N3XT Heterogeneous Integration: From Lab to Fab

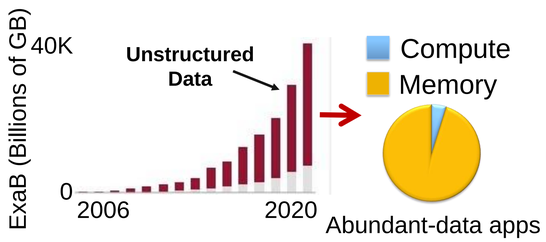
Energy Efficient Learning Machines (ENIGMA) (E2CDA:1640078)

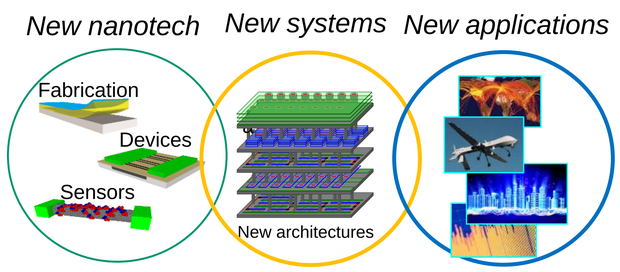
Robert M. Radway[1], Subhasish Mitra[1], H.-S. Philip Wong[1]

T. F. Wu[1], B. Q. Le[1], A. Bartolo[1], W. Hwang[1], S. Jeong[1], H. Li[1], P. Tandon[1], E. Vianello[3], P. Vivet[3], E. Nowak[3], M.K. Wootters[1], M. M. Sabry Aly[2], E. Beigne[3]

CEA-LETI[3], Nanyang Technological University[2], Stanford University[1]

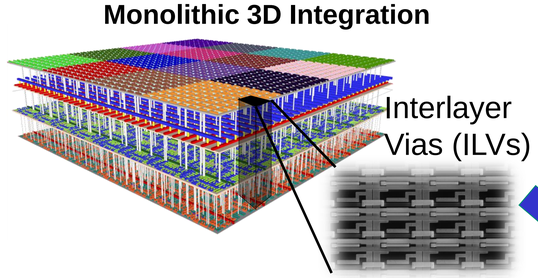
Abundant Data Applications…Need New Nano*Systems*



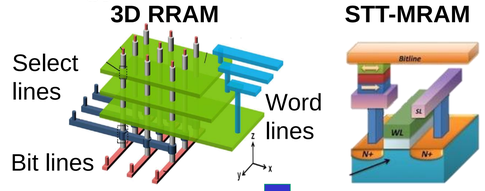


Monolithic 3D Integration of Logic and Memory

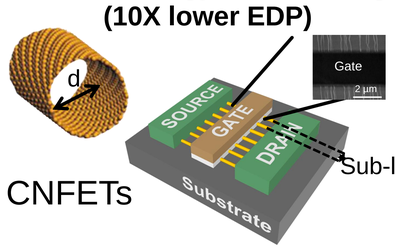
Fine-grained Ultra-dense Monolithic 3D Integration



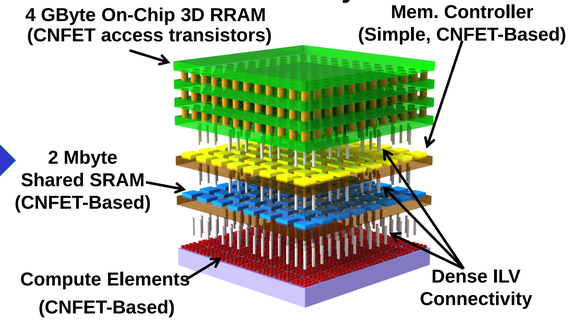
Energy-Efficient Non-Volatile Memory



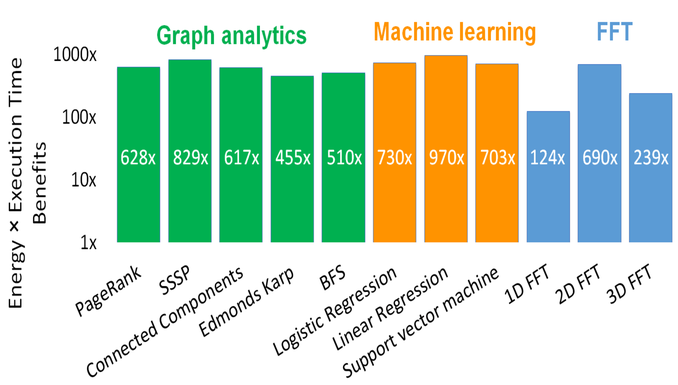
CNFETS Energy-Efficient Logic (10X lower EDP)



N3XT 3D NanoSystem



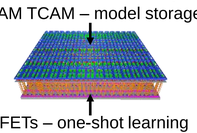
Major Benefits – Existing Algorithms



New Paradigms

Hyperdimensional Computing

RRAM TCAM – model storage

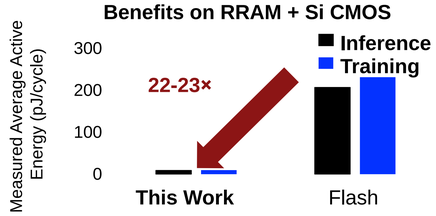


CNFETs – one-shot learning

Monolithic CNFETs + RRAM

Live ISSCC Demo

Benefits on RRAM + Si CMOS



21 Language Classification, 95% Accuracy

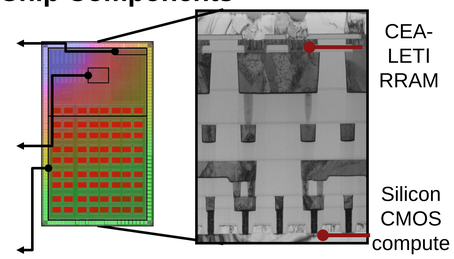
RRAM + Si CMOS (Collaboration with CEA-LETI)

