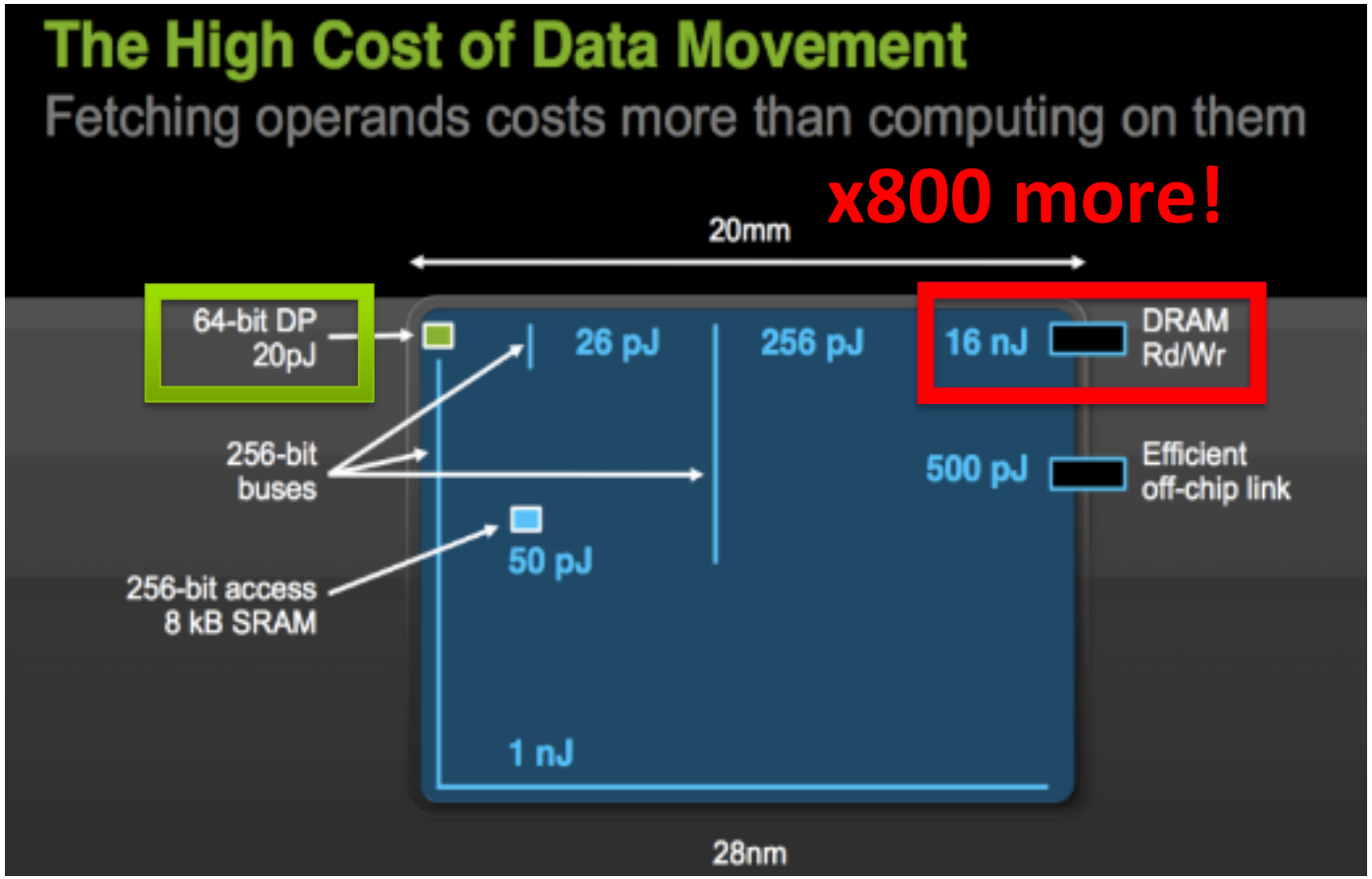
Research on FD-SOI and non-volatile memory

Olivier Faynot

Silicon Component division General Manager

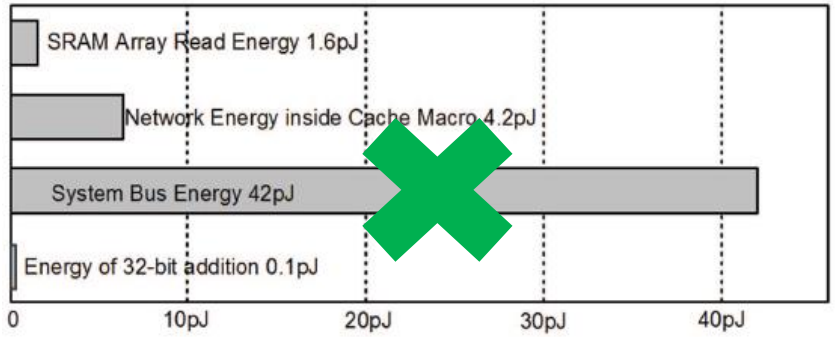
CEA-Leti

The cost of moving data



Bill Dally, "To ExaScale and Beyond", 2010

[J. Wang – ISSCC'19]



~90% of energy is in **data transfer**
→ IMC could lead **8x reduction**

Operation energy is negligible

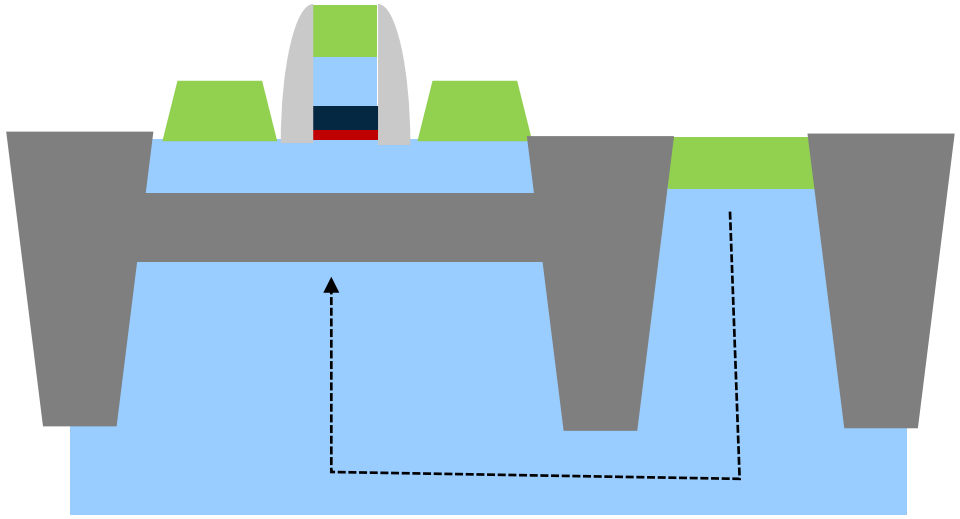
Memory access and control energies dominate

