University of Chinese Academy of Sciences

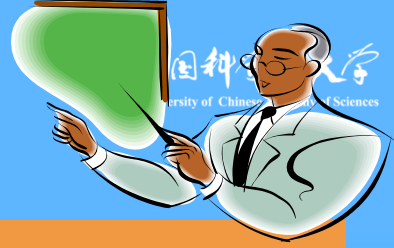
Types of IC Design, CPU and SoC

ASIC, SoC, FPGA

Chun-Zhang Chen, Ph.D.

June 28 - July 2, 2021

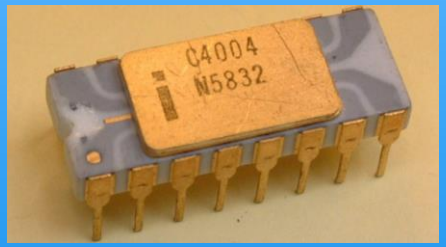
Types of IC



- CPU, MPU and DSP
- GPU and DPU
- ASIC and SoC
- FPGA and GA
- Summary

The First Microprocessor

- 1971: **Intel 4004** invented by Federico Faggin, Ted Hoff, and Stanley Mazor. (U.S. Patent #3,821,715)



- 1987: Intel's first money-making product was the **3101** Schottky bipolar 64-bit static random-access memory (SRAM) chip.

CPU Variants

- CPU | {DSP, MCU, MPU}
  - CPU is a processor; MPU
  - X86, MIPS, ARM, RISC-V
- GPU and APU
  - APU, GPU, NPU, DPU, IPU, VPU
- MCU and ECU
- FPGA

CPU Family

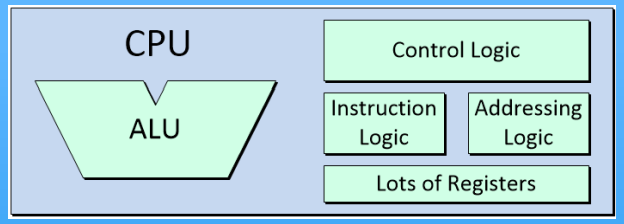
- *Obsolete or less known today*
  - IBM 701 (1953), CDC 6000 (1963), IBM 360 (1964), DEC PDP-8 (1965)
  - Motorola 6800 (1974), DEC VAX (1977), Sparc (1987), Alpha (1992), HP/Intel IA-64 (2001), AMD64/EMT64 (2003)
- X86 (Intel and AMD)
  - 8008 (1972), 8086 (1978), 80286 (1982), 80386/Am386 (1985/1991), 80486/Am486 (1989/1993), Pentium (1993), AMD K5 (1996), AMD K6 (1997), Celeron/Xeon (1998), Itanium IA-32 (2001), Core i7 (2008), Core i5 (2009), Core i3 (2010), AMD FX/APU A (2011)
- MIPS (MIPS, 1984)
  - R2000 (1985), MIPS I/ II/ III/ IV/ 32/ 64
- POWER1-POWER7 (IBM, 1990-)
- **ARM Series**: A*, R*, M* (ARM Holdings, 1990)
- **RISC-V** (2010)

RISC Family

- ARM [→ Nvidia]
  - Processors: Cortex-A710, -A510, -A78, … 45 more
  - Multimedia: Mali-G710, -G610, -G510 … 13 more
  - Physical IP
  - 2004, slogan PPA (A, C, T)
- RISC-V [Chisel]
  - SiFive [Intel]
  - Nuclei, Codasip, Cortus ..

