Software Test-Suite for Computer Energy Efficiency Measurement

How to measure the energy efficiency and verify the representativeness of the measurments

Andoni Arregui, Thomas Wucher, and Joan Roig (GTD) / Stephen Fernandes (Intertek)

2020-07-07

GTD GmbH / Intertek

Table of Contents

GTD GmbH

Who we are and what we do:

- We are a software company in the south of Germany
- Part of the European group GTD with headquarters in Barcelona

We mainly work for:

- European Space Agency (ESA)
- German Aerospace Center (DLR)
- German federal institutions (UBA, BAM)

Since 2019 we have our software in space

Software Overview, Theory, and Implementation

The Goal of Our Software

To measure computer energy efficiency we must

- Measure the **computing performance**
- Measure the **power demand**
- Compute the **energy efficiency** based on the performance and power demand

Existing benchmarks

- Focus **only** or mostly on **performance** (e.g., PCMARK)
- **Alter** the measured **performance** by having extra interfaces connected to the Unit Under Test and run client software to report the results to a controller computer (e.g., SPECpower)
- Are **not open** and do not run on different operating systems (e.g., SYSmark)

Test Suite Overview

The main characteristics of the Test Suite developed by GTD are:

- It is based on the **open-source** Phoronix Test Suite
- It runs a **series** of ***worklets***[1]
- It executes **native binaries** in Microsoft Windows, MacOSX, and Linux
- Measuring power **does not alter** the **performance** or the **power demand**
- All the results are fused together in a **single meta-efficiency metric** value

[1] A *worklet* is a set of tasks, placed in a *workflow*, that is, in a sequence of activities that can be easily reproduced on a regular basis.

Phoronix Test Suite Overview

Phoronix[2] Test Suite is an open-source framework for automated benchmarking:

- Runs on Microsoft Windows, Linux, MacOSX
- Automates the whole testing process from dependency installation to test results aggregation
- Has a huge collection of predefined test-profiles (i.e., worklets)
- Easily configurable for our use-case (e.g., new worklets)
- Results can be visualized by Phoronix Result-Viewer or exported to CSV, plain text
- Support for external measurement devices

[2] https://www.phoronix-test-suite.com/

Worklets Considered for Our Assessment

Category	# Worklets	Activities
System	18	CAD, 2/3D-image processing, office, idle, browsing, OS
CPU	11	Compressing/decompressing, video encoding, arithmetic
Disk	5	File and database reading/writing
Graphics	7	Gaming
Memory	5	Memory reading/writing and caching
Total	**46**	

Usage Profile-based worklets

They replicate a typical user activity: LibreOffice, FreeCAD, GIMP, etc.

Synthetic worklets

They target mainly specific hardware components: Cachebench, 7zip Compression, etc.

Assessment of Worklets

For the assessment and suitability of the worklets the Test Suite does the following:

- executes the worklets with native binaries for Microsoft Windows, MacOSX, and Linux
- automatically re-runs the worklets until performance falls within a certain measured tolerance of repeatability to avoid outliers
- provides the performance, power demand, and energy efficiency per worklet
- fuses all energy efficiency results into a single value

The Energy Efficiency we Compute

For the assessment of the energy efficiency we compute an efficiency metric

Worklet Efficiency

Is the division of the performance (e.g., frames per second, inverse of the time to run the worklet) by the average power demand in Watts.

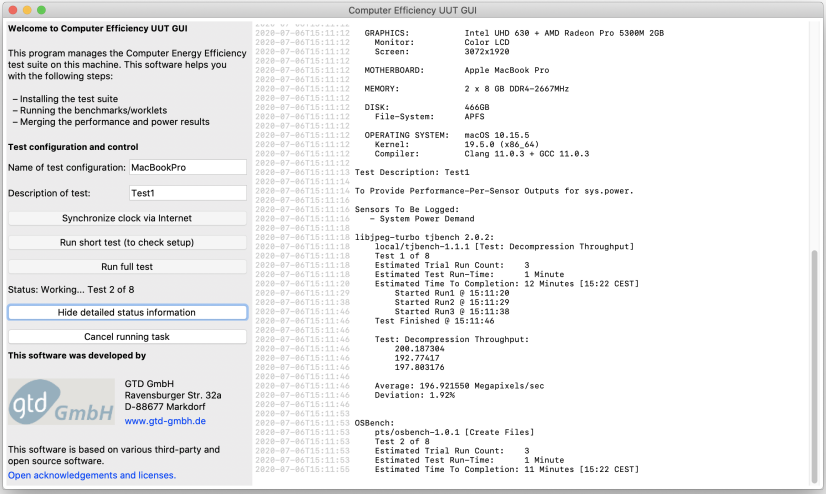
$$x_{1\,efficiency} = \frac{\left(\frac{1}{worklet\ time}\right)}{Average\ Watts}$$

