CMPE 200
Computer Architecture & Design

Lecture 1.
Computer Architecture & Design Overview (1)

Haonan Wang

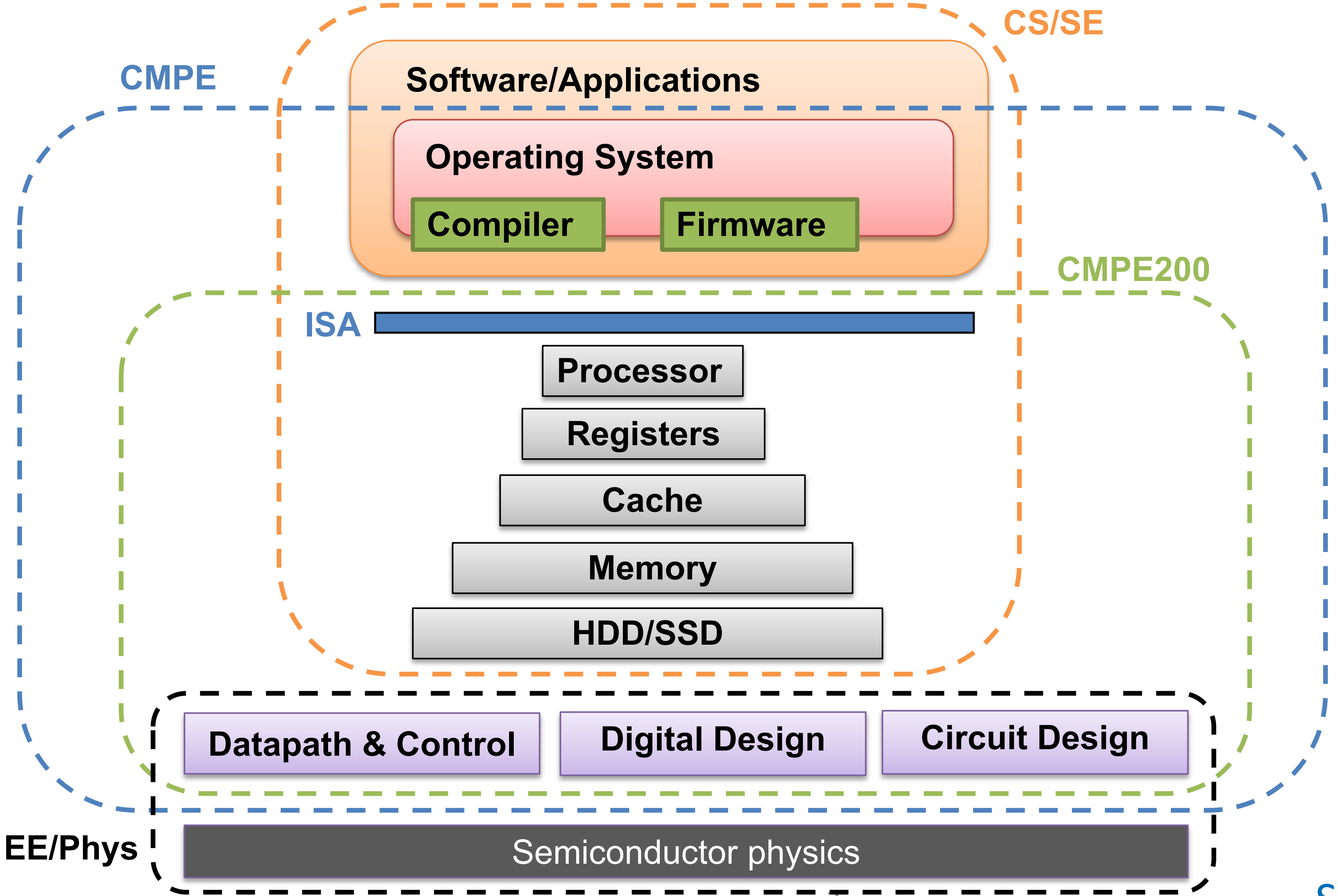
SJSU
SAN JOSÉ STATE UNIVERSITY

About Prerequisite – What to Submit?

