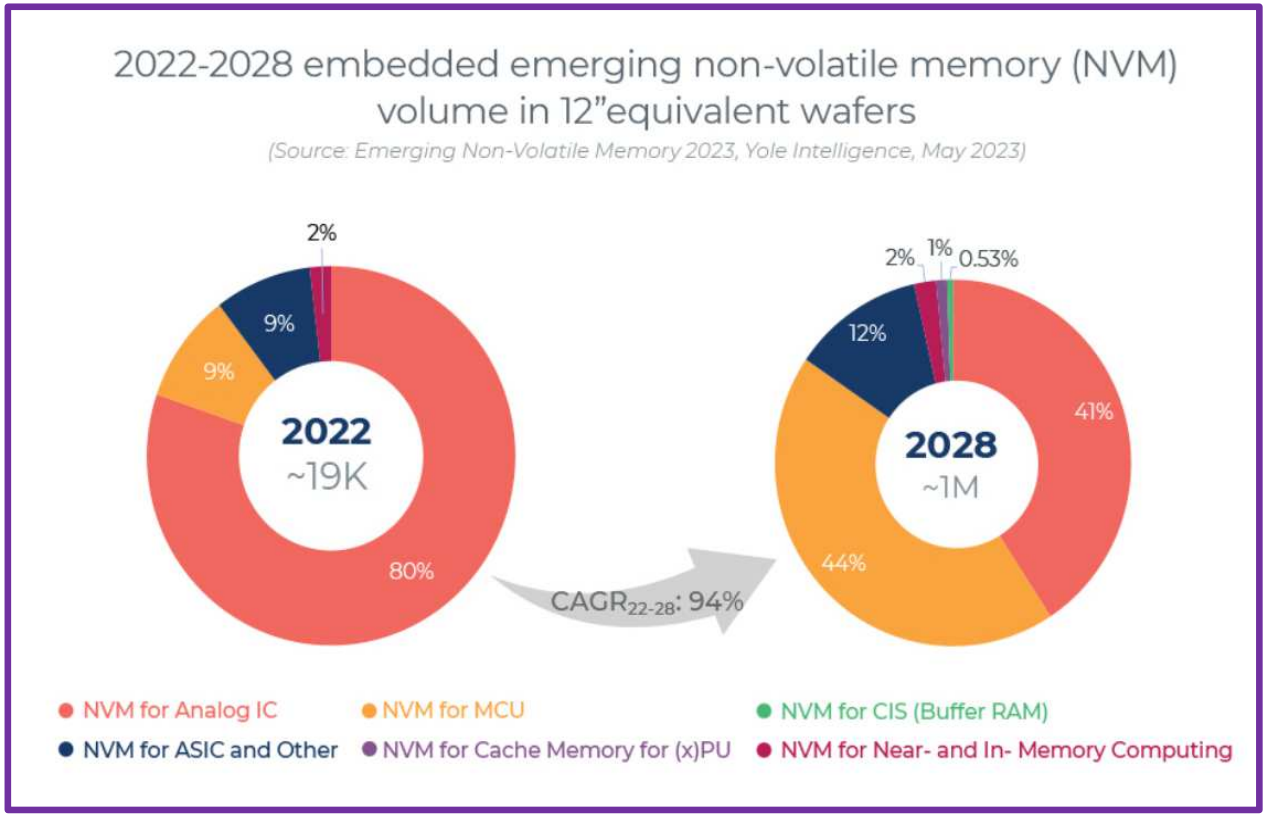
Enabling new research paths with embedded PCM

Andrea Redaelli
Fellow – ePCM Architecture Technical Director
STMicroelectronics, central TR&D, Agrate Brianza (IT)

eNVM market trend

- Expanding emerging eNVM market
- Strong IOT, Industry 4.0 and automotive trends
- Flash NVM below 28nm extremely complex
- Room for emerging NVM below 28nm

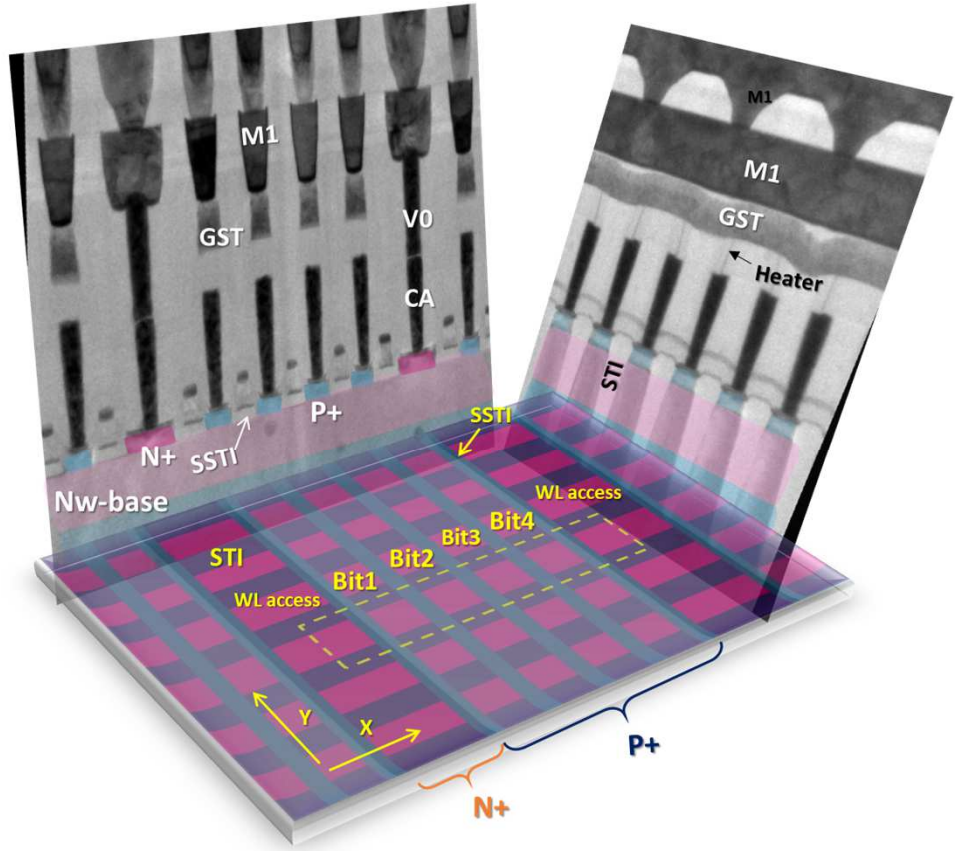


High density eNVM landscape

	PCM	STT-MRAM			RRAM	
Current flow	Uni-directional	Bidirectional			Bidirectional	
Working principle	Phase transition (temperature)	Free Layer MAG orientation (spin)			Oxide vacancies filament	
Selector element	BJT	MOS			MOS	
I_{ON}/I_{OFF} (typ)	**Large (100x)**	Small (2x)			Large (>10)	
Process complexity	Medium	High (MTJ stack)			Low-medium	
Reliability	**High**	High			Medium-Low	
Major limitations	Programming current	Magnetic Field immunity			Erratic bits / variability	
Physical cell size [μm^2]	**0.019 at 28 nm**	0.041-0.045 at 22 nm	0.033 at 16nm	0.024 at 14 nm	0.042 at 28 nm	0.02x at 12 nm
Grade	G0 / qualified	G1 / qual	G1 / qual	G0 / not qual	G1	Indust. / not qual

