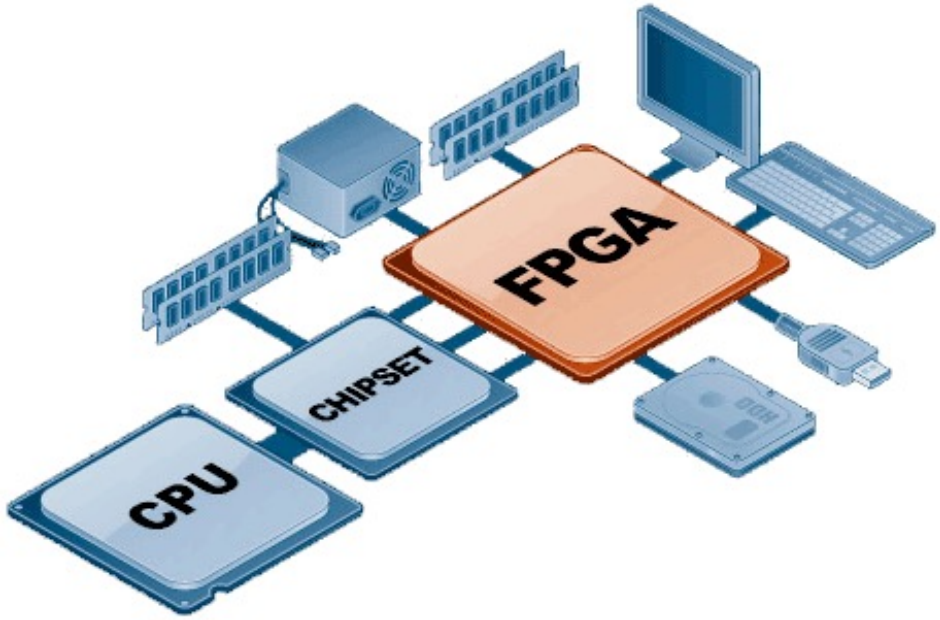
The Case for FPGAs

The Case for More FPGAs

Outline

- Introduction
- Popularity of FPGAs
- Advantages of FPGAs
- New Era for FPGAs



A Weird Idea

- Back in the early 1980s, IC designers tried to get the most out of each and every transistor on their circuits
- Ross Freeman
  - Proposed a chip packed with transistors that formed loosely organized logic blocks that in turn could be configured and reconfigured with software
  - So sometimes a bunch of transistors wouldn't be used
  - Betted that Moore's Law would eventually make transistors really cheap

World's first commercial FPGA introduced in 1985

What is an FPGA?

- Field-programmable gate array (FPGA)
  - An IC whose function and wiring can be re-programmed
  - Different from ASIC (application specific integrated circuit), it is not designed and made for a specific system

It's Hot

EE Times

News & Analysis

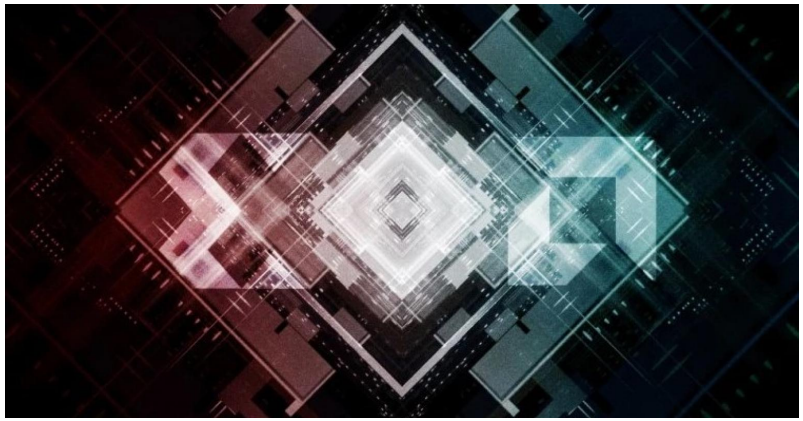
Intel Seals $16.7 Billion Altera Deal

Dylan McGrath

12/28/2015 04:02 PM EST

> Waxman said that by 2020 Intel believes a third of the data center market could be using the type of chips that Altera specializes in.

It’s Hot



AMD closes landmark $50B acquisition of chipmaker Xilinx

“Xilinx offers industry-leading FPGAs, adaptive SoCs, AI inference engines and software expertise that enable AMD to offer the strongest portfolio of high-performance and adaptive computing solutions in the industry and capture a larger share of the approximately $135 billion market opportunity we see across cloud, edge, and intelligent devices”.

Dr. Lisa Su, Chair and CEO (2022)

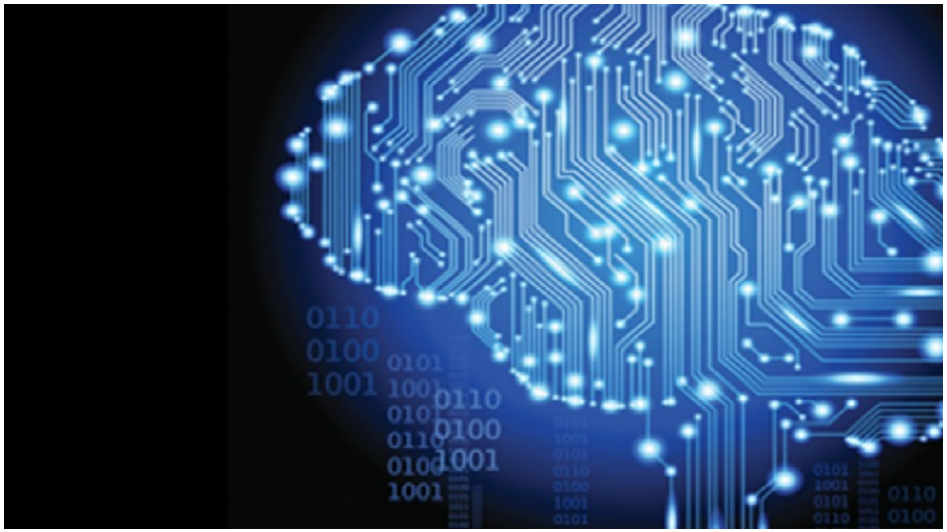
FPGA Everywhere

- FPGAs inside a lot of consumer electronics

Numerous FPGA Applications

- Medical
- Advanced driving assistance
- Emulation system
- Military
- Aerospace

