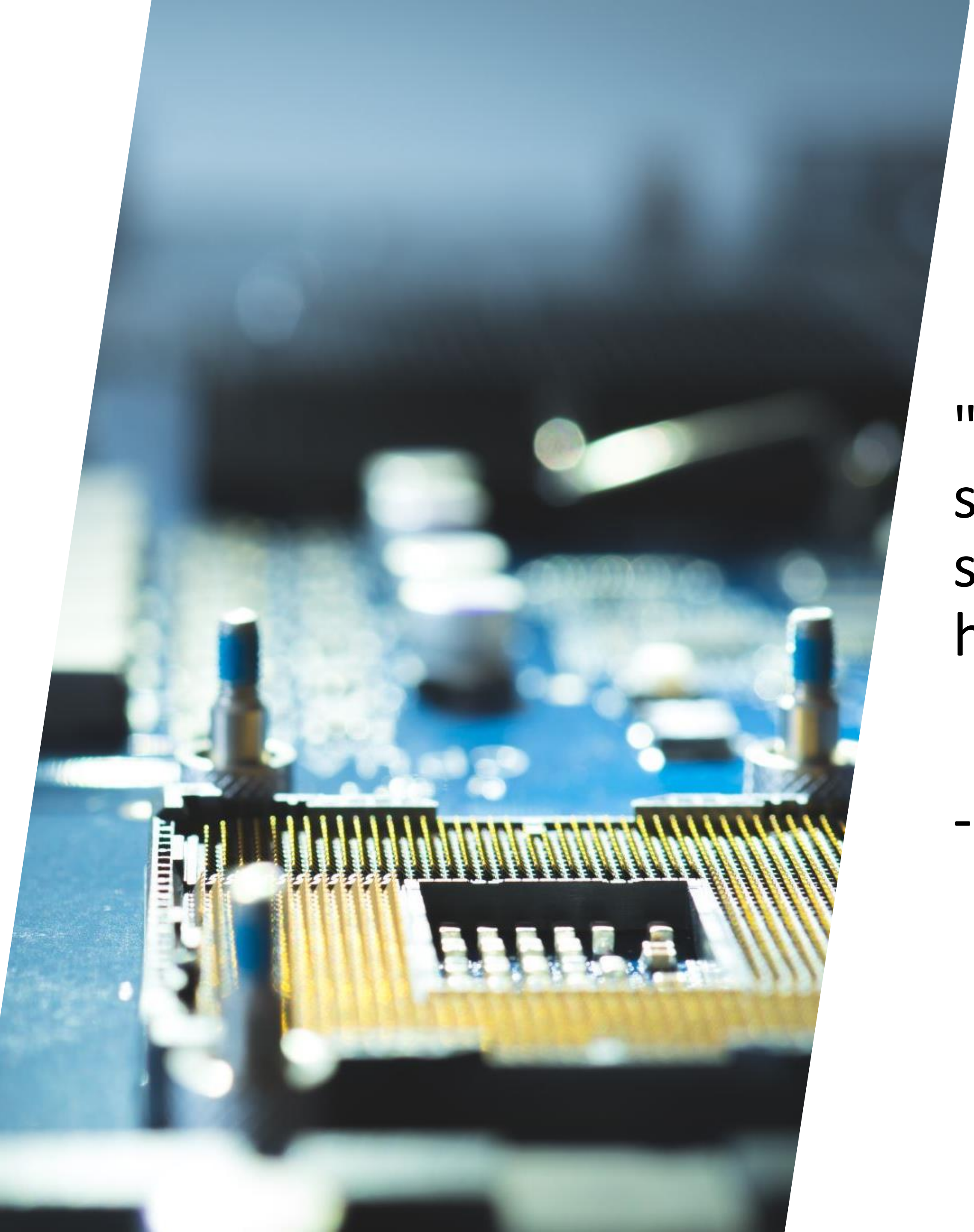
From System Software to System Hardware



"People who are really serious about software should make their own hardware."

- Alan Kay

Content

- System Software: Computer Engineers as Software Developers
- Cloud Computing, IoT, and Big Data Engineering
- Artificial Intelligence and Image Processing
- Embedded Systems, Architecture, Microprocessors, and VLSI

System Software

Computer Engineers As Software Developers

Strong mathematical skills

Understanding of software design, debugging and development cycles

Proficiency in programming languages (i.e., C++, C#)

Experience with hardware optimizations

Potential Jobs

AR/VR SOFTWARE ENGINEERS

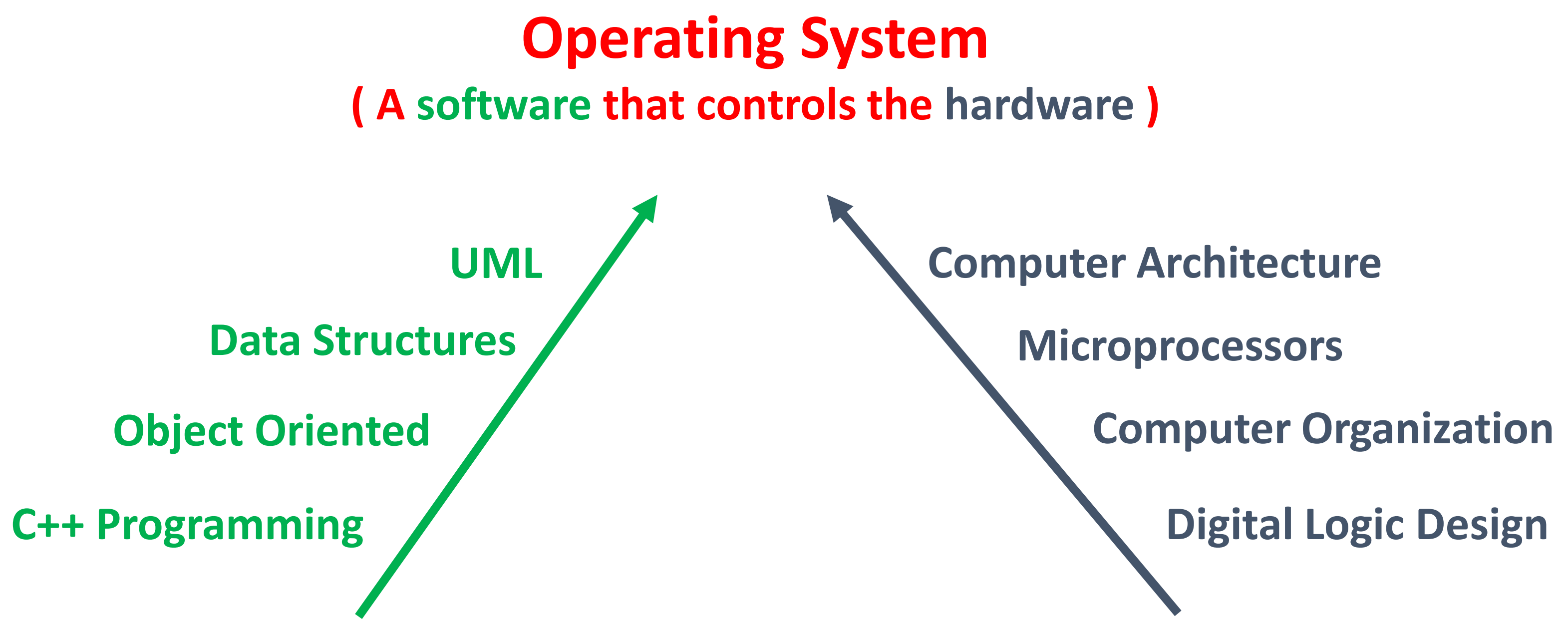
MACHINE LEARNING SOFTWARE ENGINEERS

GAMING SOFTWARE ENGINEER

BIG DATA SOFTWARE ENGINEERS

Operating Systems

- It is one of the core courses in any Computer Engineering Program
- Provides deeper understanding of how computers work



Cloud Computing

Internet of Things (IoT)

Big Data Engineering

Definition of Cloud Computing

The standardized definition of Cloud Computing by the National Institute of Standards and Technology (NIST):

- “Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model promotes availability and is composed of five essential characteristics, three service models, and four deployment models.”

Deployment Models of Cloud Computing

- **Public**
  - It's an open model, external cloud
  - Services are provided to public based on an agreement between cloud provider and client
  - Least expensive
- **Private**
  - It's an internal cloud
  - Highest degree of control over security, privacy, and performance
  - Expensive
- **Hybrid**
  - A combination of public and private clouds
  - Needs interoperability and portability between both clouds
- **Community**
  - Created between a group who shares common goals (professional, geographical, etc.)

IoT Applications

- Definition of IoT in Wikipedia:
  - The Internet of things (IoT) describes physical objects (or groups of such objects) with sensors, processing ability, software and other technologies that connect and exchange data with other devices and systems over the Internet or other communications networks. Internet of things has been considered a misnomer because devices do not need to be connected to the public internet, they only need to be connected to a network and be individually addressable.