- **If you have 180D in your admission letter**
  - You then need to submit your transcript with 180D highlighted
  - You cannot take this course before finishing 180D

- **If you do not have 180D in your admission letter**
  - You then need to submit your admission letter

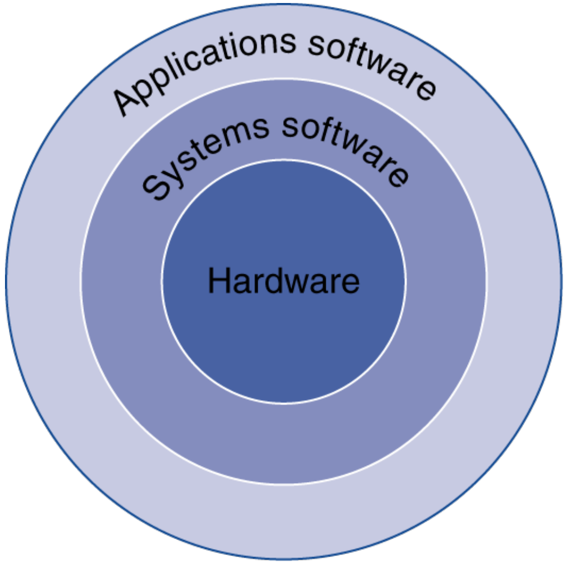
Abstraction of Computers



SJSU SAN JOSÉ STATE UNIVERSITY

The Software's Point of View

- **Application software**
  - Written in high-level language
- **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, …

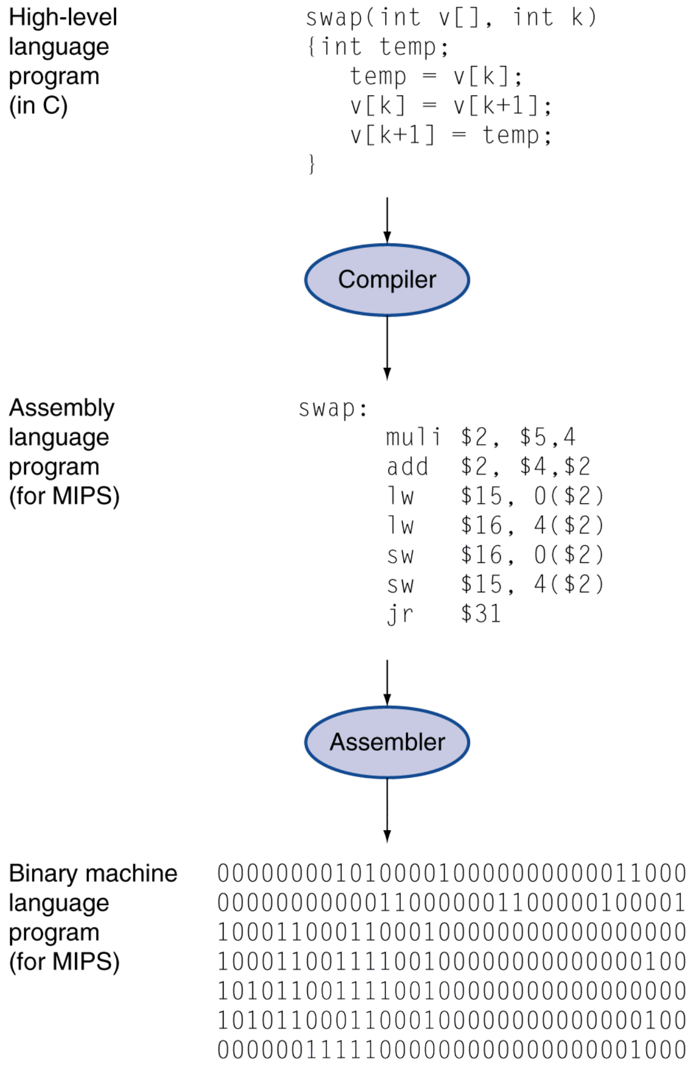


The Codes' Point of View

- **High-level language**
  - Abstraction Level closer to problem domain
  - Provides for productivity and portability

