

# Digital design: challenges and solutions

*Giovanni De Micheli*

**EPFL**



# The challenge for this decade and beyond

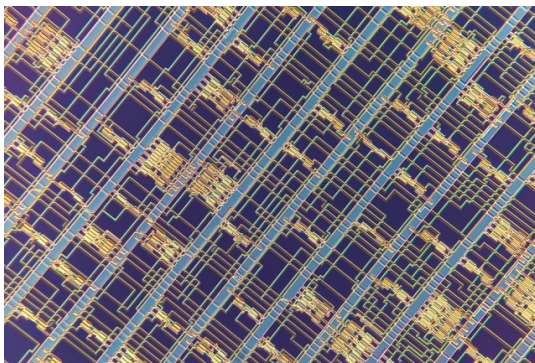
- The market is driven by AI/ML applications:
  - Large storage space
  - Large amount of energy for computation
- How do we rethink architectures, circuits and devices ?
  - To enable *edge devices* to use AI/ML
  - To curb the energy consumption
- Human/planetary dimension of AI/ML
  - Ethical questions

(c) Giovanni De Micheli

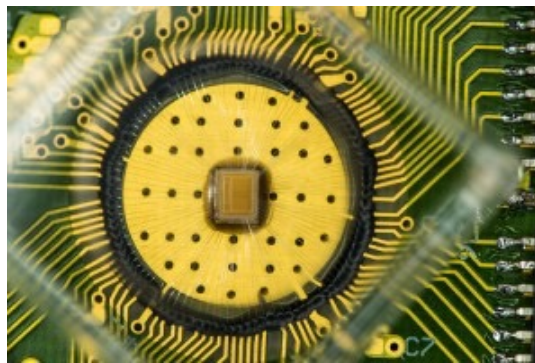


# Where is the technological space for growth?

- Growing computing and storage needs
  - *Data Centers* account 50GW in 2025 with 15% yearly growth
- Skyrocketing costs of sub-nano *fabrication*
  - TSMC's three fabs in Arizona budgeted at 65 B USD
- Diversity
  - A plurality of technologies may be combined to *accelerate* computation



[Shulaker, MIT 2021]



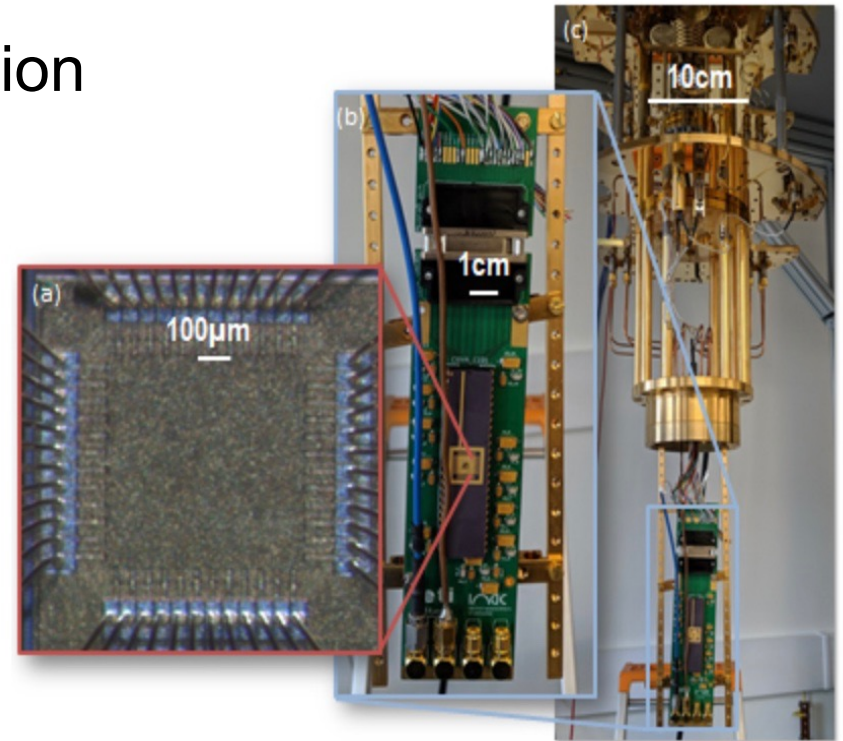
[Boybat, IBM 2020]  
(c) Giovanni De Micheli



[Ramey, Hot Chips, 2020]

# There is plenty of space at the bottom

- Quantum computing accelerates computation
  - In some restricted but relevant domains
- Error tolerance and corrections
  - Evolution from NIQC to FTQC
- QC will be provided as a service
  - Will QC reach the workplace, home, edge?



[Source: LETI]

(c) Giovanni De Micheli

# Quantum computing

- Two facets of quantum engineering:
- The physical world
  - Using matter to compute with *superposition* and *entanglement*
- The computational world
  - Designing and *compiling* algorithms into quantum circuits
- EDA can be instrumental for R&D in both domains
  - Advanced *physical design* and *quantum compilation*



