



HAL
open science

Low Energy Consumption on Post-Moore Platforms for HPC Research

Pablo Rojas, Carlos Jaime Barrios-Hernandez, Luiz Angelo Steffenel

► **To cite this version:**

Pablo Rojas, Carlos Jaime Barrios-Hernandez, Luiz Angelo Steffenel. Low Energy Consumption on Post-Moore Platforms for HPC Research. Advanced and High Performance Computing Trends in Latin America Workshop (CARLA), 2020, Cuenca, Ecuador. hal-02925055

HAL Id: hal-02925055

<https://hal.univ-reims.fr/hal-02925055v1>

Submitted on 28 Aug 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Low Energy Consumption on Post-Moore Platforms for HPC Research

Pablo Josue Rojas Yepes^{13[0000-0003-1823-7922]}, Carlos Jaime Barrios Hernandez^{13[0000-0002-3227-8651]}, and Luiz Angelo Steffanel^{2[0000-0003-3670-4088]}

¹ Universidad Industrial de Santander, Cl. 9 # Cra 27, Bucaramanga, Colombia

² Université de Reims Champagne Ardenne, CReSTIC Laboratory, Reims, France

³Supercomputación y Calculo Científico UIS, Colombia

cbarrios@uis.edu.co

pablo2198162@correo.uis.edu.co

angelo.steffanel@univ-reims.fr

Abstract. The increase in computational capacities has helped in the exploration, production and research process, this has allowed the use of applications that were infeasible years ago. This increase brings us a new Era (known as Post-Moore Era) and a wide range of promising devices, devices such as Single Board Computers (SBC) or Personal Computers (PC) that achieve performance that a decade ago was only found on a Server. This work presents high performance computing devices with low monetary cost and low energy cost that meet the needs for the development of research in Artificial Intelligent (AI) applications, in-situ data analysis and simulations that can be implemented on a large scale, these devices are compared in different tests, presenting advantages such as its performance per watt consumed, smart form, among others.

Keywords: Edge Computing, Embedded Systems, Internet of Things, Many-core and Heterogeneous Computing, Low Cost Computing.

1 Introduction

Two of the largest projects in Latin America have been the Fênix [1] supercomputer and the Santos Dumont Hybrid [2] supercomputer, both of which were manufactured by Atos but Fênix was commissioned by Petróleo Brasileiro S.A. to be used in the oil and gas industry, while the Santos Dumont was built for the Laboratório Nacional de Computação Científica for the academic segment and meets in the post 476 of the TOP500 list. Fênix is among the top three on the continent and at number 195 on the TOP500 list, provide a theoretical capacity of 4,297.42 *TFlop/s*, its Linpack result is 1,836 *TFlop/s* with a consumption of approximately 287 *kW*. These capabilities have contributed to geophysical data processing using complex algorithms to generate images that are essential in oil exploration and production processes. With these computational capabilities will be possible to provide higher resolution images, reducing operational and geological risks directly impacting the profitability of projects. The Fênix project demonstrate the great benefit that is obtained when investing in techno-

logical tools, but developing countries cannot always invest in these options, therefore, it is necessary to look for alternatives that can offer us decent computing capabilities that, in turn, give us a good cost/benefit ratio.

The evolutionary way that hardware has traveled until today has been ups and downs, as well as long, but exponentially beneficial. Moore's law [3] was the guide for planning, development and investigation for decades. Numerous advances have arisen under your guide. This trigger an improvement in the performance thousands of millions of times. The problem is that nothing is forever and leads us to find ourselves at the gates of the end of one of the main boosters of this advance. Thus, in 2016, the roadmap for semiconductor technology no longer focused on "Moore's law", it ends is expected for the period 2025-2030 [4]. This new path describes how application needs drive the advancement of technology. This leads to an "explosion" [5] of notorious computing capabilities today, algorithms that were not viable (Artificial Intelligence, Autonomous and deep learning), are at the top of research and are the ones that generate the technological evolution of countries. At the same time, there is an abundance of hardware with high computing capacity, low power consumption and an affordable price.

