Gruppe 3: Davit Melkonyan, Florian Schrittwieser, Simon Pavicic
1.1
Reviews:

Michael J. Flynn and Patrick Hung. 2005. Microprocessor Design Issues: Thoughts on the Road Ahead.

Primary contribution:
Flynn and Hung's primary contribution is to identify the key challenges facing microprocessor designers, and to discuss the potential solutions to these challenges. They argue that the increasing importance of power consumption is forcing designers to make trade-offs between performance and power, and that the optimal power-oriented architectures are not yet clear.

Key insight of the contribution:
Flynn and Hung point out that the power consumption is becoming increasingly more important for microprocessor designers. This is due to several factors, including the increasing use of mobile devices and the growing demand for high-performance computing. They also mention that the traditional approach to increasing performance by increasing clock frequency is no longer sustainable, due to the increasing power consumption. Therefore, the authors suggest several directions for future development such as more focus on parallelism in high-performance architectures as well as Extremely low-power architecture (ELPA).

Rating:
Flynn and Hung's paper is a very well-written and insightful overview of the key challenges facing microprocessor designers in the early 21st century. Many of the challenges they discussed are still relevant today, and their paper continues to be a valuable resource for anyone interested in the future of microprocessor design. Therefore, the appropriate rating for the paper is an “accept”.

Walker, 2008, benchmarking Amazon EC2 for high-performance scientific computing, The USENIX Magazine, 33(5)

Primary Contribution:
The primary contribution of the paper written by Walker consists of the benchmarking the Amazon EC2 system for high-performance scientific computing. Macro and micro benchmarks are used to compare the performance of a cluster composed of EC2 nodes and compared this against the performance of a cluster composed of equivalent processors available to the open scientific research community.

Key insight of the contribution:
As stated by the author Commercial Cloud computing for HPC may allow scientists an access to a custom budget based suitable service without the overhead of cluster maintenance. However, according to Walker the delivery of HPC performance with commercial cloud

computing services such as Amazon EC2 was not yet mature in 2008. A significant performance gap between performing HPC computations on a traditional scientific cluster and on an EC2 provisioned scientific cluster has been shown in the article. This performance gap can be found in the MPI performance of distributed-memory parallel programs as well as in the single compute node OpenMP performance for shared-memory parallel programs. The author points out that vendors will need to upgrade the service offerings, especially in high-performance network provisioning, so that it may become a viable alternative to scientific clusters.

Rating:

The paper is well-structured and has a clear focus on the benchmarking tasks, well known and relevant benchmarking tasks have been used for the comparison. Unfortunately, since the paper is from 2008, it cannot be assumed that the results still yield. Cloud computing has received a significant amount of attention during the last decade and due to the tough competition between AWS, Azure and GCP, the services had to become more affordable and offer a higher performance. The paper receives an “accept” due to its relevant takes on the AWS service for HPC, although a new analysis might be necessary.

