TRINITY COLLEGE DUBLIN, SCHOOL OF MATHEMATICS

Assignment #1, Module: MA5612

Gustavo Ramirez

October 16, 2016

1 PROBLEM DESCRIPTION

1. Find and summarise the current generation processors from each of the following manufacturers: Intel, AMD, IBM, with respect to manufacturing technology, clock speeds, core count, cache sizes. What can you find out about the Chinese Sunway processor?

2. Explain what is meant by an accelerator and provide examples of some current accelerators available.

3. Describe the top500 list of supercomputers. What sorts of programmes run on a selection of the machines in the top 10? Describe the Linpack benchmark and explain how it works. Find some other benchmarks that are sometimes used to rank computer systems.

4. Explain what is meant by Grid Computing and Cloud Computing outlining some of the benefits and risks associated with them.

1.1 CURRENT GENERATION PROCESSORS

- **Intel**: Intel's last generation, 7th (with codename Kaby Lake), has two[1] processors.

Processor	Manufacturing technology	Clock speed	Core count	Cache size
Core™ i7-7Y75	14 nm	3.6 GHz	2	4 MB SmartCache
Core™ i7-7500U	14 nm	3.5 GHz	2	4 MB SmartCache

- **AMD**: 7th Generation AMD A-Series Processors and FX Processors[2] (manufacturing technology: 28 nm; for the upcoming Zen computer processors, the manufacturing technology will be 14 nm).

Model	Cores	CPU Frequency (Max/Base)	L2 Cache
FX™ 9830P	4	3.7 / 3.0 GHz	2MB
FX™ 9800P	4	3.6 / 2.7 GHz	2MB
A12-9730P	4	3.5 / 2.8 GHz	2MB
A12-9700P	4	3.4 / 2.5 GHz	2MB
A10-9630P	4	3.3 / 2.6 GHz	2MB
A10-9600P	4	3.3 / 2.4 GHz	2MB
A9-9410	2	3.5 / 2.9GHz	1MB
A6-9210	2	2.8 / 2.4GHz	1MB
E2-9010	2	2.2 / 2.0GHz	1MB

- **IBM**: The current IBM processor is named Power8, with manufacturing technology of 22 nm (but new chips at 7 nm have already been made and tested at low volume)[3].

  There are several IMB systems using the Power8 technology[4], and depending on the specific server, the processor options are different. For example, in the case of the IBM Power S812LC, it can have 3.32 GHz and 8 cores, or 2.92 GHz and 10 cores, both cases with L2 cache per core of 512 KB.

[1] Specific information about these 2 processors taken from: http://ark.intel.com/products/family/95544/7th-Generation-Intel-Core-i7-Processors
[2] Taken from: http://www.amd.com/en-us/products/processors/laptop-processors
[3] This prior information was taken from: http://www.nextplatform.com/2015/08/10/ibm-roadmap-extends-power-chips-to-2020-and-beyond/
[4] Taken from: http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=POB03046USEN

1.1.1 Sunway Chinese Processor

The ShenWei SW26010 processor is the building block of the Sunway TaihuLight (the current world's fastest supercomputer, according to Top500). There are[5] 260 cores in the chip, and each chip delivers a performance of 3 teraflops. It is a 64-bit RISC processor.

1.2 Accelerators

An accelerator[6] is a separate architectural substructure that is designed using a different set of objectives than the base processor, where these objectives are derived from the needs of a special class of applications. Through this manner of design, the accelerator is tuned to provide higher performance at lower cost, or at lower power, or with less development effort than with the general-purpose base hardware. Depending on the domain, accelerators often bring greater than a 10× advantage in performance, or cost, or power over a general-purpose processor.

Then, in summary, an accelerator is an extension of a processor, with the idea of serving broader purposes, with better benefits (in power consumption, performance, etc.).

Examples of accelerators include floating-point coprocessors, graphics processing units (GPUs) to accelerate the rendering of a vertex-based 3D model into a 2D viewing plane, and accelerators for the motion estimation step of a video codec.

Two specific examples of accelerators are[7]:

- Intel Xeon Phi: has a less specialized architecture than a GPU, and is designed to be familiar to anyone who has experience with parallel programming in an x86 environment. The Phi contains Intel Pentium generation processors and runs a version of the Linux operating system. Thus, it can execute parallel code written for "normal" computers using a wide variety of modern and legacy programming models including Pthreads, OpenMP, MPI and even GPU software (e.g. CUDA or OpenCL).