This paper focuses on the energy efficiency of low cost devices, there are different techniques and tools such as commands, monitors, hardware, and benchmark suites that can help with these issues. Thus, the first topic to be exposed will be the context in which computing is currently, followed by referent devices, techniques and tools, to finally present the results obtained from the tests carried out.

2 Post-Moore Era

Any attempt to overcome the limitations presented by Moore's law can be considered as Post-Moore [4]. Moore's law says that around every two years the number of transistors in a microprocessor doubles, in addition to this, the improvement was characterized by an increase in the clock frequency, a decrease in lithography [4], but limitations such as size the lithography or quality of silicon have caused Moore's law to begin to falter. This end has generated new research paths such as multicores or heterogeneous computing. Furthermore, these emerging architectures do not focus on a single chip but on the combination of multiple chips [4], this allows these architectures to adapt to the needs of each application, generating a sudden appearance and rapid diversification of the hardware. In order to achieve this integration, strategies such as Heterogeneous System Architecture (HSA), languages such as CUDA, OpenCL or OpenACC, unification of CPU, GPU, FPGA, NPU, non-volatile 3D memory, reconfigurable communication grids, inductive wireless couplings, among others in a SoC [6, 7, 8, 9], must be taken into account.

As discussed above, the investment in computing resources is helpful for research and development of countries, but developing countries do not have as one of its priori-

ties, creating shortages in resources. To reduce research costs can make or buy low cost devices such as the Single Board Computer (SBC) or computer equipment heterogeneous desktop (PC) that can meet the needs of the applications used in research. Added to this, the applications must be developed or transcribed with the objective of being implemented in these architectures as a test base, these applications must maintain a portability and scalability that allows them fluidity and flexibility of deployment on architectures with greater capabilities. To verify the performance of different architectures, two reference devices are chosen that classify in the characteristics proposed by the Post-Moore Era, these devices will be treated in the next section.

3 Reference Devices

In order to compare performance, various devices are chosen to perform the experiments and compare the results. After several searches, there are several candidates, among which stand out: Raspberry Pi, Orange Pi, Asus Tinker Board, Odroid and Jetson. Of all these proposals, the Jetson family stands out above all. The Jetsons are SBCs developed by Nvidia, in this group there are cards such as TK1, TX2 or Xavier, even so, their costs are somewhat high or their production was suspended, but not everything is lost, therefore, from this family we chose to younger brother Jetson Nano.

The Nano features provide us with an ARM CPU, a Nvidia GPU. The other architecture chosen is somewhat more traditional, the features it provides us with are a Ryzen CPU, GPU Nvidia. The components that the experiments are going to focus on are the CPU and GPU, so it is good to give them a little review.

3.1 The CPUs

ARM Cortex-A57 is primarily composed of a finder, decoder, and instruction dispatcher, integer executor, load / storage unit (L1), L2 memory system, floating point unit, advanced SIMD, generic CPU interrupt control interface, generic timer, debug and trace.

AMD Ryzen 5 3600 is manufactured in a Zen 2 microarchitecture, this design is given around small 8-core chiplets separated into 2 groups of 4, this allows the cores to be organized in a "central complex" or CCX, which contains the 4 cores and an L3 cache set. Regardless of the number of chiplets, it paired with a central I/O via Infinity Fabric, this I/O acts as a central hub for all off-chip communications, as it houses all PCIe lanes, memory channels, and Infinity Fabric links to other chiplets or CPUs. This separation greatly benefits its scalability and manufacturing capabilities, plus it makes it easier to build processors with lots of cores.

3.2 The GPUs

The selected GPUs are from successive architectures, Maxwell and Pascal.

Maxwell focused primarily on energy efficiency, his SM (Streaming Multiprocessor) was restructured, partitioned and renamed SMM. The structure of the warp scheduler, along with the FP64 CUDA core and texture unit was inherited from Kepler, but the design of most execution units was partitioned so that each warp scheduler in an SMM handles a 32-core FP32 CUDA package. This enables better resource management than Kepler, saving more energy when the workload is not optimal for sharing resources. Maxwell would be succeeded by the Pascal microarchitecture.