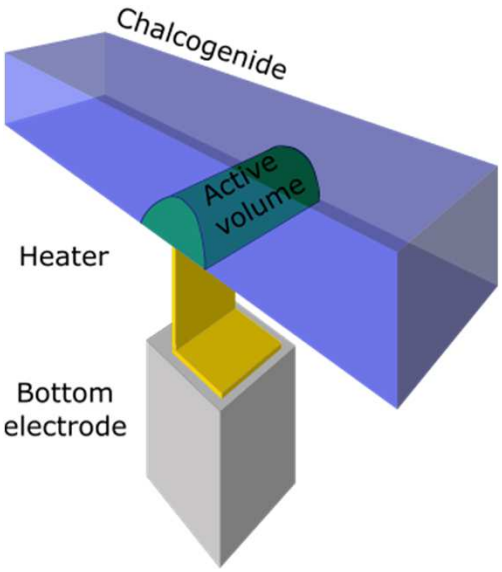
ePCM in the NVM panorama

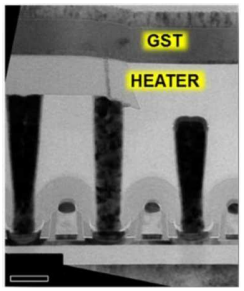
- **Compact**
- **Versatile**
- **Reliable**
- **Scalable**
- **Manufacturable**

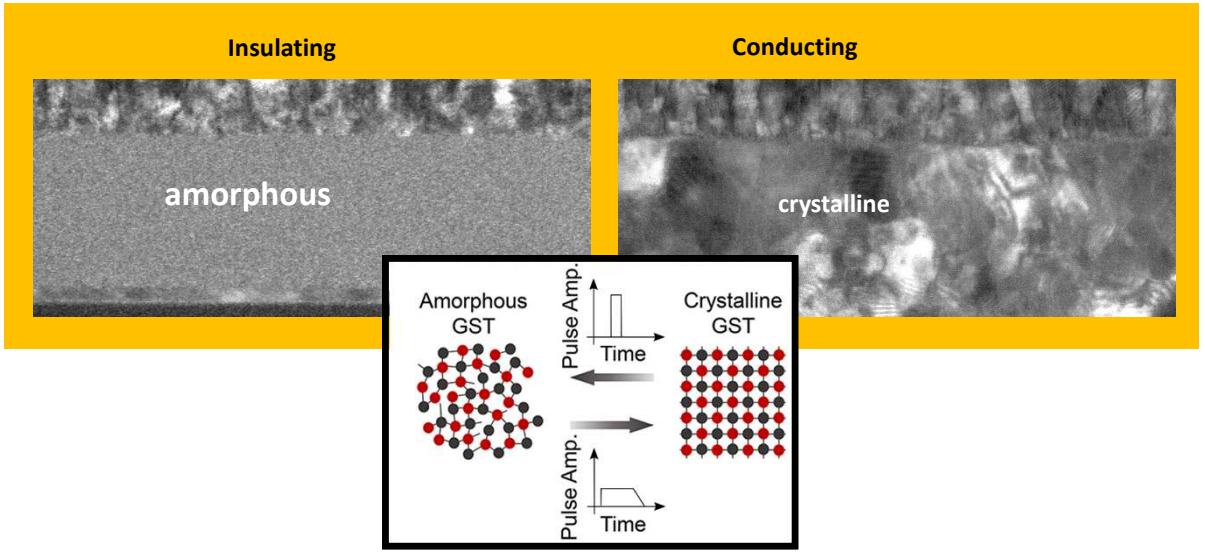


PCM Introduction

PCM device operation

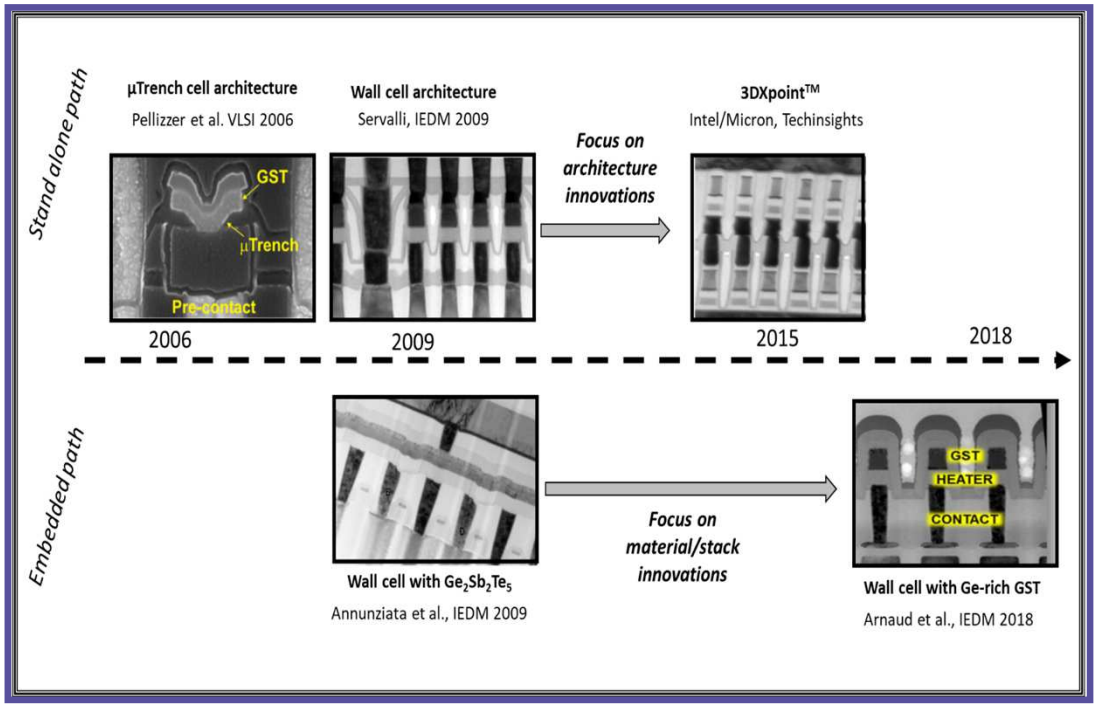






- Reversible switching of chalcogenide material phase ($Ge_2Sb_2Te_5$)
- Amorphous insulating and crystal conducting
- Temperature increase by Joule heating enable phase change
- A simple resistor can be easily integrated in the BEOL

