Benchmarking a Many-Core System

UCI I-SURF 2017

Aug 25th, 2017 – Final Presentation

Professor: Nikil D. Dutt
Graduate Mentor: Bryan Donyanavard (Donny)

Kyuhan Lee, Sehee Jeong, Kihoon Han

About Term “Benchmark”

It calculates the performance and displays the result as a score.

Original Goal

Measuring the performance and power consumption related to each operation speed.

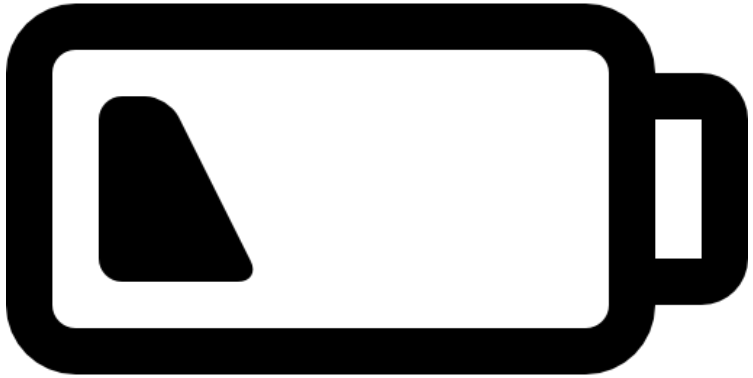
Additional Goal (If Applicable)

Calculate the power to performance ratio.

&

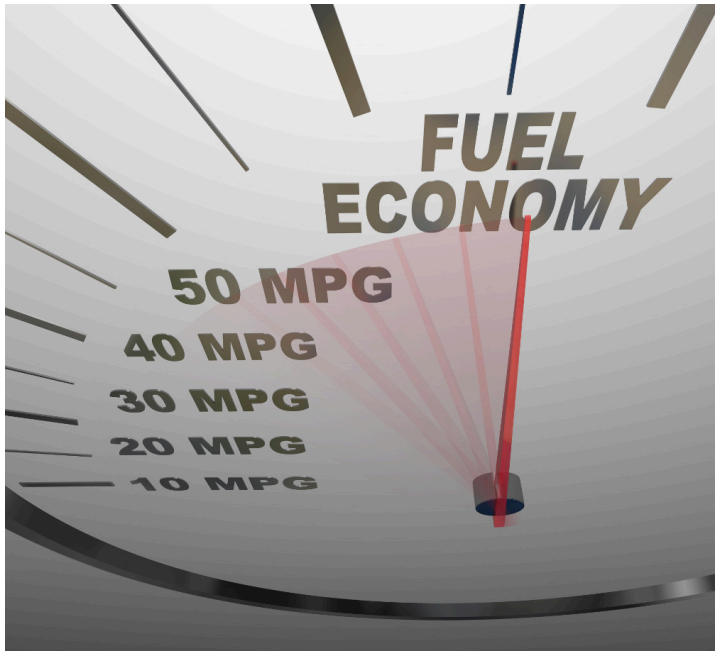
Find the most optimal one.

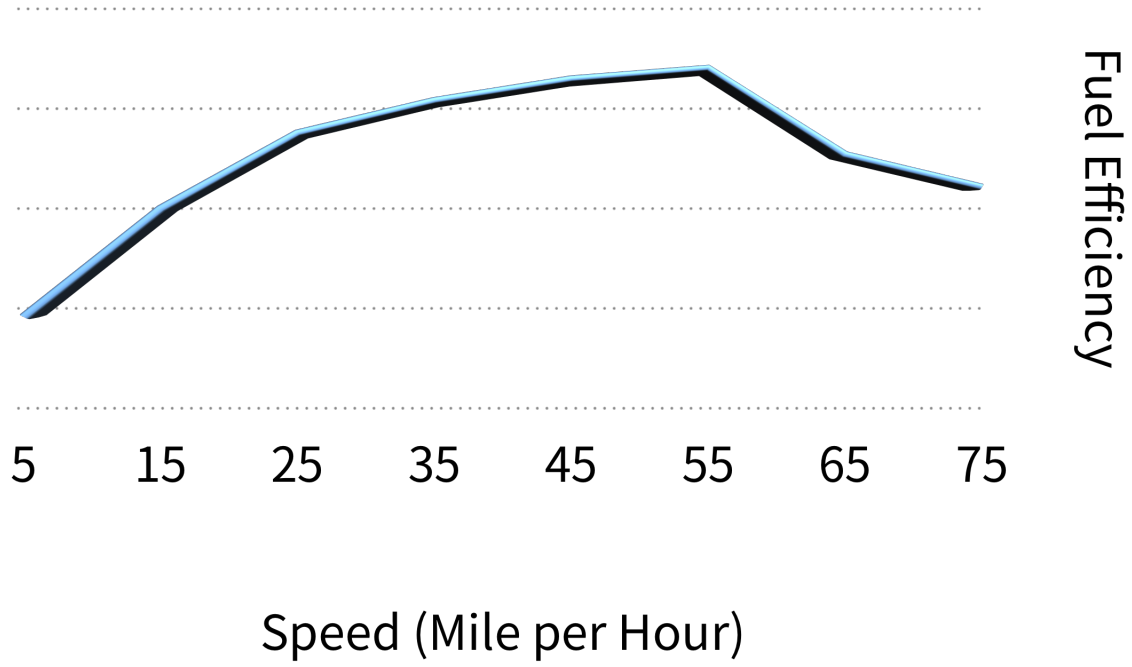
Why We Want To Find It?



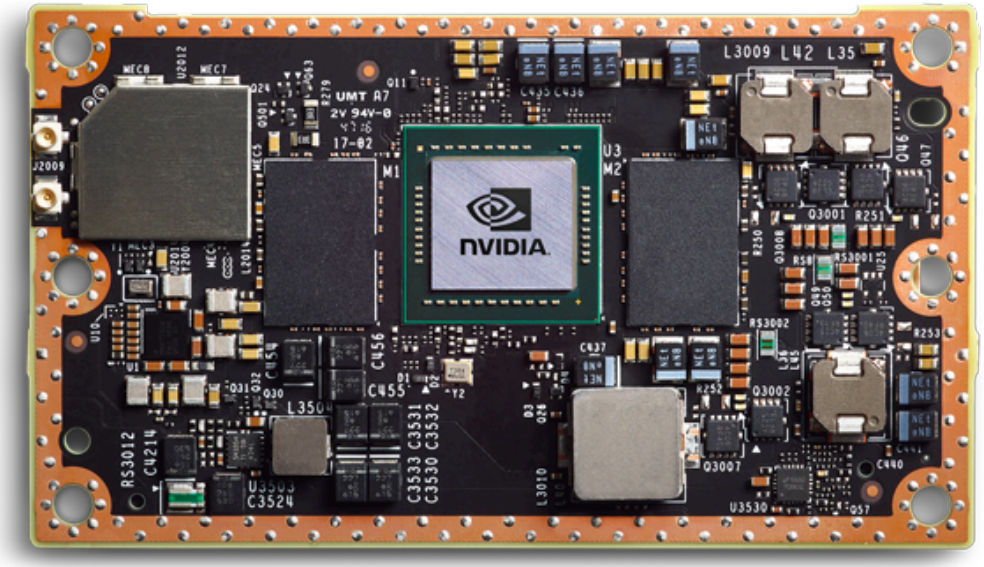
These days, there are a lot of systems that have limited power such as battery.

Example of Driving





Benchmark Environment (TX2)



- Ubuntu 16.04 LTS
- CPU with 6 cores
- GPU with 256 CUDA cores
- 12 Power Monitoring Units

How To Adjust Performance? (TX2)

```
$ cat [cpu frequency] > /sys/devices/system/cpu/cpuN/cpufreq/scaling_max_freq
```

(0 ≤ N ≤ 5)

```
$ cat [gpu frequency] > /sys/devices/17000000.gp10b/devfreq/17000000.gp10b/max_freq
```

We can use DVFS (Dynamic Voltage and Frequency Scaling).

In TX2, we can adjust performance using above command.