Pascal architecture improvements are based on five technological advancements, 1) a 16nm fabrication process that increases performance and improves energy efficiency, 2) increased double precision performance for HPC workloads (in Deep Learning offers more than 12 times of neural network training and a 7-fold increase in Deep Learning inference performance compared to previous generation GPU architectures), 3) it is the first architecture to integrate revolutionary NVIDIA NVLink™ bi-directional interconnect High Speed (this technology is designed to scale applications across multiple GPUs, delivering faster 5X acceleration of interconnect band), 4) using an innovative approach to memory design, CoWoS® (Chip-on-Wafer -on-Substrate) with HBM2 gives you a 3X boost in memory bandwidth performance over NVIDIA Maxwell™ architecture, 5) the new in 16-bit medium-precision floating-point instructions and the new 8-bit integer instructions allow AI algorithms to provide real-time responsiveness for Deep Learning inference.

Table 1. Specifications of the Reference Devices.

	Device 1	Device 2
CPU	ARM Cortex-A57 Quad-Core 64-bit @ 1.43 GHz	AMD Ryzen 5 3600 Hexa-Core 64-bit @ 3.6GHz
RAM	4GB LPDDR4 @ 1600 MHz	16GB DDR4 @ 3200 MHz
GPU	Nvidia Maxwell 128-Core @ 921 MHz	Nvidia Pascal 768-Core @ 1350~1800 Mhz (GTX 1050Ti)

With these descriptions, it gets an idea of the specifications that these devices offer. To measure performance, techniques and tools must be able to be implemented in the reference devices, this topic will be covered in the next chapter.

4 Techniques and Tools

As it was shown in the introduction, the computational capacities are immense, which requires large investments, for this fact, it cannot be directly compared to the devices chosen with these titans, based on the Post-Moore approach, more affordable partici-

pants can be chosen. In order to simplify the choice, these devices should have at least one CPU and one GPU, it should also be possible to measure their power consumption. In this way, its capabilities can be verified by means of benchmarks.

There are many benchmark tools, from the reliable Linpack (HPL) [10] to Phoronix Test Suite (PTS) [11], in terms of energy consumption, hardware or other measurement variants can be used to obtain this data in a controlled way.

4.1 Benchmarks

Stress-ng will test a computer system in various selectable ways. Execute a wide range of CPU specific stress tests that exercise floating point, integer, bit manipulation and control flow. It was designed to show thermal problems, errors in the operating system, this can result in device failure, because of this, care must be taken in the execution of the tests.

HPL (High-Performance Linpack) solves a random dense linear system in double precision (64 bits) arithmetic on distributed-memory computers. HPL provides a testing and timing program to quantify the accuracy of the obtained solution as well as the time it took to compute it. The generic implementations of MPI, the BLAS and VSIP are available for a large variety of systems.

PTS (Phoronix Test Suite) is a testing and evaluation platform. The software is designed to effectively carry out qualitative and quantitative benchmarks in a clean, reproducible and user-friendly way, it will take care of the entire testing process, from dependency management to test download/installation, execution and aggregation of results. PTS has access to more than 100 test suites through [13]. If there is a test that is not currently covered in PTS, they can be added through its extensible architecture.

4.2 Measure Power Consumption

As for measuring the power consumption, the aim is to generate the least possible load on the chosen devices, therefore, a measurement tool external to the device is chosen, which will measure the consumption second by second. For this task we choose a smart outlet of the VTA brand.

Table 2. Specifications of the Measure Power Consumption.

VTA Smart Outlet	
Wi-Fi	2.4 GHz
Operating Voltage	125 VAC 60Hz
Maximum Current	10 A
Maximum Power	1250 W

The smart outlet counts a mobile APP, which carries a monthly history consumption, but video of second-by-second consumption can be taken when testing is performed. The devices are thought to have the lowest possible cost and decent capacities, therefore, SBCs are the best option that can be found, followed by Personal Computers (PC), although their cost is not comparable to SBCs, they have higher power.

An SBC is a computer (CPU, RAM, GPU, etc.) on a single circuit or board. Its applications range from industrial environments to home IoT systems. Due to high component integration and small footprint, these devices feature a higher reliability, better power handling, less weight and an SBC can mass-produce to reduce its costs.