PCM evolution paths



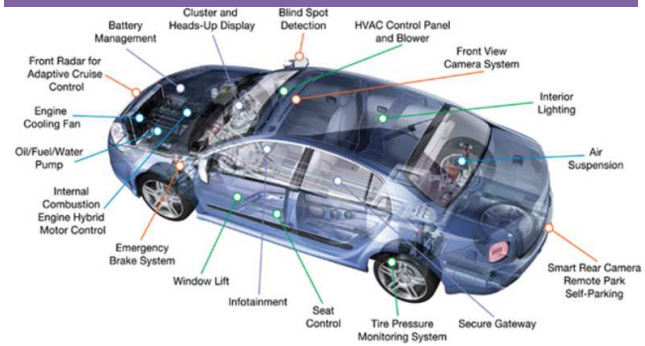
Nokia Asha phone including standalone PCM (Numonyx-Micron)

1.5 TB Optane SSD by Intel

- PCM for standalone focused on high density 3DXpoint architecture
- PCM for embedded memories instead required material innovations

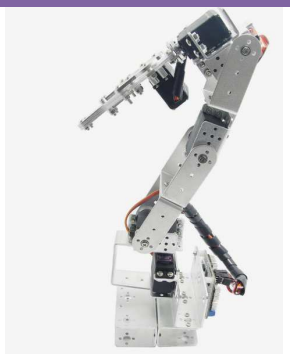
PCM in embedded applications

Automotive



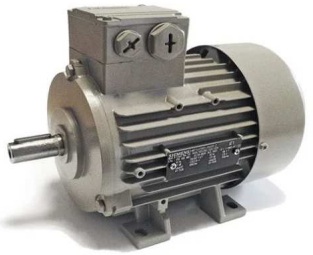
- ePCM integrated with an advanced microcontroller
- Demanding retention (years at 150C)
- Demanding reliability (1ppm chip loss on field)

Consumer/industrial



- General market, many customers → soldering compliance
- ePCM can ben integrated over advanced microcontrollers
- Most advanced technologies can be used

Smart power



- Power hungry applications → heating of the chip
- CMOS realized with older technologies
- Either consumer, industrial or automotive applications

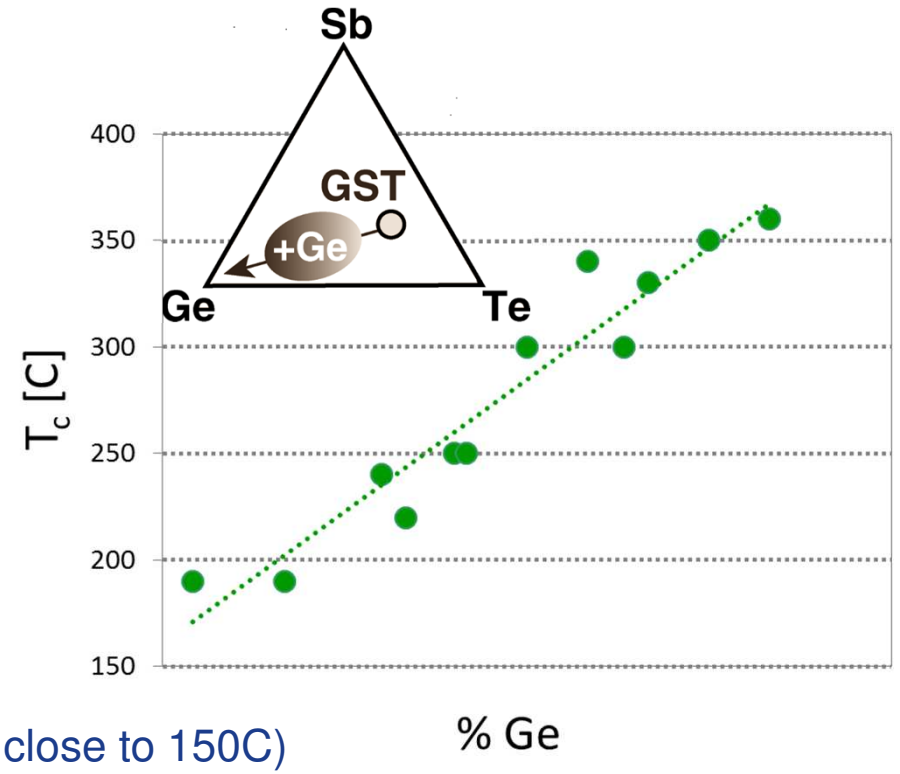
All these applications need very high temperature data retention

Semiconductor Memories and Systems, Chapter 6, P. Cappelletti, Elsevier 2022

Material engineering for ePCM

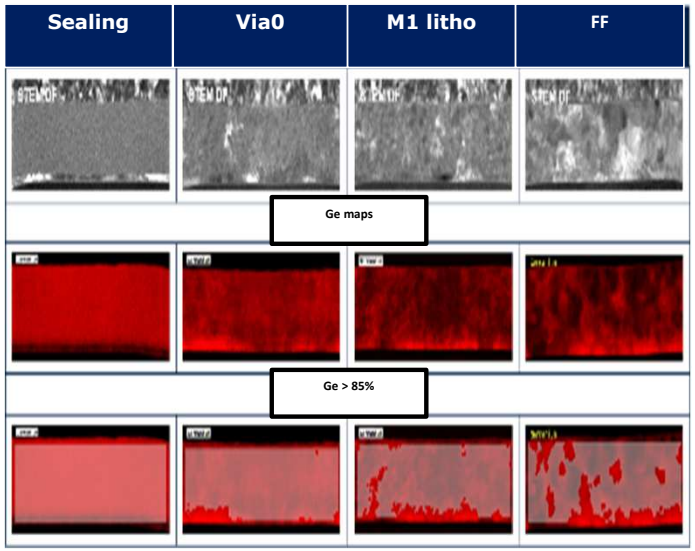
"Semiconductor Memory and systems", P. Cappelletti et al., Chapter 6, Elsevier 2022