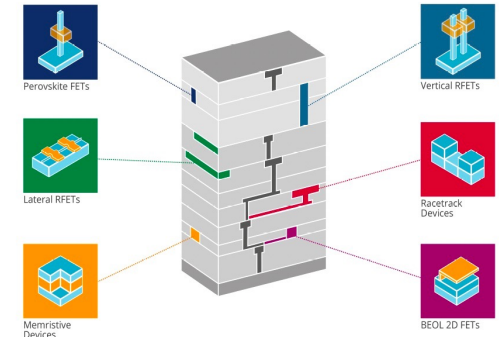
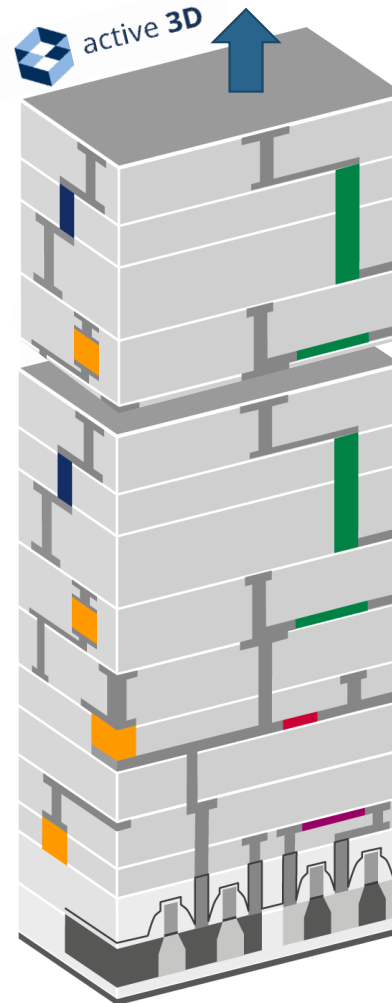
[P. Picasso]

# There is plenty of space at the top

## ABEOL Active 3D

active FEOL, active BEOL

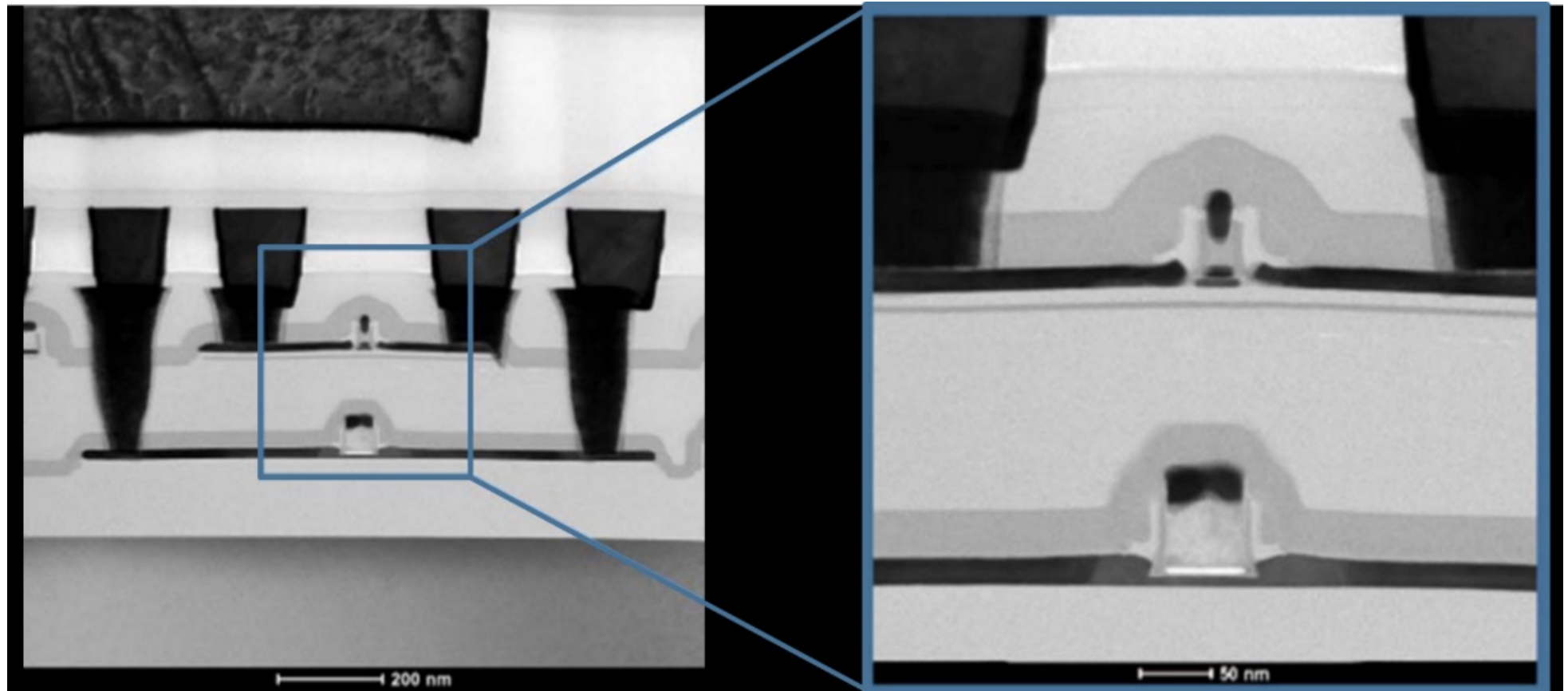
- Devices integrated into **volume** BEOL
- **Scalable approach** by adding additional layers
- Can be combined with different base technologies



[ TU Dresden and RWTH Aachen ]