On the other hand, the PC has been heterogeneous for quite some time. In addition, over time, the cost of calculations per second has been decreasing, improving access to new hardware with greater power, allowing the development and implementation of countless applications. To measure the performance of each reference device, several tests must be carried out, the process of carrying out these tests is presented in the next section.

5 Experiments

The tests perform two branches of experiments to measure the behavior of the chosen devices, the first branch focuses on the use of the CPU, the second on the use of the GPU. For CPU testing, Stress-ng and HPL were nominated. Stress-ng offers us a variety of tests such as operations with floating numbers, integers, random numbers, matrices, among others. The following configuration is used for this test:

stress-ng --cpu N --cpu-load P --cpu-method method --metrics-brief --timeout T

Where,

--cpu N starts N jobs by stressing the CPU.

--cpu-load P loads the CPU with a percentage P of load for stress jobs on the CPU. Accuracy depends on overall processor load and planner responsiveness, so the actual load may be different from the desired load. Also, the number of bogo operations may not scale linearly with load, as some systems employ CPU frequency scaling, and therefore heavier loads result in higher CPU frequency and higher bogo operations.

--metrics-brief enables metrics and only shows non-zero metrics.

--timeout T stops the stress test after T seconds. You can also specify units of time in seconds, minutes, hours, days, or years with the suffix s, m, h, d, or y.

--cpu-method method specifies a CPU stress method. By default, all stress methods are exercised sequentially, however, only one method can be specified to use if necessary. Some methods were selected to measure the behavior of the CPU:

- **cfloat** are 1000 iterations of a combination of complex floating-point operations.

- **correlate** make a correlation of random doubles: 16384×1024 .
- **union** performs integer arithmetic on a combination of bit fields in a C union. This shows how well (the compiler and the CPU) can perform loading and storing of integer bit fields.
- **hyperbolic** calculates $\sinh(\theta) \times \cosh(\theta) + \sinh(2\theta) + \cosh(3\theta)$ for hyperbolic sine and cosine functions on float, double and long double, where $\theta = 0$ a 2π in 1500 steps.
- **prime** finds all prime numbers in the range 1 to 1,000,000 using a slightly optimized brute force search.
- **matrixprod** is a matrix product of two 128×128 double floating matrices. Testing in hardware shows that this provides a good combination of memory, cache, and floating-point operations and is probably the best method of measuring CPU performance.

Bogo ops are the number of iterations of the stress test during the race. This is the metric for how much overall "work" has been accomplished in stress operations.

In addition to the Stress-ng tests, HPL tests are performed. HPL is software that solves a dense, random linear system with double-precision (64-bit) arithmetic on computers with distributed memory. For the HPL test, it is configured as presented in the Table 3:

Table 3. Configuring the HPL parameters.

		Parameter			
	Device	N	NB	P	Q
Test 1	Jetson Nano	11584	64	1	4
	PC	23168	64	1	4
Test 2	Jetson Nano	17376	96	1	4
	PC	34752	96	1	4

In Test 1, the HPL.dat was configured to use $\pm 50\%$ of the RAM capacity, Test 2 only uses 75% of the total capacity, it was configured in this way to take advantage the capabilities that each device offered.

In the next branch we have the GPU usage tests, in the GPUs tests are carried out with different tests that use OpenGL, CUDA, among others. The tests consist of: building a terrain in a random way, simulating a colloid in a liquid medium and the CUDA n-body test, using the PTS as a benchmark tool for this test. The tests are performed at 1080p for the Simulation-Visualization cases and with the power limiters disabled for all the tests.

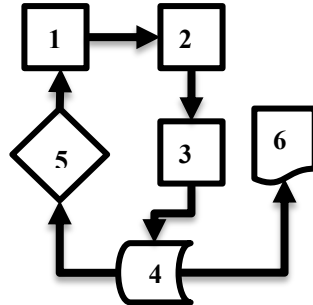


Figure 1. Diagram of General Test Workflow.

In order to better explain the tests, the Diagram of General Test Workflow is presented (Figure 1), this diagram presents six steps and is described below:

5.1 Set up the different test requirements

As explained in this same section, each test is different and you can modify values such as the duration time or the percentage of the test load, in addition, it can modify the OS configuration files to increase the Hertz or the models of consumption of the devices, for this reason it is necessary to configure the test.

