THE LINUX FOUNDATION
OPEN SOURCE SUMMIT
NORTH AMERICA

Sustainability *the Container Native Way*

Huamin Chen
Red Hat
Twitter @root_fs
Github rootfs

Chen Wang
IBM Research
Twitter @wangchen615
Github wangchen615

#ossummit

THE LINUX FOUNDATION

Agenda

- Background
- Introduction to Project Kepler
- Next Steps
- Community
- Demo

Energy Measurement

- Energy measurement theory and methodology
  - How to measure energy consumption, indirectly?
- Energy consumption attribution problem and methodology
  - How much energy used by the Pods?

Power Measurement Theory

There are several factors contributing to the CPU power consumption; they include dynamic power consumption, short-circuit power consumption, and power loss due to transistor leakage currents:

$$P_{cpu} = P_{dyn} + P_{sc} + P_{leak}$$

leakage current is a function of thermal temperature

The dynamic power consumption originates from the activity of logic gates inside a CPU. When the logic gates toggle, energy is flowing as the capacitors inside them are charged and discharged. The dynamic power consumed by a CPU is approximately proportional to the CPU frequency, and to the square of the CPU voltage:

$$P_{dyn} = CV^2 f$$

Frequency based tuning to optimize dynamic power consumption.

where C is the switched load capacitance, f is frequency, V is voltage

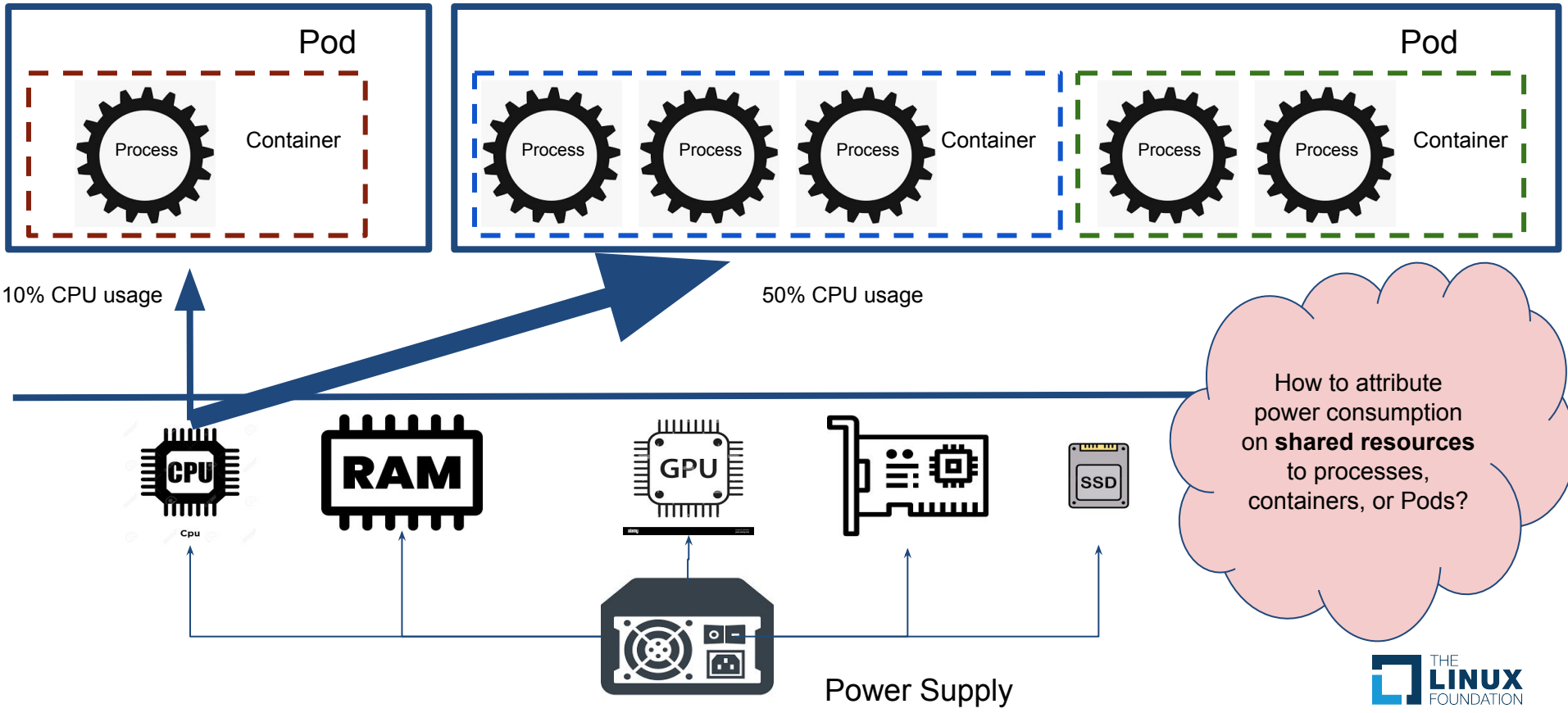
Source: https://en.wikipedia.org/wiki/Processor_power_dissipation