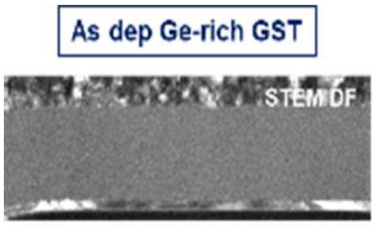
- Improved amorphous state retention
- Automotive specs (some years at 150C)
- Soldering compliance (2min at 260C)
- Automotive-grade defectivity (target of 1ppm product level)

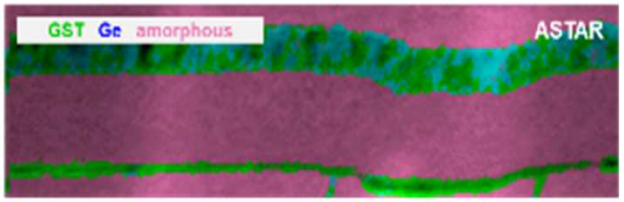


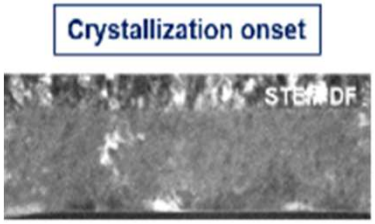
- No way with commonly used $Ge_2Sb_2Te_5$ (T_C close to 150C)
- Higher Ge content in GST alloys needed

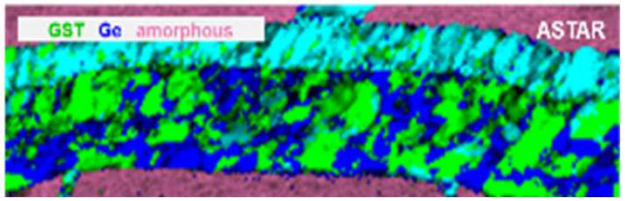
Ge enrichment implications





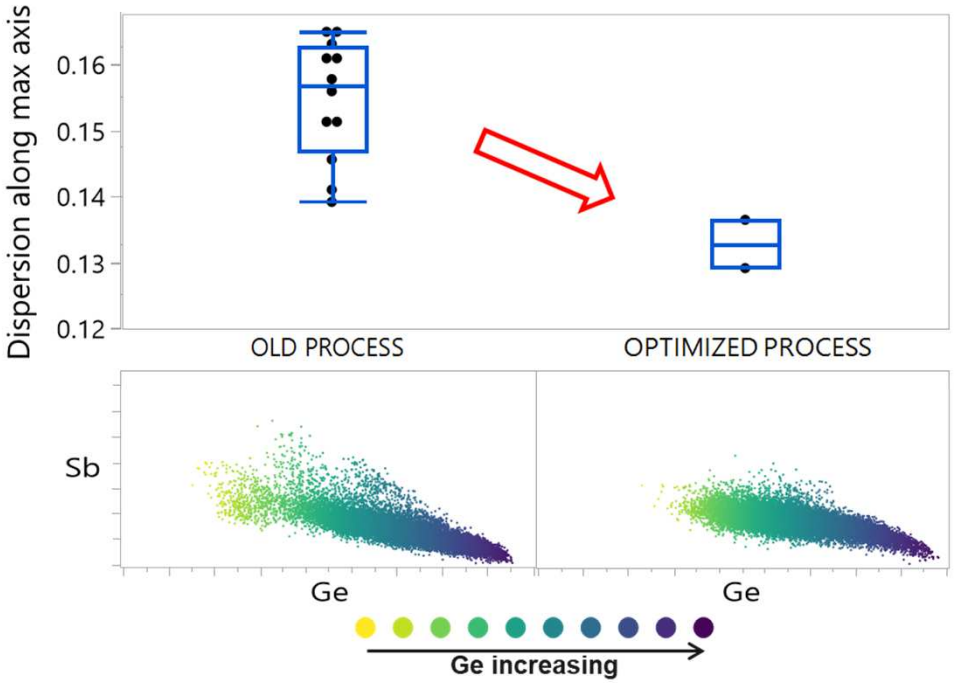


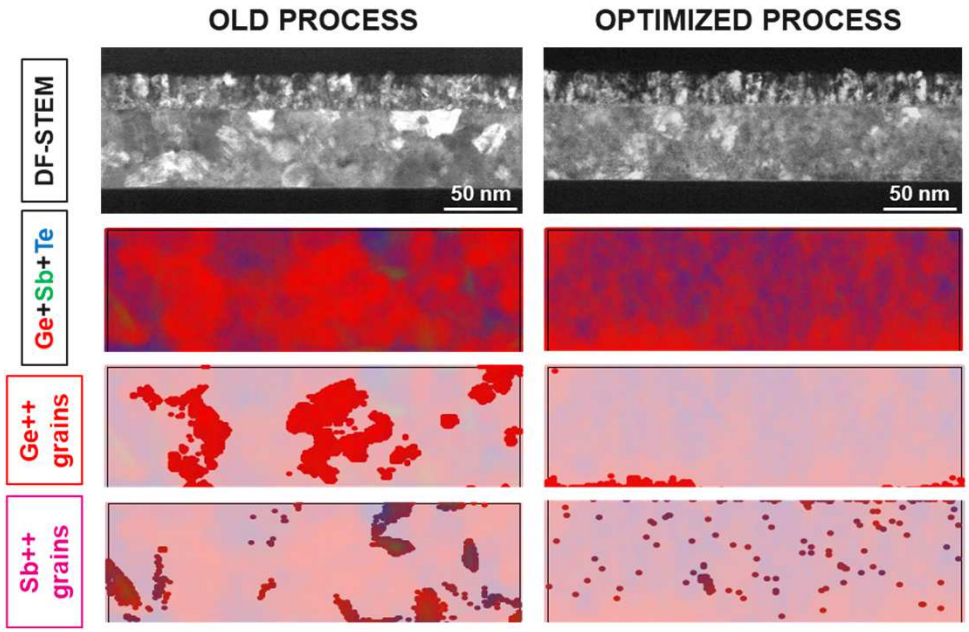




- As-dep Ge-rich GST uniform and amorphous in phase
- During crystallization, separation in $Ge_2Sb_2Te_5$ and Ge occurs
- Ge clustering is maximum at the end of the flow
- Dedicated optimizations at process and algo levels needed

Process optimization





- Dispersion along max axis reduced, indicating less segregated material
- This is confirmed by EELS images, showing finer texture and fewer clusters (both Ge and Sb)