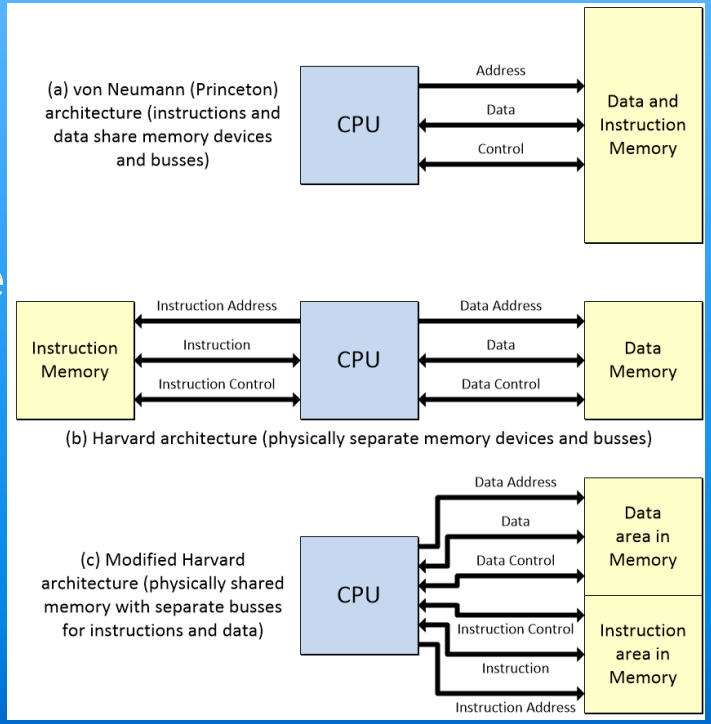
CPU, a processor

- ALU and CPU



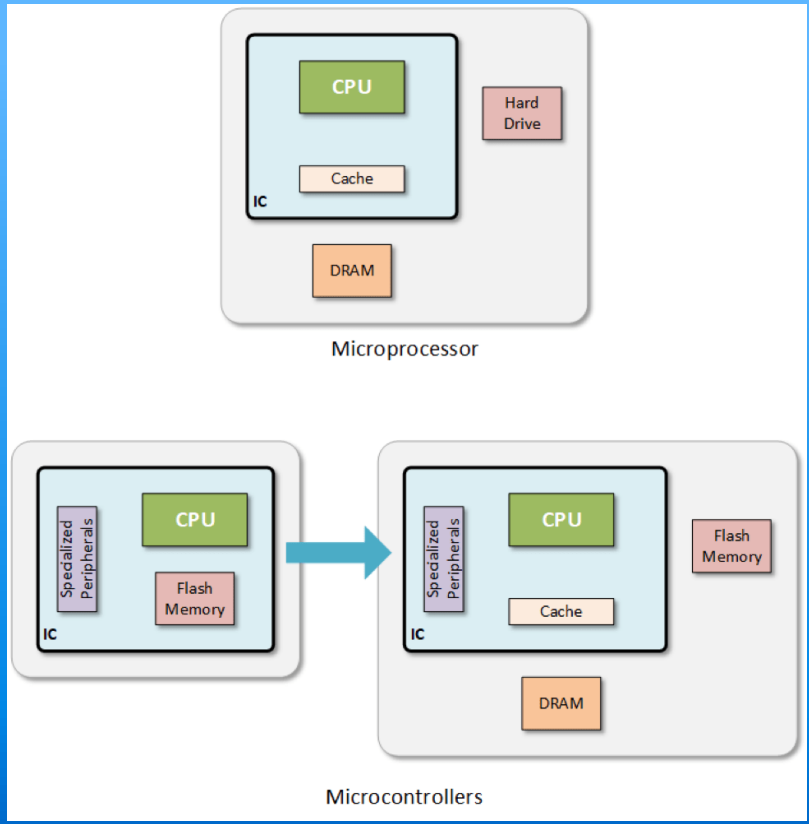
- CPU Architecture
  - a) von Neumann (Princeton) architecture
  - b) Harvard architecture
  - c) Modified Harvard architecture

(Image source: Max Maxfeld)



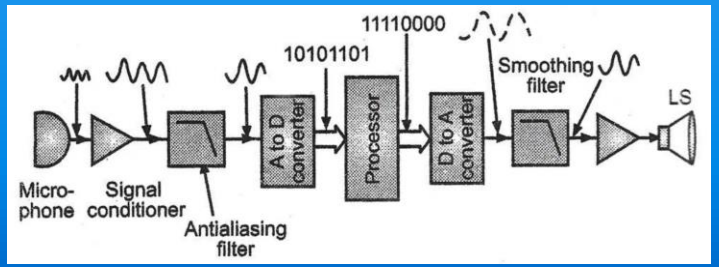
CPU, MCU and MPU

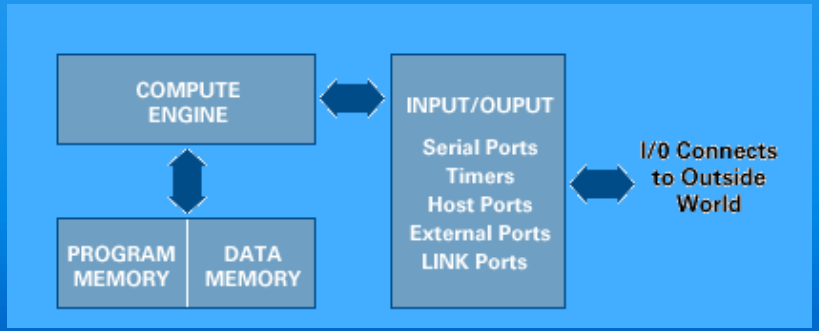
- CPU, processor
- MPU, microprocessor
- MCU, microcontroller



What is a DSP?

- What is a DSP?





DSP before 4G

- DSP 20 years ago
  - SIMD data computation, for single- | dual-MAC
  - Limited in Performance and Complex in programming (assembly-level); limited ISA
  - Used in signal proc./voice
- DSP in 3G
  - 3G modem system needs dual- | quad-MAC
  - Filter for FFT & IIR
  - VLIW introduced for multiple/ parallel comp. (C-compiler was advancing)
  - Complex data (I &Q) required; 50% signal were digital + 50% control & non-vector (scalar)

DSP in 4G (LTE-Advanced)

- DSP in 3.9G/4G
  - *(DSP were well suited for 3G voice/audio)*
  - Wireless MIMO
  - SW/Compiler for complex vector data (type)
  - Dual load store archi., Load, Load, Execution, and Store, hence 4-issue VLIW for DSP

