Greening of Streaming's February 2024 REM Hackathon Review

Participating members;
Agama, id3as, Humans, not Robots, DTG, Fraunhofer FOKUS, CTOiC, Kibo121, Simon Jones

Executive summary

Greening of Streaming believes the lack of ‘real world’ energy data impedes our ability to understand the energy demands on streaming infrastructures at scale. Today’s measurement is typically derived from lab-based extrapolation and modeling rather than real-world data. REM is Greening of Streaming’s (GoS) Remote Energy Measurement project, currently focussed on consumer device energy consumption in the real world. We will be integrating data centre and network energy measurements in the future.

Greening of Streaming has successfully put together an end-to-end scalable workflow - REM - that measures the energy consumption of devices using consumer-grade (cheap) smart plugs.

This review describes the first practical tests of the REM workflow, focusing first on the TV set. The overall architecture is defined, and high-level first results are given. Detailed information is, for now, reserved for member organisations.

Four initial outputs

The hackathons demonstrated the feasibility of remote energy measurement of streaming devices in real-world environments, showcasing an effective end-to-end workflow for energy consumption monitoring.

They demonstrated the feasibility of marker signals within video streams for synchronizing and identifying test sequences, significantly correlating energy consumption data with specific streaming content in what we term ‘energy signatures’.

The events underlined the scalability of the measurement system, successfully testing with multiple devices and indicating the potential for large-scale applications.

Challenges in universal access to streams across different devices were noted; we are now looking for the support of an established media player.

Context

Setting the scene

It is difficult to measure real-world energy in the consumer environment. Privacy issues mean we cannot, as an industry, simply configure technology to report on consumer's energy use habits—the technical challenges of where to measure and how to report that data are non-trivial.

Greening of Streaming's Working Group 8 (WG8) seeks to undertake pioneering measurement in the most challenging environment: the home.
In the long run, we aim to put probes into many homes that can inform us of the energy consumption of the users' digital media devices as they stream and consume content.

This series of hackathons is a first step in testing our ability to bring all the critical elements of the test workflow together into a usable platform.

We have identified a Wi-Fi-enabled smart plug with suitable API and cloud-sharing capabilities that a general consumer can easily opt in to and use to connect to our tests without (too much) support.

We can now remotely read the energy on these devices at regular intervals. This data is read into our central data repository, and from there, we are starting to produce real-time infographics to help us understand energy use during our test campaigns.

The smart plugs are connected between the TV and the consumer mains electricity.

Understanding how media consumption in the home drives energy consumption is essential for the industry to form a consequential energy reduction strategy.

Once the REM platform is active, it gathers the energy information. We then ask test volunteers to 'tune into' a test channel on their TV, and while they are tuned in for a few hours, we vary the streaming signals while simultaneously measuring their devices' energy consumption.

In the hackathon, we are working out how that test signal can best be sent and confirming that known signal changes cause expected energy use changes and are reflected in the data gathered.

One of the other issues in many tests of this type has been a lack of *control* of the 'actual' streaming format being tested (encoding / resolution / bitrate / dynamic range, etc.). In this experiment, we are setting up a streaming model that ensures we are testing known video streams to bring some determinism into these energy studies. So, in this test, we did not use any adaptive bitrates.

This review focuses on the feasibility of the measurement process itself. We can then shift our goal to producing actual data.

First Hackathon Jan 2024

Aim

The critical goals of this testing session were to:

1. validate an end-to-end test workflow,
2. check the accessibility of our energy readings and begin to test the challenges the model may face as we scale it to many more test locations,
3. check the feasibility of manually synchronising testing content playback from USB sticks,
4. ensure that the backend was able to capture real-time time-series data correctly,
5. find a suitable way for devices (TVs in this test) to connect to an online ('streaming') source,
6. form initial insights into consumer device capabilities and compatibility with test streams,
7. define a **marker signal** from a combination of black and white screens to identify test sequences from their **energy signature**.