E. Petroni et al. Frontiers of Physics (2022)

Available technologies and applications

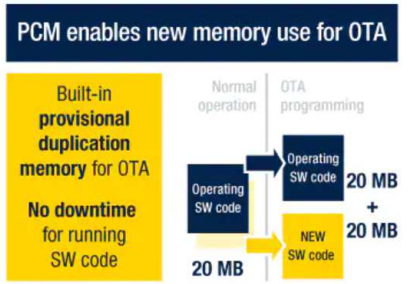
Available technologies

- ePCM in **110 nm/90 nm** BCD platforms
- ePCM in **28 nm FD-SOI** developed for automotive applications
- ePCM in **18 nm FD-SOI** in development for general purpose MCU

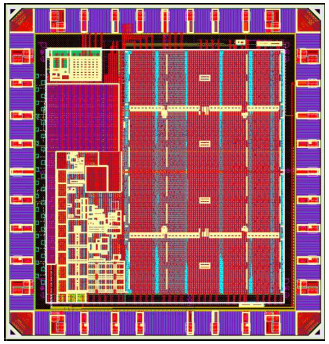
G. Croce,2022 IEEE IMW doi: 10.1109/IMW52921.2022.9779244

N. Grossier et al., 2023 VLSI, doi: 10.23919/VLSITechnologyandCir57934.2023.10185252

A. Conte et al, 2023 IEEE IMW, doi: 10.1109/IMW56887.2023.10145983



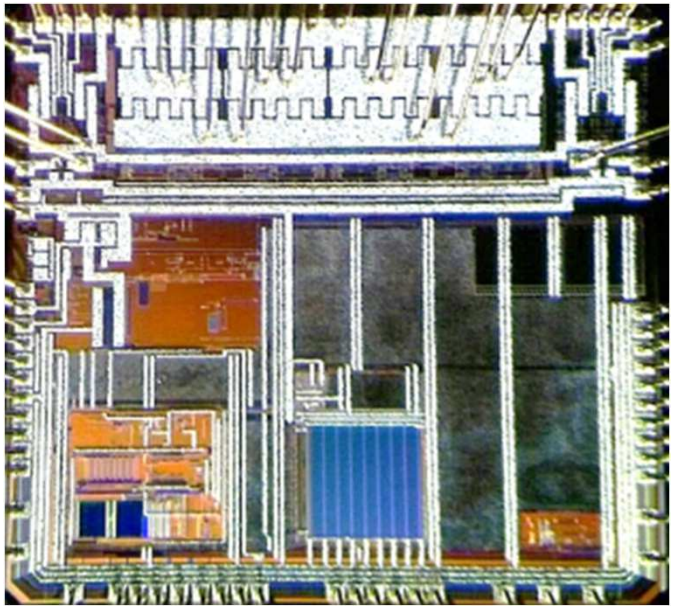
https://www.st.com/en/automotive-microcontrollers/stellar-integration-mcus.html



2MB ePCM NVM IP layout for GMP product
Designed in 18nm FDSOI technology

BCD motor control

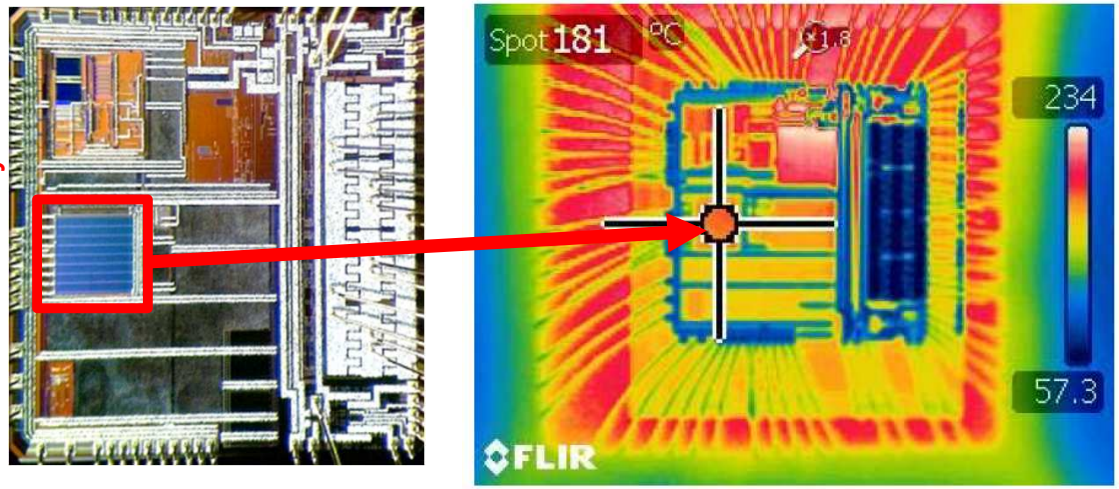
Fully integrated SOC for miniaturized motion control with embedded microcontroller, non-volatile memory (ePCM) and 40V power device



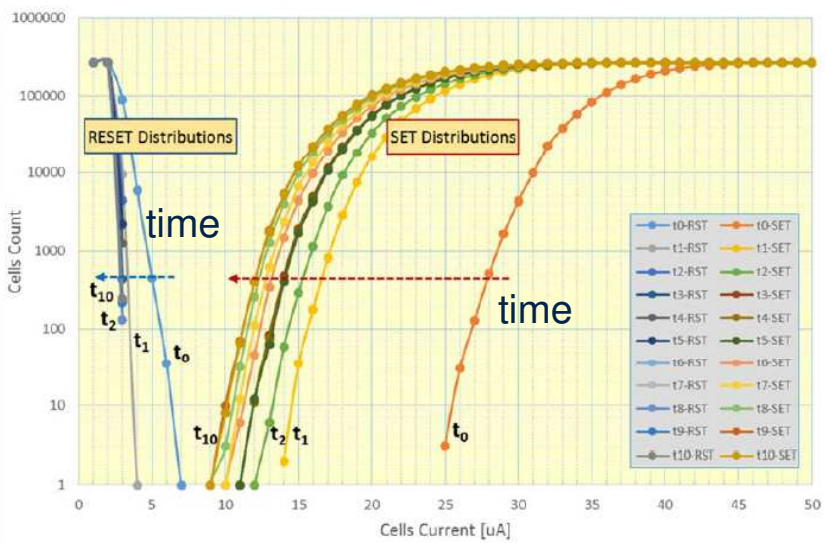
- 110 nm BCD technology platform
- ARM Cortex-M4 with DSP for numeric torque generator
- Memory: 32KB ePCM and 8KB RAM
- Converters: 2x DACs (10 bits and 15 bits), 1x ADC (12 bits)
- 1.8V/ 3.3 V/ 5 V voltage regulators
- Integrated power stages: 4x half-bridges (40 V max 2.5A)
- 8x 40 V gate drivers for external power devices