Energy Measurement Methodology: Ideal vs Reality

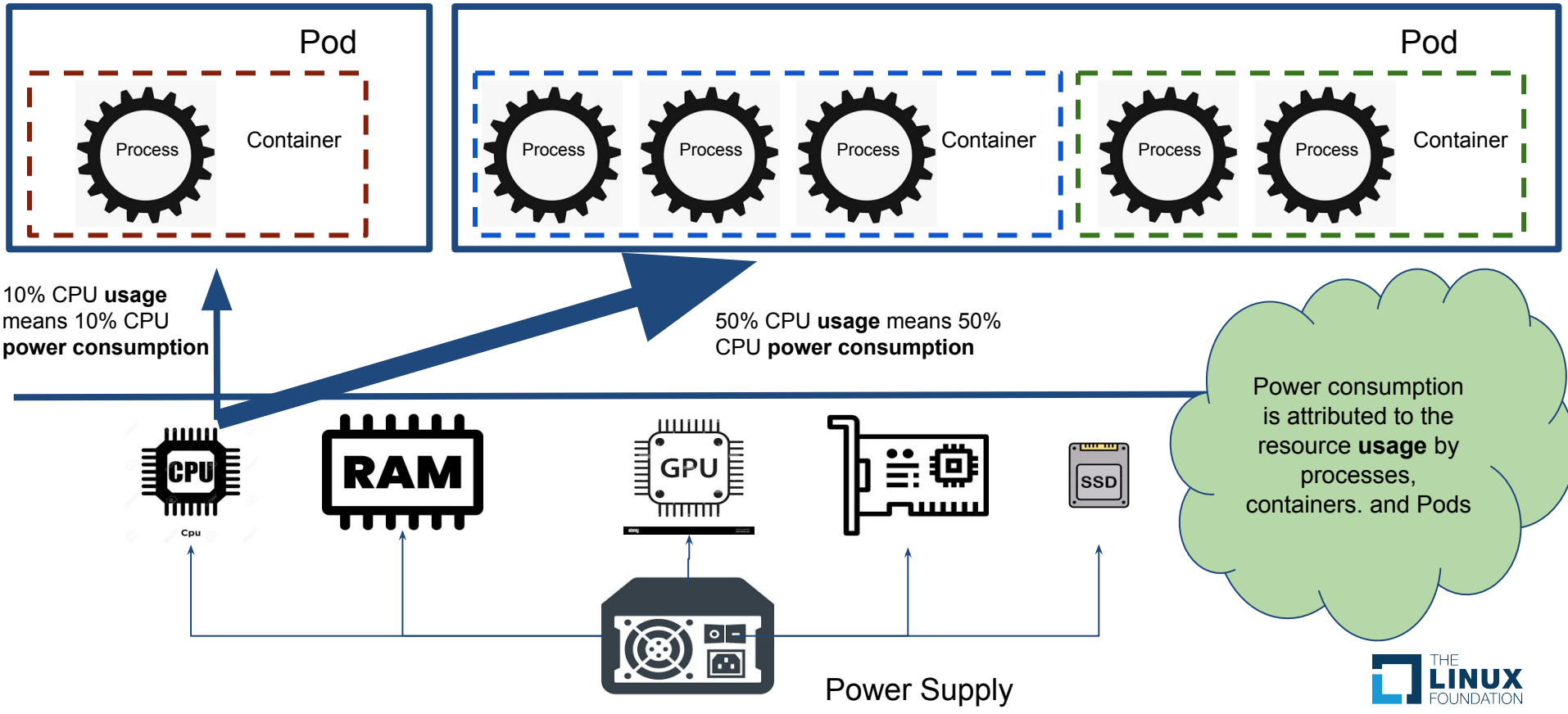
Measure Target	Ideal Way	Reality	Solution
Frequency	Monitor each **circuit frequency**	Linux kernel CPU frequency governor dynamically changes frequency for energy saving or performance objectives	Use **average frequency** (aperf) to approximate
Capacitance	Monitor the **number of circuits** powered on	There is no such way on CPUs	Use the **number of CPU instructions** to approximate
Execution Time	Monitor the **duration of circuits** staying powered on	There is no such way on CPUs	Use the **CPU cycles** to approximate

