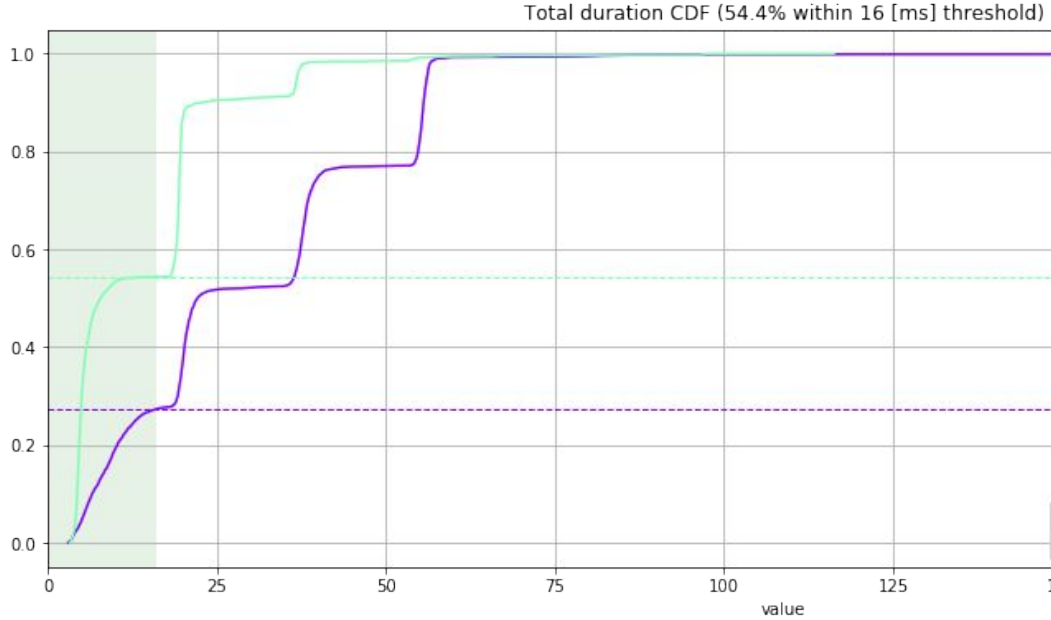
How to be Better Citizens

From change review to change testing



Patrick Bellasi
<patrick.bellasi@arm.com>

arm

Problem description

How to validate changes for power/performance?

Google has an internal validation CI support

- not accessible outside
- even if it runs on submitted patches, it does not report power/perf metrics

We would like to **verify** the **impact** of **new contributions** and **backported** patches

- especially for subsystems affecting power/performance
  - scheduler, power-management frameworks, PowerHAL are the main areas of interest
- possibly before changes get merged
- either to improve the change or raise new defects to follow up

A proper set of representative benchmarks are required

- an open and freely available framework could be on hand
- analysis results could be posted/linked to the gerrit pull request