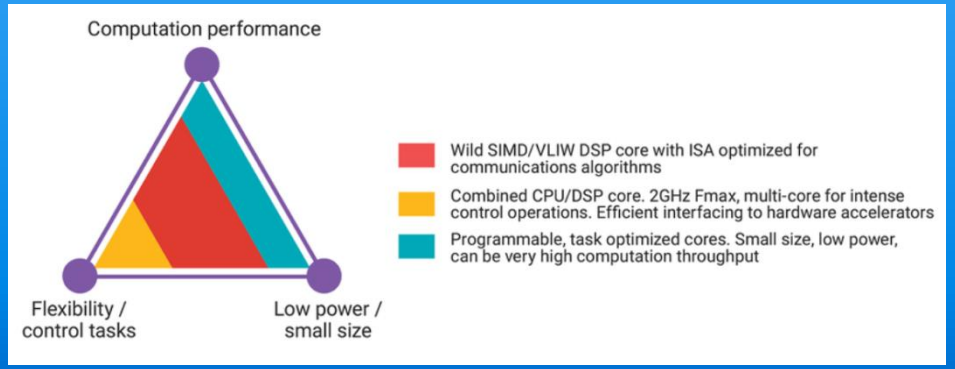
vector operations, thus →

- DSPs integrated vector and scalar engines
- Multiple cores, 16-bit data, 10x comp. → floating-point support
- 4G and Higher …

DSP in 4G and Higher

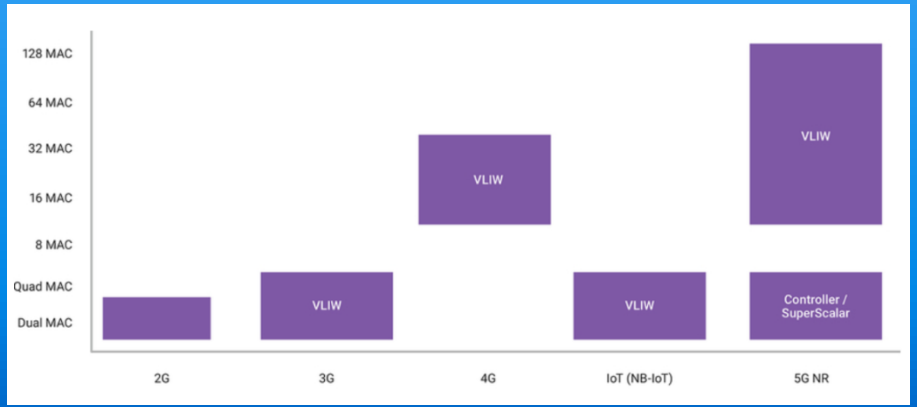
- 4G and Higher …
  - data-intensive tasks, such as front-end FFT, turbo codec, and channel estimation
  - Heterogeneous architecture
  - Low-cost, low-power requirements, and the ability to implement multiple DSPs in one DSP core

- These new types of DSPs have an architecture and ISA that bring together audio, wireless communications, control, and image/motion detection algorithms.

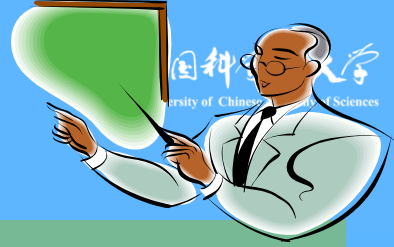


DSP in 5G

- 5G Complexity
  - Further 10x-20x comp.
  - Massive MIMO with high-precision floating-point comp.
  - DSP extensions supporting single-, dual- and quad-MAC level DSP computation for noncritical DSP functions
  - A superscalar architecture is better suited than VLIW for a processor that targets controller algorithms as the code is control, with branching, interrupts, and exceptions.



