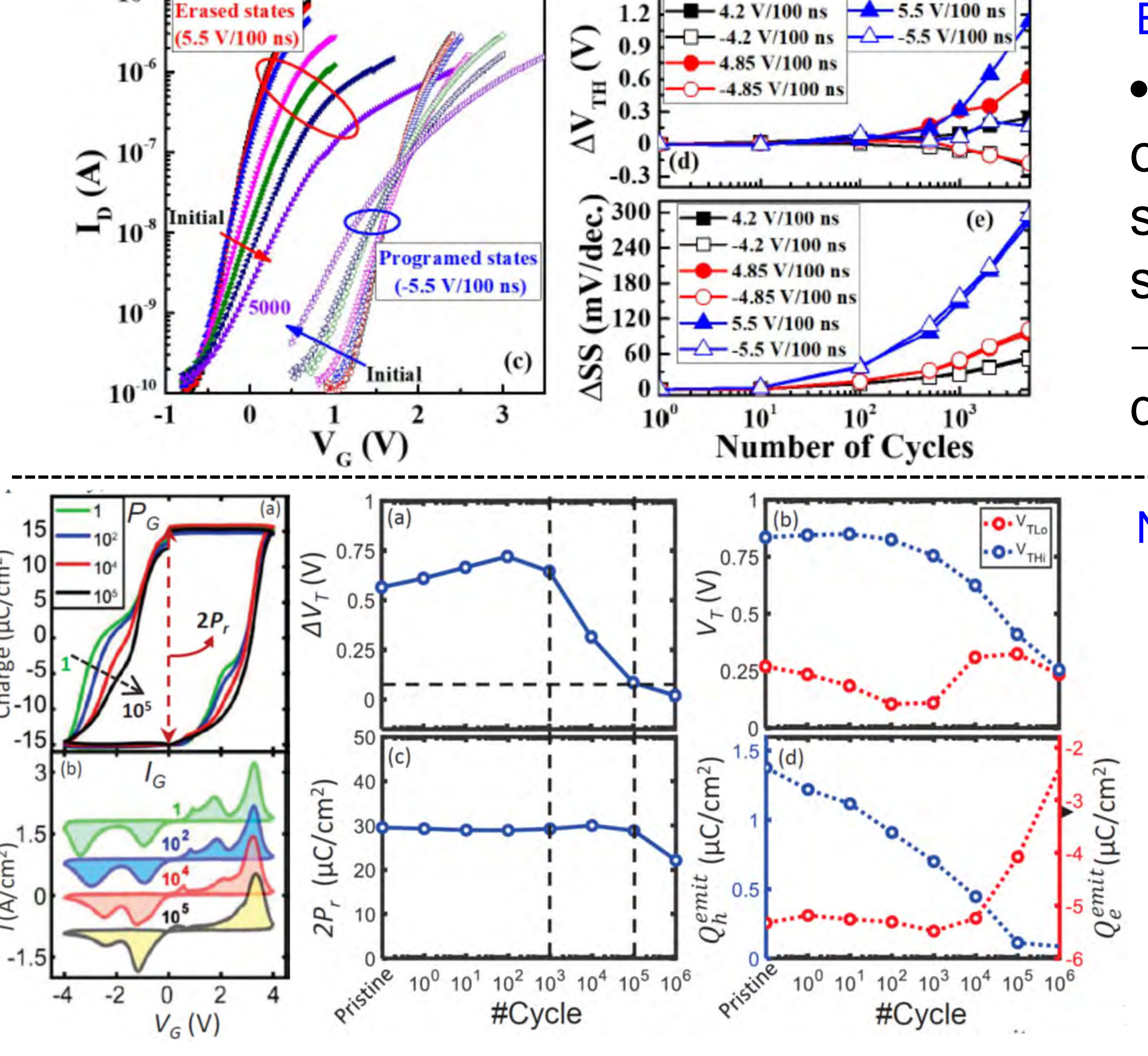
⇒ How about bipolar stress in reservoir computing?

Many reports on endurance properties



- Endurance properties of low V_{th} and high V_{th} have some variations among reports, though window narrowing is the common characteristic

Typical I-V and p-V characteristics of FeFET with write/erase cycling stress



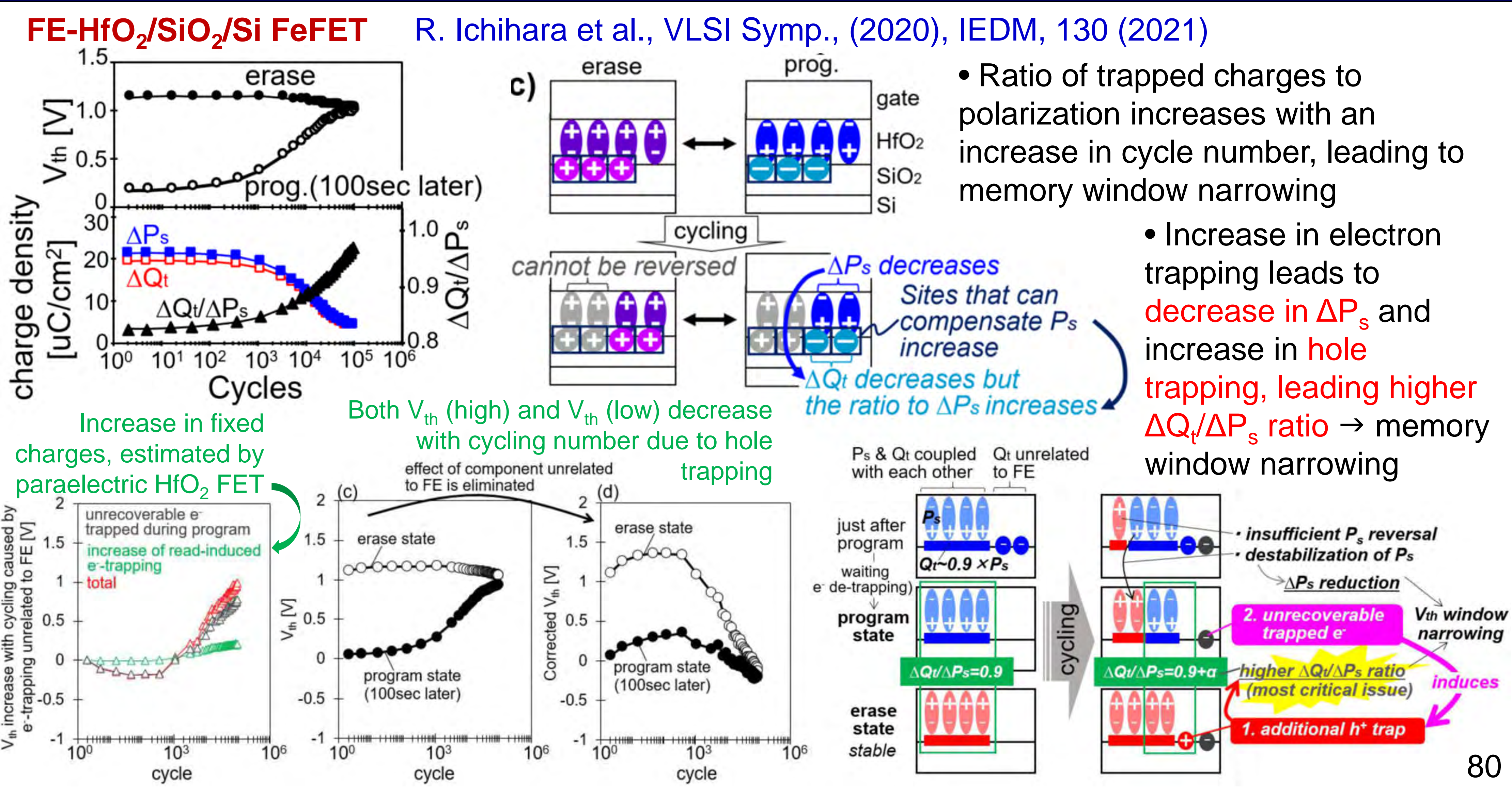
B. Zeng et al., IEEE EDL 98, 40, 2019

- Repeated cycling stress causes not only a voltage shift but also a significant degradation of sub-threshold swing (S factor) of I-V characteristics → generation of interface traps and degradation of interface properties

N. Tasneem et al., IEDM, p. 122, 2021

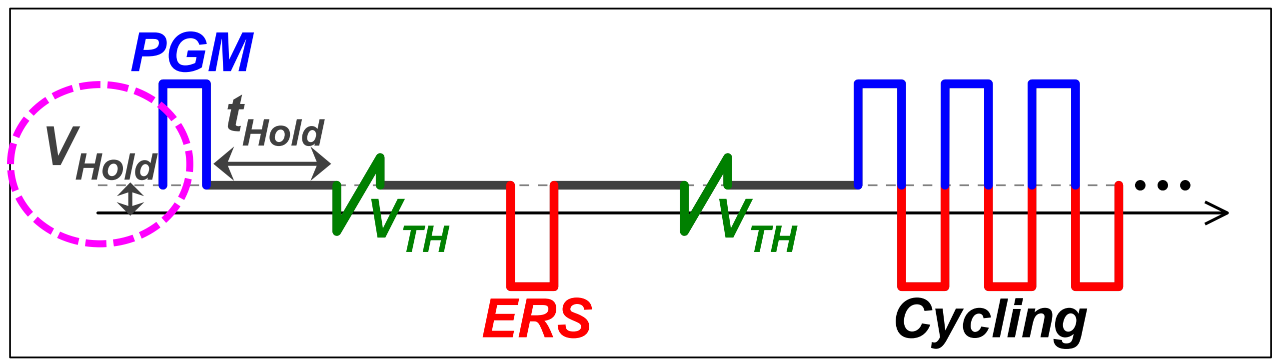
- Some papers reports no degradation in polarization even after cycling stress causing a significant reduction in memory window
- Polarization degradation in FE films might not be a dominant factor for memory window narrowing

Another report on trap-assisted polarization in FeFET

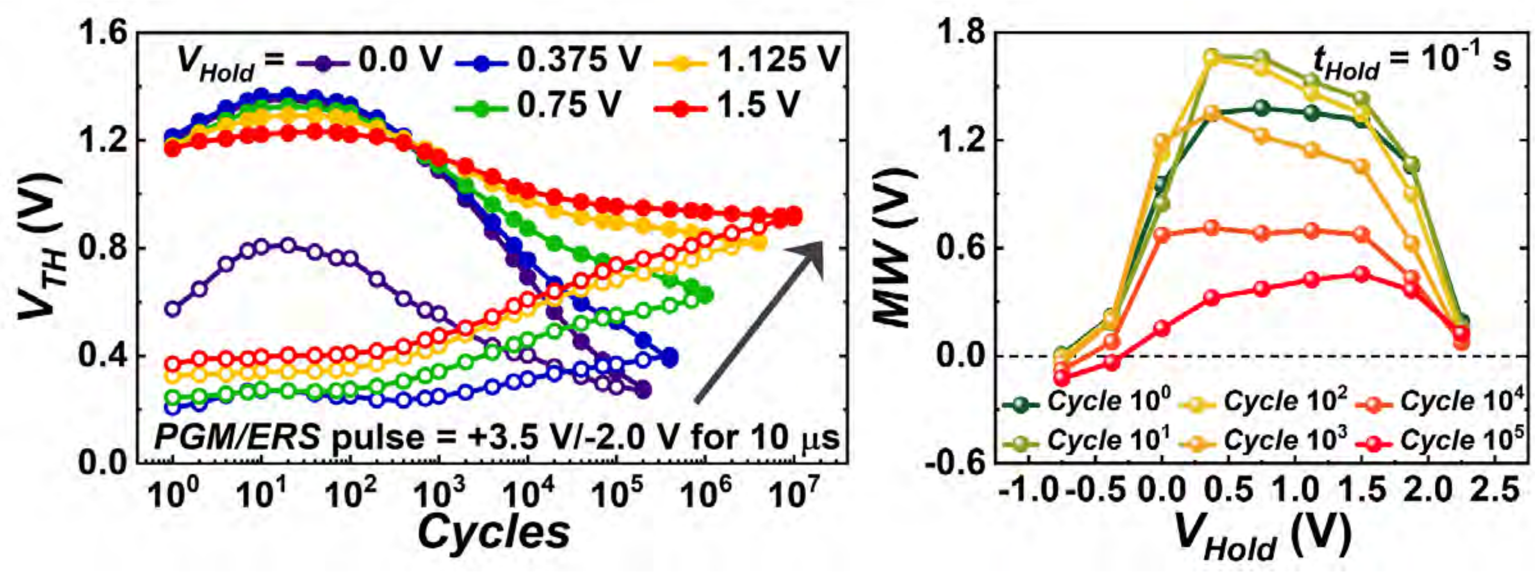


Effect of hold voltage on memory window narrowing of FeFET

S.-K. Cho et al., IEDM, 4-7 (2024)