The required gain in energy efficiency

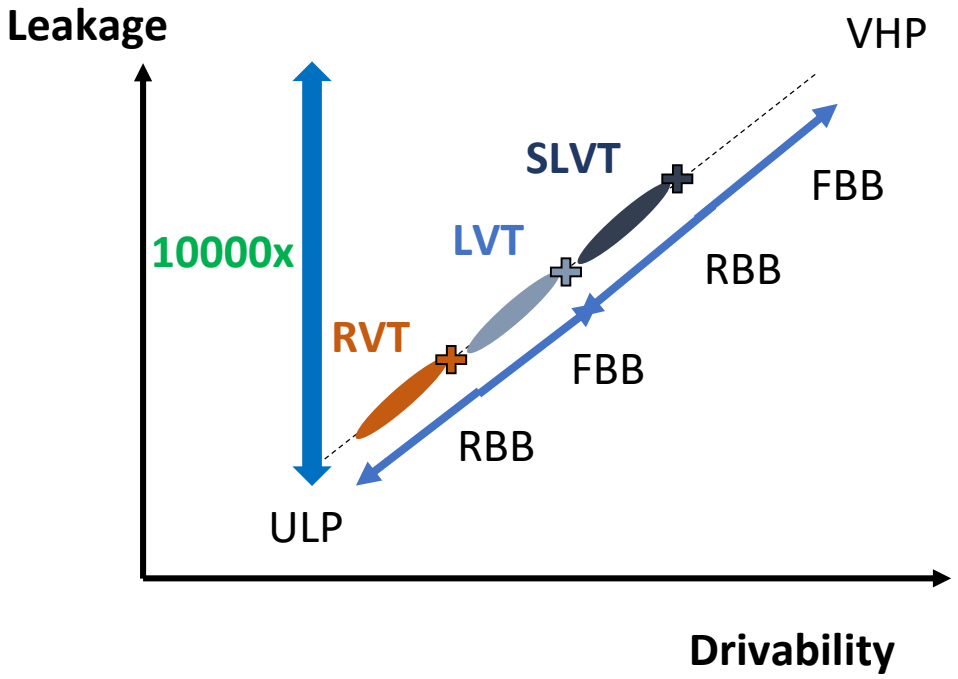
>1000x
by
2030

- CMOS scaling
- Memory technologies
- Disruptive Computing
- Chiplet & 3D System

FD-SOI Technology



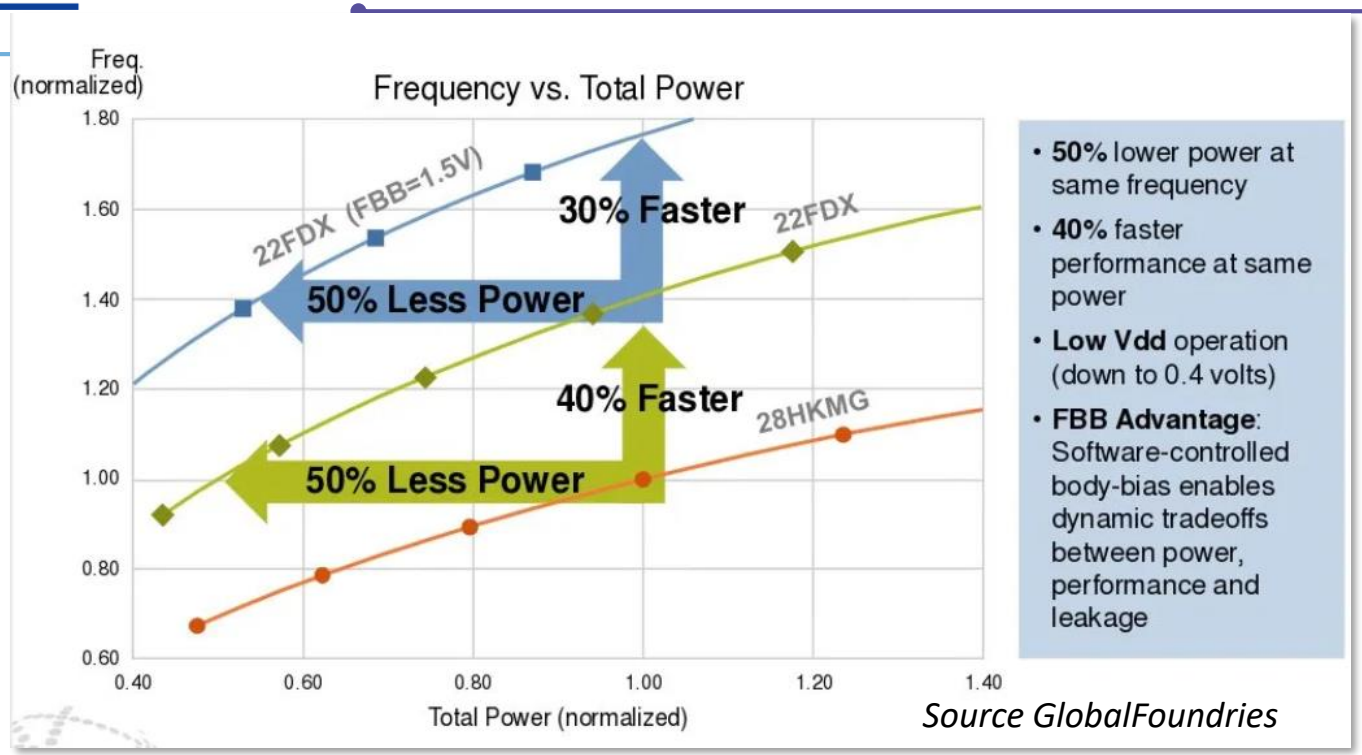
	22FDX	14nm FInFET	28nm Bulk	45nm PDSOI
f_T n-FET [GHz]	347	314	310	296
f_{max} n-FET [GHz]	371	180	161	342
f_T p-FET [GHz]	242 275 (mmWave)	285	185	-
f_{max} p-FET [GHz]	288 299 (mmWave)	140	104	-



RBB: Reverse Body Bias
FBB: Forward Body Bias
ULP: Ultra-Low Power
VHP: Very High Performance

Energy Efficiency gain with Adaptative Back Bias

Dolphin & CEA-Leti, ISSCC 2021



Source GlobalFoundries

Technology	FDX 22nm
Die size	2.5x2.5 mm²
# measured die	316
Corners lots	TT,FF,SS,FS,SF
ABB power	14 µW@25C,TT
ABB area	21 000 µm²
Biased well area	2mm²