5.2 Configure the energy consumption monitor

The consumption monitor is always active and presenting data, therefore, after configuring the test we must find a way to capture the data at the time of the test, if possible we should look for monitors that keep a record of their measurements, if it's not possible, we can capture on video or by hand the behavior of the monitor during the test.

5.3 Start energy monitoring and launch test

Once the test is configured and it is ready to capture the data from the monitor, we proceed to run the test, we must be very aware of the beginning and end of the test since at this time is when the probability of failure is highest.

5.4 Store and label test results

This step is as important as the first step, this is due to the labeling, almost always when it is done wrong, it results in confusion when comparing the results of each device and the test should be repeated, if possible, the label should bear the name of the device, the test carried out, % of workload, its duration or the consumption model.

5.5 Repeat the test to average the results or start a new test

This step is a bifurcation, in certain cases it is necessary to have greater certainty in the results, it is good to repeat the test several times and vary its configuration, this allows us to see patterns of behavior in the devices, the test can also be changed, these Modifications can show problems like the thermal bottle. collars, memory saturation, poor performance when loading or storing data types, etc.

5.6 Group the results and generate the graphs of the tests

Based on the labels, the data is processed to generate the graphs. For this document, the number of operations (Ops) per second performed was taken together with the number of watts to generate the Ops/W.

With this description, the following section shows the results obtained in each of the tests carried out for the different configurations.

6 Results

In the previous section, the tests carried out for the analysis of the two reference devices were explained, it should be clarified that the ARM Cortex-A57 is presented with the A57 label, the AMD Ryzen 3600 uses the 3600 label. As for GPUs, the tag used for the GPU of the Jetson Nano is Nano and the GP107 tag is used for the GTX 1050ti. Each test is performed five times per core (5 x 1 core, 5 x 2 core, etc.), the results presented are the average of the values obtained in each test. The first tests to be carried out are those carried out on the CPU. The first test performed is Cfloat, the unit of measurement will be Operations per watt (Ops / W), as indicated above, Cfloat is 1000 iterations of a combination of complex floating-point operations. The completion of these 1000 iterations is counted as an Operation (Ops), the following graph (Figure 2) shows the results of this test:

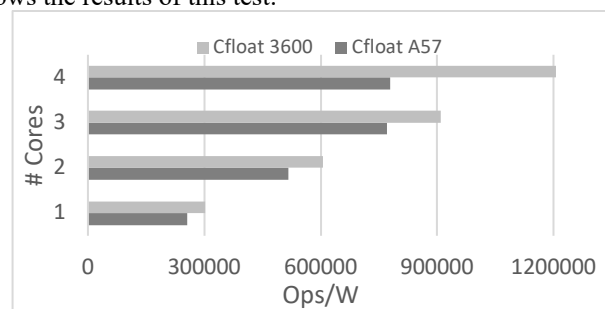


Figure 2. Cfloat test on CPU.

As we can see in Figure 2, the 3600 has superior performance and better scalability in this test than the A57, it should be noted that the A57 suffers from a bottleneck when using all its processors for the task, which makes scalability growth flattened out. The

second test run is Correlation, an Ops consists of performing a random double correlation (16384×1024). We see the results in Figure 3.

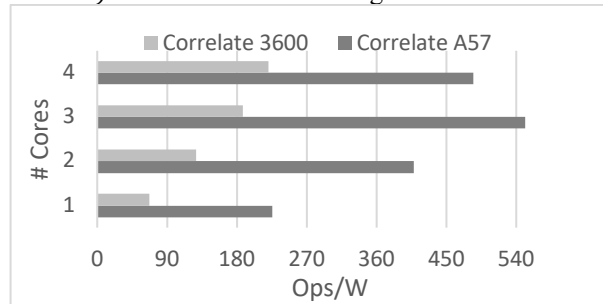


Figure 3. CPU Correlation Test.

In this test, the A57 stands out notably in its performance, but it is still affected by the bottleneck, even so, it almost triples the performance of the 3600. The third test is union; the results are shown in Figure 4.