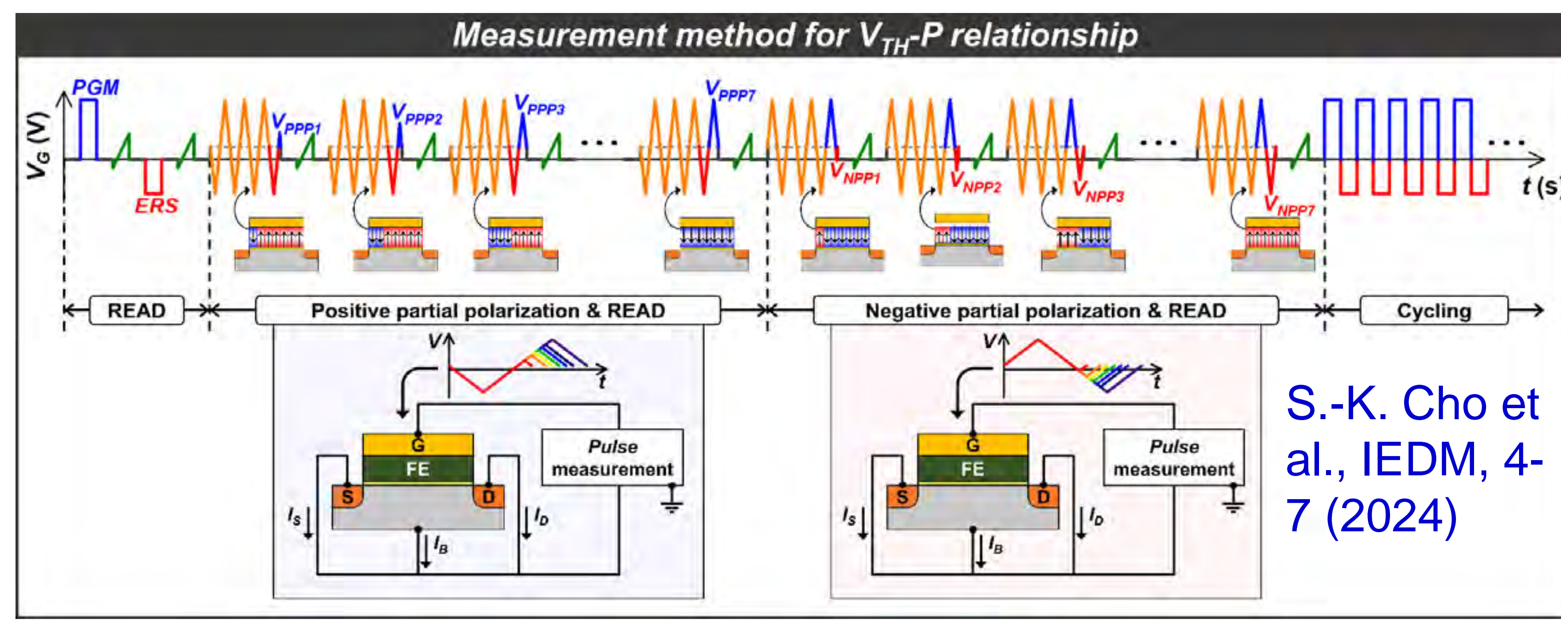


11 nm HZO/SiO_2(0.6 nm)/Si FeFET

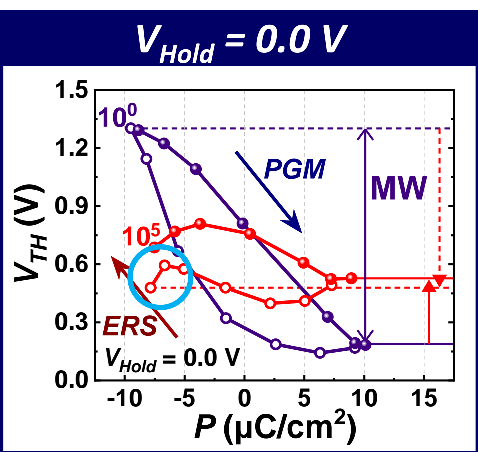


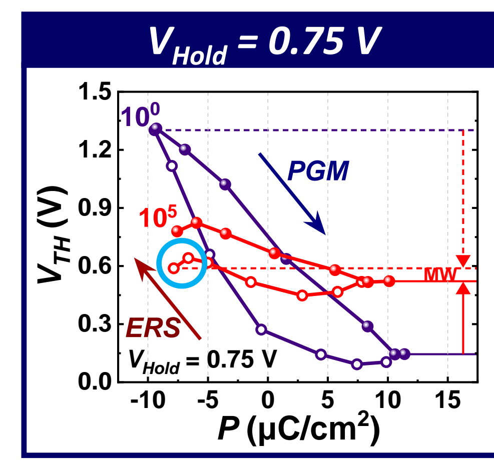
- ❑ It is found that endurance characteristics and memory window after stress are dependent on hold voltage
- ❑ Memory window appears for degraded FeFETs without memory window by applying positive hold voltage
- ❑ Optimum hold voltage is shifted to positive direction with increasing cycle numbers, indicating influence of hole traps

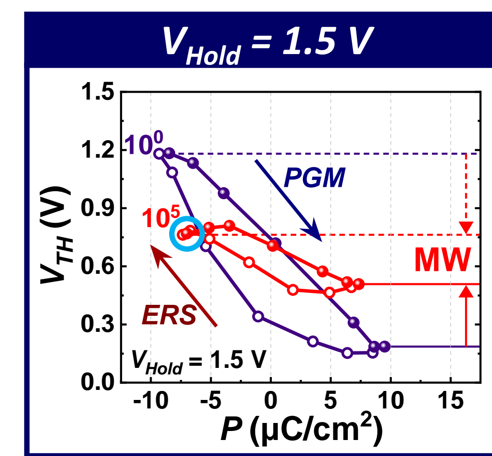
New method for evaluating P-V_{th} relation



S.-K. Cho et al., IEDM, 4-7 (2024)



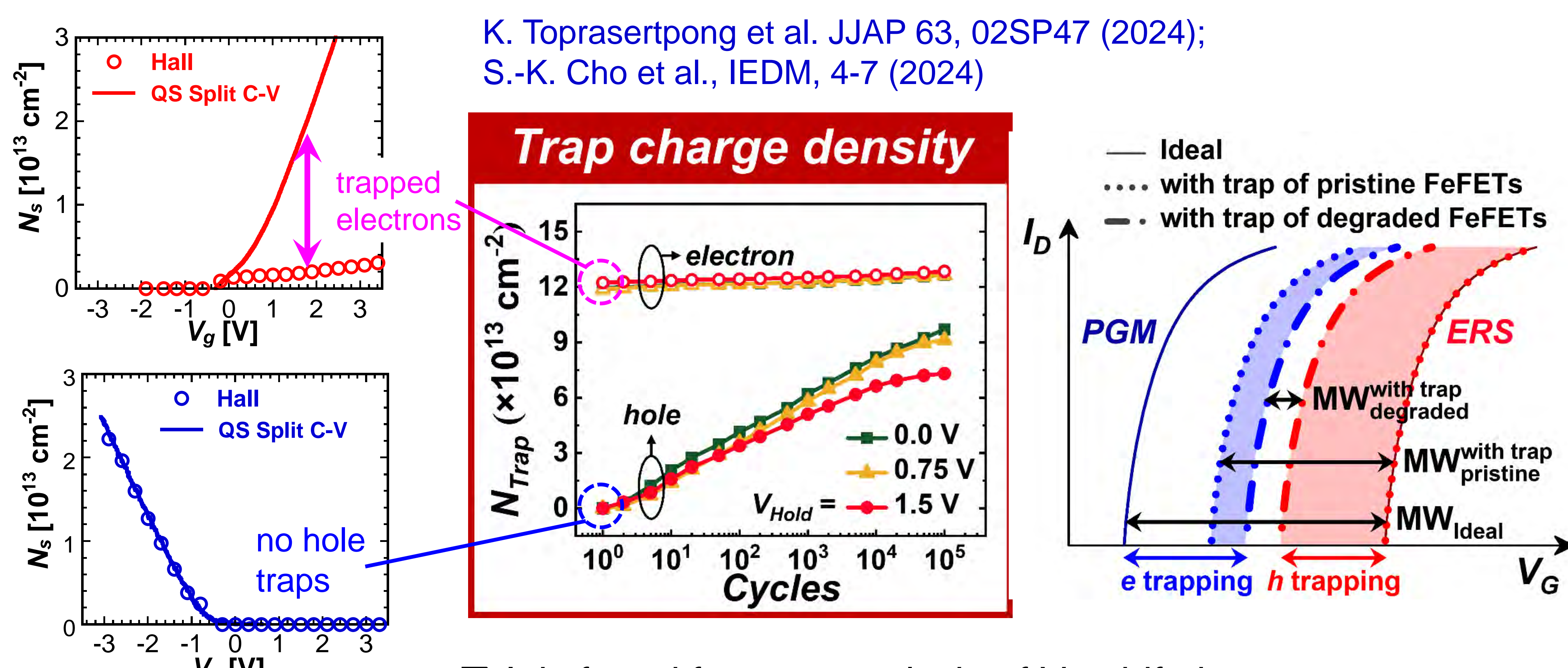




- We have proposed and demonstrated a method to evaluate P-V_{th} characteristics to identify the physical origin of memory window reduction
- No degradation in polarization is observed for FeFETs with no memory window
- Positive V_{hold} changes V_{th} after erase under the same polarization, indicating importance of hole de-trapping

Physical understanding of memory window narrowing mechanism

K. Toprasertpong et al. JJAP 63, 02SP47 (2024);
S.-K. Cho et al., IEDM, 4-7 (2024)



Pristine condition
(Hall measurement)

- It is found from an analysis of V_{th} shift that memory window reduction is mainly caused by hole trap generation

Interpretation of memory window narrowing of FeFET

• Amounts of trapped electrons and trapped holes at readout of V_{th} are controlled by hold voltage and hole time (read time after write)

• Note that no fixed charge generation is assumed in this figure

V_{th}

hole trap generation

electron trap generation

Stress cycle number

Insufficient hole de-trapping
⇨ more holes trapped at V_{hold}
⇨ V_{th} becomes lower with cycling

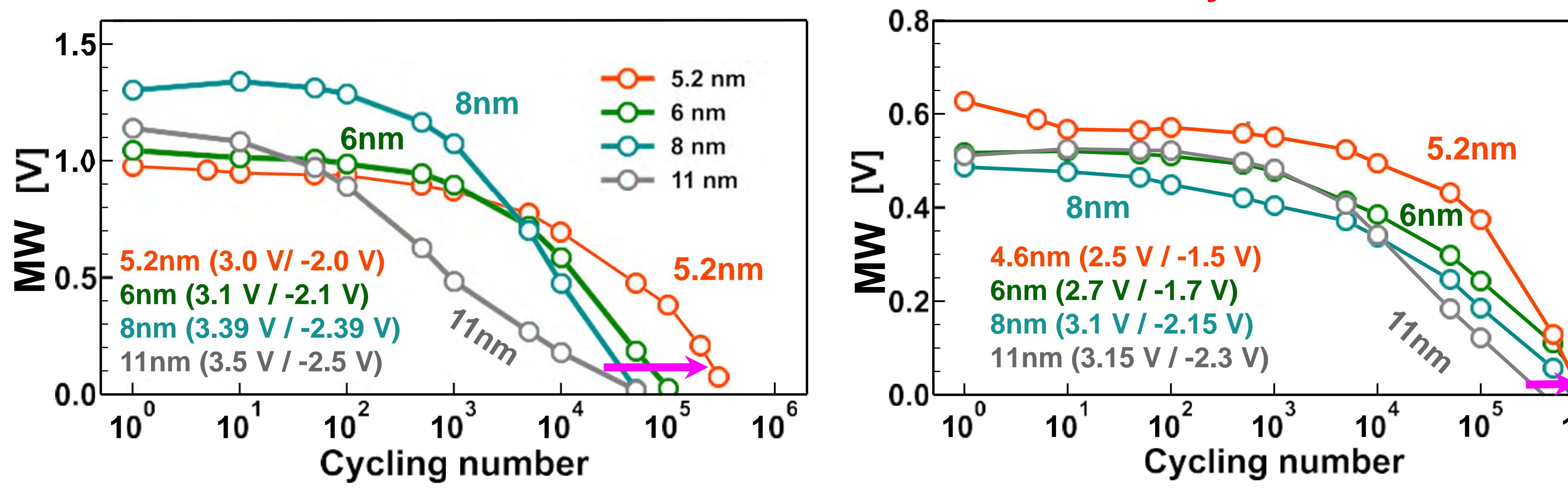
Insufficient electron de-trapping
⇨more electrons trapped at V_{hold}
⇨ V_{th} becomes higher with cycling

S.-K. Cho et al., IEDM, 4-7 (2024)

❑ FeFETs have two aspects in memory operation: polarization switching memory function and trap memory function, both of which compete in terms of memory window
❑ The physical origin of memory window narrowing is that trapped charges increases with increasing cycle number and that these charges, which have the opposite impact to polarization, are not de-trapped at the V_{th} readout time

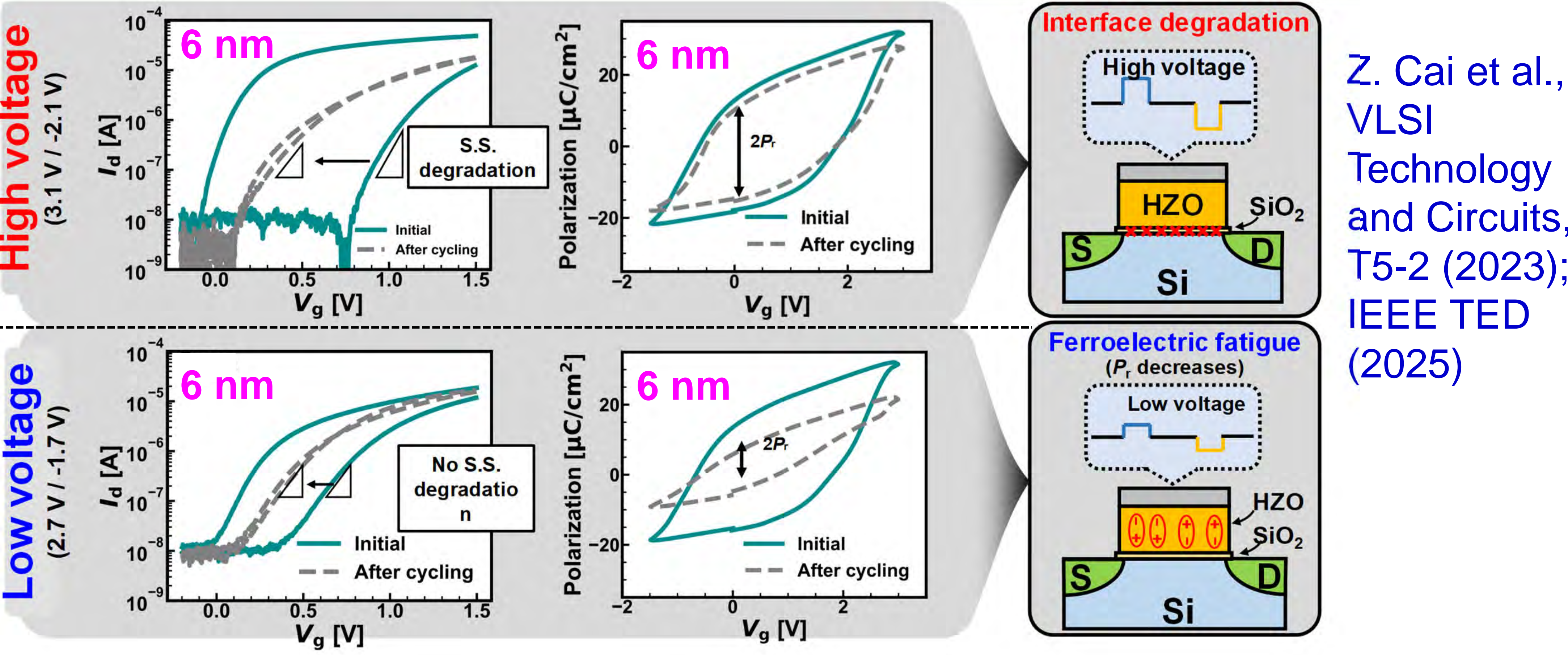
Endurance with different HZO thickness

Z. Cai et al., VLSI Technology and Circuits, T5-2 (2023); publishd in TED (2025)



- ❑ Dominant failure in endurance of FeFETs is known to be fatigue (memory window narrowing), rather than gate stack breakdown
- ❑ HZO scaling and lower voltage operation lead to smaller fatigue
- ❑ Improvement in endurance due to HZO scaling is still limited, implying the existence of degradation mechanism under low voltage

