Assessment of software-based energy measurement tools using hardware probes at component level

Hongjian Huo
PRM 2025

Motivation

"Global data center energy consumption expected to triple by 2035 to 1,200 terawatt hours" - International Energy Agency

- Data centers represent a significant fraction of global energy consumption
    - Represents significant opportunity for emission and cost reduction

Challenge:

- Large variety of proposed energy measurement tools for software and AI

Challenges

What tool should I use for optimizing energy use of a given system?

- Multitudes of tools available
- No unified reporting framework

Does the tool have the right scope for my system?

- Tools may not be able measure energy on a by-process basis or inside containers

Does the tool have sufficient fidelity to allow energy optimization?

- Tools may not measure frequently enough to allow fine-grained optimization or experience execution drift at required frequencies.
- Tools may not accurately model or measure power consumption
- Tools may not take into account components beyond CPU/GPU

Objectives

Survey of pre-existing literature to identify tools for measuring energy consumption

- Compile information such as scope, ease of use, system requirements

Assessment of energy measurement tools

- Measure achievable sampling frequency, accuracy and overhead
- **Challenge**: how do we acquire the ground truth for measurement accuracy?

Methodology: acquiring ground truth at component level

Prior solutions:

Motherboard / component integrated sensors: (example: nvidia GPU, Jetson TX1)

Not component-agnostic, Most components/motherboard do not have such features

Power supply measurement:

System-scope measurements only, lack component-scope measurements

Hardware specific interface (example: intel RAPL):

Not component-agnostic, blackbox, limited applicability to other components

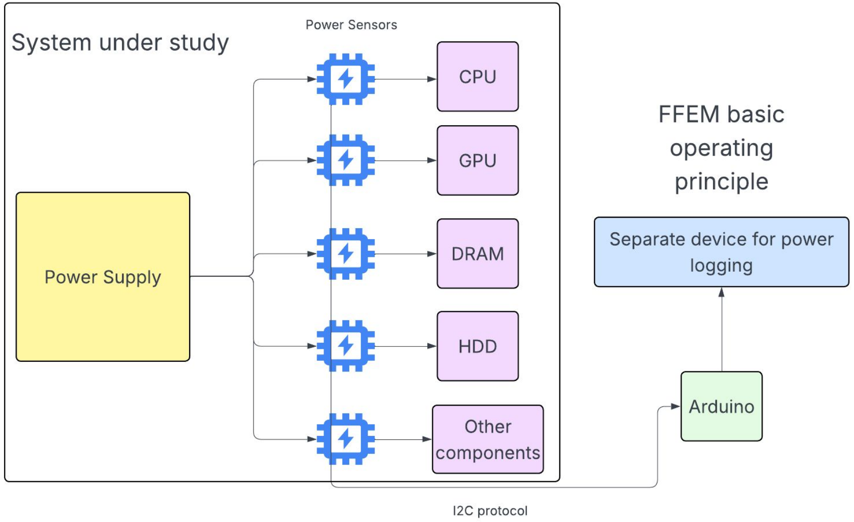
Proposed solution:

Component level probes on cuttable wires as ground truth for assessment

Tool intro and background

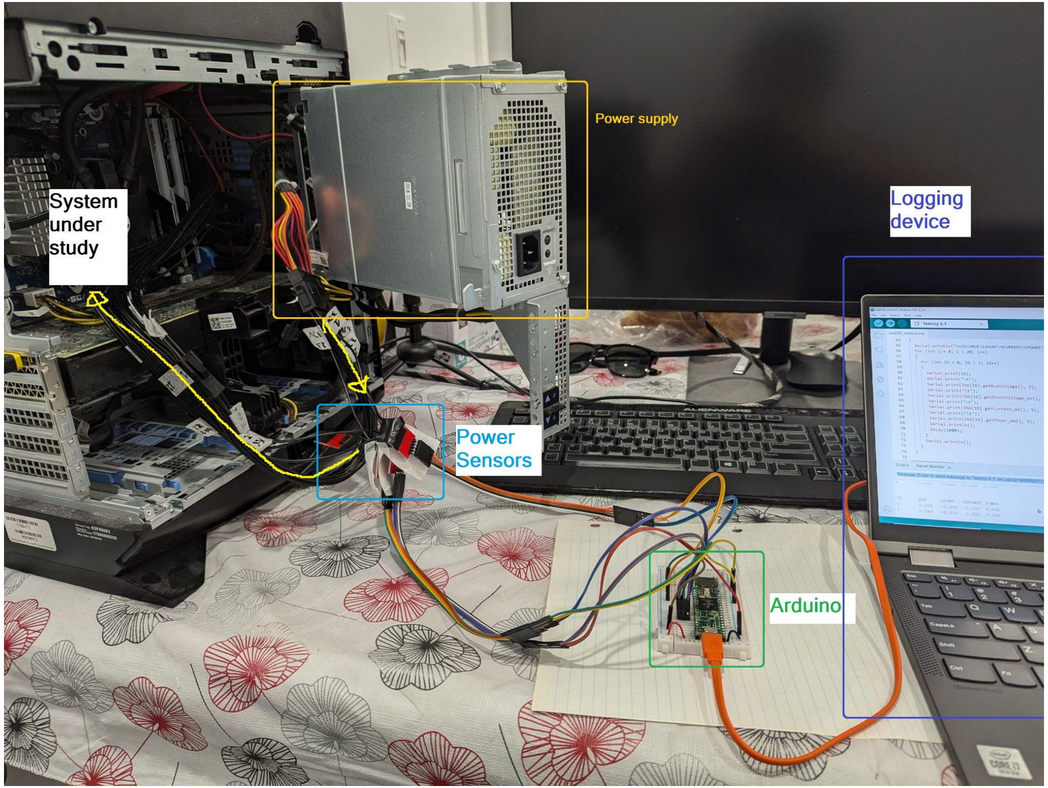
FFEM tool (Weber et al 2025)