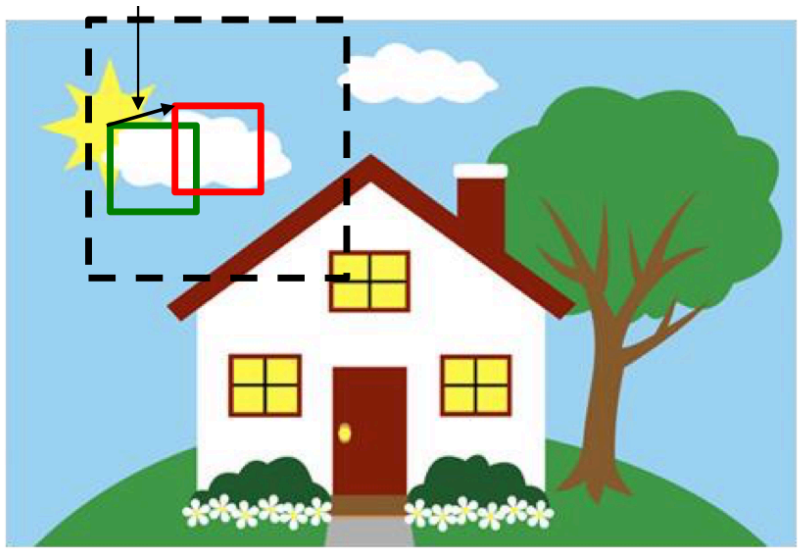
How To Measure Power Consumption? (TX2)

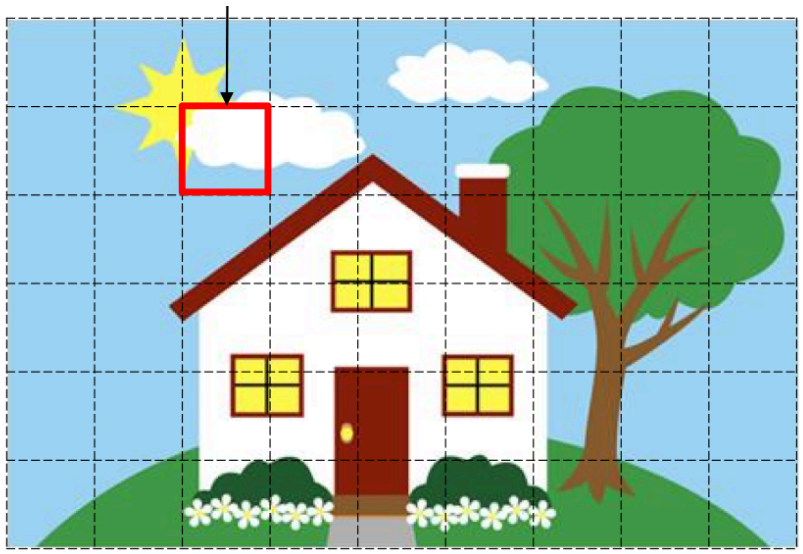
```
$ cat /sys/devices/3160000.i2c/i2c-0/0-004N/iio_device/in_powerM_input
```

(0 ≤ N ≤ 3, 0 ≤ M ≤ 2)

We can measure power consumption via I2C using above command.

How To Benchmark? (TX2)





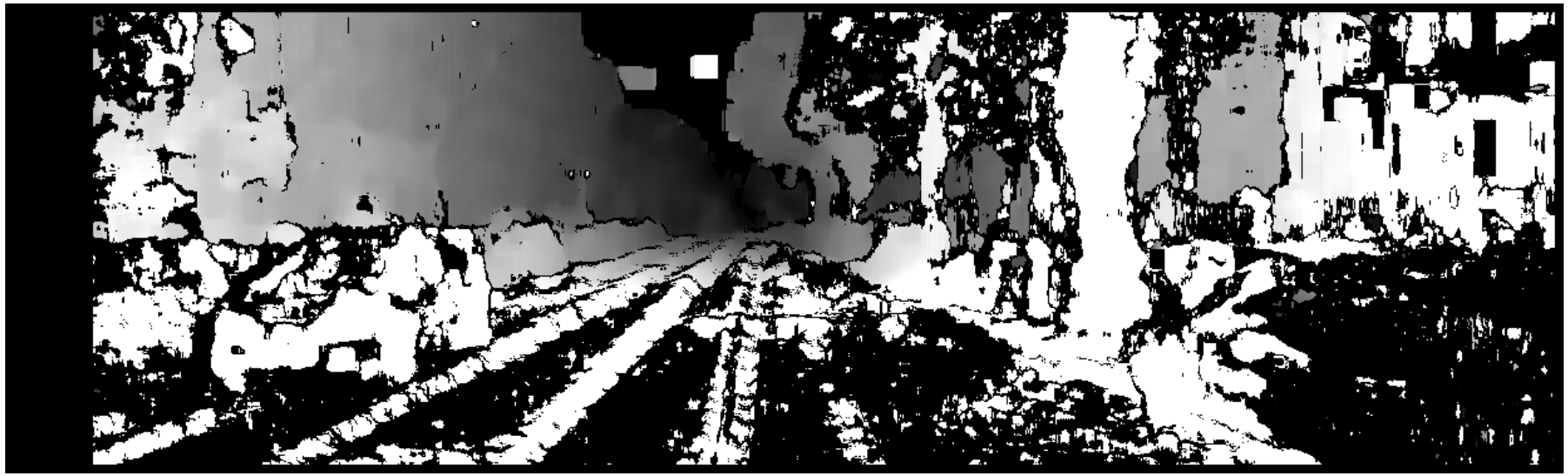
I used Stereo Block Matching algorithm.

This algorithm can run on both CPU and GPU, and generates equal results.

Benchmark Prodecure (TX2)

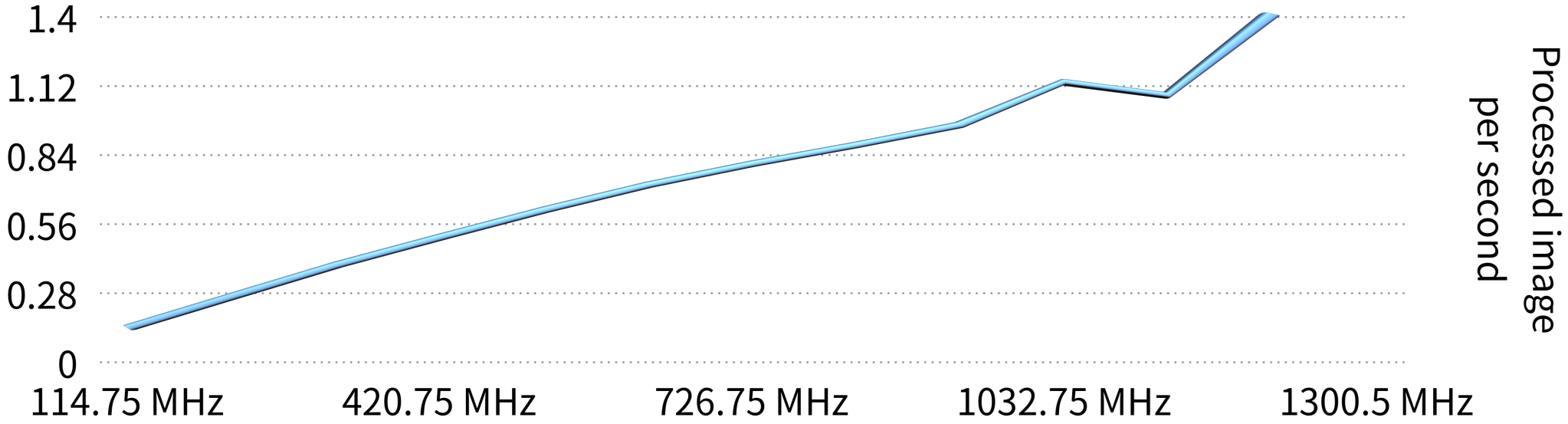
Using two 6000*4000px resolution images.

Benchmark Prodecure (TX2)



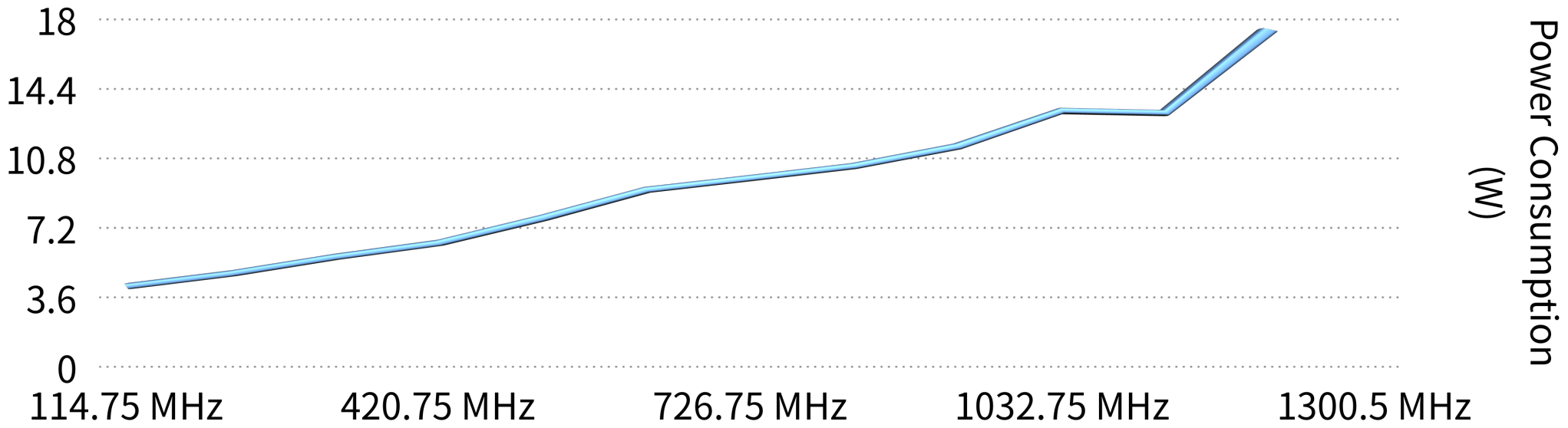
Generates result image using StereoBM algorithm.