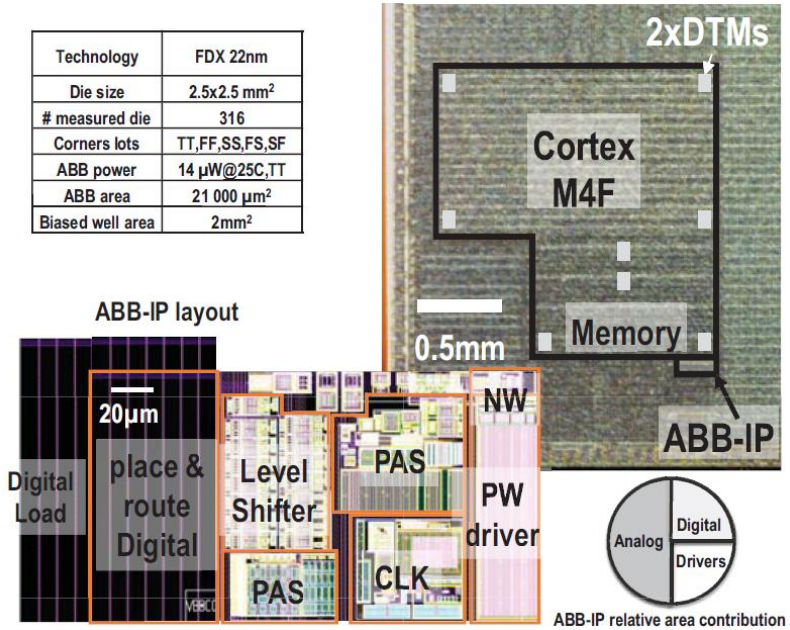


ABB IP: 0.0021mm²
Biased Die area: 2mm²

- Back bias effect gain similar as switching to the next node (55% Power gain)
- Required ABB area almost negligeable

Future Needs

- Available/planned technologies:

28nm	22nm	18nm	12nm	…
ST life.augmented	GLOBAL FOUNDRIES		GLOBAL FOUNDRIES	
SAMSUNG SEMICONDUCTORS		SAMSUNG SEMICONDUCTORS		

- Customer needs:
  - Denser node for competitivness
  - High Voltage devices, eNVM
  - Improvement of RF performance (F_T, F_{max} >450GHz)
  - Enhanced Back Bias effect

FD-SOI : Selected by Worldwide Key Players

COMMUNICATIONS

MAIN DRIVERS

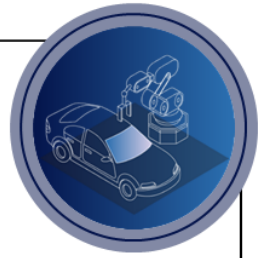
- **FD-SOI is ideal for 5G mmWave**
- 5G sub-6 GHz
- Mobile infrastructure
- WiFi 6

Google SAMSUNG — Nov 2021: Google Pixel6 5G mmW with Samsung FD-SOI

Qualcomm — Aug 2022: Next gen 5G mmW RFIC by Qualcomm will use 22FDX technology

MEDIATEK — Jul 2022: Mediatek 5G mmWave platform uses 22FDX

AUTOMOTIVE



MAIN DRIVERS

- Autonomous cars
- Infotainment

BOSCH — March 2021: Bosch mmW radar ECSEL project OCEAN12

NXP — March 2021: NXP for ultra low power in automotive

STM — STM unveil its new MCU Stellar MCU for auto

SMART DEVICES

MAIN DRIVERS

- Edge computing
- 3D sensing & Healthcare
- Smart home & Smart cities
- Data centers

LATTICE SEMICONDUCTOR — Jan 22 : Low power Lattice FPGAs and computer vision software

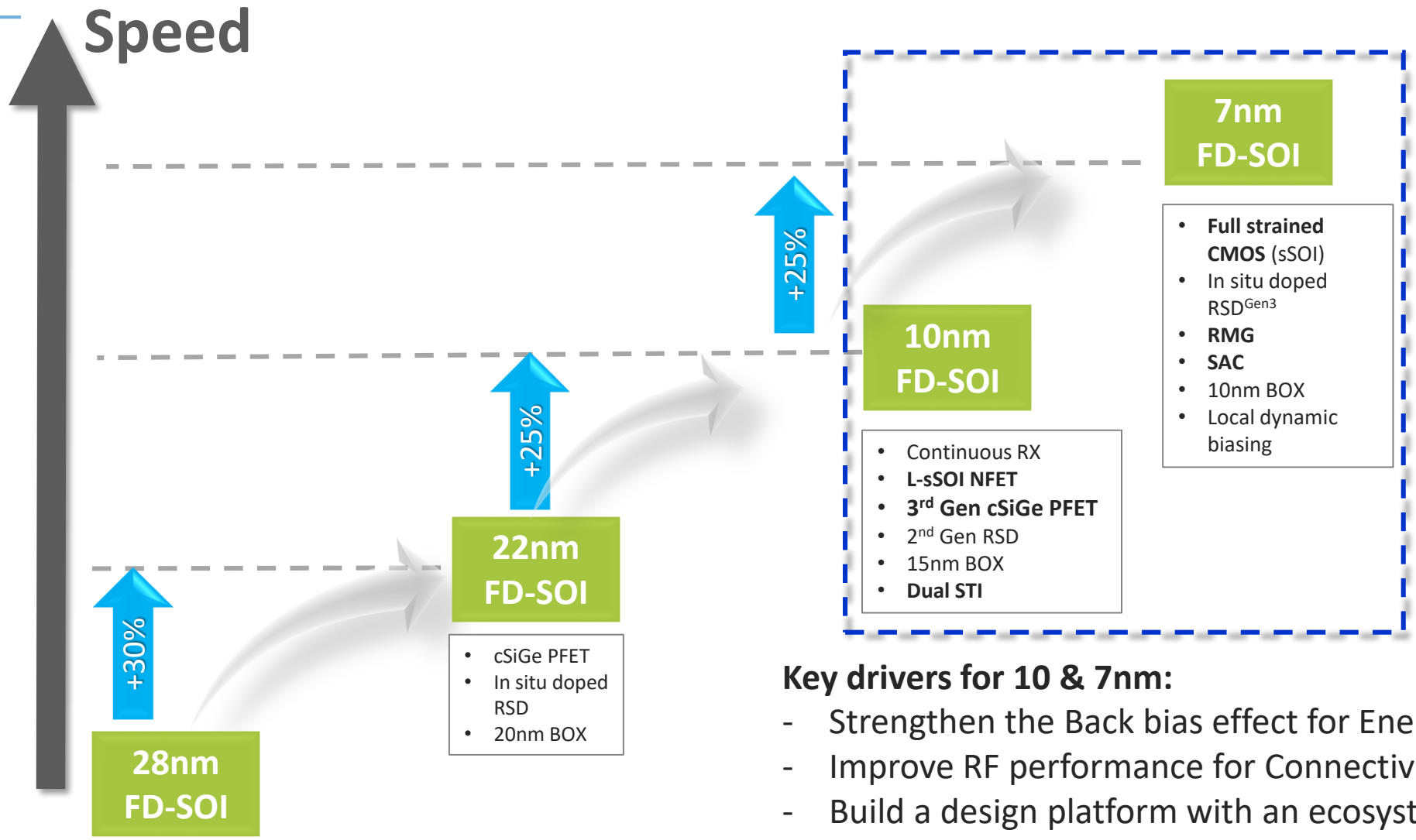
SONY GNSS

https://www.researchgate.net/publication/300409342_265_A_07V_15-to-23mW_GNSS_receiver_with_25-to-38dB_NF_in_28nm_FD-SOI