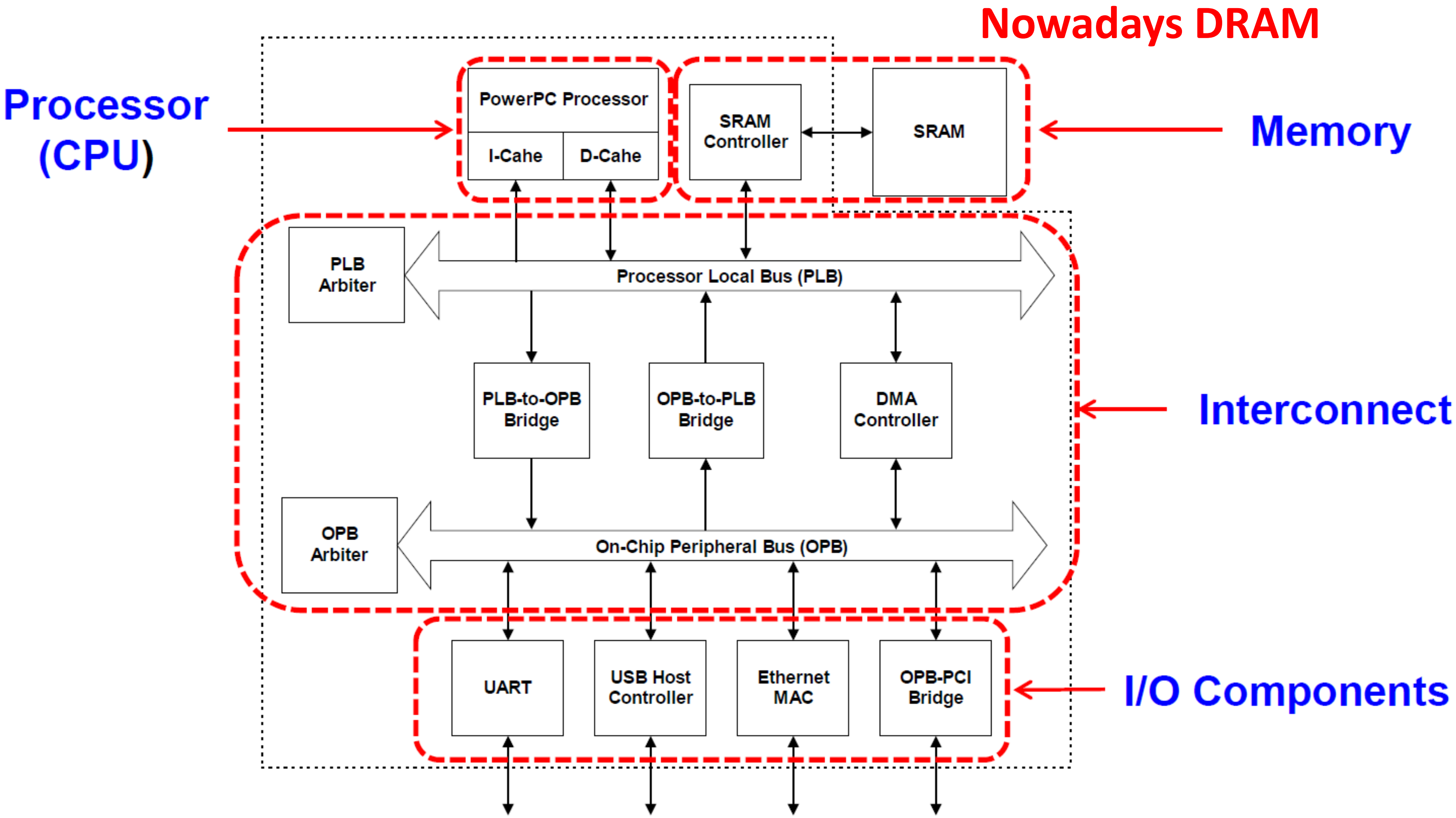
- **Assembly language**
  - Textual representation of instructions

- **Hardware representation**
  - Binary digits (bits)
  - Encoded instructions and data



Hardware's Point of View

Computer Organization:



Your Thoughts

Try to conclude with as few words as possible:

Computer Organization vs. Computer Architecture

– What is the key difference?