Benchmark Results (TX2)



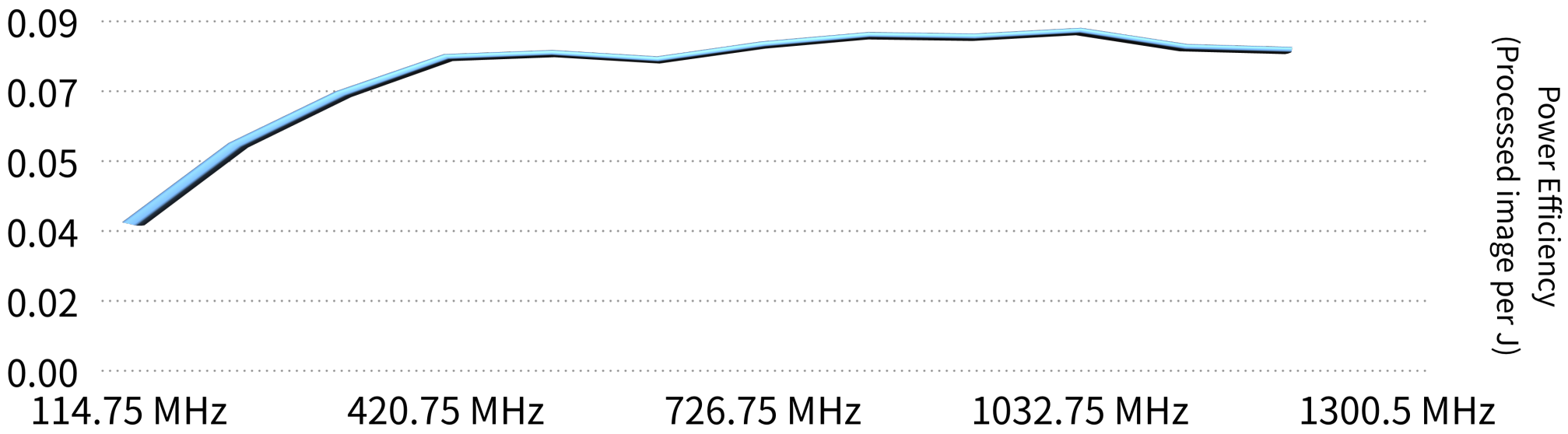
Performance of GPU according to the frequency of it

Benchmark Results (TX2)



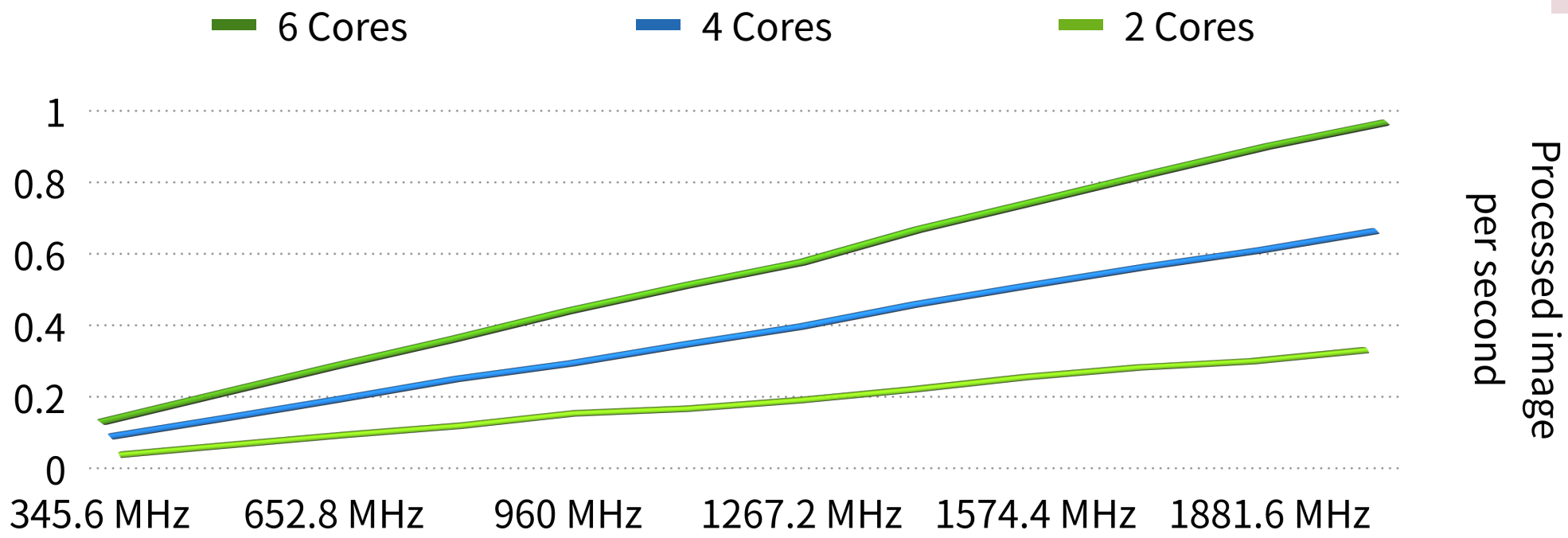
Power Consumption of GPU according to the frequency of it

Benchmark Results (TX2)



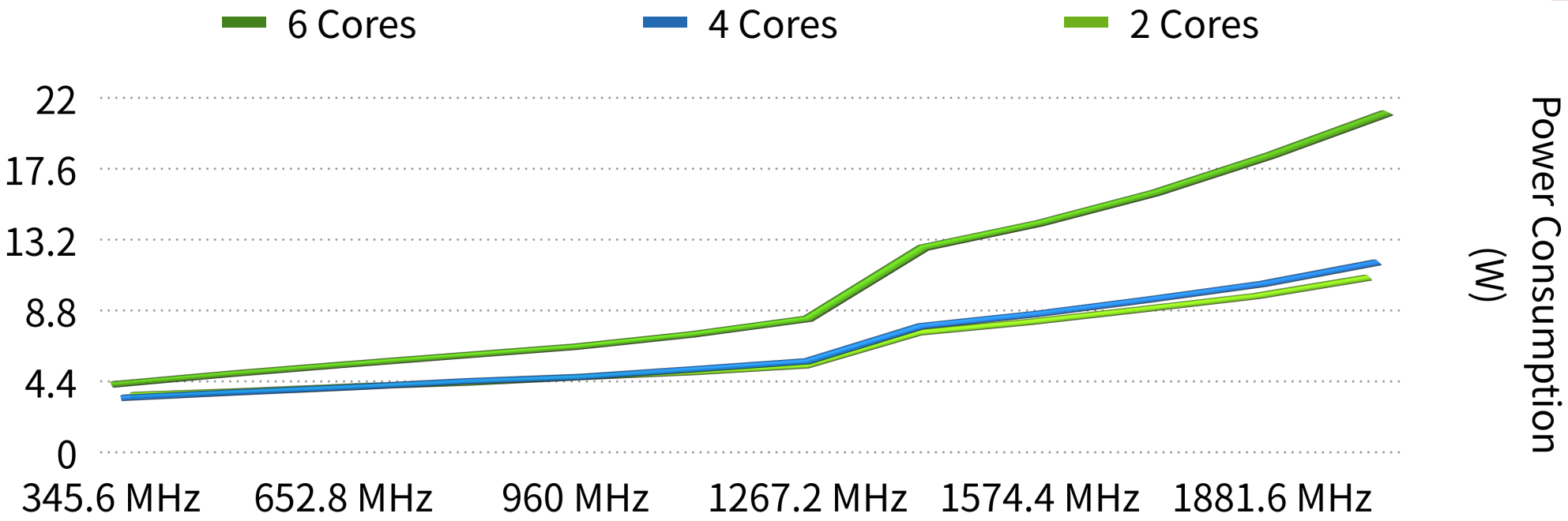
Power Efficiency of GPU according to the frequency of it

