

Power Modeling for Virtual Reality Video Playback Applications

Christian Herglotz, Stéphane Coulombe
Department of Software Engineering and IT
École de technologie supérieure (ÉTS)
Montréal, Canada

Ahmad Vakili
Summit Tech Multimedia
Montréal, Canada
www.summit-tech.ca

André Kaup (*Fellow IEEE*)
Multimedia Communications &
Signal Processing,
Friedrich-Alexander University
Erlangen-Nürnberg (FAU)
Erlangen, Germany

Abstract—This paper proposes a method to evaluate and model the power consumption of modern virtual reality playback and streaming applications on smartphones. Due to the high computational complexity of the virtual reality processing toolchain, the corresponding power consumption is very high, which reduces operating times of battery-powered devices. To tackle this problem, we analyze the power consumption in detail by performing power measurements. Furthermore, we construct a model to estimate the true power consumption with a mean error of less than 3.5%. The model can be used to save power at critical battery levels by changing the streaming video parameters. Particularly, the results show that the power consumption is significantly reduced by decreasing the input video resolution.

Index Terms—360 degree, video coding, smartphone, power

I. INTRODUCTION

Due to the recent advances in virtual reality (VR) techniques as well as the rise in the computing power of mobile devices, most modern smartphones are capable of executing real time VR applications. These applications can implement local playback, online streaming, or interactive gaming scenarios. One of the downsides of running these applications on portable devices is that the battery drains quickly because of high computational requirements.

This paper presents an initial work on power modeling of video playback in VR platforms using Android-operated smartphones. We consider the scenario of a smartphone user watching a video on his device using headsets like Google’s cardboard or Samsung’s Gear VR. The application performs pure playback such that no user interaction is possible except for basic user commands like play, stop, or pause. We consider two video source scenarios: local playback and online streaming via a Wi-Fi connection. Finally, various use cases like traditional 2D videos, 360° panorama videos, as well as 3D-360° videos are considered. Testing two different applications ensures that the proposed power model is not restricted to a single application.

Fig. 1 shows a high-level diagram of the main processes being performed during VR video playback. The control process sends commands to the processing entities and ensures synchronization. This control is usually performed by the central processing unit (CPU). The video data can be received by either a network interface like Wi-Fi or long-term evolution

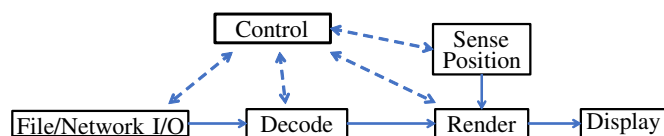


Fig. 1. Flowchart of the processing pipeline for VR video playback.

(LTE), or it can be read from an internal storage like a secure digital (SD) card. The compressed video data is then transferred to the decoder that reconstructs the output image. In this study, we only consider hardware decoding because most modern devices provide such functionality. Furthermore, decoding of high resolution videos (e.g., 4K) with software is often not feasible in real time.

Concurrently, a motion sensor like a gyroscope or an accelerometer, which is often available on motion processing units (MPUs), delivers information on the current orientation and motion of the smartphone in real time. The rendering process exploits the motion information and the raw video data to render the output images for the display. These processes are usually performed by the graphics processing unit (GPU). Finally, the output frames are sent to the display.

Similar studies have already examined various applications like online browsing, video filming, or standard video playback [1], where the focus was put on a high variety of applications instead of a detailed analysis on a single application. Furthermore, different hardware modules being active in the VR streaming pipeline such as the network interface [2], [3] and the display [4] were investigated. Many studies focused on the video decoding process which can be software-bound [5], [6] or hardware-bound [7], [8]. A detailed study on the power consumption during online VR video streaming has been performed in [9], where the power consumption of a smartphone was analyzed in detail. However, only few test cases with a restricted set of input videos were considered, such that no dependencies on high-level parameters like the frame rate or the resolution of the video were discussed. To the best of our knowledge, this is the first work analyzing the power consumption depending on sequence properties and constructing a respective power model.

In this work, we focus on the power of data processing such that peripheral devices like the display are neglected. Further work will include such considerations in our power model.

TABLE I
MAIN SPECIFICATIONS OF QUALCOMM'S SNAPDRAGON 820 SOM [10].