Test environment description

Planning

1. Energy measurements were exclusively taken from TVs.
2. Each TV was plugged into the electrical mains through a smart plug to measure real-time energy consumption and collect it centrally.
3. The wattage resolution was one Watt (we expect to improve this in the long run, especially for mobile device energy measurements).
4. The Smart Plug data was polled every 20 seconds.
5. Eight testers across Europe joined a conference call for coordination, and a virtual drive containing test sources was shared for all members to prepare their USB sticks.
6. The files included black and white screens and a test sports sequence in HDR and SDR.
7. The TV's Energy modes and display settings were not adjusted; this gave a power measurement for typical personal viewing settings.
8. For file playback, the TV's native media player was used whenever possible.
9. For stream playback, the URL of the packaged source was entered into the TV's native Internet browser where available. For others, we had alternative engineering strategies.

FIRST SESSION (Morning)

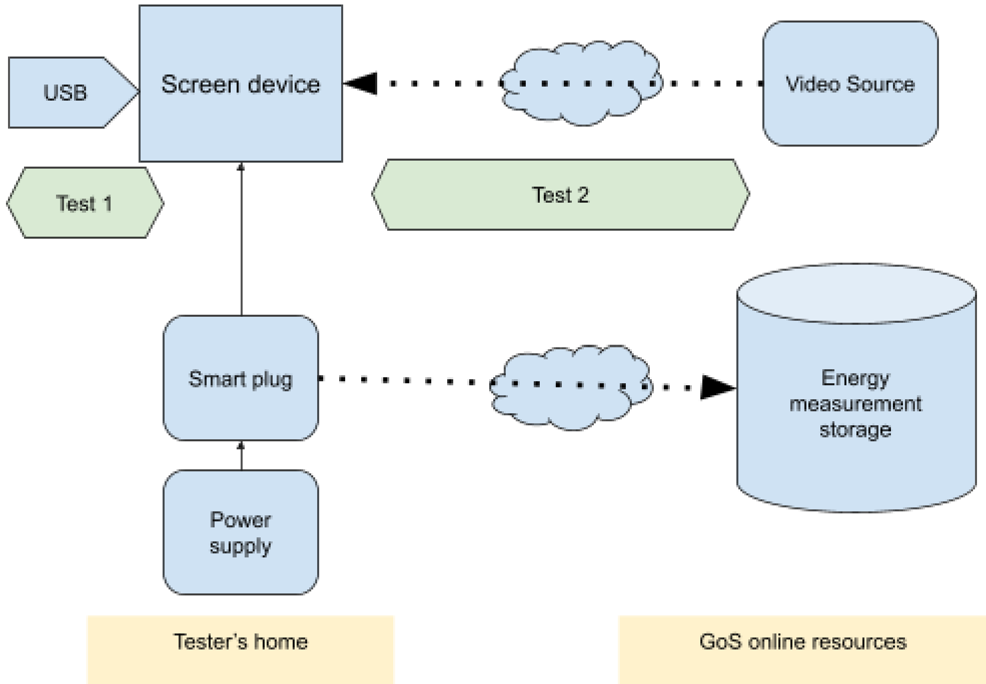
10. Manual testing was carried out by synchronised playback from the USB sticks.
11. Energy measurements were taken with the Smart TV on its static home page.
12. Measurements were manually captured for the six video files.

SECOND SESSION (Afternoon)

13. A source was set up online to send a controlled test signal to which the TVs were connected.
14. Five of the TVs were successfully connected to a live SDR stream.
15. Measurements were successfully automatically collected.