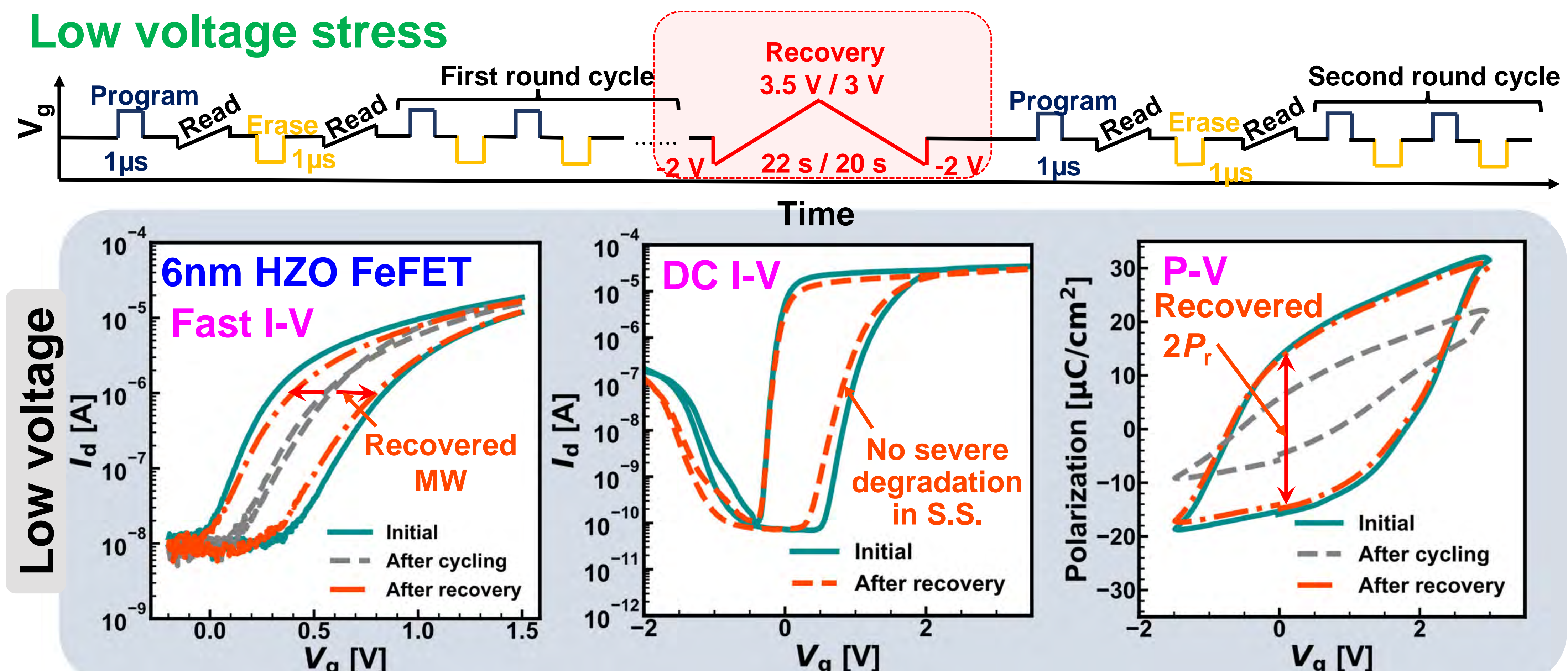
Difference in MW narrowing in high and low voltage operations



Z. Cai et al., VLSI Technology and Circuits, T5-2 (2023); IEEE TED (2025)

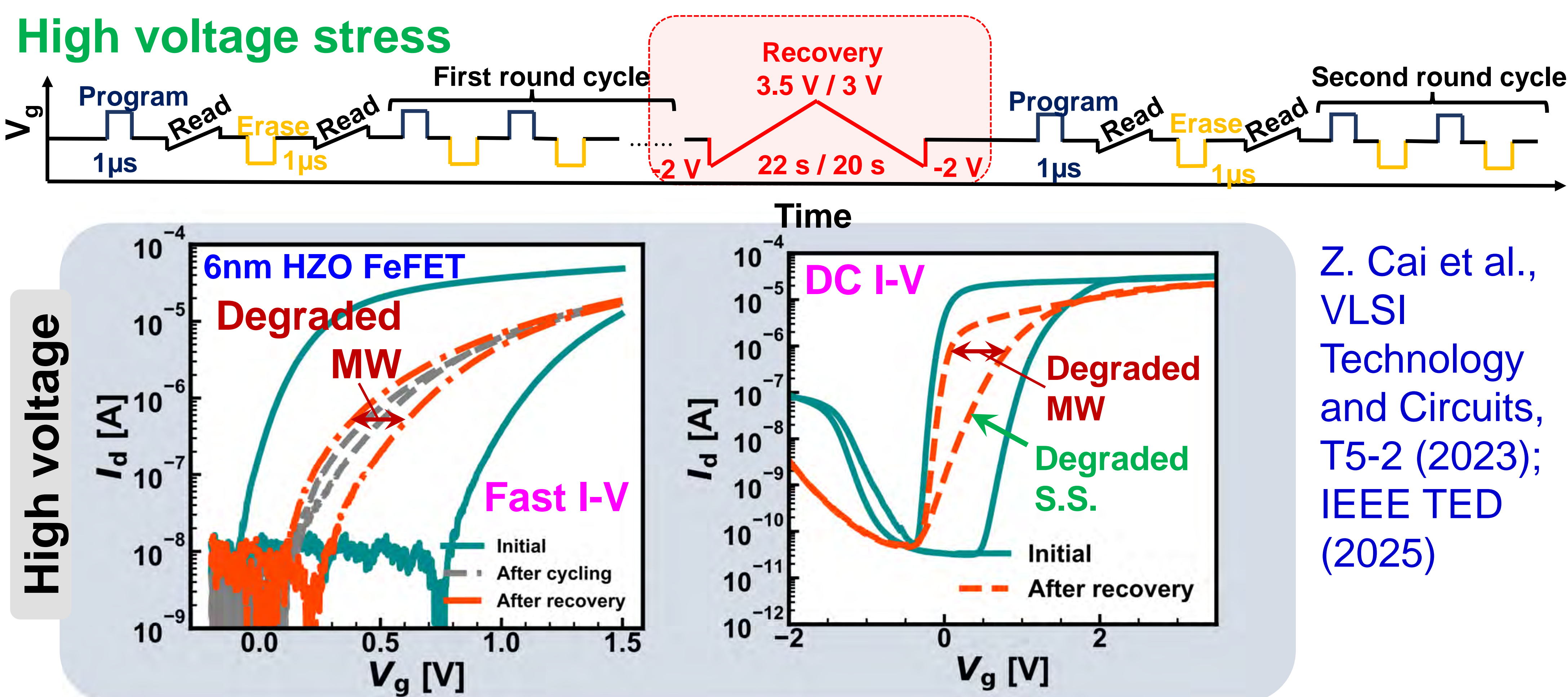
- Fatigue under low voltage is found to be caused by Pr degradation
- Pr degradation is not observed for fatigue under high voltage, which is caused by MOS interface degradation

Recovery of fatigue after low voltage operation



- P_r degradation and resulting MW narrowing under low voltage operation can be recovered by applying a high voltage pulse

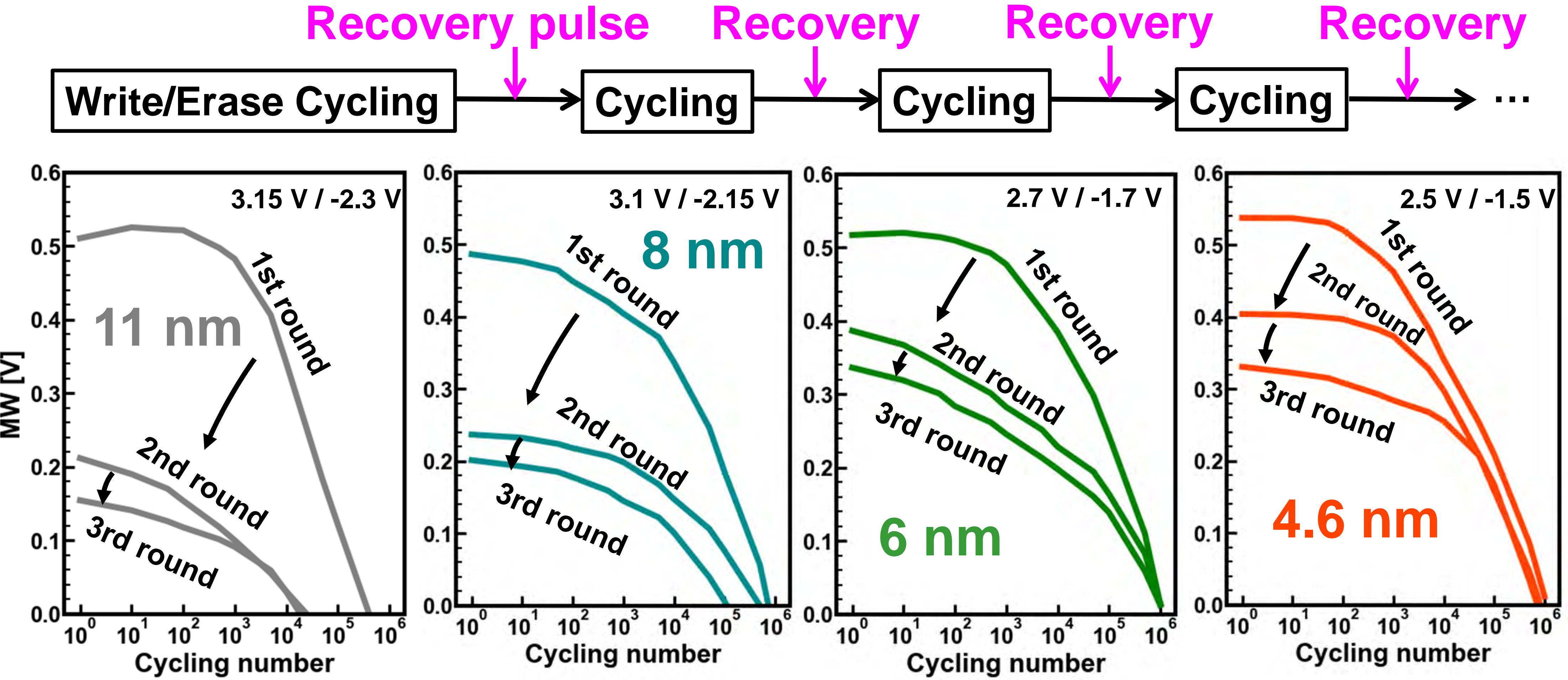
Recovery of fatigue after high voltage operation



Z. Cai et al., VLSI Technology and Circuits, T5-2 (2023); IEEE TED (2025)

- MW narrowing under high voltage operation, which is caused by interface degradation, cannot be recovered by a high voltage pulse

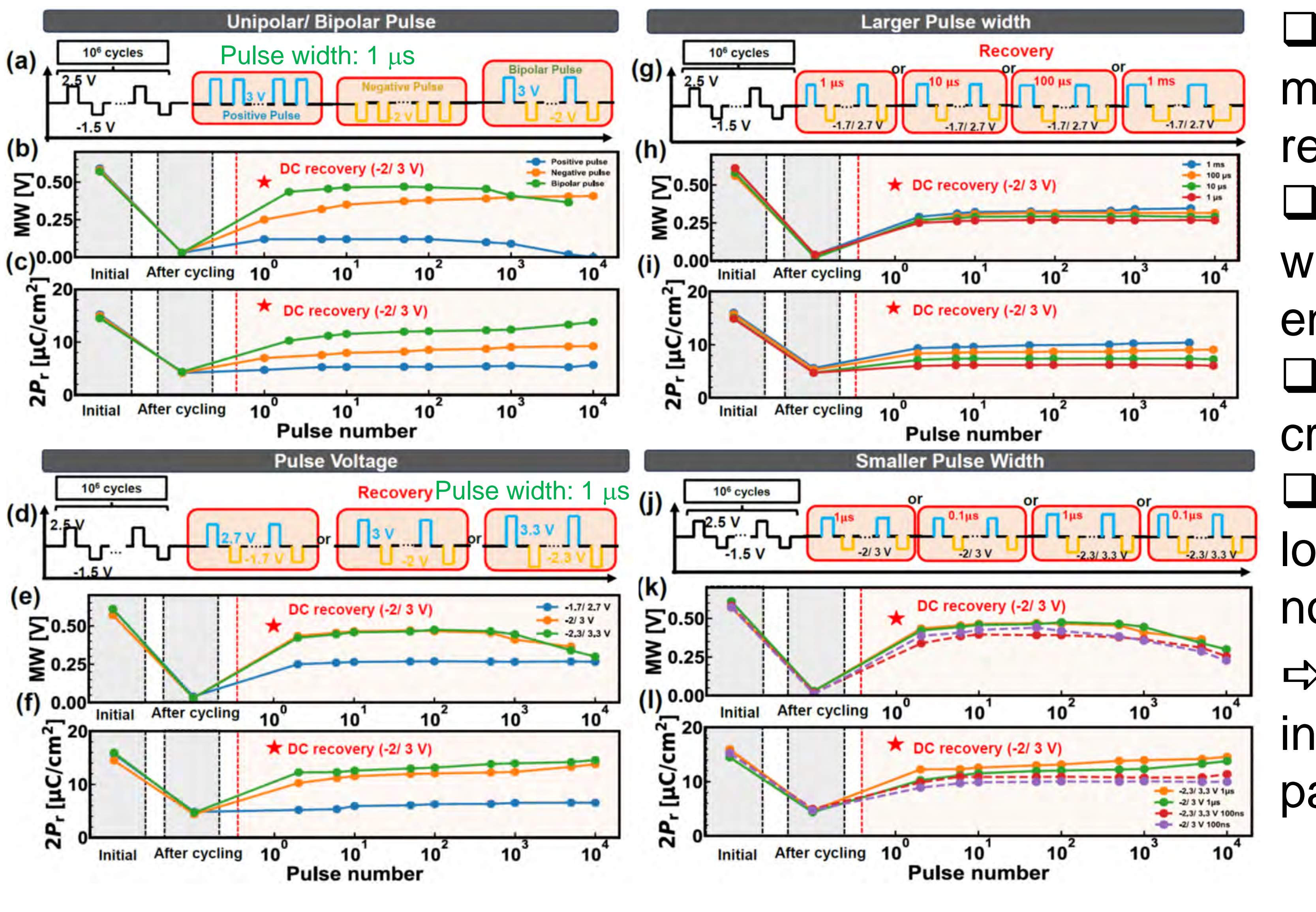
Multiple recovery scheme of FeFETs with different HZO thickness



- A multiple recovery scheme is more effective in improving endurance of thinner HZO FeFETs under lower-voltage operation

Cai, VLSI (2023)

Effect of recovery pulse condition on recovery characteristics



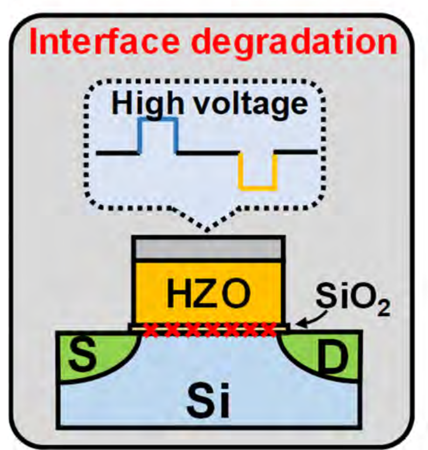
- ❑ Bipolar pulse is most effective for recovery
- ❑ Recovery pulse width of 1 μs is enough for recovery
- ❑ Pulse voltage is critical
- ❑ Longer pulse with lower voltage does not work well