Module	Properties
CPU	Qualcomm Kryo quad core (64 bit) 2 low power cores (max 1.593 GHz) 2 high-performance cores (max 2.15 GHz)
GPU	Adreno 530 (up to 624 MHz)
Memory	4 GB of LPDDR 4 (up to 1866 MHz)
Multimedia	4K video decoding at 60 fps (H.264 or HEVC)

However, as the tested device in this paper uses common portable architecture, reported absolute power savings are also valid for portable devices with a display attached.

In Section II, we introduce the power measurement setup using a smartphone-like test board. Afterwards, in Section III, we present the modeling approach that is inspired by power modeling for video streaming. Then, in Section IV, we validate the proposed models on various video streams and playback settings. The trained models are analyzed and give insights into the power characteristics of the VR video process that are also discussed. Finally, Section V concludes this paper.

II. POWER MEASUREMENT SETUP

For the measurement of the power consumption, we chose the Eragon 820 software development kit (SDK) evaluation board [10]. It is equipped with a Qualcomm Snapdragon 820 system-on-module (SOM). The main specifications of the SOM are summarized in Table I. This system includes a quad-core processor that is used in many modern smartphones of manufacturers.

Android is used as the operating system and the impact of simultaneously running subroutines is minimized by uninstalling, disabling, or removing all unnecessary system processes from the application cache. Furthermore, services such as Bluetooth, global positioning system (GPS), and Wi-Fi are switched off when not used for measurements. To simulate the display output, an external screen is attached to the HDMI output port.

The power itself is measured through the main power supply jack using an external power meter as shown in Fig. 2. The power meter is a Monsoon High Voltage Power Meter (HVPM) [11]. It provides the input voltage of 12 V and reads the power consumption with a sampling rate of 5 000 kHz. To simulate online streaming, the video sequence is streamed via Adobe's RTMP network protocol. The evaluation board receives the live stream through the Wi-Fi interface.

As tested applications, we choose VaR's VR Player [12] and VRTV [13]. The former one is used to measure local playback, the latter one to measure streaming through a Wi-Fi network. Both applications allow various playback options as explained in the following. For each option, a unique identifier is defined.

First of all, classic 2D video playback is performed in which the video stream is just displayed on the device. Second, a stereo ('st') view is enabled, in which the same image is

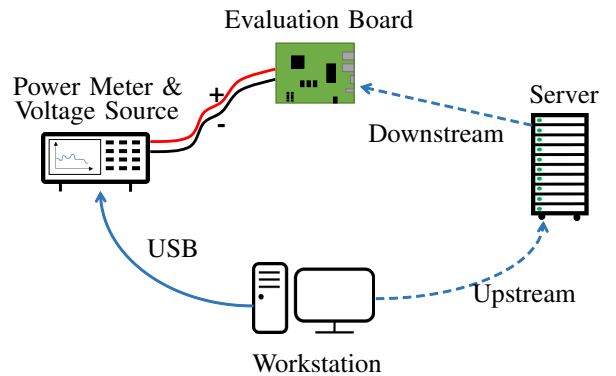


Fig. 2. Measurement setup with evaluation board, power meter, streaming server, and workstation for data analysis.

displayed on the left and the right part of the display, such that a VR headset can be used. In the basic case, head tracking is disabled such that the video is always located exactly in front of the eyes. Next, head tracking is enabled such that the displayed content changes dynamically based on the tracked position ('dyn'), although the screen is at a fixed position in the VR environment. The next option allows to watch 360° sequences, in which the VR user is immersed into the sequence ('360'). Finally, 3D VR is tested by using a left and a right view ('3D') to include depth information.

Further playback options were tested in the dynamic case. Regarding head tracking, in the standard setting, the sensor is chosen automatically. Additional measurements were performed for the explicit use of the gyroscope ('gyro'), the accelerometer ('accel'), and the magnetometer ('magn'). It is worth mentioning that head tracking only worked correctly in standard and in gyroscope mode.

Furthermore, lens distortion, zooming, and physical head movement using a turntable were tested. As the power measurements indicate that all these options have a very small influence on the power (the mean power for all test sequences differed by less than 1% in comparison to the default setting), they are not used and tested in the proposed modeling approaches. Table II summarizes the eight playback option sets that were measured.