SONY GPS

https://www.semi.org/en/big-win-for-fd-soi-sony-gps-in-huamixiaomi-smartwatch-eetimes

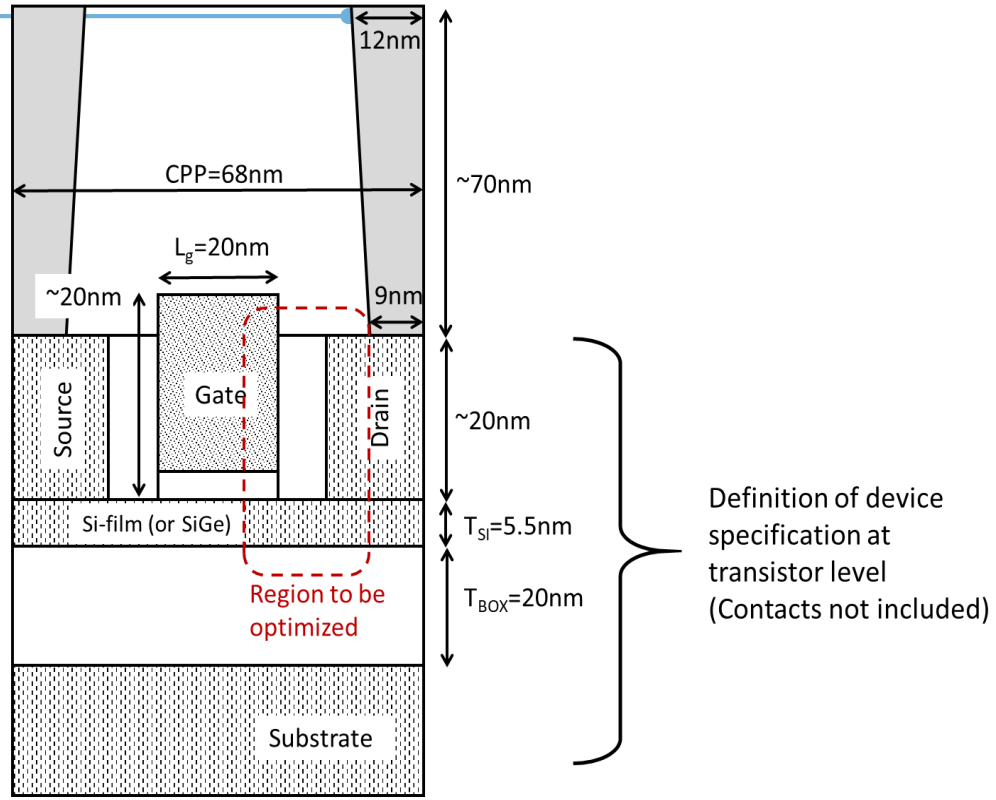
FD-SOI Technology Roadmap



Key drivers for 10 & 7nm:

- Strengthen the Back bias effect for Energy saving
- Improve RF performance for Connectivity
- Build a design platform with an ecosystem

FD-SOI NextGen: Key features



	10FD
Vnom	0.75V
Pitches (CPP/Mx)	68/48nm
Body biasing	Bi-directional
Embedded NVM	PCRAM
High Voltage dev.	Up to 3.3V
RF/Connectivity	F_T/F_{max}>450GHz

Intrinsic gain vs 28FD	
Speed at same power	**1,9x**
Power at same speed	**÷5**
Transistor Density	**x4**

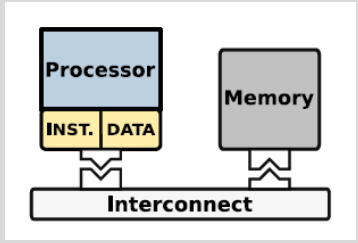
Will be the most advanced node with Gate First approach!

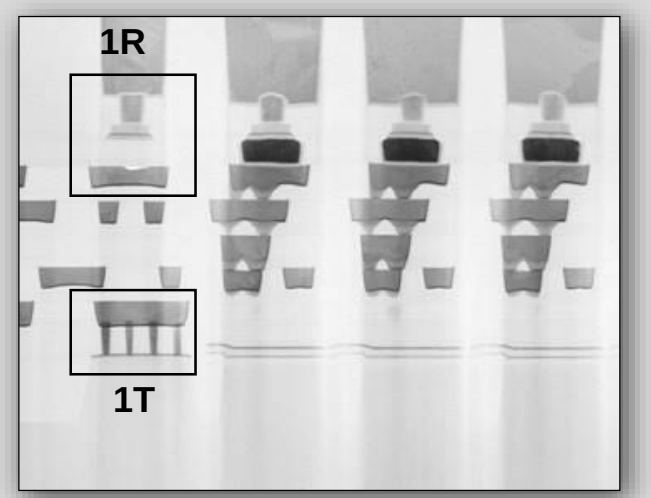
Why Emerging Resistive Memories?

High dense on-chip memory

DRAM access is at least **1500x** more costly than a MAC operation in NN accelerators

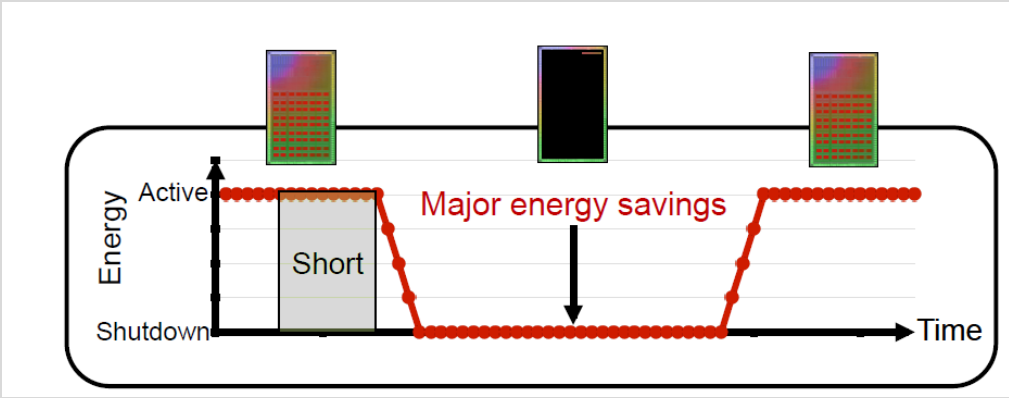
[F. Tu, et al., 2018 ACM/IEEE]