⇨ these results are included in the patent

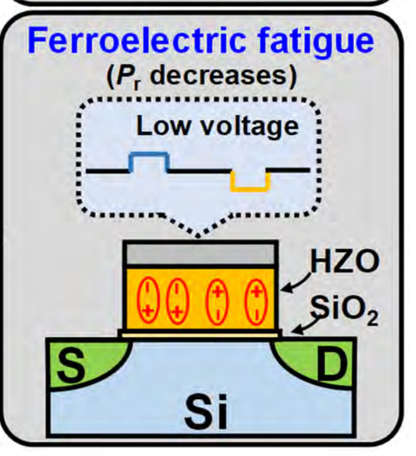
Z. Cai et al., IEEE TED (2025)

Summary of impact of HZO thickness scaling on FeFET endurance

Memory window narrowing → Mixture of two mechanisms



Interface degradation
→ high V_g driven
(hot carrier induced)
→ permanent degradation



Ferroelectric fatigue
(Pr degradation)
→ low E_{FE} driven
(domain pinning)
→ recoverable degradation

Endurance improvement strategy:

Thinner HZO film
(HZO thickness scaling)
→ Use lower V_g
→ Reduce interface permanent degradation
+
→ Use recovery pulse for recoverying remaining P_r degradation

Z. Cai et al., VLSI Technology and Circuits, T5-2 (2023)

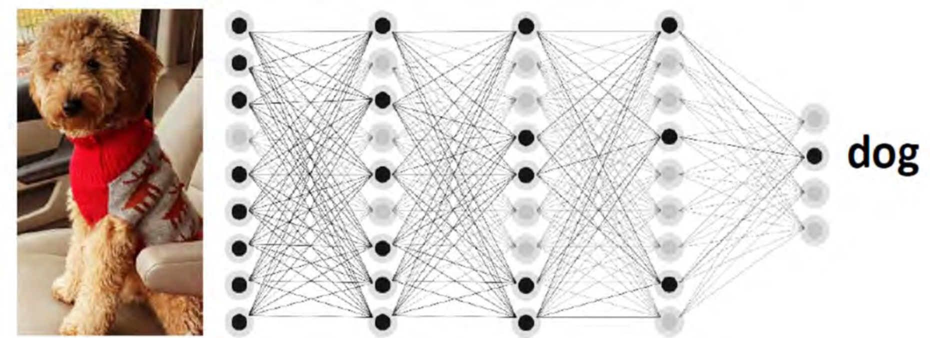
AI applications

Expectation of FeFETs for AI applications

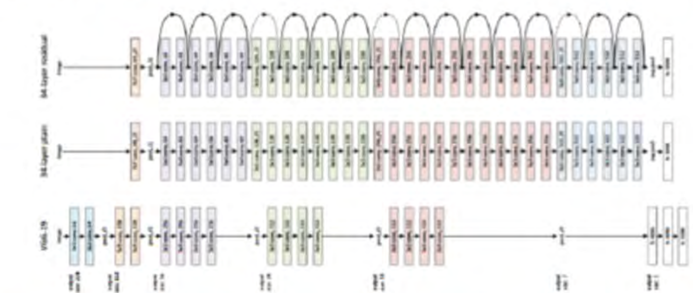
AI revolution by deep neural networks

A. Sebastian, IEDM short course, 2.3. (2023)

Artificial deep neural networks

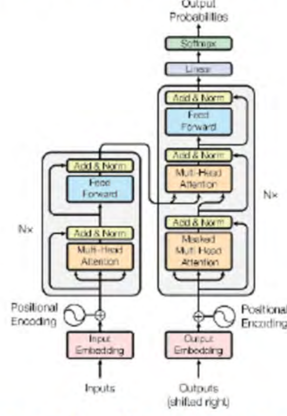


Convolutional neural networks



He et al., "Deep Residual Learning for Image Recognition", CVPR (2016)

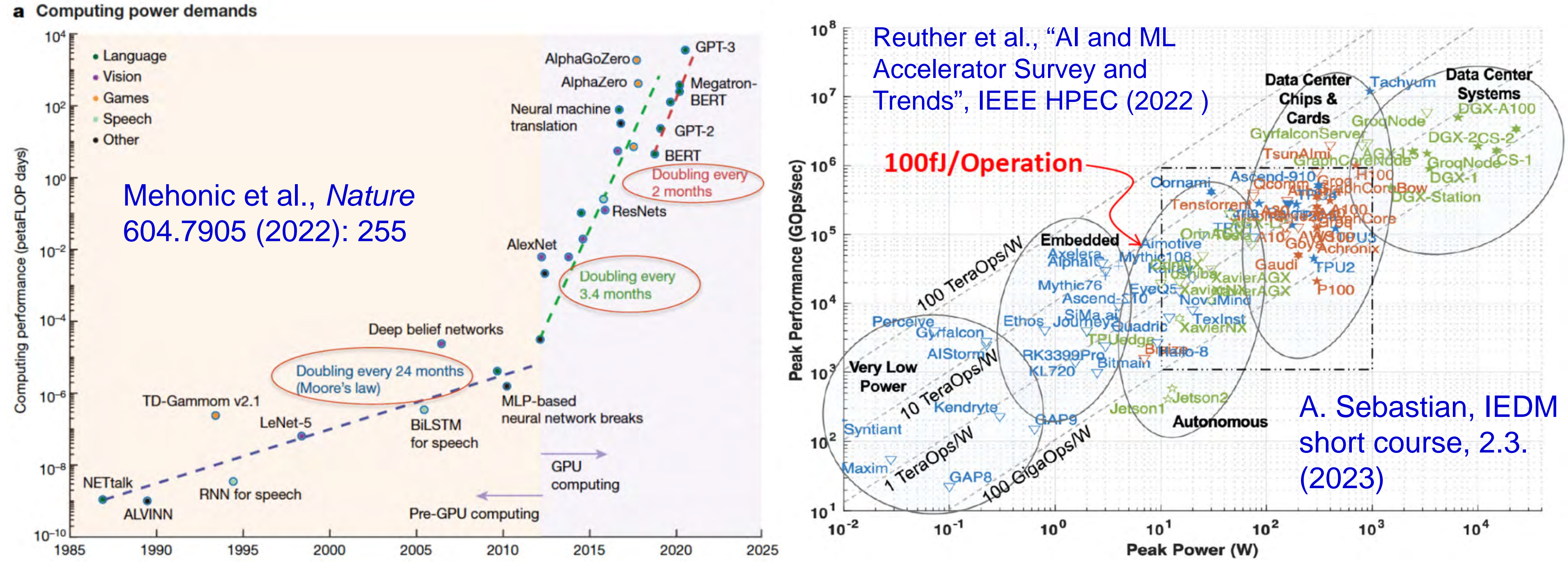
Transformers



Vaswani et al., "Attention is all you need", NeurIPS (2017)

- The machine learning paradigm of artificial deep neural networks (DNNs) has revolutionized AI in recent years
- Powerful tool to learn representations from unstructured data
- Recent consolidation of DNN architectures around convolutional neural networks (CNNs), long-short term memory (LSTM) networks and transformers

Power crisis of AI system

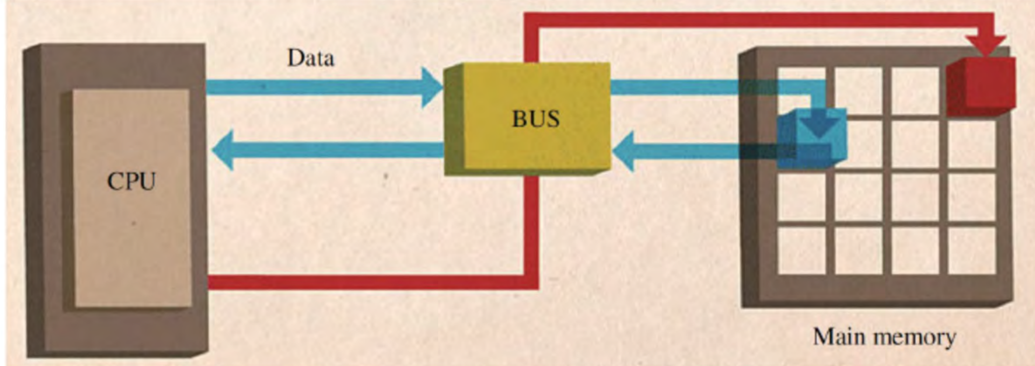


- Software for ML/AI strongly demands rapid growth computing power
- Peak performance vs. power plot, based on present CMOS system, suggests that achieving energy efficiencies better than 100 fJ/Operation is challenging
- Innovation of AI software/hardware is mandatory, particularly for edge AI

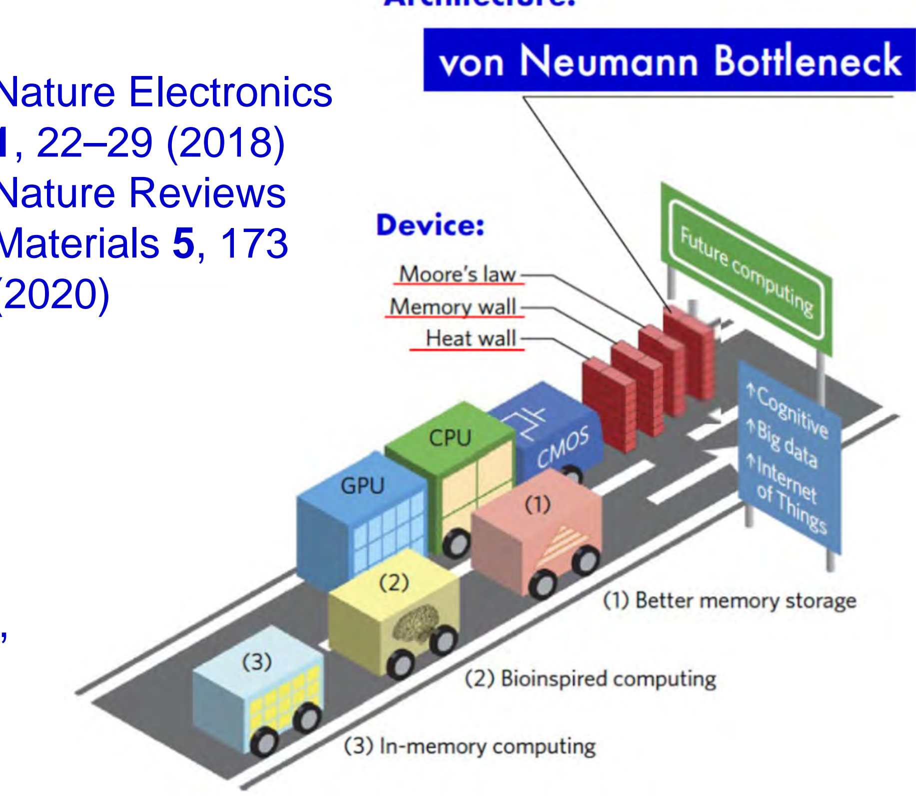
Possible innovation for mitigating power issues of AI system

J. J. Yang, VLSI symp. short course, T-5 (2023)

Classical computer



Nature Electronics **1**, 22–29 (2018)
Nature Reviews Materials **5**, 173 (2020)

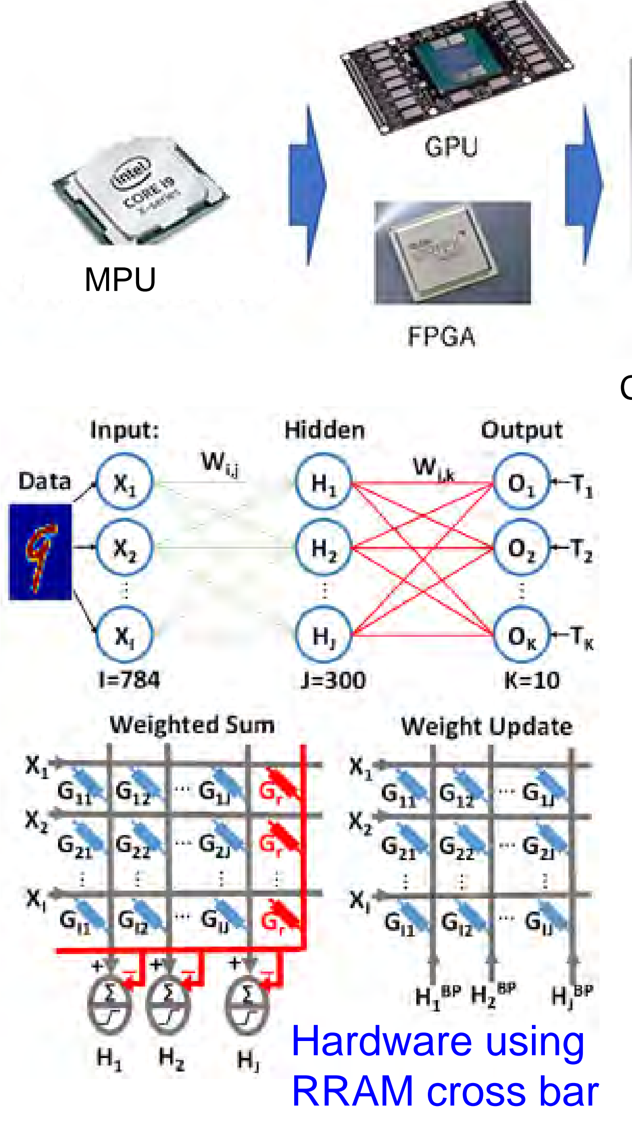


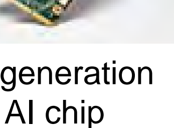
1. Separate processor and storage (**Von Neumann bottleneck**)
2. Sequential process
3. Analog/Digital conversion

Q. Xia, J. Joshua Yang, Nature Materials **18**, 309 (2019)

- In hardware level, AI-specific hardware design, in-memory computing and utilization of analog computing are key technologies
- In architecture level, new algorithm with low computational power and high energy efficiency is strongly expected for edge AI applications