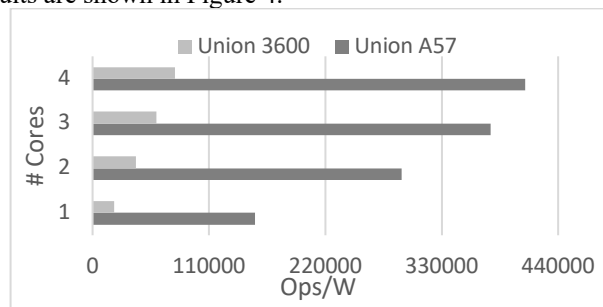


Figure 4. CPU Union test.

The A57 shows an excellent handling (Compiler-CPU) of the loading and storage of bit fields, which is the strength of this test, almost quadrupling the performance of the 3600. The fourth test carried out is Hyperbolic, it calculates hyperbolic sine and cosine functions with 1500 steps in each Ops, Figure 5 shows these results.

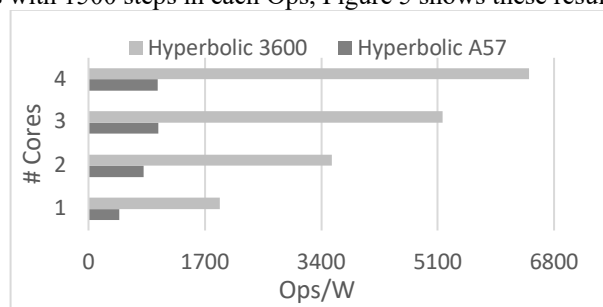


Figure 5. CPU Hyperbolic Test.

The 3600 offers excellent performance for this test, showing great performance for hyperbolic operations with float, double and long double, almost five times more than the A57. The fifth test is a calculation of the primes between 1 and 1,000,000, this

calculation is carried out by means of brute force, once it is finished, it compares the results and if the results coincide, it is marked as an Ops, the results are seen in Figure 6.

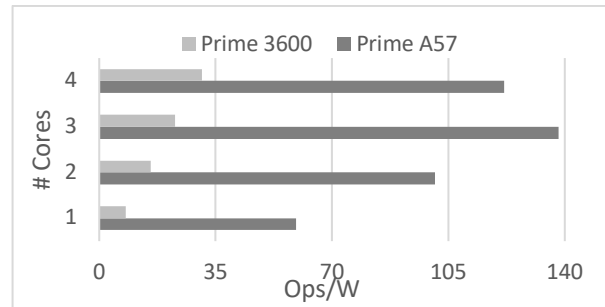


Figure 6. CPU Prime Test.

In this test, the A57 gives us superior performance per watt used, making it a good candidate for brute force tasks. The penultimate test performed is a multiplication of two 128x128 matrices, each element in the matrix is in a double precision floating point format, once the task is completed it is marked as an Ops, the results can be seen in Figure 7.

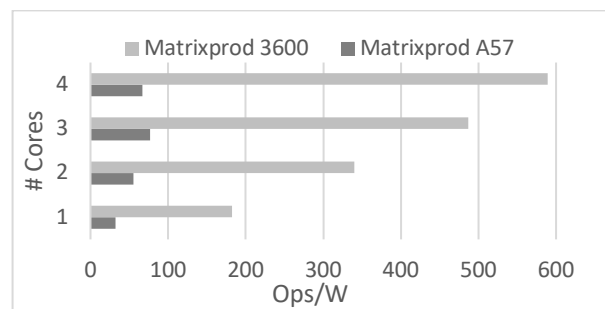


Figure 7. CPU Matrix Multiplication Test.

The Results shown by the 3600 overwhelm the A57, with this preliminary matrix multiplication test we move on to the next and last test, the last test is HPL. HPL is one of the best-known benchmarks to test the performance of a CPU, we see the results in Figure 8.