L. Grenouillet et al., 2021

Zero stand-by power thanks to non-volatility



10x better energy efficiency than embedded flash thanks to resistive memories

T. Wu et al., 2019

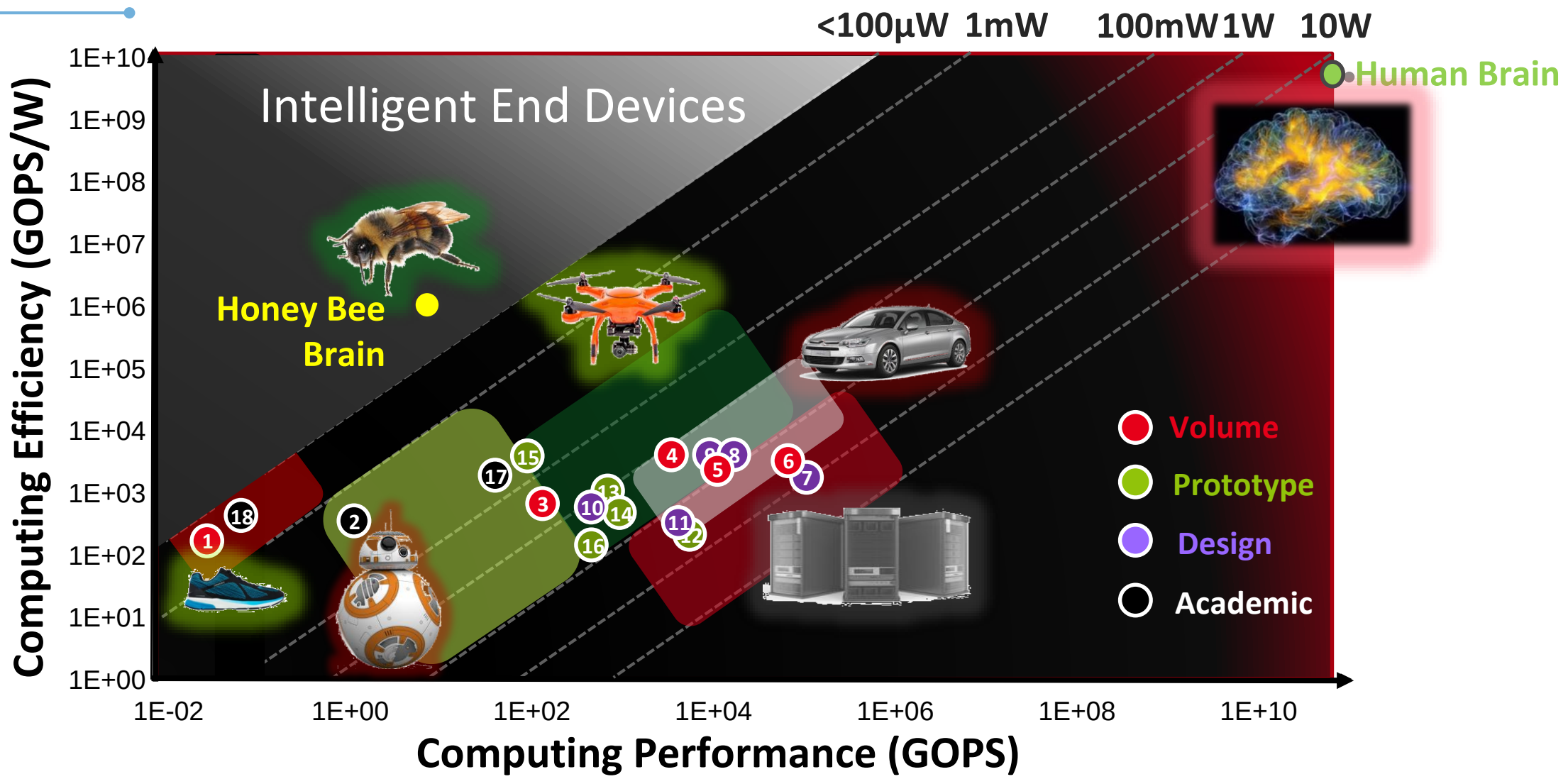
Emerging Non-Volatile Memories

	NOR FLASH	MRAM	PCRAM	OxRAM	FeRAM (PZT)	FeRAM (HfO_2)
Programming power	~200pJ/bit	~20pJ/bit	~300pJ/bit	~100pJ/bit	**~10fJ/bit**	**~10fJ/bit**
Write speed	20 µs	**20 ns**	**10-100 ns**	**10-100 ns**	<100ns	**14ns @ 2.5V (SONY)** **4ns @ 4.8V (LETI)**
Endurance	10^5 - 10^6	**10^6-10^{15}**	10^8	10^5 – 10^6 on 16 kbit	**> 10^{15}**	**> 10^{11} single device** **10^6 – 10^7 on 16 kbit**
Retention	> 125°C	85°C - 165 °C	**165°C**	> 150°C	125°C	125°C
Extra masks	Very high (>10)	Limited (3-5)	Limited (3-5)	**Low (2)**	**Low (2)**	**Low (2)**
Process flow	Complex	Medium	Medium	**Simple**	**Simple**	**Simple**
Multi-Level Cell	Yes	No	**Yes**	**Yes**	No	No
Scalability	Bad	Medium	**High**	**High**	Medium	**Poor (2D)** **High (3D)**

Power Reduction by 10000!

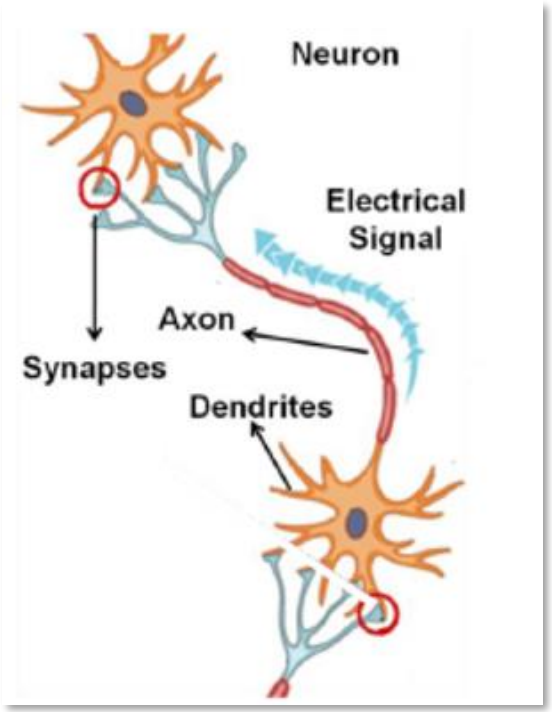
Memory activity focus on embedded NVM for NOR flash replacement