(c) Giovanni De Micheli

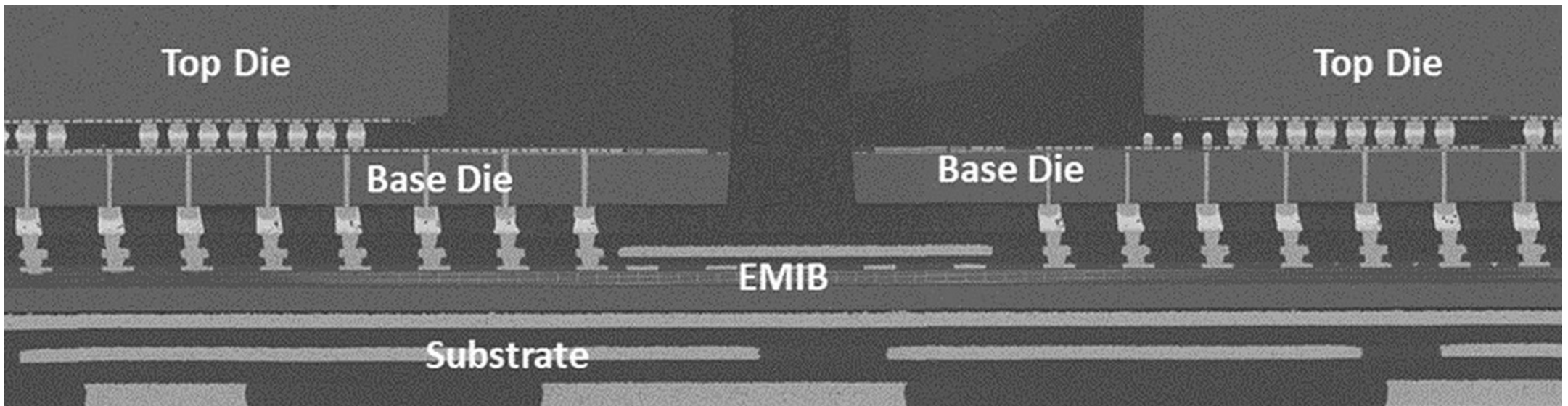
# 3D Integration



(c) Giovanni De Micheli

[Batude et al., VLSI 2011]

# Chiplets

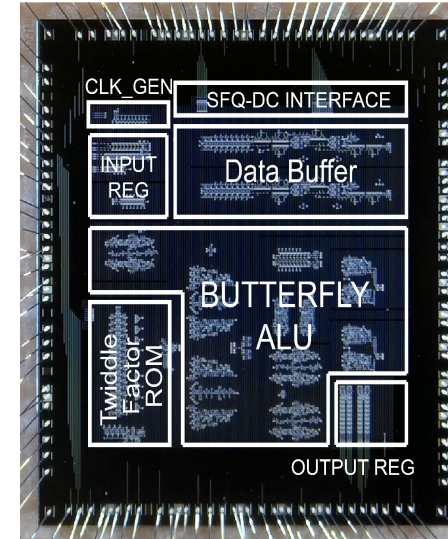


[Intel, 2022]

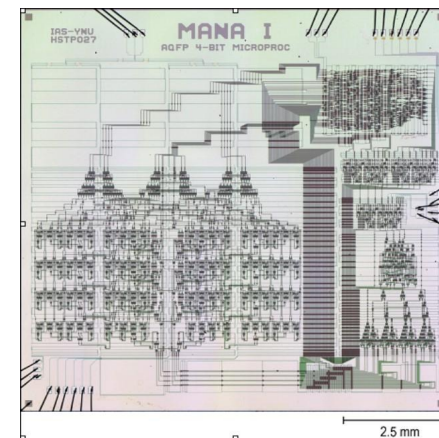
(c) Giovanni De Micheli

# There is plenty of space in the cold

- Superconducting electronics
  - No parasitic resistance at low temperature (4K)
  - Information by *quantized pulses*  $\int V(t)dt = \phi_0 = h/2e = 2.07 \text{ mv ps}$
- Design features
  - *Classic computing paradigm with deep pipelined logic*
  - Many variants including adiabatic operation



47 GHz SFQ FFT Processor [Ke et al., 2021]



1.5 GHz 4-bit RISC [Ayala et al., 2021]

(c) Giovanni De Micheli

# Superconducting electronics

Logic	Clock Freq. [GHz]	$E_{\text{bit}} / I_c \Phi_0$	Typical $I_c$ [mA]	EDP [aJ·ps]	
CMOS	4	-	-	$\sim 10^5$	
RSFQ [1]	50	19	150	120	
eSFQ [2]	20	0.8	150	12	← Single-flux
RQL [3]	10	0.33	150	10	
LV-RSFQ [4]	20	3.5	150	54	
AQFP [5]	5	0.0083	50	0.086	← Adiabatic
Quantum limit	-	-	-	$5.3 \times 10^{-5}$	

- [1] X. Peng et al., IEICE Trans. Electron. **E97.C**, 188 (2014).  
 [2] M. H. Volkmann et al., Supercond. Sci. Technol. **26**, 015002 (2013).  
 [3] Q. P. Herr et al., J. Appl. Phys. **109**, 103903 (2011).  
 [4] M. Tanaka et al., IEEE Trans. Appl. Supercond. **23**, 1701104 (2013).  
 [5] N. Takeuchi et al., Supercond. Sci. Technol. **28**, 015003 (2015).