IoT Applications

- Smart Cities
- Intelligent Transportation Systems
- Inter-connected Cars
- Smart Homes/Buildings
- Industrial Systems
- Agricultures
- Smart Devices
- Environment Monitoring
- Medical and Health Systems

Big Data Engineering

Data engineering is about designing robust/scalable data-processing systems that work with large datasets

Big data engineer designs systems that collect and extract data

Big data scientist analyzes generated data using predictive models to create insights

Cloud computing focuses on scalability, elasticity, on-demand allocation of resources, and pay-per-use self-service models

Cloud computing provides big data with massive on-demand computation power, networking, and distributed storage capacity required for analysis and visualization of big data

Certificates – Cloud Computing

CompTIA Cloud+

Amazon Web Services (AWS) Solutions Architect – Associate

AWS Certified Cloud Practitioner

AWS Certified Developer – Associate

AWS Certified SysOps Administrator – Associate

Microsoft Certified: Azure Fundamentals

Microsoft Certified: Azure Administrator Associate

IBM Certified Technical Advocate - Cloud v3

Cloud Security Alliance: Certificate of Cloud Security Knowledge (CCSK)

Google Associate Cloud Engineer

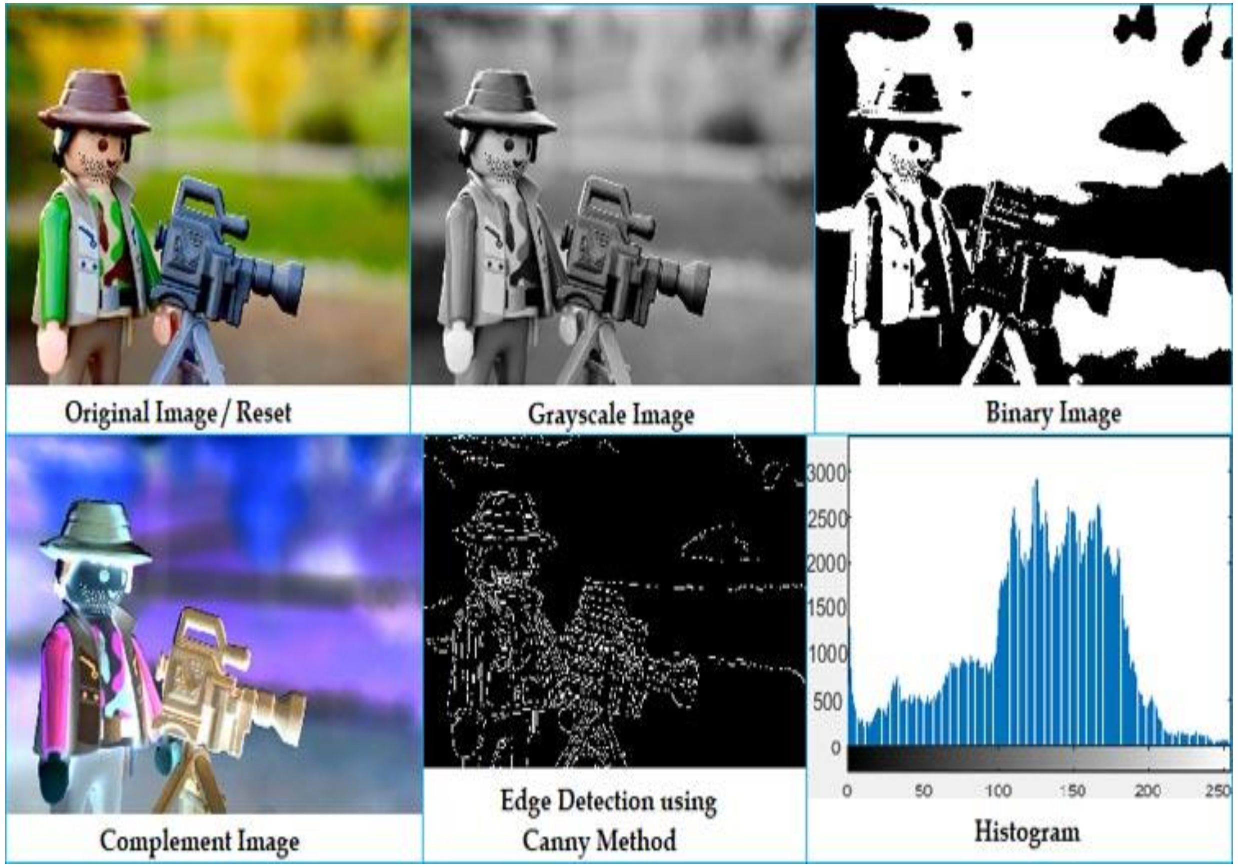
Google Professional Cloud DevOps Engineer

Certified Cloud Security Professional (CCSP)

Introduction to Image Processing

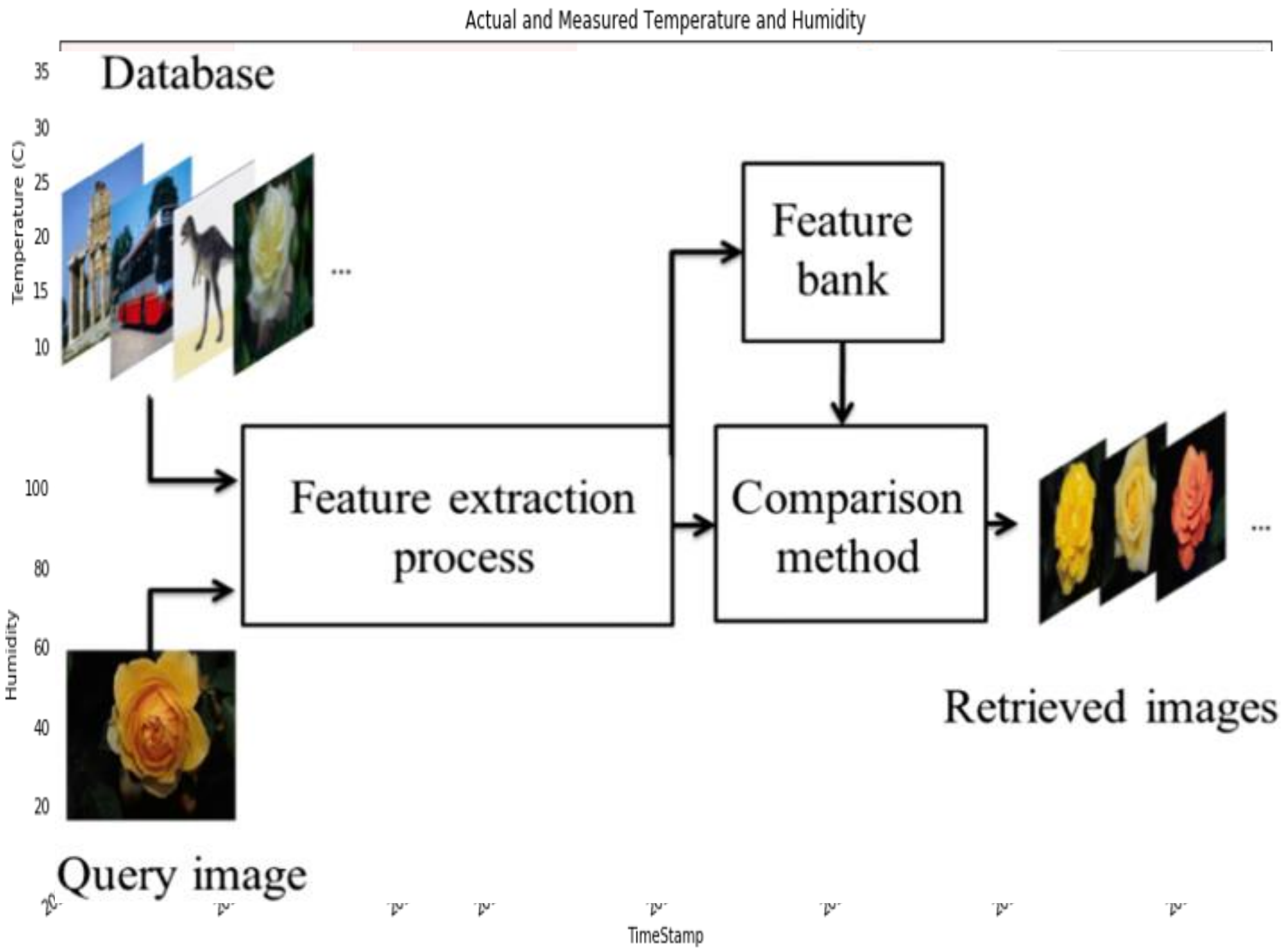
What is Image Processing?

- Image processing is the process of transforming an image into digital form and perform operations to get useful information
- The type of operations depends on what we need to achieve and what we need to enhance in the image



