Modeling the Energy Consumption of the HEVC Decoding Process

Christian Herglotz, Dominic Springer, Marc Reichenbach, Benno Stabernack, and André Kaup, *Fellow, IEEE*

***Abstract*— In this paper, we present a bit stream feature-based energy model that accurately estimates the energy required to decode a given High Efficiency Video Coding-coded bit stream. Therefore, we take a model from literature and extend it by explicitly modeling the in-loop filters, which was not done before. Furthermore, to prove its superior estimation performance, it is compared with seven different energy models from the literature. By using a unified evaluation framework, we show how accurately the required decoding energy for different decoding systems can be approximated. We give thorough explanations on the model parameters and explain how the model variables are derived. To show the modeling capabilities in general, we test the estimation performance for different decoding software and hardware solutions, where we find that the proposed model outperforms the models from the literature by reaching framewise mean estimation errors of less than 7% for software and less than 15% for hardware-based systems.**

***Index Terms*— Decoder, energy, estimation, High Efficiency Video Coding (HEVC), model, power.**

I. Introduction

DURING the last two decades, video coding technologies have gone through a remarkable evolution. Former, rather simplistic approaches aiming at low resolution and low-quality video content have evolved into high-performance, complex, and efficient coding techniques that are capable of reducing bitrate for ultrahigh-definition (UHD) resolution videos. In the meantime, the common use case for video coding switched from traditional desktop personal computer (PC) applications to various different specialized devices such as televisions, cameras, tablet PCs, smartphones, or other embedded devices. These place new constraints on the encoding and decoding solutions and range from low complexity and portability to ease of integration and low-energy consumption.

In this paper, we focus on the decoding energy consumption that is typically important for portable devices such as tablet PCs or smartphones. Nowadays, these devices are able to perform real-time streaming and decoding of video content such that the user can watch videos anywhere and anytime. Due to the high complexity of the decoder, this process is one of the portable device's major power consumers, shortening the operating time until the battery has to be recharged. In [1], it is shown that a smartphone requires about a quarter of its power consumption merely for the decoding part, for a rather outdated H.263 decoder solution (display at medium brightness). Our measurements show that for the state-of-the-art High Efficiency Video Coding (HEVC) codec [2], the real-time software decoding of an HD sequence would drain the smartphone's battery in ∼4 h, not taking into account the peripheral energy consumers such as the display.

Hence, research aiming at reducing the amount of energy required to decode a video stream is a worthwhile task. Therefore, a common approach is to develop specialized decoding hardware such as the HEVC-decoder chip presented in [3]. Furthermore, for the different modules of the decoder, several optimized hardware solutions have been proposed [4]–[8]. Other approaches aim at the optimization of existing solutions [9], [10]. To this end, one proposal was to determine the decoder's complexity, which can be seen as an approximation to the energy consumption, to then be incorporated into the rate–distortion optimization (RDO) process on the encoder side. In this context, van der Schaar and Andreopoulos [9] proposed a generic complexity model for typical decoder processes, and Lee and Kuo [10] derived a dedicated model for the H.264 decoder that considers cache misses, interpolation filters, and the number of motion vectors per macroblock. Using this information inside the RDO process helps the encoder to construct a video bit stream of low decoding complexity. As a drawback, the encoder only has incomplete information about the decoder's behavior when using this model.

Although complexity can be seen as a close approximation to energy consumption, different solutions have been proposed using the actual power or energy consumption. Li *et al*. [11] proposed a power model based on the high-level bit stream parameters resolution, frames per second, and quantization. In [11], it is shown that the bit stream can be encoded in such a way as to match an energy requirement given by the decoding device by adaptively changing these parameters in the encoder. This approach was expounded upon by Raoufi and Peters [12] by incorporating the transmission of the bit stream via a wireless channel and concentrating on bitrate and rate of intra-coded frames.

The models presented before were all established for the H.264/AVC (advanced video coding) standard [13].

Manuscript received October 22, 2015; revised January 8, 2016 and June 14, 2016; accepted July 28, 2016. Date of publication August 10, 2016; date of current version January 5, 2018. This work was supported by the Research Training Group 1773 Heterogeneous Image Systems through the German Research Foundation. This paper was recommended by Associate Editor T.-S. Chang.

C. Herglotz, D. Springer, and A. Kaup are with the Chair of Multimedia Communications and Signal Processing, Friedrich-Alexander University Erlangen–Nürnberg, 91058 Erlangen, Germany (e-mail: christian.herglotz@fau.de; dominic.springer@fau.de; andre.kaup@fau.de).

M. Reichenbach is with the Chair of Computer Science 3, Friedrich-Alexander Universität Erlangen–Nürnberg, 91054 Erlangen, Germany (e-mail: marc.reichenbach@informatik.uni-erlangen.de).

B. Stabernack is with the Fraunhofer Institute for Telecommunications–Heinrich Hertz Institute, 10587 Berlin, Germany (e-mail: benno.stabernack@hhi.fraunhofer.de).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2016.2598705

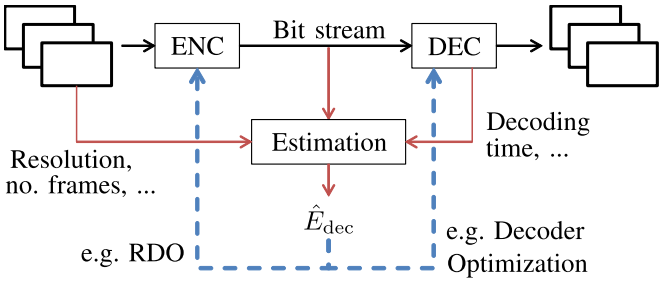

Fig. 1. Possible input variables (red arrows) and potential applications (blue dashed arrows) of the estimated energies. An input sequence is encoded in the encoder (ENC), and the resulting bit stream is transmitted to the decoder (DEC). The estimator can use information derived from the bit stream, from sequence specific parameters, or from decoding parameters. $\hat{E}_{\text{dec}}$ is the estimated decoding energy.

To estimate the decoding energy of HEVC decoders, Ren *et al.* [14] selected the processor-level parameters instructions, cache misses, and hardware interrupts as a foundation and achieved average relative estimation errors of less than 10%. In a different approach, we proposed a model based on high-level bit stream features [15] such as bitrate, resolution, and number of frames. Another work showed that for software solutions, the decoding time is highly correlated with the energy [16]. The more sophisticated model that we extend in this paper is presented in [17] (intra-frame decoding) and [18] (P- and B-frame decoding). It is based on bit stream features and aims at modeling the energy that certain parts of the decoding process require.

Hence, a high amount of different models was proposed in the past to estimate decoding energy or power with the goal of reducing the power consumption on the decoder side. Typical and potential applications of these models are the following.

1) Real-time encoding parameter adaptation to meet the decoder requirements. Due to the possibility to perform decoding energy estimations on the encoder side, the encoding parameters can be adapted to encode low-energy bit streams if the end user's device reports that battery power is low.
2) Decoding energy optimization of the video bit streams inside the RDO process. By incorporating the estimated decoding energy into the rate–distortion function, energy saving bit streams could be generated targeting the decoding system of the end user.
3) Comparison between different decoder solutions. A new algorithm can be compared with other solutions to check if it is energetically more efficient.
4) Information about the energy consumption of different decoder modules that correspond to certain bit stream properties. This information can, e.g., be used by developers to identify the module with the highest energy consumption, such that further development focuses on this module.

Fig. 1 visualizes possible input variables for the energy estimator, the output energy estimation, and applications.

In this paper, we adopt an incomplete model for HEVC decoding and extend it, such that it is able to estimate the energy consumption of any HEVC-conform bit stream. Furthermore, we compare it with several models from the literature and discuss their estimation accuracies for HEVC coded sequences as well as the advantages and disadvantages. Explanation on how the models can be applied will be given, and Section IV will show the performance of the modeling approaches for different software and hardware-based decoding systems.

The rest of this paper is organized as follows. Section II introduces the energy models and describes the parameters that are used to estimate the decoding energy. A major focus is put on the proposed feature-based model presented in Section II-A as the descriptions in prior work [17], [18] are incomplete. Afterward, a detailed description of the measurement setup is given in Section III. Section IV discusses the estimation accuracy of the different models by showing their estimation performance for different decoder implementations on different hardware platforms. Finally, Section V summarizes the conclusions of this paper.

II. Energy Estimation Models

This section introduces and explains the tested models. Our goal is to show the models' energy estimation capabilities for complete decoding systems including central processing unit (CPU) and memory access like random access memory (RAM). Peripheral devices that are not directly related to decoding, like a network connection or a display, are not considered.

Section II-A presents the proposed model based on bit stream features. In the model in Section II-B, the energy is estimated using the number of processor events. A similar model that aims at modeling the energy consumption of the RAM is presented in Section II-C. We include this model in our evaluations to be able to compare the RAM-modeling capabilities of the proposed model to a dedicated reference model from the literature.

Another approach related to the properties of the decoding process is shown in Section II-D, where the decoder processing time is used as a basis. At last, Section II-E shows four models that are based on high-level bit stream parameters (e.g., resolution, number of frames, and bit stream file size).

A. Energy Model Based on Bit Stream Features

The model presented in this section was first introduced for intra coding in [17] and extended to inter-coded frames in [18], where no in-loop filters were taken into consideration. In this paper, we finalize this model by including the in-loop filters [the deblocking filter (DBF) [19] and sample adaptive offset (SAO) [20]] such that any HEVC-conform bit stream can be modeled. Specific hardware properties of the decoding system are not modeled.