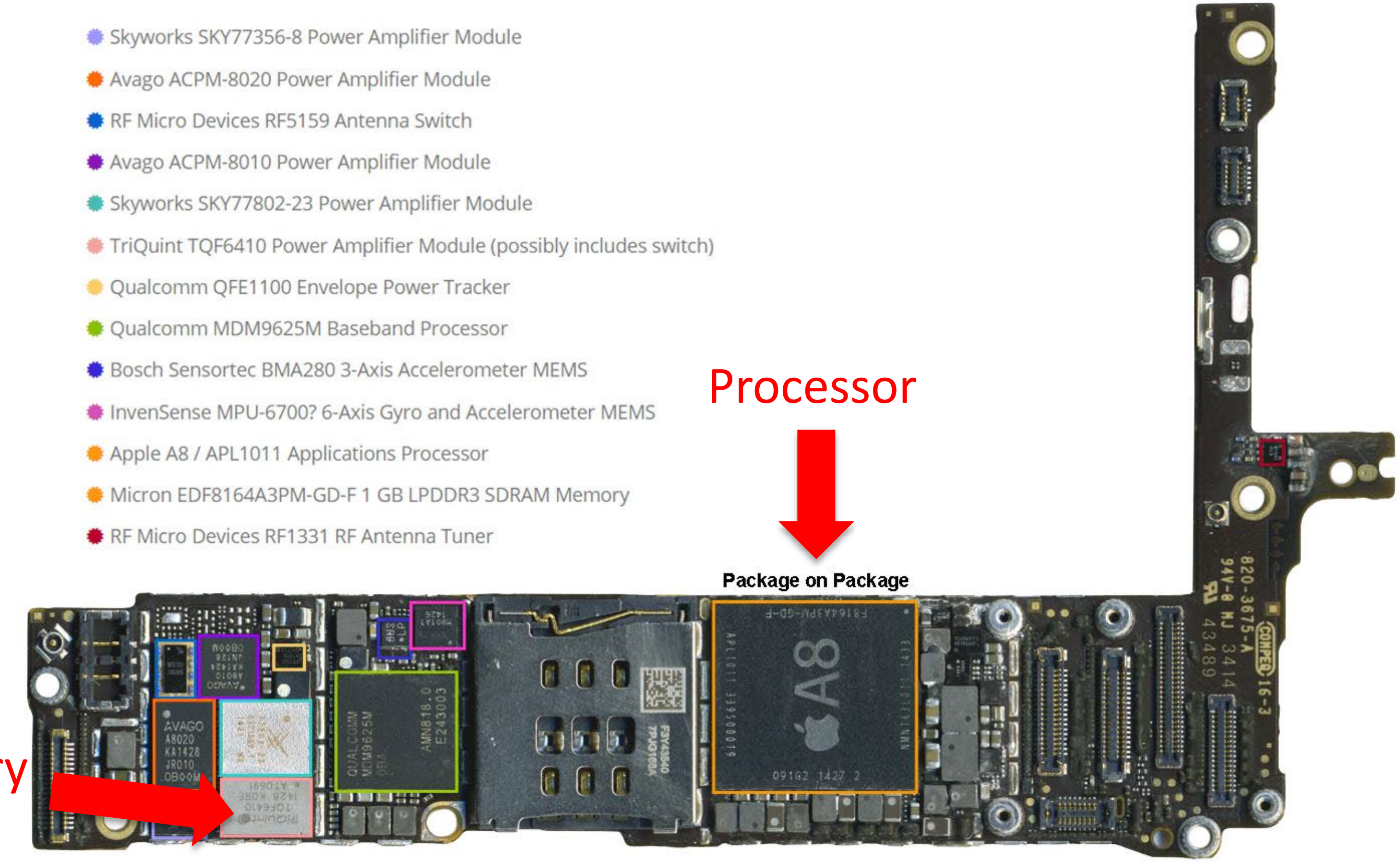
Connection vs. interface

Hardware Examples

- Skyworks SKY77356-8 Power Amplifier Module
- Avago ACPM-8020 Power Amplifier Module
- RF Micro Devices RF5159 Antenna Switch
- Avago ACPM-8010 Power Amplifier Module
- Skyworks SKY77802-23 Power Amplifier Module
- TriQuint TQF6410 Power Amplifier Module (possibly includes switch)
- Qualcomm QFE1100 Envelope Power Tracker
- Qualcomm MDM9625M Baseband Processor
- Bosch Sensortec BMA280 3-Axis Accelerometer MEMS
- InvenSense MPU-6700? 6-Axis Gyro and Accelerometer MEMS
- Apple A8 / APL1011 Applications Processor
- Micron EDF8164A3PM-GD-F 1 GB LPDDR3 SDRAM Memory
- RF Micro Devices RF1331 RF Antenna Tuner

Processor

Main memory (SDRAM)