A showcase power measurement for watching 10 s of video using VaR's VR Player with head tracking enabled and a 360° video format is shown in Fig. 3. The x-axis shows the time in seconds (s) and the y-axis the measured power values in watts (W). At the beginning, the app is launched and waits for the play command. The playback starts at approximately 0.1 s. A short initialization phase follows until 2 s, during which the maximum power is less than 6 W. Finally, after 2 s, the power reaches a static behavior with minimum and maximum powers of 2 W and more than 6 W, respectively. The playback ends at approximately 11.5 s.

Further tests showed that for longer playback cases, the power characteristics do not change. To make sure that the mean power is averaged over the static playback process, only 7 s of the complete process, which begins after the initialization phase and ends before the stop, are considered.

TABLE II

LIST WITH MEASURED PLAYBACK SETTINGS I TO VIII. 'A1' CORRESPONDS TO VaR'S VR PLAYER AND 'Ab' TO BOTH APPLICATIONS. 'x' AND '-' INDICATE THAT THE OPTIONS ARE SWITCHED ON AND OFF, RESPECTIVELY. THE INDICATOR 'a' FOR THE ROWS 'gyro', 'accel' AND 'magn' INDICATE THAT THE HEAD-TRACKING SENSOR IS CHOSEN AUTOMATICALLY BY THE APPLICATION.

	I	II	III	IV	V	VI	VII	VIII
Apps	A1	A1	Ab	Ab	Ab	A1	A1	A1
st	-	x	x	x	x	x	x	x
dyn	-	-	x	x	x	x	x	x
360	-	-	-	x	x	x	x	x
3D	-	-	-	-	x	x	x	x
gyro	-	-	a	a	a	x	-	-
accel	-	-	a	a	a	-	x	-
magn	-	-	a	a	a	-	-	x

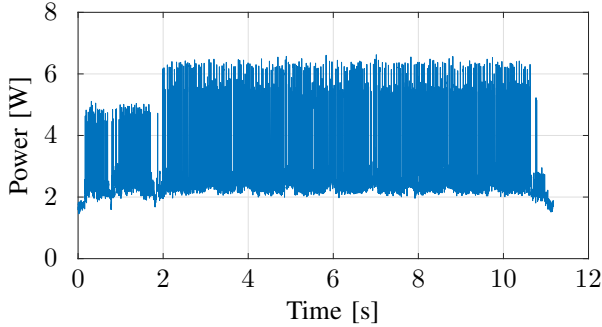


Fig. 3. Power consumption of the VR playback process including head tracking using VaR's VR Player (sequence AerialCity at 3840×1920 pixels, 30 fps, crf 28, cf. Table III).

To have more realistic power results, a constant offset power is subtracted from all power measurements. This constant offset corresponds to the power the evaluation board consumes in idle mode. In this case, idle mode is considered to be the case when no user application is running, i.e., the display shows the home screen. The advantage is that the differential power values are comparable to the power consumption of real smartphones that are not affected by overhead hardware modules like Ethernet connectors or additional USB ports.

III. POWER MODELING

We assume that each power component adds to the complete power linearly and write the complete power \hat{P}_{VR} as

$$\hat{P}_{VR} = \hat{P}_{cont} + \hat{P}_{rec} + \hat{P}_{dec} + \hat{P}_{sens} + \hat{P}_{rend} + \hat{P}_{disp}, \quad (1)$$

where the circumflex indicates that the power is estimated. \hat{P}_{cont} is the power needed for controlling the process, \hat{P}_{rec} is the receiver power, \hat{P}_{dec} the decoding power, \hat{P}_{sens} the sensing power, \hat{P}_{rend} the rendering power, and \hat{P}_{disp} the power of the display. This approach corresponds to the main processing steps identified in Fig. 1.