The model is based on typical features that an HEVC-coded bit stream can incorporate. From an abstract point of view, these features describe which subprocesses are executed to reconstruct the sequence. Such a subprocess can, e.g., be the prediction process of one block, the residual transformation of one block, or the DBF performed on a block boundary. Usually, such subprocesses are implemented in high-level

languages such as C++ and encapsulated in functions that are called multiple times during the decoding process. These functions show a rather constant processing flow; nevertheless, their energy consumption may differ significantly upon different calls. For instance, the input data that has an impact on the number of clipping operations or the current state of the cache system can lead to a varying energy consumption. In spite of this observation, it is assumed that this variance can be neglected due to averaging when executed multiple times during the decoding process. As an advantage of this concept, concrete knowledge about the implementation of these subprocesses is not required. Hence, the first property of a bit stream feature can be described as follows.

Property 1 (Bit Stream Feature): A bit stream feature can be associated with a subprocess during the decoding process. During the decoding of a single bit stream, these subprocesses can be executed multiple times. The subprocess consumes a specific and nearly constant processing energy e_i upon each execution.

In the next step, we define the set of features. A typical subprocess is, e.g., the prediction of a block of a fixed size. An example would be the bit stream feature "pla($d = 1$)" (see Table IX), which corresponds to the prediction of a block at depth $d = 1$ with the planar intra-prediction mode [21].

By such means, the complete decoding process is split up into subprocesses that are associated with the bit stream features. The second important property of these features is that they can be linked to syntax elements or variables as defined in the standard [22]. The occurrence and value of these syntax elements and variables are used to determine how often the bit stream features occur, and hence, how often the corresponding subprocesses are executed.

Property 2 (Bit Stream Feature): By analyzing the occurrence and value of certain HEVC syntax elements or variables, it is possible to find out how often a bit stream feature occurs. The number of occurrences is denoted by the *feature number* n_i.

Hence, for a given HEVC-coded bit stream, it is counted how often a subprocess is executed. Multiplying this number with a corresponding specific energy yields the complete required decoding energy related to this subprocess. Summing up these energies for all the bit stream features yields the estimated decoding energy

$$\hat{E}_{\text{dec}} = \sum_{\forall i} n_i \cdot e_i \tag{1}$$

where the index i denotes the bit stream feature index, n_i the feature number, and e_i the corresponding feature-specific decoding energy. Later, in Section IV-C, we will show how these feature-specific energies can be determined.

The set of features is summarized in Table IX and can be divided into five categories.

1) The general features (ID 1, …, 3) comprise the global processes (initialization of the decoding process and the slices).
2) Intra-frame coding features (ID 4, …, 10) comprise intra prediction for different block sizes and parsing corresponding flags.
3) Inter-frame coding features (ID 11, …, 24) describe inter prediction for different block sizes, parsing, and fractional pel filtering.
4) Residual coding features (ID 25, …, 34) comprise coefficient parsing and transformation.
5) In-loop filtering features (ID 35, …, 45) comprise the in-loop filtering processes DBF and SAO.

The feature number derivation is explained in detail in the Appendix. There, we show how feature counters implemented at positions that correspond to subclauses in the standard [22] can be used to derive the numbers. Using the syntax elements and variables used in the subclause, a condition can be defined that must be maintained to increment the corresponding feature number. With the help of this list, it is possible to implement the counters into any HEVC-conform decoder solution. For instance, every time a skip flag that holds the value "true" is decoded, the corresponding feature number $n_{\text{skip}(d)}$ at the current block's depth d (which relates to the block size) is incremented.

The feature numbers can be derived by a modified decoder. A corresponding tool based on the HEVC-test model (HM) reference software [23] is available online [24]. The tool incorporates counters that increment the feature numbers every time the feature occurs. To derive the feature numbers, we put special attention that no reconstruction, such as motion compensation or fractional pel interpolation, has to be performed. Hence, depending on the input sequence, the tool requires 20% to 60% less processing time than the complete decoding process. In comparison with the pure decoding process, when performing simultaneous decoding and feature number derivation, the complexity of the decoder rises by 4.5% in average.

As the complete set of bit stream features is rather detailed and complex, two different models are investigated as proposed in [18]. The accurate model (feature-based accurate *FA*) considers the complete set of bit stream features (in total 90 features including the depth dependencies), and the simple model (feature-based simple *FS*) considers a reduced set of 27 features including the depth dependencies. Table IX in the Appendix indicates which features are used in which model. The two models are proposed for the following reasons.

1) The accurate model (*FA*) is supposed to be used on the encoder side. When finding the best coding mode, next to the classic rate and distortion criteria, the estimated decoding energy can serve as a third decision criterion. Hence, energy saving bit streams could be constructed. For this purpose, a model using a high amount of features is better suited as more information on the energetic differences between the coding modes can be exploited.
2) The simple model can be used for applications where good estimates shall be achieved using low effort. Hence, this model is proposed for convenience and ease of application.

Thus, the accurate model incorporates a high amount of features, though some of them may only contribute little to estimation accuracy. Furthermore, other features, like weighted prediction or the use of temporal motion vector prediction,

have been tested, too, but showed a very low or even negative impact on the estimation accuracy. For the simple model, all the features that showed a rather low impact were discarded or merged. Note that the feature selection process for the models was rather intuitive, further work could aim at optimizing the feature set.

The outcome of the energy estimator can be tested in an online demonstrator [24]. It calculates the estimated decoding energy for any HEVC-conform bit stream. For further analysis, it also shows the details about the distribution of the energy among the bit stream features. We encourage visiting the Web site and testing the functionality for different bit streams and decoder solutions.

B. Energy Model Based on Processor Events

The first reference model (short *PE*) is proposed by Ren *et al.* [14]. It estimates the decoding energy using processor level information. Different processor events (PEs) are counted during the decoding process and interpreted as independent variables. Afterward, the energy is estimated using a multivariate adaptive regression spline (MARS) [25] model by

$$\hat{E}_{\text{dec}} = \sum_{i=0}^{M} \sum_{j=0}^{n_i} c_j \cdot B_j(x_i). \tag{2}$$

In this model, i is the PE index, M the number of PEs, and x_i the number of occurences of the ith PE. For each PE, n_i slopes are assigned with an associated basis function B_j. Ren *et al.* [14] use the basis functions as a constant or a hinge function of the form

$$B_j(x_i) = \max(0, x_i - k) \quad \vee \quad B_j(x_i) = \max(0, k - x_i) \tag{3}$$

where k is a constant. Hence, the MARS model constructs a piecewise linear relation between the independent variables and the target function.

Ren *et al.* [14] considered the three PEs instruction fetches, level-1 data cache misses, and hardware interrupts. In this paper, we only consider the former two PEs as all the peripheral hardware components are disabled in the measured devices. The construction process usually returns 13 parameters that describe the function.

A major drawback of this model is the fact that only pure software decoder solutions can be modeled where suitable profiling tools are available. Hence, the model could only be evaluated for decoding system (a) (see Section IV-A), where we used Valgrind [26] as a profiling tool. Furthermore, the derivation of the PE numbers is highly complex due to the instruction level analysis that needs to be performed during the decoding process of the considered bit stream.

C. Energy Model for the RAM

The model to estimate the energy consumption of the RAM is proposed in [27] and reads

$$\hat{E}_{\text{dec}} = e_{\text{ra}} \cdot n_{\text{ra}} + e_{\text{wa}} \cdot n_{\text{wa}} + e_{\text{rf}} \cdot n_{\text{rf}} + V_{\text{dd}} \cdot i_{\text{ss}} \cdot T_{\text{cp}} \cdot n_{\text{cycles}} \tag{4}$$

where the parameters e and the variables n represent specific energies and numbers, respectively, ra stands for read-access, wa for write-access, and rf for memory refresh. The last summand corresponds to the idle energy, where V_{dd} is the voltage, i_{ss} the steady-state current, T_{cp} the clock period, and n_{cycles} the number of clock cycles that are considered.

As we are solely interested in the pure decoding energy, we disregard the idle and refresh energy, which we assume to be stationary and reduce (4) to

$$\hat{E}_{\text{dec}} = e_{\text{ra}} \cdot n_{\text{ra}} + e_{\text{wa}} \cdot n_{\text{wa}}. \tag{5}$$

We obtain the number of read- and write-accesses using Valgrind [26], where the number of RAM read-accesses n_{ra} (instruction and data reads) corresponds to the last-level cache misses. For the write-accesses, we take the complete number of data writes as an approximation. Due to the write-back cache policy implemented in the tested system, data that are available in the cache are not always written to the RAM. A tool returning the exact number of writes was unfortunately not available for this paper; hence, the model only returns rough estimates for the RAM energy consumption.

Note that this model does not aim at estimating decoding energies, but at modeling the general RAM behavior. It is included in this paper to have a reference for memory modeling. As a similar analysis as for model PE has to be performed, the derivation of the number of read- and write-accesses is highly complex, and therefore only evaluated for system (a). In the following, this model is referred to by *M*.

D. Energy Model Based on Processing Time

The simplest model to estimate the energy consumption is presented in [16]. In this paper, it was shown that the processing time of the decoder is approximately linear to the decoding energy, such that the energy can be estimated by

$$\hat{E}_{\text{dec}} = E_0 + P_{\text{mean}} \cdot t_{\text{dec}} \tag{6}$$

where t_{dec} is the sequence-dependent processing time as, e.g., returned by the linux `time`-function or the C++ `clock()`-function. The parameter P_{mean} can be interpreted as the mean processing power and E_0 as a constant offset. For software that is not real-time constrained, this approach can return highly accurate estimates. As a drawback, estimates can only be obtained when the decoding process is executed once on the target device, because the processing time needs to be measured. Furthermore, real-time constrained decoders (such as typical hardware decoders) cannot be modeled using this approach as their processing time is equal to the playback time. In the following, this model is referred to by *T*.