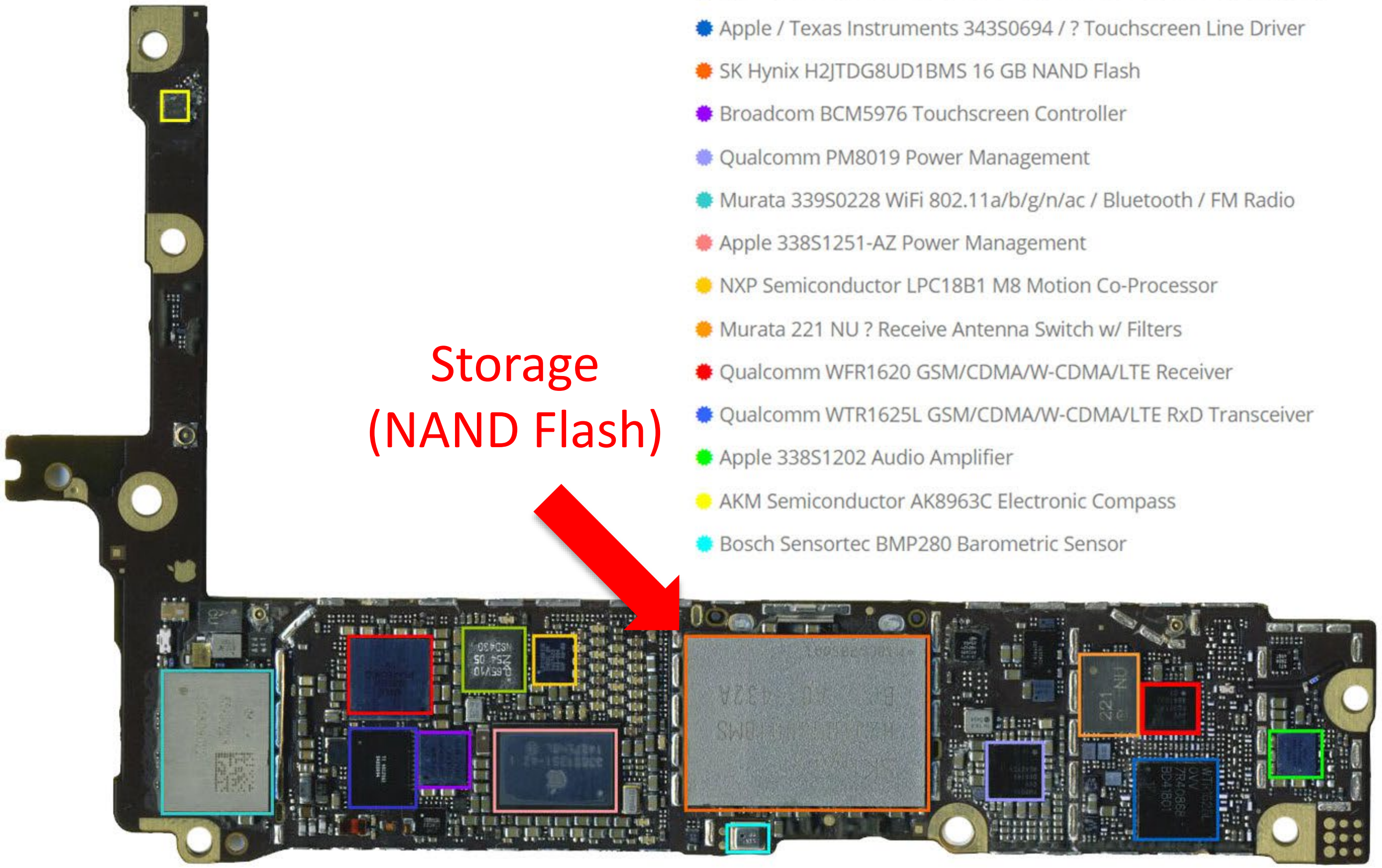


Figures from http://www.techinsights.com/teardown.com/apple-iphone-6/

Hardware Examples

Input/output device
(touch screen)

Storage
(NAND Flash)



- NXP Semiconductor PN548 NFC Controller w/ Secure Element Chip
- Apple / Texas Instruments 343S0694 / ? Touchscreen Line Driver
- SK Hynix H2JTDG8UD1BMS 16 GB NAND Flash
- Broadcom BCM5976 Touchscreen Controller
- Qualcomm PM8019 Power Management
- Murata 339S0228 WiFi 802.11a/b/g/n/ac / Bluetooth / FM Radio
- Apple 338S1251-AZ Power Management
- NXP Semiconductor LPC18B1 M8 Motion Co-Processor
- Murata 221 NU ? Receive Antenna Switch w/ Filters
- Qualcomm WFR1620 GSM/CDMA/W-CDMA/LTE Receiver
- Qualcomm WTR1625L GSM/CDMA/W-CDMA/LTE RxD Transceiver
- Apple 338S1202 Audio Amplifier
- AKM Semiconductor AK8963C Electronic Compass
- Bosch Sensortec BMP280 Barometric Sensor

Figures from http://www.techinsights.com/teardown.com/apple-iphone-6/

Hardware Examples

What is the advantage of having them in the same package?

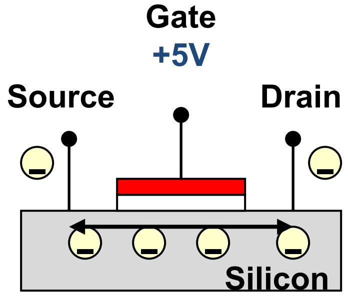
SJSU SAN JOSÉ STATE UNIVERSITY

Figures from http://www.anandtech.com/show/8562/chipworks-a8

Electronic Components' Point of view

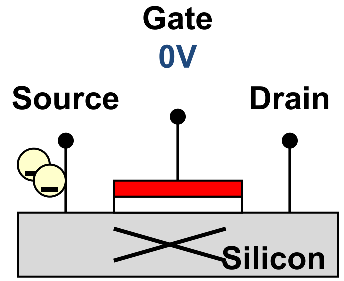
- **Transistors**
  - 3-terminal device
  - Gate input: the control input; its voltage determines whether current can flow
  - Source & Drain: terminals that current flows from/to

- **Many transistors can be fabricated on one piece of silicon** (i.e. an integrated chip, IC)



Transistor is 'on'

High voltage at gate allows current to flow between source and drain



Transistor is 'off'

Low voltage at gate prevents current from flowing between drain and source

Integrated Circuit

Actual silicon wafer is quite small but can contain several billion transistors

Silicon wafer is then packaged to form the chips we are familiar with

More transistors = faster core?