Expectation of AI-specific hardware and AI chip



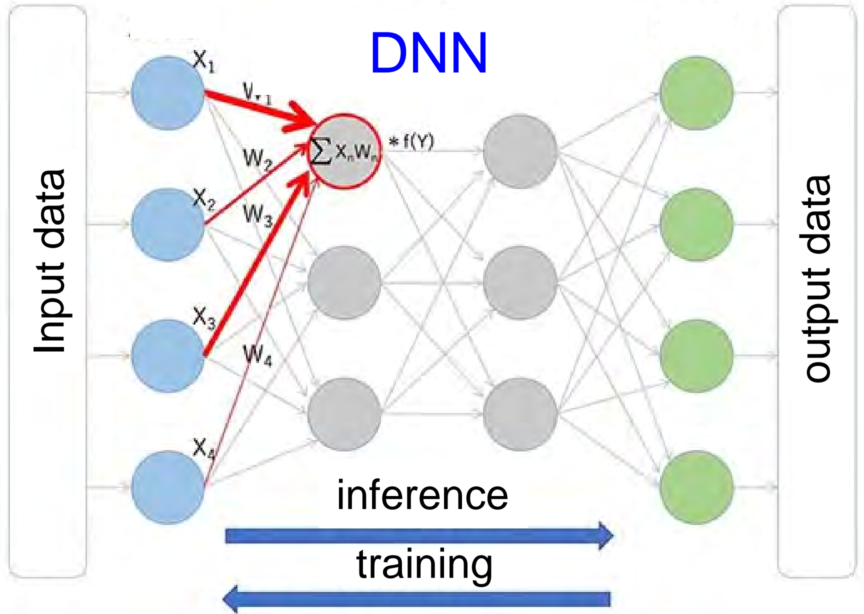


1st generation AI chip

Chip specialized for AI processing

2nd generation AI chip

Chip mimicing computation in brain

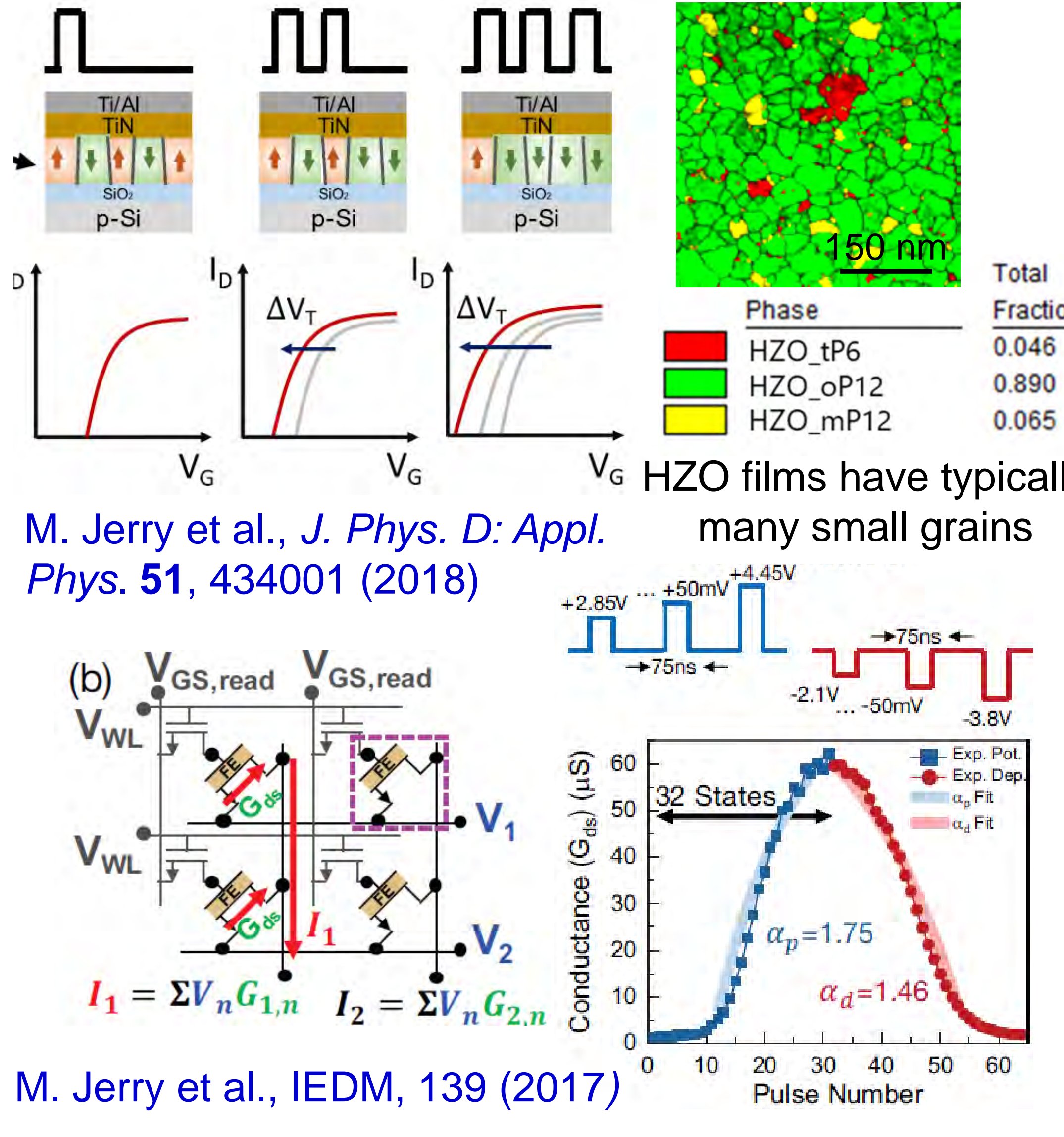


Heavy multiply-add calculation needed

Hardware using RRAM cross bar

- Demands of AI applications are expanding rapidly
- A variety of deep neural network (DNN)-based information processing are mainly performed by software at present
- Von-Neumann bottleneck between logic and memory functions is critical
- Implementation of hardware to perform NN-based calculations is expected, in terms of low power consumption and computation efficiency
- Devices with both memory and switch functions are promising
- Cross-bar arrays using non-volatile memories like RRAM, PCRAM, MRAM and FeRAM etc. are promising for this application

FeFET analog memory (synaptic nature)



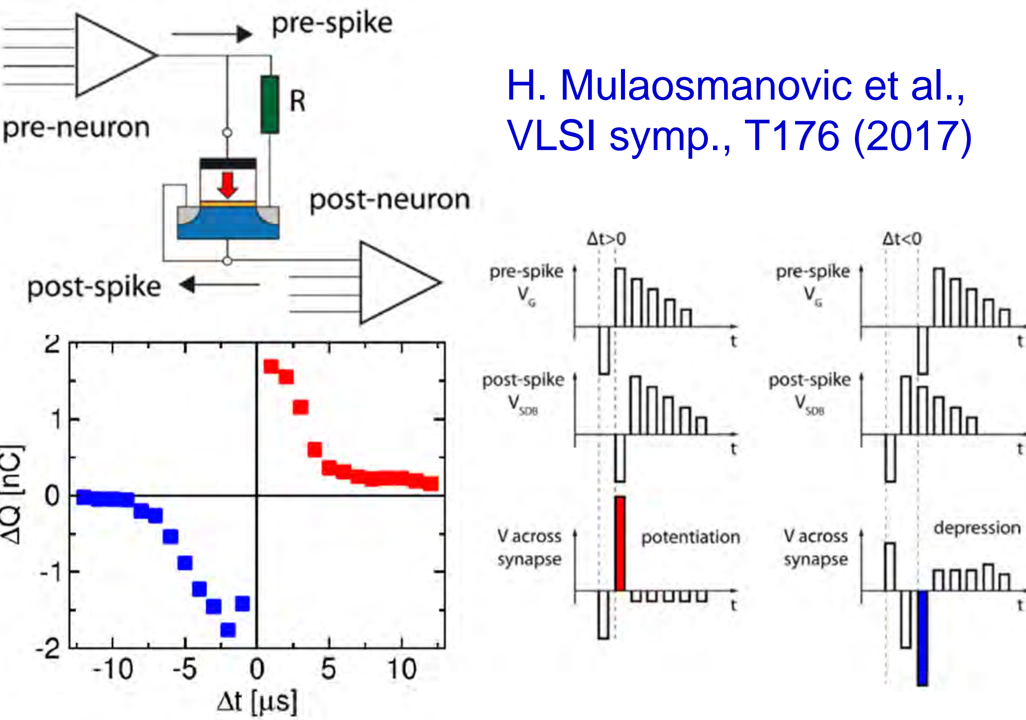
HZO films have typically many small grains

M. Jerry et al., *J. Phys. D: Appl. Phys.* **51**, 434001 (2018)

$$I_1 = \Sigma V_n G_{1,n} \quad I_2 = \Sigma V_n G_{2,n}$$

M. Jerry et al., IEDM, 139 (2017)

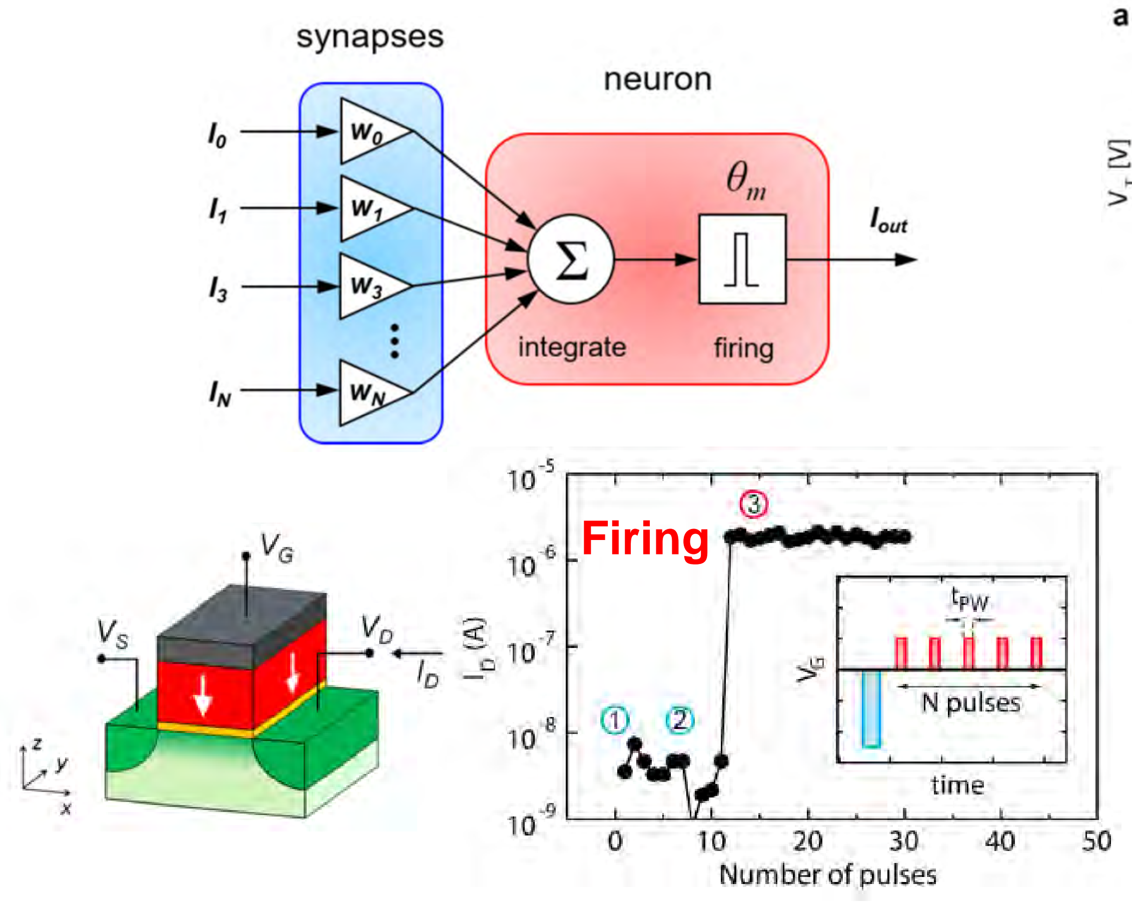
Spike timing dependent plasticity (STDP)

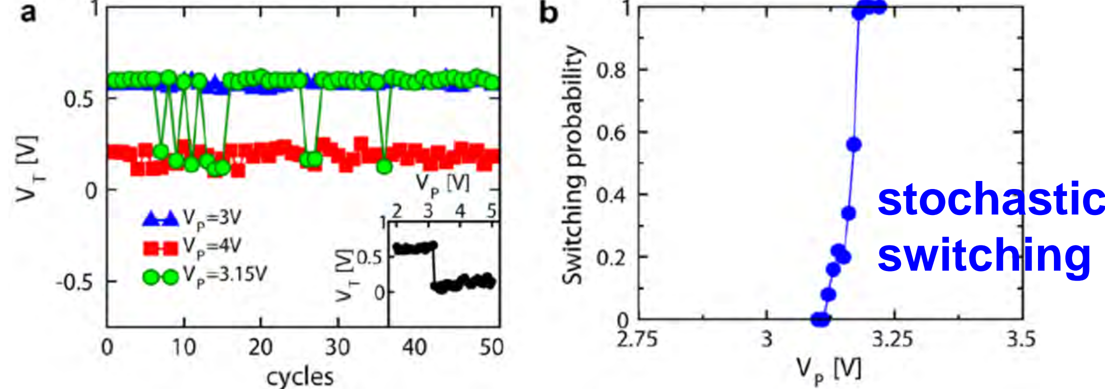


H. Mulaosmanovic et al., VLSI symp., T176 (2017)

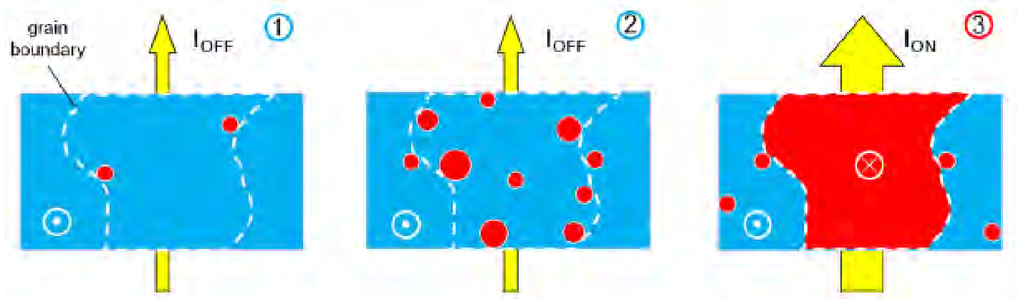
- FeFETs with large areas shows analog memory characteristics, attributable to many localized polarization domains
→ FeFETs can be employed for artificial synapse

Accumulative and stochastic polarization (neuron-like nature)