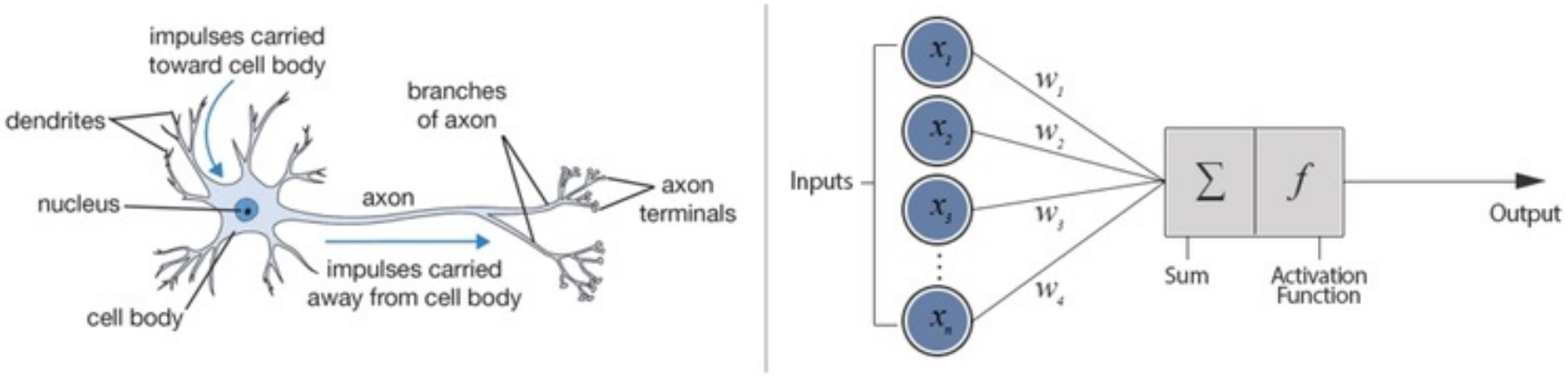
More FPGA Deployment



- Data center
- Machine learning
- Telecommunication

Why FPGA is a good match for Convolution Neural Network?

Biological neuron vs Artificial neuron



Why FPGA is a good match for Convolution Neural Network?

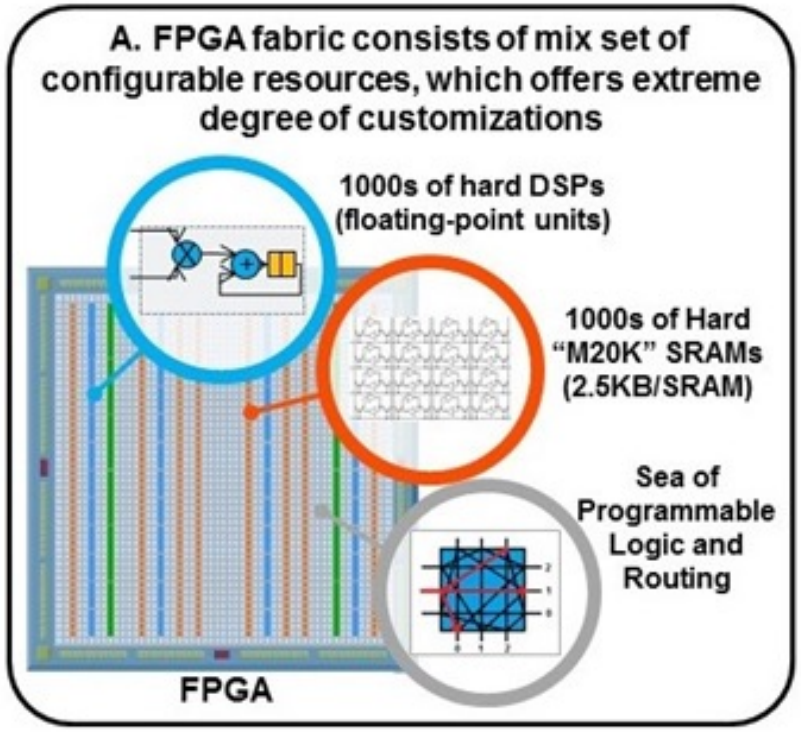
- FPGA offers
  - massively parallel architectures
  - efficient DSP resources (for numerical operations like dot-product accumulation)
  - adaptable to any numerical representations and sizes
  - large amount of on-chip memory

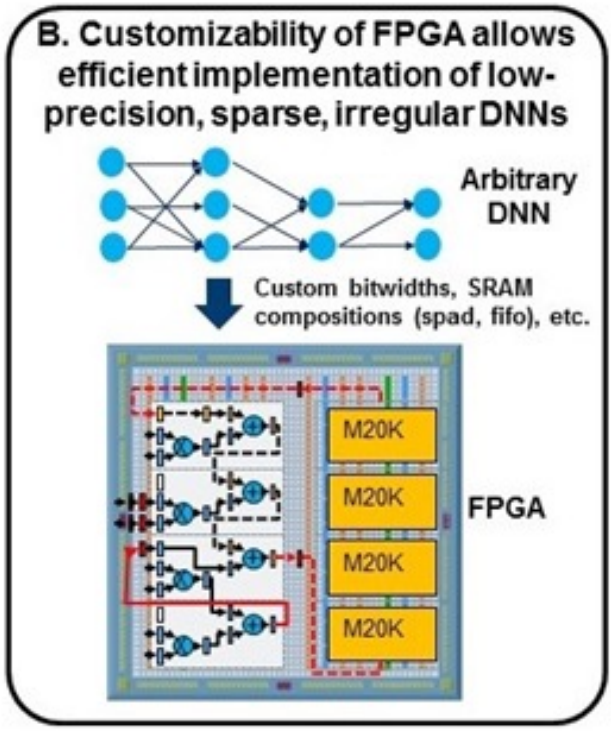
Primary DNN Calculation is Input Vector * Weight Matrix = Output Vector

Input Data | Neuron Weights | Outputs Equations

$$[X_0 \quad X_1 \quad \cdots \quad X_N] * \begin{bmatrix} A_0 & B_0 & C_0 \\ A_1 & B_1 & C_1 \\ \cdots & \cdots & \cdots \\ A_N & B_N & C_N \end{bmatrix} = \begin{bmatrix} Y_A = X_0A_0 + X_1A_1 + X_2A_2 \\ Y_B = X_0B_0 + X_1B_1 + X_2B_2 \\ Y_C = X_0C_0 + X_1C_1 + X_2C_2 \end{bmatrix}$$