Meta Energy Efficiency

A geometric mean is then calculated based on the individual efficiencies of all the executed worklets. Geometric mean calculations are most appropriate for combining values which have not a defined weighting.

$$Meta\ Energy\ Efficiency = \sqrt[n]{x_1 \cdot x_2 \cdots x_n}, \qquad n : Number\ of\ worklets$$

Test Suite GUI



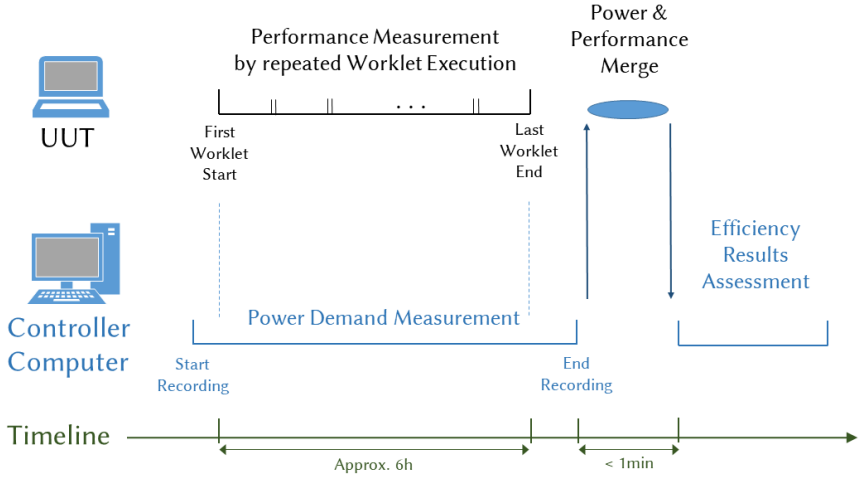
Result Computation

Efficiency is computed taking into account that:

- the Power Meter is connected to a controller computer that records the power measurements
- the UUT is not connected to any external devices to avoid a false increase in power demand
- power is measured continuously during the whole time the Test Suite is running, including all of the worklets
- however only the power measured during the execution of the worklets goes into the efficiency computation for each individual worklet
- UUT and controller computer clocks are synchronized via NTP before starting the test

This represents an important improvement wrt. existing Test Suites (Phoronix, SPECpower, etc.)

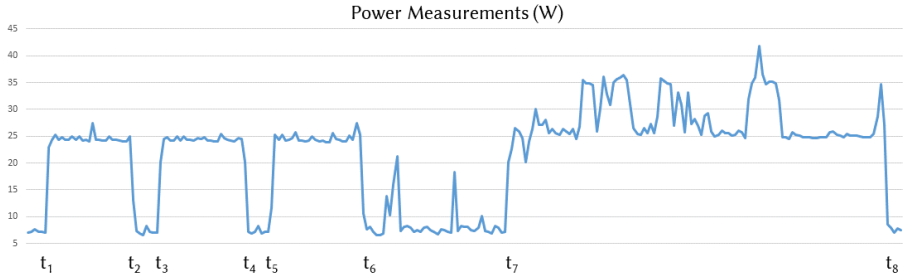
Test Suite Execution Workflow



Power Measurements Example

Measurements between worklet tries are discarded for efficiency computation

- t_1: Worklet1 try1 start 15:47:57
- t_2: Worklet1 try1 end 15:48:04
- t_3: Worklet1 try2 start 15:48:06
- t_4: Worklet1 try2 end 15:48:12
- t_5: Worklet1 try3 start 15:48:14
- t_6: Worklet1 try3 end 15:48:21
- t_7: Worklet2 try1 start 15:48:31
- t_8: Worklet2 try1 end 15:48:59