Schematic view of accumulative polarization switching



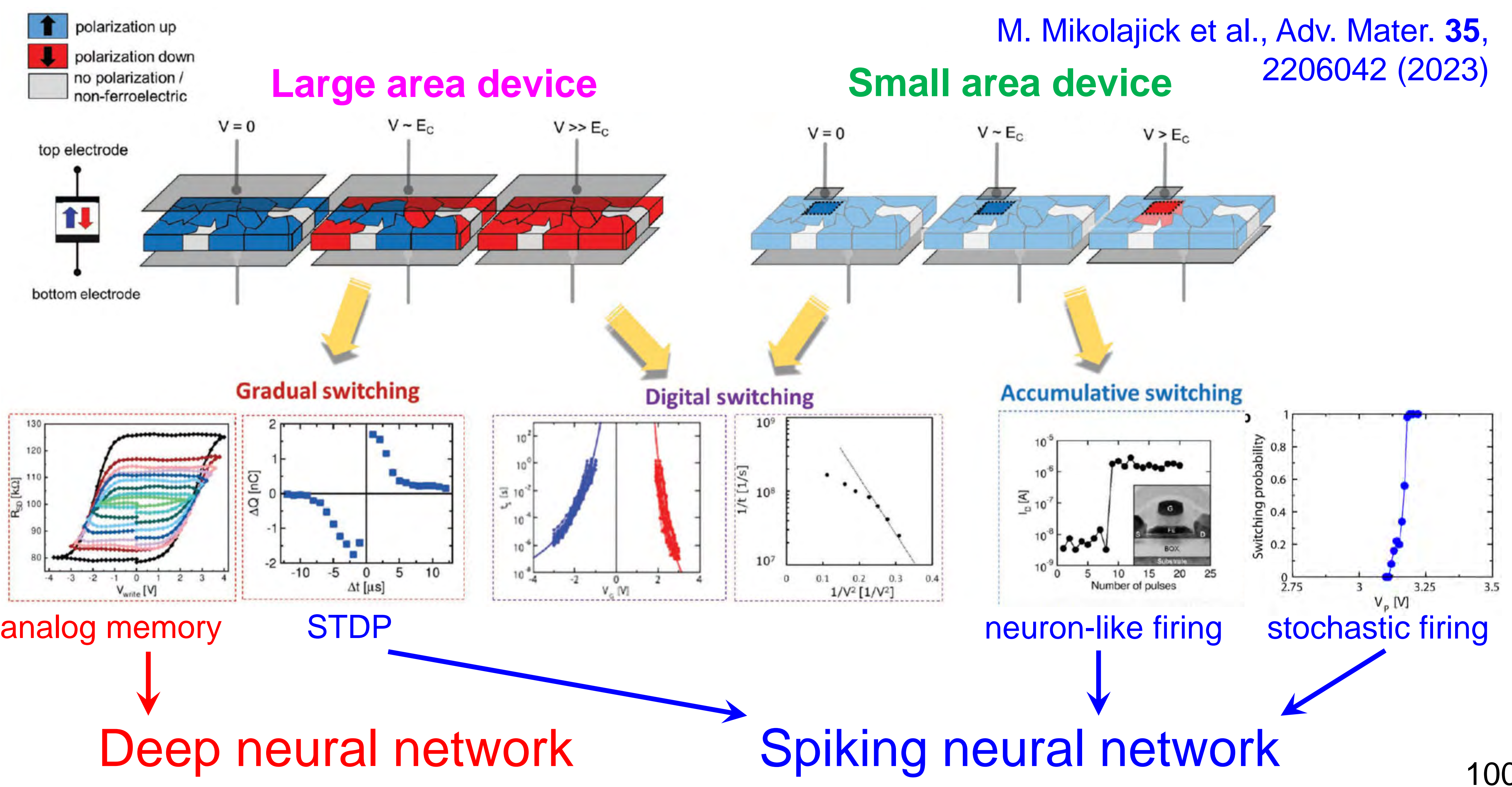
H. Mulaosmanovic et al, ACS Appl. Mater. Interfaces 2017 **9**, 3792 (2017); ACS Appl. Mat. Interfaces **10**, 23997 (2018)

- Polarization reversal occurs through an invisible “polarization domain nucleation” process, resulting in accumulative and stochastic behavior

→ utilized as artificial neurons

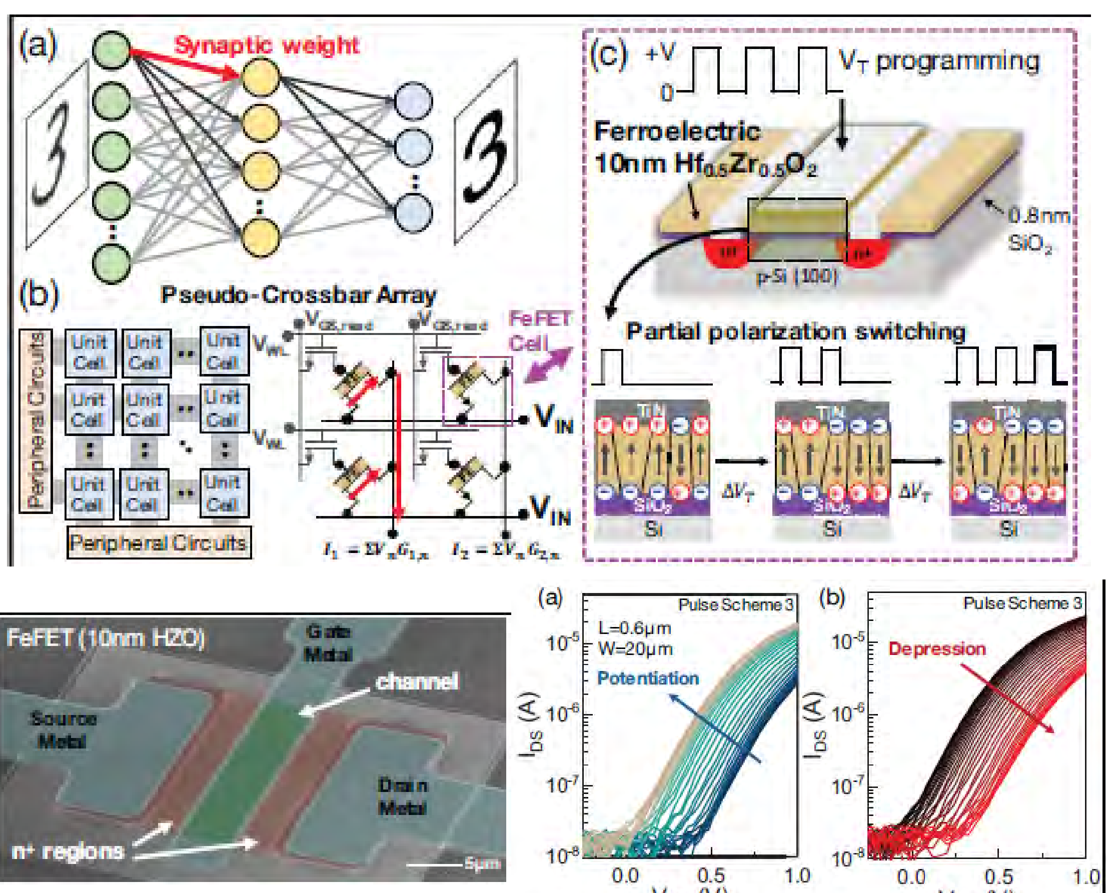
Different natures of FeFET and their applications to neuromorphic computing

M. Mikolajick et al., Adv. Mater. **35**, 2206042 (2023)



Application of FeFET to cross-bar array for DNN calculation

M. Jerry et al., IEDM 139 (2017)

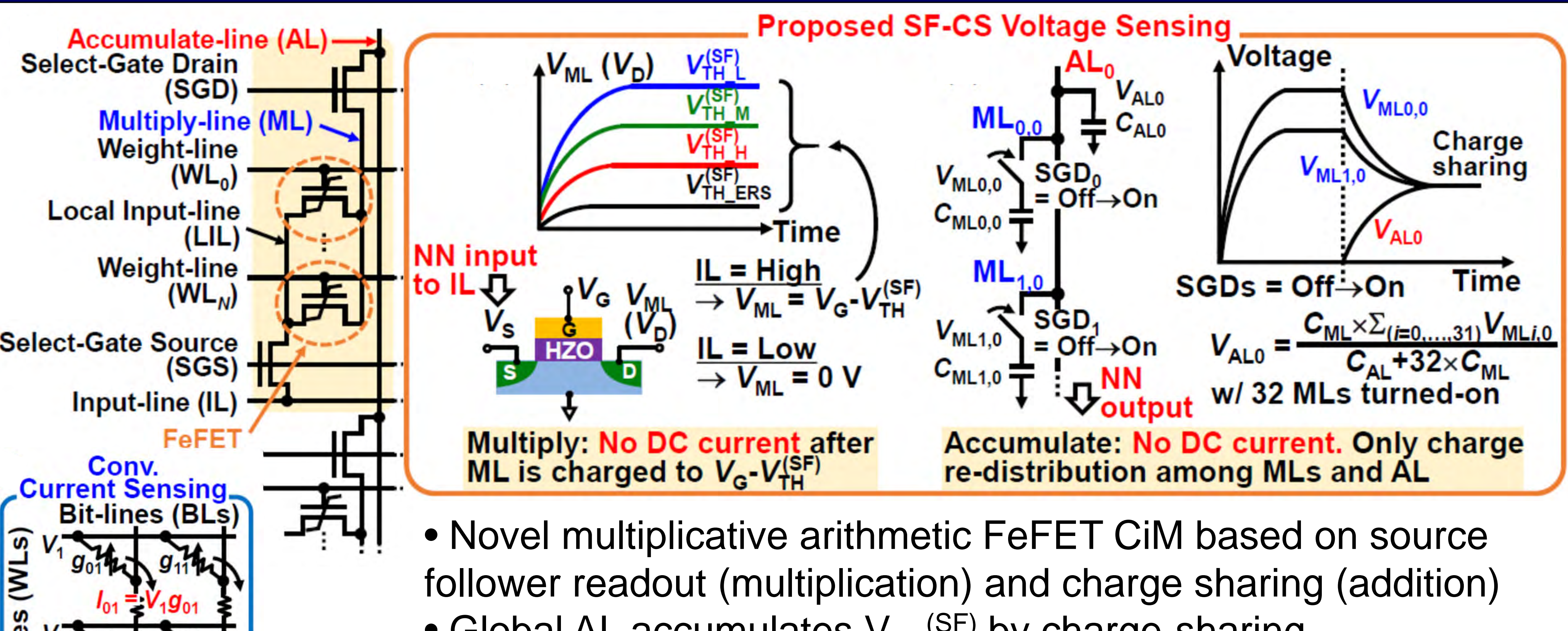


Analog Synapse Devices			
Type	Filamentary RRAM	Interfacial RRAM	Ferroelectric
Mechanism	current + field assist	current + field assist	electric-field
Prototypical behavior	identical pulses; Conductance vs. Pulse # [2]	identical pulses; Conductance vs. Pulse # [3] [4] [5]	non-identical pulses; n+ p-Si n+; Conductance vs. Pulse # This Work
Symmetry	Low	High	High ✓
Accuracy	Low	High	High ✓
Speed	ns	ms	ns ✓

Analog eNVM type	TaO_x/TiO_2 [3]	PCMO [4]	Ag:a-Si [5]	AlO_x/HfO_2 [2]	FeFET (This Work)	6-bit SRAM
# of conductance states	102	50	97	40	32	—
Nonlinearity (weight increase/decrease)	0.66/-0.69	3.68/-6.76	2.4/-4.88	1.94/-0.61	1.75/1.46	—
Asymmetry	1.35	10.44	7.28	2.55	0.29	
R_{ON}	5 MΩ	23 MΩ	26 MΩ	16.9 kΩ	559.28 kΩ	
ON/OFF ratio	2	6.84	12.5	4.43	45	—
Weight increase pulse	3V/40ms	-2V/1ms	3.2V/300μs	0.9V/100μs	3.65V (avg.)/75ns	—
Weight decrease pulse	-3V/10ms	2V/1ms	-2.8V/300μs	-1V/100μs	-2.95V (avg.)/75ns	—
Weight update cycle-to-cycle variation (σ)	<1%	<1%	3.5%	5%	<0.5%	—
Accuracy for online learning	~10%	~10%	~73%	~41%	~90%	~94%
Area	1,071.3 μm^2	1,071.3 μm^2	1,072.0 μm^2	3,657.2 μm^2	1,190.4 μm^2	10,311 μm^2
Latency for online learning (1M images)	1132 years (3.57×10^{10} s)	22.19 years (7.00×10^8 s)	13.3 years (4.20×10^8 s)	1.77 years (5.60×10^7 s)	9.33 hours (3.36E4 s)	7.76 s
Energy for online learning (1M images)	65.86 mJ	29.4 mJ	87.94 mJ	150 mJ	98.01 mJ	6.98 mJ
Leakage power	35.29 μW	35.29 μW	35.29 μW	35.29 μW	35.29 μW	1.1 mW

- Early demonstration of application of FeFET analog memory to synaptic weights in cross-bar array for deep neural network calculations with emphasis on high symmetry and fast writing speed

High energy efficiency FeFET CiM with source follower readout

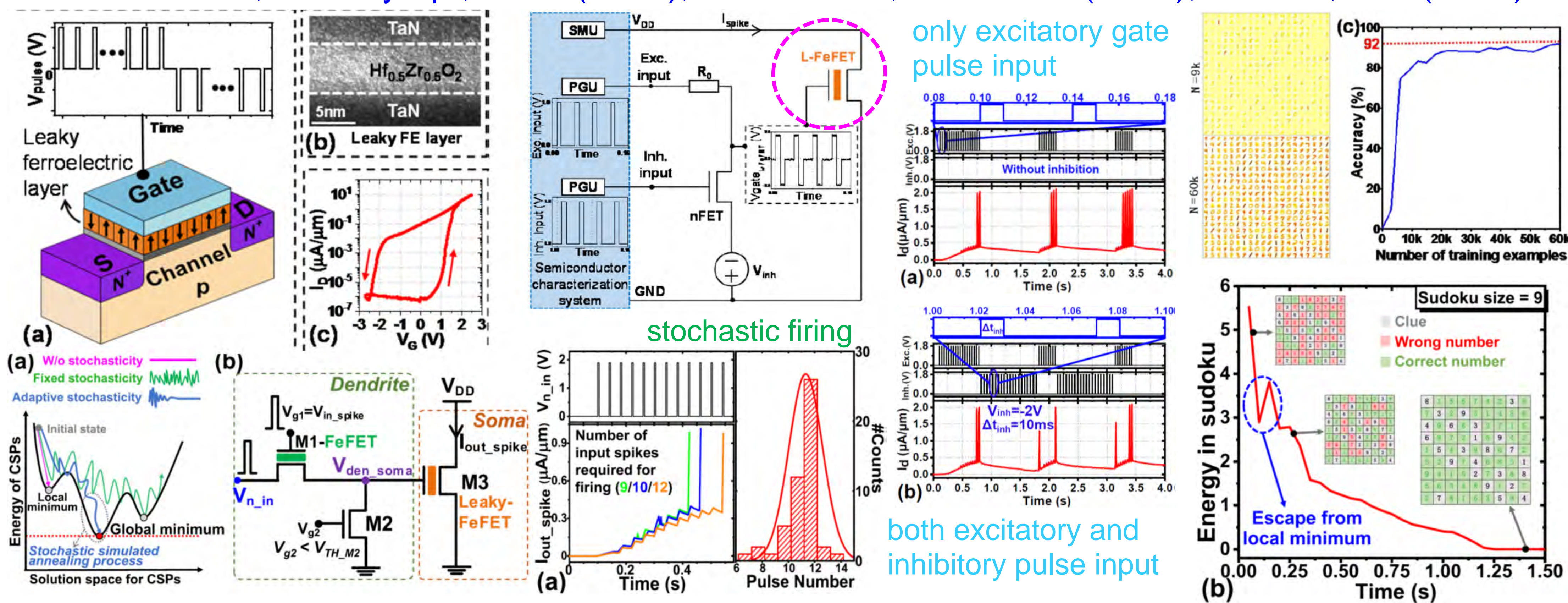


- Novel multiplicative arithmetic FeFET CiM based on source follower readout (multiplication) and charge sharing (addition)
- Global AL accumulates $V_{TH}^{(SF)}$ by charge-sharing
- Achieve massively parallel operations without DC current → low power (high energy efficiency) & accurate inference

C. Matsui et al., VLSI symp. (2021) TJFS2-8; Trans. VLSI System **32**, 468 (2024)

Application of Leaky-FeFET to spiking neural network

C. Chen et al., VLSI symp., T136 (2019), J. Luo et al., IEDM 122 (2019); EDL **43**, 308 (2022)



• Leaky-FeFET with thin HZO and larger gate leakage current was demonstrated to show the leaky integrate-and-fire (LIF) function, which was utilized spiking-neural network