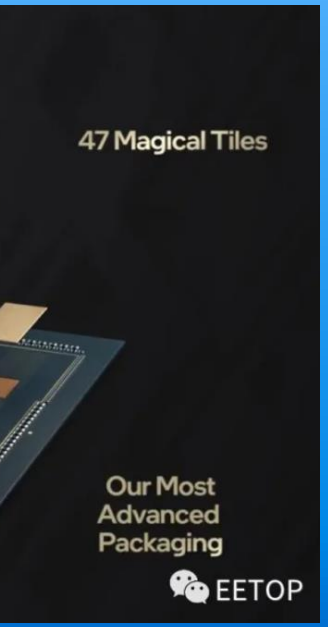
Types of IC

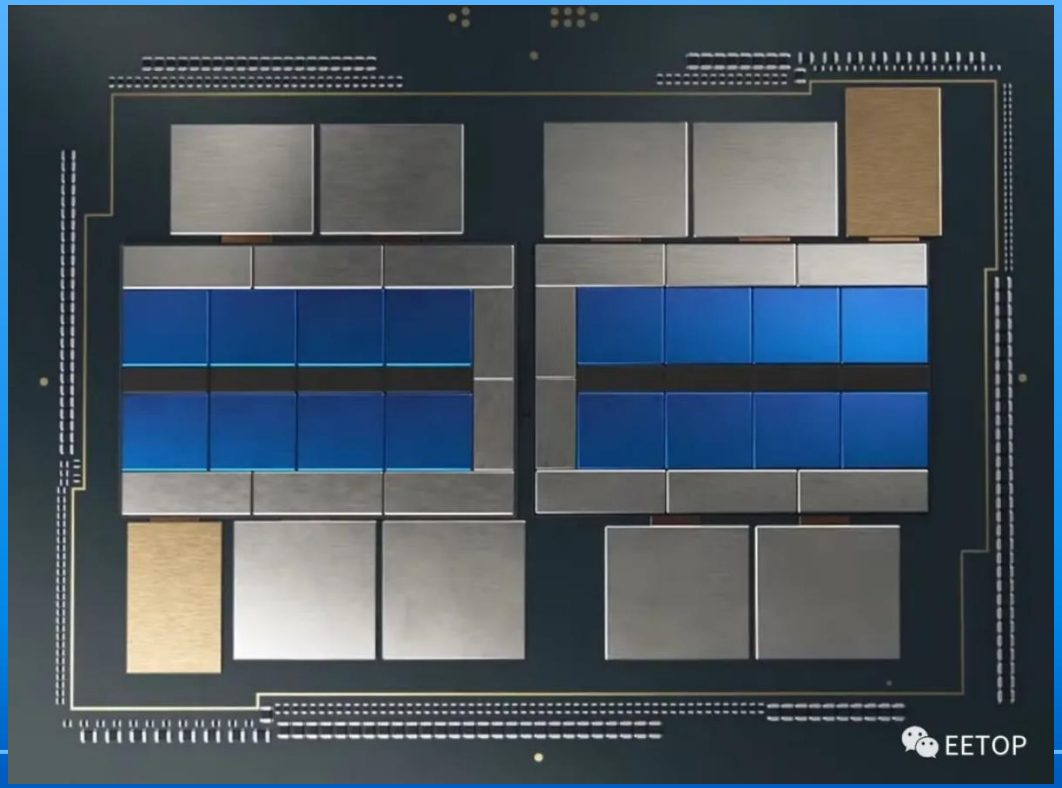


- CPU, MPU and DSP
- GPU and DPU
- ASIC and SoC
- FPGA and GA
- Summary

Xe-HPC GPU from Intel

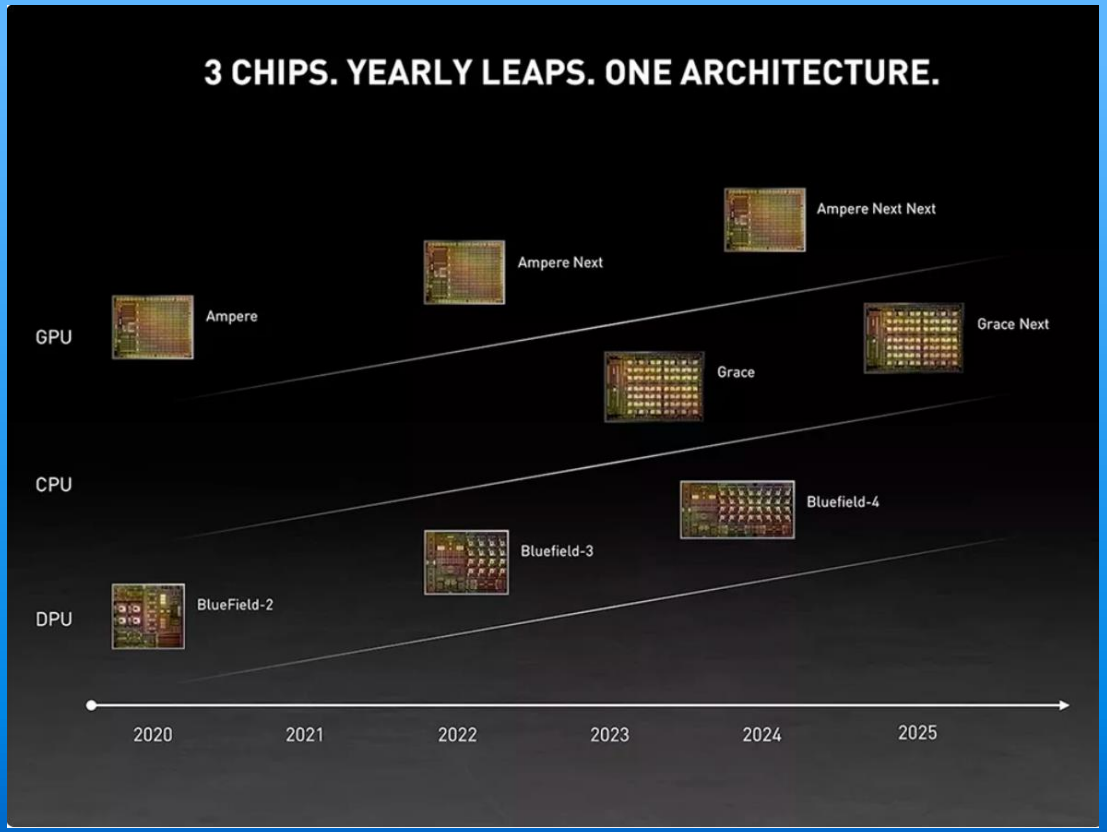
- The largest GPU 7nm, 100B xtors! Code name “Ponte Vecchio”