Benchmark Results (TX2)



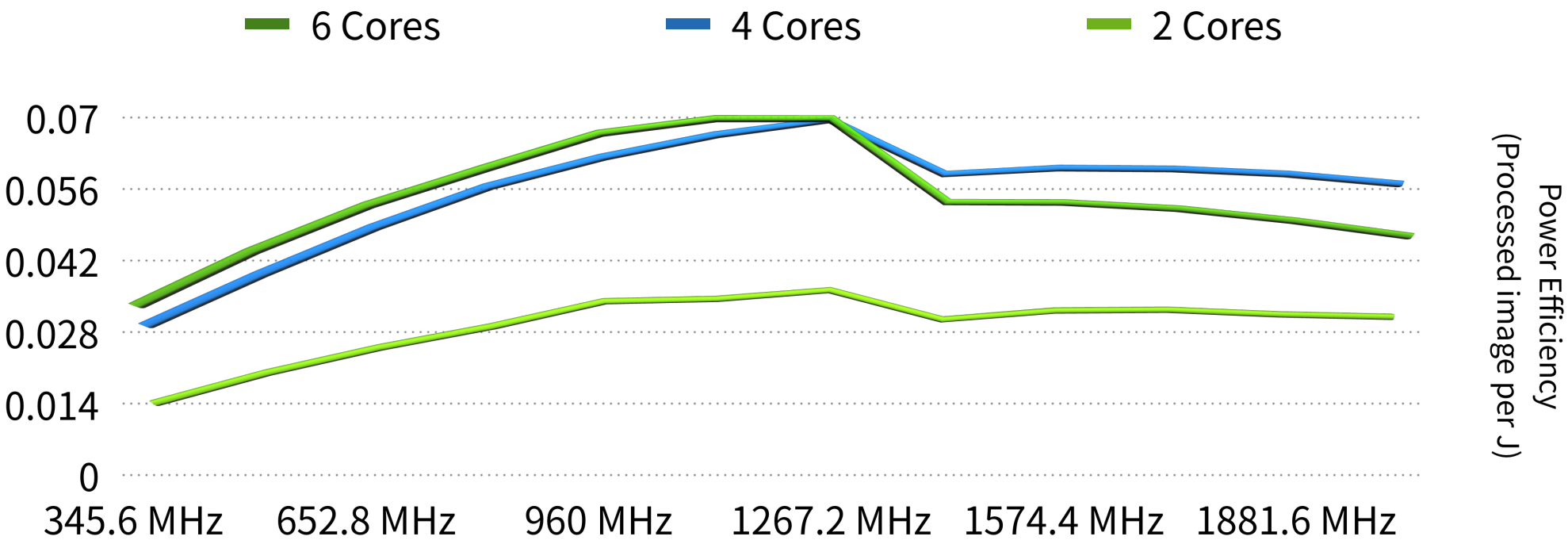
Performance of CPU according to the frequency of it

Benchmark Results (TX2)



Power Consumption of CPU according to the frequency of it

Benchmark Results (TX2)



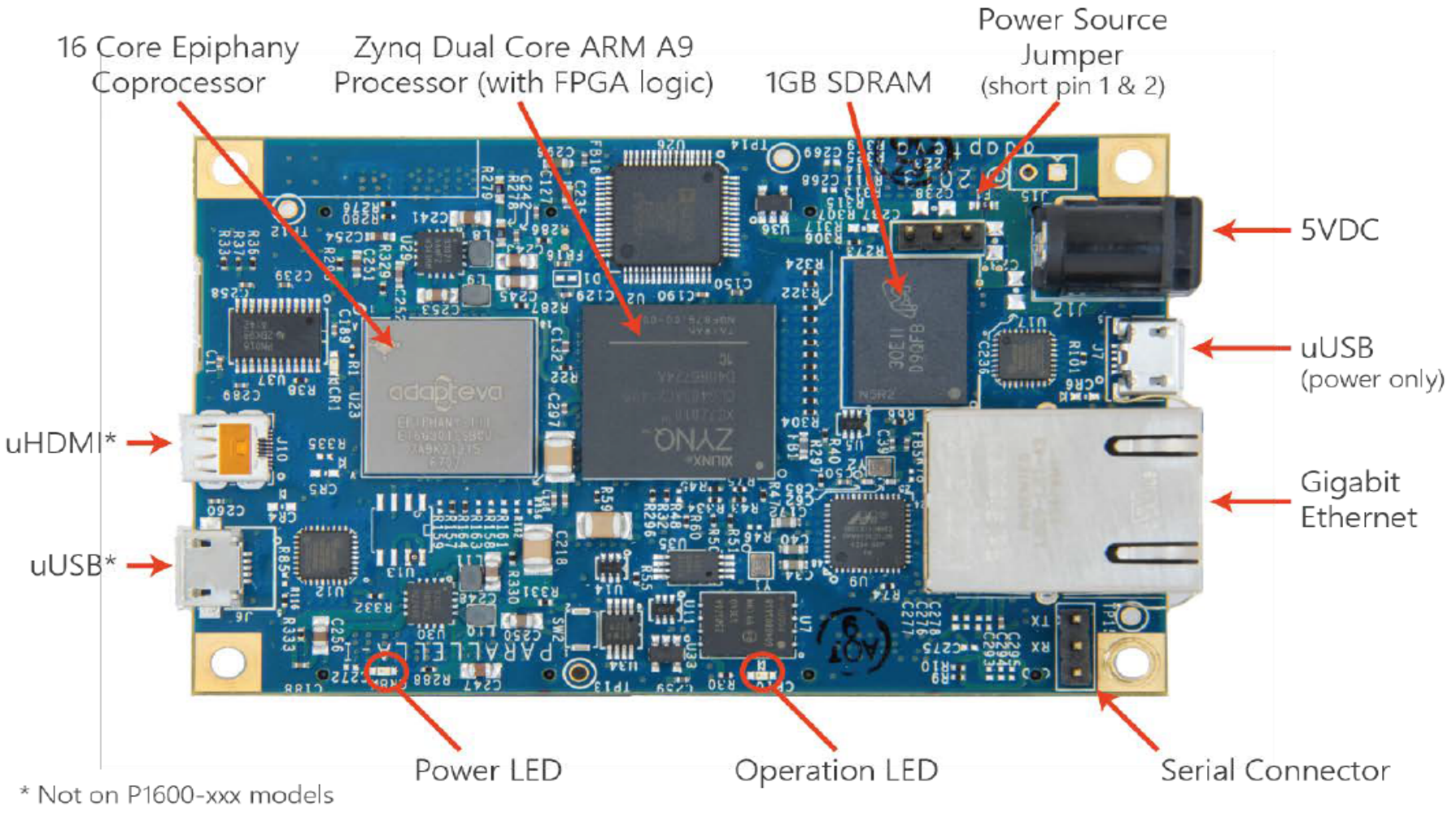
Power Efficiency of CPU according to the frequency of it

Benchmark Results (TX2)

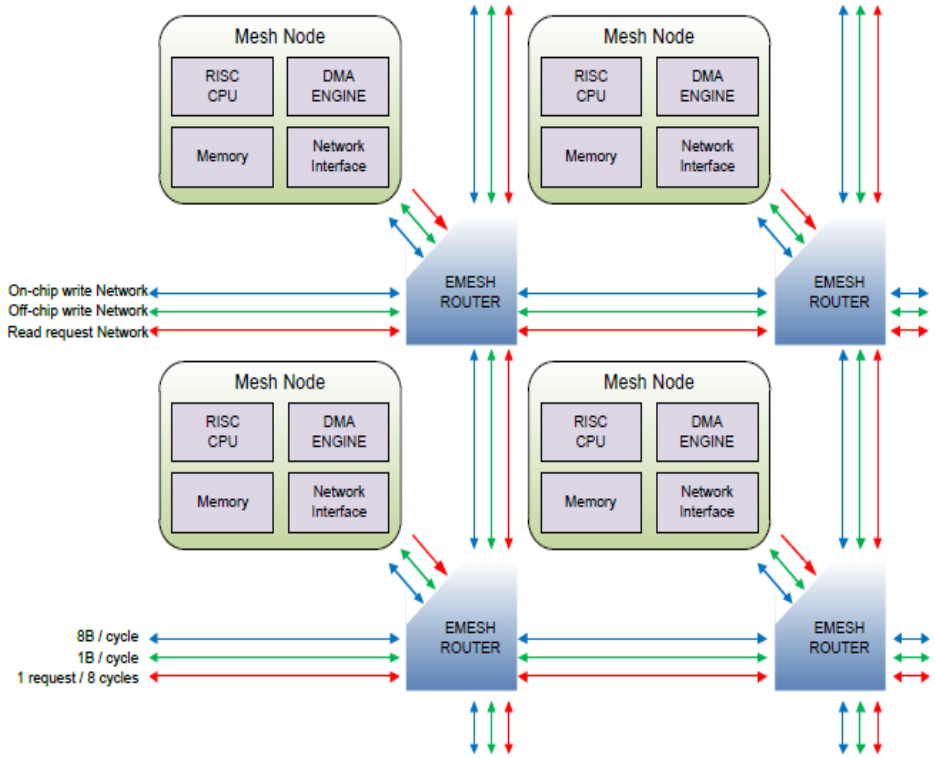
High Performance doesn’t guarantee Power Efficiency.