Moore’s Law (1965)

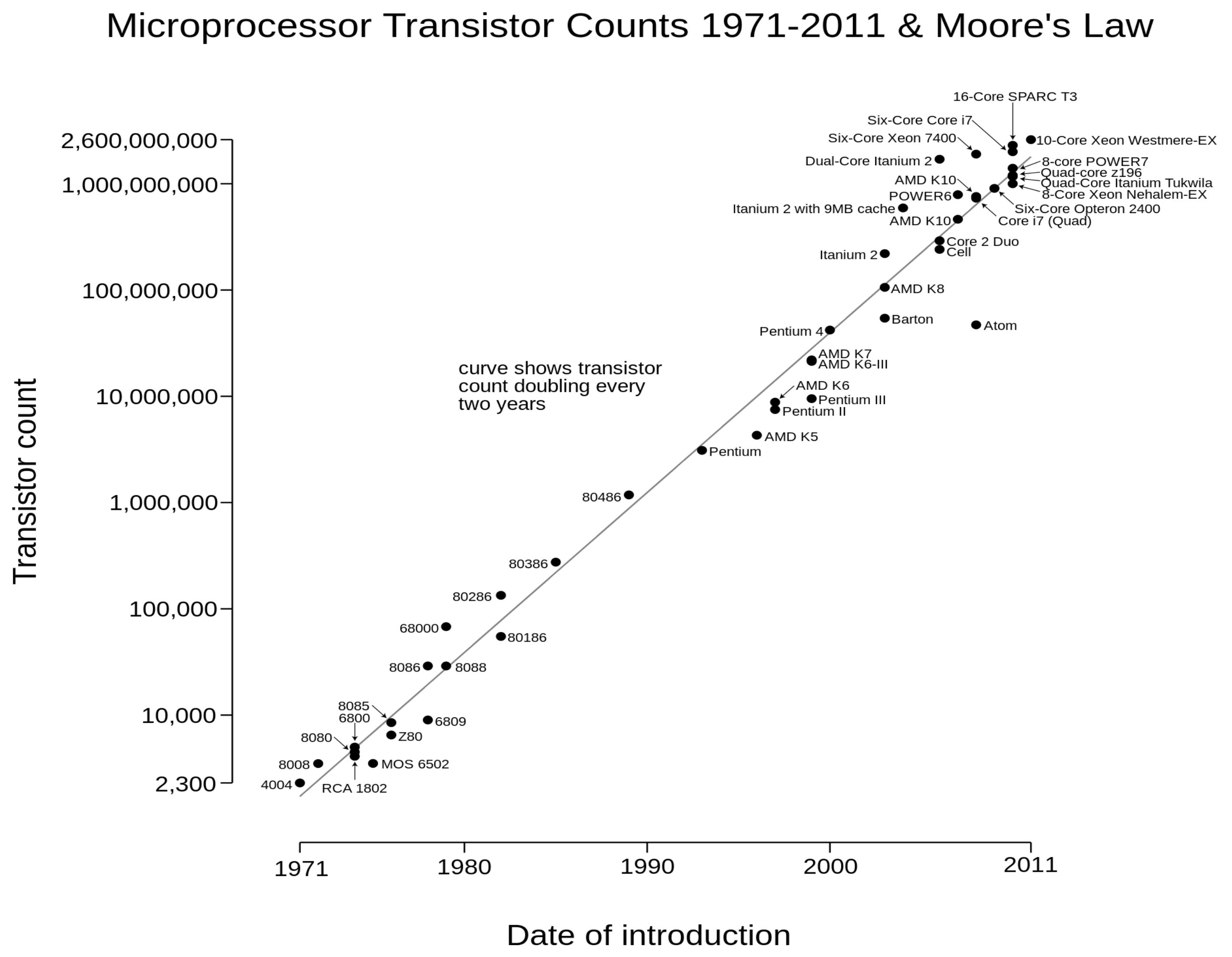
In 1975, he recalibrated it to every two years. Later, it is recalibrated again to 18 months, as is widely known as Moore’s Law.