(c) Giovanni De Micheli

# Superconducting electronics

## **Support for ultra fast acceleration**

General purpose accelerators but also for AI/ML

## **Low power operation**

Limited by cooling technology

## **Bridge to Quantum Computing processors**

Intermediate temperature range

## **Need for new EDA tools**

Activities in China, Japan and USA

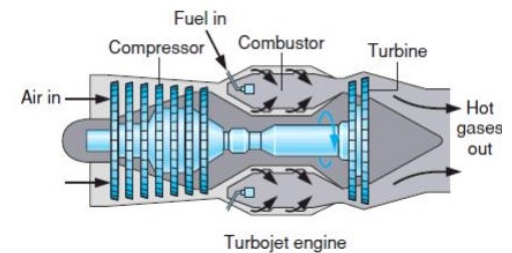
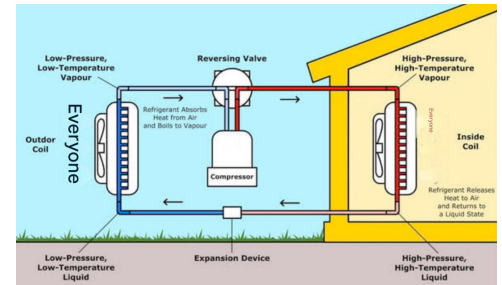
x

# There is not much space in the hot !

- Thermodynamics of computing systems is the limiting factor
  - Reduce/recycle energy cost and heat dissipation of computation
  - Current roadblock for AI/ML and data centers
- The search for energy-efficient computing systems may involve:
  - *Reversible* and/or adiabatic computation
- Computing systems are still very low in terms of efficiency
  - Heat and CO<sub>2</sub> generation per computational task

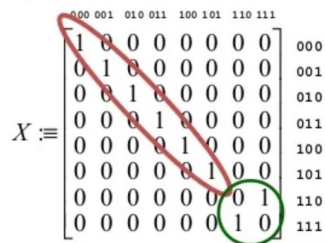
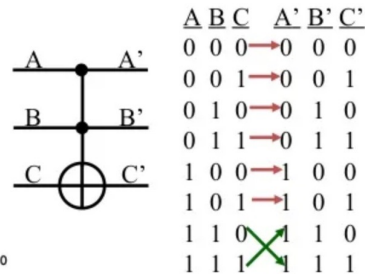
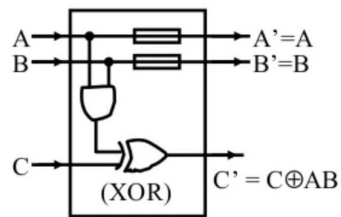
# Thermodynamics of computing

- Reversible logic
  - A reversible process is more efficient than an irreversible one (Carnot's Theorem)
  - Quantum computing leverages reversible processes
  - Reversible computing also in CMOS [Varie 2025]
- Adiabatic logic
  - An adiabatic process has no heat exchange with the exterior
  - Adiabatic computing in CMOS:
    - Charge/discharge capacitors over a small  $\Delta V$
  - Adiabatic computing in SCE
    - Charge/discharge inductors over a small  $\Delta I$



# Example: reversible logic gates

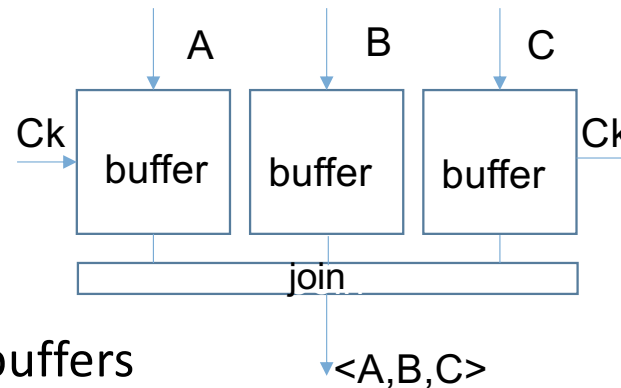
- Toffoli gate (QC)
  - Permutation



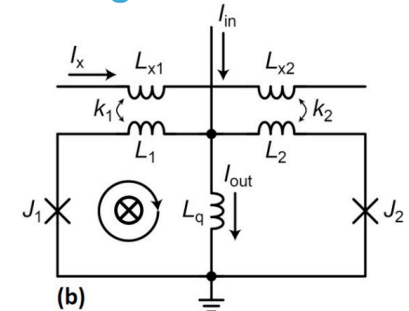
- Vaire computing (CMOS)
  - Reversible adder embeded in LC resonator
  - Adder is C; Sine-wave supply
- Energy recovery
  - Adiabatic operation

# Example: Superconducting AQFP circuits

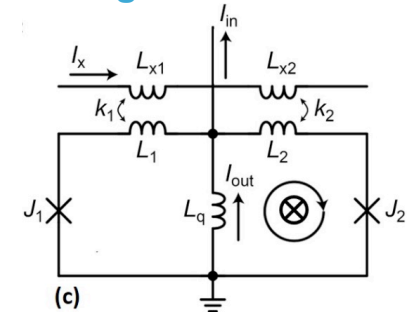
- *Adiabatic Quantum Flux Parametron* [Takeuchi 13]
- Very small dynamic power consumption
- Majority-based logic
  - Free inversion
- Clocked logic
  - Clock at  $\approx 5$  GHz
  - Path balancing using clocked buffers



Logic state '1'



Logic state '0'



# Summary:

## Diversity of Architectures and Devices

- The competitive advantage of emerging technologies:
  - Adapting computational model to computational fabric
- Application-specific accelerators are best examples
  - Very desirable for both cloud and edge systems
- What is the effort to achieve successful diversity?
  - New computational thinking models
  - Adapting/creating EDA tools and flows