For the controlling power, no explicit model is known from the literature. For simplicity, we assume a constant power

$$\hat{P}_{cont} = p_{cont,0}. \quad (2)$$

In terms of the receiver power \hat{P}_{rec} , when fixed network settings are used, Sun et al. [3] found that a constant offset and a linear term depending on the bitrate b are the main factors describing the receiver power in Wi-Fi networks. A similar observation was made by Huang et al. [14] for 4G LTE networks. Hence, we model the receiver power as

$$\hat{P}_{rec} = p_{rec,b} \cdot b + p_{rec,0}, \quad (3)$$

where both parameters can have different values for different networks.

For the decoding power, we consider a model that originally targets energy estimation [8]. The model considers the variables frame rate f , resolution S , and bitrate b . We rewrite the model to enable power estimation and obtain

$$\hat{P}_{dec} = p_{dec,0} + p_{dec,f} \cdot f + p_{dec,S} \cdot S + p_{dec,b} \cdot b. \quad (4)$$

In this model, the parameter $p_{dec,0}$ is a constant offset power and the parameters $p_{dec,f}$, $p_{dec,S}$, and $p_{dec,b}$ describe linear contributions of f , S , and b . The resolution S is the product of the pixel width and the pixel height.

For rendering, we consider the power consumption of the GPU. A corresponding power analysis was performed by Chen et al. [15]. However, a suitable model was not proposed. In this work, we propose to also use the frame rate and the resolution of the sequence for power modeling because these two metrics are directly related to the computational complexity of the rendering process. As both parameters are already included in the decoder power models, we assume that no additional modeling parameter is required for \hat{P}_{rend} . As a consequence, the decoder parameters cover both decoding and rendering powers.

Next, the options of the applications discussed in Section II are considered ('dyn', '360', etc.). The settings affect both the rendering and the sensing process. Our measurements indicate that it is sufficient to model these settings using constants as

$$\hat{P}_{sens} = p_{gyro} \cdot F_{gyro} + p_{accel} \cdot F_{accel} + p_{magn} \cdot F_{magn} \quad (5)$$

and

$$\hat{P}_{rend} = p_{st} \cdot F_{st} + p_{dyn} \cdot F_{dyn} + p_{360} \cdot F_{360} + p_{3D} \cdot F_{3D}. \quad (6)$$

The parameters $F_{(\cdot)}$ represent flags showing with the values 1 and 0 whether the setting is activated. For example, $F_{st} = 1$ means that two views for the use with a VR headset are rendered and $F_{st} = 0$ means that only the classic single view (like in regular video streaming) is shown on the display.

Finally, as the display is not part of the measurements, we neglect its power consumption ($\hat{P}_{disp} = 0$). We gather the information from (1) to (6) and combine the parameters that are linearly redundant. These are the constant offsets

$$p_0 = p_{cont,0} + p_{rec,0} + p_{dec,0} \quad (7)$$

and the bitrate dependent parameters

$$p_b = p_{rec,b} + p_{dec,b}. \quad (8)$$

The resulting model reads

$$\begin{aligned} \hat{P}_{VR,a} = & p_0 + p_b \cdot b + p_{dec,f} \cdot f + p_{dec,S} \cdot S \\ & + p_{st} \cdot F_{st} + p_{dyn} \cdot F_{dyn} + p_{360} \cdot F_{360} + p_{3D} \cdot F_{3D} \\ & + p_{gyro} \cdot F_{gyro} + p_{accel} \cdot F_{accel} + p_{magn} \cdot F_{magn}. \end{aligned} \quad (9)$$

It includes 10 variables and $K = 11$ parameters and in the following, is referred to as advanced model.

Evaluations showed that within the limits of the content we tested, several variables only contribute marginally to the modeling accuracy such that they can be dropped. Hence, we construct a simplified model with the following method. We use (9) as a baseline, drop one by one each of the input variables, and discard those that cause an estimation error increase by less than 0.5%. The resulting model reads

$$\hat{P}_{VR,s} = p_0 + b \cdot p_b + p_{dec,S} \cdot S + p_{360} \cdot F_{360}, \quad (10)$$

which includes 3 variables and $K = 4$ parameters. In the following, this model is called simplified model. It comprises the parameters with the highest impact on the power consumption.

IV. EXPERIMENTAL RESULTS

The evaluation is split into four parts. First, we present the set of tested input video sequences in Section IV-A. Afterwards, we discuss the evaluation method which makes use of the mean estimation error in Section IV-B. Third, we evaluate the models from Section III in Section IV-C and finally, the modeling results are interpreted in Section IV-D.