Indirect measurement to approach the truth in an imperfect world

Energy Consumption Attribution Problem



Energy Consumption Attribution Methodology



Put Together

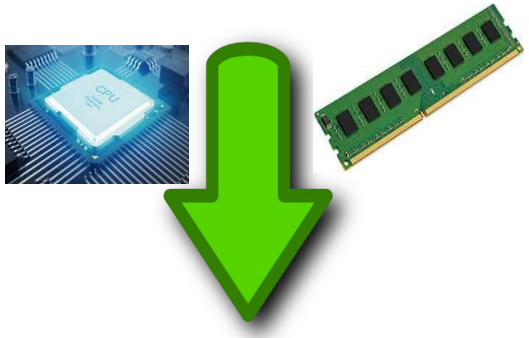
- Use software counters to measure power consumption by hardware resources
  - **ML** models are used for approximation.
- Use hardware resource utilization to attribute power consumption by processes, containers, and Pods.

Kepler: Kubernetes-based Efficient Power Level Exporter

https://github.com/sustainable-computing-io/kepler

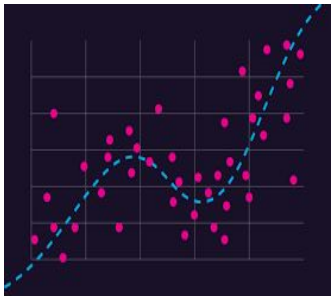
Reporting

- Per Pod level energy consumption reporting, including **CPU/GPU, RAM**
- Support **bare metal** as well as **VM**
- Support **Prometheus**



Reduction

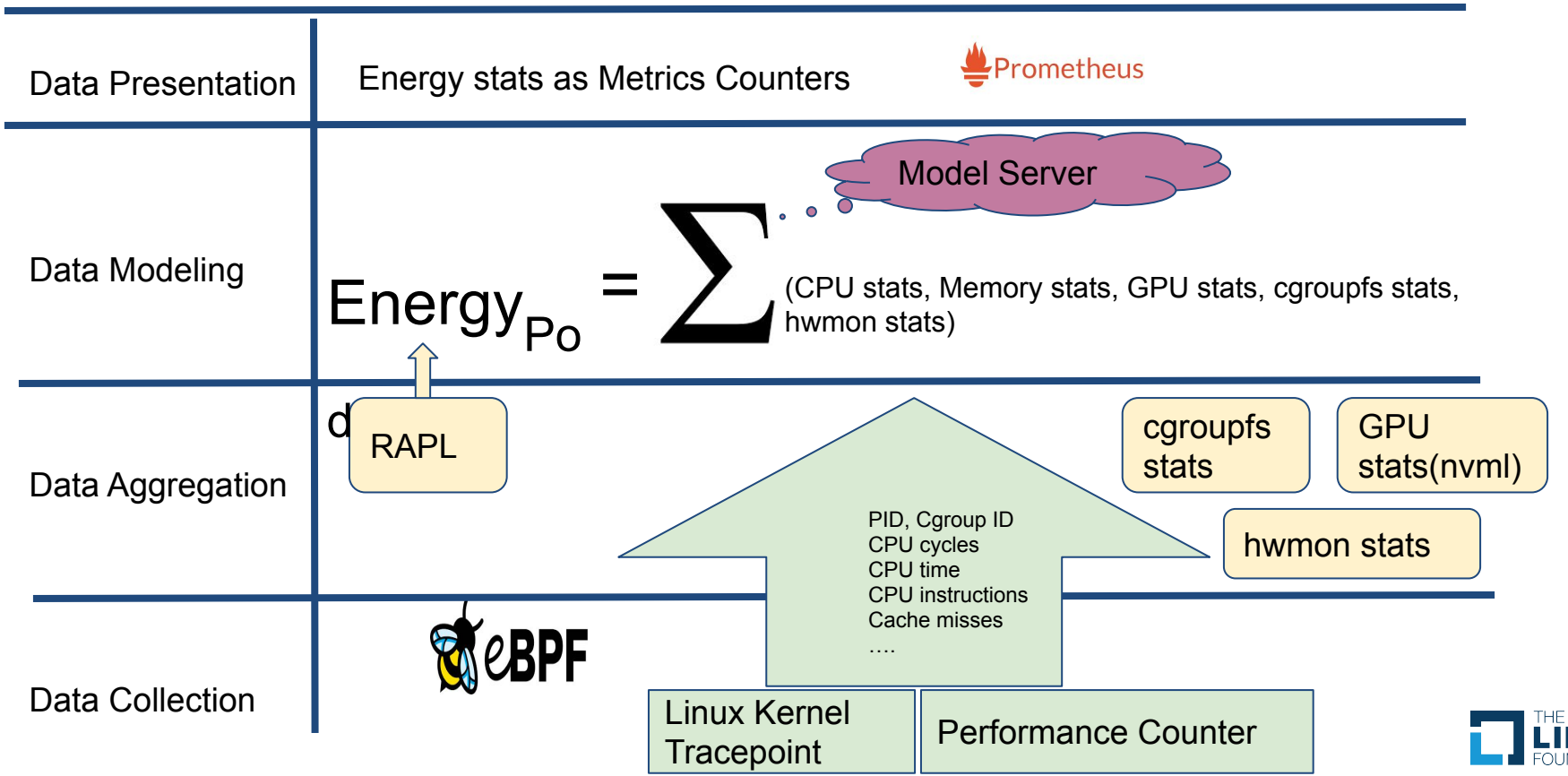
- Reduced computational resource used by the probe
- Using **eBPF**