E. Energy Models Based on High-Level Parameters

Further research has been conducted on the estimation using high-level, bit stream specific variables. To this end, Li *et al.* [11] estimated the decoding power of a real-time decoder using the bit stream properties frame size in pixels per frame S, frame rate f, and quantization q. We adapt this approach to estimate the decoding energy by

$$\hat{E} = P_{\max} \cdot \left(\frac{S}{S\max}\right)^{c_S} \left(\frac{f}{f\max}\right)^{c_f} \left(\frac{q}{q_{\min}}\right)^{c_q} \cdot t_{\text{dec}} \tag{7}$$

where we multiply the estimated power with the decoding time t_{dec}. $S_{\max}$, $f_{\max}$, and $q_{\min}$ are the maximum and minimum values for the frame size, the frame rate, and the quantization, respectively. The exponents c_s, c_f, and c_q and the maximum power $P_{\max}$ are the system specific parameters. Apparently, this model (high level with timing $H1_T$) is similar to the processing time-based model from Section II-D, but incorporates more variables. Note that we are targeting general decoding systems that are not necessarily real-time constrained. Hence, the variables frame size and rate, which have a very high influence on the processing power in real-time systems, can show a much lower impact in our configuration.

In later work, Raoufi and Peters [12] identified bitrate b and intra-refresh rate α (fraction of intra-coded frames of all the coded frames) as the suitable variables to determine the processing power. To estimate the processing energy, we again multiply with the processing time and obtain

$$\hat{E} = (c_1 \cdot \alpha \cdot b + c_2 \cdot \alpha + c_3 \cdot b + c_4) \cdot t_{\text{dec}} \tag{8}$$

where $c_1 \ldots c_2$ are the system specific parameters. This is the second high-level feature-based model using the processing time *($H2_T$)*.

In a further step, we tried to modify the model such that it better fits the properties of an energy and not a power estimator. Therefore, we exchange the bitrate b with a differently normalized variable, the mean bits per pixel b_{pixel}, and multiply the result with the complete number of pixels (the product of the number of frames N with the frame size S) and obtain

$$\hat{E} = (c_1 \cdot \alpha \cdot b_{\text{pixel}} + c_2 \cdot \alpha + c_3 \cdot b_{\text{pixel}} + c_4) \cdot N \cdot S. \tag{9}$$

As an advantage, we do not need to measure the decoding time t_{dec} on the decoding system to be able to estimate the energy. This model is referred to by *H2*.

As a last high-level model *(H3)*, we tested the estimation accuracy of the model presented in [15]. It is built upon the variables bits per pixel b_{pixel}, number of frames N, and frame size S and reads

$$\hat{E} = C + S \cdot N \cdot (\alpha + \beta \cdot b_{\text{pixel}}^{\gamma}) \tag{10}$$

where the constants C, α, β, and γ are the system specific parameters.

All the models presented in this section are summarized in Table I including their main properties.

TABLE I

DECODING ENERGY MODELS AND THEIR MAIN PROPERTIES. THE PARAMETERS (par.) WILL BE FITTED TO THE DECODING SYSTEMS, AND THE VARIABLES (var.) ARE FIXED FOR A GIVEN BIT STREAM. THE LAST COLUMN SHOWS IF THE DECODING PROCESS OF THE BIT STREAM HAS TO BE EXECUTED ON THE TARGET DEVICE TO BE ABLE TO OBTAIN THE ENERGY ESTIMATES

Model	based on	# par.	# var.	Ex. required
(*FA*)	bit stream features (accurate)	90	90	no
(*FS*)	bit stream features (simple)	27	27	no
(*PE*)	processor events	~ 13	2	yes
(*M*)	memory accesses	2	2	yes
(T)	processing time	2	1	yes
($H1_T$)	high-level features, time	7	4	yes
($H2_T$)	high-level features, time	4	3	yes
(*H2*)	high-level features	4	4	no
(*H3*)	high-level features	4	3	no

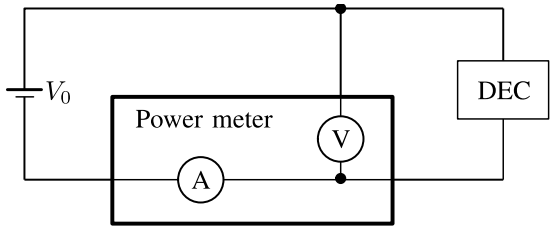

Fig. 2. Measurement setup with voltage source V_0, decoding device (DEC), and power meter including an ampere meter (A) and a voltmeter (V).

III. ENERGY MEASUREMENTS

In order to obtain the energy required for decoding complete sequences, a dedicated test setup was constructed, capable of measuring the energy consumption of the different video decoding solutions. These measurements are used to train and validate the investigated energy models.

The test setup used to measure the energy consumption of a decoding device is shown in Fig. 2. It consists of a voltage source, a high precision power meter (ZES Zimmer's LMG95), and the decoding system (DEC). To obtain the power consumption of a device, we measure voltage across and current through the main supply jack, such that the measured energy represents the combined energy consumption of all the modules (CPU, RAM, periphery, and so on) on the tested board. Furthermore, two measurements were conducted to determine the energy that the CPU and the RAM consume. Therefore, we identified the corresponding supply pins on the schematics of the tested board. The voltage was measured across the input and the output of the CPU and RAM power supply, respectively. The current was determined indirectly by dividing the voltage drop across the induction at the input of the power supply by the internal resistance of the induction.

Fig. 3 shows an example of the power that a complete decoding system consumes (measured through the main supply jack including the main energy consumers CPU and RAM). Fig. 3 (bottom curve) shows the power that the system consumes in idle mode when no user process is running. It is not perfectly constant due to, e.g., memory refreshes and background processes. Fig. 3 (top curve) shows the power consumption when a decoding process is started at $\sim$0.5 s. Here, we can see that the power rises to a higher level until the decoder finishes at $\sim$22 s.

If we now calculate the time-integral over the curves, we obtain the complete energy that the decoding system consumes in the observed time interval. As we are only interested in the pure decoding energy E_{dec}, we calculate the difference between both energies as

$$E_{\text{dec}} = \int_{t=0}^{T} P_{\text{dec}}(t)dt - \int_{t=0}^{T} P_{\text{idle}}(t)dt \tag{11}$$

where the time interval T must be equal for both integrals, $P_{\text{dec}}(t)$ is the power when the decoding system performs the decoding, and $P_{\text{idle}}(t)$ is the power in idle mode. Visually, the

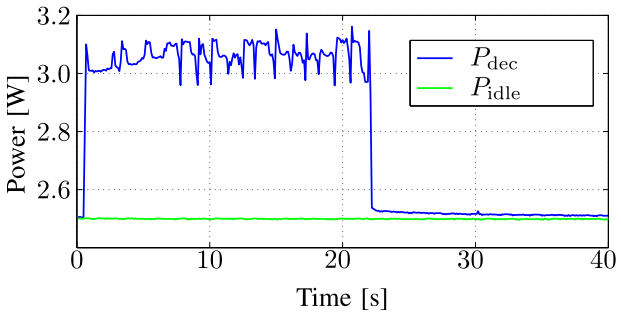

Fig. 3. Power consumption of the decoding system (a) (see Section IV-A) during the decoding process (top curve) and the idle state (bottom curve). The decoding starts at about 0.5 s and ends at 22 s.

resulting energy E_{dec} corresponds to the area between both curves.

To measure the energy consumption of an Intel CPU (see Section IV-A), we used the running average power limit system [28], which has been recently proven to return very accurate approximations of the true processing energy [29]. The energy values can be read from registers that are continuously updated. We obtain the energy of the complete decoding process by subtracting the energy value immediately before from the value immediately after the decoding process. In addition, we subtract the idle energy that we obtain by multiplying the mean processing time of the decoder with the mean idle power that we obtained in a further measurement. In our evaluation, we will use three energy values: package that includes the energy consumption of the complete chip, core that corresponds to the CPU cores, and uncore that corresponds to the energy consumption of the RAM.

As the measured energies for both measurement setups are subject to different noises resulting from temperature variations, kernel processes, hardware interrupts, power meter accuracy, and other factors, we found that it is not sufficient to perform merely a single measurement for each bit stream. Hence, we perform a measurement series and incorporate a statistical test to ensure validity.

To this end, we perform a confidence interval test as proposed in [30]. Consider a test sequence consisting of m measurements, where we assume the samples to be normally distributed: The mean of these samples is $\bar{x}$, and σ is the measurement's standard deviation. Given these parameters, the real mean value μ is located within probability α inside the interval

$$c_{\text{left}} = \bar{x} - \frac{\sigma}{\sqrt{m}} \cdot t_\alpha(m-1) < \mu < \bar{x} + \frac{\sigma}{\sqrt{m}} \cdot t_\alpha(m-1) = c_{\text{right}} \tag{12}$$

where $t_\alpha(m-1)$ denotes the critical t-value for m samples.