PCM robustness: a real case

Product thermal map (extreme operation conditions)



Embedded PCM distributions (acquired during product operations)

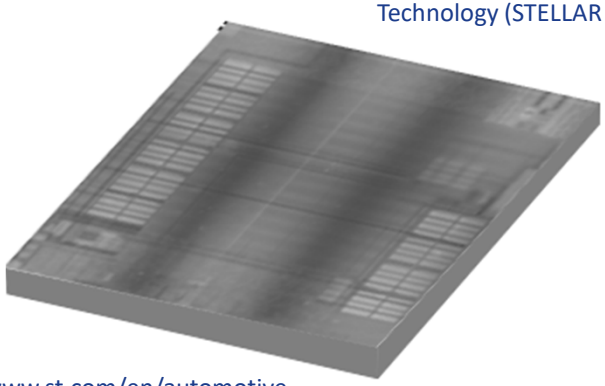


Data Retention secured (> 24 hrs) under very extreme thermal conditions (Tj,max > 200 C, PCM Array T = 180 C!)

28 nm - 20.5 MB for automotive

Figure 1. **Block diagram**

Silicon on 28 nm ePCM Technology (STELLAR G7)



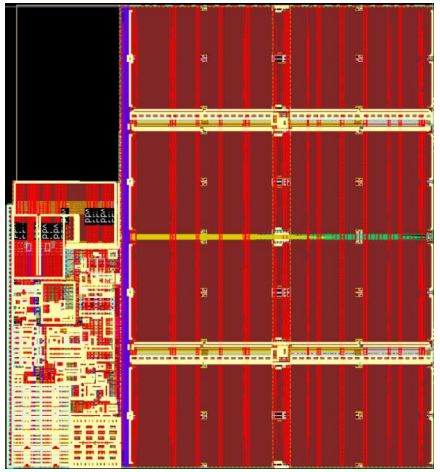
https://www.st.com/en/automotive-microcontrollers/sr6g7c4.html#documentation

- Microcontroller designed in 28 nm FD-SOI platform
- ePCM NVM qualified for automotive applications
- OTA feature available for code update
- Operating temperature up to 150C

BGA476/292 packages

18 nm - 2 MB ePCM for GPM

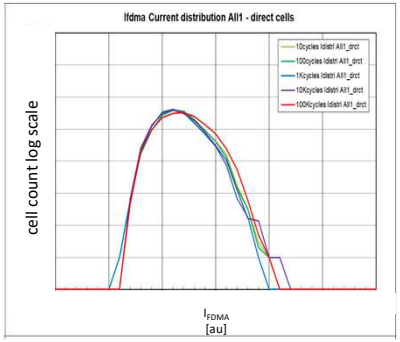
Condition	Spec
External operating voltage V_{CC}	1.62 V - 3.6 V
Cycling temperature (T_J)	-40ºC: +140ºC
JEDEC soldering compliancy	Yes
Code memory reliability	Cycling: 1K-write
	Retention: 10y @ 140C after 1K-write
Data memory reliability	Cycling: 100K-write
	Retention: 2y @140C after 100K-write

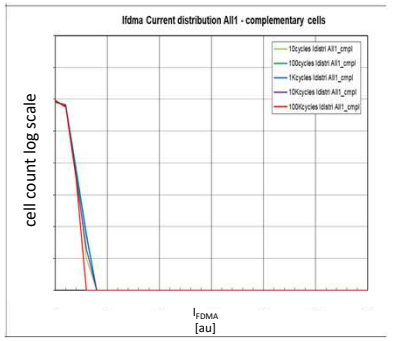


2MB ePCM NVM IP layout for GMP product
Designed in 18 nm FDSOI technology

A. Conte, IMW 2023, Monterey

- Macro designed in 18 nm FD-SOI platform
- ePCM measured and under qualification
- Macro IP available for product roadmap