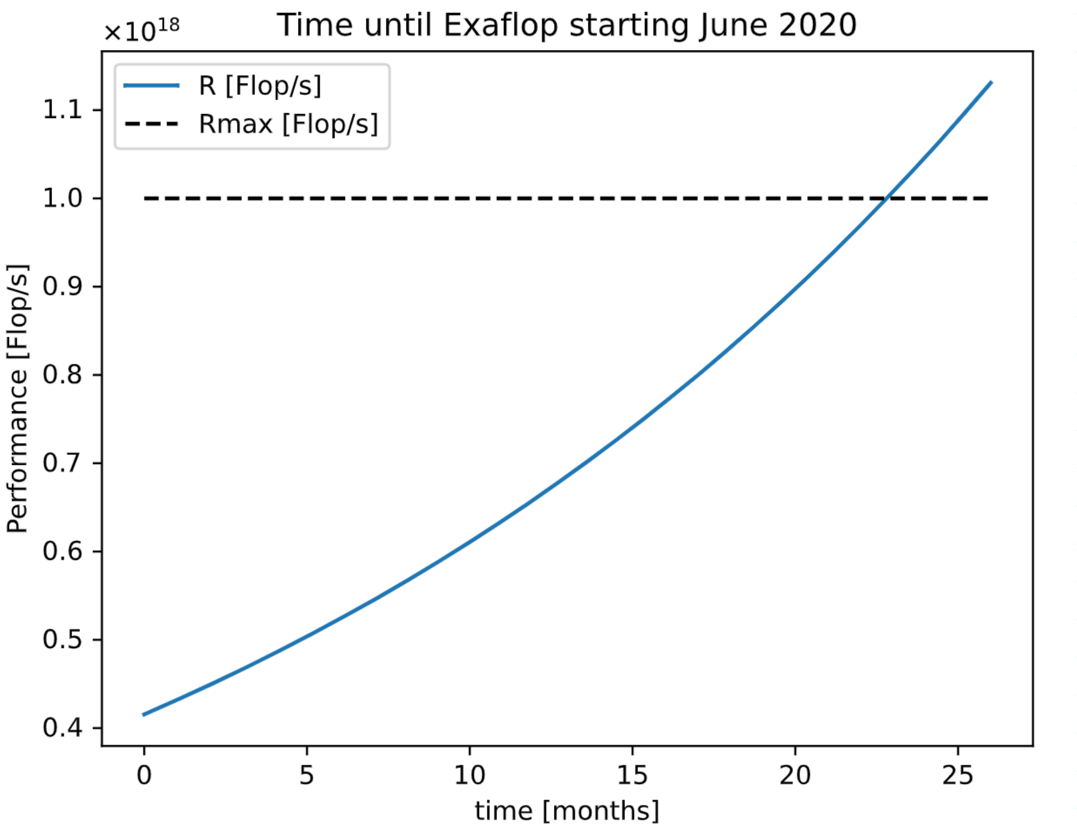
1.2 Moore's Law

1.2.1

double performance every 18 months: $R(t) = N_0 \cdot 2^{\frac{1}{18} t}$

where N0 is 415.53 PetaFlops/s (Supercomputer Fugaku) (1) data from 06/2020 $N_0 := 415.53 \cdot 10^{15}$ "Flops/s"

solve for t, when performance reaches one Exaflop $10^{18} = N_0 \cdot 2^{\frac{1}{18} t} \xrightarrow{solve, t \; float, 3} 22.8$ "months"



summary: the datapoint June 2020 was choosen, because it is the first appearance of the supercomputer Fugaku. Ttaking this data in consideration, it is expected to reach the one Exaflop era after 22.8 months, concluding in the year 2022.

Checking the result against available data, it shows that the one Exaflop threshold was indeed achieved in june 2022 by the supercomputer Frontier

1 https://www.top500.org/lists/top500/2020/06/
2 https://www.top500.org/lists/top500/2022/06/

1.2.2

11/2007 Blue Gen	478.2 TFlops/s 3
11/2011 K computer	10510.0 TFlops/s 4

growthrate: $100 \cdot \frac{10510 - 478.2}{478.2} \xrightarrow{float,3} 2097.0$ "%"