To make sure that the true mean value is not too far from the measured mean value, we check whether the calculated confidence interval is smaller than a constant fraction of the current mean value

$$\Delta c = c_{\text{right}} - c_{\text{left}} = 2 \cdot \frac{\sigma}{\sqrt{m}} \cdot t_\alpha(m-1) < \beta \cdot \bar{x}. \tag{13}$$

If the condition is not satisfied, we repeat the measurement until it is met. As the parameters we chose $\alpha = 99\%$ and $\beta = 0.02$. Doing so ensures that with a probability of at least 99%, the true decoding energy is not lower than 0.99 times and not higher than 1.01 times the measured mean energy E_{dec}. Note that for the measurement on the Intel CPU, the measured values are subject to much higher variation than in the other measurement setup, because no background processes were switched OFF. Hence, for this decoding system, we drop outlying values from the calculation of the mean energy.

IV. Evaluation

In this section, we verify the models for several different decoding systems (Section IV-A). The data set, i.e., the bit stream set that was evaluated, is given in Section IV-B. Section IV-C presents our evaluation methodology, and Section IV-D analyzes the estimation accuracies for all the tested decoding systems and all the presented estimation methods.

A. Decoding Systems

We set up six different test systems to validate the estimation accuracy of the models. Since one focus is on energy estimation in embedded systems, we decided on the Pandaboard [31] as a representative of a typical 32-bit ARM platform used in state-of-the-art smartphones and tablets. Three different software solutions were tested on this device. Furthermore, a PC featuring an Intel CPU was investigated aiming at the energy consumption of desktop PCs or portable tablet PCs. For the remaining two test systems, we selected field-programmable gate array (FPGA) platforms. On the first FPGA platform, a software decoder was executed on a soft intellectual property (soft IP) core to replicate the energy consumption on embedded, non-ARM CPU architectures. The second platform featured a decoder written in a hardware description language (VHDL). The results to this test series are supposed to show if the model is also valid for common, energy efficient hardware architectures like application specific integrated circuits (ASICs), which were not available for our research.

Three software decoder solutions were executed on the Pandaboard featuring Ubuntu Server 12.04 32-bit as an operating system. The board represents a typical, embedded, multimedia platform, equipped with an OMAP4430 SoC (System-On-Chip) [32] with a 45-nm feature size. The board incorporates two CPU cores on the SoC, each running at 1 GHz, and a variety of additional peripheral units [wireless LAN, LCD, camera, universal serial bus, HD multimedia interface (HDMI), Ethernet, and so on] that are commonly available on portable devices. The on-chip RAM is 1 GB, low-power double data rate synchronous DRAM (LPDDR2). For our tests, we restricted the UNIX runlevel to 1 in order to keep the energy consumption noise of the user-level processes and peripheral units to a minimum. The main properties of the software implementations are given in the following list.

(a) *HM-13.0 Software on the Pandaboard:* The HM software is the reference implementation of the HEVC coder. We compiled the software on the

Pandaboard using gcc for ARM with o3-optimization.

(a_{CPU}) *CPU Usage of the HM Software:* The measurement of the CPU aims at the same process as in system (a). In contrast, here we only measured the energy consumption of the CPU as explained in Section III to check if the energy consumption of this item can be modeled separately.

(a_{RAM}) *RAM Usage of the HM Software:* The same properties as for system (a_{CPU}), but this time the measurements were performed for the RAM.

(b) *Libde265 on Pandaboard:* This is an open-source implementation of the HEVC coder designed for real-world applications [33]. We tested version v0.7 that was compiled using gcc.

(c) *FFmpeg Single Core on Pandaboard:* For our third software implementation, we tested the HEVC decoder provided in the FFmpeg multimedia framework [34], version 2.8. An interesting feature of this framework is its ability to perform multicore decoding. Hence, we split up our evaluation into single and dual core processing. For this system, the decoder is executed on a single core.

(c_{dual}) *FFmpeg Dual Core on Pandaboard:* Here, we used the same configuration as with the preceding system, except that we allowed dual core decoding. Note that no voltage or frequency scaling, which helps in optimizing the energy consumption, has been tested.

The Intel-based decoding system has the following properties.

(d) *HM-16.6 Software on Intel i7:* For this test, we used a standard desktop PC featuring an Intel i7-4790 Quadcore (3.6 GHz with 16-GB RAM). To obtain the measurements from a realistic environment, we did not switch OFF any background processes, but only disconnected the network. The operating system was CentOS 6.X. Similar to system (a), we give results for the complete chipset, the CPU (d_{CPU}), and the RAM energy consumption (d_{RAM}).

The main properties of the two FPGA-based systems are as follows.

(e) *HM-16.5 on Scalable Processor Architecture (SPARC) Leon 3 Processor:* This decoding system features a SPARC Soft IP core that is a ready-to-use hardware architecture synthesized from VHDL. It can be modified [e.g., changes in cache size, enabling/disabling FPU (floating point unit) support, and so on] and customized at will. Afterward, the architecture is synthesized as a custom IC or on a reconfigurable hardware, like an FPGA. The Leon3 soft IP core is a 32-bit reduced instruction set computer processor created by Aeroflex Gaisler (now Cobham Gaisler) [35]. It is fully compliant with the SPARC-V8 instruction set. The processor was implemented onto an Altera Cyclone IV E FPGA, which was mounted on a Terasic DE2-115 board at a clock frequency of 50 MHz. For synthesis and implementation, Quartus 13 was used. All unnecessary components [e.g., FPU, memory management unit (MMU)], except for an instruction level cache, were removed. We used the Bare-C Cross-Compiler System for Leon3 to compile the decoding software. To remove all uncertainty regarding power estimation, no operating system ran on this processor. Therefore, no system for submitting jobs (e.g., a shell) is provided. For controlling purposes, we used direct hardware debugging via serial port (UART) and the debugging software GRMON2 [36]. Due to file in and out operations, the HM code cannot run directly on the Leon3; therefore, all calls were modified to use string buffers directly compiled in the executable.

(f) *Hardware Implementation on Altera Stratix V FPGA:* In contrast to the aforementioned platforms, which are based on CPU cores, a completely hardwired HEVC-decoder implementation was used to validate the energy modeling concept on hardware-based decoders. This decoder is based on a pure VHDL implementation, which reflects a pipeline architecture consisting of three stages: entropy decoding, motion compensation, and deblocking/SAO. The stages are coupled via coding tree unit double buffers, guaranteeing simultaneous processing at all stages. For memory access to the frame buffer and the reference frame memory, a 32-bit wide DDR3 memory interface is used, which is not supported by a cache, as is typical for other implementations and can be found in literature. The decoder is compatible with HM-15.0 and fully real-time capable for HEVC Main Profile @ Level 4.1. The decoder runs at a clock frequency of ∼150 MHz, and the bit stream interface is implemented on the basis of a simple hardware Ethernet stack, which only supports user datagram protocol packets using a simple streaming server on a standard PC as the sender. The output of the decoder features an HDMI interface for the connection to standard monitors. Further descriptions of the decoder can be found in [37].

B. Evaluation Sequences

For the evaluation of our model, we encoded ten different sequences from the HEVC test set (class A to class F) with three different quantization parameters (QPs 10, 32, and 45) and four configurations (intra, lowdelay, lowdelay_P, and randomaccess). In this standard data set (960 bit streams), the low number of QPs was chosen to avoid an excessive duration of measurements. Therefore, system (a) was more thoroughly tested using an extended data set (2880 bit streams) with a higher number of QPs ranging from 5 to 50 in steps of 5. For all the sets, one to eight frames were coded for each sequence. The sequences and their main properties are listed in Table II.

For the measurement of the Leon3 on the Altera Board (e), only the sequences taken from the classes B, C, D, and F were evaluated, due to the relatively small size of the RAM (613 bit streams). Furthermore, a different set of test sequences was evaluated for the hardware implementation (f) as it features some architectural constraints (412 bit streams). One of them is the fact that it can only decode and display sequences with HD-resolution (1920×1080 pixels). Besides, a QP lower than 20 and, depending on the content, QPs up to 35 could not be tested as the resulting bitrate was too high for the

TABLE II

PROPERTIES OF STANDARD EVALUATION DATA SET. THE SEQUENCES WERE TAKEN FROM THE HEVC TEST SET AND ENCODED WITH HM-13.0 USING THE CONFIGURATIONS INTRA, LOWDELAY_P, LOWDELAY, AND RANDOMACCESS. QP WAS SET TO 10, 32, AND 45

Name	Class	Resolution
PeopleOnStreet	A	2560 × 1600
Traffic	A	2560 × 1600
Kimono	B	1920 × 1080
RaceHorses	C	832 × 480
BasketballPass	D	416 × 240
BlowingBubbles	D	416 × 240
BQSquare	D	416 × 240
RaceHorses	D	416 × 240
vidyo3	E	1280 × 720
SlideEditing	F	1280 × 720

TABLE III

PROPERTIES OF EVALUATION DATA SET FOR THE HARDWARE IMPLEMENTATION [SYSTEM (f)]. THE SEQUENCES WERE TAKEN FROM THE HEVC TEST SET AND WERE ENCODED WITH HM-13.0. QP WAS SET IN THE RANGE FROM 25 TO 50

Name	Class	Resolution
BasketballDrive	B	1920 × 1080
BQTerrace	B	1920 × 1080
Cactus	B	1920 × 1080
Kimono	B	1920 × 1080
ParkScene	B	1920 × 1080

implemented entropy decoding engine. The list of sequences is given in Table III.

C. Evaluation Methodology

As a quality metric to rate the estimation performance of the models, we choose the relative estimation error as

$$\varepsilon = \frac{\hat{E}_{\text{dec}} - E_{\text{dec}}}{E_{\text{dec}}} \tag{14}$$

where $\hat{E}_{\text{dec}}$ is the estimated and E_{dec} the measured (true) decoding energy. This approach returns more meaningful results than using the absolute error as we aim at estimating the decoding energy accurately independent of the sequence resolution and length. Then, we average the estimation error for a given decoding system and a given model over the measured bit stream set by

$$\overline{\varepsilon} = \frac{1}{M} \sum_{m=1}^{M} |\varepsilon_m| \tag{15}$$

where m is the bit stream index and M the total number of evaluated streams in the corresponding data set. By averaging over the absolute value of the relative error, we obtain the mean value of the deviations.