“The number of transistors per square inch on integrated circuits had doubled every year since the integrated circuit was invented and this trend would continue for the foreseeable future.”

-- Gordon E. Moore, “Cramming More Components onto Integrated Circuits,” Electronics, pp. 114–117, April 19, 1965.

Moore’s Law (1965)

The law has been preserved over four decades..

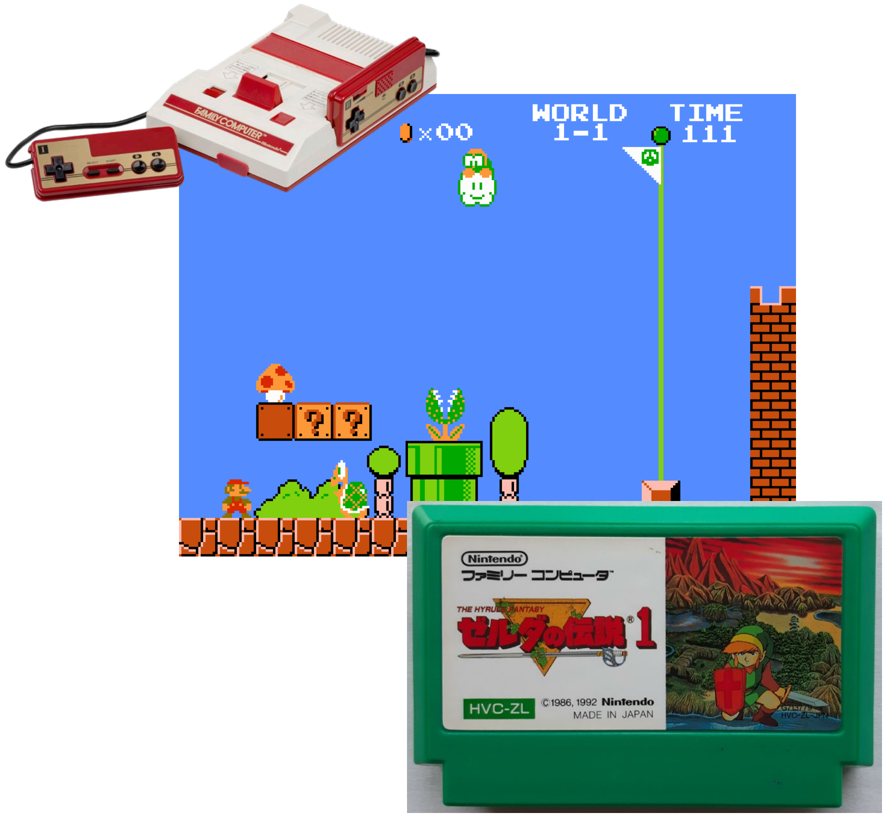


SJSU SAN JOSÉ STATE UNIVERSITY

What Does It Enable?

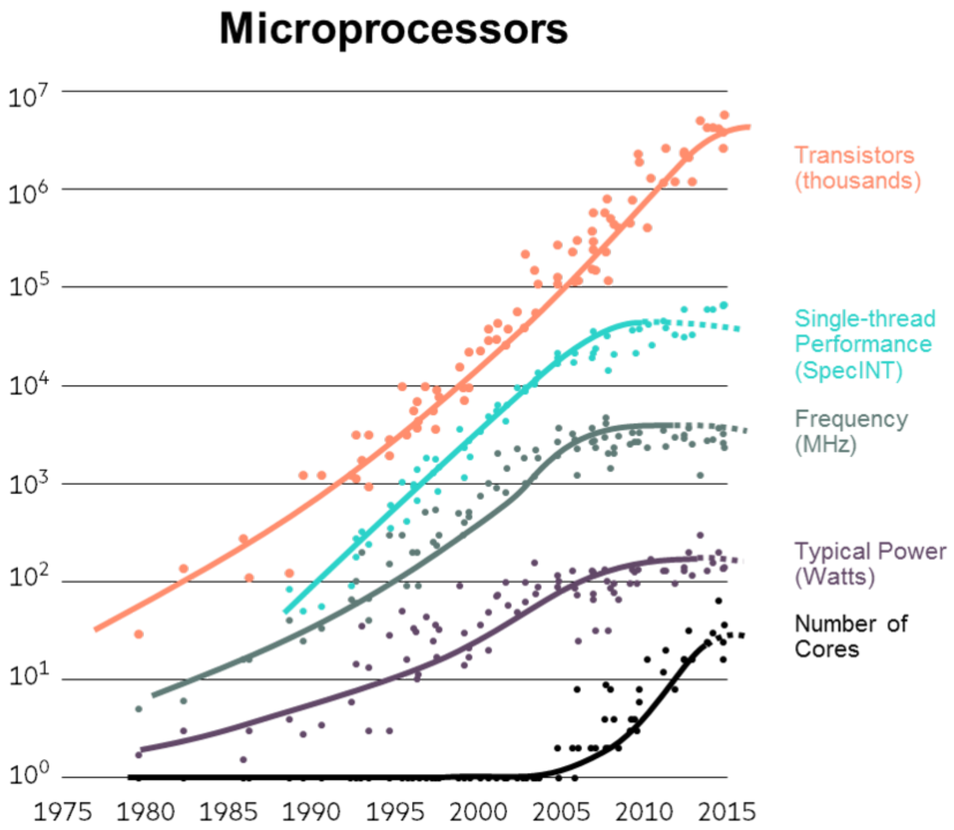
Applications that were impossible to run on the computers before