Low Performance doesn’t guarantee Power Efficiency too.

Parallella Board



Epiphany Network-On-Chip (NOC)



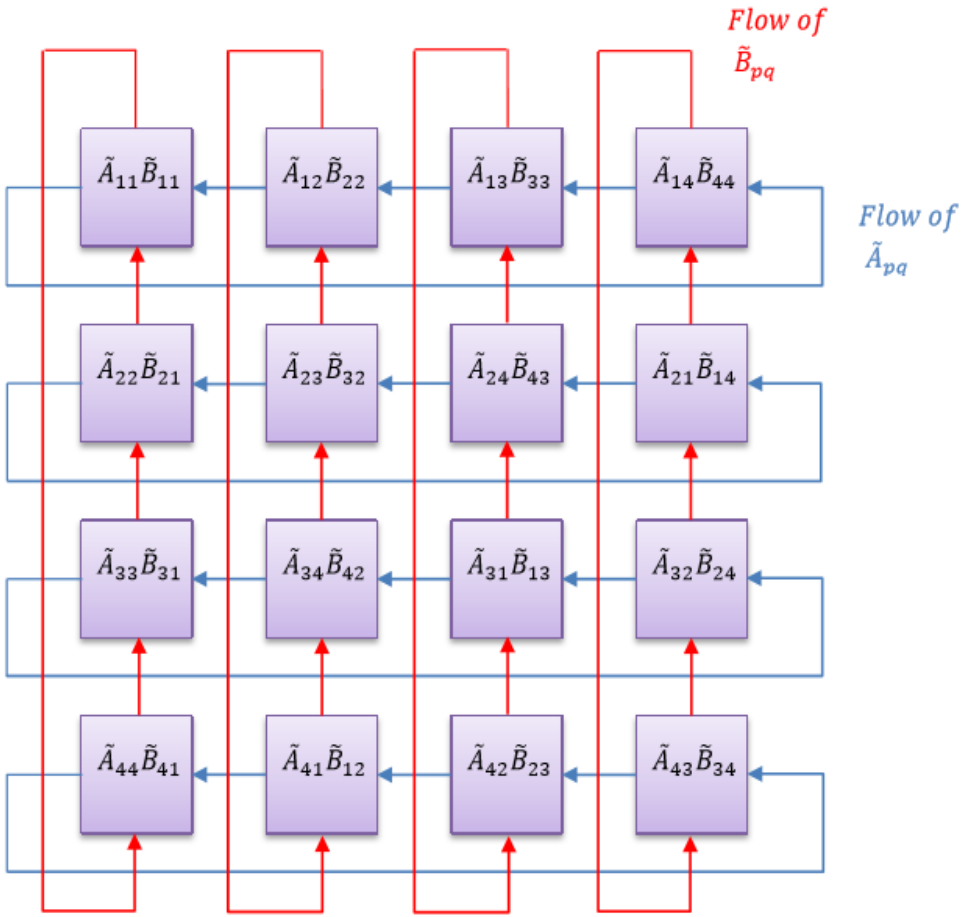
- GPU can only be used for data parallelism(No communication between cores).
- Epiphany cores have full-fledged CPUs that supports parallelism(Cores can communicate with each other as a workgroup)
- Epiphany IV draws about 2W for 100GFlops. While top end NVIDIA GPU draws about 50W.

Matrix Multiplication

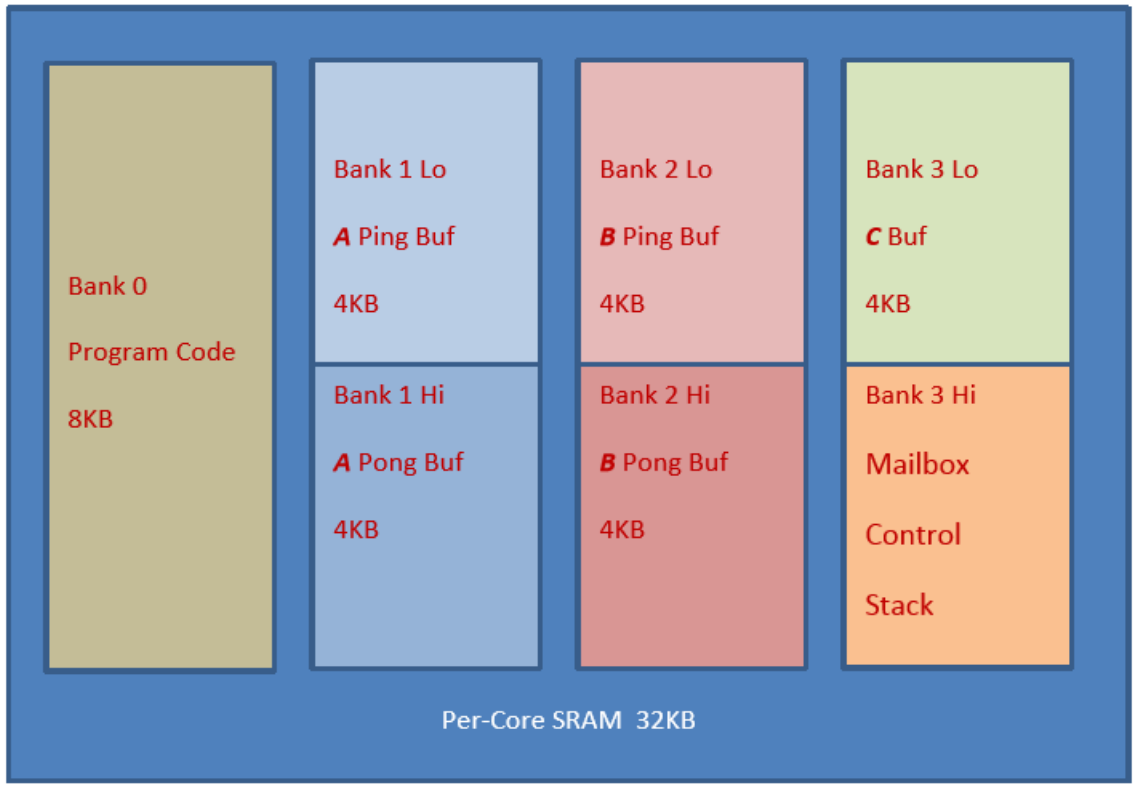
$$C = A \cdot B \rightarrow c_{ij} = \sum_{k=1}^{N} a_{ik} \cdot b_{kj}$$

$$A = \begin{pmatrix} \tilde{A}_{11} & \tilde{A}_{12} & \cdots & \tilde{A}_{1M} \\ \tilde{A}_{21} & \tilde{A}_{22} & \cdots & \tilde{A}_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{A}_{M1} & \tilde{A}_{M2} & \cdots & \tilde{A}_{MM} \end{pmatrix}$$