A. Test Sequences

We measure the power consumption of the two applications with a high number of input sequences with different properties. These sequence-related properties comprise the content of the sequence, the frame rate, the resolution, and the format projection. As for the format projections, we choose classic rectilinear (recti) videos and equirectangular (equi) videos representing 360° content. Additionally, the top-bottom equirectangular format that allows 3D-360° playback is used. The rectilinear and equirectangular sequences are taken from the HEVC common test conditions [16] and from VVC documents [17], respectively. A personal archive was used for the 3D sequences, which were all recorded from a fixed camera position. The sequences and their main properties are listed in Table III.

The 'Cars02' sequence was recorded from a sidewalk showing several cars passing by. The 'Kitchen2' sequence was recorded inside a kitchen showing several people working. The 'Skatedance' sequence was recorded in the center of an ice rink showing girls practicing figure skating. Finally, the 'Wall6' sequence was recorded in a climbing hall.

The sequences are encoded with the x265 and the x264 encoders [18], [19]. These encoders were selected to compare the behavior of the two codecs H.264 and HEVC. The standard encoder settings are used in general; however, for each sequence, four instances are coded with the constant rate factors (crf) 18, 23, 28, and 33 to take different bitrates into

TABLE III
TEST SEQUENCES USED FOR POWER MEASUREMENTS. ALL SEQUENCES HAVE A DURATION OF 10 S.

Name	S	f	Projection	3D
BQSquare	416 × 240	60	recti	no
BlowingBubbles	416 × 240	50	recti	no
BasketballPass	416 × 240	50	recti	no
RaceHorses	416 × 240	30	recti	no
BQMall	832 × 480	60	recti	no
BasketballDrill	832 × 480	50	recti	no
PartyScene	832 × 480	50	recti	no
FlowerVase	832 × 480	30	recti	no
FourPeople	1280 × 720	60	recti	no
Johnny	1280 × 720	60	recti	no
SlideEditing	1280 × 720	30	recti	no
SlideShow	1280 × 720	20	recti	no
BQTerrace	1920 × 1080	60	recti	no
BasketballDrive	1920 × 1080	50	recti	no
Cactus	1920 × 1080	50	recti	no
Kimono	1920 × 1080	24	recti	no
AerialCity	3840 × 1920	30	equi	no
DrivingInCity	3840 × 1920	30	equi	no
DrivingInCountry	3840 × 1920	30	equi	no
PoleVault	3840 × 1920	30	equi	no
Cars02	3840 × 2160	30	equi	yes
Kitchen2	3840 × 2160	30	equi	yes
Skatedance	4096 × 2048	30	equi	yes
Wall6	3840 × 1920	30	equi	yes

account. In total, we use 192 sequences for our measurements, i.e., 96 for each codec.

Bit streams with a bitrate higher than 12 Mbps are discarded in measuring the VRTV online streaming case because a stable stream could not be established through the Wi-Fi connection. The resulting set includes 78 bit streams with at least two bit streams for each sequence from Table III. Furthermore, all sequences are tested for all configurations although some configurations would not constitute a reasonable use case, e.g., using a rectilinear video in a 360° environment. This enables to strictly separate the 360°-rendering power from the sequence specific decoding power.

B. Evaluation Method

For model evaluation, we transform the advanced model and the simplified model into a vector notation and obtain

$$\hat{P}_{VR,i} = \mathbf{A}_i \cdot \mathbf{p}_i, \quad (11)$$

where i indicates the model index ($i \in \{a, s\}$), \mathbf{p}_i is a vector containing all K parameters $p_{(\cdot)}$, \mathbf{A}_i is a matrix containing the values of all variables for all N measurements, and $\hat{P}_{VR,i}$ a vector containing all N estimated powers when using model i .

To separate the training data from the evaluation data, we perform cross-validation. For each validation iteration, we use the measurements corresponding to one source sequence for

TABLE IV

ESTIMATION ERRORS FOR THE MODELS INTRODUCED IN SECTION III. THE NUMBER OF TRAINED PARAMETERS OF THE MODEL IS INDICATED IN THE THIRD COLUMN. THE FOURTH COLUMN INDICATES THE APPLICATION THAT WAS TARGETED DURING CROSS-VALIDATION.