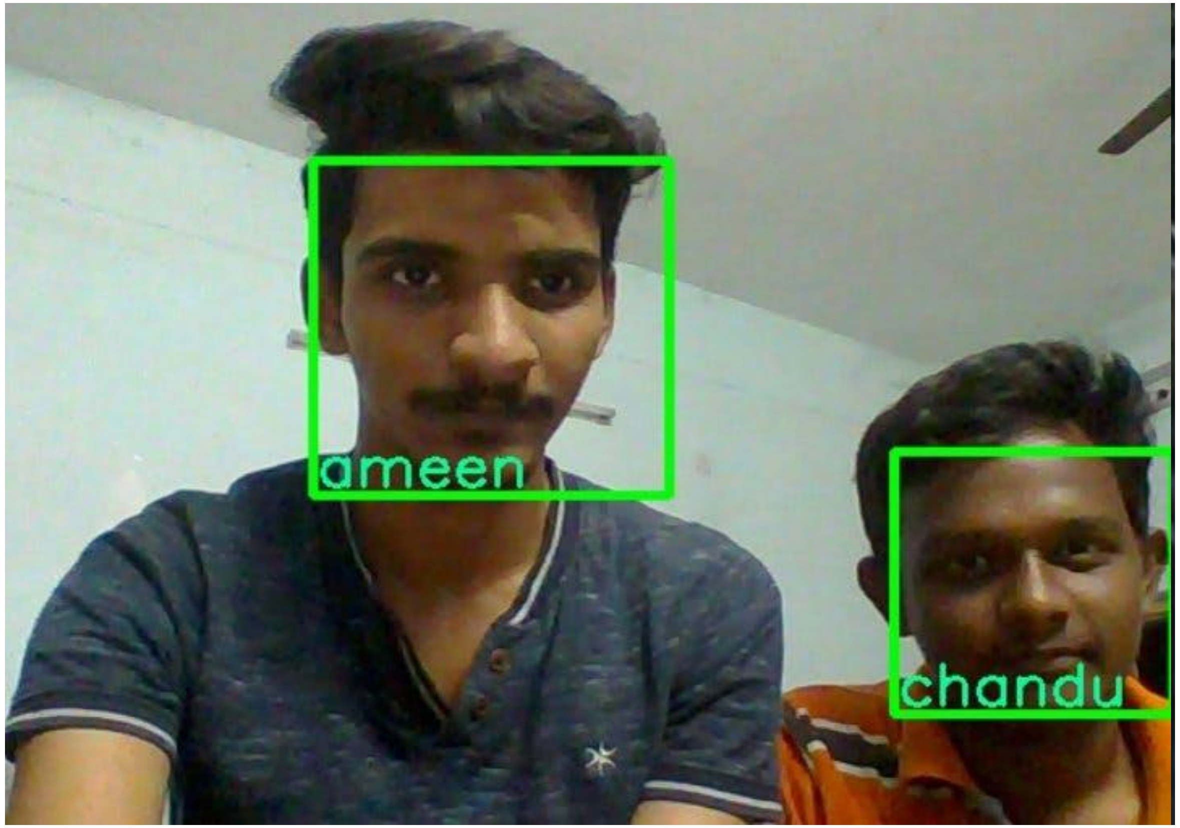
Types of image processing

- Image processing has five main types
  - Visualization: finding hidden objects in an image
  - Recognition
  - Sharpening and restoration
  - Pattern recognition
  - Retrieval



Why do we need it?

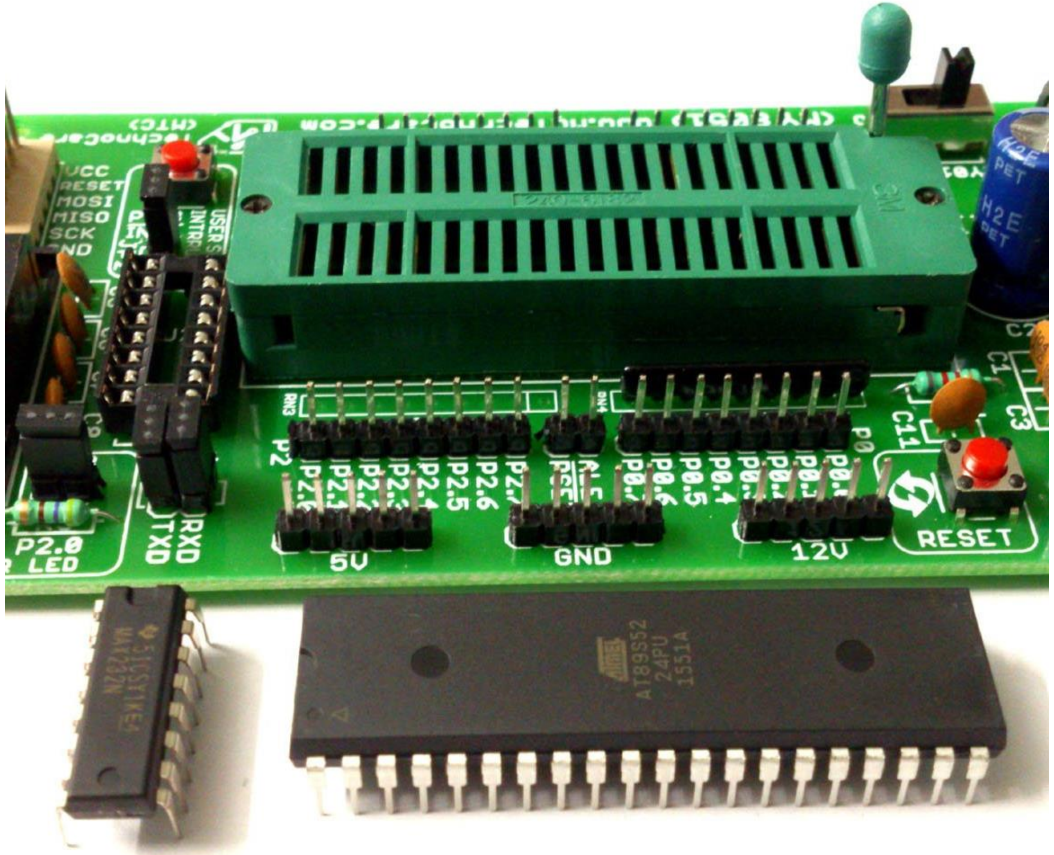
- Medical image enhancement
- Traffic sensing
- Image reconstruction
- Face detection