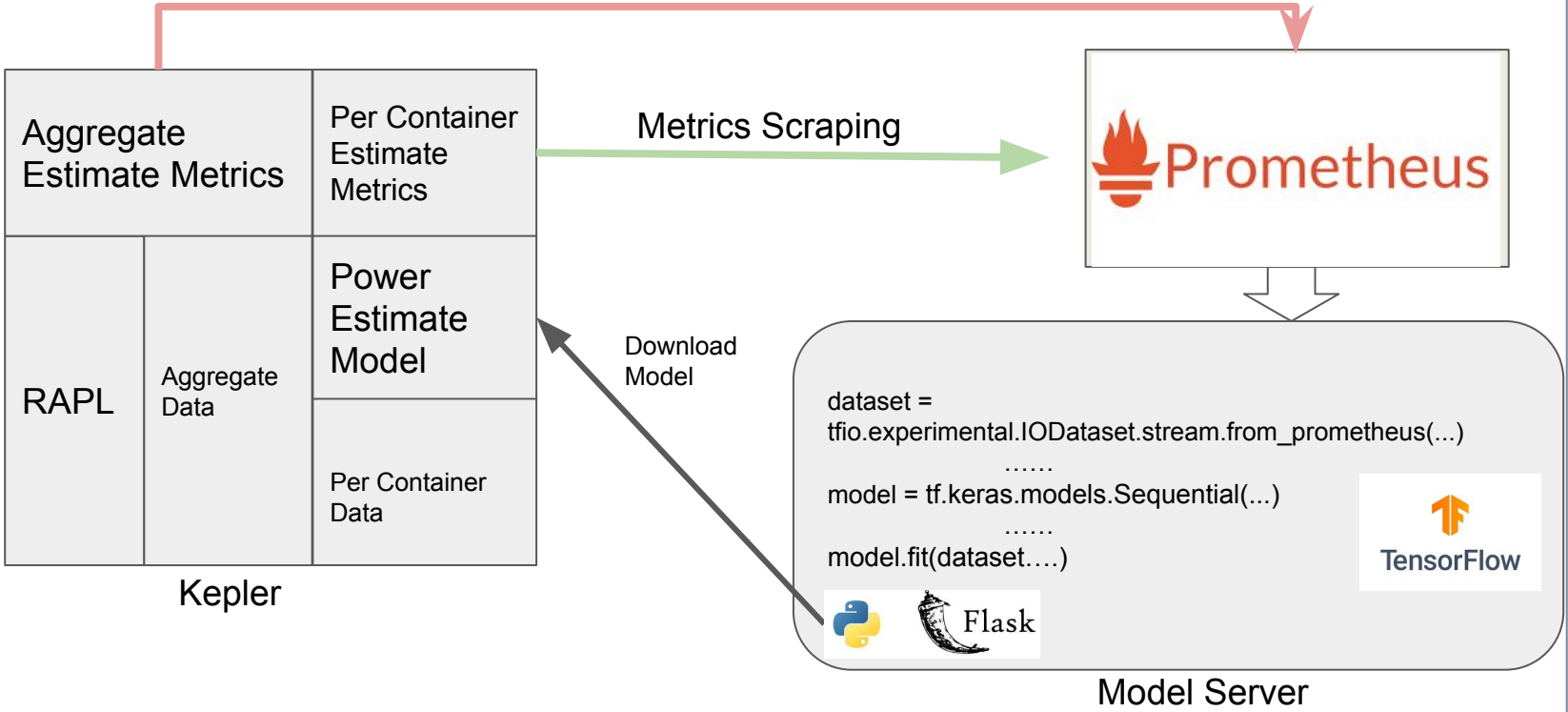


Regression

- Support **ML** models to estimate energy consumption
- Science based approach

Kepler Architecture





Model Server trains ML models that use software counters to approximate power consumption.

Demo

Setup

- Bare Metal Machine
  - Intel Xeon 4210
  - NVIDIA T4 Graphics Card / 20 cores / 2.20 GHz / 32 GB RAM / 20 TB bandwidth
- MicroShift
- NVIDIA GPU Operator
- Kepler
- kube-prometheus

Demo

Steps:

1. Deploy Kepler Exporter as a DaemonSet
2. Check Kepler exported data in Kepler logs and endpoint
3. Configure Prometheus and Metric Scape
4. Load Kepler Grafana Dashboard
5. Test CPU intensive workload
6. Test GPU intensive workload

Supported Data Stats

- CPU Stats
  - Energy, model, time and frequency
  - Hardware Performance Counters
- Memory
  - Cache misses, resident memory size
- GPU
  - Energy, resident memory size
- Block Device
  - Read/write stats

Next Steps

- Energy consumption and carbon emissions telemetry and dashboard.
- Kepler and Model Server provide tuning and scheduling heuristics.
  - Dynamic power reduction
    - Energy Aware Pod Scheduling
    - DVFS based Vertical Pod Scaling
    - Energy Efficient Node Tuning
  - Leaky power reduction
    - Thermal Temperature Aware Scheduling and Scaling