Model	Eq.	# Parameters	App	$\bar{\epsilon}$	ϵ_{\max}
$\hat{P}_{\text{VR},a}$	(9)	10	VaR	2.32%	13.3%
$\hat{P}_{\text{VR},s}$	(10)	4	VaR	2.25%	13.7%
$\hat{P}_{\text{VR},a}$	(9)	10	VRTV	3.28%	14.0%
$\hat{P}_{\text{VR},s}$	(10)	4	VRTV	3.47%	13.5%

validation and all the remaining measurements for training. For model training in each iteration of the cross-validation, the vector of model parameters is determined to minimize the sum of the squared estimation errors as

$$\min_{\mathbf{p}_i \in \mathbb{R}^K} \left[e_i = \sum_{j=1}^n \left(\hat{P}_{\text{VR},i}(j) - P(j) \right)^2 \right], \quad (12)$$

where j is the measurement index, n the cardinality of the training set, $\hat{P}_{\text{VR},i}(j)$ the estimated power for the j -th measurement using model i , and $P(j)$ the measured power of the j -th measurement. The training algorithm is a trust-region reflective algorithm proposed by Coleman et al. [20].

Finally, the models are evaluated using the mean relative estimation error and the maximum relative estimation error from validation as

$$\bar{\epsilon}_i = \frac{1}{N} \sum_{j=1}^N \left| \frac{\hat{P}_{\text{VR},i}(j) - P(j)}{P(j)} \right| \quad (13)$$

and

$$\epsilon_{i,\max} = \max_{1 \leq j \leq N} \left\{ \left| \frac{\hat{P}_{\text{VR},i}(j) - P(j)}{P(j)} \right| \right\}. \quad (14)$$

C. Estimation Errors

Table IV summarizes the estimation errors for VaR's VR player and VRTV separately. The table shows that the mean estimation error is below 3.5% for all tested cases. The maximum estimation error is below 15%.

The estimation errors for the VRTV application are generally larger (over 3%), which can be explained by Wi-Fi streaming which is less predictable than local memory reading. The observation that estimation errors of the second model only differ slightly in comparison with the first model confirms that the dropped parameters can be neglected for accurate modeling.

D. Interpretation

To assess the contribution of each modeling parameter in detail, we calculate their maximum contribution on all measured powers $P(j)$ using the advanced model. It is obtained by

$$C_{k,\max}[\%] = \max_{1 \leq j \leq N} \left| \frac{\mathbf{A}(j,k) \cdot \mathbf{p}(k)}{P(j)} \right| \cdot 100\%, \quad (15)$$

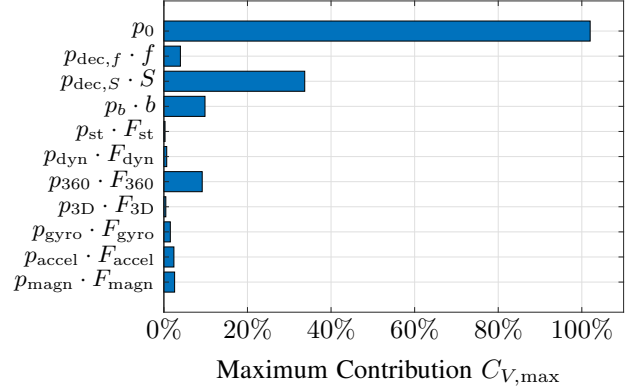


Fig. 4. Relative maximum contributions $C_{V,\max}$ of parameter-variable products to the complete power.

where k is the index of the variable used for modeling (e.g., the frame rate f , the resolution S , or a flag $F_{(\cdot)}$). $\mathbf{p}(k)$ is the corresponding model parameter as returned by training. The resulting values for all parameters in the case of VaR's VR Player are shown in Fig. 4. We can see that within the limits of the content we tested, the offset power is most important as in the maximum case, the estimated contribution is even higher than the complete measured power ($> 100\%$). The corresponding measurement was made for a low-resolution sequence with a small bitrate and a small frame rate.