Top-level diagrams of our test environment

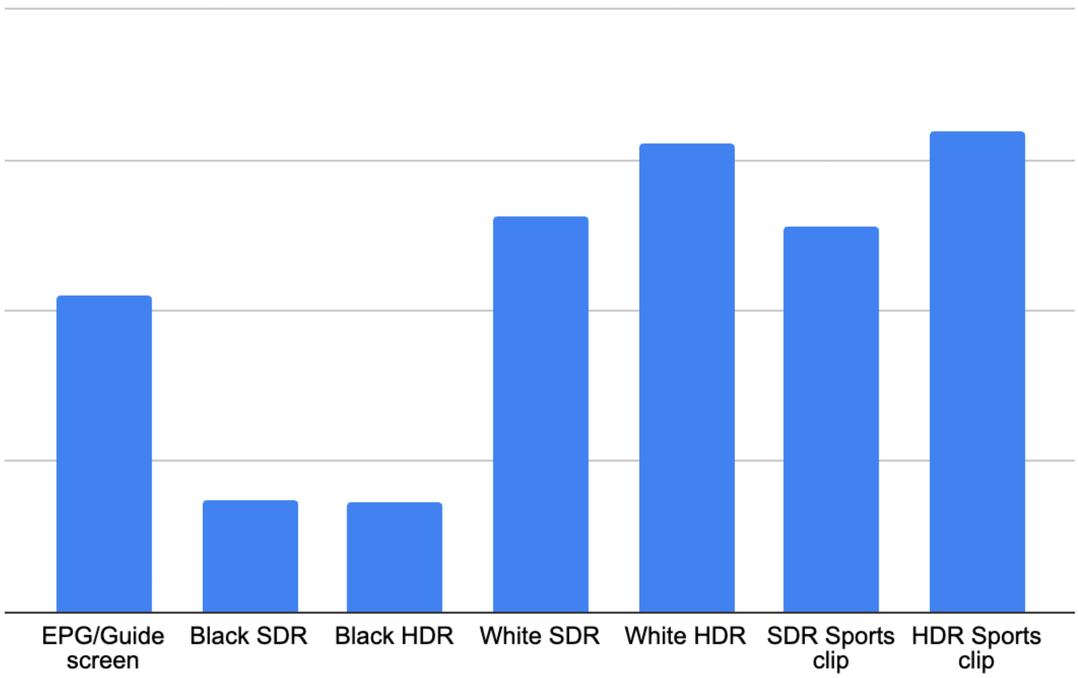
Logical Schematic (**public**)



Results of the first Hackathon:

First Session

The graph below represents the mean wattage of 8 TVs playing six video clips from a USB stick.



We have yet to include specific numbers on the graph since we do not believe our sample size was representative at this stage. However, the proportionate energy readings between the test signals during this hackathon are meaningful:

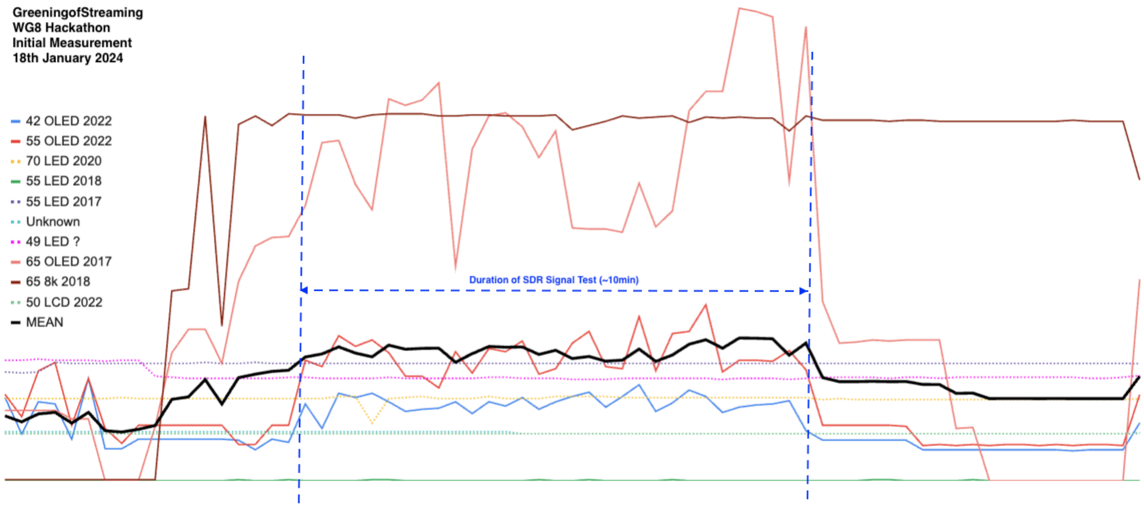
- The variation in energy that could be used to drive marker signal variations and create ‘energy signatures’ showed up clearly.
- The variation between Smart TV’s home screen or EPG (not playing video) and Black (both SDR and HDR) is significant and roughly equal to that between EPG and White and Sports HDR peaks.
- The White and Sports SDR content generated proportionately lower energy footprints than their HDR counterparts, so manual results were deemed effective to use as signatures for more comprehensive testing sources.

We can reasonably confidently play a combination of scheduled black and white **marker signals** within the stream and expect to see this “**signature**” in the collected energy data to synchronise more granular testing.