Opportunities and future developments

AI as a key growing trend

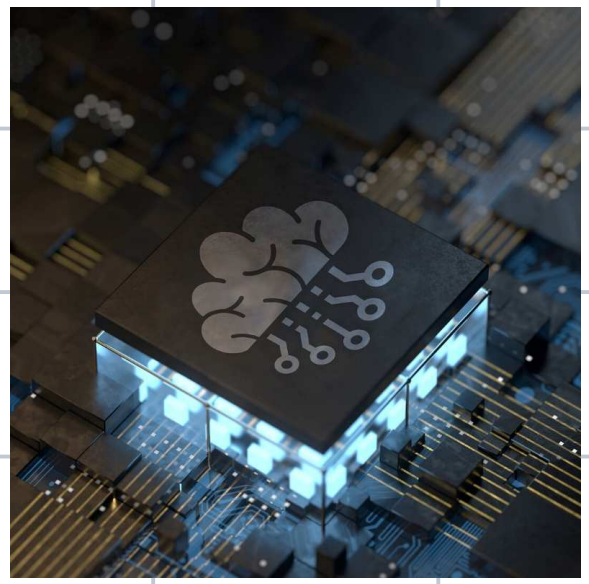
Drones
Flying & landing

Smart industry
Predictive maintenance

Smart home
Event detection

Smart farming
Optimizing conditions

Personal healthcare
Body measurements



White goods
Smart control

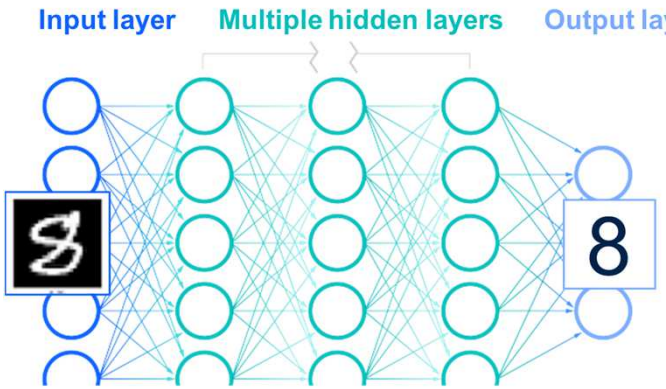
Self-driving cars
Environment sensing

Smart Buildings
Energy-saving

Robots
Collision detection

Personal electronics
Activity recognition

CNN and hardware



https://www.ibm.com/it-it/topics/neural-networks

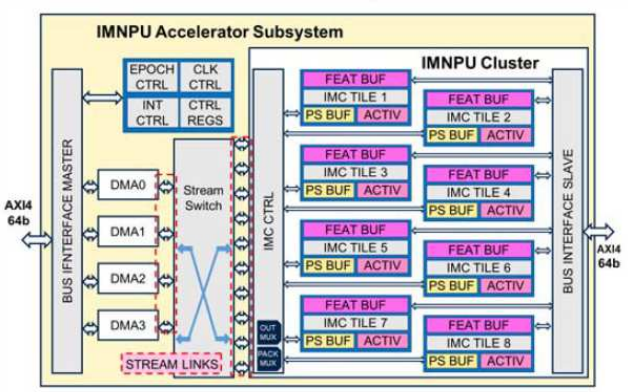
Function we want to realize in hardware

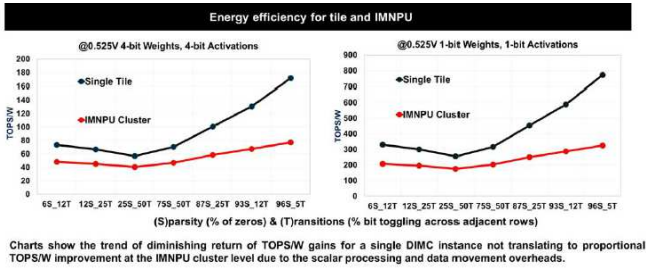
$$\begin{bmatrix} Y_1 \\ Y_2 \\ \\ Y_M \end{bmatrix} = \begin{bmatrix} W_{11} & W_{12} \cdots W_{1M} \\ W_{21} & W_{22} \cdots W_{2M} \\ \vdots & \vdots \quad \vdots \\ W_{N1} & W_{N2} \cdots W_{NM} \end{bmatrix} \times \begin{bmatrix} X_1 \\ X_2 \\ \\ X_N \end{bmatrix}$$

- Neural network as a sequence of neuron layers connected by synapses
- Once the NN learning phase has been done, a set of weights is defined and used for inference for classification
- In both learning and inference phases, the MVM is a key operation that is power and time inefficient when operated in the CPU
- Goal: **embedded MVM multiplication in a dedicated hardware**

PCM for SRAM-based AI accelerators

- D-IMC done in SRAM as a short-time MVM enabler in AI accelerators
- Very good performances obtained at NPU level in 18 nm FD-SOI ST technology
- This architecture can take advantage of ePCM:
  - ePCM can be embedded in a SoC together with the processor (no interface bottleneck in data movement)
  - ePCM today guarantees the best density among the eNVM (0.019um^2), thanks to the unipolar mechanisms of operation (BJT/diode selection)



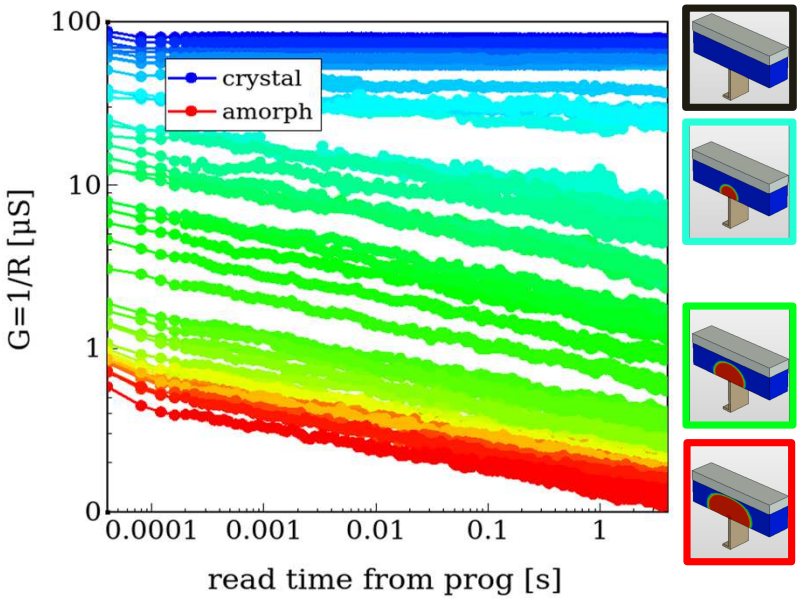


Charts show the trend of diminishing return of TOPS/W gains for a single DIMC instance not translating to proportional TOPS/W improvement at the IMNPU cluster level due to the scalar processing and data movement overheads.

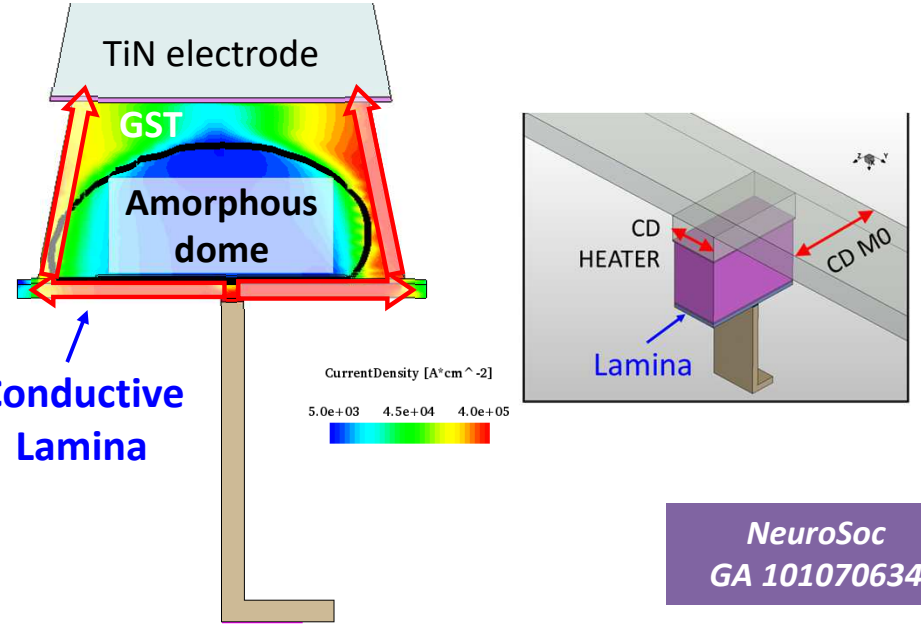
G. Desoli et al., "16.7 A 40-310TOPS/W SRAM-Based all-digital Up to 4b In-Memory Computing Multi-Tiled NN Accelerator in FD-SOI 18nm for Deep-Learning Edge Applications," 2023 ISSCC, doi: 10.1109/ISSCC42615.2023.10067422.