This counter-intuitive observation is obtained because the measurement method does not allow to measure the offset separately from other variables. If VR playback is running, other variables like frame rate and resolution must be non-zero, too. As a consequence, the offset value, which fits the measurement best in terms of the error criterion (12), can be larger than the smallest measured power.

Further important variables are the resolution, the bitrate, and the flag indicating 360° processing, which all show a maximum contribution of more than 9%. A striking observation is that the influence of the frame rate is rather small. From observations in decoder modeling [7], [8], a higher influence would have been expected. Presumably, the reason is that for VR applications, the output frame rate is always set to the highest level to be able to follow head rotation as quickly as possible. Hence, differences in the frame rate of the input sequence only affect the decoding process.

To visualize the power consumption for two representative test cases, we plot the measured power and the modeled power distribution in Fig. 5. The measured power is depicted by the dark blue bars. The stacked bars below show the power distribution as returned by model 4. Assuming that modeling is accurate, most of the power is attributed to the offset (almost 1 W). For the 4K sequence, the resolution still accounts for approximately 0.4 W. The bitrate has a rather small influence (due to the high constant rate factor) and the projection format conversion requires approximately 100 mW.

These results indicate that a significant amount of power can be saved by switching to a low resolution for the input video. In case a smartphone recognizes that its battery is low,

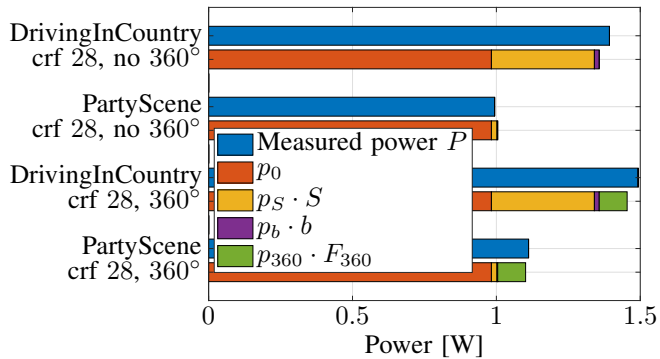


Fig. 5. Measured and estimated power consumption for four test cases in VaR’s VR Player. The blue bars correspond to the measured power, the stacked bars below the blue bars represent the corresponding power estimates by model 4, split up into the summands. Head tracking is switched on with standard sensors, the codec is HEVC.

it could switch from a 4K to a 2K sequence, which would result in estimated power savings of

$$\Delta p = (3840 \cdot 1920 - 1920 \cdot 1080) \cdot p_S. \quad (16)$$

Taking the example of the DrivingInCountry sequence in the 360° mode (Fig. 5), this would save 0.266 W, which accounts for 17.2% of the total power consumption.

To verify this claim, we encoded the corresponding sequence with the lower resolution of 2K and measured the corresponding power. A true power reduction of 18.3% could be observed. As similar observations could be made for other sequences, we can conclude that reducing the resolution is a valid and preferred way to reduce the power consumption.

V. CONCLUSION

This paper demonstrates that the power needed for live VR applications on mobile devices mainly depends on few parameters: the input video resolution, the bitrate, and the projection format conversion from ERP to rectilinear. This result is verified for two different VR applications. Two models were tested that both reach mean estimation errors below 3.5%. It is also shown that a significant amount of power can be saved by decreasing the input video resolution, which may, however, negatively affect the user experience.

Future studies may attempt to replicate and test our proposed methods and settings on other devices and software applications including wired professional VR headsets like the Oculus Rift or the HTC Vive. Interactive applications including user feedback can also be modeled and analyzed. Finally, the resulting models and power measurements are helpful tools to develop energy and power efficient VR solutions.

ACKNOWLEDGMENT

This work was supported by Mitacs and Summit Tech Multimedia (<https://www.summit-tech.ca/>) through the Mitacs Accelerate Program.