- Power monitor attached between computer power supply and component
  - Component-level fidelity
  - Component-agnostic (can be attached to any cuttable wire)
  - High measurement fidelity: 1.5khz measurement frequency and 0.1% maximum gain error

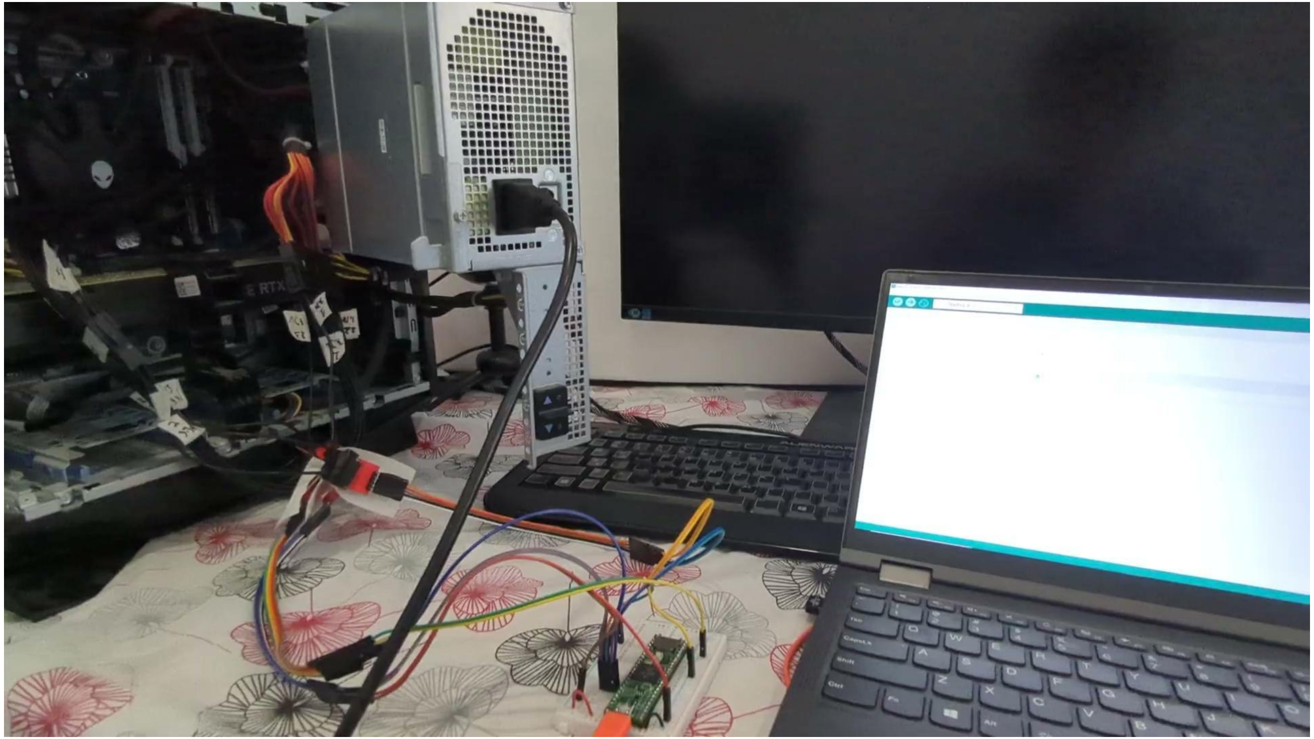


Tool Progression

- An INA226 sensor is attached to an extension cable between the power supply and components of interest.
- The Teensy 4.1 is a fast, arduino-compatible device enabling high sampling frequency.
- The INA226 sensors communicate with the Teensy 4.1 through I2C protocol.



Tool Demo



What’s next

1. **Collect power measurement accuracy data on software based tools**
   - The selected workload should represent a sufficiently wide variety of tasks
   - Measurement on multiple hardware platforms is required to ensure validity.

2. **Collect data on achievable sampling frequency for software based tools**

Methodology:

Use SPEC 2017 and MLcommons benchmarks as workload for collecting power measurement and frequency data.