Nvidia CPU

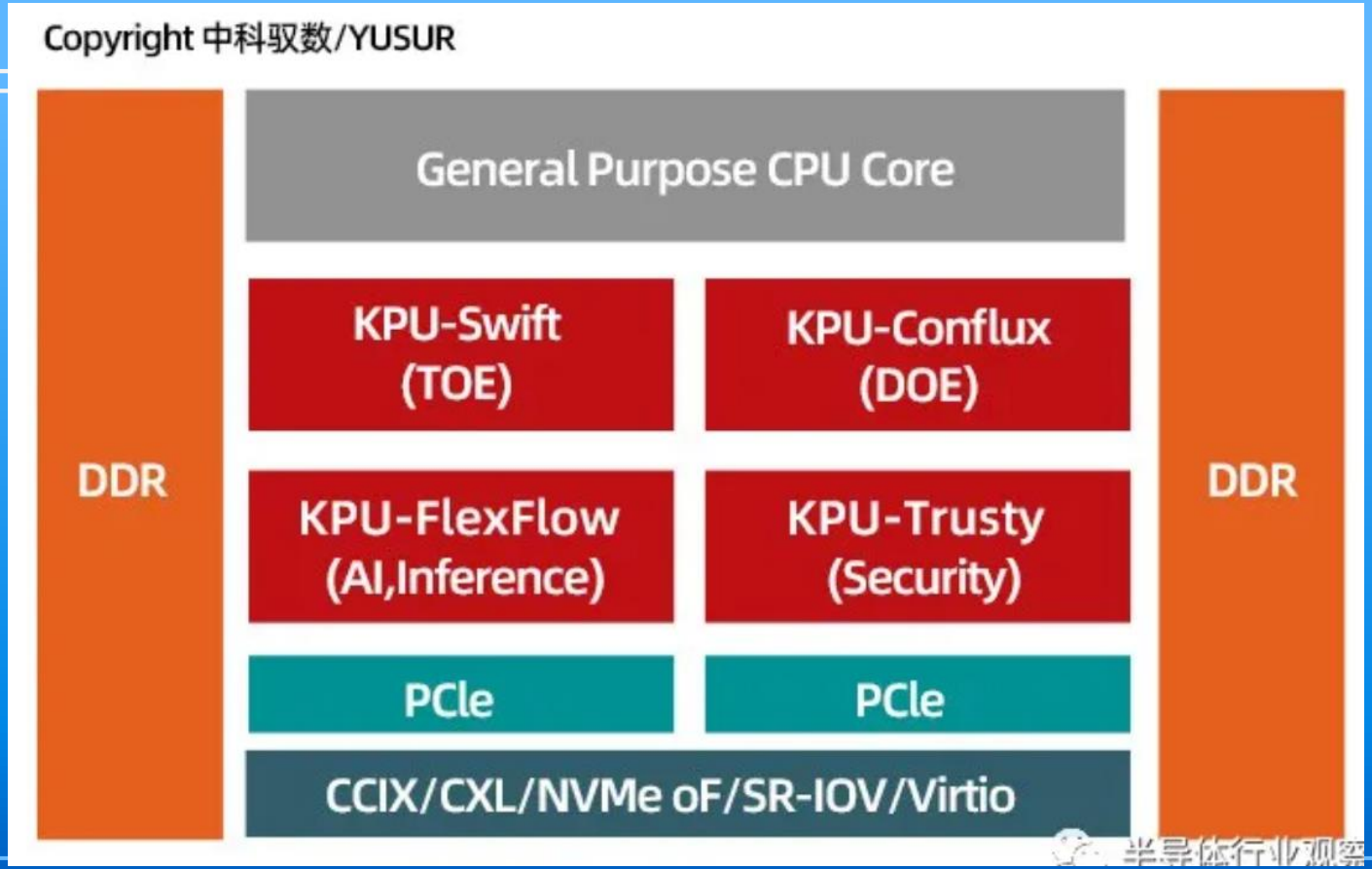

ROCESSING
nsistors


3 CHIPS. YEARLY LEAPS. ONE ARCHITECTURE.
GPU
CPU
DPU
Ampere
Ampere Next
Ampere Next Next
Grace
Grace Next
BlueField-2
Bluefield-3
Bluefield-4
2020
2021
2022
2023
2024
2025


1000
800
600
400
200
0
TOPS
Parker
2018
1 TOPS
7 SPECint

DPU

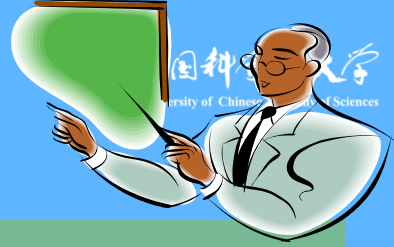
- DP


Copyright 中科驭数/YUSUR
DDR
General Purpose CPU Core
KPU-Swift
(TOE)
KPU-Conflux
(DOE)
KPU-FlexFlow
(AI,Inference)
KPU-Trusty
(Security)
PCIe
PCIe
CCIX/CXL/NVMe oF/SR-IOV/Virtio
DDR

IPU from Intel

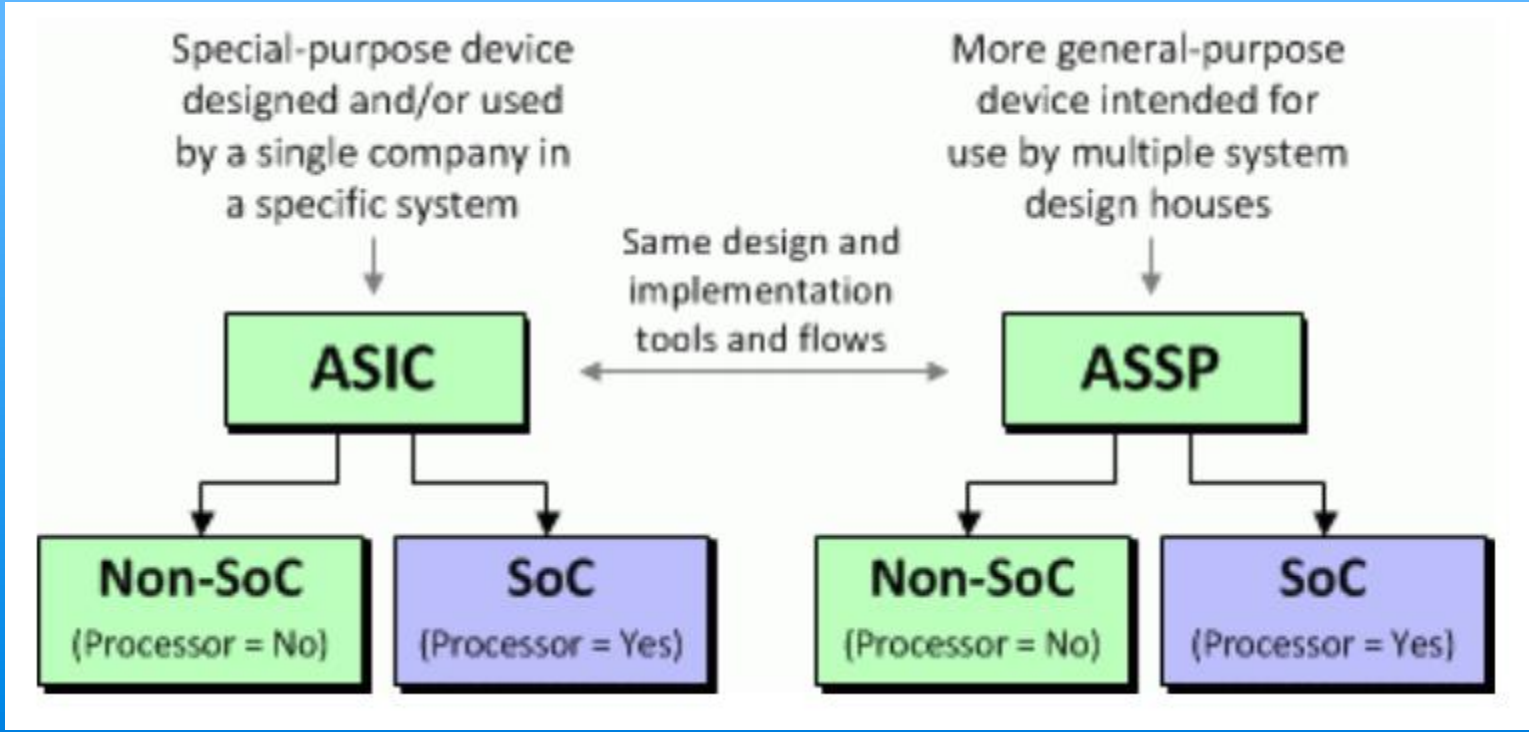
- Services: High-speed, low latency, and a secure networking infrastructure
- IPUs accelerate network infrastructure, freeing up CPU cores … improving data center utilization

Types of IC



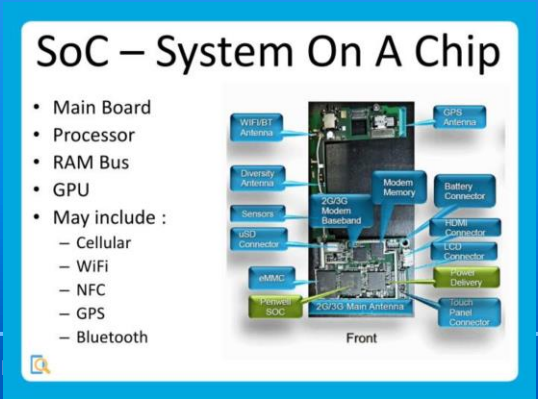
- CPU, MPU and DSP
- GPU and DPU
- **ASIC and SoC**
- FPGA and GA
- Summary

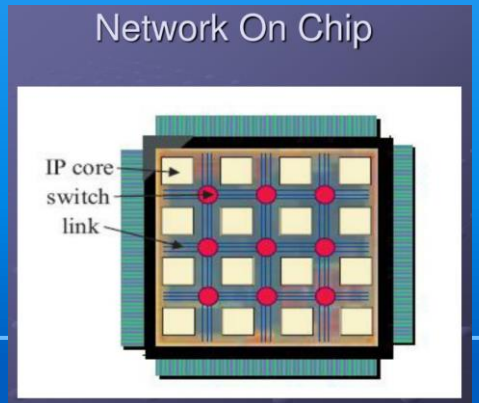
ASIC and ASSP vs SoC



From SoC to NoC

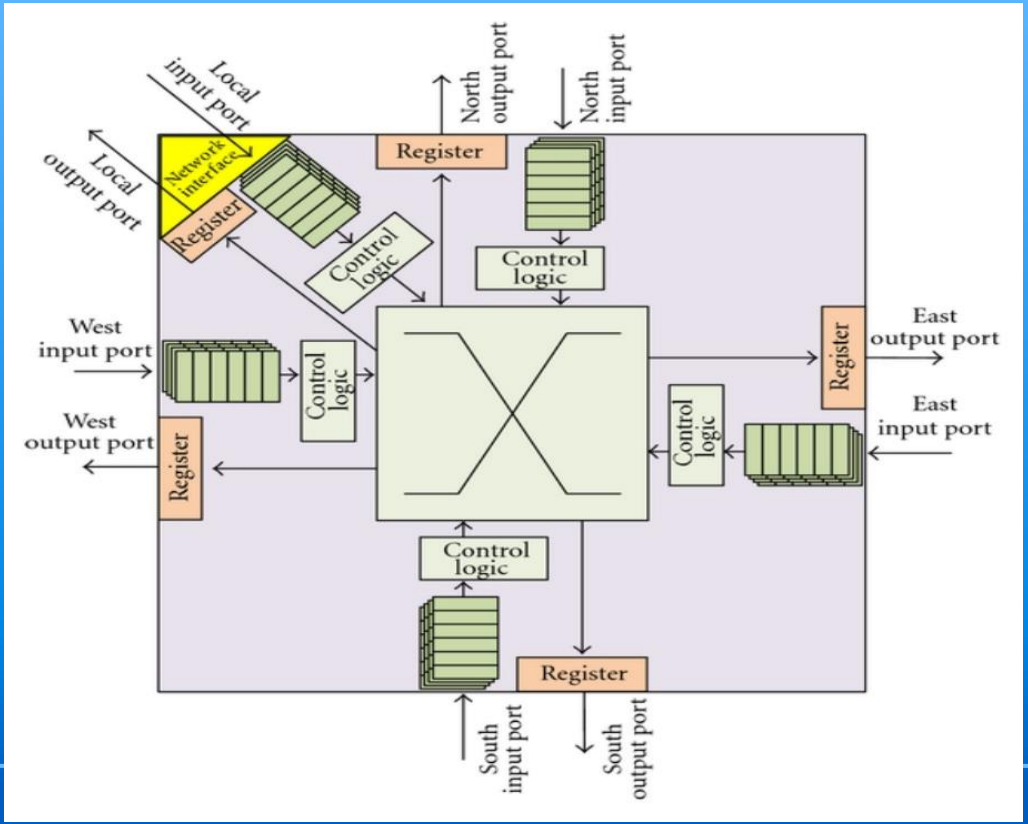
- SoC, system-on-chip (system-on-a-chip)
  - 2001, ARM-based SoC (used in iPod)
  - 2007, iPhone (ARM core + GPU)
- NoC, network-on-chip
  - Fig: Typical NoC architecture in a mesh topology.





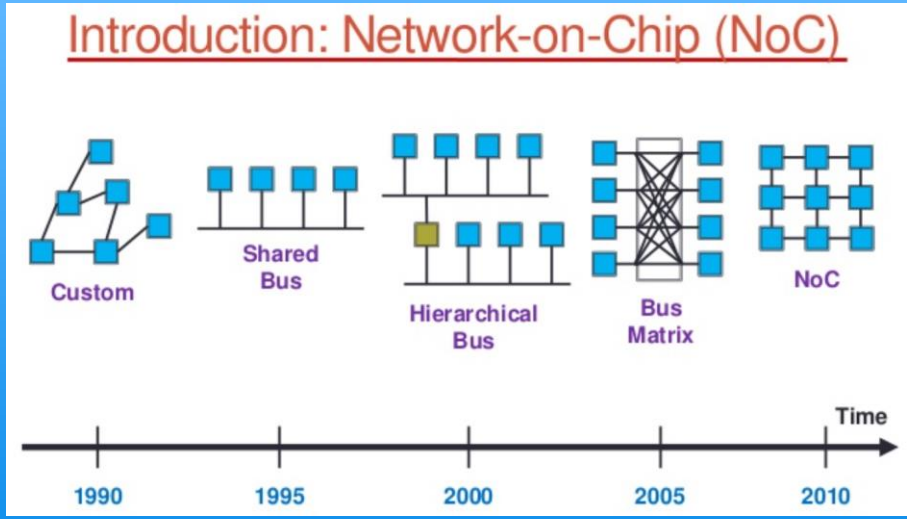
NoC Router Architecture

- Typical NoC router architecture



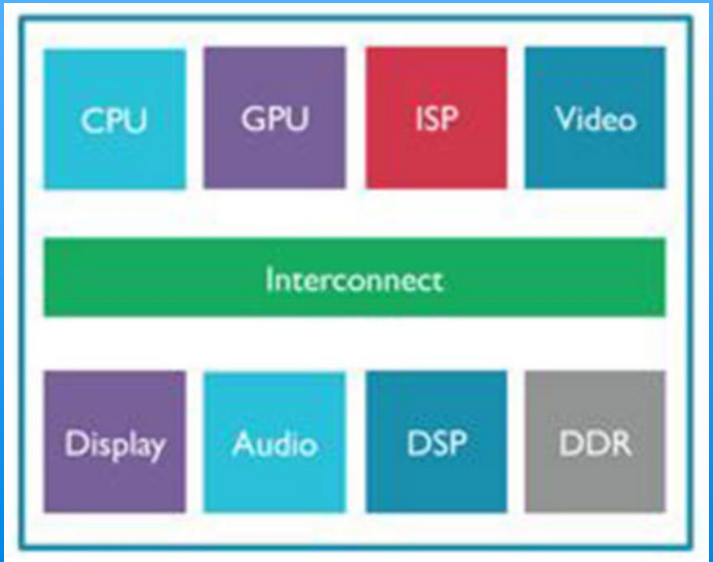
NoC vs SoC

- **NoC Advantages**
  - Higher Operation Frequencies
  - Reduced Wiring Congestions
  - Change IP at an Extreme Ease
  - Ease Timing Closure
- Difference in Design Flows