Embedded Systems

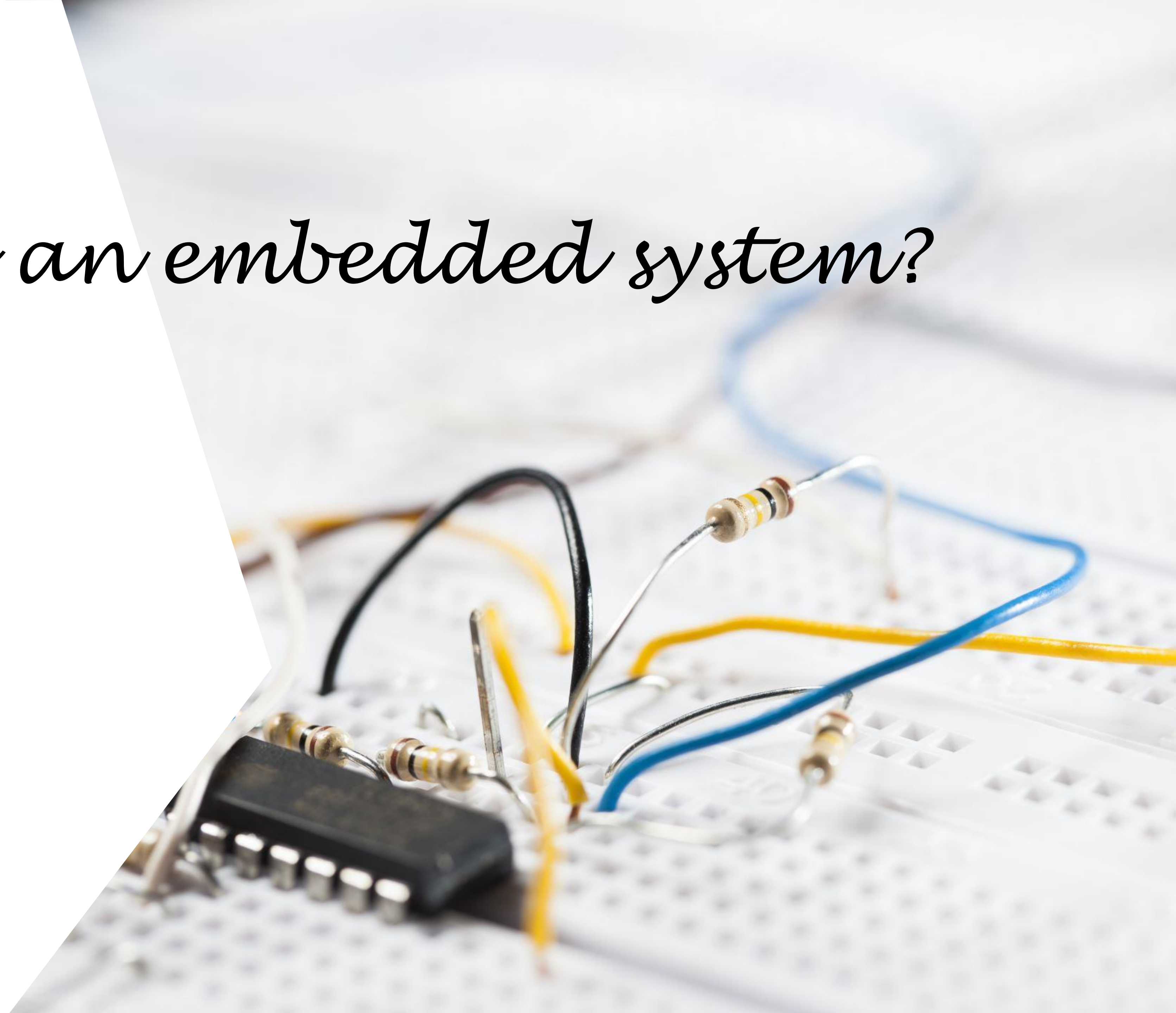
Architecture

Microprocessors



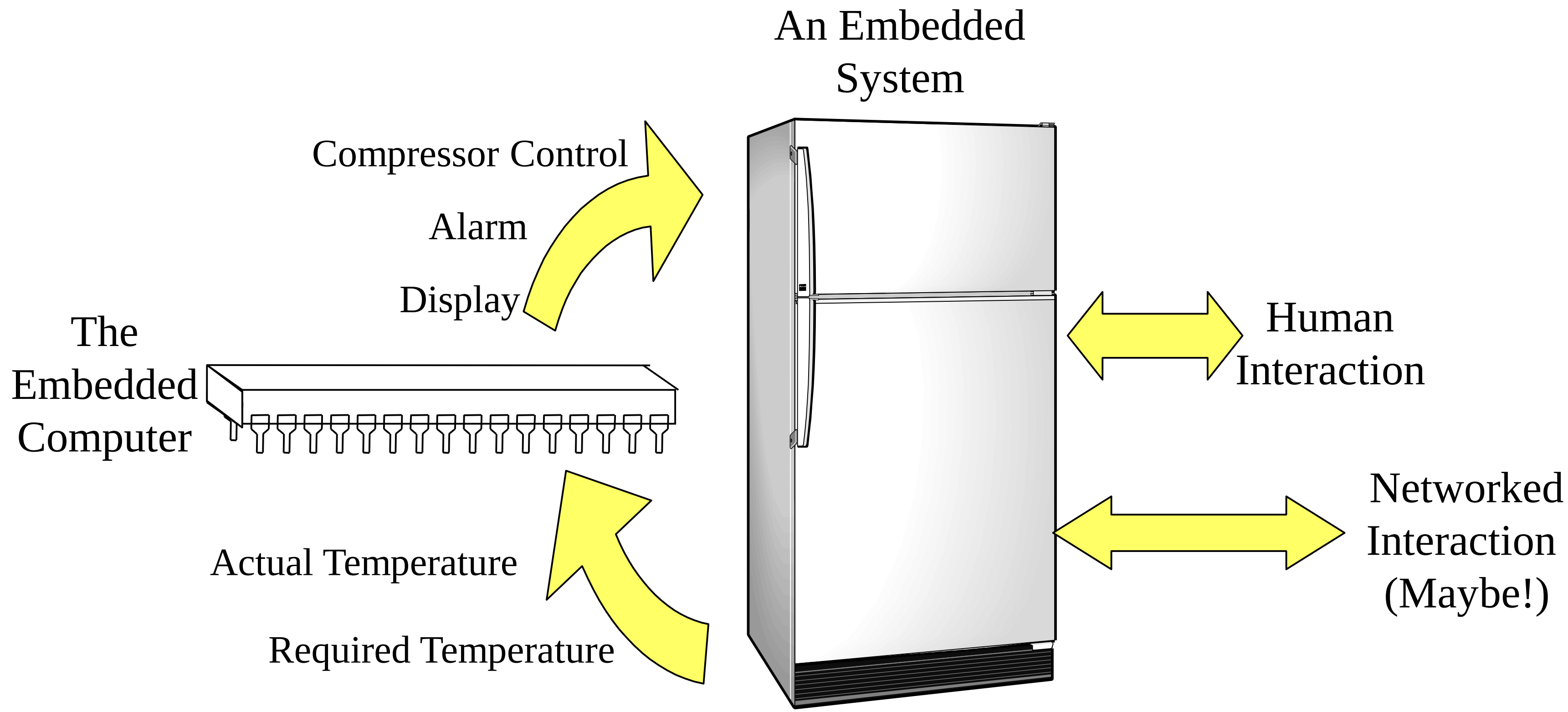
Embedded Systems

What is an embedded system?

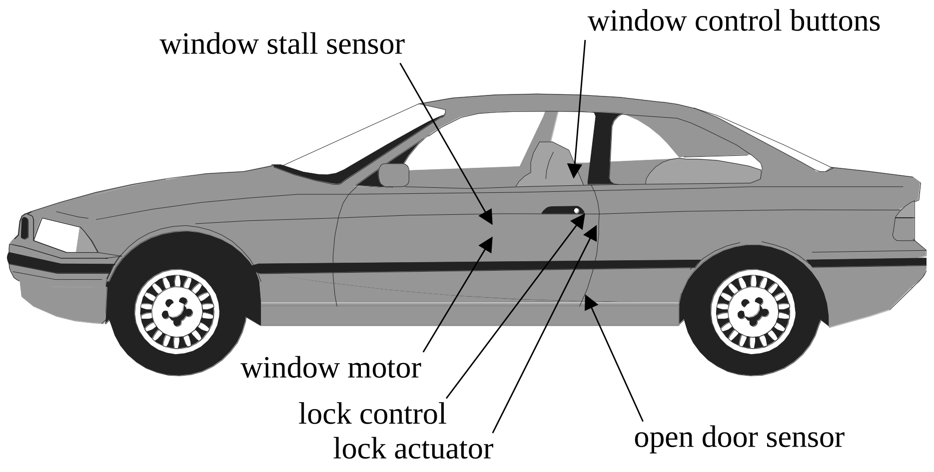


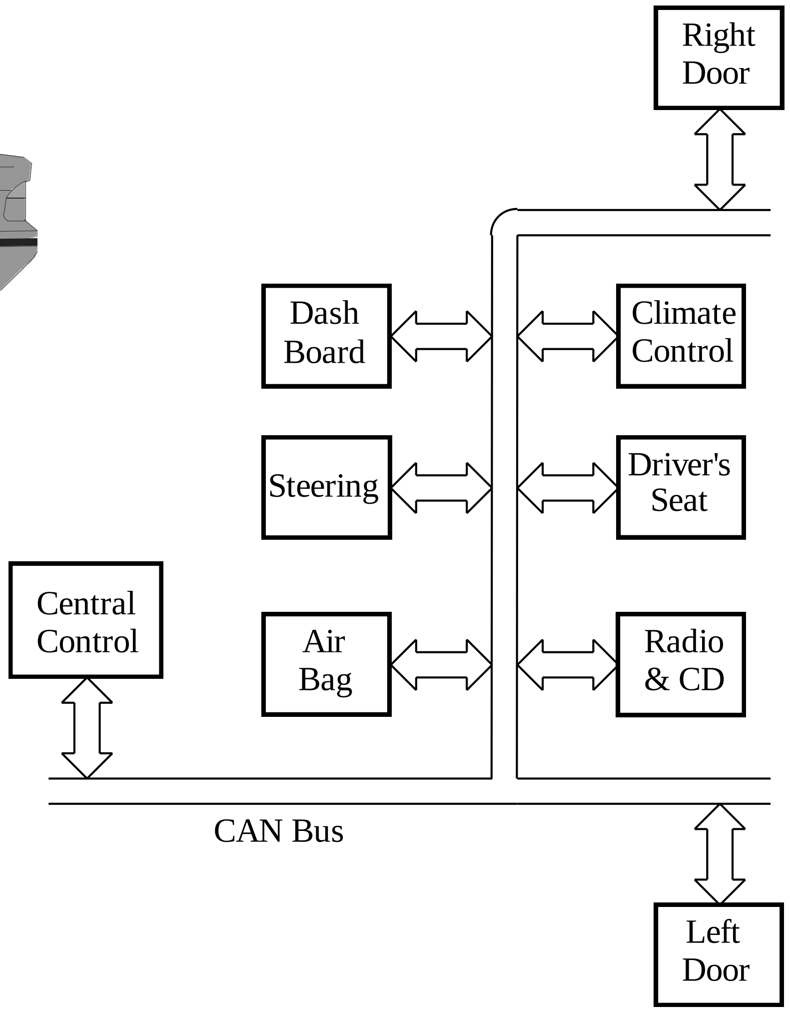
- What examples of embedded
- systems do you know?

Example 1: The Domestic Fridge



Example 2: Car Door Control, within a Larger Network





The Microcontroller

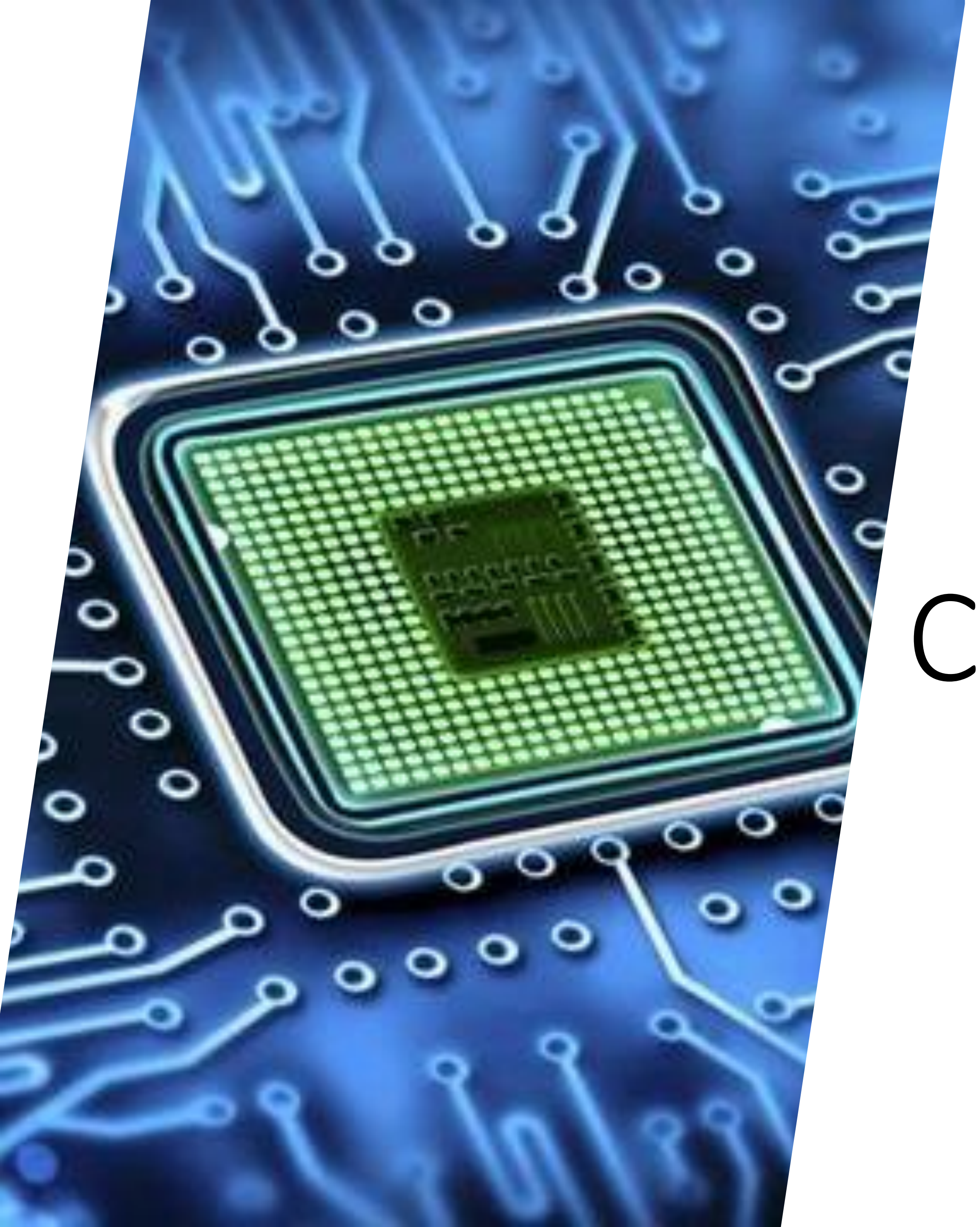
A microcontroller is a microprocessor designed primarily to perform simple control functions.