# EDA for emerging technologies and paradigms



# There's still tomorrow for *Logic Synthesis*

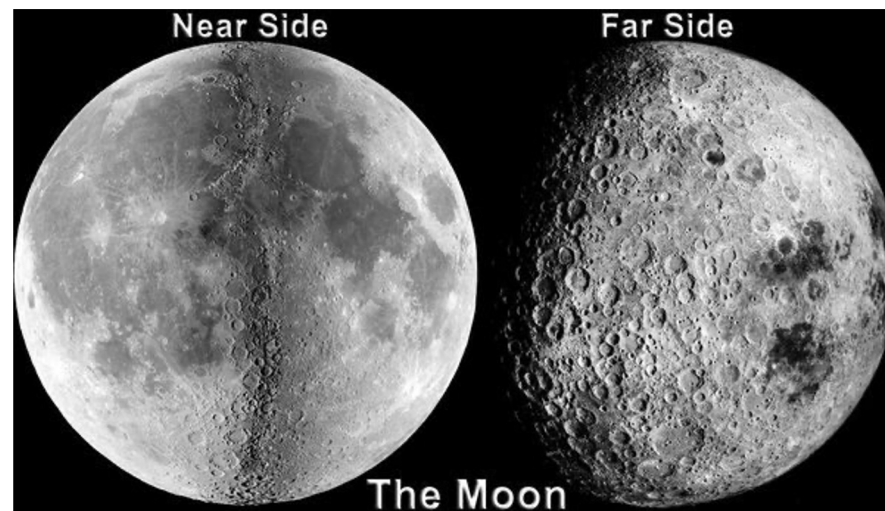


[P. Cortellesi 2023 Film]

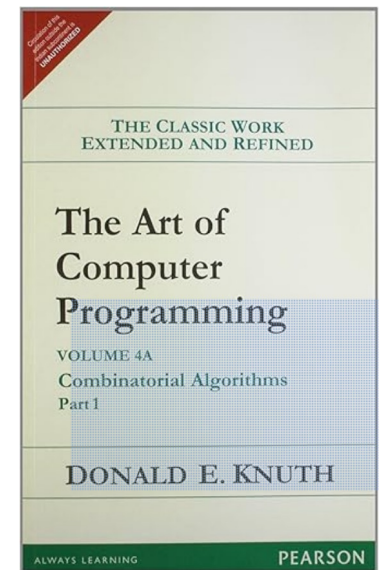
(c) Giovanni De Micheli

# There's still tomorrow for *Logic Synthesis*

- The main fabrics of most chips is a network of gates
- We know a lot about logic networks and synthesis
- But there is more that we *don't know* than we *know*
  - Is that relevant?



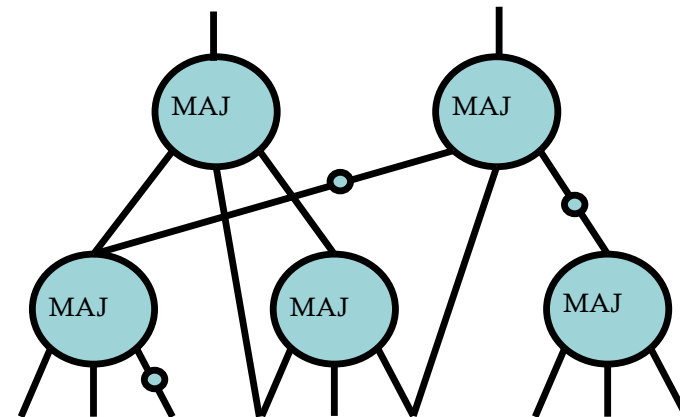
(c) Giovanni De Micheli



# Majority logic synthesis

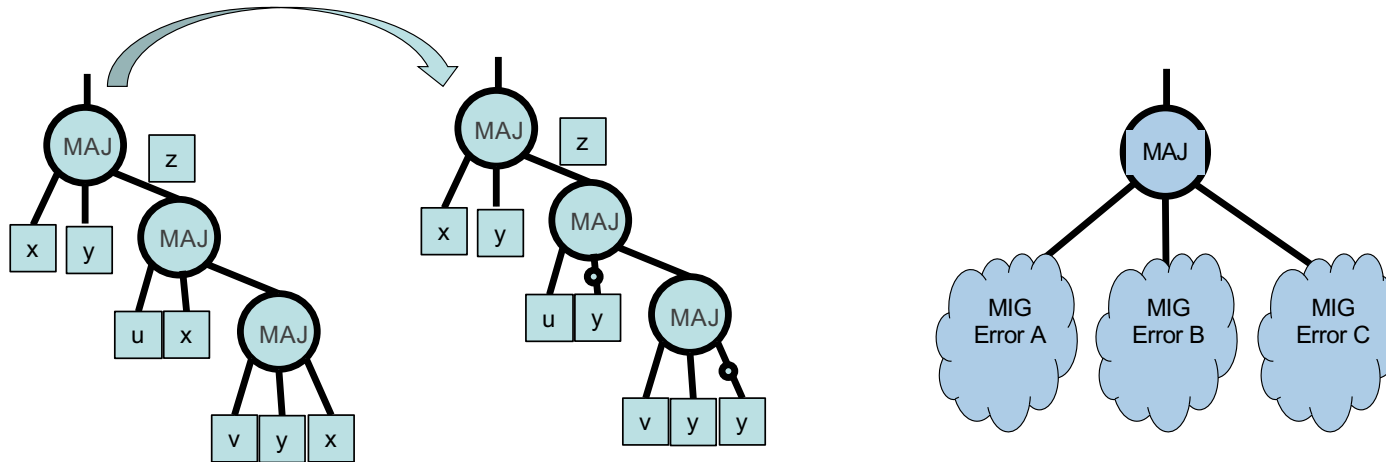


- Majority logic axioms
  - Sound and complete
  - Reachability property



# Algorithms for MIG optimization

- Optimization algorithms for area and delay
  - Algebraic methods based on rewriting
  - Boolean methods exploiting don't cares