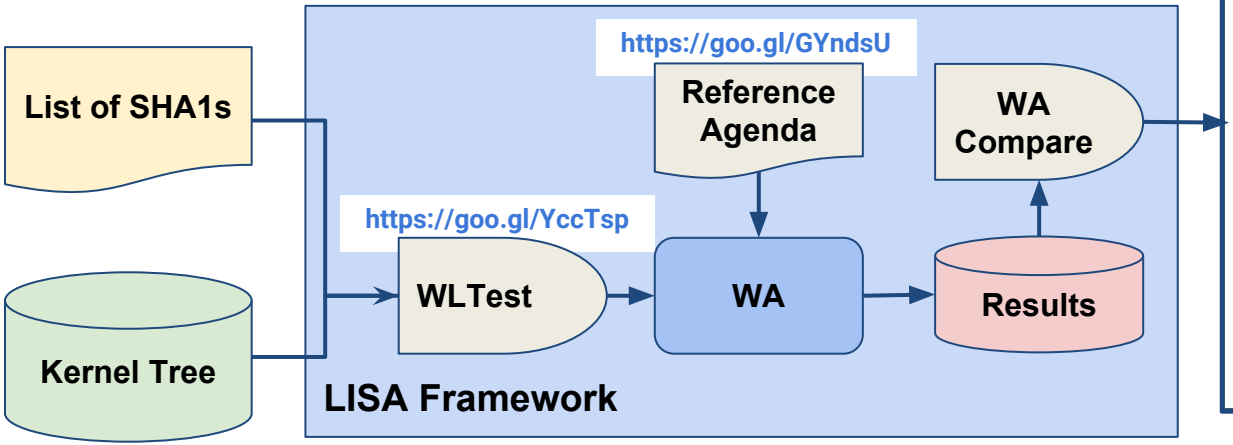
Example
https://goo.gl/qJPNQZ

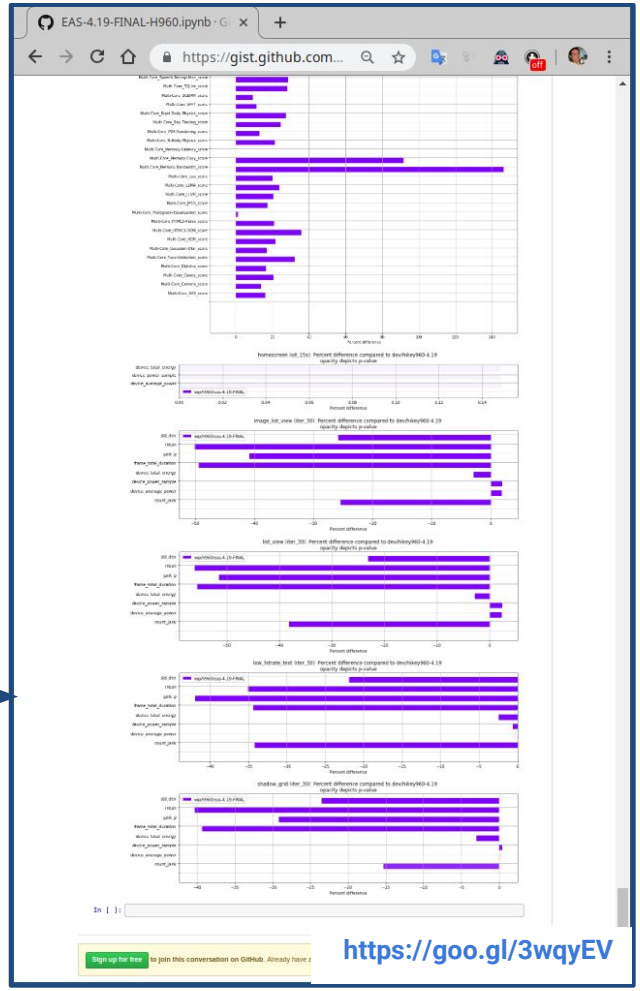
Proposed solution (1/2)

The Arm approach: WLTest in a nutshell

Automation support:

- **compile, flash** and **boot** a series of test kernels e.g. w/ and w/o a feature
- run a representative **set of benchmarks** and collect **power/perf** figures
- **plot** and **compare** collected metrics