Result Viewer

Xonotic

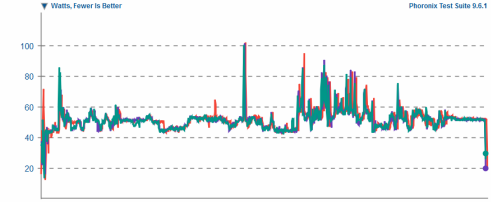
This is a benchmark of Xonotic, which is a fork of the DarkPlaces-based Nexuiz game. Development began in March of 2010 on the Xonotic game. *Learn more at OpenBenchmarking.org.*

Performance | System Power per Performance (Efficiency) | System Power Demand

Xonotic v0.8.2

System Power Demand

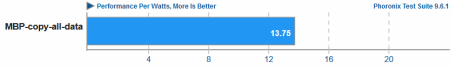
	Min	Avg	Max
MBP-2 (try 1)	12.83	51.2	101.35
MBP-2 (try 2)	16.46	50.9	100.66
MBP-2 (try 3)	13.46	51.0	99.82



Meta Performance Per Watts

Meta Performance Per Watts

Performance Per Watts

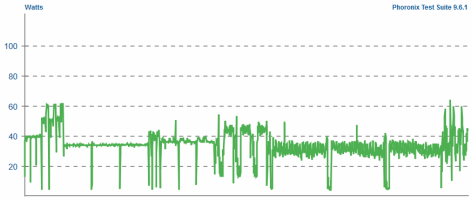


System Power Demand Monitor

System Power Demand Monitor

Phoronix Test Suite System Monitoring

	Min	Avg	Max
MBP-copy-all-data	4.26	34.2	63.72



Test Set-up and Procedure

GTD and Intertek Test Results Interpretation and Assessment

Tests Carried Out

- Intertek has run the Test Suite on 7 computers
- GTD has run it on more than 6 computers
- The complete Test Suite runs in ca. 6 hours in its current configuration
- We use the Yokogawa WT310 Power Meter

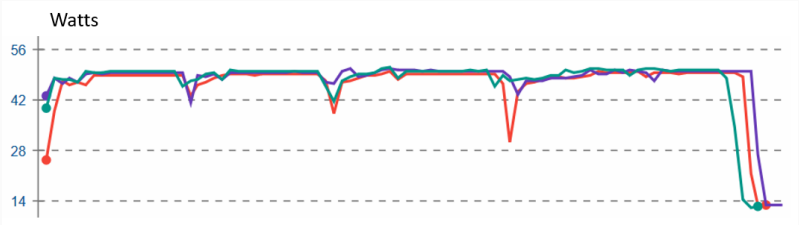
Meta Efficiency Comparison

Computer	Test 1	Test 2	Test 3	Average	Deviation
MacBook Pro 2019[3]	3.74	3.78	3.95	**3.82**	0.09
Windows Desktop 2018[4]	2.19	2.23	2.27	**2.23**	0.03
Windows Laptop 2019[5]	2.80	3.0	2.96	**2.92**	0.09
Ubuntu Linux Laptop 2020[6]	3.37	3.43	3.44	**3.41**	0.03
Windows Laptop 2020[6]	1.85	1.86	1.76	**1.82**	0.04
Arch Linux Desktop 2015[7]	1.98	2.0	N.A.	**1.99**	N.A.
Windows Desktop 2015[7]	1.32	1.34	N.A.	**1.33**	N.A.
Ubuntu Linux Desktop 2015[8]	1.57	1.58	N.A.	**1.58**	N.A.

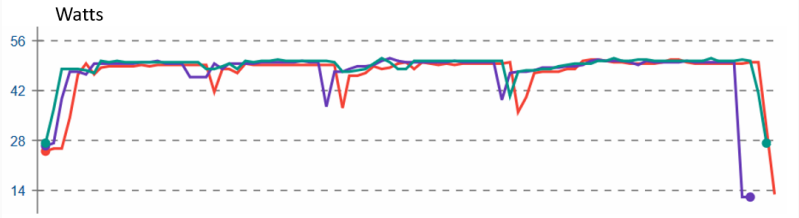
Higher is better!