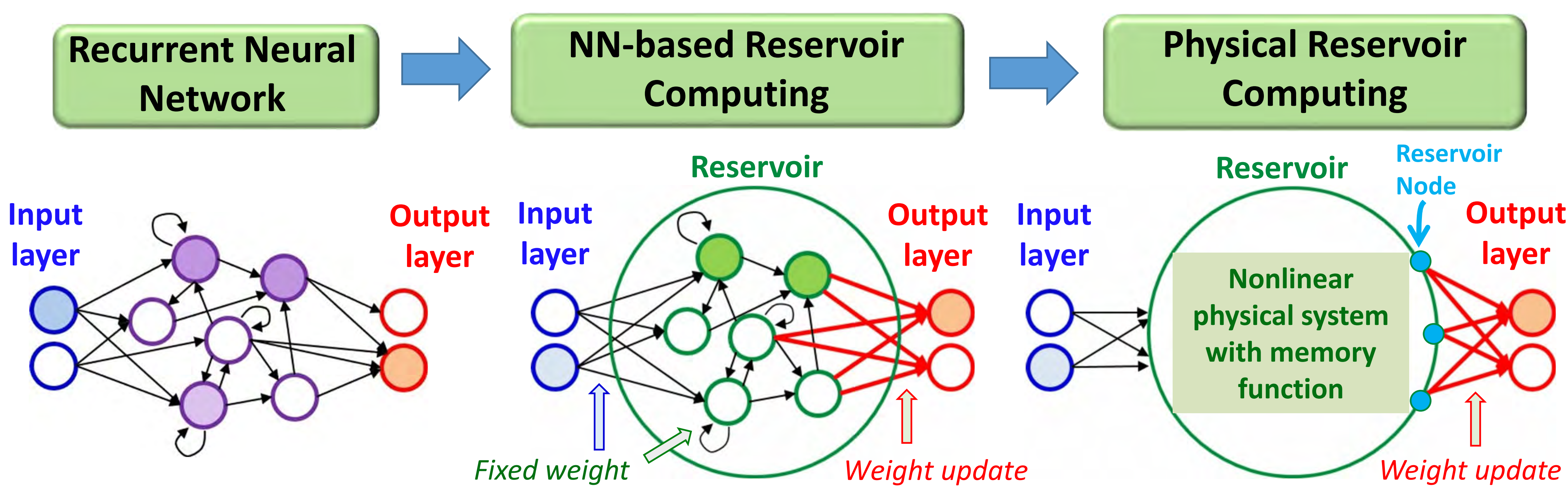
• Neuron firing due to polarization switching is controlled by excitatory and inhibitory gate pulse input

• Stochastic polarization function of FeFET is utilized for SNN-based inference and optimization problems

Physical reservoir computing using FeFETs

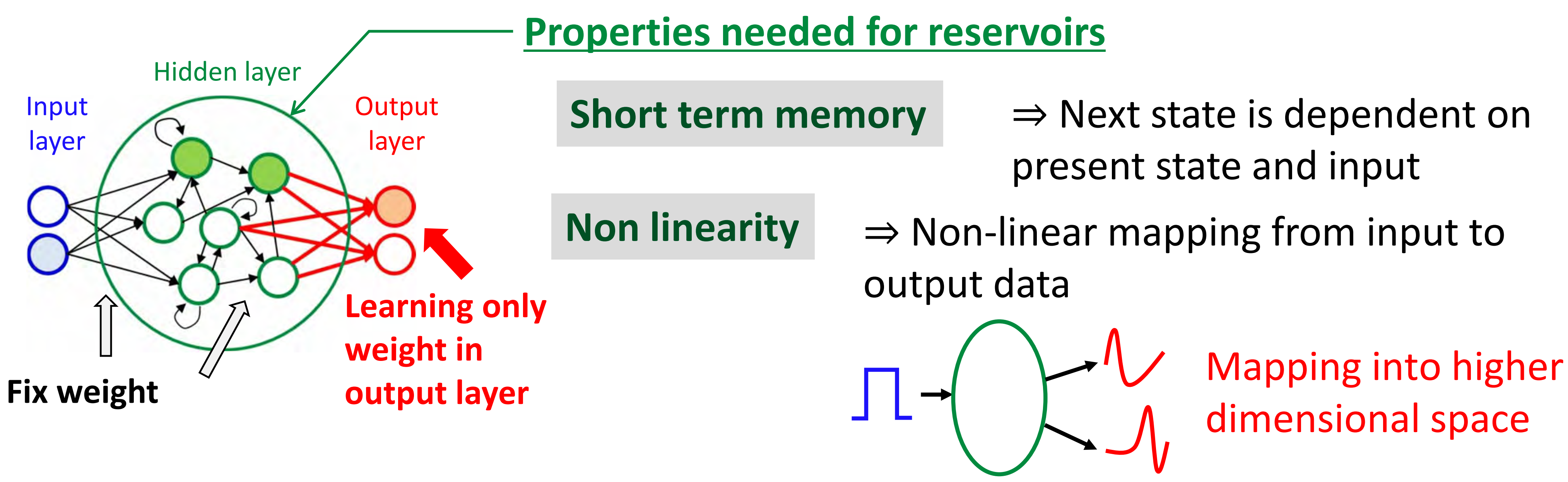
Physical reservoir computing

G. Tanaka et al., Neural Networks 115, 100 (2019)



- Recurrent neural network ⇒ suitable for processing time-series data
- Reservoir computing ⇒ Only weights of a single output layer are trained (other weights are fixed) ⇒ Training with high speed and low energy consumption
- Reservoir can be implemented by a nonlinear physical system (hardware) ⇒ Further reduces computational and hardware cost ⇒ promising for edge AI applications

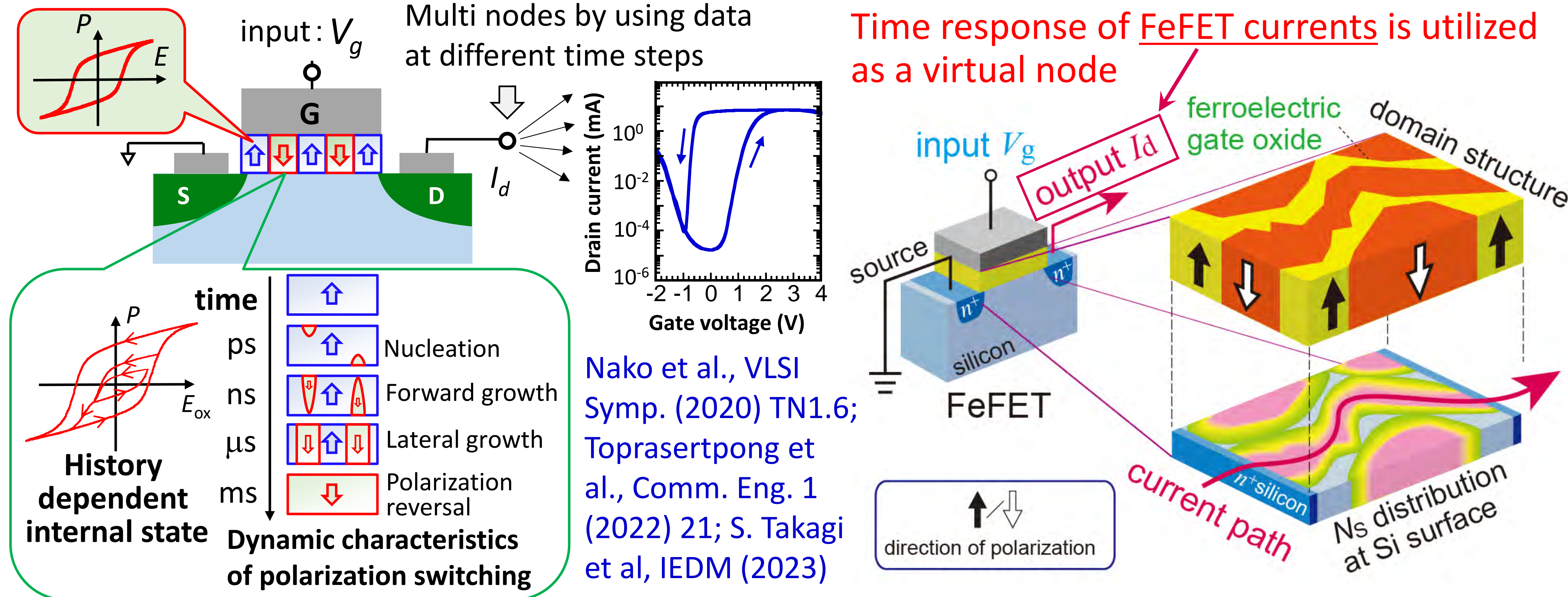
Requirements for physical reservoir



- Reservoir can be implemented by a physical system with short-term memory and rich non-linear functions through dynamics
- We propose FeFET as a reservoir

⇒ realize CMOS-friendly reservoir computing system (high compatibility with Si standard process and easy integration with CMOS circuitry)

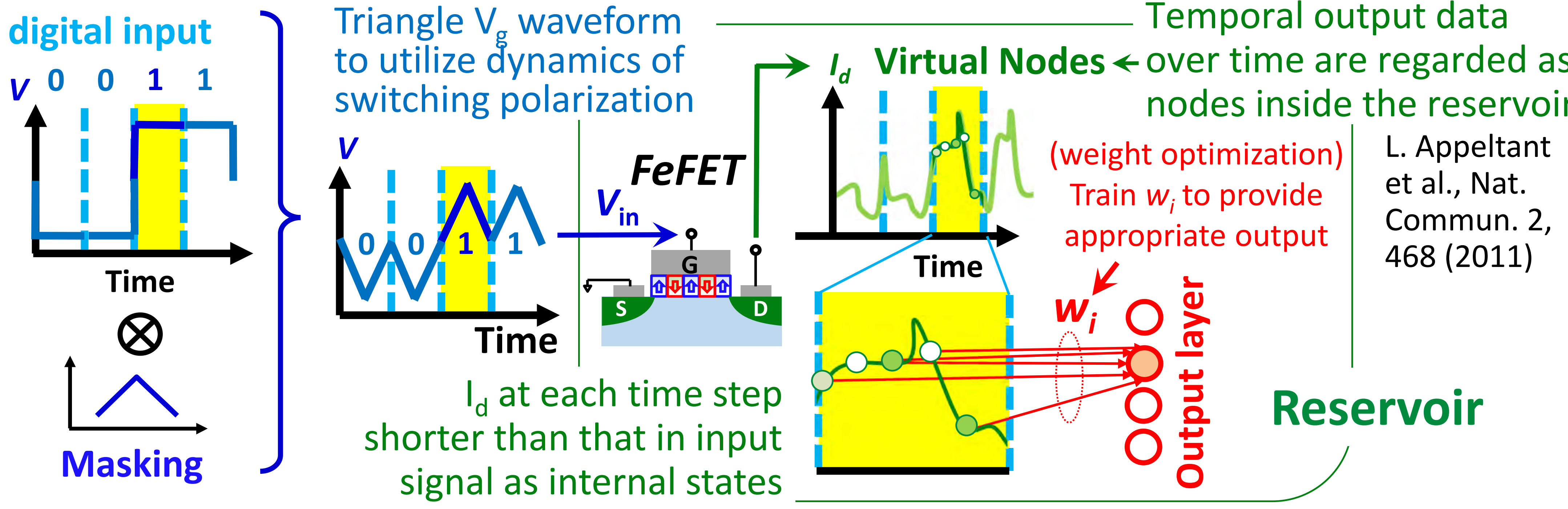
Expectation for FeFET reservoir computing



- Memory function due to polarization and rich non-linearity due to complex time responses of polarization domains
- Enhancing these functions is critical to computing performance improvement

Operation of reservoir computing using FeFET

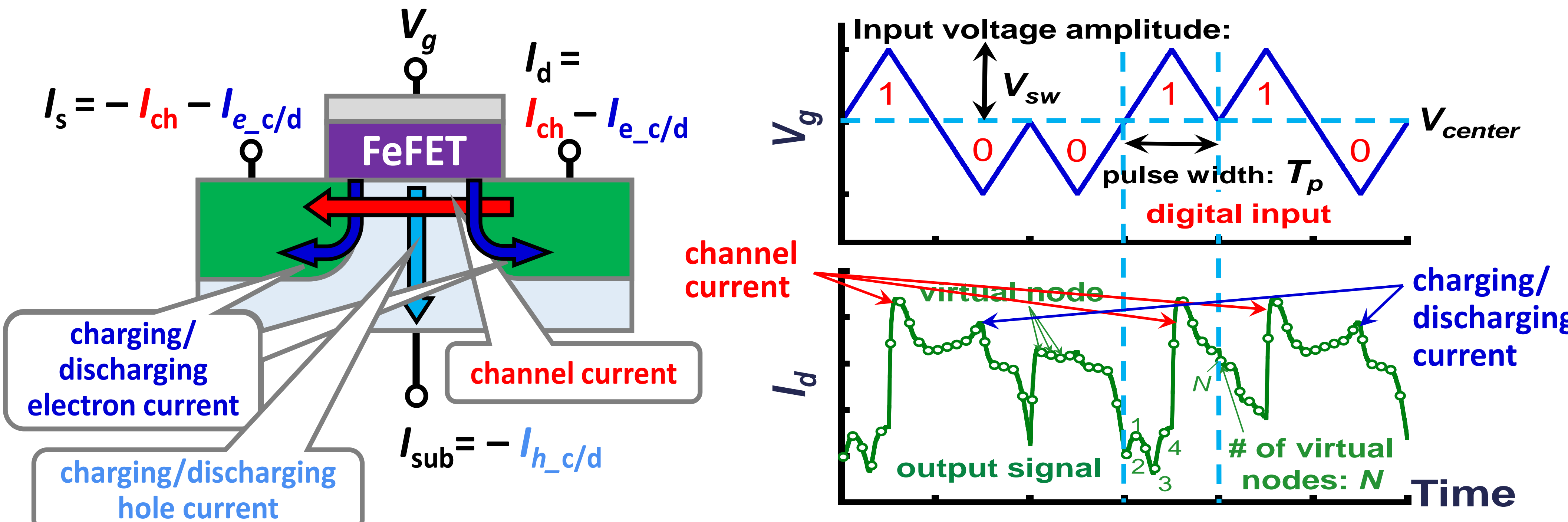
Nako et al., VLSI Symp. (2020) TN1.6; Toprasertpong et al., Comm. Eng. 1 (2022) 21



- How to extract rich information from output signal is a key to successful reservoir computing
- It is essential to design and optimize the operating scheme to maximize the performance of the reservoir computing system with FeFETs