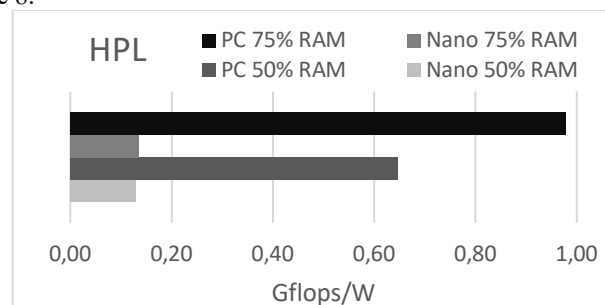


Figure 8. HPL for CPU.

The test was performed with two different loads and was done conservatively so as not to have the losses presented in the previous tests, as can be seen, the CPU 3600 achieves a performance of almost one Gflop per Watt consumed, after completing the task the 3600 has consumed almost 8 Watts while the A57 has consumed about 7 Watts. The CPU 3600 provides us with a greater computational force than the A57 but it should be noted that the AMD Ryzen 3600 works at a frequency of 3.6 Ghz, while the ARM Cortex-A57 works at a frequency of 1.4 Ghz among other differences.

The previous tests were carried out on the CPU, the next round of tests was carried out on the GPUs of the chosen devices, the GPUs are the GPU of the Jetson Nano (Nano) and the GTX 1050ti (GP107), Figure 9, 10 and 11 present the results. Terrain (Figure 9) is a simulation of a field that is generated randomly, this generation is measured by frames per second (FPS), during this test the ratio of FPS per Watt consumed during the task is shown, the Jetson Nano is far superior to that of the GTX 1050Ti (GP107).

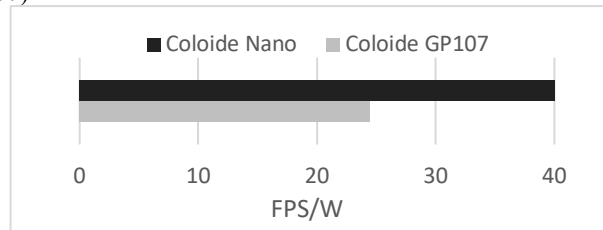


Figure 9. Colloid Simulation Tests.

The Colloid test has the same configuration as Terrain and presents its result in the same unit of measurement, Figure 10. shows the results. Colloid like Terrain shows us an excellent FPS/W ratio, almost double the FPS per Watt consumed, in addition these two simulations were performed at 1080p.

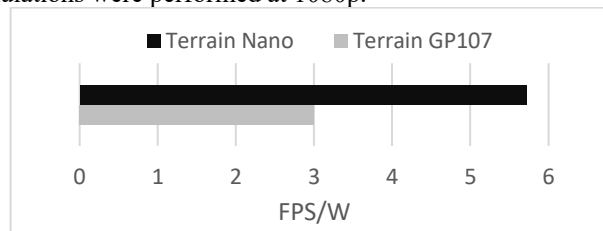


Figure 10. Terrain Simulation Test.

As the last test, a deployment of the Phoronix Test Suite tool is performed, this tool provides us with several benchmarks, Mini-Nbody is selected as the test benchmark, the result is shown in Figure 11.

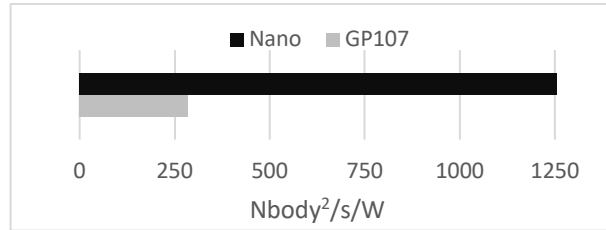


Figure 11. N-body test.

The Nbody per Watt consumed ratio is notably better in the Jetson Nano, it should be noted that each of the tests was performed in the maximum consumption mode for the Jetson Nano, as for the GP107, the configuration given by the manufacturer Asus in its ROG Strix model is maintained. Once these tests are performed according to Figure 1. Several conclusions are reached, in the next segment these and other conclusions that this work has generated are presented.

7 Conclusions

As discussed at the beginning, the objective of this work is to propose low-cost computational options that take advantage of the characteristics of the Post-Moore Era devices to increase the benefit / cost ratio. With the abundance of devices this era has brought, two different devices are selected to perform the same tests. SBCs are devices that stand out for their low cost and PCs for their wide use in different fields. Both offer computational capabilities that less than a decade ago were only found in servers or supercomputers, the investment in these devices is in most cases less than 5 ~ 10% of what it would be in equipment for HPC.