- SPEC 2017:
  Provides a collection of task-specific benchmarks
- MLcommons:
  Provides various machine learning model benchmarks

TABLE VIII: Classification of benchmarks based on application domains.

INT Benchmarks	
App domain	SPEC 2017
Compiler	**502.gcc_r**, 602.gcc_s **500.perlbench_r**, 600.perlbench_s
Compression	**525.x264_r**,**557.xz_r**, **625.x264_s**, 657.xz_s
AI	**531.deepsjeng_r**, 631.deepsjeng_s, **541.leela_r**, 641.leela_s, **548.exchange2_r**, 648.exchange2_s
Combinatorial optimization	**505.mcf_r**, 605.mcf_s
DE Simulation	**520.omnetpp_r**, **620.omnetpp_s**
Doc Processing	**523.xalancbmk_r**, **623.xalancbmk_s**
FP Benchmarks	
App domain	SPEC 2017
Physics	**507.cactuBSSN_r**, **549.fotonik3d_r**, 607.cactuBSSN_s, **649.fotonik3d_s**
Fluid dynamics	**519.lbm_r**, **503.bwaves_r**. **619.lbm_s**, **603.bwaves_s**
Molecular dynamics	**508.namd_r**,**544.nab_r**, 644.nab_s
Visualization	**511.povray_r**,526.blender_r, **538.imagick_r**,**638.imagick_s**
Biomedical	**510.parest_r**
Climatology	**521.wrf_r**, 527.cam4_r, 628.pop2_s, **554.roms_r** 621.wrf_s, 627.cam4_s, **654.roms_s**

Methodology:

Maximum achievable sampling frequency

- Using the same workload, configure tools to sample at greater frequencies until clock drift or loss of accuracy is observed.
- Higher stable sampling frequency is crucial for the utility of these tools.

Available hardware

- Alienware R8 PC
- Other lab computers

Future work and directions

Next steps

- Complete data collection and assess tools

Future directions

- Improving component level power consumption prediction models using hardware probe data.
- Better reporting framework for software based power measurement tools

Conclusion

Motivation

"Global data center energy consumption expected to triple by 2035 to 1,200 terawatt hours" - International Energy Agency

- Data centers represent a significant fraction of global energy consumption
  - Represents significant opportunity for emission and cost reduction

Challenge:

- Large variety of proposed energy measurement tools for software and AI

Objectives

Survey of pre-existing literature to identify tools for measuring energy consumption

- Compile information such as scope, ease of use, system requirements

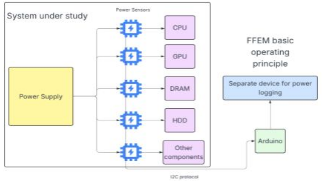
Assessment of energy measurement tools

- Measure achievable sampling frequency, accuracy and overhead
- **Challenge**: how do we acquire the ground truth for measurement accuracy?

Tool intro and background

FFEM tool (Weber et al 2025)

- Power monitor attached between computer power supply and component
  - Component-level fidelity
  - Component-agnostic (can be attached to any cuttable wire)
  - High measurement fidelity: 1.5khz measurement frequency and 0.1% maximum gain error



Methodology:

Use SPEC 2017 and MLcommons benchmarks as workload for collecting power measurement and frequency data.

- SPEC 2017:
  Provides a collection of task-specific benchmarks
- MLcommons:
  Provides various machine learning model benchmarks

TABLE VIII: Classification of benchmarks based on application domains.

INT Benchmarks	
App domain	SPEC 2017
Compiler	**502.gcc_r**, 602.gcc_s **500.perlbench_r**, 600.perlbench_s
Compression	**525.x264_r**,**557.xz_r**, **625.x264_s**, 657.xz_s
AI	**531.deepsjeng_r**, 631.deepsjeng_s, **541.leela_r**, 641.leela_s, **548.exchange2_r**, 648.exchange2_s
Combinatorial optimization	**505.mcf_r**, 605.mcf_s
DE Simulation	**520.omnetpp_r**, **620.omnetpp_s**
Doc Processing	**523.xalancbmk_r**, **623.xalancbmk_s**
FP Benchmarks	
App domain	SPEC 2017
Physics	**507.cactuBSSN_r**, **549.fotonik3d_r**, 607.cactuBSSN_s, **649.fotonik3d_s**
Fluid dynamics	**519.lbm_r**, **503.bwaves_r**, **619.lbm_s**, **603.bwaves_s**
Molecular dynamics	**508.namd_r**,**544.nab_r**, 644.nab_s
Visualization	**511.povray_r**,526.blender_r, **538.imagick_r**,**638.imagick_s**
Biomedical	**510.parest_r**
Climatology	**521.wrf_r**, 527.cam4_r, 628.pop2_s, **554.roms_r** 621.wrf_s, 627.cam4_s, **654.roms_s**

Thank you

Email

hongjian.huo@polymtl.ca