Second Session

The graph below shows the real-time data captured every 20 seconds. [The graph was generated manually from the automatically collected dataset ingested in .csv format.

At this stage, the automated graphing was still being engineered and was unavailable at the hackathon.



As seen above, when the test of the SDR signal was run, all the actively connected TVs consumed more energy.

TVs varied their energy consumption in line with the streaming activity to various degrees. For example, the prototype 8K TV didn't lower its consumption once the stream was started.

The flat lines represent TVs that were not connected to the stream, and as such, we treated them as 'noise' - however, these *are* included in the *mean* measurement (black), which is, therefore, flatter than we expected.

Since streaming start times varied in the hackathon, the sharp definition of the marker signal needs to be better defined in this graph. This test required a clear, programmatically driven start point. Part of the ongoing engineering work focuses on ensuring that programmatic synchronisation and coordination can be introduced (see Hackathon 2 below).

Notes:

- All TVs that could playback from USB could process HEVC.
- Only one of the eight used TVs refused to play back HDR content, which was PQ with HDR10 metadata, highlighting HLG's backward compatibility issues.
- Only one TV of the eight could *not* play video files from a USB stick.

Second Hackathon - February 2024

Aim

After the success of the first hackathon, the second hackathon's goal was to obtain clean energy signatures, complete testing of streaming SDR again, and compare it to HDR.

A key aim was to test the hypothesis that signal markers from black and white video sources would be identifiable and act as correlating signatures in the energy measurement. Reaching this goal should enable us to identify the test sequence segments and potentially synchronise them for data analysis.

Real-time graphing was, therefore, one of the central requirements under test.

Another further operational goal was to explore the challenges of universally accessing the test streams in a controlled yet simple way across various devices.

Test environment

As previously, a conference call was set up to synchronise the measurements.
The video streams were iteratively tested and prepared beforehand (thanks, Simon).

Source inputs to the test stream

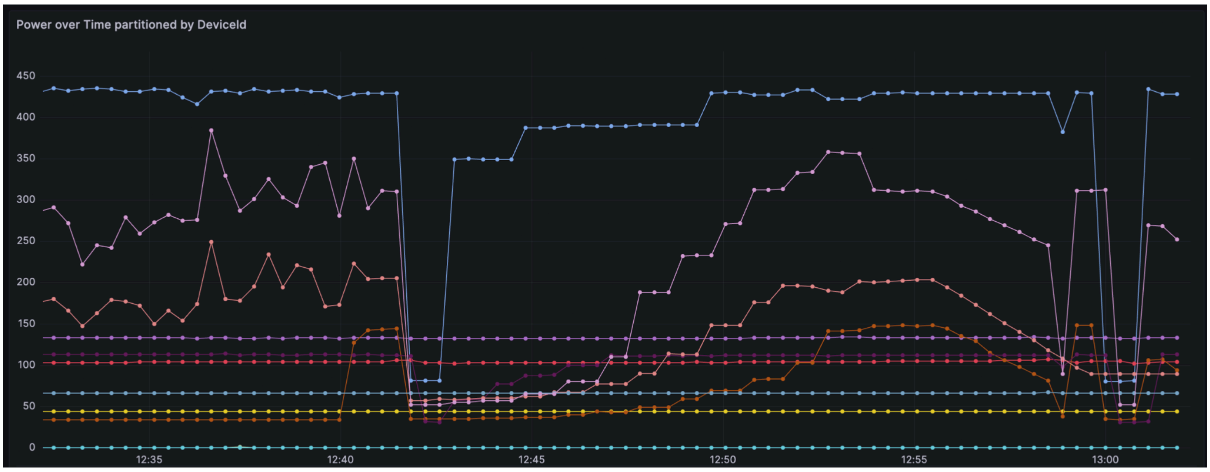
- The source video was carefully prepared to ensure a clearly defined reference throughout the streaming workflow.

Programmatic contents of test stream ~50-minute duration, identical for SDR and HDR).
- SDR and HDR sources in different streams
- Ten-minute countdown to allow for everyone to line up
- Signal marker sequence (White-black)
- Sports sequence
- Signal marker sequence (White-black)
- Luma Ramp up (whole screen)
- Luma ramp down (White square covering screen area, and reducing by half at each step)
- Signal marker sequence (White-black) - White full screen, then reduce the area by 50% for each step. See the PowerPoint attached to Simon Jones's WG6 post from 21/11/23

Streams were run several hours in advance to allow people to prepare their setup and check they could “tune in”.