The results presented in the previous section show that SBCs like the Jetson Nano are a great option to be used as the computational force of an investigation. These devices provide researchers with computational strength on par with more traditional options. Due to its shape, we can take advantage of its energy efficiency to be implemented in different tasks and environments.

Both SBCs and PCs are excellent options from an economic point of view to encourage investment in research and development. The current PC offers us considerable computational strength at affordable prices, while SBCs take advantage of their low energy consumption, both options can be implemented in multiple tasks, achieving great performance.

In addition to this, the specifications of these devices make them essential for the development of applications designed to be scalable, portable, simple and efficient. Since, when developing an application for these devices, tests can be carried out with different amounts of data or their deployment on different platforms, improving the fluidity of the applications when implemented on servers or supercomputers.

8 Further Work

The most intense tests were performed on the CPU. For future work, a series of easy-to-implement, high computational effort tests will be developed on the GPU. In the first instance, the tests are written in CUDA, then be transcribed into other languages. Once this stage is complete, an implementation methodology will be proposed to measure the impact on performance when testing is performed non-natively (using methods such as containers).

References

1. TOP500 The List. Fênix - SYS-1029GQ-TRT, <https://www.top500.org/system/179681>, last accessed 2020/2/20.
2. TOP500 The List. Santos Dumont Hybrid - Bullx B710, <https://www.top500.org/system/178569>, last accessed 2020/2/20.
3. M. Waldrop.: The chips are down for Moore's law. *Nature*. 530 (7589): 144–147. DOI:10.1038/530144a. ISSN 0028-0836. PMID 26863965.
4. S. Matsuoka et al.: From FLOPS to BYTES: Disruptive change in High-Performance Computing towards the Post-Moore Era. In *CF '16 Proceedings of the ACM International Conference on Computing Frontiers*. 2016-05-16. ACM New York, NY, USA. DOI: <http://dx.doi.org/10.1145/2903150.2906830>.
5. S. Matsuoka.: Cambrian explosion of computing and big data in the Post-Moore era. In *HPDC '18 Proceedings of the 27th International Symposium on High-Performance Parallel and Distributed Computing*. 2018-06-11. ACM New York, NY, USA. DOI: <https://doi.org/10.1145/3208040.3225055>.
6. K. Barker et al.: On the feasibility of optical circuit switching for high performance computing systems. In *Proc. of IEEE/ACM SC 2005*, pages 16–16, 2005.
7. Take, Y., Matsutani, H., Sasaki, H., Koibuchi, M., Kuroda, T. and Amano, H.: 3D noc with inductive-couplings for building-block SiPs. In *IEEE Trans. on Computers*, pages 748–763. 63 (3), 2014.
8. Kagami, T., Matsutani, H., Koibuchi, M., Take, Y., Kuroda, T. and Amano, H.: Efficient 3-D bus architectures for inductive-coupling ThruChip Interfaces. In *IEEE Trans. on VLSI systems*, pages 493–506. Vol.24, No.2, Feb. 2016.
9. Inadomi, Y., Patki, T., Inoue, K., Aoyagi, M., Rountree, R., Schulz, M., Lowenthal, D., Wada, Y., Fukazawa, K., Ueda, M., Kondo, M., and Miyoshi, I.: Analyzing and mitigating the impact of manufacturing variability in power-constrained supercomputing. In *Proc. of IEEE/ACM SC15*, 2015.
10. HPL- A Portable Implementation of the High-Performance Linpack Benchmark for Distributed-Memory Computers, <https://www.netlib.org/benchmark/hpl/>, last accessed 2020/2/2.
11. Phoronix Test Suite, <https://www.phoronix-test-suite.com/>, last accessed 2020/2/4.
12. Stress-ng, <https://wiki.ubuntu.com/Kernel/Reference/stress-ng>, last accessed 2020/2/4.
13. Open benchmarking, <https://openbenchmarking.org/>, last accessed 2020/2/4.