$$\tilde{C}_{pq} = \sum_{r=1}^{M} \tilde{A}_{pr} \cdot \tilde{B}_{rq}$$

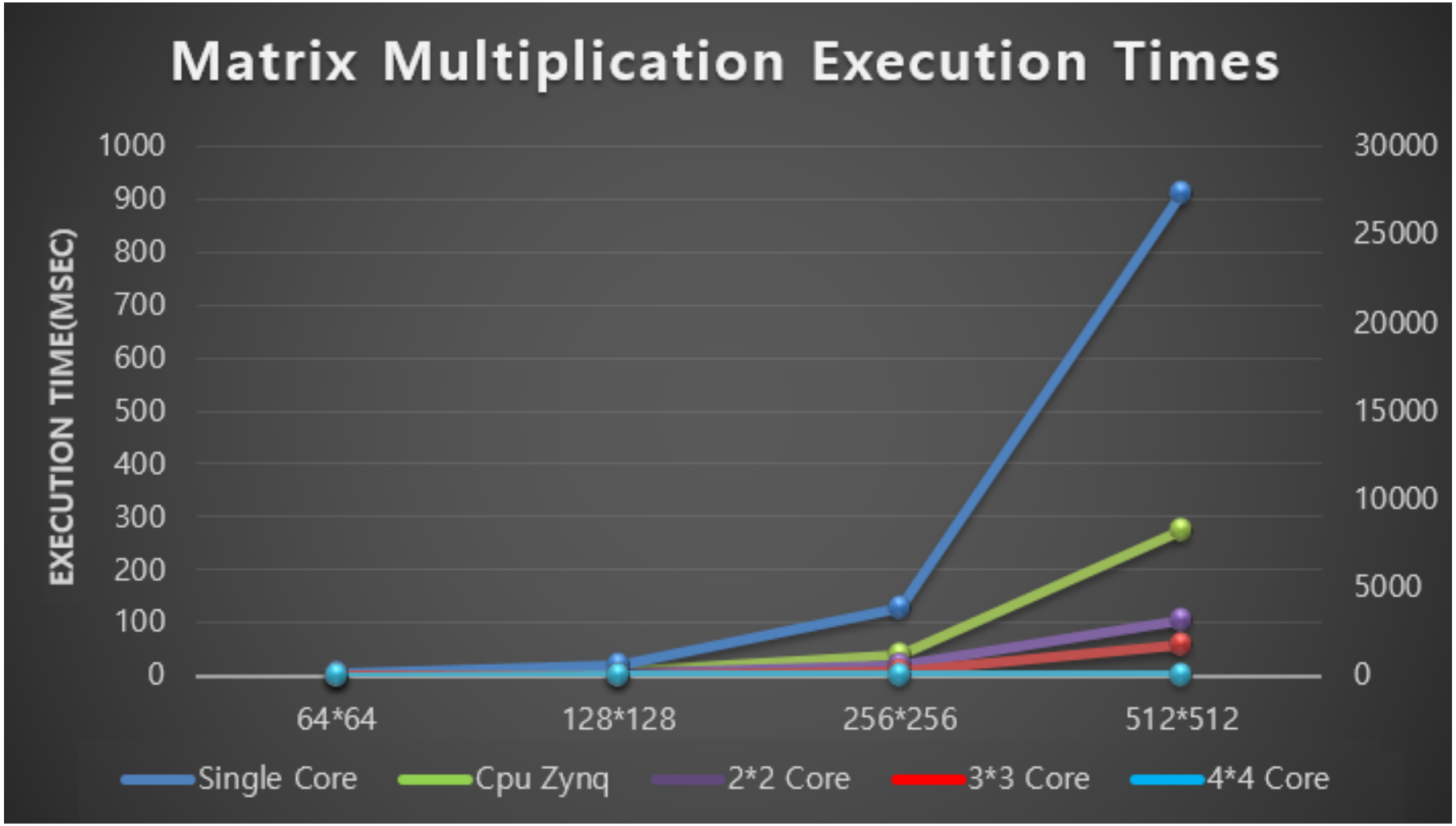


E-core Local Memory Arrangement

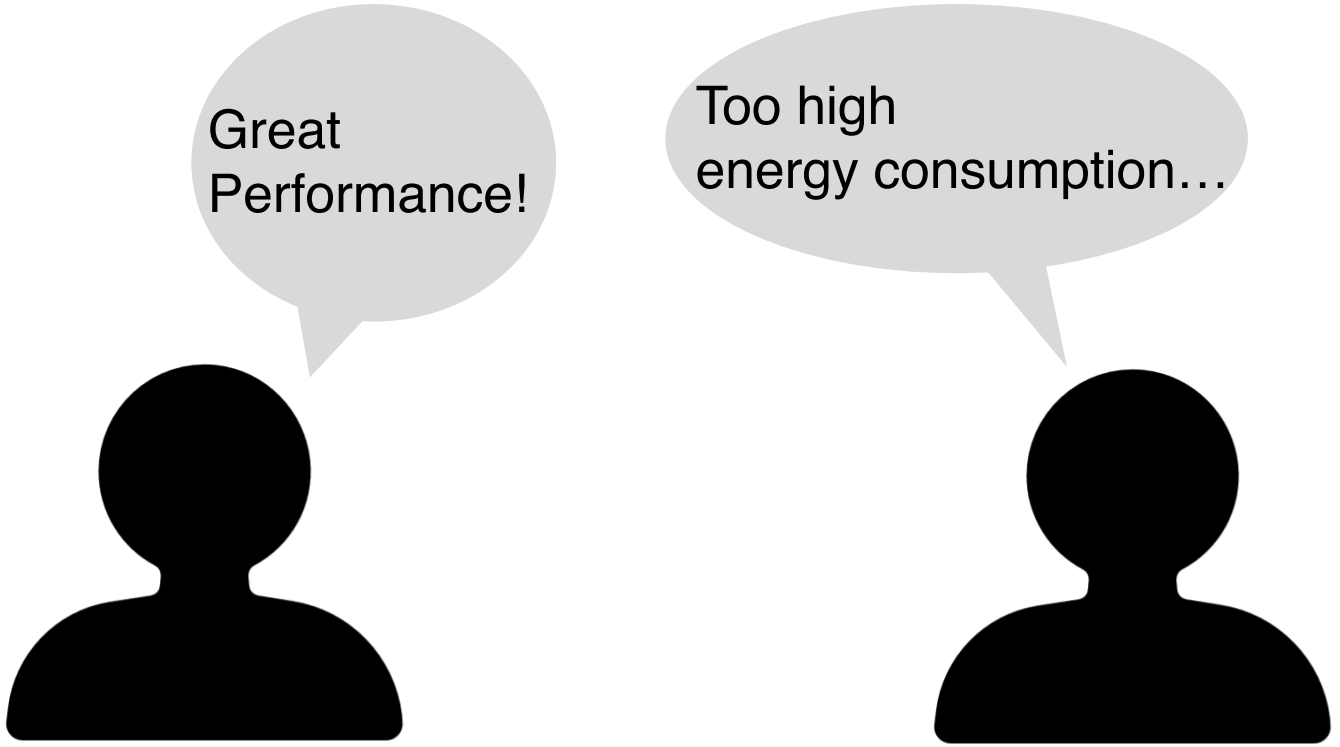


- 32KB of per-core SRAM has four 8KB banks
- First bank is for program code
- Second bank contains operand A sub-blocks
- Second bank contains operand B sub-blocks
- Higher half of the fourth bank contains per-core control structure, inter-core mailbox, program stack