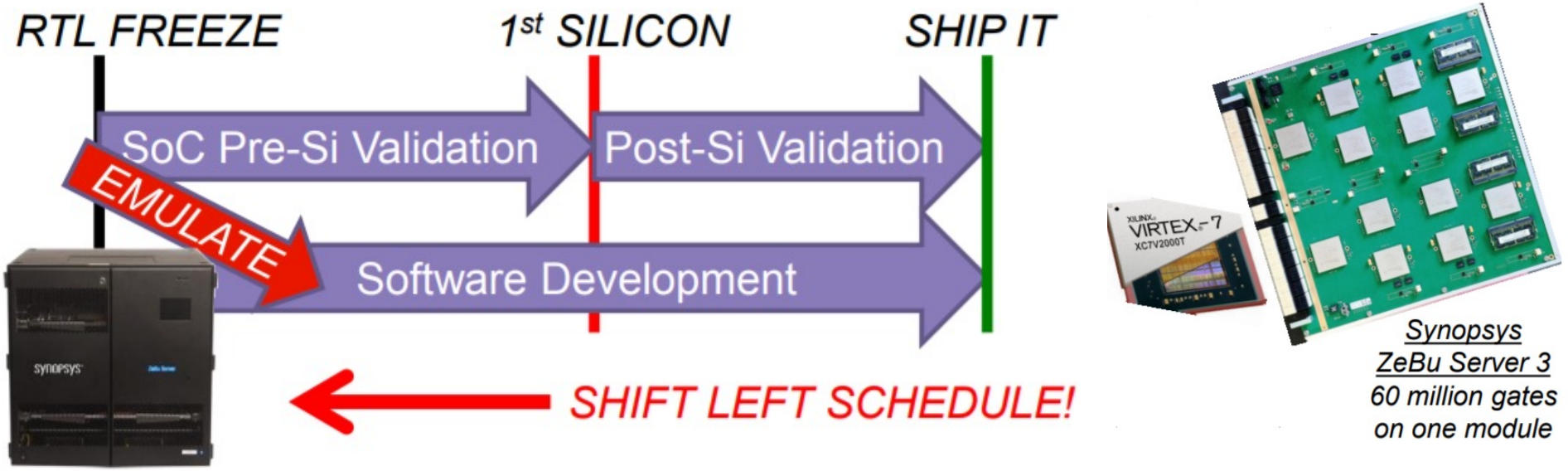
Key Operation: Multiply-Accumulate, or "MAC"





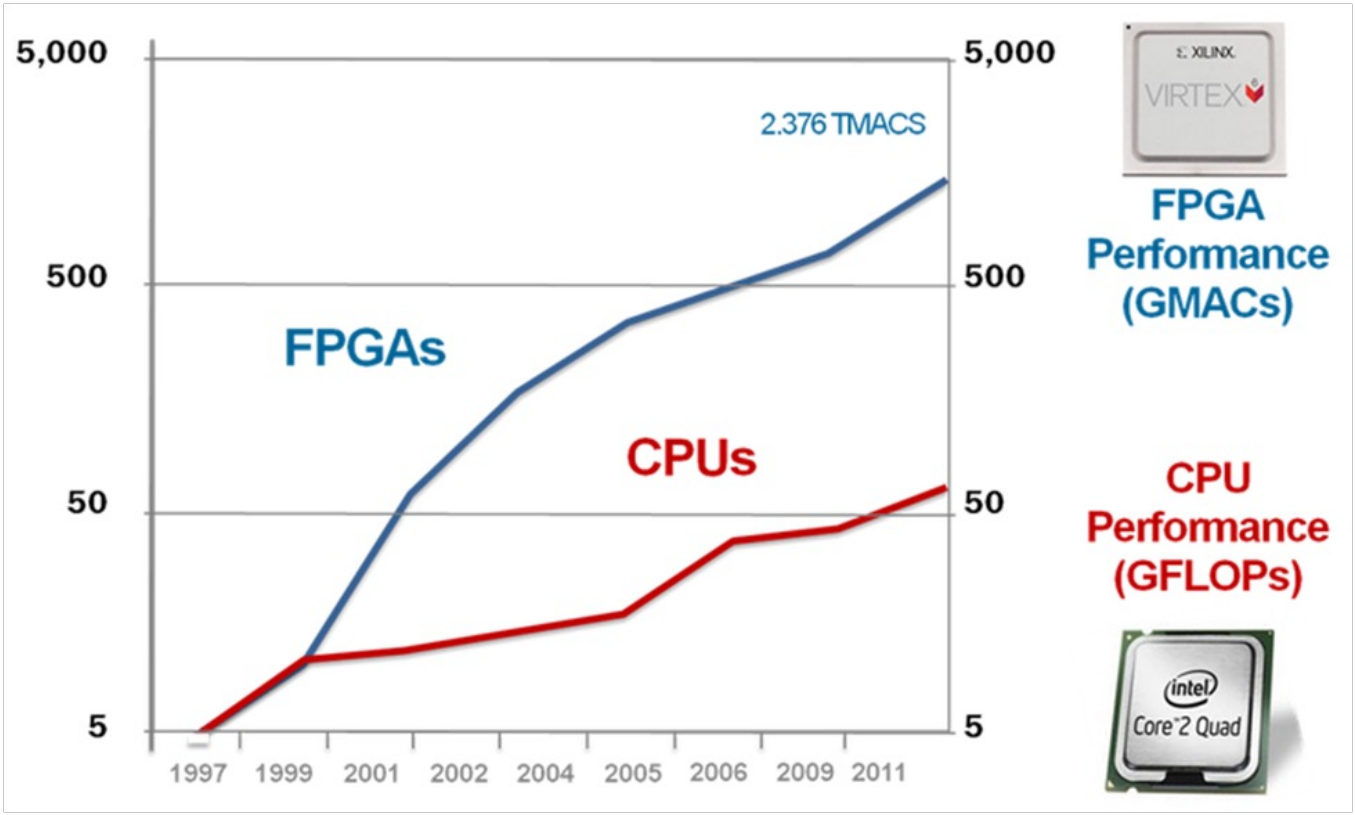
FPGAs for Logic Emulation

- State-of-the-art electronic system
  - Huge gate count (over hundreds of millions)
  - Before committing to silicon, need to verify that it can function correctly under all operating conditions
- FPGA-based logic emulation of electronic system can
  - Provide orders-of-magnitude faster debugging and functional verification compared to software based simulation
  - Bringup/develop/validate software part before silicon is ready