(c) Giovanni De Micheli

# Algorithms for MIG optimization

## Majority-based algorithms:

improve commercial and academic logic synthesis for **CMOS**  
15-20% delay reduction as compared to earlier methods

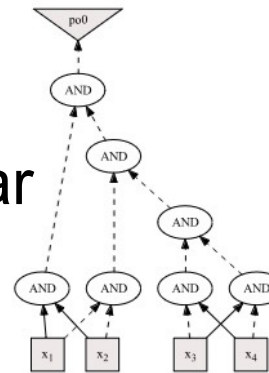
are very successful for many **emerging technologies**, like:

superconducting electronics  
optical/wave based computing  
logic in memory

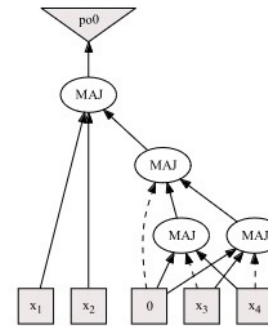
# The role of linearity

- AND-OR logic is inherently non-linear
- MAJority-Inverter logic as well
- EXOR are linear operators
  - EXOR play a leading role in realizing adders and multipliers
  - Adders require at least one AND gate per bit
- It is useful to decompose functions into linear and non-linear parts

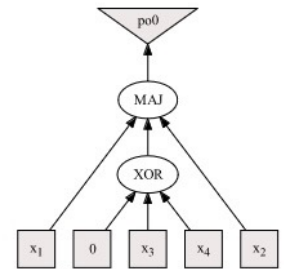
AIG



MIG

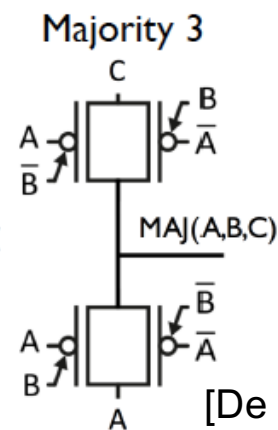
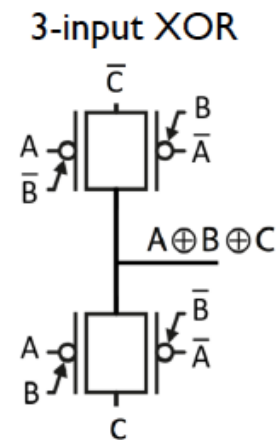
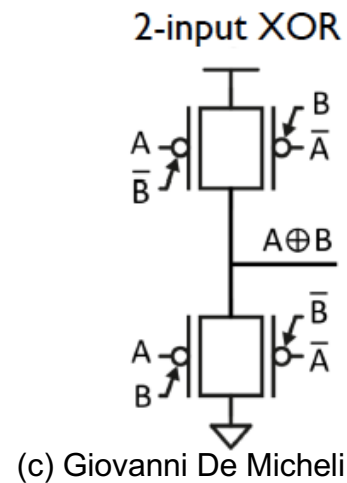
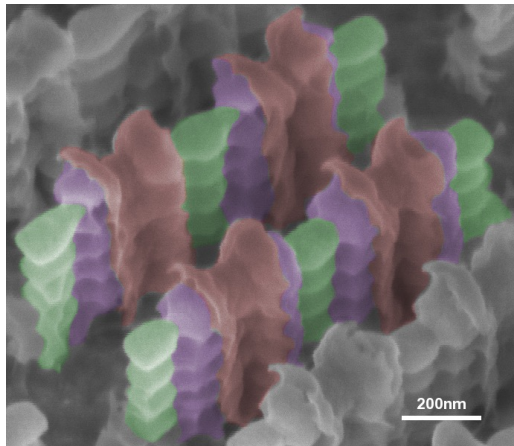
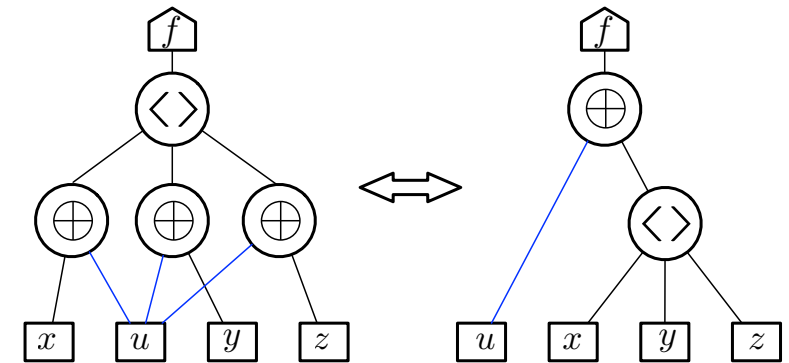