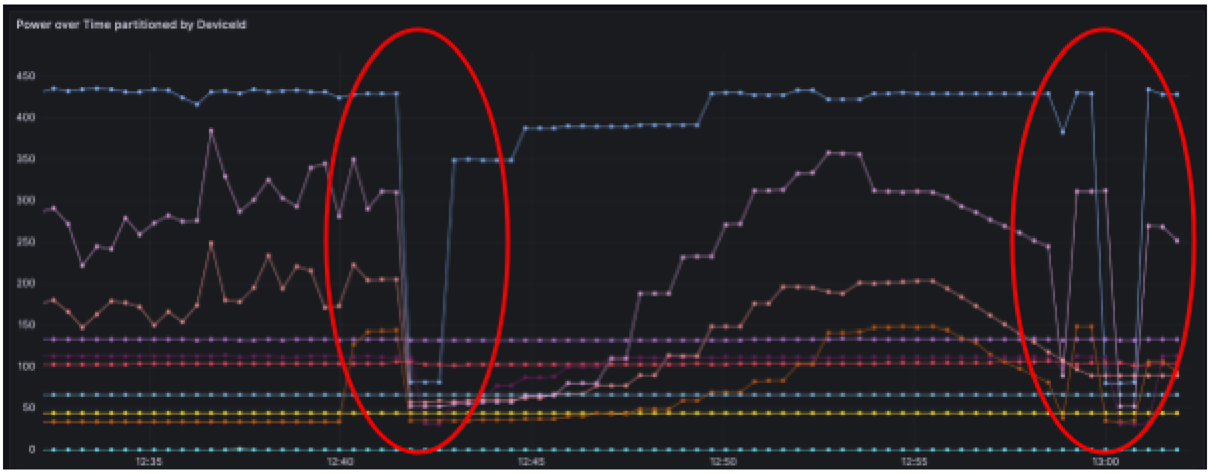
Observations

The following diagrams represent real-time energy readings only (not bandwidth or brightness). The lower yellow and blue flat lines represent TVs that didn't participate in the measurement process but have been left to show the Hackathon's un-doctored results (we have removed them from diagram 5 onwards).

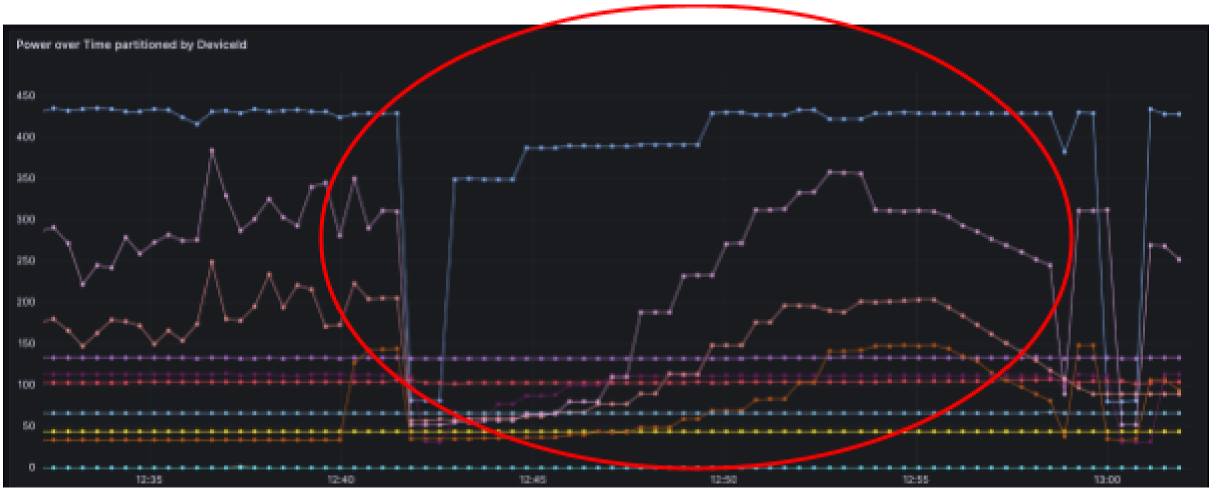


Graphing of results from HDR streaming.