All the steps that are performed to obtain these errors are summarized in Fig. 4. In the first step, to determine the

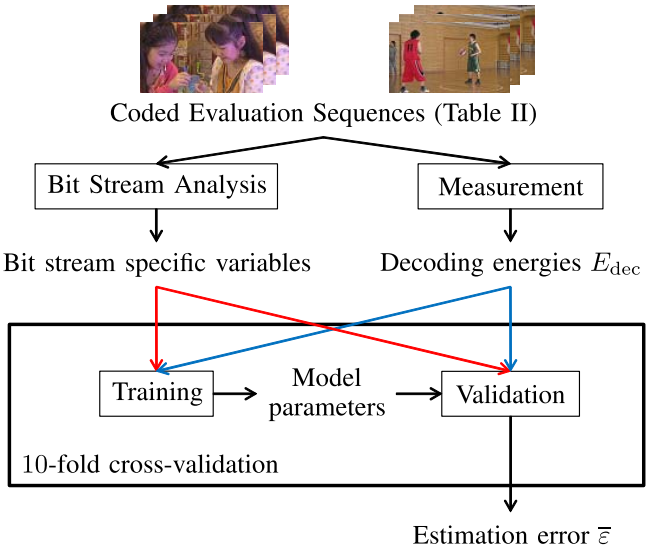

Fig. 4. Evaluation flow. The coded evaluation sequences (see Section IV-B) are analyzed for bit stream specific variables like feature numbers, high-level features, or decoding times as introduced in Section II. Furthermore, the decoding energy of these sequences is measured for all the decoding systems (see Section IV-A). Afterward, for each model and each decoding system, we feed the bit stream specific variables and the decoding energies into a tenfold cross-validation loop to train the model parameters and validate the estimation accuracy. As an output, we obtain the mean absolute estimation errors $\overline{\varepsilon}$ [see (15)] for the complete data set.

model parameter values (i.e., the specific energies), for each decoding system and each model, we perform a least squares fit using a trust-region-reflective algorithm as presented in [38]. As an input, we use the measured energies for a subset of the sequences (the training set) and their corresponding variables, i.e., the sequence-dependent feature numbers. As an output, we obtain the least squares optimal parameters (i.e., the optimal specific energies) for the input training set, where we train the parameters such that the mean relative error as shown in (15) is minimized. These model parameters are then used to validate the accuracy of the model on the remaining validation sequences.

The training and validation data set is determined using a ten-fold cross-validation as proposed in [39]. In this technique, we randomly divide the complete set of measured energies into ten approximately equally sized subsets. Then, for each subset, we use the other nine subsets to train the models as explained earlier. The trained parameters are then used to validate the remaining subset by calculating the relative estimation errors [see (14)] for all the sequences.

During our investigations, we observed that the estimation accuracy of the models highly depends on the used video sequences. Particularly, we noticed that the high-resolution sequences with a large number of frames are estimated much more accurately than the short sequences with a low resolution. Hence, to make sure that the comparisons performed in this section are fair, we only compare estimation errors for the exact same set of measured video data. We perform the analysis on frame level as all the presented models perform sufficiently well for large sequences. Frame-level values are obtained by subtracting the measured energy of a sequence

TABLE IV

MEAN ESTIMATION ERRORS FOR DECODING SYSTEM (a) AND ALL THE TESTED MODELS. THE MEAN RELATIVE ERRORS $\overline{\varepsilon}$ ARE GIVEN IN PERCENTAGE. THE RIGHT COLUMN GIVES THE AVERAGE ROW-WISE ESTIMATION ERRORS AND THE BOTTOM ROW THE AVERAGE COLUMN-WISE ESTIMATION ERRORS

	FA	*FS*	*PE*	*M*	*T*	$H1_T$	$H2_T$	*H2*	*H3*	∅
(a)	6.21%	6.27%	8.58%	8.86%	6.04%	5.96%	6.13%	20.30%	14.74%	9.23%
(a_{CPU})	6.41%	6.77%	9.30%	10.82%	7.30%	6.93%	7.90%	21.76%	15.72%	10.32%
(a_{RAM})	16.20%	15.80%	17.55%	18.54%	37.82%	31.38%	20.04%	23.57%	24.19%	22.79%
∅	9.61%	9.61%	11.81%	12.74%	17.05%	14.76%	11.36%	21.88%	18.22%	

TABLE V

MEAN ESTIMATION ERRORS FOR THE DECODING SYSTEMS (b) TO (d) AND A REPRESENTATIVE SUBSET OF THE TESTED MODELS. NOTE THAT THE MODELS *PE* AND *M* COULD NOT BE EVALUATED FOR THESE SYSTEMS

	FS	$H2_T$	*H2*	*H3*
(b)	4.49%	9.30%	29.06%	22.46%
(c)	6.77%	14.79%	20.93%	16.95%
(c_{dual})	7.48%	24.67%	21.03%	16.87%
(d)	12.83%	45.09%	24.69%	19.74%
(d_{CPU})	9.00%	35.86%	22.89%	16.20%
(d_{RAM})	50.54%	102.29%	99.25%	62.65%

containing $n-1$ frames from the energy of the corresponding sequence containing n frames. Note that consequently, the reported estimation errors from literature may differ significantly from those presented in this paper.

As shown in Section IV-B, we used different data sets for different decoding solutions. Hence, it is not possible to directly compare the estimation error for different decoding systems. As a counter measure, we evaluate the decoding process of each data set on system (a) such that the relative estimation performance can be compared with this system.

D. Discussion

Due to the high amount of data, we split the discussion of the results into three parts. First, we compare the estimation errors of the different models based on decoding system (a) to point out their strengths and weaknesses. Second, we talk about the different decoding systems and how accurately they can be modeled by the given approaches. Third, further observations are summarized.

1) Model Performance: The estimation errors of all the models for system (a) are summarized in Table IV. Looking at the average estimation errors of the models *FA* and *FS*, we can see that in general, they perform best (average errors smaller than 10%), which can be explained by the high amount of parameters. As both the models show a similar estimation error for all the decoding systems, and the model *FS* requires fewer modeling parameters, we will only evaluate the model *FS* in the following.

In general, all the models using the processing time (*T, $H1_T$,* and $H2_T$) return similar results for all the decoding systems. Sound estimations (around 6%, which is close to the errors of the models *FA* and *FS*) are especially obtained for decoding device (a). In contrast, modeling the RAM of

TABLE VI

MEAN ESTIMATION ERRORS FOR THE DECODING SYSTEMS (a) AND (e). THE DATA SET CONSISTS OF THE SEQUENCES FROM TABLE II THAT WERE SMALL ENOUGH TO BE DECODED ON THE FPGA

	FS	$H2_T$	*H2*	*H3*
(a)	6.69%	7.38%	45.75%	42.20%
(e)	6.07%	8.77%	59.10%	34.09%

TABLE VII

MEAN ESTIMATION ERRORS FOR THE DECODING SYSTEMS (a) AND (f). THE VIDEO DATA SET CONSISTS OF THE HD SEQUENCES FROM TABLE III

	FS	*H2*	*H3*
(a)	6.38%	14.55%	11.00%
(f)	14.89%	18.32%	27.51%

TABLE VIII

MEAN ESTIMATION ERRORS FOR A SUBSET OF 120 BIT STREAMS FROM THE STANDARD TEST SET CONTAINING EIGHT FRAMES (SEE TABLE II). WE CAN SEE THAT IN COMPARISON WITH TABLE IV, THE ESTIMATION ERRORS DECREASE SIGNIFICANTLY

	FS	$H2_T$	*H2*	*H3*
(a)	1.39%	2.40%	16.65%	15.00%
(a_{CPU})	2.91%	7.34%	18.49%	14.54%
(a_{RAM})	3.47%	9.91%	14.92%	10.50%
(d)	2.61%	13.49%	21.07%	15.19%
(d_{CPU})	2.94%	10.31%	19.64%	16.61%
(d_{RAM})	4.91%	33.90%	22.58%	8.96%

system (a) only works well with the model $H2_T$ (20.04%), which can be explained by the fact that this model takes into account the fraction of intra coded frames, which has a direct influence on the memory usage (inter frames require more memory access to reference frames). For the RAM, the other decoding time-based models fail which can be explained by, e.g., the influence of the RAM's unknown initial state. Hence, in the following, we will only evaluate the model $H2_T$ as in general, it returns the best results.