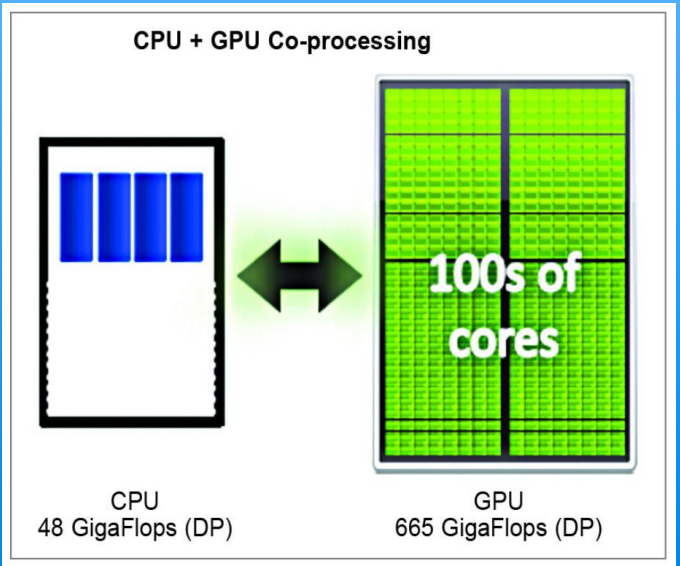


SoC and HSA

- *Heterogeneous multiprocessing; CPU + GPU*

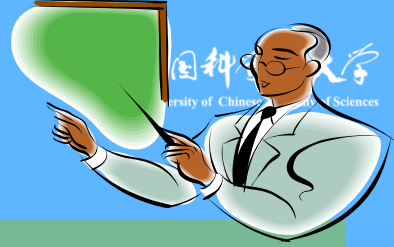


A generic heterogeneous multiprocessing (HMP) compute system



Heterogeneous computing with GPUs

Types of IC



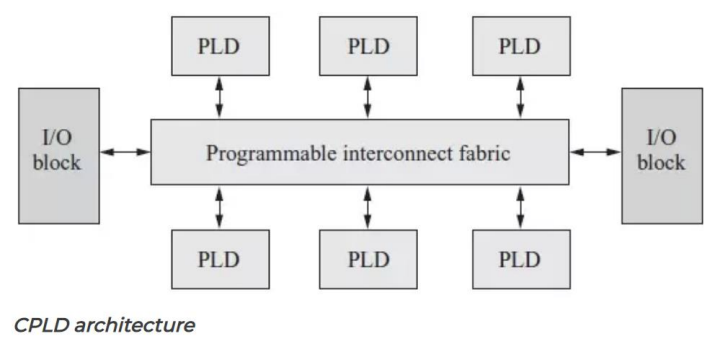
- CPU, MPU and DSP
- GPU and DPU
- ASIC and SoC
- FPGA and GA
- Summary

Programmable Logic

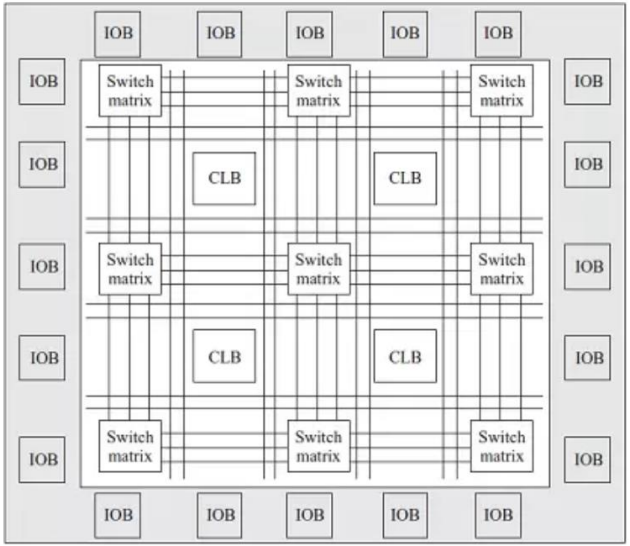
- PAL
  - The first by PAL: PLD
- Types of PLD
  - GAL
  - CPLD
  - FPGA
- EPLD
- PSoC

CPLD and FPGA

- Complex Programmable Logic Device (CPLD)
  - CPLD is PLD based EPROM or EEPROM
- Gate array (GA) and FPGA development board

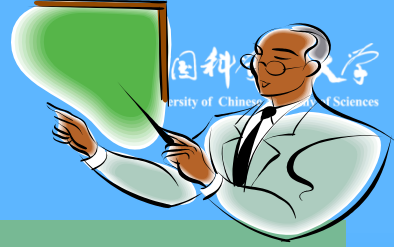


CPLD architecture



FPGA architecture

Types of IC



- CPU, MPU and DSP
- GPU and DPU
- ASIC and SoC
- FPGA and GA
- Summary

Nvidia and Intel

- NVIDIA announced its own CPU
  - 3 Chips, Yearly Leaps, One Architecture!
  - CPU + GPU + DPU
- Intel DNN w/ Metrix, research paper said
  - CPU is 4-15x faster than GPU

CPU, GPU and FPGA

- Intel: CPU, GPU, FPGA (Altera)
- AMD/ATI: CPU, APU, FPGA (Xilinx)
- Nvidia: CPU (ARM), GPU, FPGA
- RoW: VPU, NPU, DPU…