One immediate observation is that using a polling time ***t*** with video test segments of duration ***3t*** works well (***t*** was 20 s in this test). The finer the granularity (the lower ***t***), the better for understanding content effects on energy consumption. However, there are likely limits to reducing the interval within which we can collect the data, both technically and cost-wise.



Above, we see the energy signature correlating to the white-black signal marker at either side of the luma ramp-up. Ultimately, this was a prominent indicator we were looking for in this hackathon, and it is a strong, clear signal marker represented as a 'signature' in the energy measurements.

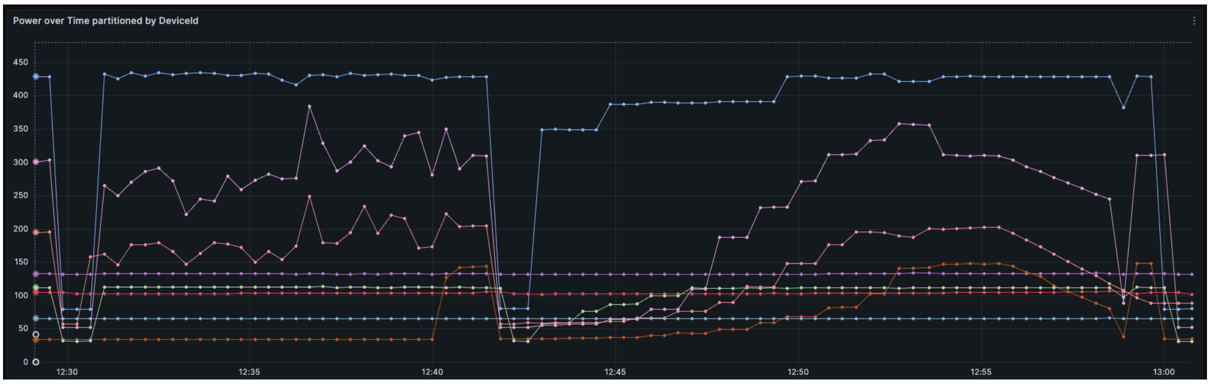


Above, we see the luma ramp-up.

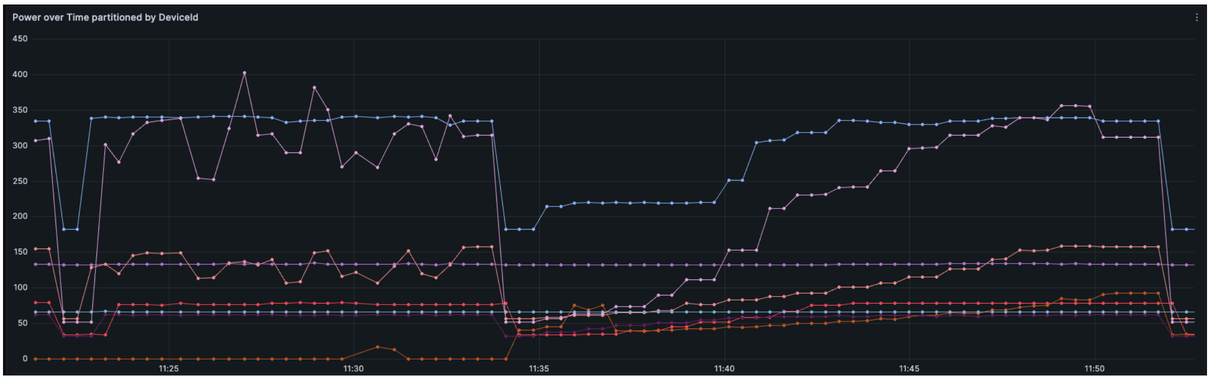
- The top line (large 8K prototype LCD screen) shoots up at the first step when the content is only one step up from black and reaches its maximum after only two more steps.
- The three OLED TVs in the middle step up almost proportionally to the screen brightness. Their relative consumption seems to be proportional to their screen size.