Energy Efficiency is far from biological Systems

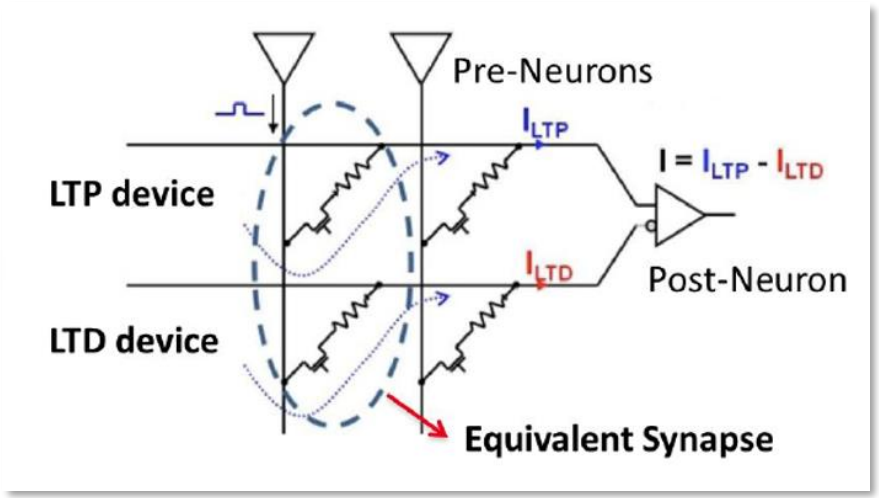


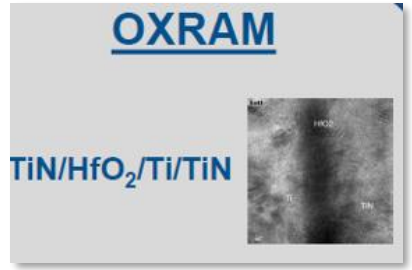
Neuromorphic based RRAM circuit

M. Suri et al, IEDM 2011.

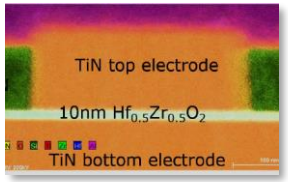


2 PCRAM Example:



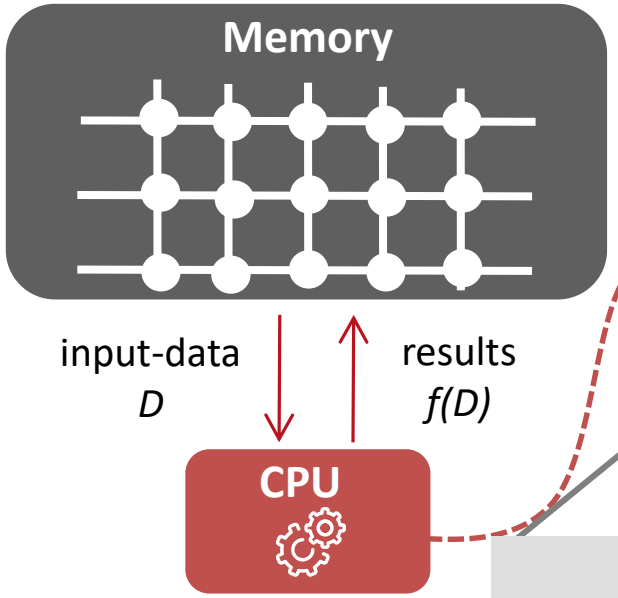


FeRAM

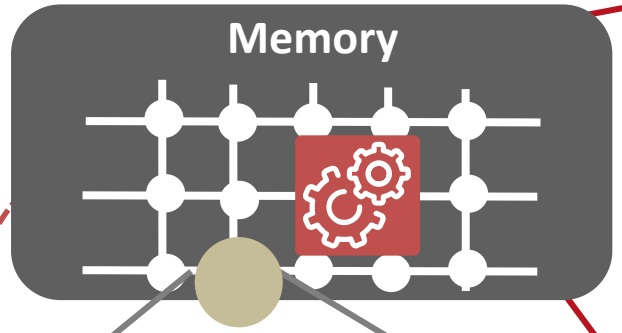


In memory computing

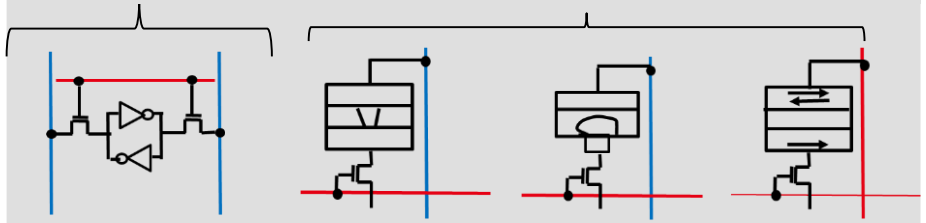
Von Neumann computing



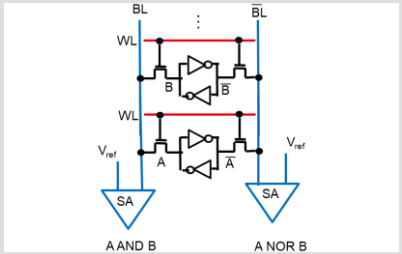
In-memory computing



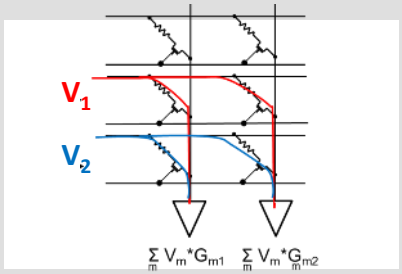
Memory technologies



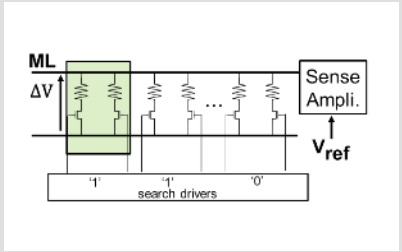
Logic and arithmetic operations



Digital operations using SRAM



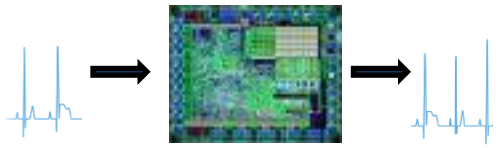
Analog: multiply and accumulate



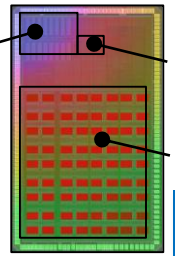
Searching / matching (CAM)