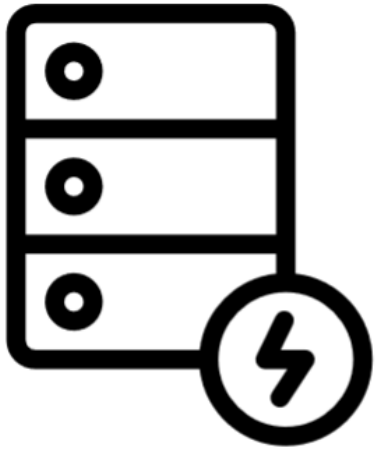
Matrix Multiplication Execution Times



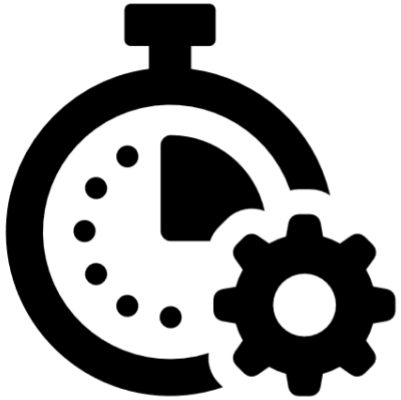
Benchmarking Parallella



Benchmarking Parallella

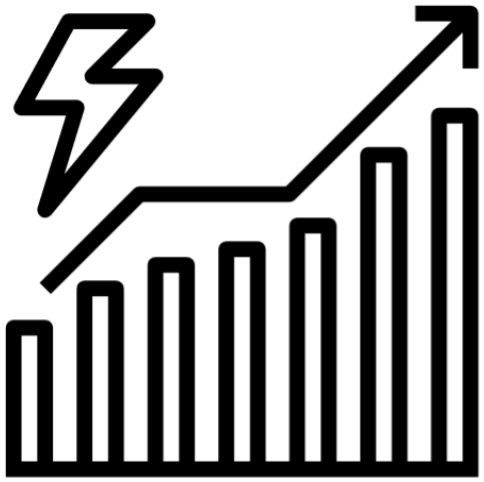


Power Consumption



Performance

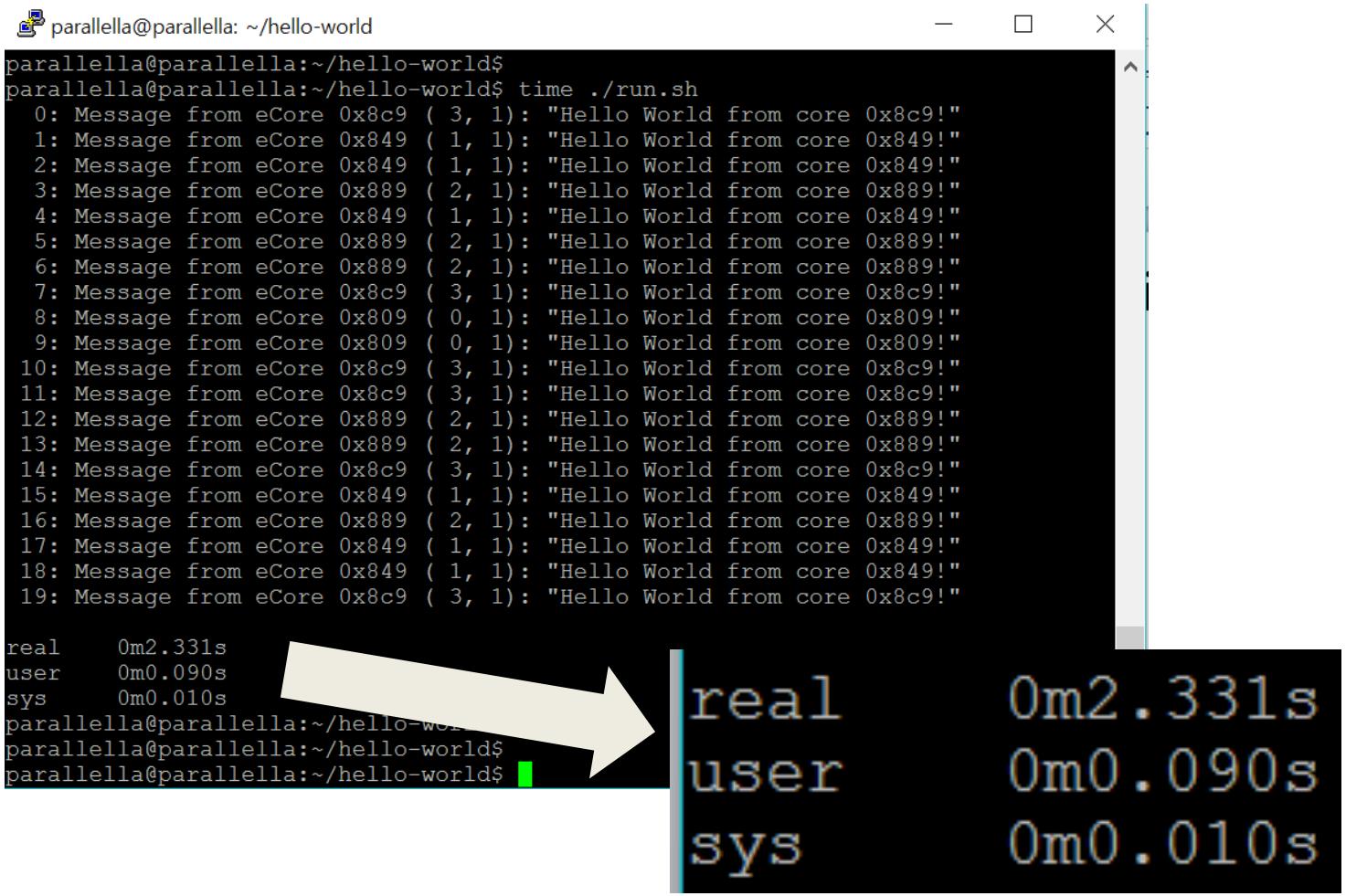
Benchmarking Parallella



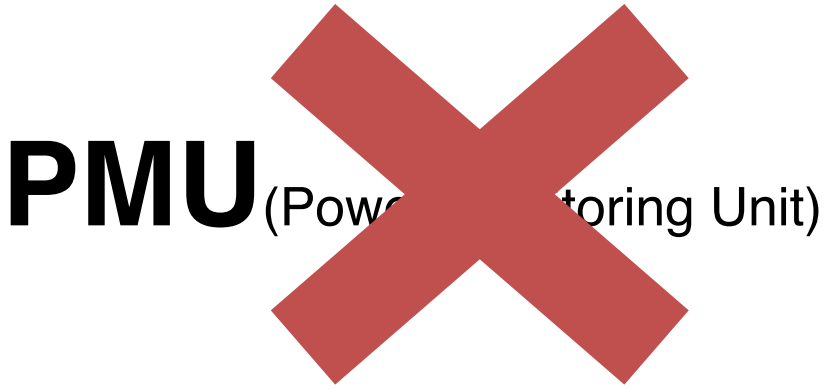
Power Efficiency

How To Check Execution Time

- Using “**Time**”



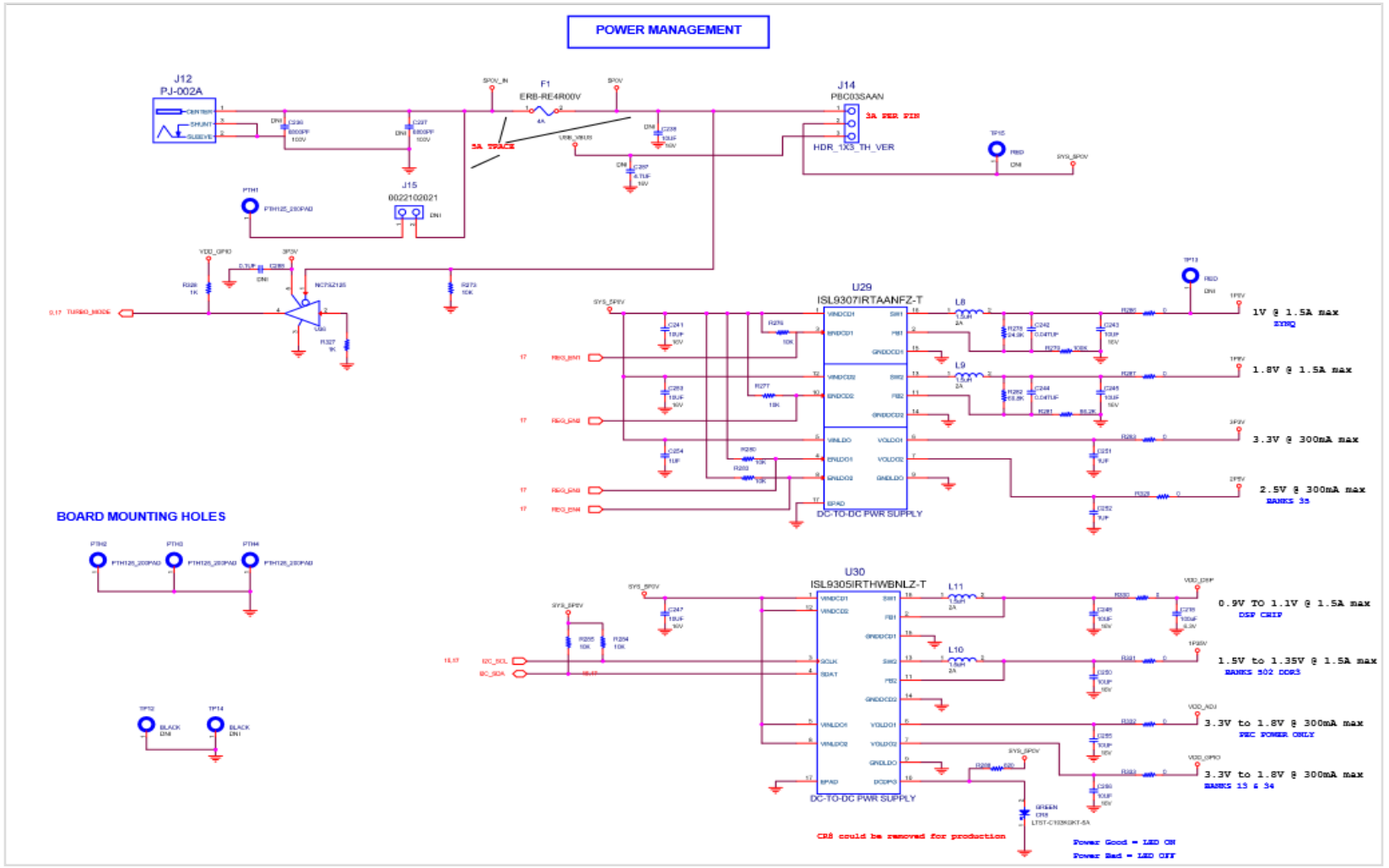
How To Measure Power Consumption



How To Measure Power Consumption

$$P_{(W)} = V_{(V)} * I_{(A)}$$

How To Measure Power Consumption



How To Measure Power Consumption

- Power Monitoring Test Point

TP5 (Ground of Board)