Above, we see the transition from an SDR signal to an HDR one. The red line is from an HDR-capable TV with simplistic backlight management. Unlike with the SDR content, this TV has no possible energy savings once it displays HDR; the backlight is “always on,” so the red line stays flat (unlike in the previous diagram when the same TV displayed SDR).

The two graphs below show the SDR and HDR measurements using our end-to-end workflow, with the removal of TVs that did not provide actual measurements (the straight lines are explained above):

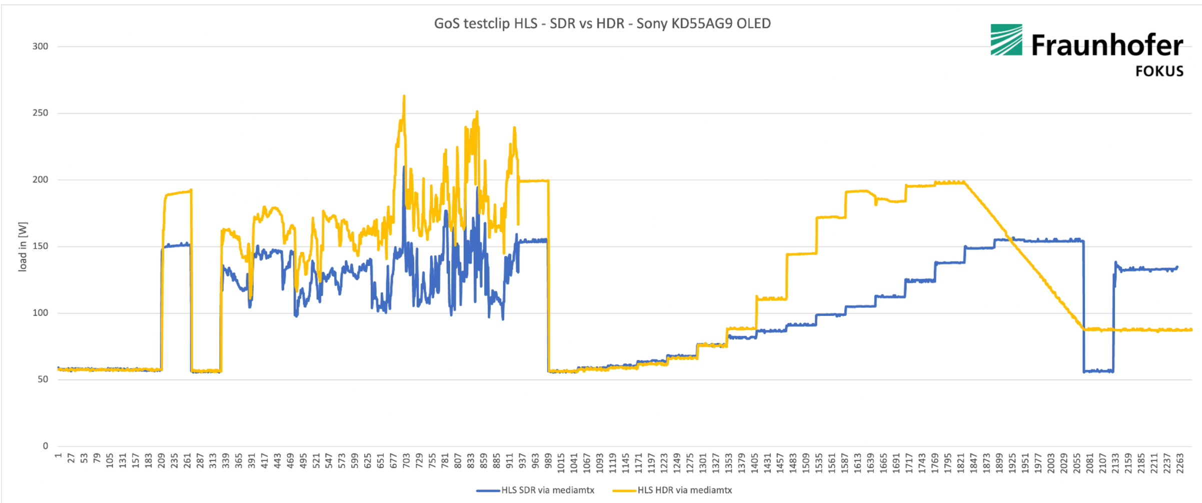


HDR



SDR

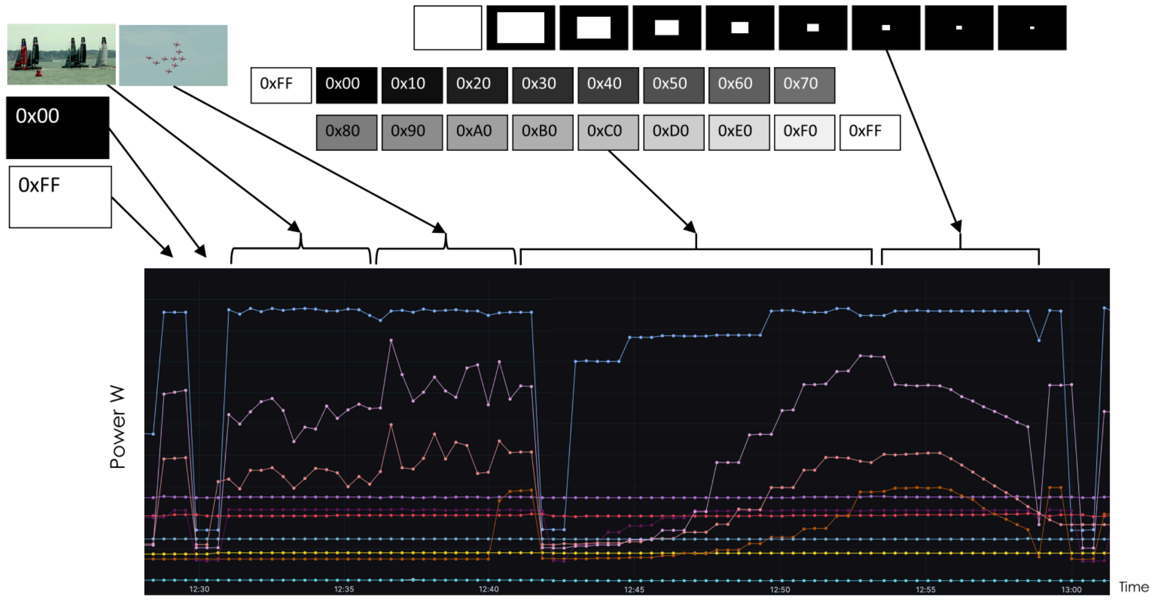
For calibration purposes, Fraunhofer, who took part in the Hackathon, also measured the consumption of one TV with lab-based measurement equipment, where the correlation is even more apparent in the following diagram.