REFERENCES

- [1] A. Carroll and G. Heiser, “The systems hacker’s guide to the galaxy - energy usage in a modern smartphone,” in *Proc. 4th Asia-Pacific Workshop on Systems (APSys)*, Singapore, 2013.
- [2] L. Zou, A. Javed, and G. Muntean, “Smart mobile device power consumption measurement for video streaming in wireless environments: WiFi vs. LTE,” in *Proc. IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, Cagliari, Italy, June 2017, pp. 1–6.
- [3] L. Sun, R. Sheshadri, W. Zheng, and D. Koutsonikolas, “Modeling WiFi active power/energy consumption in smartphones,” in *Proc. IEEE 34th International Conference on Distributed Computing Systems (ICDCS)*, Madrid, Spain, Jun 2014, pp. 41–51.
- [4] Q. Liu, Z. Yan, and C. W. Chen, “Cloud-based video streaming with systematic mobile display energy saving: Rate-distortion-display energy profiling,” in *Proc. IEEE International Conference on Image Processing (ICIP)*, Phoenix, AZ, USA, Sept 2016, pp. 1504–1508.
- [5] T. Mallikarachchi, D. S. Talagala, H. K. Arachchi, and A. Fernando, “A feature based complexity model for decoder complexity optimized HEVC video encoding,” in *Proc. IEEE International Conference on Consumer Electronics (ICCE)*, Las Vegas, USA, Jan 2017, pp. 366–369.
- [6] C. Herglotz, D. Springer, M. Reichenbach, B. Stabernack, and A. Kaup, “Modeling the energy consumption of the HEVC decoding process,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 1, pp. 217–229, Jan 2018.
- [7] X. Li, Z. Ma, and F. C. A. Fernandes, “Modeling power consumption for video decoding on mobile platform and its application to power-rate constrained streaming,” in *Proc. Visual Communications and Image Processing (VCIP)*, San Diego, USA, Nov 2012.
- [8] C. Herglotz and A. Kaup, “Decoding energy estimation of an HEVC hardware decoder,” in *Proc. International Symposium on Circuits and Systems (ISCAS)*, Firenze, Italy, May 2018, pp. 1–5.
- [9] N. Jiang, V. Swaminathan, and S. Wei, “Power evaluation of 360 VR video streaming on head mounted display devices,” in *Proc. 27th Workshop on Network and Operating Systems Support for Digital Audio and Video*. ACM, 2017, pp. 55–60.
- [10] eInfochips, “Eragon 820 SOM development kit - technical datasheet,” <https://eragon.einfochips.com/pub/media/Productattachments/e/r/qualcomm-snapdragon-820-apq8096-eragon-820-technical-datasheet.pdf>, 2018.
- [11] Monsoon Solutions, Inc. High voltage power monitor - mobile device power monitor manual. <http://msoon.github.io/powermonitor/PowerTool/doc/PowerMonitorManual.pdf>. Accessed 2018-11.
- [12] AfterBreakdownGames. VaR’s VR Player. <http://afterbreakdowngames.com/>. Accessed 2018-11.
- [13] Chai Software. VRTV. <https://chaisoftware.wordpress.com/>. Accessed 2018-11.
- [14] J. Huang, F. Qian, A. Gerber, Z. M. Mao, S. Sen, and O. Spatscheck, “A close examination of performance and power characteristics of 4G LTE networks,” in *Proc. 10th International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys ’12. New York, NY, USA: ACM, 2012, pp. 225–238.
- [15] X. Chen, Y. Chen, Z. Ma, and F. C. Fernandes, “How is energy consumed in smartphone display applications?” in *Proc. 14th Workshop on Mobile Computing Systems and Applications*. ACM, 2013.
- [16] F. Bossen, “JCTVC-L1100: Common test conditions and software reference configurations,” Joint Collaborative Team on Video Coding (JCTVC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Geneva, Switzerland, Tech. Rep., Jan 2013.
- [17] P. Hanhart, J. Boyce, and K. Choi, “JVET common test conditions and evaluation procedures for 360° video,” Joint Video Exploration Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JVET-K1012, Jul 2018.
- [18] x265: H.265 / HEVC video encoder application library. x265.org. Accessed 2018-11.
- [19] x264: Encoder for H.264/MPEG-4 AVC video compression. x264.org. Accessed 2018-11.
- [20] T. F. Coleman and Y. Li, “An interior trust region approach for nonlinear minimization subject to bounds,” *SIAM Journal on optimization*, vol. 6, no. 2, pp. 418–445, 1996.