TP13 (Voltage for ZYNQ)

TP14 (Voltage supplied to Epiphany)

TP15 (5V input)

Voltage

J14
(Entire board)

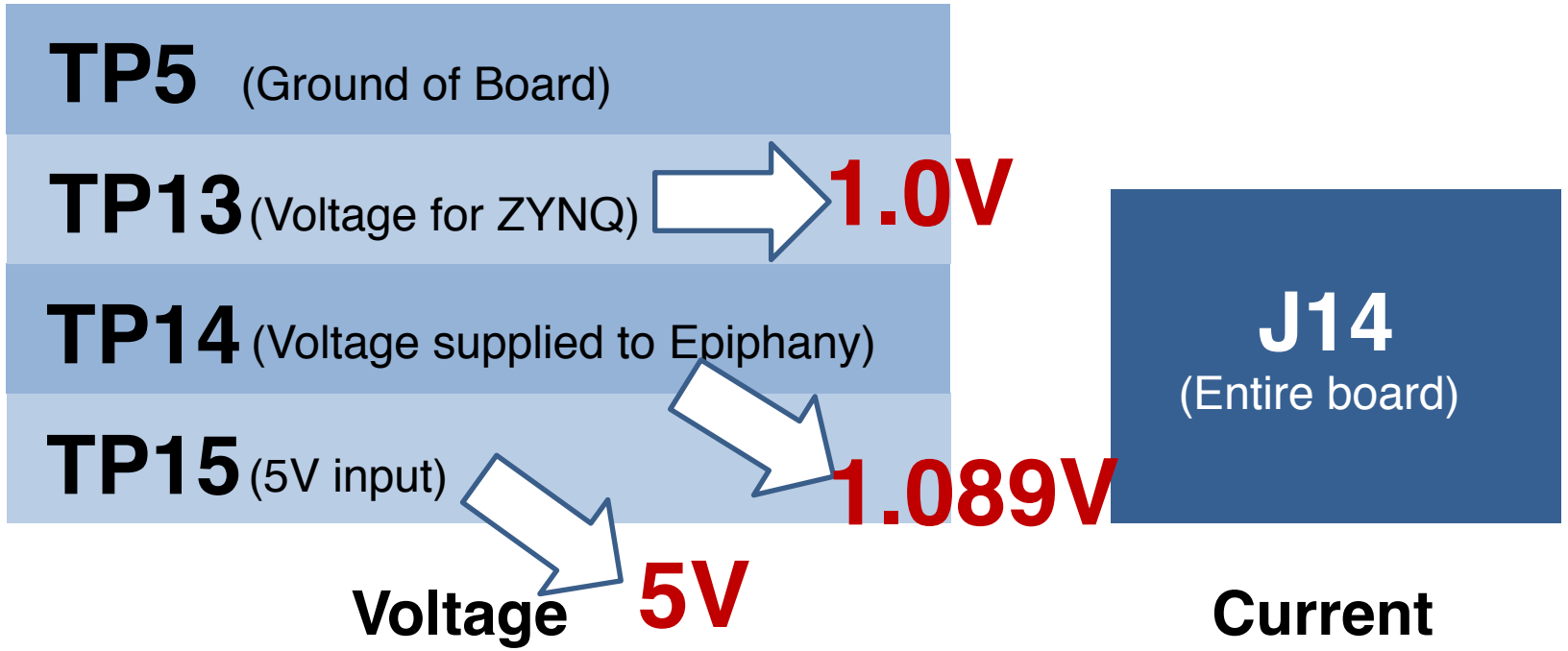
Current

How To Measure Power Consumption

DVFS (Dynamic V… …nd Frequency Scaling)

How To Measure Power Consumption

- Power Monitoring Test Point



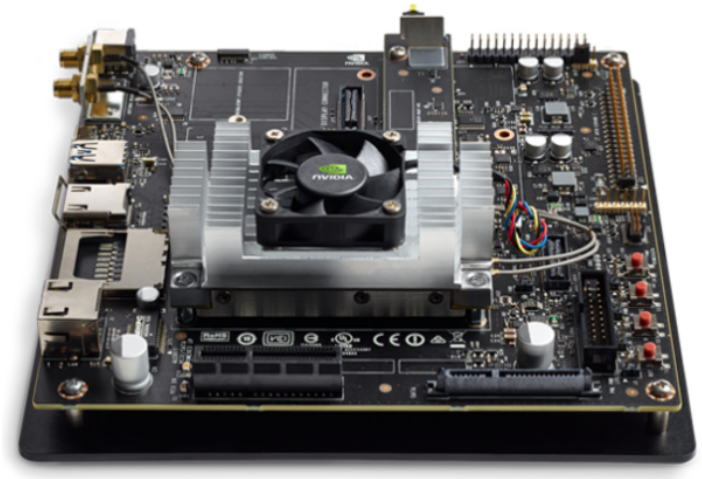
How To Measure Power Consumption

$$P_{(W)} = 5_{(V)} * I_{(A)}$$

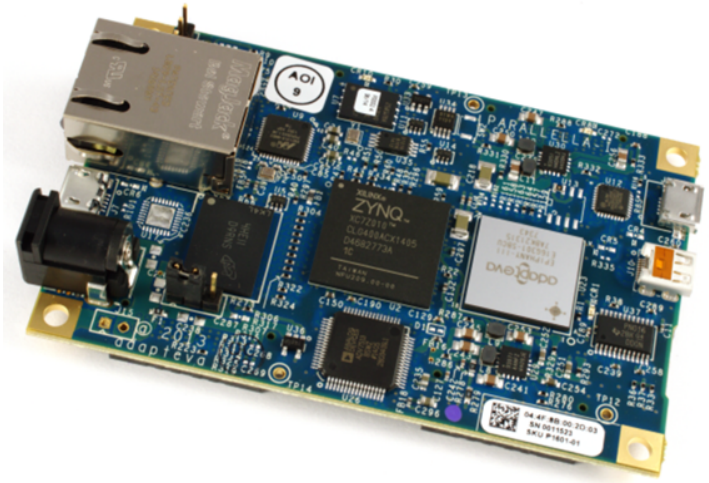
Results For Power Consumption

	CPU Zynq Only	Single Core	2*2 Core	2*4 Core	4*4 Core
Power Consumption	**TBD**	**TBD**	**TBD**	**TBD**	**TBD**

Summary



NVIDIA Jetson TX2

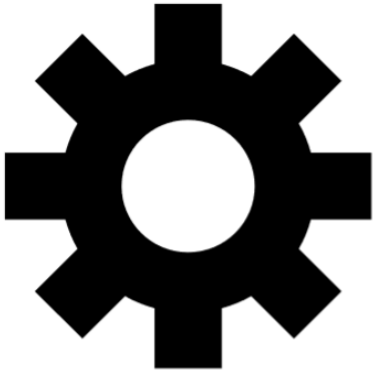


Parallella-16

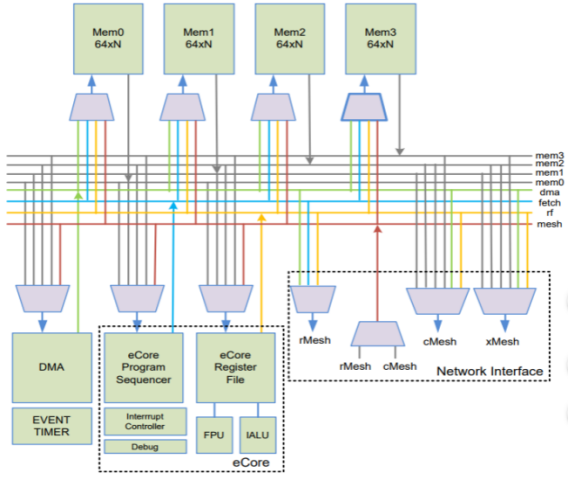
Future Work

- Parallella

Measuring Current and Calculating Power Consumption.



Find the most Optimal state of Parallella.



Measuring IPS(Instructions per second) by using Event Timer.

Q&A