Microcontrollers usually have these features

- low cost,
- physically small,
- input/output intensive, and capable of easy interfacing,
- limited memory capability for program and data,
- instruction set leading to compact code, with limited arithmetic capability,
- ability to operate in a real-time environment.

In certain applications the following further features are essential:

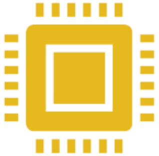
- ability to operate in hostile environment, e.g. high or low temperature, tolerant to electromagnetic interference,
- low power, with features adapted to battery power.



Computer Architecture

Computer Architecture: overview

Computer architecture is a set of rules and methods that describe the functionality, organization, and implementation of computer systems



Design goal: improving performance and power efficiency of computer systems

Topics of interest to computer architects

How programs are translated into the machine language and how the hardware executes them

The hardware/software interface – Instruction Set Architecture (ISA)

What determines program performance and how it can be improved

Techniques used by hardware designers to achieve the design goals

Relation between performance resources

Application software – realization of an algorithm

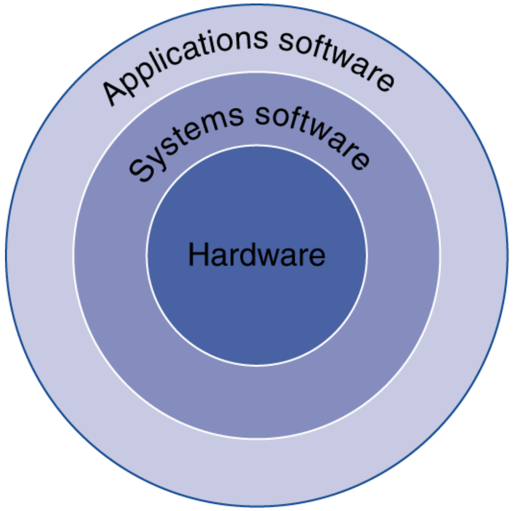
- Written in high-level language (HLL)

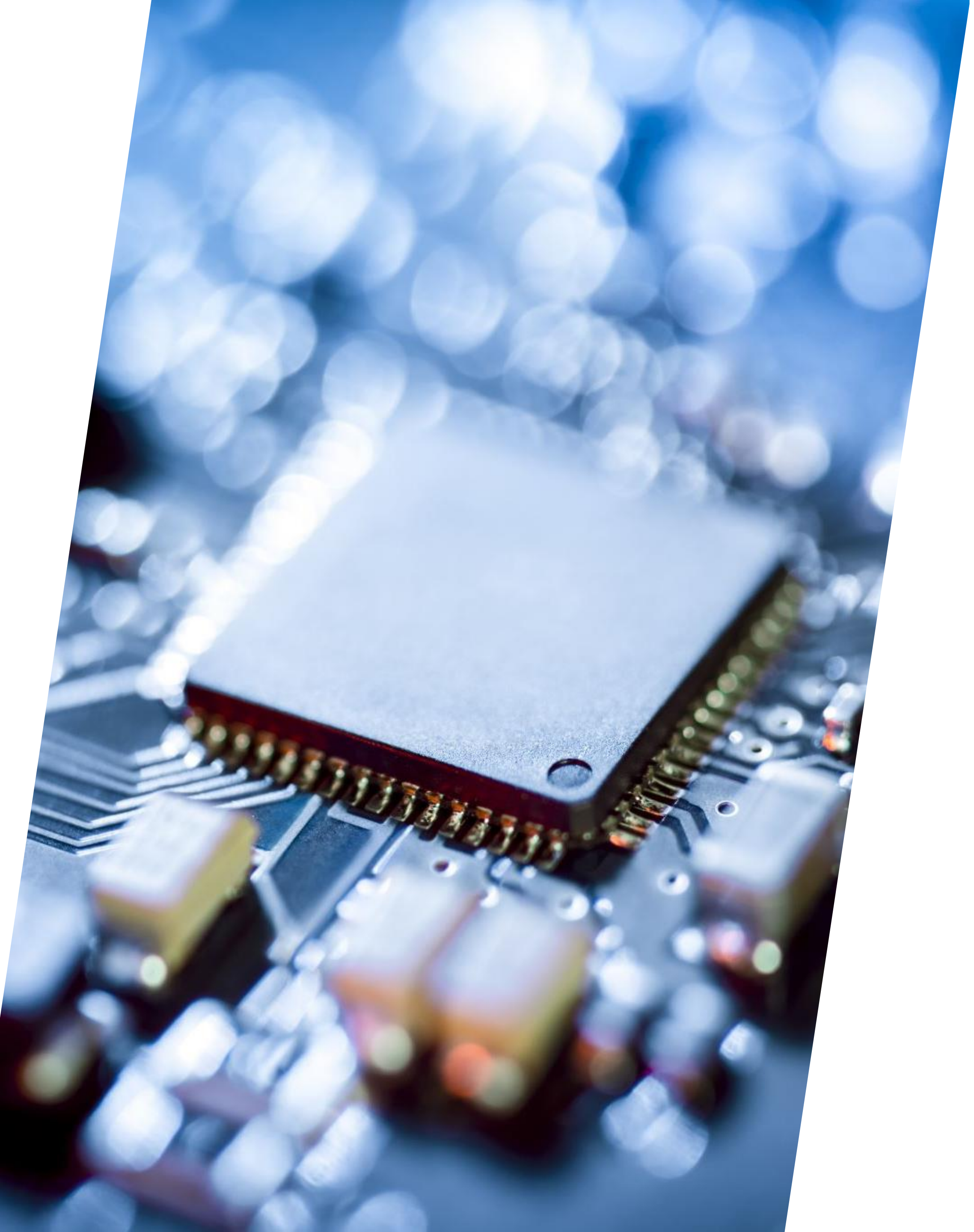
System software

- Compiler: translates HLL code to machine code
- Operating System: service code
  - Handling input/output
  - Managing memory and storage
  - Scheduling tasks & sharing resources