Rheostatic PCM cell for AIMC

ePCM wall cell



PCM-AI rheostatic cell



NeuroSoc
GA 101070634

PCM-AI rheostatic cell cell → conduction mainly in crystalline GST thanks to the lamina

Programmable OTS (SOM)

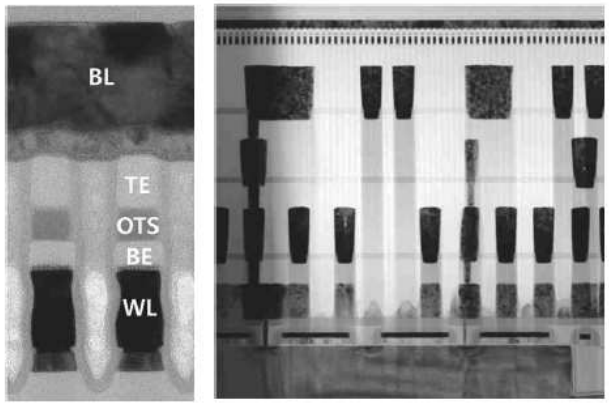


Fig. 1. Cross-sectional TEM Image of (a) 16nm SOM Cell and (b) the entire device of 64Gb SOM. (not including the metal lines of back-end.)

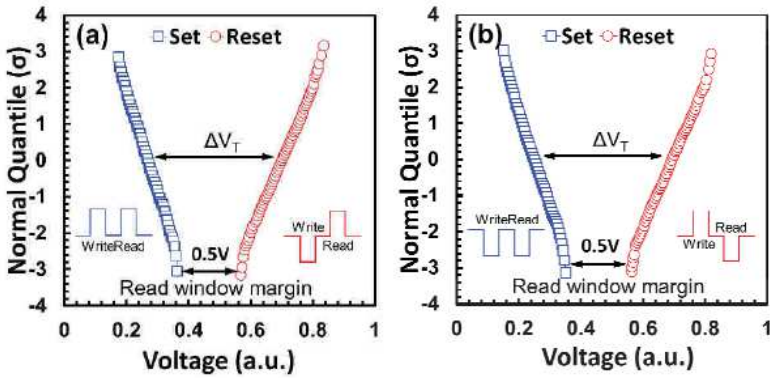
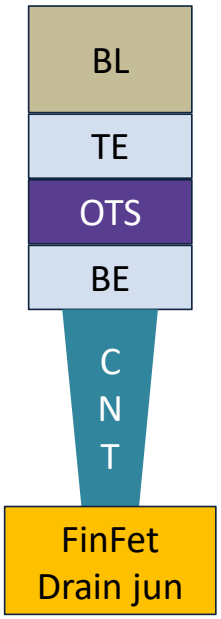


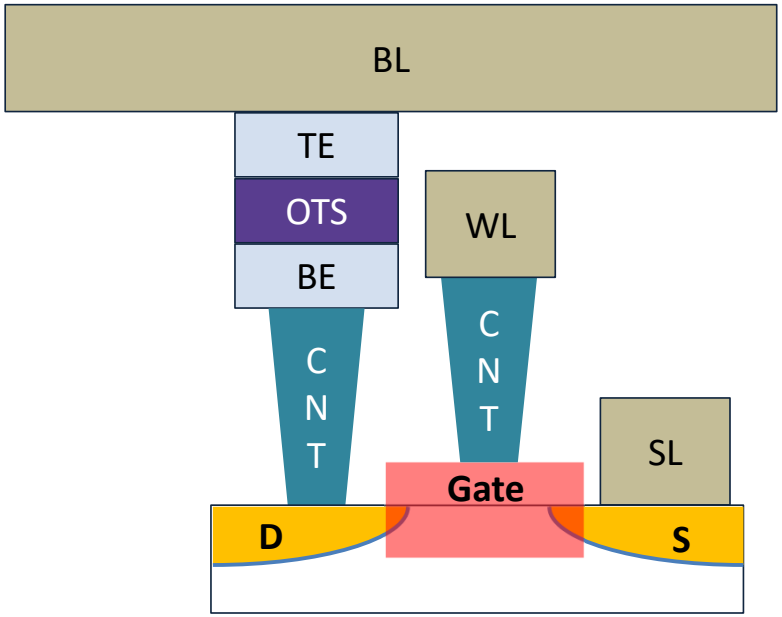
Fig. 2. The distribution of V_T and read window margin depending on the bias pulse for set and reset states: (a) Case of positive direction in read pulse. (BL is cathode.) (b) Case of negative direction in read pulse. (BL is anode.)

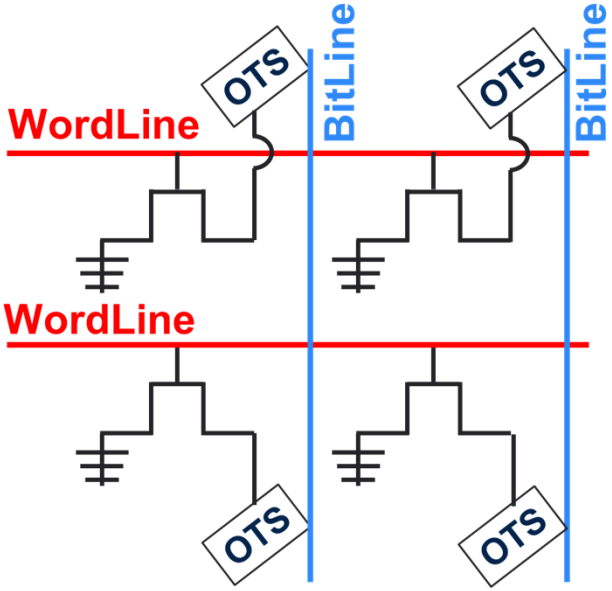
Hynix – IEDM 2018, 2022
Samsung – IEDM 2023
Micron – IEDM 2023
IMEC – IEDM 2021

- OTS working as **memory** (Selector Only Memory)
  - Same polarity between prog and read → low Vth
  - Opposite polarity between prog and read → high Vth

FinFET selected eSOM







Very low process cost (1/2 added masks)
Very high density: low Iprog → small selector area
Coupling with a MOS selector enable low voltage operation

Need to partner with a FinFET provider

Conclusions

ePCM is one of the candidates for eNVM below 28 nm

Material and algo innovations to achieve demanding reliability

ePCM is a reality, ST offers solutions at various technologies

Future developments on AI (IMC and spiking NN) and integration with FinFET

THANK YOU

ICOS International Cooperation On Semiconductors

European Commission

This project has received funding from the European Union's Horizon Europe research and innovation programme under GA Nº 101092562

EU – SOUTH KOREA – Joint Researchers Forum on Semiconductors

www.icos-semiconductors.eu