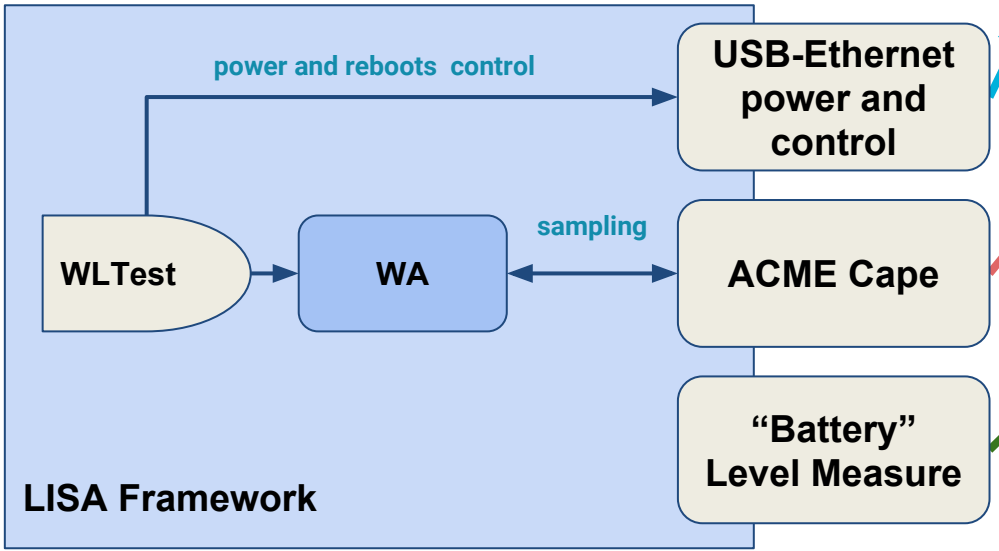
arm

Proposed solution (2/2)

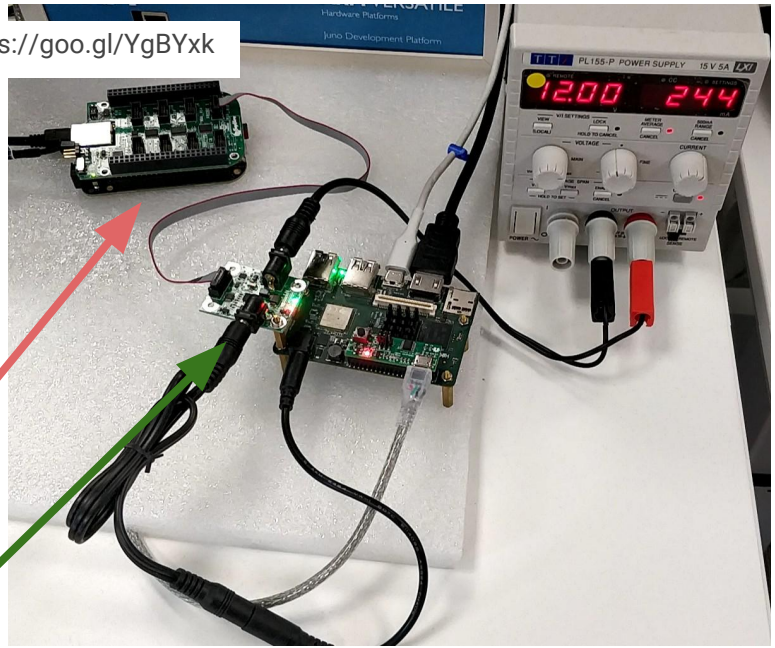
Example setup: Hikey960 + ACME CAPE Energy Meter

Power measurements are as important as performance metrics

- we don't need high accuracy
  - we found cheap energy meters to be up to the job
- we use an ACME Cape EM
  - which supports up to 16 channels, +1KHz



https://goo.gl/YgBYxk



arm

Main Discussion Points

Are we heading in the right direction?

How can we improve collaboration around this idea?

- the current proposal is mostly a set of guidelines to setup a simple and cheap on-desk testing solution

What's the best reference board/device?

- we need consistent and stable support, i.e max performance don't care
  - we usually focused on **Hikye960** and commercially available **Pixel devices**
- reasonably stable support for **AOSP** and recent **common kernel**
  - ACK 4.14 is going to be the reference kernel in 2019

Which benchmarks is better to use?

- Interactivity: Jankbench
- Energy-efficiency: homescreen, audio and video playback (exoplayer)
- Performance: PCMark and Geekbench

Battery power or rails power?

- battery power is easier to measure and represents the actual device juice
  - ... but requires care in properly setting the device for experiments (e.g. wireless connections and screen backlight)

Should we care about results anonymization?

- allows usage of new/secrete platforms

Thanks for the discussion

That's all... for Today