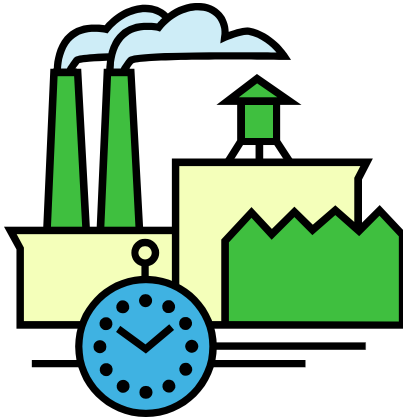
Moore's Law and FPGAs

- With technology scaling, FPGAs achieve
  - Increasing capacity
  - Faster performance
  - More functionality

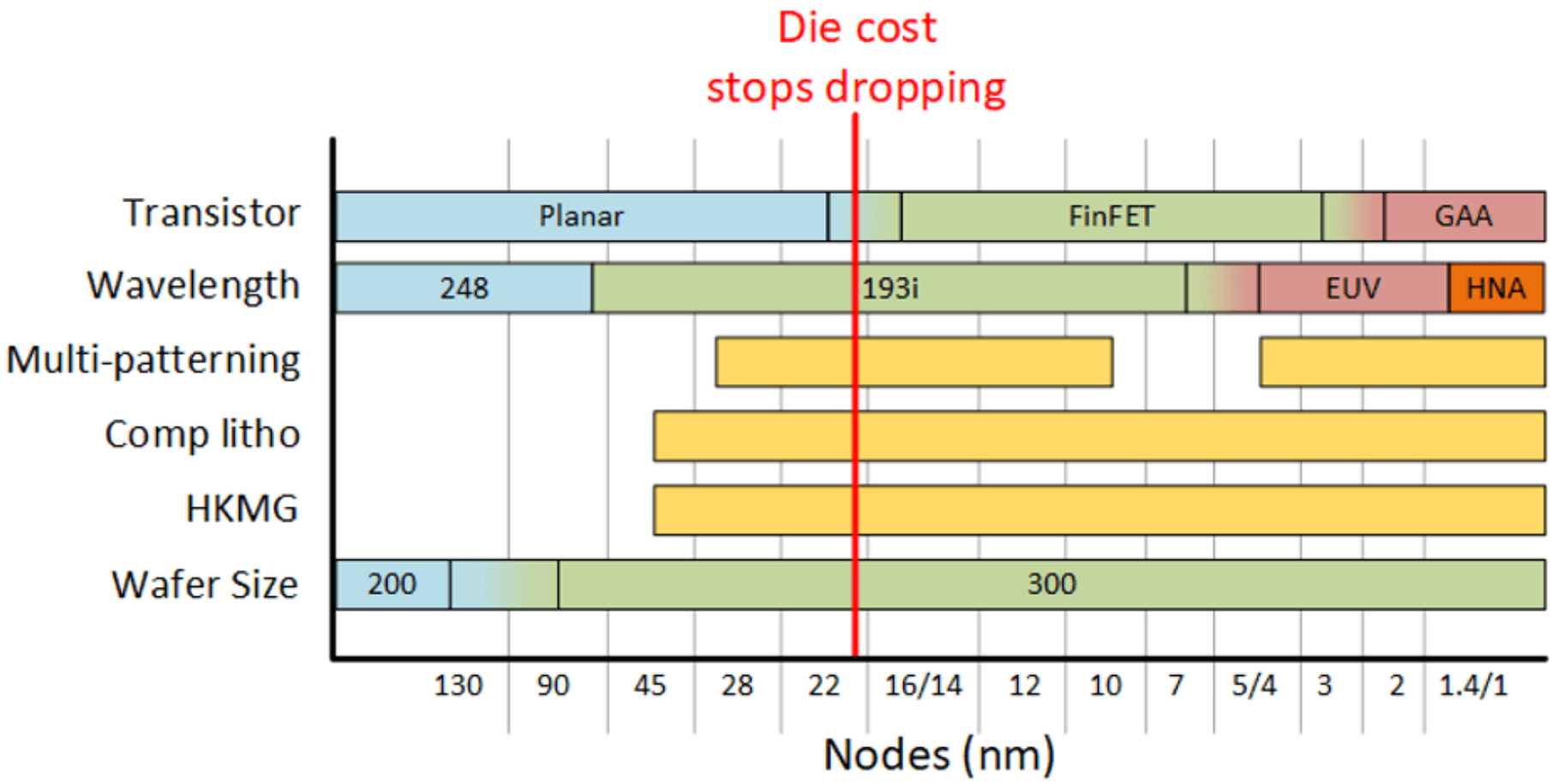


Cost of IC Fabrication

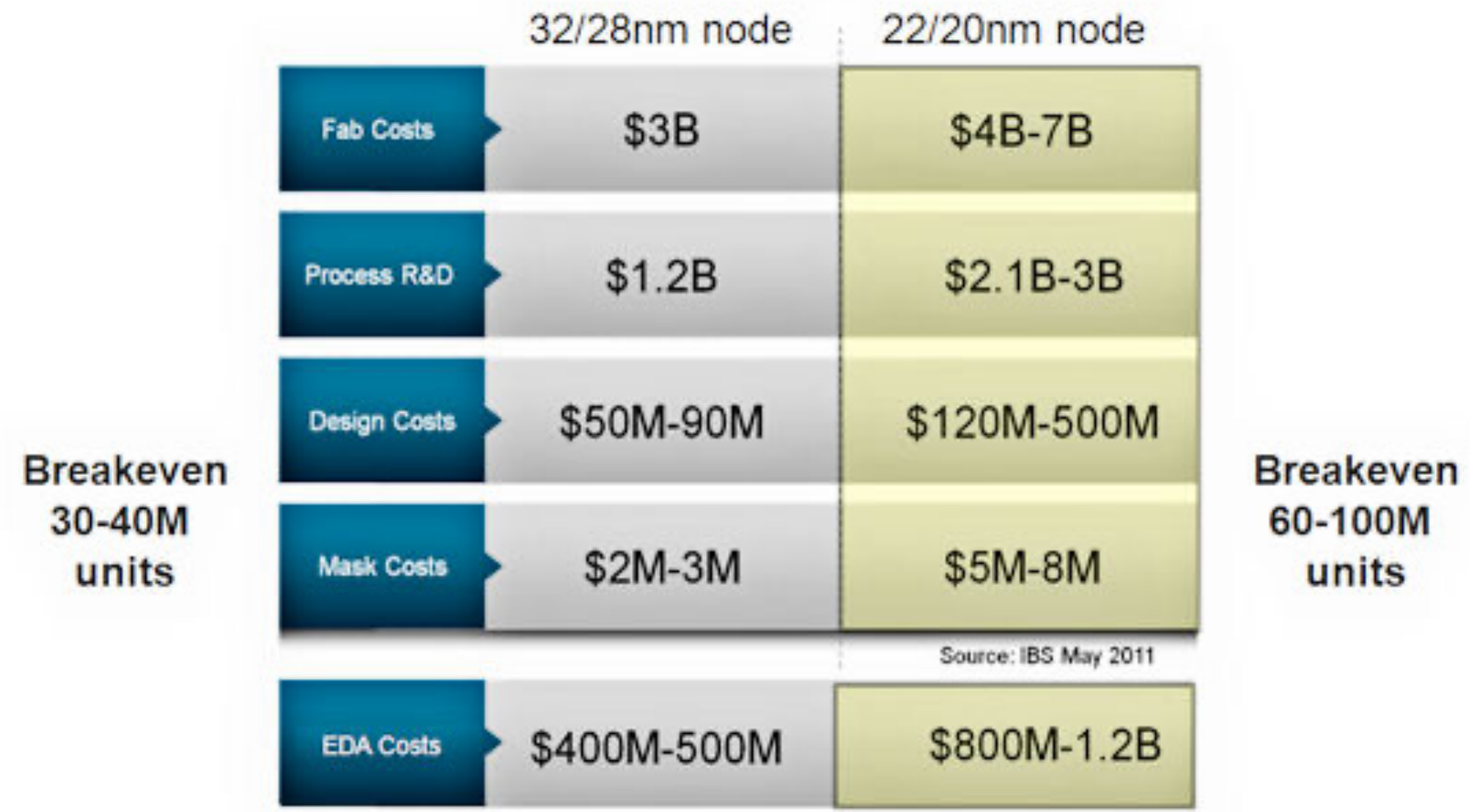
- Cost: billions of US dollars
- Typical fab line occupies about 1 city block, employs a few hundred people.