Community

https://github.com/sustainable-computing-io

https://sustainableco-ese6814.slack.com/archives/C02VCJ8K8LT

slack

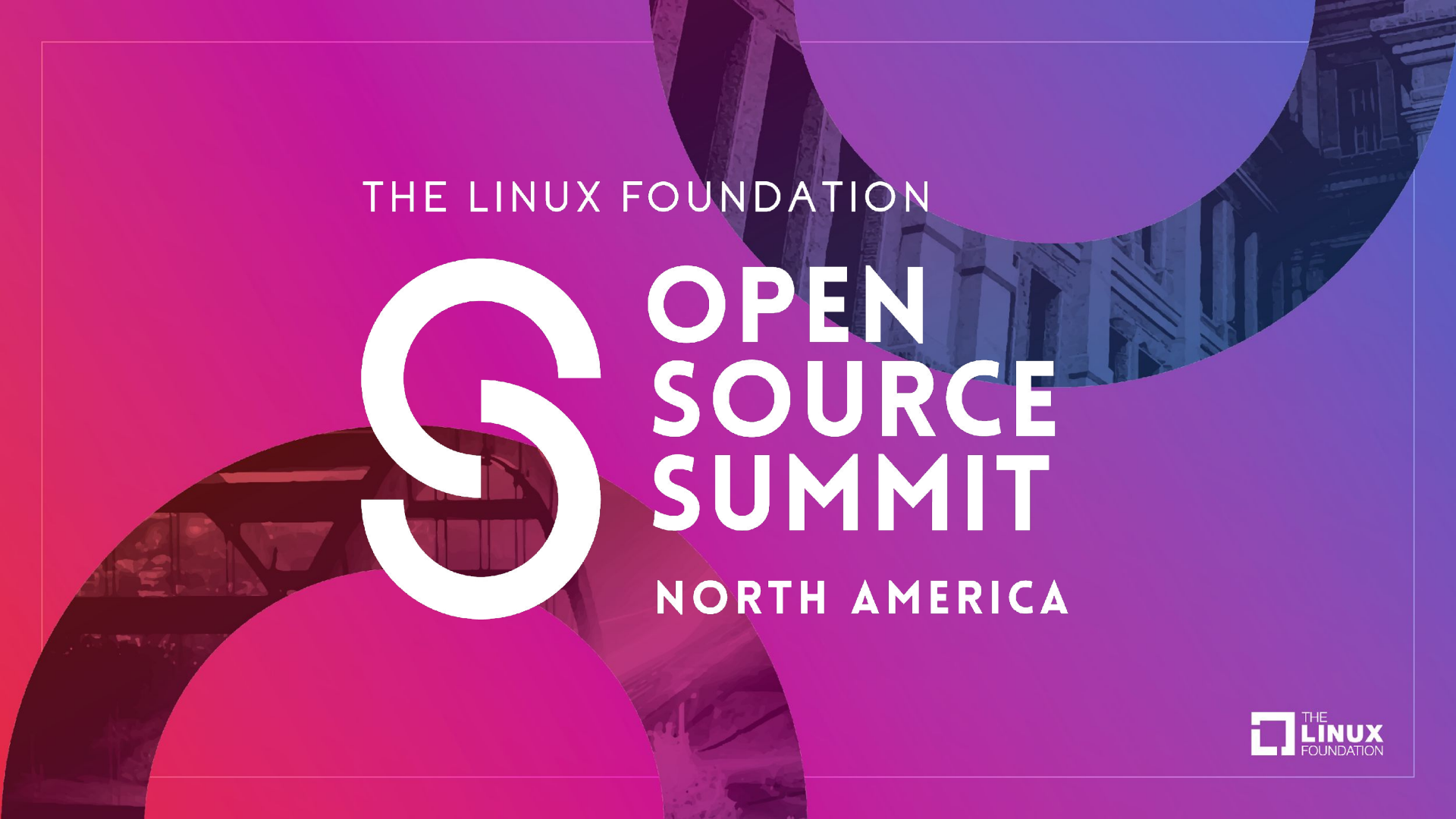
Red Hat

IBM

intel

weaveworks

THE LINUX FOUNDATION


THE LINUX FOUNDATION
OPEN SOURCE SUMMIT
NORTH AMERICA
THE LINUX FOUNDATION

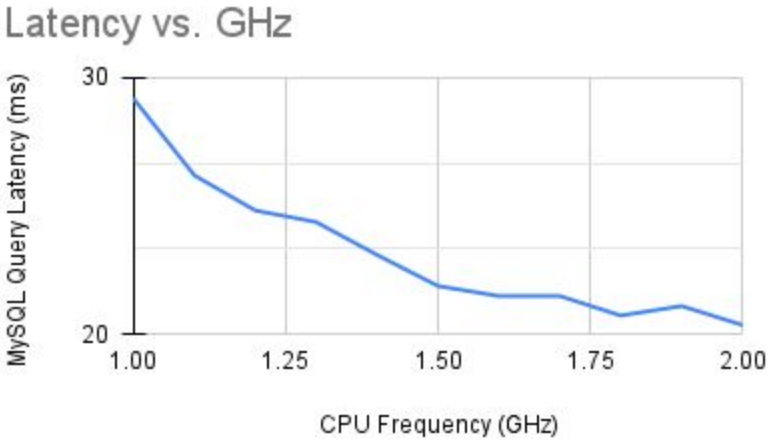
Backup Slides

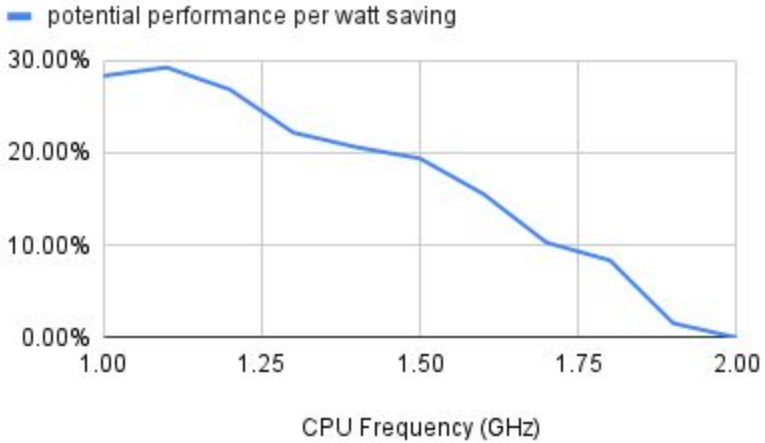
Existing Landscape and Use Cases

- Open Source Projects
  - PowerAPI
    - https://github.com/powerapi-ng
  - Scaphandre
    - https://github.com/hubblo-org/scaphandre
- Kepler Design Principle:
  - Container Native, Cloud Native
  - Lightweight, Expansible
  - Accurate and Fair

Performance Per Watt

Workload: sysbench
Platform: Intel Xeon Sandy Bridge





DVFS based performance per watt tuning yields promising energy savings