Pattern recognition with current waveform

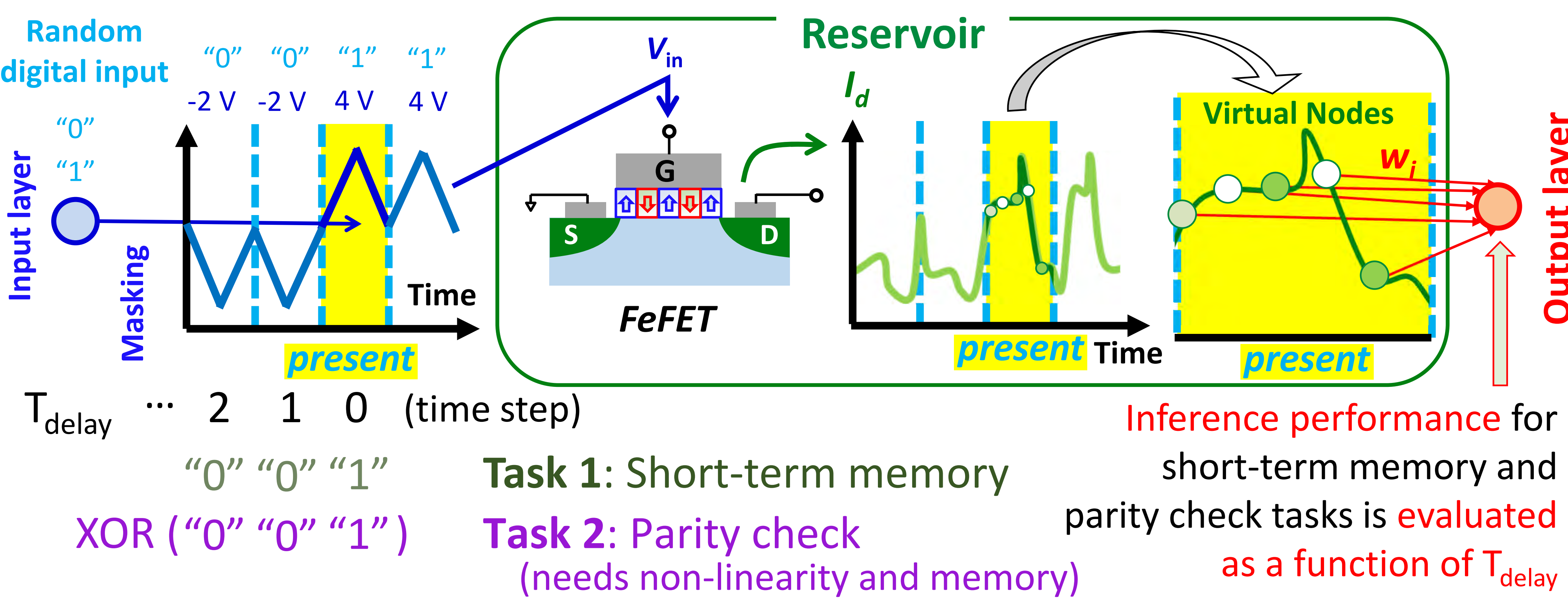
Nako et al., VLSI Symp. (2020) TN1.6; Toprasertpong et al., Comm. Eng. 1 (2022) 21



- Polarization switching can affect both channel current and charging/ discharging current
- The time-series current waveform patterns of FeFETs, which depend on the input history, are utilized for inference

Evaluation of computing performance by basic tasks

Nako et al., VLSI Symp. (2020) TN1.6; Toprasertpong et al., Comm. Eng. 1 (2022) 21



"0" "0" "1" **Task 1**: Short-term memory

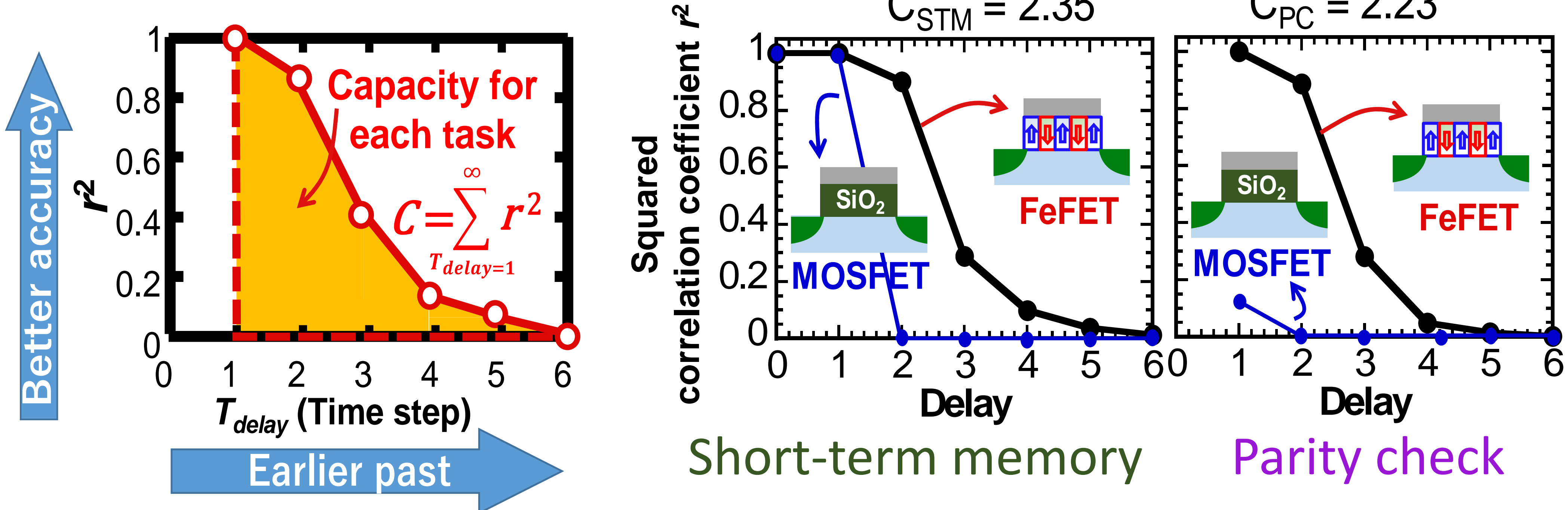
XOR ("0" "0" "1") **Task 2**: Parity check (needs non-linearity and memory)

Inference performance for short-term memory and parity check tasks is evaluated as a function of T_{delay}

- Two basic tasks are performed to evaluate the "short-term memory" and "non-linearity" of the FeFET reservoir as a function of the time delay step

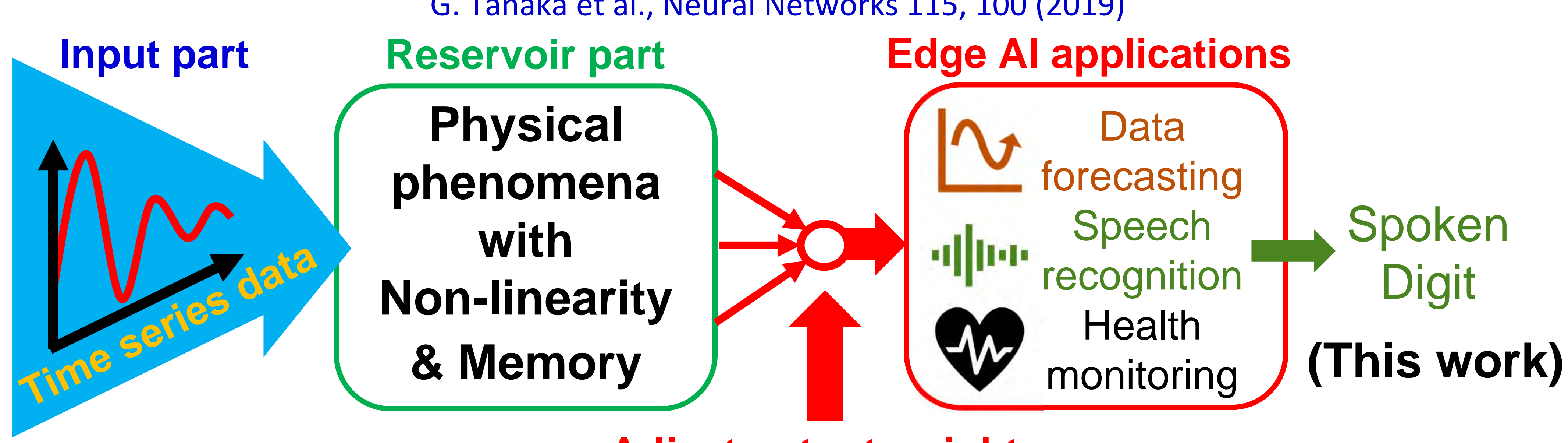
Basic task performance of a single FeFET reservoir

Nako et al., VLSI Symp. (2020) TN1.6; Toprasertpong et al., Comm. Eng. 1 (2022) 21



- Reservoir computing performance is estimated by correlation coefficient as a function of the time delay step and the integrated value (capacity)
- MOSFETs have no reservoir computing performance
- FeFETs exhibit much higher performance, attributed to polarization in HZO

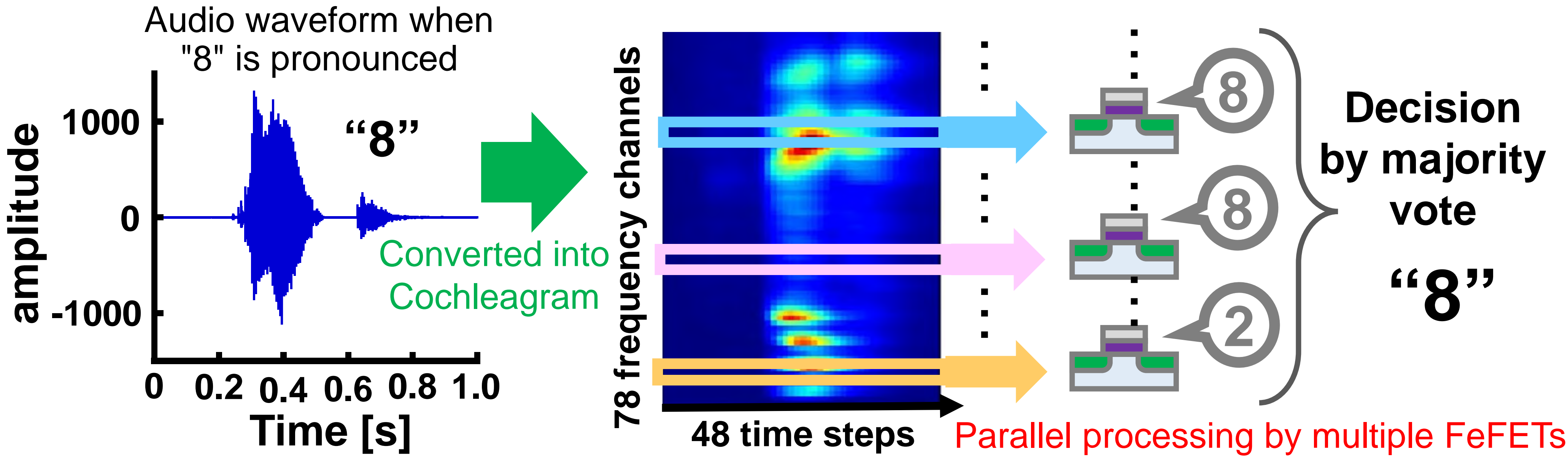
Application of physical reservoir computing



- Potentially important edge AI applications include data forecasting, speech recognition, and health monitoring
- In this study, we apply FeFET reservoir computing to two applications, NARMA as a data forecasting application and spoken digit as a speech recognition application

Spoken digit recognition by FeFET reservoir computing

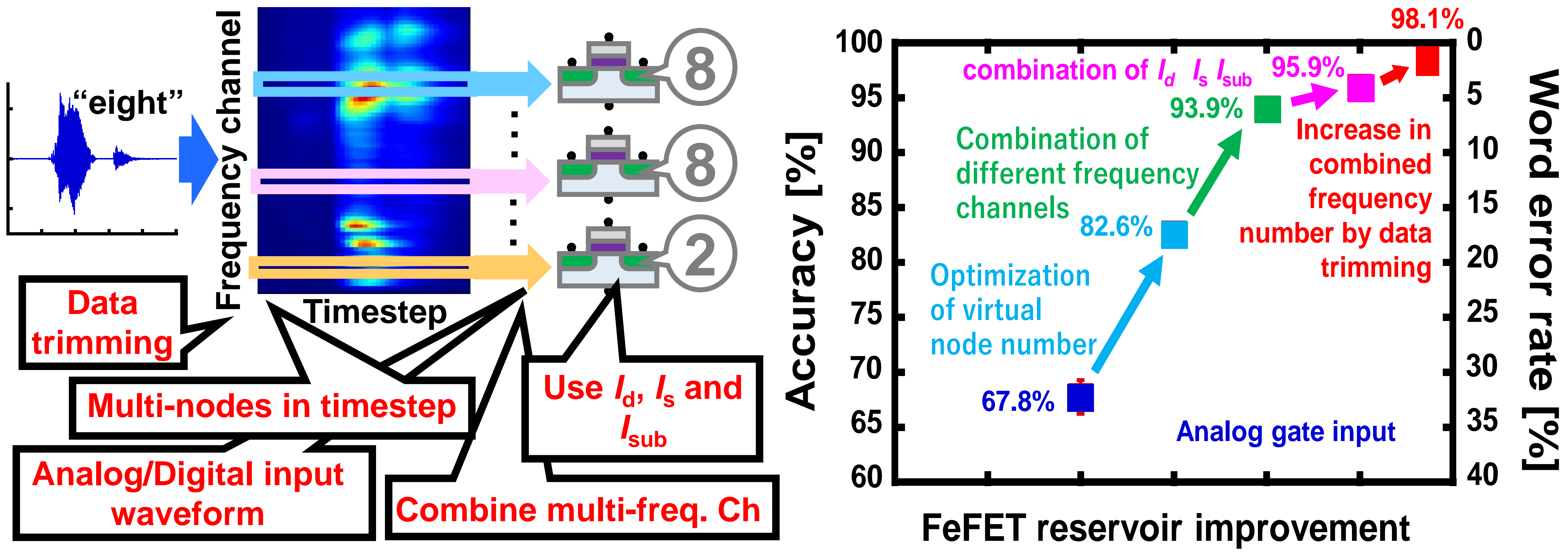
Nako et al., VLSI Symp., 220 (2022) TN1.6; IEEE TED, 70, 5657 (2023)



- Spoken 0-9 digit speech data are converted into cochleagram, which is composed of time series data with multi-frequency channels
- We have proposed a new reservoir computing scheme using parallel processing by multiple FeFETs for spoken digit recognition
- Final decision is made by a majority vote of inference by multiple FeFETs

Improvement of accuracy in FeFET reservoir computing

E. Nako et al., Symp. VLSI Tech., 220 (2022); S. Takagi et al., IEDM (2023)



- Optimization of virtual node number, an adaptation of analog input, and the optimum combination of multi-frequency channel data additively contribute to an increase in recognition accuracy
- FeFET reservoir computing achieves 98.1% classification accuracy

Acknowledgement

This work was supported by JST CREST (JPMJCR20C3) and JSPS KAKENHI (21H01359).

I have appreciated Kasidit Toprasertpong, Zuocheng Cai, Zhenhong Liu, Seong-Kun Cho, Masaki Otomo, Kento Tahara, Ryosho Nakane, Eishin Nako, Koichiro Iwashige, Tsung-En Lee, Shin-Yi Min, Rikuo Suzuki, Koshuke Ito, Xuan Luo, Makoto Kawano, Zeyu Wang, Zaoyang Lin, Mingxia Wan, Mitsuru Takenaka in the University of Tokyo, and Yukinobu Hikosaka, Ko Nakamura, and Hitoshi Saito in Fujitsu Semiconductor Memory Solution for their cooperations and technical supports

Thank you very much for your kind attention !