Chip Components

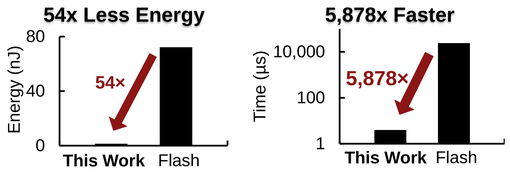
Scheduler
5,000× quicker shutdown vs. Flash chip

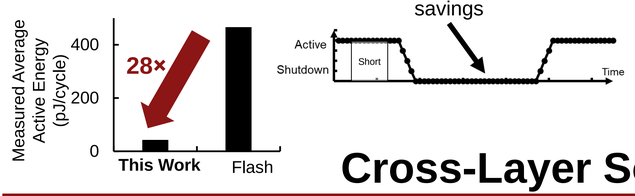
Processor core
Apples-to-apples comparison vs. Flash chip

Multiple bits/cell RRAM on-chip
10× lower energy vs. Flash chip

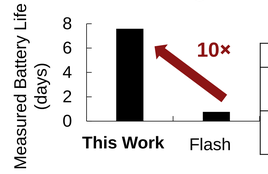


Fine-Grained Temporal Power Gating





10x Longer Battery Life

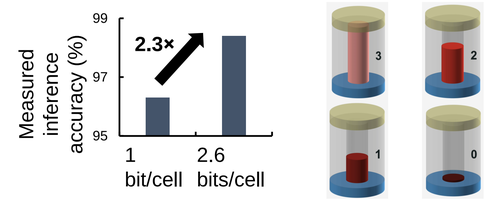


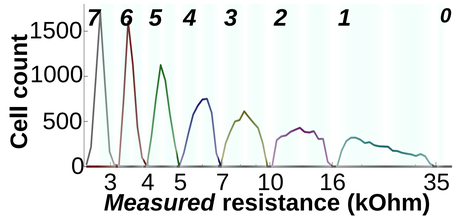
	Our Chip	Flash chip
Active	3.5%	99.9%
Active-to-Shutdown	4.9%	*Infeasible*
Shutdown (Retention)	0.02%	0.1%

Application: Kalman Filter (control)

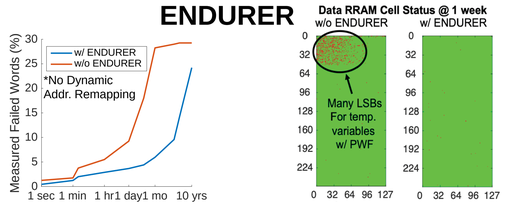
Cross-Layer Solutions

Multiple Bits Per Cell

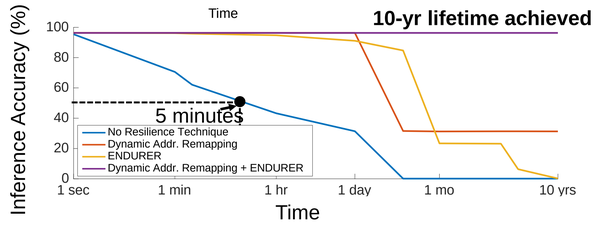




ENDURER

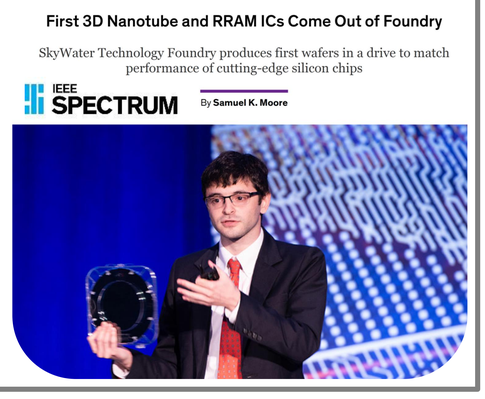


Application: MNIST CNN running inference continuously — Healthy, Permanent Write Failure (PWF)



Foundry CNFET + RRAM + Monolithic 3D

DARPA 3DSoC Program:



References

[1] M.M.S. Aly et al., "Energy-efficient Abundant-data Computing: The N3XT 1,000x", Computer Magazine, 2015.
[2] M.M. Shulaker et al., "Three-dimensional integration of nanotechnologies for computing and data storage on a single chip", in Nature, 2017.
[3] T. Wu et al., "Brain-Inspired Computing Exploiting Carbon Nanotube FETs and Resistive RAM: Hyperdimensional Computing Case Study", in ISSCC 2018.
[4] T. Wu et al., "A 43pJ/cycle Non-volatile Microcontroller with 4.7µs Shutdown/Wake-up integrating 2.3 bits-per-cell Resistive RAM and Resilience Techniques", in ISSCC 2019.
[5] M.M.S. Aly et al, "The N3XT Approach to Energy-Efficient Abundant-Data Computing", in Proc. IEEE 2019
[6] B.Q. Le et al, "Resistive RAM With Multiple Bits Per Cell: Array-Level Demonstration of 3 Bits Per Cell", in IEEE TED 2019.

DARPA · NSF · SRC Semiconductor Research Corporation · Stanford SystemX Alliance · STARnet · E2CDA - ENIGMA · leti ceatech