Thus, precise observation of the signal marker's resultant 'energy signature' was visible through the REM platform.

The following diagram shows not just the signal marker but also starts to indicate the sensitivity to other early tests. It is too early to draw any conclusions on these small samples, but a correlation between video signal and energy signature can be associated at this stage.

Mapping Video Test Sequence to Energy Measurements



The video test sequence consisted of the following elements:

- The White-Black marker
- Video sequence - sailing
- Video sequence - air display
- White-Black marker
- Full-screen brightness ramp for Black to White
- Decreasing area White block, each step reducing by 50%
- White-Black marker

Preliminary conclusions

Proof of remote monitoring

Remote Energy Monitoring works! *The hackathons demonstrated its feasibility on streaming devices in real-world environments.*

After many months of trials of different measurement devices and streams to measure with, we are excited by the success of the end-to-end workflow that successfully measured energy in devices in real-time.

Scalability

Measurement

We have tested up to 10 devices simultaneously reading the same stream and reporting data to the same central repository. We will have an initial limit with part of the data aggregation to about 300 devices. We can run several groups of up to three hundred devices, so scalability is already potentially in the low thousands. If this turns out to be a limiting factor, it can be addressed.

Video Players

There were complexities in universally accessing streams across even our initial small sample set of devices.

In particular, some devices proved non-trivial to set up and had some issues with stream playback (although connectivity issues could not be ruled out in those cases).

Getting a simple single bitrate stream to play on on the different TVs using different players was sometimes challenging. This clear scalability challenge must be addressed to adopt the same approach for true scalability. To this end, we are seeking the support of an established video player.

Marker signal

In initial manual USB-stick-based tests, we used audio bleeps for synchronisation, which isn't practical for our goal of automated measurement.

Manual and automatic energy measurements proved that the smart plugs gave 'energy signatures' aligned with expectations for given marker signals. For black, we expected energy to drop, and it did, and for white signals, the energy consumption seemed to peak, too. Our first sample HDR energy reading was above the SDR version - again in line with our expectations. Still, as stated elsewhere, this is our first data set and requires further scaling and validation before we publish specific values.

Test Sequence

The test sequence was developed using an LCD TV for display; the results show significantly more variation in power consumed concerning content for OLED TVs. See the figure “Mapping Video Test Sequence to Energy Measurements” above, where the top trace is for a 65” LCD, and the one immediately below is for a similar-sized OLED.

Future REM enhancements

An improvement for the future will be to have the video seamlessly loop to improve preparation.

Our backend data aggregation should be able to be turned on and off very quickly to run ad hoc sessions.

For in-depth analysis, we would benefit from information on the test conditions (ambient light, TV settings, etc.); however, for establishing a picture of 'average domestic use', these details are not essential to provide critical insight into consumer media electronic energy use at scale.

We must test that the energy signature can be easily picked up from large data sets.
We would like to access the screen sizes of all devices to express our results in Watts per Square Metre.

We will make the white-black signal marker more distinct, with a black-white-black signal, with each segment set at 3x the polling interval (currently 20s). This should enable the identification of the start and end of sequences within the power measurements.

The test sequence was 50 minutes long to support the manual synchronisation of testing to the wall clock. We will explore changes to this to find optimums for both manual and automated testing.

The test sequence will be enhanced to explore the differences in the variation of power consumed concerning content between LCD and OLED TVs. It may also be possible to show the effectiveness of regional dimming between different LCDs.

Afterthought

Performing tests on a typical TV, we expect this to create a power variation of about 80-100W per household.

[Food for thought: even across our test group of just ten TVs, we could turn on and off a similar energy demand to boiling a single kettle with a simple change in streaming signal from black to white.]