XMG



# XMGs – Controlled-polarity transistors

- XOR – MAJority graph
  - Extension of MIGs (replace INVerter by XOR)
  - Self-duality of XORs (odd) and Majority
- XOR can flow through MAJ nodes
- Data structure for controlled-polarity gates



[De Marchi, IEDM 12]  
[Rai et al., TCAD 22]

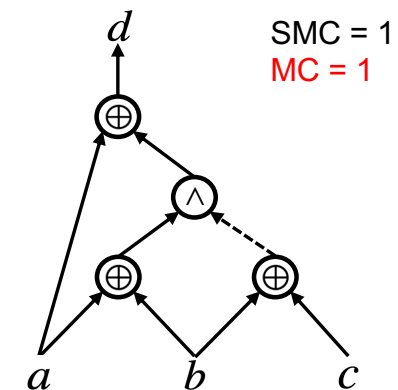
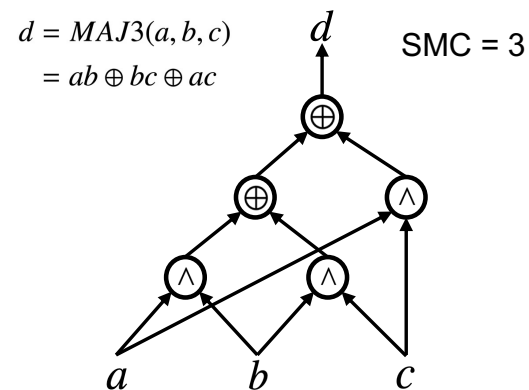
# The role of multiplicative complexity

- Network model: XOR-AND graph
  - Abstraction of arithmetic addition and multiplication
  - Separate linear from non-linear component
- Multiplicative complexity (MC) is the minimum #AND to realize a function
  - Intractable problem
  - MC is known for functions with up to 6 inputs
- Multiplicative depth (MD) is the longest path of AND gates

# Multiplicative complexity

- The *structural* MC ( # ANDs) is an upper bound to MC
- Example
  - Majority-3

$a$	$b$	$c$	$d = MAJ3(a, b, c)$
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	1



[ Ç. Çalık, 2019]

(c) Giovanni De Micheli

# Using multiplicative complexity and depth

- Some problems in logic design, quantum compilation and security can be reduced to MC minimization
  - Synthesis of circuits robust to *Side-channel Attacks* [Testa 2019]
  - Minimizing *T-Cells in Quantum Compilation* [Meuli, Nature QI, 2022]
  - Reducing cost of implementing and executing *Garbled Circuits* [Yu, 2023]
  - Improving efficiency of computation with *Fully Homomorphic Encryption*
- Efficient logic synthesis algorithms for MC reduction
  - MC is invariant over affine transformations
  - Various heuristic methods

# Fully Homomorphic Encryption (FHE)

- Compute functions using encrypted data
  - Speed of computation is essential
- Apply logic synthesis to restructure computation
  - While keeping noise under a threshold level
- Key metrics:
  - Multiplicative complexity – multipliers generate noise
  - Multiplicative depth – critical path
- Approaches:
  - MC-aware MD optimization ( $MC * MD^2$ )
  - ESOP balancing [Haner et al., TQC 22]