- New fabrication processes require 6-8 month turnaround.
- Most profitable period is first 18 months-2 years.



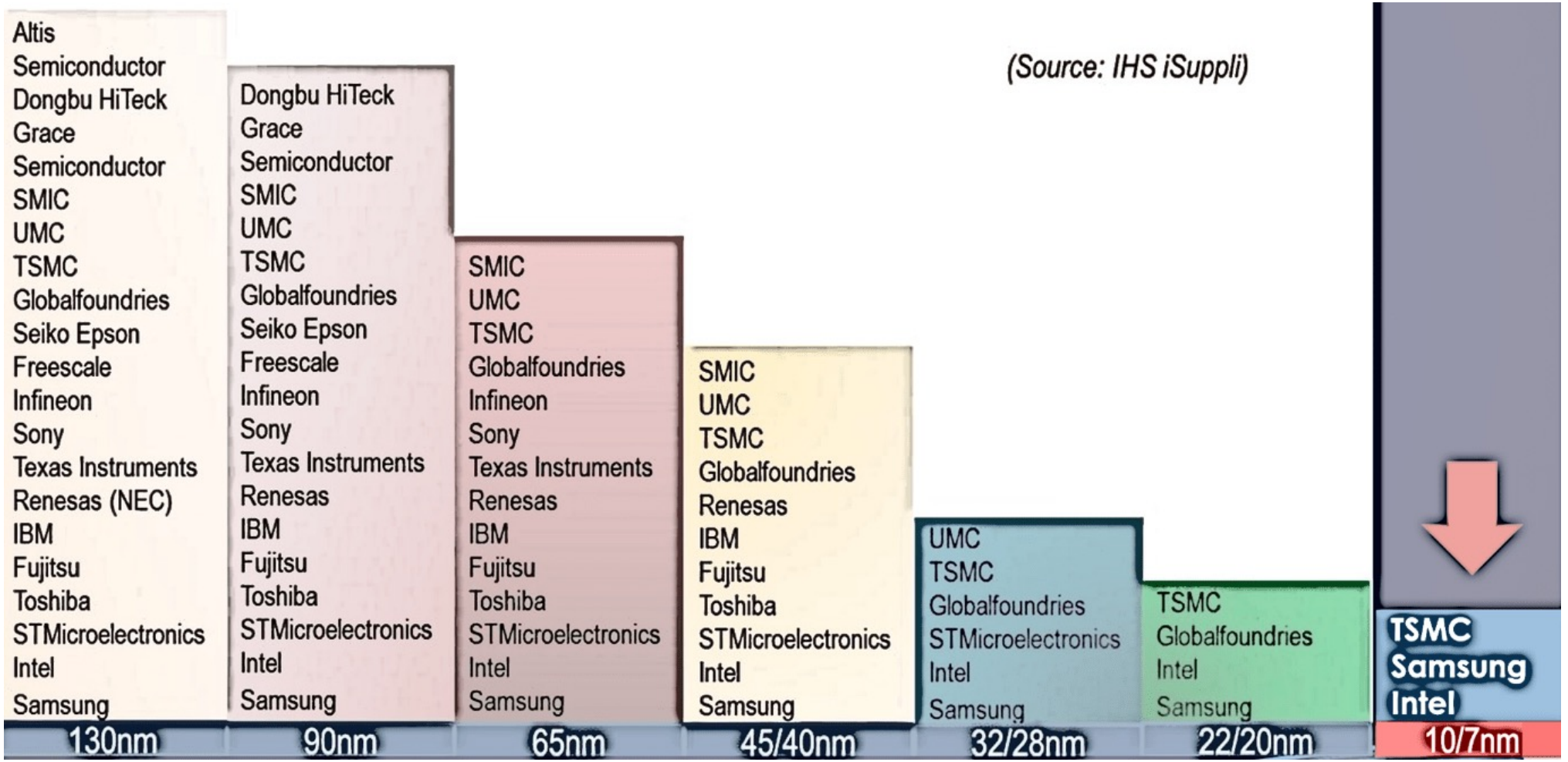
Changes in Silicon Processing



Massive Investment for Advanced-Node Design and Manufacturing



Severe Reduction in Number of Fabs for Advanced Nodes



Why FPGAs in Consumer Products?

