Lecture 1: Understanding our computer

Jin-Guo Liu[1,*]

[1] *Hong Kong University of Science and Technology, Guangzhou 510000, China*

I. COMPUTER ARCHITECTURE

A. CPU

You CPU information can be obtained by typing `lscpu`.

```
$ lscpu
Architecture:            x86_64
  CPU op-mode(s):        32-bit, 64-bit
  Address sizes:         39 bits physical, 48 bits virtual
  Byte Order:            Little Endian
CPU(s):                  8
  On-line CPU(s) list:   0-7
Vendor ID:               GenuineIntel
  Model name:            Intel(R) Core(TM) i7-10510U CPU @ 1.80GHz
    CPU family:          6
    Model:               142
    Thread(s) per core:  2
    Core(s) per socket:  4
    Socket(s):           1
    Stepping:            12
    CPU max MHz:         4900.0000
    CPU min MHz:         400.0000
    BogoMIPS:            4599.93
    Flags:               fpu vme de pse tsc msr pae mce cx8 apic sep mtrr pge mc
                         a cmov pat pse36 clflush dts acpi mmx fxsr sse sse2 ss
                         ht tm pbe syscall nx pdpe1gb rdtscp lm constant_tsc art
                          arch_perfmon pebs bts rep_good nopl xtopology nonstop_
                         tsc cpuid aperfmperf pni pclmulqdq dtes64 monitor ds_cp
                         l vmx est tm2 ssse3 sdbg fma cx16 xtpr pdcm pcid sse4_1
                          sse4_2 x2apic movbe popcnt tsc_deadline_timer aes xsav
                         e avx f16c rdrand lahf_lm abm 3dnowprefetch cpuid_fault
                          epb invpcid_single ssbd ibrs ibpb stibp ibrs_enhanced
                         tpr_shadow vnmi flexpriority ept vpid ept_ad fsgsbase t
                         sc_adjust sgx bmi1 avx2 smep bmi2 erms invpcid mpx rdse
                         ed adx smap clflushopt intel_pt xsaveopt xsavec xgetbv1
                          xsaves dtherm ida arat pln pts hwp hwp_notify hwp_act_
                         window hwp_epp md_clear flush_l1d arch_capabilities
Virtualization features:
  Virtualization:        VT-x
Caches (sum of all):
  L1d:                   128 KiB (4 instances)
  L1i:                   128 KiB (4 instances)
  L2:                    1 MiB (4 instances)
  L3:                    8 MiB (1 instance)
NUMA:
  NUMA node(s):          1
```

[*] jinguoliu@hkust-gz.edu.cn

```
  NUMA node0 CPU(s):      0-7
Vulnerabilities:
  Itlb multihit:          KVM: Mitigation: VMX disabled
  L1tf:                   Not affected
  Mds:                    Not affected
  Meltdown:               Not affected
  Mmio stale data:        Mitigation; Clear CPU buffers; SMT vulnerable
  Retbleed:               Mitigation; Enhanced IBRS
  Spec store bypass:      Mitigation; Speculative Store Bypass disabled via prctl
                           and seccomp
  Spectre v1:             Mitigation; usercopy/swapgs barriers and __user pointer
                           sanitization
  Spectre v2:             Mitigation; Enhanced IBRS, IBPB conditional, RSB fillin
                          g, PBRSB-eIBRS SW sequence
  Srbds:                  Mitigation; Microcode
  Tsx async abort:        Not affected
```

The computing power of a device can be measured by the number of floating point operations your computing device and perform in one second, namely, in floating point operations per second (FLOPS).

```
The power of a single thread CPU = 2.9 GHz (CPU clock speed, we use the maximum Turbo frequenc
                            * 2 (multiplication and add can happen at the same CPU clock)
                            * 2 (number of instructions per cycle)
                        * 4 (avx instruction set has a 256 with register, it can
                    crunch 4 vectorized double precision floating point
                                  operations at one CPU cycle)
                          = 46.4 GFLOPS
```

B. GPU

The GPU information of your computer can be obtained using the `nvidia-smi` command.

C. Storage Hierachy

By the descreasing order of accessing speed, the storage can be classified as registers, random access memory (RAM) and hard disk. Registers are tightly related to the instruction sets of a CPU. Before doing any arithematic operation, data are always loaded from the RAM to a specific register and call the specific instruction and then copy the result back to the RAM.

When we talk about memory in our daily life, we usually talk about the main memory, or the dynamic random access memory (DRAM). It works in a much slower clock speed and is in general very slowr to access (comparing to CPU clock time). Due to the fact that the frequeny access to RAM is inevitable, people developed the 3-level caching system, or the static random access memory, that existing in majority of our modern computers. They are fastest L1 cache working in the same speed as CPU, slower L2 cache, and slowest L3 cache that only slightly faster than the DRAM. The wisdom behind the caching system is data locality, i.e. whenever a data at some address is used, the data physically close to it has much higher probability to be used than the rest. Locality is particularly true when a program enumerates the items in an array with contiguous storage.

```
$ lsmem
RANGE                                  SIZE  STATE REMOVABLE  BLOCK
0x0000000000000000-0x000000007fffffff    2G online       yes   0-15
0x0000000088000000-0x000000008fffffff  128M online       yes     17
0x0000000100000000-0x0000000a6fffffff 37.8G online       yes 32-333

Memory block size:       128M
Total online memory:    39.9G
Total offline memory:      0B
```