Model *PE* returns higher estimation errors than the decoding time-based models [8% in comparison to 6% for system (a)]. As it additionally requires a very complex profiling on the decoding system, the decoding time-based approaches seem to be more convenient in practical use. The values for the model *M* indicate that even a dedicated memory model does not estimate the memory energy more accurately (18.54%) than the proposed models *FA* and *FS* (around 16%), which

TABLE IX

BIT STREAM FEATURES USED TO ESTIMATE THE ENERGY CONSUMPTION OF A VIDEO DECODER. THE FIRST COLUMN SHOWS THE FEATURE ID, THE SECOND COLUMN ITS LABEL. IF THE LABEL IS FOLLOWED BY (*d*), THE FEATURE IS DEPTH-DEPENDENT AND DEFINED FOR THE DEPTHS SHOWN IN THE COLUMN "DEPTHS." THE NEXT THREE COLUMNS ("SUBCLAUSE" TO "INC") GIVE INFORMATION ABOUT THE POSITION AND THE CONDITION OF THE CORRESPONDING FEATURE NUMBER COUNTERS. THE CHECKMARK IN THE COLUMNS *FS* AND *FA* STATE IF THE FEATURE IS USED IN THE SIMPLE AND ACCURATE FEATURE-BASED MODEL, RESPECTIVELY

ID	Feature label	Depths	Subclause	Condition for syntax element or variable	Inc	*FS*	*FA*
1	E_0	-	-	-	-	√	√
2	Islice	-	7.4.7.1	`slice_type == I`	√	√	√
3	PBslice	-	7.4.7.1	`slice_type == P || slice_type == B`	√	√	√
4	intraCU	all	7.4.9.5	`pred_mode_flag == 1`	√	√	√
5	pla*(d)*	1..4	8.4.4.2.1	`IntraPredModeY == 0`	√	-	√
6	dc*(d)*	1..4	8.4.4.2.1	`IntraPredModeY == 1`	√	-	√
7	hvd*(d)*	1..4	8.4.4.2.1	`IntraPredModeY` $\in \{2, 10, 26, 34\}$	√	-	√
8	ang*(d)*	1..4	8.4.4.2.1	`IntraPredModeY` $\in \{3, .., 33\} \backslash \{10, 26\}$	√	-	√
9	all*(d)*	1..4	8.4.4.2.1	always	√	√	-
10	noMPM	all	7.4.9.5	`prev_intra_luma_pred_flag == 0`	√	-	√
11	skip*(d)*	0..3	7.4.9.5	`cu_skip_flag == 1`	√	√	√
12	merge*(d)*	0..3	7.4.9.6	`merge_flag == 1 && PartMode == 0`	√	-	√
13	mergeSMP*(d)*	0..3	7.4.9.6	`merge_flag == 1 && PartMode` $\in \{1, 2\}$	√	-	√
14	mergeAMP*(d)*	0..2	7.4.9.6	`merge_flag == 1 && PartMode` $\in \{4, 5, 6, 7\}$	√	-	√
15	inter*(d)*	0..3	7.4.9.6	`merge_flag == 0 && PartMode == 0`	√	-	√
16	interSMP*(d)*	0..3	7.4.9.6	`merge_flag == 0 && PartMode` $\in \{1, 2\}$	√	-	√
17	interAMP*(d)*	0..2	7.4.9.6	`merge_flag == 0 && PartMode` $\in \{4, 5, 6, 7\}$	√	-	√
18	interCU*(d)*	0..3	7.4.9.5	`CuPredMode != MODE_INTRA && cu_skip_flag == 0`	√	√	-
19	fracpelHor*(d)*	0..3	8.5.3.3.3.1	`mvLX[0] % 4 != 0 (&& mvLX[1] % 4 != 0)`	-	-	√
20	fracpelVer*(d)*	0..3	8.5.3.3.3.1	`mvLX[1] % 4 != 0`	-	-	√
21	fracpelAvg	all	8.5.3.3.3.1	`mvLX[0] % 4 != 0 || mvLX[1] % 4 != 0`	-	√	-
22	chrHalfpel*(d)*	0..3	8.5.3.3.3.1	`mvLX[0] % 8 == 4; mvLX[1] % 8 == 4`	-	-	√
23	bi	all	8.5.3.2	`predFlagL0 == predFlagL1 == 1`	-	√	√
24	MVD	all	7.4.9.9	`abs_mvd_greater1_flag == 1`	-	-	√
25	coeff	all	7.4.9.11	`significant_coeff_flag == 1`	√	√	√
26	coeffg1	all	7.4.9.11	`coeff_abs_level_greater1_flag == 1`	√	-	√
27	CSBF	all	7.4.9.11	`coded_sub_block_flag == 1 && ( xS, yS )!=( 0, 0 )`	√	-	√
28	val	all	7.4.9.11	`coeff_abs_level_remaining`	-	√	√
29	TrIntraY*(d)*	1..4	7.3.8.10	`CuPredMode == MODE_INTRA && cbf_luma == 1`	√	-	√
30	TrIntraC*(d)*	1..4	7.3.8.10	`CuPredMode==MODE_INTRA && (cbf_cb==1; cbf_cr==1)`	√	-	√
31	TrInterY*(d)*	1..4	7.3.8.10	`CuPredMode != MODE_INTRA && cbf_luma == 1`	√	-	√
32	TrInterC*(d)*	1..4	7.3.8.10	`CuPredMode!=MODE_INTRA && (cbf_cb==1; cbf_cr==1)`	√	-	√
33	Tr*(d)*	1..4	7.3.8.10	`cbf_luma==1; cbf_cb==1; cbf_cr==1`	-	√	-
34	TSF	4	7.4.9.11	`transform_skip_flag == 1`	√	-	√
35	Bs0	all	8.7.2.4	`bS == 0`	√	-	√
36	Bs1	all	8.7.2.4	`bS == 1`	√	-	√
37	Bs2	all	8.7.2.4	`bS == 2`	√	-	√
38	Bs	all	8.7.2.4	always	√	√	-
39	SAO_Y_BO	0	7.4.9.3	`SaoTypeIdx[Y] == 1`	√	-	√
40	SAO_Y_EO	0	7.4.9.3	`SaoTypeIdx[Y] == 2`	√	-	√
41	SAO_Y	0	7.4.9.3	`SaoTypeIdx[Y] != 0`	√	√	-
42	SAO_C_BO	0	7.4.9.3	`SaoTypeIdx[Cb] == 1; SaoTypeIdx[Cr] == 1`	√	-	√
43	SAO_C_EO	0	7.4.9.3	`SaoTypeIdx[Cb] == 2; SaoTypeIdx[Cr] == 2`	√	-	√
44	SAO_C	0	7.4.9.3	`SaoTypeIdx[Cb] != 0; SaoTypeIdx[Cr] != 0`	√	√	-
45	SAO_allComps	0	7.4.9.3	`SaoTypeIdx[Y] && SaoTypeIdx[Cb] && SaoTypeIdx[Cr]`	-	-	√

can be explained by the simplicity of the model. Furthermore, the memory variables (reads and writes) seem to have a very high correlation with the complete energy as the estimation errors for the systems (a) and (a_{CPU}) are below 11%.

Considering the high-level models *H2* and *H3*, the estimation errors are significantly higher than the other models [15%–20% in comparison with less than 9% for system (a)]. Furthermore, model *H3* returns better results than the model *H2* (around 5% lower estimation error).

2) Decoding Systems: Table V gives the mean estimation errors for decoding systems (b)–(d). First of all, considering the Pandaboard decoders (b) and (c), we can see that their estimation errors are comparable with the values of (a). In contrast, the time-based model $H2_T$ fails for system (c) (value greater than 14%). Furthermore, we can see that the high-level models H_2 and H_3 are more accurate in modeling the Intel system than the decoding time-based model $H2_T$ (around 20% in comparison to more than 30%), which indicates that processing time is an unsuitable predictor for complex instruction set architectures.

Considering the energy consumption for CPU and RAM of the systems (a) and (d) as shown in Tables IV and V, we can see that the CPU estimates are sound. In contrast, the estimation of the RAM is weak as its error is higher than 10% and 50%, respectively. Apparently, the influence on the combined error of (a) and (d) is rather low, which can be explained by the RAM occupation, which is less than 10% of the complete energy consumption for both systems, as indicated by our measurements. Please note that as stated in Section III, we only consider the additional processing energy, the RAM occupation of the complete energy is much higher when the idle energies of the RAM and the CPU are included.

The estimation errors for the FPGA-based systems (e) and (f) are given in Tables VI and VII, respectively. For system (e), we can see that decoding on the SPARC returns similar estimation errors as system (a). The high-level models perform worse in this case, because only low-resolution sequences were tested. Estimating the hardware decoder (f) returns rather high estimation errors (more than 14%). This observation has two reasons: First, the fraction of the RAM on the complete power consumption is higher than for the software-based systems (∼30% as indicated by measurements using the tool power monitor [40]). Second, the structure of such an implementation is completely different from a software decoder such that a different model that respects the different characteristics may be more suitable. Unexpectedly, the high-level model H_2 performs relatively well (18% estimation error), which indicates that the fraction of intra-coded frames is an important factor in the energy consumption of this hardware implementation.

3) Further Observations: To show that the tested models are independent of the QP, we tested system (a) for a higher number of different QPs (QPs 5 to 50 in steps of 5). The resulting estimation errors turned out to be very close to the values given in Table IV such that our evaluation method using only three QP values is confirmed.

Next, we would like to discuss the results shown in Table VIII. Here, we can see how the estimation errors change when using sequences containing multiple frames. First, we can observe that the estimation errors drop significantly (in average, the error is halved for all the tested models and decoding systems). This can be explained by the law of large numbers where estimation inaccuracies are averaged. Particularly, we can see that sound estimations can even be obtained for the RAM (model *FS* lower than 5%). For the hardware implementation [system (f)], the mean error does not change significantly in comparison with the single frame approach.

Finally, to show that our results are consistent, we would like to discuss the reported values from the literature. Basically, these values are close to the values given in Table VIII that correspond to multiple frame sequences (e.g., the models *FS, T,* H_3, and $H1_T$). Intuitively, this corresponds to our expectation as in related work, only sequences containing multiple frames were tested. For the feature-based models, we could even achieve a lower estimation error (3.63% in [18]), which is caused by the more sophisticated training method. In addition, reported errors for the model *PE* correspond to our findings (more than three quarters of the tested sequences in [14] showed an estimation error between 5%–10% where we found a mean error of ∼ 8%). In contrast, Raoufi and Peters [12] reported estimation errors of less than 1% (models *H2* and $H2_T$), which is much better than our results. However, they used an H.264 decoder and only tested a single sequence such that a lower error can be expected.