– 3D games, Augmented reality, Virtual reality, Deep learning…



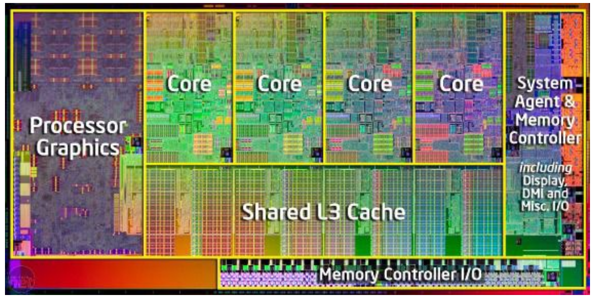
Moore's Law (1965)

Will it last forever?



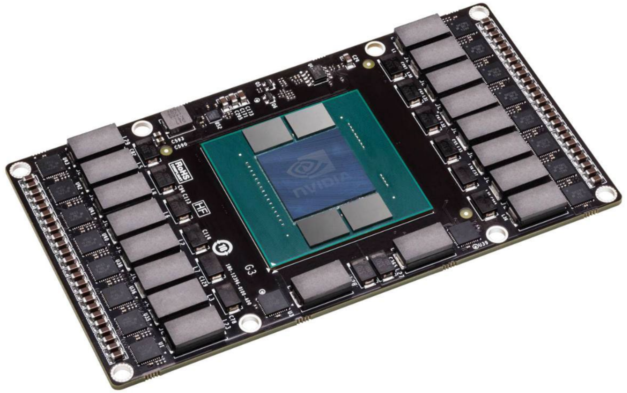
Dead?

Current Trend of Computer Systems



- **Multi- or Many-core Processors**
  - Embedding multiple cores in one CPU

- **Complex microarchitecture**
  - Out-of-order execution, branch predictor, memory prefetcher, etc..



- **Accelerators**
  - GPUs, FPGAs, etc..
  - Application specific designs

- **Novel architectures**
  - Dark silicon, 3D-stacking, neuromorphic computing, quantum, etc..

Metrics: How Do We Evaluate a System?

- **Performance (speed)**
  - Response time (Latency):
    - How long it takes to do a task
  - Throughput :
    - Total work done per unit time
- **Power efficiency**
  - How much power it consumes to do a task
- **Cost**

What Determines Performance?

- **Algorithm**
  - Determines number of operations executed
- **Programming language, compiler, architecture**
  - Determine number of machine instructions executed per operation
- **Processor and memory system**
  - Determine how fast instructions are executed
- **I/O system (including OS)**
  - Determines how fast I/O operations are executed

Relative Performance

- **Define Performance = 1/Execution Time**
- **"X is n times faster than Y"**

$$\text{Performance}_X / \text{Performance}_Y$$
$$= \text{Execution time}_Y / \text{Execution time}_X = n$$

Example: time taken to run a program -- 10s on a computer A, 15s on computer B

Execution Time_B / Execution Time_A = 15s / 10s = 1.5

A is 1.5 times faster than B.

Measuring Execution Time

- **Elapsed time**
  - Total response time, including all aspects
    - Processing, I/O, OS overhead, idle time
  - Determines system performance
- **CPU time**
  - Time spent processing a given job on CPU
    - Discounts I/O time, other jobs' shares
  - Estimating CPU time is relatively easy based on the number of lines of code and processing speed of the CPU

Conclusion Time

Why do we need multi-core processors to keep the Moore's Law alive?

- **Adding transistors no longer works**

How do we evaluate a system?

- **Performance**
- **Power efficiency**
- **Cost**

Conclusion Time

Why do we use CPU time to measure the performance?

- **Avoid influences**

SAN JOSÉ STATE UNIVERSITY *powering* SILICON VALLEY