Hardware

- Processor, memory, I/O controllers





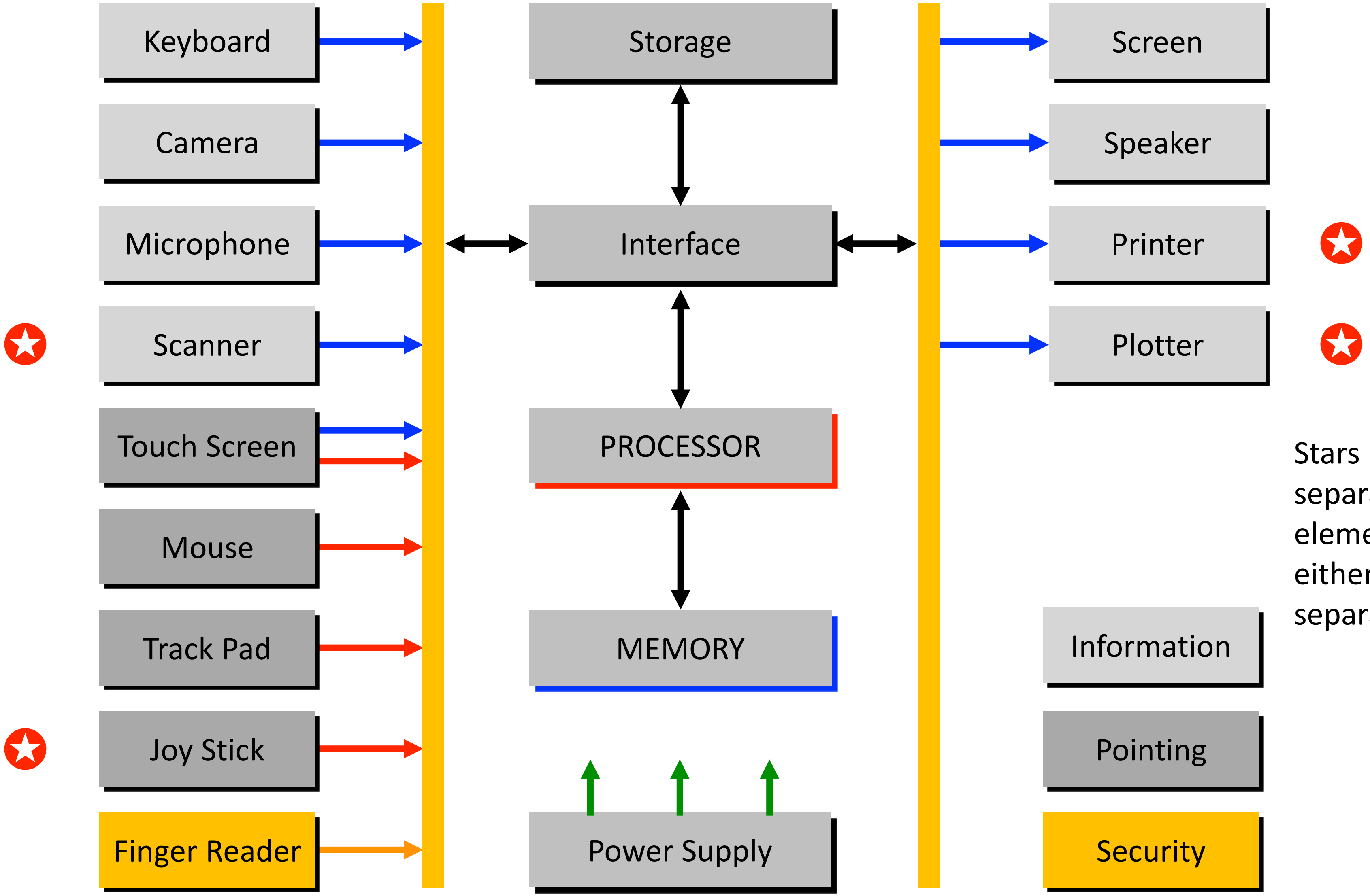
Microprocessors Systems

“Study the past if you would define the future” - Confucius

µP Systems

- Digital systems have a prominent role in our life; social, commercial, industrial, scientific, etc.
- Performance, cost, size, power, etc. vary; depending on what they are designed to do
- Computing platforms like Desktops, Laptops & Tablets represent a fraction of the world's computing; the industrial and embedded platforms are numerous
- On the average, a person in the USA faces computing devices of some sort more than 70 times in a typical day; Security doors, Car, Phone, Tickets, ATM, etc.
- Artificial Intelligence (AI) has a touch on everything today, and this requires sophisticated high performance processors; autonomous vehicles as an example
- Internet of Things (IoT) emerged recently, and everything is now connected
- A typical microprocessor system consist of:
  - **Microprocessor**, and support logic like clock & reset circuits
  - **Storage**, to store code and data
  - **Peripherals**, to interact with the outside world via information exchange
  - **Connections**, to connect all devices for signal transmission
  - **Glue Logic**, to arbitrate operations using decoders, encoders, buffers, etc.
- The hardware runs under the control of low level software (Firmware) that provides services to the operating systems, which in turn serves the end user applications

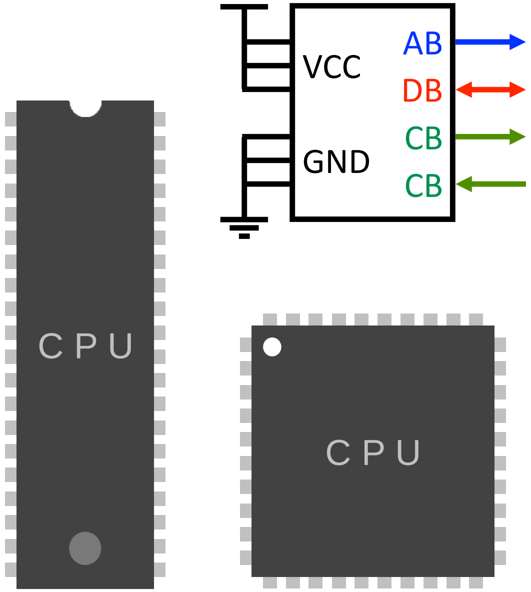
General Purpose Computer



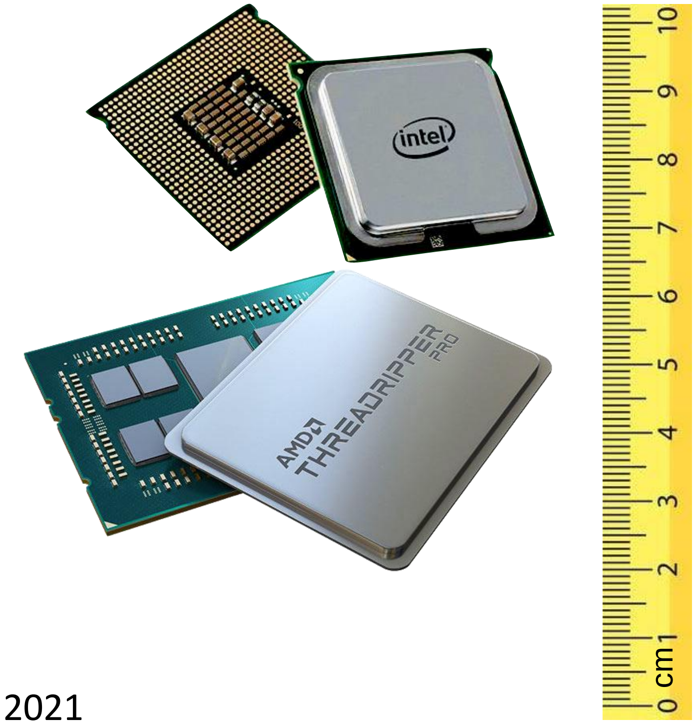
Stars indicate separate elements, rest are either separate or integrated

Integrated Circuits

- In the late 1950, transistors replaced the electronic valves as a switch to build gates and use in making computers as discrete elements, since they are smaller, faster and more reliable
- Todays transistors are built using Complementary Metal Oxide Semiconductor (CMOS) type for density and power consumption
- Integrated Circuits (ICs) is about placing the whole circuit, transistors and connections, on a single die yielding smaller space and more reliability
- Dies are packaged in what we called chips, with many contacts through with it communicates with the other components
- Contacts represent: Address, Data, Control & Power lines
- Contacts started with 16, 18, 40, 64, … and now 6096 ? !
    - How does the number of transistors affect performance?
    - How does wider data bus affect performance?
    - How does wider address bus affect performance?
    - How does higher clock frequency affect performance?
    - How does power consumption relate to performance?
    - Why do we have many VCC& GND inputs ?

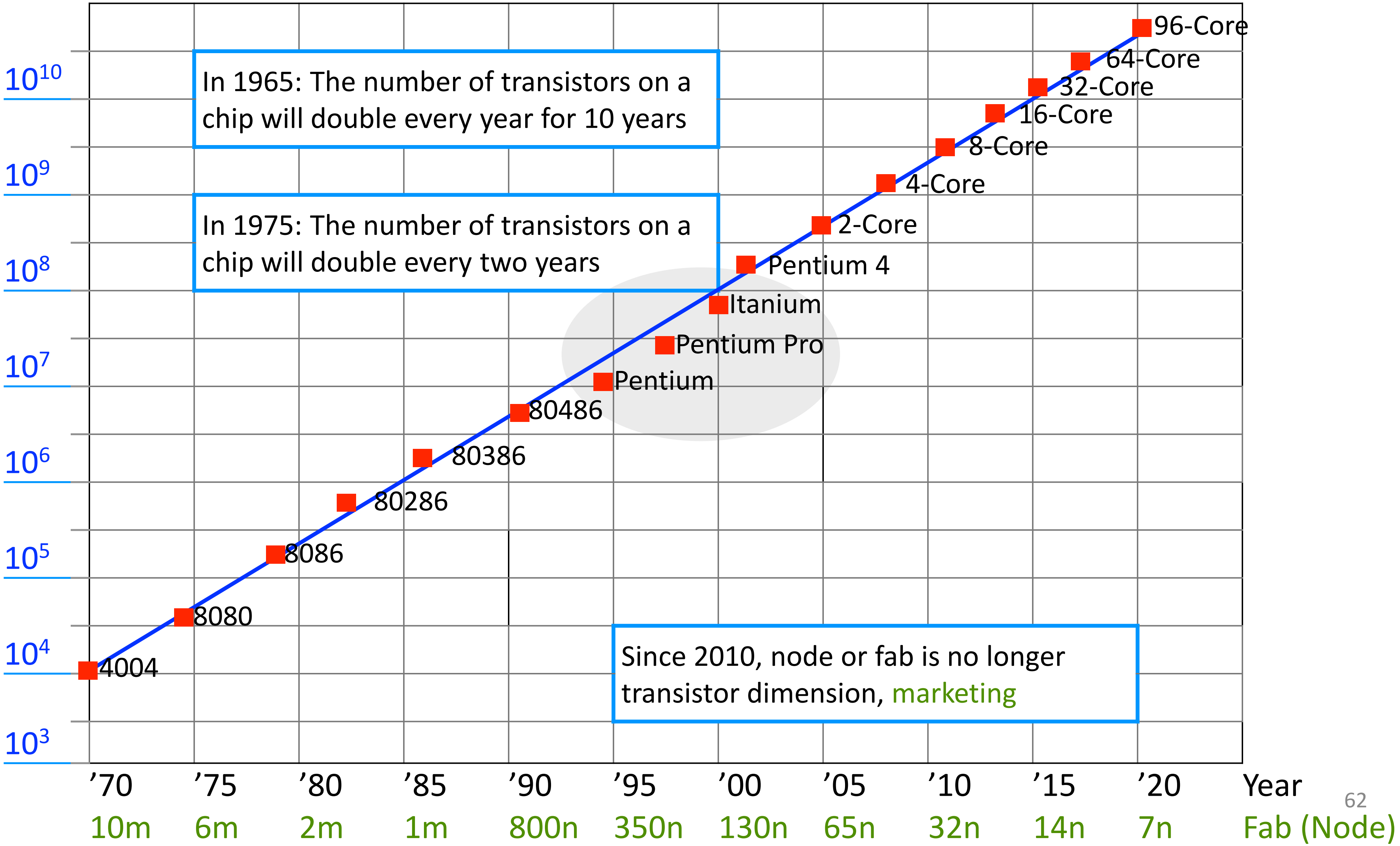


Modern Systems



- Integration Levels (Tr for Transistor, K, M & B for Kilo, Million & Billion)
  - 1975: 10 KTr, Very Large Scale Integration (VLSI)
  - 1980: 100 KTr, Ultra Large Scale Integration (ULSI)
  - 1990: 1 MTr, Extremely Large Scale Integration (ELSI)
  - 2000: 10 MTr, VLSI for all as a generic name
  - 2010: 1 BTr
  - 2013: 10 BTr
  - 2015: 20 BTr
  - 2018: 40 BTr
  - 2020: 50 BTr
  - 2021: 57 BTr
- Wafer Scale Integration (WSI)) integrates 2,600 BTr on nearly 22 cm x 22 cm in 2021
- Making chips from Silicon is a long, complex and extremely precise process
- It uses lithography to print dies on wafers to cut them later and package them as chips
- Process, node or fab, used to refer to the transistor dimension (10 µm was the beginning)
- Today, the node is more of marketing number although smaller still means smaller transistors
- Today's 5 nm process has 130 to 230 MTr/mm^2; cells of 100 x 100 nm^2 (including wiring …)