- Nvidia Kepler GPU: specifically designed for solving problems that can be expressed in a single-instruction, multiple thread (SIMT) model. For example, processing a large vector of data where each element of the vector can be treated independently can be easily matched to the SIMT model.

[5] Taken from: http://www.pcworld.com/article/3086107/hardware/chinas-secretive-super-fast-chip-powers-the-worlds-fastest-computer.html

[6] Taken from: http://home.deib.polimi.it/sami/architetture/accelerators.pdf

[7] Taken from: http://www.hpc.mcgill.ca/index.php/starthere/81-doc-pages/255-accelerator-overview

Both types of devices are cards that interface with a node through the PCI-express bus (Peripheral Component Interconnect Express, a high-speed serial computer expansion bus standard, designed to replace the older PCI, PCI-X, and AGP bus standards), and are designed to accelerate a computation through massive parallelization. These accelerator devices contain a large number of processing cores, as well as internal memory. They are most often used in conjunction with the CPUs of the node to accelerate certain "hot spots" of a computation that requires a large amount of algebraic operations.

1.3 TOP500

The top500 is[8] a list of the 500 most powerful computer systems. Their list has been compiled twice a year since June 1993 with the help of high-performance computer experts, computational scientists, manufacturers, and the Internet community in general. In the present list (which they call the TOP500), they list computers ranked by their performance on the LINPACK Benchmark.

The Linpack Benchmark is a measure of a computer's floating-point rate of execution. It is determined by running a computer program that solves a dense system of linear equations. Over the years the characteristics of the benchmark has changed a bit. Nowadays, there are three benchmarks included in the Linpack Benchmark report. The benchmark used in the LINPACK Benchmark is to solve a dense system of linear equations. For the TOP500, they used that version of the benchmark that allows the user to scale the size of the problem and to optimize the software in order to achieve the best performance for a given machine. This performance does not reflect the overall performance of a given system, as no single number ever can. It does, however, reflect the performance of a dedicated system for solving a dense system of linear equations. Since the problem is very regular, the performance achieved is quite high, and the performance numbers give a good correction of peak performance.

Another good (and recently growing) benchmark is the HPCG[9]; it that ranks supercomputers on their ability to solve complex problems rather than on raw speed alone.

Other benchmarks are[10]: NAS parallel benchmark suites, SPEC benchmark, HPC Challenge benchmark, STREAM.

[8] Taken from: https://www.top500.org/

[9] Taken from: http://insidehpc.com/2015/12/supercomputer-benchmark-gains-adherents/

[10] This list is taken from: http://www.ctwatch.org/quarterly/articles/2006/11/metrics-for-ranking-the-performance-of-supercomputers/

1.4 GRID & CLOUD COMPUTING

First of all, it's important to introduce the definitions[11] associated to these two concepts. The problem with Grid and Cloud computing (and in particular the latter), is that their definitions are not very well stablished. Here, specific definitions are taken, in order to be able to make a comparison between them.

- cloud: a style of computing where massively scalable IT-related capabilities are provided as a service across the cyber infrastructure to external users. It has been claimed for some aspects that cloud systems are narrow Grids (to be defined next), in the sense of exposing reduced interfaces.
- grid: a system that coordinates resources which are not subject to centralized control, using standard, open, general-purpose protocols and interfaces to deliver nontrivial qualities of service.

Another good but also simple and practical definition of cloud computing[12], is the following: *using the internet to allow people to access technology-enabled services; those services must be massively scalable.*
Instead of presenting some of the main characteristics of Grid and Cloud computing as benefits of risks, in the following table are presented some (not all) of the main characteristics of both schemes, in order to make a direct comparison. Whether or not this properties are benefits or risks, in general, depends on the specific application in which the comparison is taken.

Feature	**Grid**	**Cloud**
Resource Sharing	Collaboration	Assigned resources are not shared
Architecture	Service oriented	User chosen architecture
Virtualization	Virtualization of data and computing resources	Virtualization of hard. and soft. platforms.
Security	Security through credential delegations	Security through isolation

[11] The following two definitions were taken from: http://airccse.org/journal/ijccsa/papers/2412ijccsa01.pdf
[12] This definition, and the following table, were both taken from: http://airccse.org/journal/ijccsa/papers/2412ijccsa01.pdf