Near- & in-memory computing

Bayesian Neural Networks

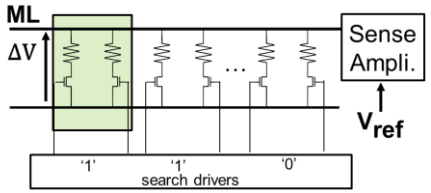


Spiking Neural Network



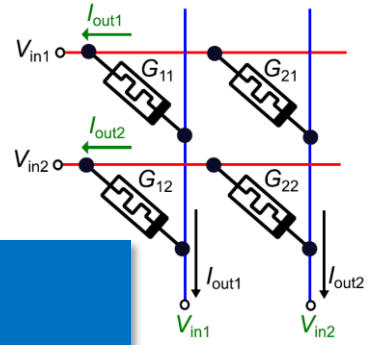
Energy-efficient NN accelerator

RRAM for near- and in-memory computing

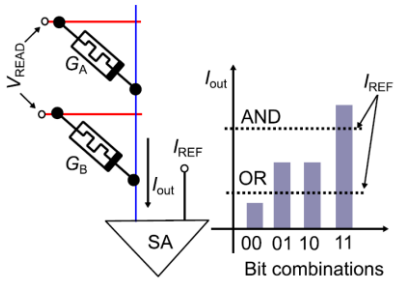


Search/match operations (CAM)