Moore’s Law



In 1965: The number of transistors on a chip will double every year for 10 years
In 1975: The number of transistors on a chip will double every two years
Since 2010, node or fab is no longer transistor dimension, marketing
10^{10}
10^9
10^8
10^7
10^6
10^5
10^4
10^3
4004
8080
8086
80286
80386
80486
Pentium
Pentium Pro
Itanium
Pentium 4
2-Core
4-Core
8-Core
16-Core
32-Core
64-Core
96-Core
’70 ’75 ’80 ’85 ’90 ’95 ’00 ’05 ’10 ’15 ’20 Year
10m 6m 2m 1m 800n 350n 130n 65n 32n 14n 7n Fab (Node)

RISC vs. CISC - Specifications

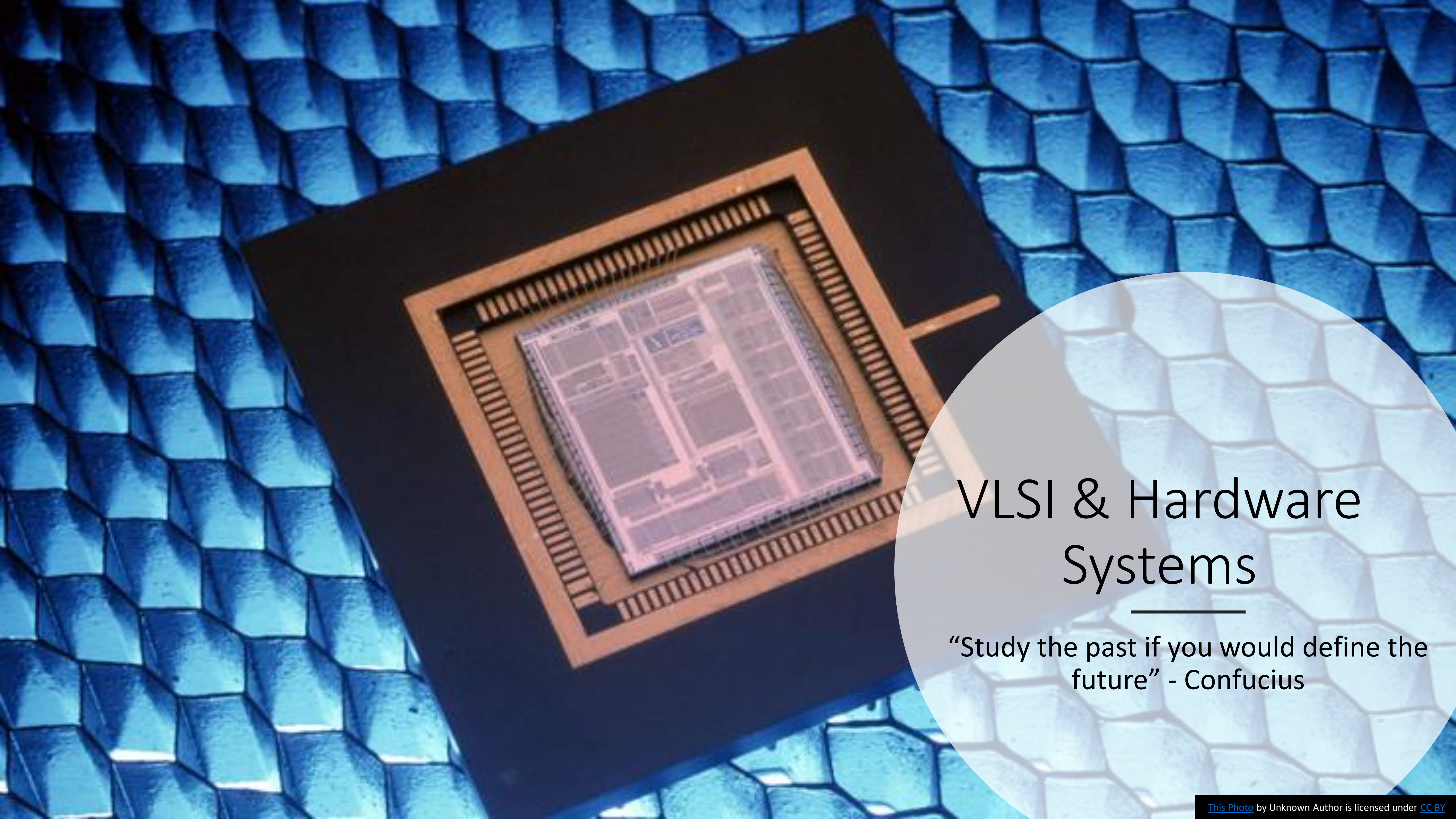
- RISC compilers are harder to design, take more time to compile, generate larger binary codes; around 25% - 50% more
- RISC implements conditional instructions, reducing the number of instructions to less than double or triple as might be deduced from the example below
- Example: when compiled, 1 KiloLine of high level language code may produce:
  - 1500 to 2000 instructions for a CISC processor
  - 2000 to 2500 instructions for a RISC processor
- RISC compilers are hard to design
- RISC compilers take longer to run
- RISC / CISC code segments
  - A & B are variables
  - R1 & R2 are registers

CISC Code		RISC Code	
AND	R1, R2	AND	R1, R2
MOVE	A, B	LOAD	A, R1
		STORE	R1, B
ADD	R1, A	LOAD	A, R2
		ADD	R1, R2
		STORE	R2, A
ADD	A, B	LOAD	A, R1
		LOAD	B, R2
		ADD	R1, R2
		STORE	R2, B

State of the Art - Specifications

- Compute Cores (CPUs): General purpose compute cores with variety of instructions
  - Around 300 MTr per core (Speed oriented; Do it fast)
  - 2 - 8 cores, Laptops class µPs
  - 4 - 12 core, Desktops class µPs
  - 12 - 24 core, Workstations
  - 24 - 64 core, Servers
  - 96 - 100 cores now available for Data Centers

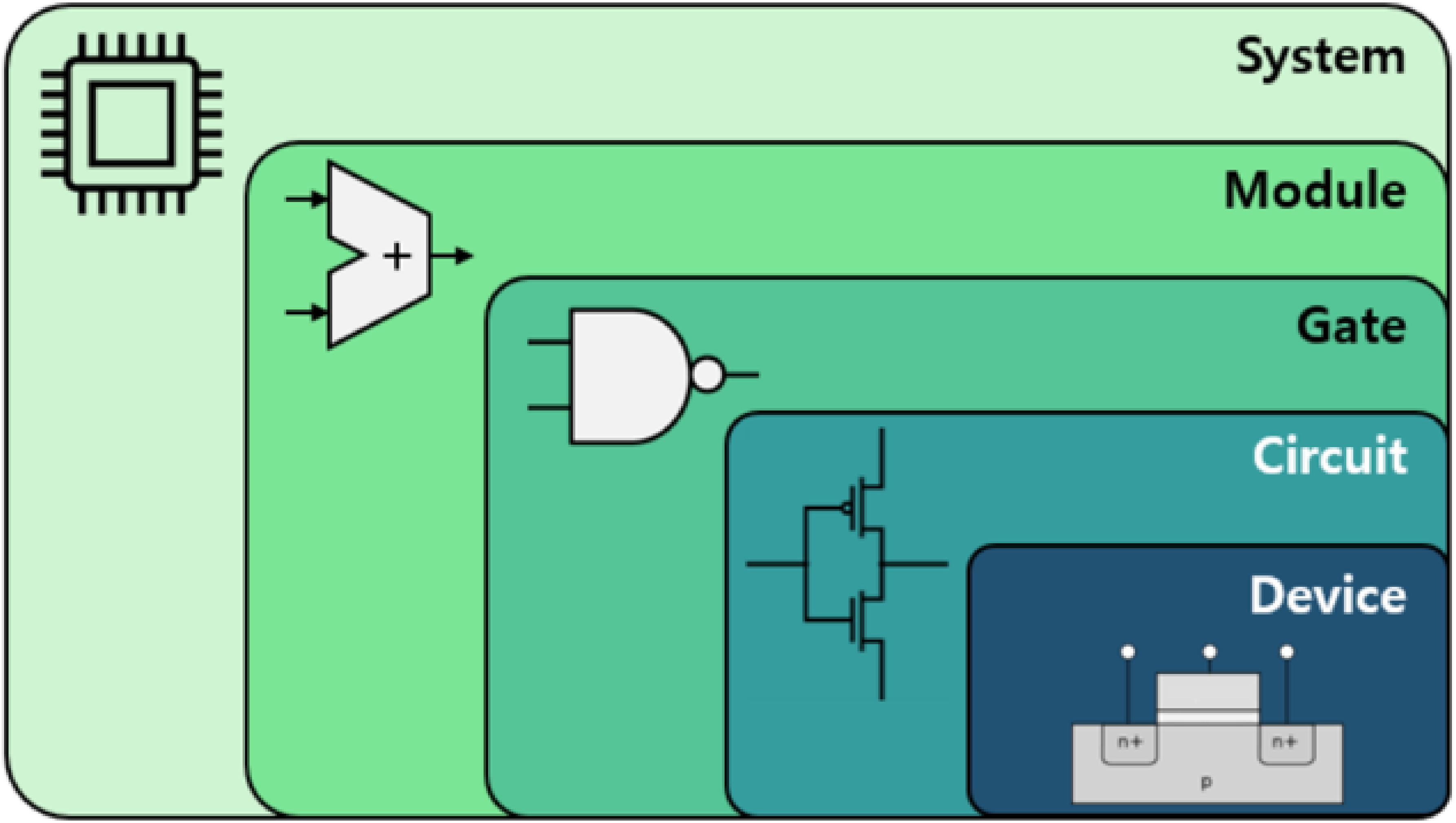
- Graphic Cores (GPUs)
  - Around 3 MTr per core (Throughput oriented; Do more)
  - Special purpose graphic cores, limited to Math & Logic instructions
  - 100s or 1000s of GPUs on chip as co-processors add on
  - 100s of stages in every pipeline; Single Instruction Multiple Data (SIMD)
- Special Cores; Neural Processing Units (NPUs) & Tensor Processor Units (TPSs)
  - Special purpose neural processing cores, limited to specific functions for neural network algorithms, for training and recognition (Throughput oriented; Do more)
  - Like GPU Cores; few Millions of transistors per core. Typically, 10s NPUs on an So