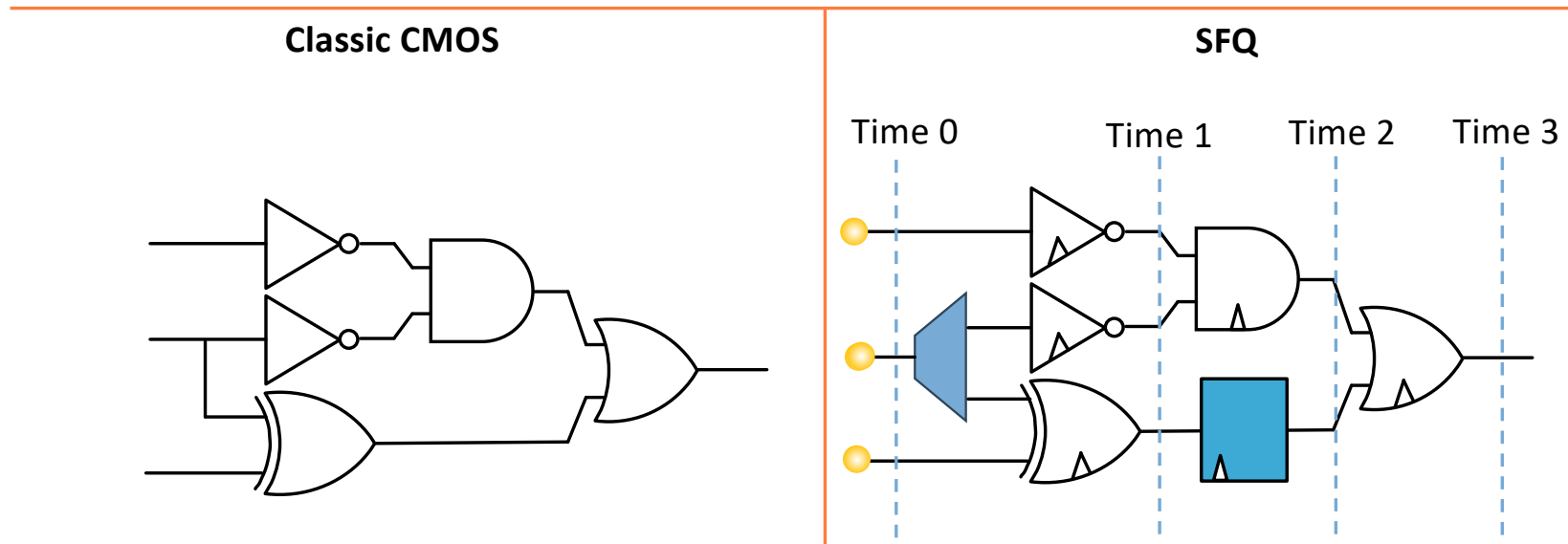


# Pipelining in new logic synthesis paradigms

- Leveled FHE circuits
  - FHE operations amplify noise (especially multipliers)
  - Above a certain noise level, data cannot be reconstructed
  - Bootstrapping is used to recondition data
- Bootstrapping has a cost and needs to be distributed in the circuit
  - Leveled FHE circuits are reminiscent of pipelined logic
  - Retiming-like techniques can be used to best place bootstraps

# Pipelining SCE circuits

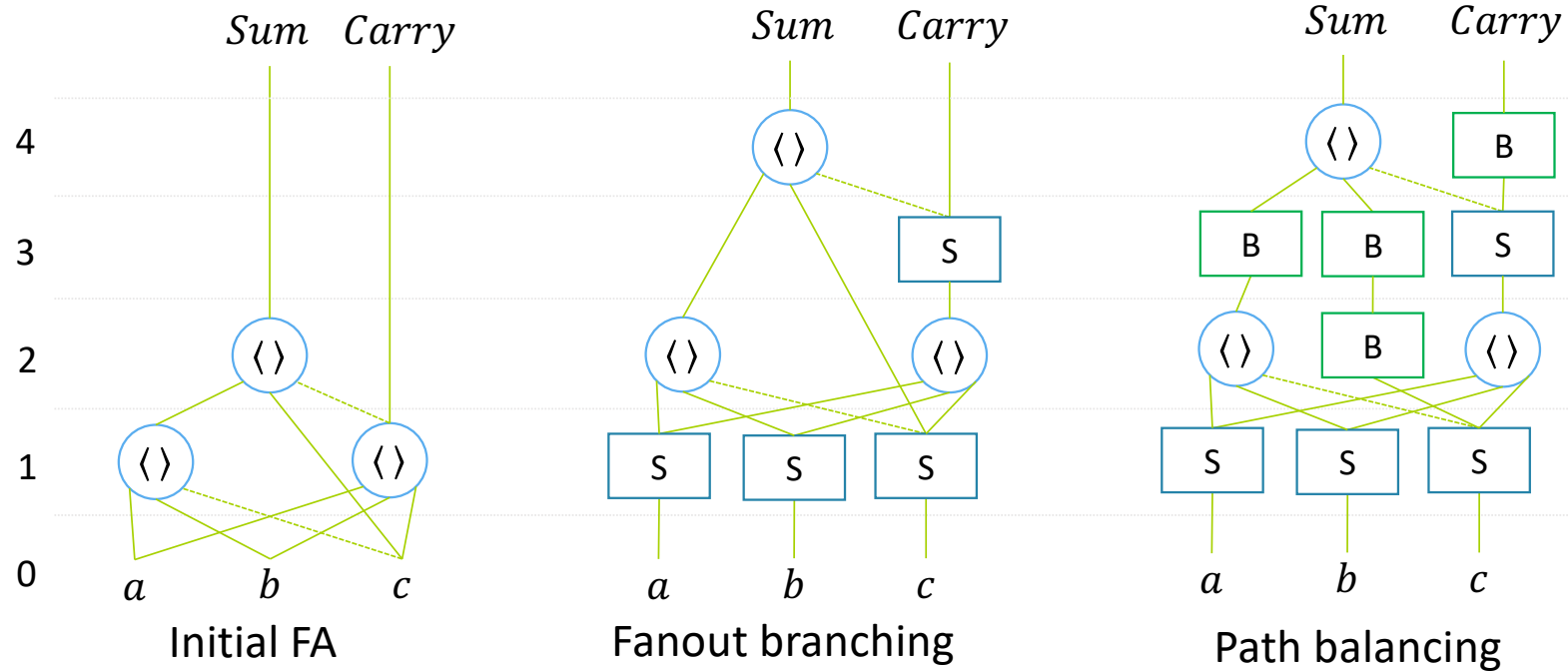
- SCE circuits are pipelined at the gate level (gates are clocked)
- The entire circuit need to be *balanced* (correct input when gate execute)
- Limited gate driving capability requires *splitter* insertion



(c) Giovanni De Micheli

# Balancing and splitting in AQFP

- Splitters are clocked
  - Balancing and splitting are interdependent



(c) Giovanni De Micheli

# The elephant in the technology room



- What is the role of AI/ML ?
- Applicable almost everywhere!
- *Design space* exploration mainly

(c) Giovanni De Micheli

## **Summary:** **EDA is the enabler of future computing**

- AI/ML applications require design of novel architectures by EDA tools
- EDA theory and practice enables new technologies
- Synthesis allows designers to reach multiple goals beyond PPA
- Verification is essential, especially in view of new modeling methods

# Conclusions

- Efficient computing needs a match between architecture and technology
  - Accelerators will exploit the diversity of materials and circuits
- Logic modeling and reasoning is a fundamental tool
  - Algorithms for diverse design may leverage a wealth of techniques
- Chip design is entering a new exciting epoch
  - To sustain the fast evolution of algorithms, software and services

# Thank You