exponential growing process: $R(t) = N_0 \cdot a^{\lambda \cdot t}$

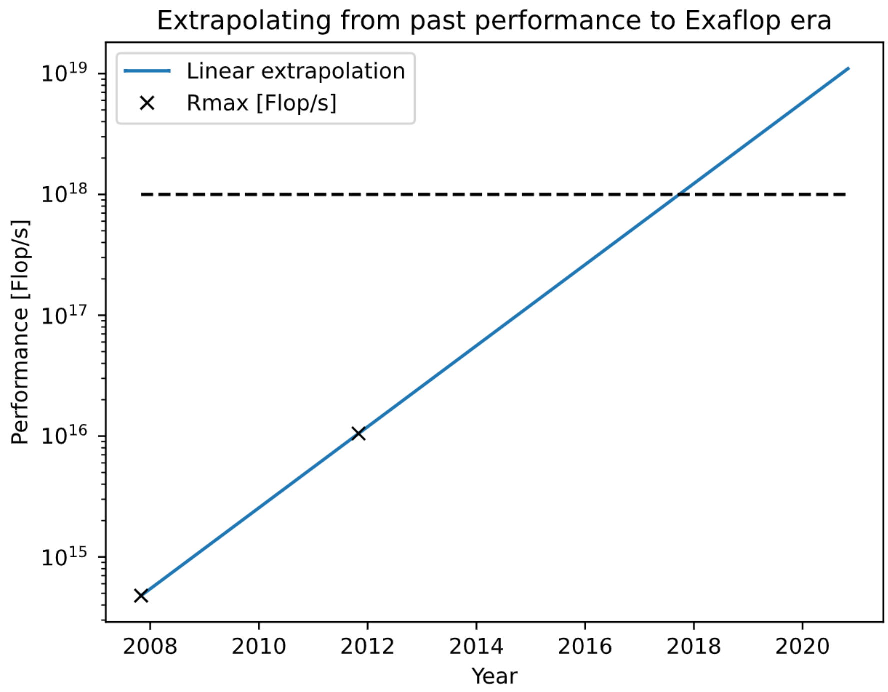
start value: $N_0 := 478.2 \cdot 10^{12}$ "Flops/s"

second datapoint: year after start and value $t := 4$ "year" $R := 10510 \cdot 10^{12}$ "Flops/s"

solve for λ: $\lambda := R = N_0 \cdot a^{\lambda \cdot t} \xrightarrow[float,3]{solve,\lambda} \frac{0.773}{\ln(a)}$

calculate the amount of years to reache one Exaflop/s $10^{18} = N_0 \cdot a^{\lambda \cdot t_{exa}} \xrightarrow[float,3]{solve,t_{exa}} 9.89$ "years"

summary: extrapolate from give data, the one Exaflop/s threshold will be overcome after 9.89 years. The calculation shows that the one ExaFlop/s era is predicted for End of 2017. It need to be mentioned that extrapolation is not a robust method.



3 https://www.top500.org/lists/top500/2007/11/
4 https://www.top500.org/lists/top500/2011/11/

1.3

1) New CPU 10 times faster

$S = 60\%$ 10

$P = 40\%$ calculations

$N = 10$

old system:

$$a = \frac{1}{S + \frac{P}{N}} = \frac{1}{(1-p) + \frac{p}{N}}$$

$$= \frac{1}{0{,}6 + \frac{0{,}4}{10}} = 1{,}563$$

Increase in performance of 1,563.

2) 20% time for calc of square root. $\rightarrow p = 0{,}2$

Options
(1) improve implementation of square root by 10
(2) improve everything by 1.6

50% of execution time is spent on FP operations

Option (1)

$$a_1 = \frac{1}{(1 - (\underline{0{,}5 \cdot 0{,}2}) + \frac{0{,}5 \cdot 0{,}2}{10}} = 1{,}0989$$

assuming that calculations take 50% of execution time and of that 20% are FPSQR.

Option (2)

$$a_2 = \frac{1}{(1 - 0{,}5) + \frac{0{,}5}{1{,}6}} = 1{,}2308$$

Accelerating all FPs by 1,6 speeds up everything by 23,08%
This is the optimal solution compared to 9,89%

3) $N = 128$

$a = 100$

$$a = \frac{1}{(1-p) + \frac{p}{N}} \Rightarrow 100 = \frac{1}{(1-p) + \frac{p}{128}}$$

$$a = \frac{1}{1 - \frac{127p}{128}}$$

$p = 0,9978$

$s = 1 - p = 0,0022$

The serial fraction s must be $<= 0,22\%$

1.4

Willing to present all three exercises.