[3] Intel Core i7, 16GB RAM, Intel UHD 630 + AMD Radeon Pro 5300M Graphics
[4] Intel Core i5-7500, 8GB RAM, Intel HD 630 Graphics
[5] Intel Core i5-8265U, 16GB RAM, NVIDIA GeForce GTX 1050 + Intel UHD 620 Graphics
[6] AMD Ryzen 7 3700X, 32GB RAM, NVIDIA GeForce GTX 1660 Ti Graphics
[7] Intel Core i7-6700K, 16GB RAM, NVIDIA GeForce GTX 980 Graphics
[8] AMD A8-6410, 8GB RAM, AMD Radeon R4 Graphics

Worklet Reproducibility Assessment

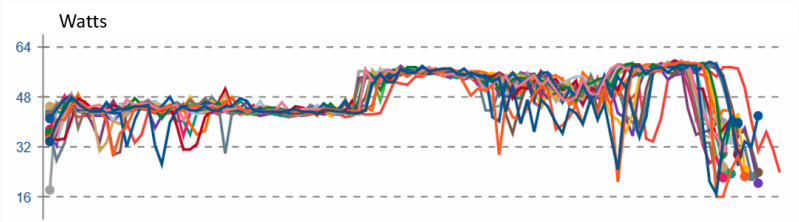


Day 1: 3 executions of 7-Zip Compression benchmark (Avg: 47.0 W, Dev.: 0.5 W, 26s)

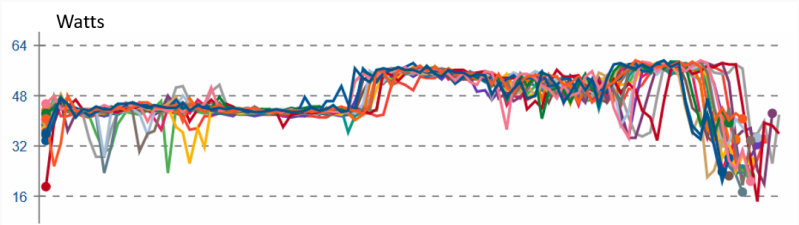


Day 2: 3 executions of 7-Zip Compression benchmark (Avg: 47.4 W, Dev.: 0.7 W, 26s)

Worklet Reproducibility Assessment

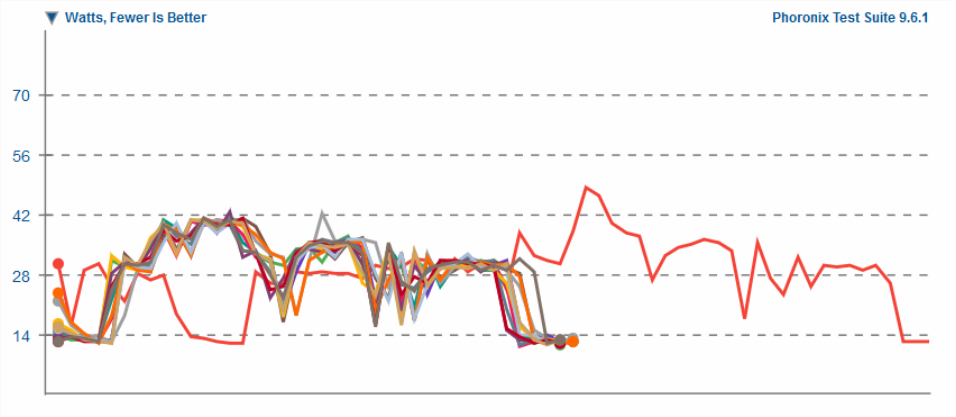


Day 1: 20 executions of DaCapo Tradebeans benchmark (Avg: 47.5 W, Dev.: 1.0 W, 30s)



Day 2: 20 executions of DaCapo Tradebeans benchmark (Avg: 47.1 W, Dev.: 0.6 W, 30s)

Worklet Reproducibility Assessment



15 executions of Gimp benchmark (17s the longest, 10s the other ones)

Software Dissemination and Feedback Tool

The test software, test procedure, this presentation and other resources are available via the following website:

gtd-gmbh.de/pceet