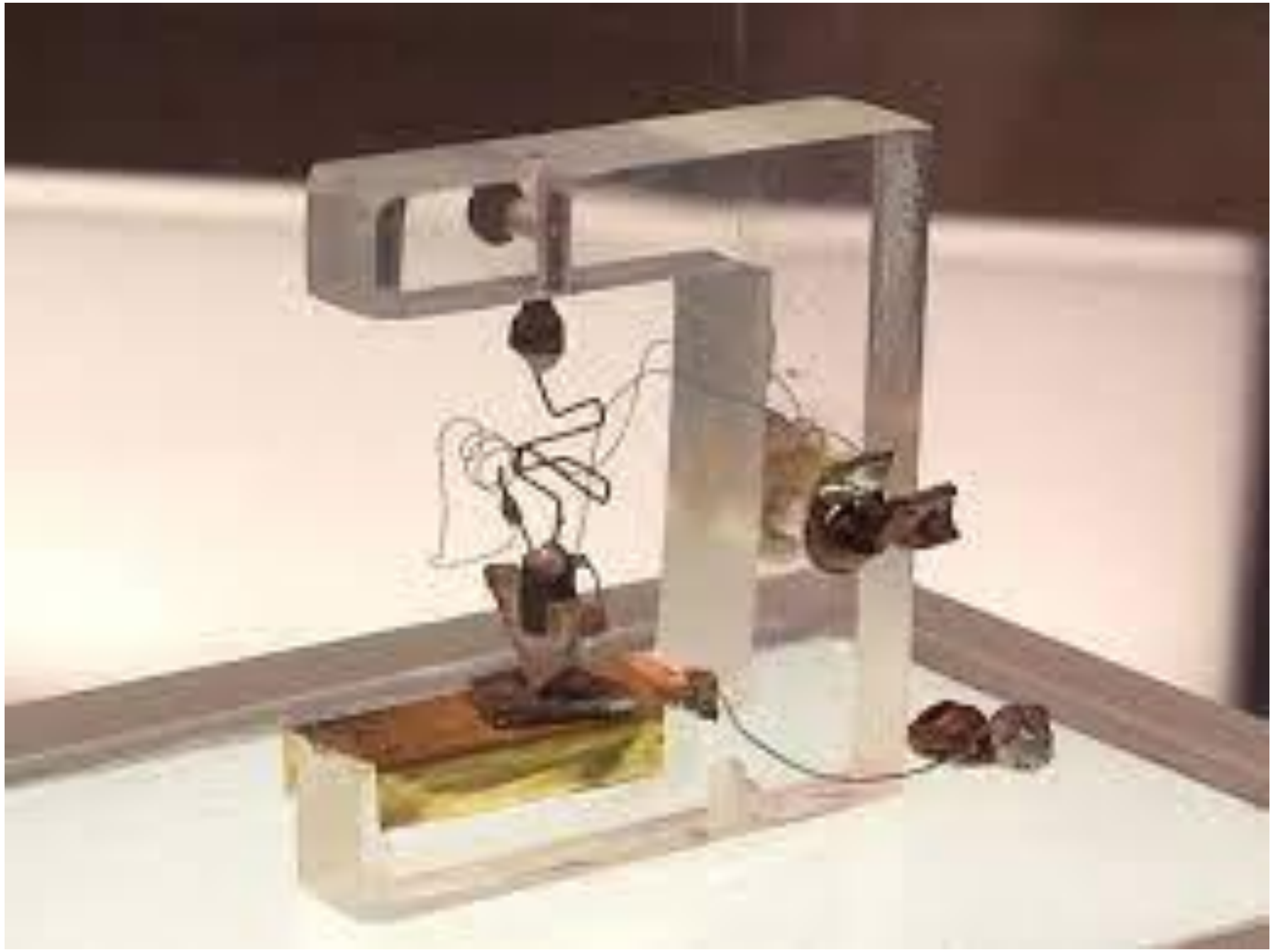
VLSI & Hardware Systems

"Study the past if you would define the future" - Confucius

Hardware Design Abstractions

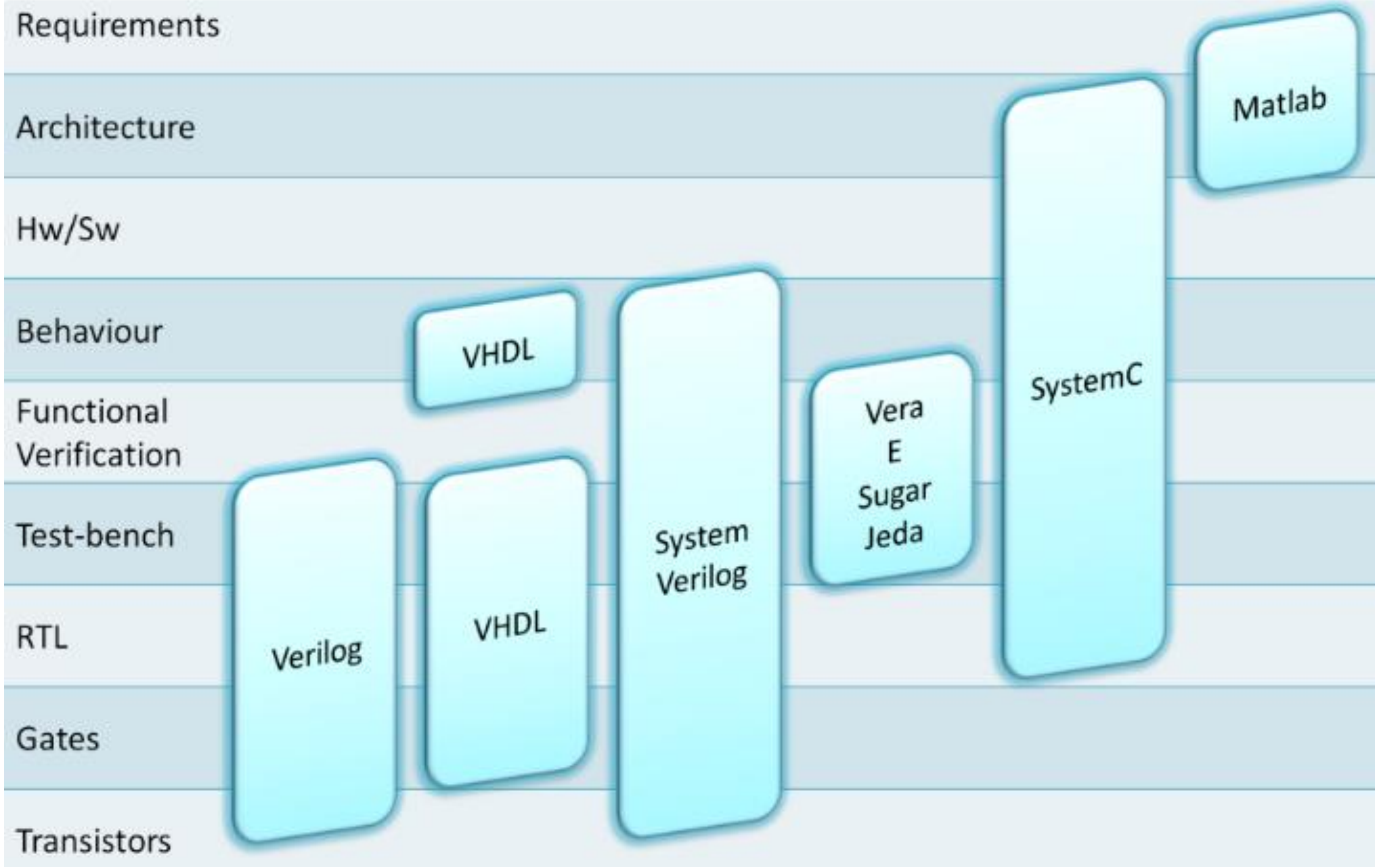


First Invented Transistor



Fabrication Room

Hardware Description Languages



System on a Chip Design