V. Conclusion

In this paper, we proposed a bit stream feature-based model and revisited some models capable of estimating the required decoding energy for a given HEVC-coded bit stream. We tested the models meticulously for different software solutions, different processing devices, and one hardware implementation with a large amount of test sequences to obtain meaningful estimation errors. As a result, we have seen that best results can be obtained using the proposed simple feature-based model using 20 parameters and that the software decoder solutions can be best modeled. Accepting estimation errors of up to 20%, the model can also be applied to hardware-based decoders.

In future work, we plan to validate, test, and adapt the model for specific hardware decoders like ASICs. Furthermore, we will utilize the feature-based model inside the RDO process to encode sequences such that the decoding energy can be minimized. Another application would be to provide a real-time task scheduler with the gathered information. As the energy estimation can be calculated before the particular decoding takes place, for instance, in a pipelined decoder structure, sophisticated power saving strategies can be applied to the software and hardware-based decoder systems with the presented results.

Appendix
Bit Stream Feature Set for the Feature-Based Model

The bit stream features used in the models *FA* and *FS* are listed in Table IX. The table indicates how the feature numbers can be derived and to which model they belong.

The feature label helps to elucidate the meaning of the feature. The column "depths" gives information about the depths at which the feature can occur. The columns "subclause," "condition for syntax element or variable," and "Inc" provide information about the derivation of the feature numbers. The last two columns "*FS*" and "*FA*" indicate whether the feature is used for the simple or the accurate model (see Section II-A), respectively.

First, the column "depths" in Table IX specifies whether the feature numbers are counted for the individual depths of the quadtree decomposition. In this context, depth 0 stands for a 64×64 block and goes down to a block size of 4×4 at depth 4. "all" specifies that the feature is counted at every depth.

The following explanation is directly related to the recommendation ITU-T H.265 [22]. The column labeled "subclause" specifies at which point the counter of the feature number can be placed during the parsing process, where the subclause refers to the subclause in the recommendation. Furthermore, the fifth column "condition for syntax element or variable" describes the syntax element or variable that must be checked to determine whether the number shall be increased.

If there is a checkmark in the column "Inc," the feature number can simply be incremented so long as the condition described in the fifth column is met. For instance, the feature number n_{coeff} is incremented every time the `significant_coeff_flag` is set during the parsing process described in subclause 7.4.9.11 in the recommendation [22].

The remaining feature numbers without a checkmark in the column "Inc" are derived as follows.

1) E_0*:* As this feature describes the initialization of the decoding process, it always occurs once for each bit stream. Hence, its number is constantly $n_{E_0} = 1$.
2) *fracpelVer:* This feature describes the filtering process of one fractional luma pel in a vertical direction. Each time the given condition (the %-sign denotes the modulo operator) is met during the parsing of the given subclause; the number $n_{\text{fracpelVer}(d)}$ at the current depth d has to be increased by the number of luma pels to be filtered. In terms of the variables used in the standard, the equation reads

$$n_{\text{fracpelVer}(d)} = n_{\text{fracpelVer}(d)} + \texttt{nPbW} \cdot \texttt{nPbH} \quad (16)$$

where `nPbW` describes the number of luma pels in a horizontal and `nPbH` in a vertical orientation in the current prediction unit (PU).
3) *fracpelHor:* To derive this number, the same equation holds as for vertical filtering

$$n_{\text{fracpelHor}(d)} = n_{\text{fracpelHor}(d)} + \texttt{nPbW} \cdot \texttt{nPbH}. \quad (17)$$

Note that any filtering performed in both the vertical and horizontal directions (denoted by the condition in brackets) must be checked. In this case, further horizontal filtering operations must be considered as for the subsequent vertical filtering, horizontally filtered fractional pels outside of the current PU have to be calculated. We consider this fact by

$$n_{\text{fracpelHor}(d)} = n_{\text{fracpelHor}(d)} + 6 \cdot \texttt{nPbW} \quad (18)$$

as six additional rows of the fractional pels are required, three above and three below the current PU.
4) *fracpelAvg:* This feature number comprises all the horizontal and vertical luma pel filterings

$$n_{\text{fracpelAvg}} = \sum_{\text{d}=0}^{3} n_{\text{fracpelVer}(d)} + n_{\text{fracpelHor}(d)}. \quad (19)$$

5) *chrHalfpel:* As the chroma pels are subsampled, we must consider the case in which the motion vector points to a chroma-half pel position (which is a full pel position for luma samples). Thus, each time the condition is true, we apply the following equation:

$$n_{\text{chrHalfpel}(d)} = n_{\text{chrHalfpel}(d)} + \frac{\texttt{nPbW}}{2} \cdot \frac{\texttt{nPbH}}{2}. \quad (20)$$

Evaluations showed that a further decomposition into horizontal and vertical direction as well as the calculation of outlying pels has a minor impact on the estimation accuracy, so that we neglect these possibilities.
6) *bi:* This feature number counts the number of bipredicted 4×4 blocks. If the condition is met, it is calculated as

$$n_{\text{bi}} = n_{\text{bi}} + \frac{\texttt{nPbW}}{4} \cdot \frac{\texttt{nPbH}}{4}. \quad (21)$$

7) *Motion vector difference (MVD):* If the given condition is met, this feature number is increased by

$$n_{\text{MVD}} = n_{\text{MVD}} + \log_2(\texttt{abs\_mvd\_minus2} + 2). \quad (22)$$

The logarithm of the value was chosen due to the exponential properties of the Exp-Golomb-Code that was used to code the MVD [41]. The real parsing energy is assumed to be nearly linear to this expression.[1]
8) *val:* The residual coefficient value is coded in the same way as the MVD. Hence, the equation reads

$$n_{\text{val}} = n_{\text{val}} + \log_2(\texttt{coeff\_abs\_level\_remaining} + 2). \quad (23)$$

9) *Tr:* This feature comprises all the transformation processes depending on the transform size. Hence, depth 1 relates to a 32×32 transform, depth 2 to a 16×16 transform, and so on. Hence, if a coded block flag (CBF) is set for one of the chroma components, the counter has to be incremented for depth $d = \min(d + 1, 4)$. Thus, the smaller transform block size for the chroma blocks is taken into account.
10) *SAO_allComps:* As a final feature number $n_{\text{SAO\_allComps}}$ is incremented if all the color components are filtered using SAO. Measurements showed that in this case, less energy is required in contrast to the processing of single color components, which is reflected in a negative value of the specific energy.

REFERENCES

[1] A. Carroll and G. Heiser, "An analysis of power consumption in a smartphone," in *Proc. USENIX Conf.*, vol. 14. Oct. 2010, p. 21.
[2] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.
[3] C.-T. Huang, M. Tikekar, C. Juvekar, V. Sze, and A. Chandrakasan, "A 249Mpixel/s HEVC video-decoder chip for quad full HD applications," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers (ISSCC)*, Feb. 2013, pp. 162–163.
[4] J. Zhu, D. Zhou, G. He, and S. Goto, "A combined SAO and de-blocking filter architecture for HEVC video decoder," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2013, pp. 1967–1971.
[5] E. Kalali, Y. Adibelli, and I. Hamzaoglu, "A high performance and low energy intra prediction hardware for HEVC video decoding," in *Proc. DASIP*, Oct. 2012, pp. 1–8.
[6] Y. Jiang, D. Llamocca, M. Pattichis, and G. Esakki, "A unified and pipelined hardware architecture for implementing intra prediction in HEVC," in *Proc. IEEE Southwest Symp. Image Anal. Interpretation (SSIAI)*, Apr. 2014, pp. 29–32.
[7] D. Palomino, F. Sampaio, L. Agostini, S. Bampi, and A. Susin, "A memory aware and multiplierless VLSI architecture for the complete intra prediction of the HEVC emerging standard," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep./Oct. 2012, pp. 201–204.
[8] S. Park and K. Ryoo, "The hardware design of effective SAO for HEVC decoder," in *Proc. IEEE 2nd Global Conf. Consum. Electron. (GCCE)*, Oct. 2013, pp. 303–304.
[9] M. van der Schaar and Y. Andreopoulos, "Rate-distortion-complexity modeling for network and receiver aware adaptation," *IEEE Trans. Multimedia*, vol. 7, no. 3, pp. 471–479, Jun. 2005.
[10] S.-W. Lee and C.-C. J. Kuo, "Complexity modeling for motion compensation in H.264/AVC decoder," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, vol. 5. Sep./Oct. 2007, pp. 313–316.
[11] 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. Vis. Commun. Image Process. (VCIP)*, Nov. 2012, pp. 1–6.
[12] P. Raoufi and J. Peters, "Energy-efficient wireless video streaming with H.264 coding," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Jul. 2013, pp. 1–6.

[1] Equation (22) can be approximated using fixed point arithmetics. In our implementation, we chose to use the position of the highest nonzero bit of `abs_mvd_minus2` + 2 to approximate the logarithm. Furthermore, if the bit next to the highest bit is also set, we add a refining value of $\log_2(1.5) = 0.585$.