- Prohibitive fabrication cost for custom silicon in advanced process nodes

  → not viable except for large volume requiring extremely high performance

- Need a cheaper *alternative*!
- FPGAs come to the rescue

FPGA vs ASIC

- FPGAs (field-programmable gate arrays) are standard parts:
  - Off-the-shelf
  - Pre-manufactured
  - Millions of customers share manufacturing costs
- ASIC:
  - Tailored to your application
  - Require own set of masks for manufacturing
  - High startup cost

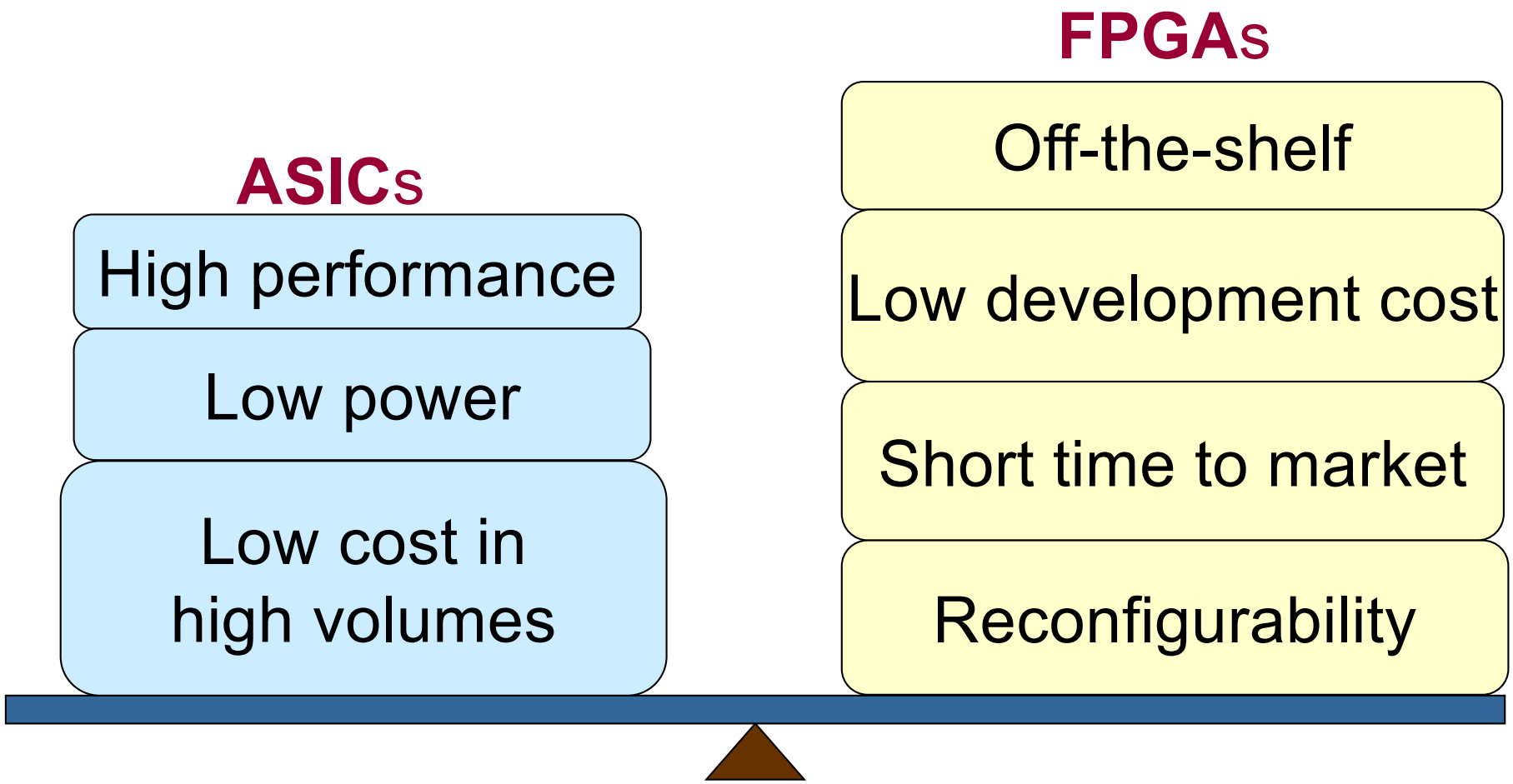
FPGA vs ASIC

Source: Xilinx

Standard Parts vs Custom Silicon

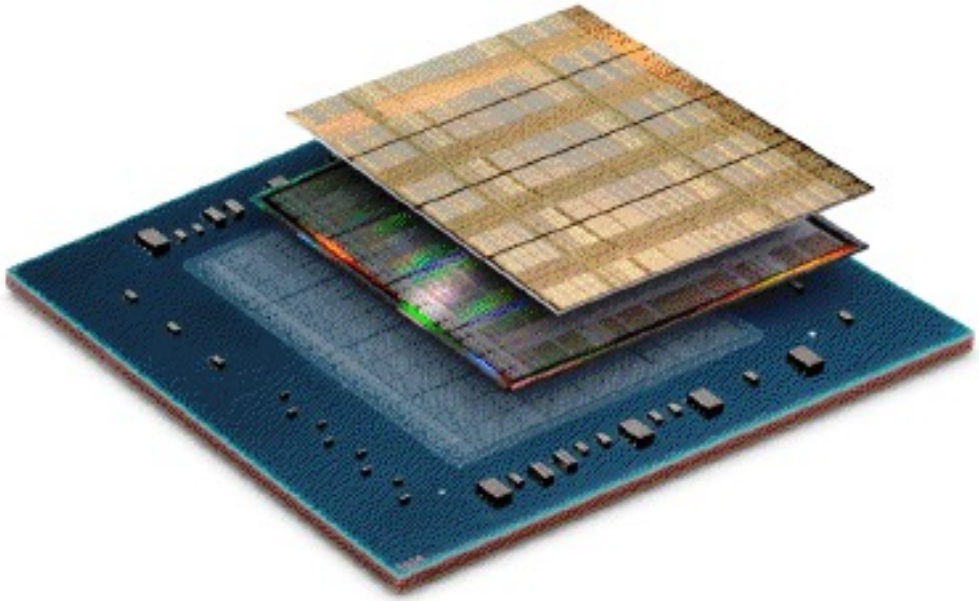
- When to build your design with an FPGA or with custom silicon?
  - FPGAs have shorter design cycle
  - FPGAs have no manufacturing delay
  - FPGAs reduce inventory
  - FPGAs are slower, larger, more power-hungry

Which Way to Go?



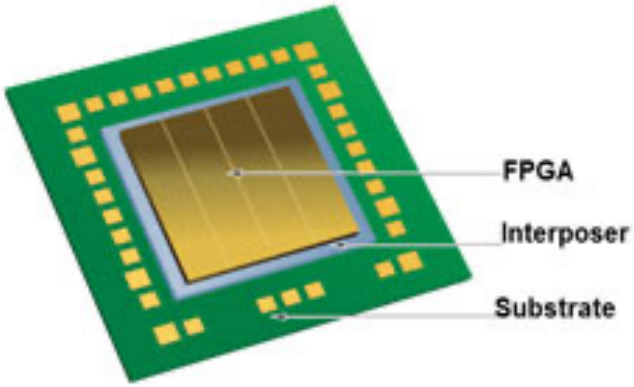
2.5D FPGAs

- Xilinx's interposer-based 2.5D FPGAs
  - First commercial 2.5D IC: Virtex-7 2000T
    - 4 dies
    - Die-to-die connection through interposer-based 2.5D technology
    - 6.8 billion transistors (~ 20 million ASIC gates)

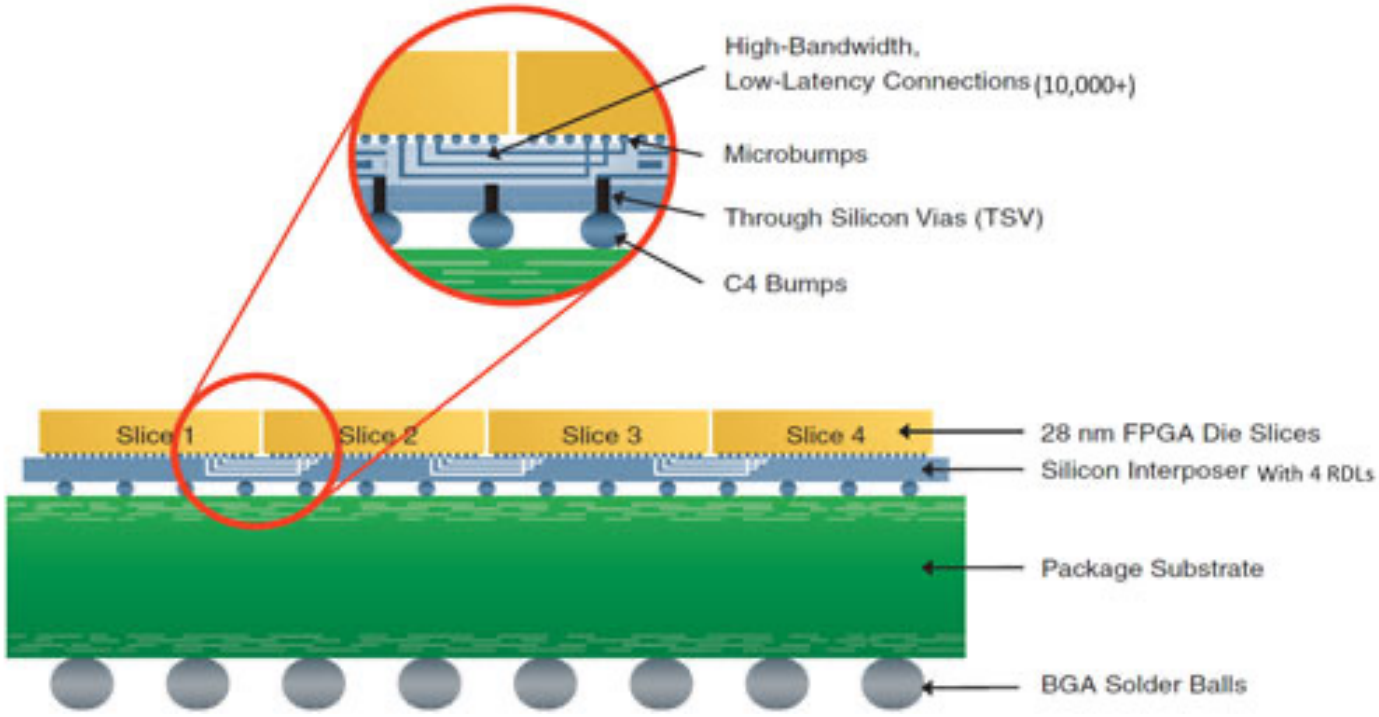


2.5D FPGAs

- Interposer-based inter-die connection

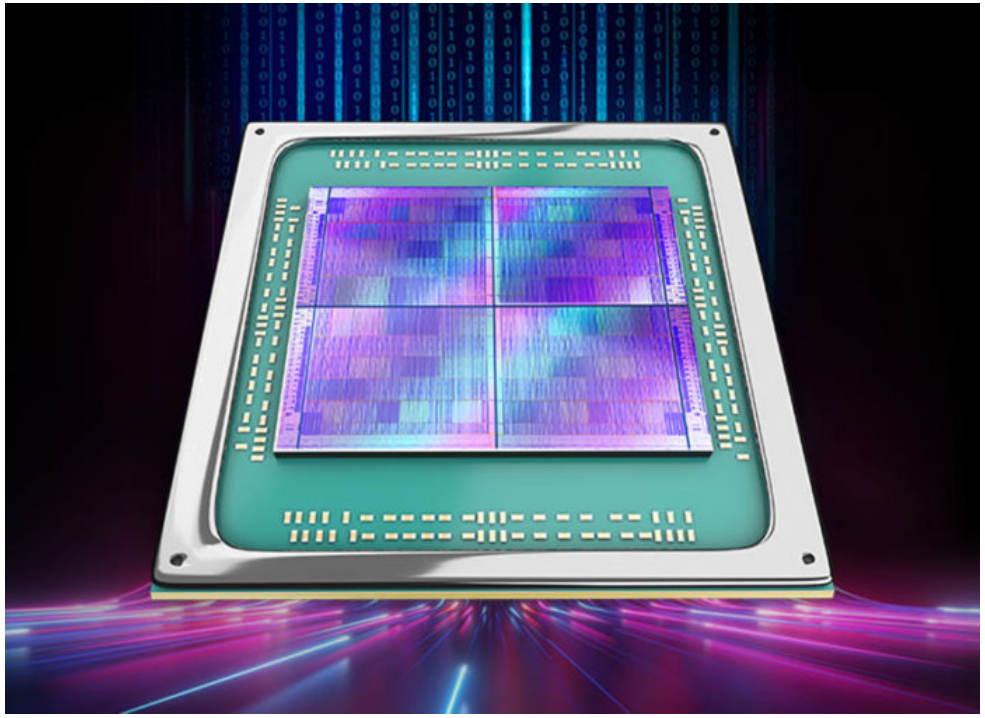


For better manufacturing yield (to save cost), a very large SoC has been sliced into 4 smaller FPGA chips made by TSMC's 28nm process technology.



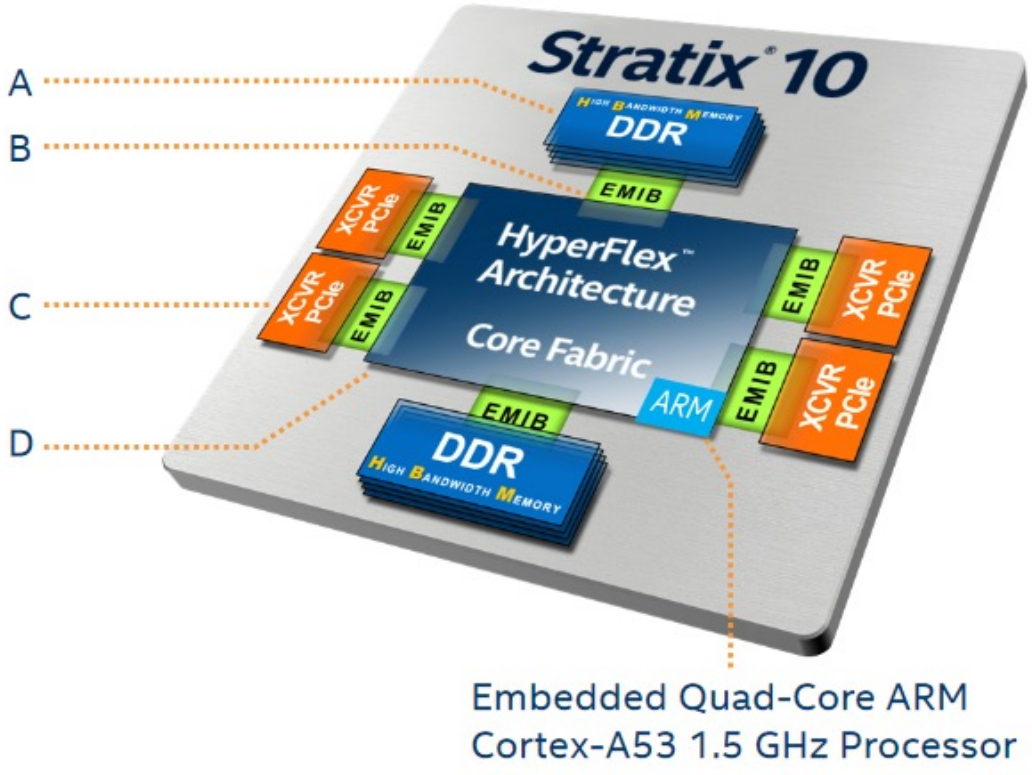
2.5D FPGAs

- Recent AMD VP1902 by TSMC's 7nm process



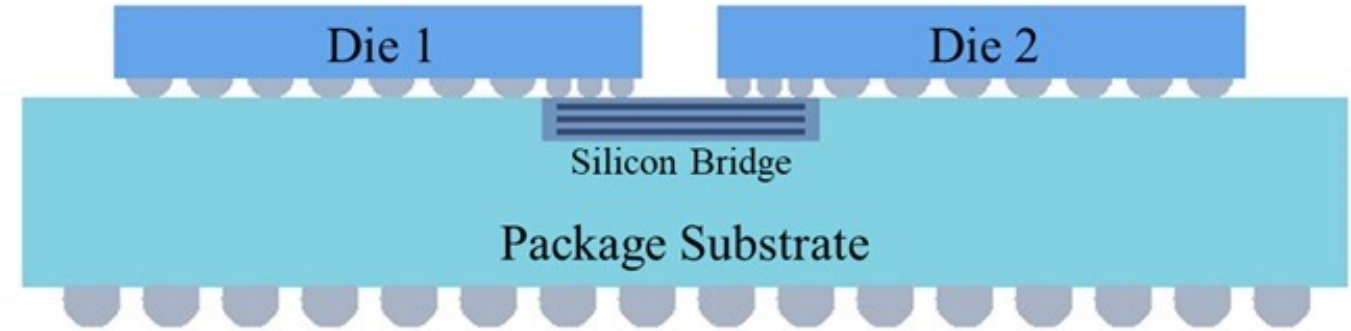
2.5D FPGAs

- Intel EMIB-based 2.5D FPGAs
  - Stratix 10 uses Embedded Multi-Die Interconnect Bridge (EMIB) to connect two adjacent dies



2.5D FPGAs

- Silicon-bridge based connection



Hybrid CPU-FPGA Device

- Hybrid Xeon CPU-Arria 10 FPGA chip

References

- S. M. Trimberger. 2015. “Three ages of FPGAs: A retrospective on the first thirty years of FPGA technology”, *Proc. of IEEE*, 318-331
- E. Nurvitadhi et al. 2017. “Can FPGAs beat GPUs in accelerating next-generation deep neural networks?” in *Proc. of the International Symposium on Field-Programmable Gate Arrays*, 5-14.