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$



Deep learning, matrix-vector multiply



In-memory logic

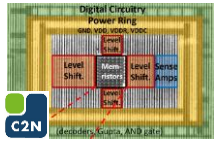
A possible Edge IA Roadmap

Courtesy of E. Vianello, Leti

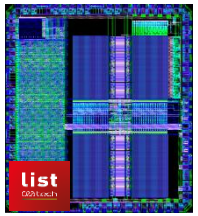
Functions

Performance

Smart sensor



Bayesian machine



Esperanto
Spiking recurrent

Classification

BNN
Fully distributed 2T2R

SPIRIT
Analog spiking

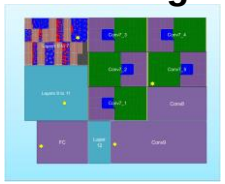
X10000 neurons
Power ÷10

<1pJ / synaptic event

Spiking CNN

Largo

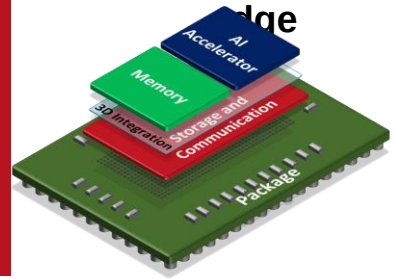
Feature extractor of HD images



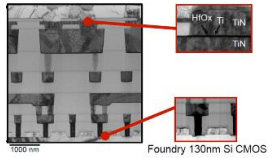
NeuroCorgi ASIC

Heterogeneous architecture

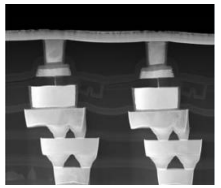
Next generation AI at edge



Multi-die (chiplet) integration on base die

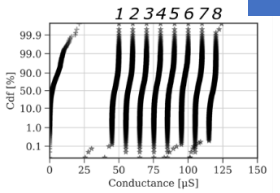


1T1R 130nm

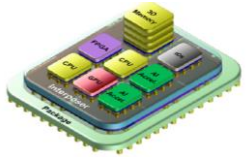


1T1R 28nm

1T1R 22nm

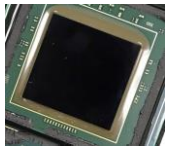


Multi-bit



Multi-die integration

Chiplets: the new IC design paradigm

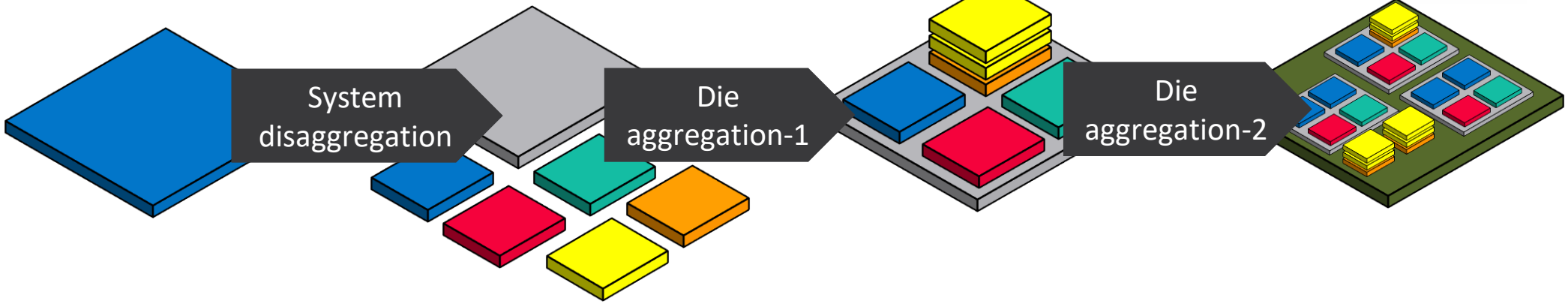


Monolithic die

Chiplets

3D Integrated Circuit

Package



Architecture

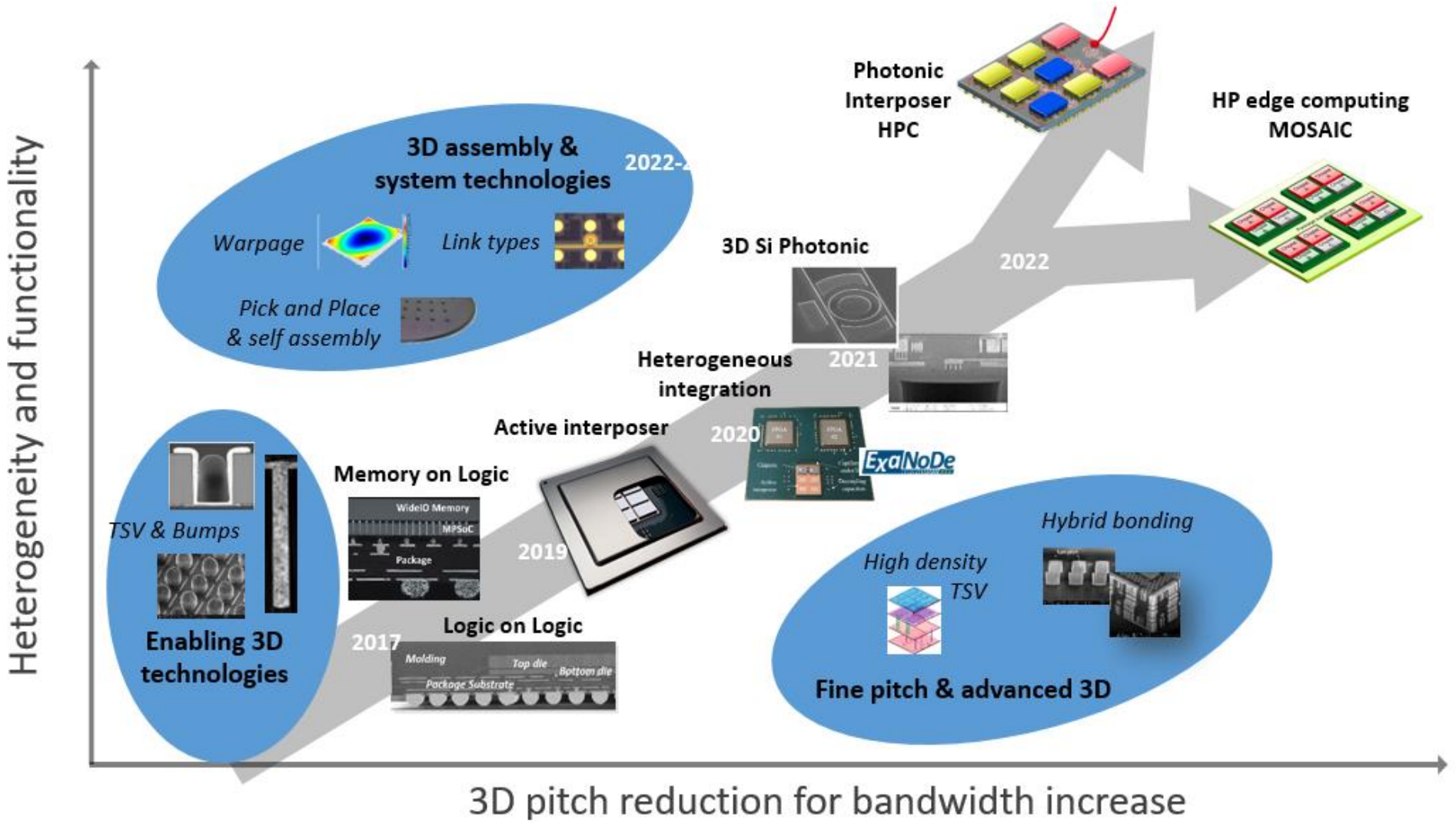
Design
Design flow
Wafer-level integration

Design
Design flow
Package-level integration

System Technology Co-optimization

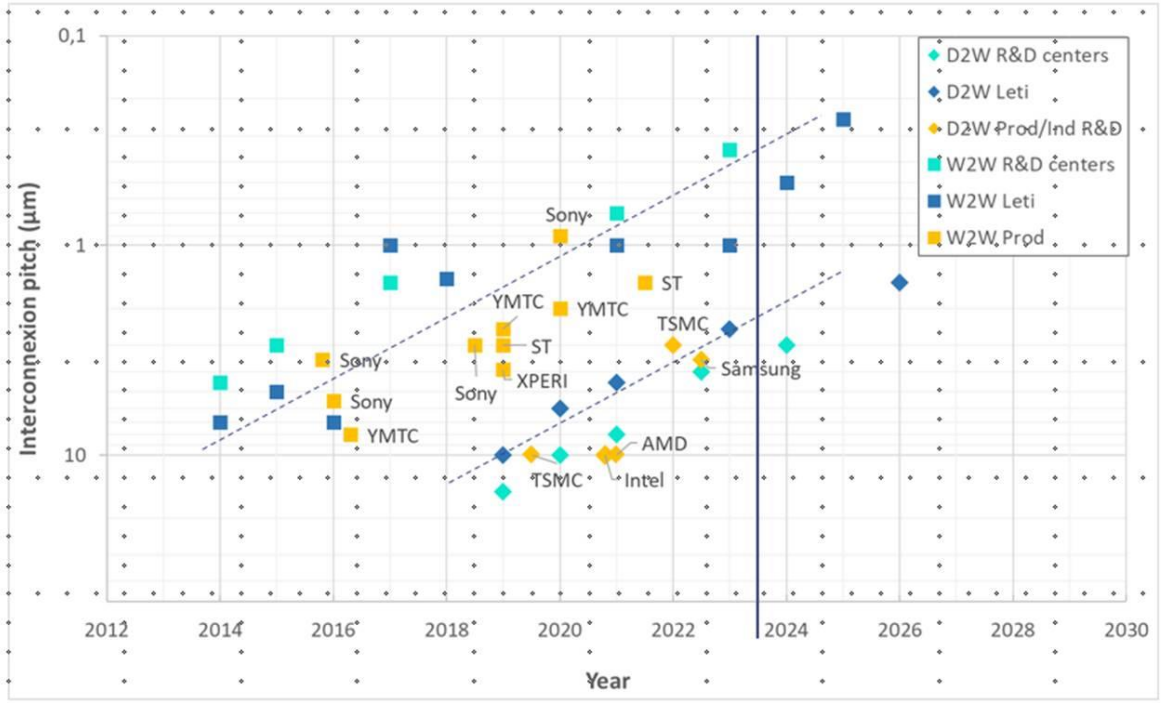
Up to 100x gain on Power Efficiency with 3D

3D Tool Box for Chiplet integration

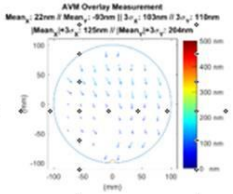


Benchmarking on 3D

Hybrid bonding Road-Map @ Leti

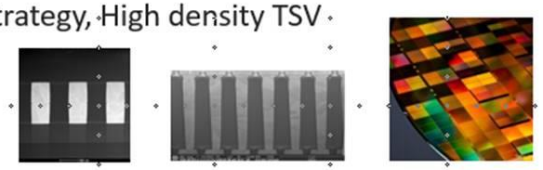


Placement accuracy from tool suppliers ≠ Alignment accuracy ≠ Pitch validated with full functional wafers connexion

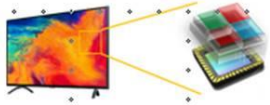


Our main developments drivers :

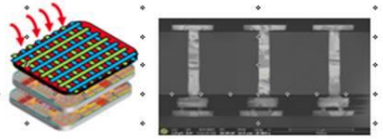
- **Performance** : Innovative bonding process, Interconnection density increase, KGD strategy, High density TSV



- **Heterogeneous 3D** : Adaptation to non-CMOS technologies, exotic material integration, Low temperature processes (200°C and below)



- **Architecture** : Multi-stacking layers, Chiplets on bottom die, Co-integration with Si interposer



Key Messages

- FD-SOI technology well adapted for Power efficient Analog/RF applications
- 10 and 7nm nodes under definition, with 4 major constraints:
  - Improvement of RF performance
  - Enhanced Back Bias effect
  - High Voltage devices
- eNVM essential for SoC and AI applications

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