[13] *Advanced Video Coding for Generic Audio-Visual Services*, ITU-T Rec. H.264 and ISO/IEC 14496-10 (AVC), ITU-T and ISO/IEC JTC 1, May 2003.
[14] R. Ren, E. Juárez, C. Sanz, M. Raulet, and F. Pescador, "Energy-aware decoders: A case study based on an RVC-CAL specification," in *Proc. Conf. Design Archit. Signal Image Process. (DASIP)*, Oct. 2014, pp. 1–6.
[15] C. Herglotz and A. Kaup, "Estimating the HEVC decoding energy using high-level video features," in *Proc. 23rd Eur. Signal Process. Conf. (EUSIPCO)*, Aug./Sep. 2015, pp. 1591–1595.
[16] C. Herglotz, E. Walencik, and A. Kaup, "Estimating the HEVC decoding energy using the decoder processing time," in *Proc. Int. Symp. Circuits Syst. (ISCAS)*, May 2015, pp. 513–516.
[17] C. Herglotz, D. Springer, A. Eichenseer, and A. Kaup, "Modeling the energy consumption of HEVC intra decoding," in *Proc. 20th Int. Conf. Syst., Signals Image Process. (IWSSIP)*, Jul. 2013, pp. 91–94.
[18] C. Herglotz, D. Springer, and A. Kaup, "Modeling the energy consumption of HEVC P- and B-frame decoding," in *Proc. Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 3661–3665.
[19] A. Norkin *et al.*, "HEVC deblocking filter," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1746–1754, Dec. 2012.
[20] C.-M. Fu *et al.*, "Sample adaptive offset in the HEVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1755–1764, Dec. 2012.
[21] J. Lainema, F. Bossen, W.-J. Han, J. Min, and K. Ugur, "Intra coding of the HEVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1792–1801, Dec. 2012.
[22] *High Efficiency Video Coding*, ITU-T Rec. H.265 and ISO/IEC 23008-2, ITU-T and ISO/IEC JTC 1/SC 29/WG 11 (MPEG), Apr 2013.
[23] *HEVC Test Model HM-13.0*. [Online]. Available: https://hevc.hhi.fraunhofer.de/
[24] *Decoding Energy Estimation Tool (DENESTO)*, accessed Aug 2016. [Online]. Available: http://lms.lnt.de/denesto
[25] J. H. Friedman, "Multivariate adaptive regression splines," *Annu. Statist.*, vol. 19, no. 1, pp. 1–67, Mar. 1991.
[26] *Valgrind Profiling Tool*, accessed on Aug. 2016. [Online]. Available: http://valgrind.org/. Accessed 2016-08.
[27] V. Konstantakos, A. Chatzigeorgiou, S. Nikolaidis, and T. Laopoulos, "Energy consumption estimation in embedded systems," *IEEE Trans. Instrum. Meas.*, vol. 57, no. 4, pp. 797–804, Apr. 2008.
[28] H. David, E. Gorbatov, U. R. Hanebutte, R. Khanna, and C. Le, "RAPL: Memory power estimation and capping," in *Proc. ACM/IEEE Int. Symp. Low-Power Electron. Design (ISLPED)*, Aug. 2010, pp. 189–194.
[29] D. Hackenberg, R. Schöne, T. Ilsche, D. Molka, J. Schuchart, and R. Geyer, "An energy efficiency feature survey of the intel Haswell processor," in *Proc. IEEE Int. Parallel Distrib. Process. Symp. Workshop (IPDPSW)*, May 2015, pp. 896–904.
[30] J. S. Bendat and A. G. Piersol, *Random Data: Analysis and Measurement Procedures*. New York, NY, USA: Wiley, 1971.
[31] *Pandaboard Development Platform*, accessed on Aug. 2016. [Online]. Available: http://pandaboard.org/
[32] Texas Instruments. *OMAP4430 Multimedia Device Data Manual [Public] Version D (Rev. D)*, accessed on Aug. 2016. [Online]. Available: http://www.ti.com/product/omap4430
[33] *Libde265 Open Source HEVC*, accessed on Aug. 2016. [Online]. Available: http://www.libde265.org/
[34] *FFmpeg Multimedia Framwork*, accessed on Aug. 2016. [Online]. Available: http:// ffmpeg.org/
[35] Cobham Gaisler. (Jan. 2016). *GRLIB IP Core User's Manual*, accessed on Aug. 2016. [Online]. Available: http://www.gaisler.com/products/grlib/grip.pdf
[36] Cobham Gaisler. *Grmon2 Debug Monitor*, accessed on Aug. 2016. [Online]. Available: http://www.gaisler.com/index.php/products/debug-tools/grmon2
[37] D. Engelhardt, J. Möller, J. Hahlbeck, and B. Stabernack, "FPGA implementation of a full HD real-time HEVC main profile decoder," *IEEE Trans. Consum. Electron.*, vol. 60, no. 3, pp. 476–484, Aug. 2014.
[38] T. F. Coleman and Y. Li, "An interior trust region approach for nonlinear minimization subject to bounds," *SIAM J. Optim.*, vol. 6, no. 2, pp. 418–445, 1996.
[39] M. J. Zaki and W. Meira, *Data Mining and Analysis*, 1st ed. Cambridge, U.K.: Cambridge Univ. Press, 2014.
[40] Altera Corporation. (Sep. 2015). *Stratix V Edition DSP Development Board-Reference Manual*, accessed on Aug. 2016. [Online]. Available: https://www.altera.com/content/dam/altera-www/global/en_US/pdfs/literature/manual/rm_svgx_fpga_dev_board.pdf
[41] V. Sze and M. Budagavi, "High throughput CABAC entropy coding in HEVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1778–1791, Dec. 2012.

Christian Herglotz received the Dipl.-Ing. degree in electrical engineering and information technology and the Dipl.-Wirt.Ing. degree in business administration and economics from RWTH Aachen University, Aachen, Germany, in 2011 and 2012, respectively.

He has been a Research Scientist with the Chair of Multimedia Communications and Signal Processing, Friedrich-Alexander Universität Erlangen–Nürnberg, Erlangen, Germany, since 2012. His current research interests include energy efficient video coding and fast encoding techniques.

Mr. Herglotz received the Paul Dan Cristea Special Award in 2013 for his paper titled "Modeling the Energy Consumption of HEVC Intra Decoding on the 20th International Conference on Systems, Signals, and Image Processing."

Dominic Springer received the Dipl.-Ing. degree in electrical engineering from Friedrich-Alexander Universität Erlangen–Nürnberg, Erlangen, Germany, in 2007, where he is currently pursuing the Ph.D. degree with the Chair of Multimedia Communications and Signal Processing.

He was involved in Digital Cinema hardware/software solutions with Fraunhofer IIS, Erlangen, until 2010. He is currently with the automotive industry. His current research interests include image processing, motion estimation, video compression, and embedded processing.

Marc Reichenbach received the Diploma degree in computer science from the University of Jena, Jena, Germany, in 2010.

He is currently a Research Assistant with the Chair of Computer Architecture, Friedrich-Alexander Universität Erlangen–Nürnberg, Erlangen, Germany. His current research interests include embedded systems, especially smart sensor architectures for different application fields.

Benno Stabernack received the Diploma and Dr.-Ing. degrees in electrical engineering from the Technical University of Berlin, Berlin, Germany, in 1996 and 2004, respectively.

He joined the Fraunhofer Institute for Telecommunications–Heinrich Hertz Institute (HHI), Berlin, in 1996. As the Head of the Embedded Systems Group with the Image Processing Department, HHI, he is currently responsible for the research projects focused on hardware and software architectures for image and video processing algorithms. He has been giving a lecture on the design of application specific processors with the Technical University of Berlin since 2005. Since 2008, he has been a Lecturer of Multimedia Signal Processing with the University of Potsdam, Potsdam, Germany. Since 2016, he has been the Chair of Embedded Systems Architectures for Signal Processing with the University of Potsdam. He has been engaged in several European research projects. His current research interests include VLSI architectures for video signal processing, application specific processor architectures for embedded media signal processing and system-on-chip designs.

André Kaup (M'96–SM'99–F'13) received the Dipl.-Ing. and Dr.-Ing. degrees in electrical engineering from RWTH Aachen University, Aachen, Germany, in 1989 and 1995, respectively.

He was with the Institute for Communication Engineering, RWTH Aachen University, from 1989 to 1995. He joined the Networks and Multimedia Communications Department, Siemens Corporate Technology, Munich, Germany, in 1995, where he became the Head of the Mobile Applications and Services Group in 1999. He was the Head of the German MPEG delegation from 1997 to 2001. He served as an Adjunct Professor with the Technical University of Munich, Munich, from 1998 to 2001. He has been a Full Professor and the Head of the Chair of Multimedia Communications and Signal Processing with the Friedrich-Alexander Universität Erlangen–Nürnberg, Erlangen, Germany, since 2001. He was a Vice Speaker of the DFG Collaborative Research Center 603 from 2005 to 2007. He has been serving as the Head of the Department of Electrical Engineering and the Vice Dean of the Faculty of Engineering since 2015. He has authored over 300 journal and conference papers. He has over 50 patents granted and patent applications published. His current research interests include image and video signal processing and coding, and multimedia communication.

Prof. Dr. Kaup is a member of the IEEE Multimedia Signal Processing Technical Committee, and a member of the Scientific Advisory Board of the German VDE/ITG. He was a Siemens Inventor of the Year in 1998 and received the 1999 ITG Award. He has received several best paper awards, including the Paul Dan Cristea Special Award from the International Conference on Systems, Signals, and Image Processing in 2013. His group received the Grand Video Compression Challenge at the Picture Coding Symposium in 2013. He serves as an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY and a Guest Editor of the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING.