Efficient Mode Decision Schemes for HEVC Inter Prediction

Jarno Vanne, *Member*, *IEEE*, Marko Viitanen, *Member*, *IEEE*, and Timo D. Hämäläinen, *Member*, *IEEE*

***Abstract*— The emerging HEVC standard reduces the bit rate by almost 40% over the preceding state-of-the-art standard AVC with the same objective quality but at about 40% encoding complexity overhead. The main reason for HEVC complexity is inter prediction that accounts for 60-70% of the whole encoding time. This paper analyzes the rate-distortion-complexity (RDC) characteristics of the HEVC inter prediction as a function of different block partition structures and puts the analysis results into practice by developing optimized mode decision schemes for the HEVC encoder. The HEVC inter prediction involves three different partition modes: square motion partition (Square), symmetric motion partition (SMP), and asymmetric motion partition (AMP) out of which the decision of SMPs and AMPs are optimized in this work. The key optimization techniques behind the proposed schemes are 1) a conditional evaluation of the SMP modes, 2) range limitations primarily in the SMP sizes and secondarily in the AMP sizes, and 3) a selection of the SMP and AMP ranges as a function of the quantization parameter. These three techniques can be seamlessly incorporated in the existing control structures of the HEVC reference encoder (HM) without limiting its potential parallelization, hardware acceleration, or speed-up with other existing encoder optimizations. Our experiments show that the proposed schemes are able to cut the average complexity of HM encoder by 31-51% at a cost of 0.2-1.3% bit rate increase under the random access (RA) coding configuration. The respective values under the low-delay *B* (LB) coding configuration are 32-50% and 0.3-1.3%.**

***Index Terms*— High Efficiency Video Coding (HEVC), inter prediction, mode decision, rate-distortion-complexity (RDC).**

I. INTRODUCTION

THE latest international video coding standard *HEVC* (*High Efficiency Video Coding*) [1], [2] is targeted for efficient transmission and storage of next-generation video. It has been developed by *Joint Collaborative Team on Video Coding* (*JCT-VC*) as a joint activity of ITU-T *Video Coding Experts Group* (*VCEG*) and ISO/IEC *Moving Picture Experts Group* (*MPEG*). The final draft of HEVC standard has been approved in January 2013. It includes three profiles called *Main Profile* (*MP*), *Main 10 Profile*, and *Main Still Picture Profile* [1]. MP is considered in this paper.

Manuscript received April 15, 2012. This work was supported in part by the European Celtic-Plus project H2B2VS.

J. Vanne, M. Viitanen, and T. D. Hämäläinen are with the Department of Pervasive Computing, Tampere University of Technology, FI-33101 Tampere, Finland (e-mail: jarno.vanne@tut.fi).

The compression performance of HEVC is significantly improved from the *AVC* (*Advanced Video Coding*) standard [3]. Compared to AVC *High Profile* (*HiP*), HEVC MP is able to reduce the bit rate by close to 40% with an equivalent objective quality and around 1.4× coding complexity when all essential coding tools are applied [4]. Furthermore, the bit rate savings of HEVC are reported to be close to 50% when subjective visual quality is used as the quality measure [5]. The benefits of HEVC are emphasized with low bit-rate, high-resolution, and low-delay applications [4], [5].

HEVC adopts the conventional *hybrid video coding scheme* (inter/intra prediction, transform coding, and entropy coding) [2] used in the prior video coding standards since H.261. As a new feature, its coding structure has been extended from a traditional *macroblock* (*MB*) concept to an analogous block partitioning scheme [2], [6], [7] that supports block sizes up to 64 × 64 pixels. This new content-adaptive scheme can be efficiently adjusted between large homogeneous and highly textured regions of the picture.

The modified coding structure is the primary factor for the HEVC coding gain [6], but it also introduces the majority of the computational overhead. The complexity is particularly leveraged in the inter prediction due to its numerous prediction modes and block partitions that have to be evaluated during the *rate-distortion* (*RD*) optimization.

The block partitioning of the HEVC inter prediction and associated mode decision schemes have recently been examined in [5]-[8]. However, these works focus more on overall HEVC encoder evaluation instead of investigating the mode decision in detail. In addition, they cover only a small fraction of the possible mode decision schemes for the HEVC inter prediction. A couple of other works have focused on speeding up the mode decision through early termination mechanisms [9]-[15], but the techniques proposed in this paper are shown to outperform them in the majority of the test cases. A part of these existing approaches are orthogonal to ours so merging them with the proposed ones is also possible.

The dominant role of the inter prediction in HEVC [4] motivates us to analyze its *rate-distortion-complexity* (*RDC*) characteristics more thoroughly and develop efficient mode decision schemes for it. As in [4]-[8], the experiments have been conducted with a HEVC reference codec called *HEVC test model* (*HM*). The codec version utilized in our development work is HM 8.0 [16], [17], which was the latest available version in the beginning of our experiments. The efficiency of the proposed mode decision schemes is also

validated in HM 11.0 [18] which was the most up-to-date version at the time our work was finished. The common test conditions [19] for HM include four predefined coding configurations: *all-intra* (*AI*), *random access* (*RA*), *low-delay P* (*LP*), and *low-delay B* (*LB*). This paper considers HM MP under the RA and LB cases. Although HM is targeted for research and conformance testing, the guidelines derived from this well-known, publicly available, and platform-independent codec can be adopted by practical real-time codecs as well.

The remainder of this paper is organized as follows. Section II presents the block partitioning structure of HEVC and the associated mode decision process of HM MP. The recent HEVC encoder analyses and the contemporary optimization approaches for the HEVC block partitioning structure/mode decision are surveyed in Section III. Section IV describes our analysis environment and evaluation policy used to validate our design decisions. Section V presents our optimization techniques through which the proposed mode decision schemes have been developed. Section VI summarizes our optimization strategy, gives configuration guidelines for the HEVC encoder, and compares the RDC characteristics of our proposal with other existing approaches. Section VII concludes the paper.

II. HEVC CODING STRUCTURE AND MODE DECISION

In HEVC, pictures are partitioned into *coding tree units* (*CTUs*) [2], [6]-[8]. Each CTU consists of one square *coding tree block* (*CTB*) of luma samples, two corresponding CTBs of chroma samples, and associated syntax elements. The size of the luma CTB can be defined as $2N_{MAX} \times 2N_{MAX}$, where $N_{MAX} \in \{8, 16, 32\}$. The selection of N_{MAX} is made by the encoder. Allowing variable-size CTBs eases meeting the different computational and memory demands of the encoder.

A. Coding units

The *coding tree* of HEVC utilizes a quadtree structure to partition the CTBs into smaller *coding blocks* (*CBs*). A CTU represents a root node (depth 0) of the quadtree. At depth 1, CTBs of a CTU can be optionally divided into four equal-sized square CBs. The division can be recursively continued until the maximum hierarchical depth (h_{max}) of the quadtree is reached. Through this adjustable quadtree depth, HEVC enables content-adaptive granularity for CBs.

The leaf nodes of the quadtree are called *coding units* (*CUs*). For each CU, the size of its luma CB can be defined as $2N \times 2N$, where $N \leq N_{MAX}$. In HEVC, $N \in \{4, 8, 16, 32\}$, so $N_{MIN} = 4$ and $h_{MAX} = 4$. The CUs of the coding tree are processed in a depth-first manner.

A CU defines a picture region that shares the same prediction mode (*Intra*, *Inter*, *Skip* or *Merge*) [2]. Hence, it acts as a root for a *prediction tree* and a *transform tree* as well.

B. Prediction units

A *prediction unit* (*PU*) represents a picture region that shares the identical prediction information. That is, the same prediction process is applied for its luma and chroma *prediction blocks* (*PBs*). The prediction tree of HEVC contains only a single level at which luma/chroma CBs can be further divided into one, two, or four rectangular-shaped PBs. For this purpose, HEVC adopts the well-known *Square motion partition* (*Square*) and *symmetric motion partition* (*SMP*) modes from AVC and extends them to conform to larger CB sizes. In addition, HEVC introduces *asymmetric motion partition* (*AMP*) modes to present irregular object boundaries with fewer bits than possible with the Square and SMP modes.

Fig. 1 depicts the partition modes and associated luma PB shapes specified for the inter-coded CBs of size $2N \times 2N$. The supported mode set includes two Square modes (*PART_2N×2N* and *PART_N×N*), two SMP modes (*PART_2N×N* and *PART_N×2N*), and four AMP modes (*PART_2N×nU*, *PART_2N×nD*, *PART_nL×2N*, and *PART_nR×2N*). With the AMP modes, *n* represents smaller one of the partitions, whereas *U*, *D*, *L*, and *R* denote up, down, left, and right, respectively. Using CU as a prediction root ensures that the size of a PB can never exceed that of a CB.

HEVC sets certain limitations for PBs. Firstly, the inter-coded CBs can be coded in PART_N×N mode at h_{MAX} only, since equivalent block splitting (division to four quadrants) is also conducted at a CU level if $h < h_{MAX}$ holds for the quadtree depth h. Secondly, HEVC disables the inter-coded PBs of size 4×4 to minimize the worst-case memory bandwidth, so PART_N×N mode is entirely omitted in the HEVC inter prediction with $N_{MIN} = 4$. Thirdly, the AMP modes are disabled with $N = 4$ to avoid block dimensions smaller than four. Finally, only the Square modes are available for the intra prediction whereas PART_2N×2N mode is always used for the Skip mode.

The HEVC inter prediction is further improved by integrating a *leaf merging concept* [20] into it. Through the Merge mode, PBs of adjacent PUs sharing the identical motion information can be merged into one larger region for which motion information is signaled only once. This reduces coding cost, since only the prediction error needs to be coded individually for each PB. The Skip mode can be treated as a special case of the Merge mode except that the prediction error is not coded for it.

C. Transform units

A *transform unit* (*TU*) specifies a picture region (residual block) that shares the same transformation. Its luma and chroma *transform blocks* (*TBs*) are obtained through a transform tree which is a nested quadtree rooted to CUs (leaf nodes) of a coding tree. TBs may only take a square size of 4×4, 8×8, 16×16, or 32×32 and the decision among them depends on the prediction mode. This paper concentrates on the inter prediction so further TU analysis is out of its scope.

D. Mode decision

Fig. 2 depicts a mode decision process of the HM MP encoder. This process is recursively conducted for each inter-coded CU at every quadtree depth h after which a final assignment of coding modes is accomplished at CTU level. The presented flowcharts have been simplified from [17] by including only the coding options that are enabled by default.

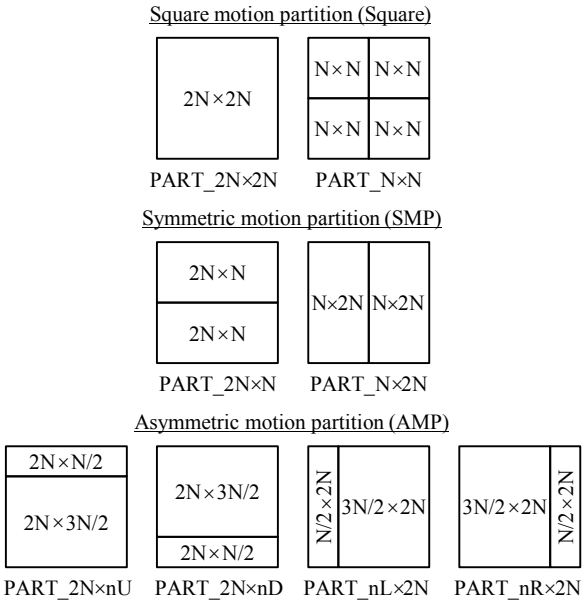

Fig. 1. Partition modes in the HEVC inter prediction.

Let us consider an inter-coded CU at depth h. At first, the HM MP encoder evaluates the Skip, Merge, Square, and SMP modes for CU of size $2N \times 2N$ and selects the *best interim mode* (M_h') among them. If $N \in \{8, 16\}$, M_h' is used to assign a mode set for the subsequent AMP mode decision (see Fig. 2(a)). The mode set contains none of the AMP modes if $M_h' \in$ {Skip, Merge}, half of the AMP modes if $M_h' \in$ {PART_2N×N, PART_N×2N}, and all AMP modes if M_h' = PART_2N×2N. After the AMP modes, the HM MP encoder evaluates the Intra mode and resolves the *best mode* (M_h).

The HM MP encoder takes advantage of a low-complexity *Merge mode only testing* in AMP mode evaluation if M_h' and the *coding mode of the parent CU* (M_{h-1}) meet the conditions [17] listed in Table I. This Merge mode only testing procedure is excluded from the flowcharts for clarity. With $N = 32$, the HM MP encoder restricts the assessment of the AMP modes to the Merge mode only testing in any case, so its flowchart for $N = 32$ converges to that for $N = 4$ (see Fig. 2(b)).

HM MP also offers three optional early termination mechanisms to speed-up its mode decision: *early CU* (*ECU*) [9], *early skip detection* (*ESD*) [10], and *CBF fast mode* (*CFM*) [11], where CBF denotes a *coded block flag*. ECU monitors M_h at the end of each quadtree depth h and terminates the mode decision process for the further depths ($h + 1 \ldots h_{max}$) if M_h = Skip. When ESD is enabled, the evaluation of PART_2N×2N is conducted before the Skip/Merge modes in order to discover its motion information and CBFs at the earliest possible stage of each quadtree depth h. If PART_2N×2N at depth h contains only luma/chroma PBs whose CBF = 0 and no supplementary motion information, ESD sets M_h = Skip and terminates the remaining mode evaluations at depths $h \ldots h_{max}$. CFM monitors CBFs of the Square, SMP, and AMP modes at each quadtree depth h. If an evaluated mode at depth h contains only PBs having CBF = 0, it is selected as M_h without evaluating the remaining modes at that depth and the mode decision proceeds to depth $h + 1$.

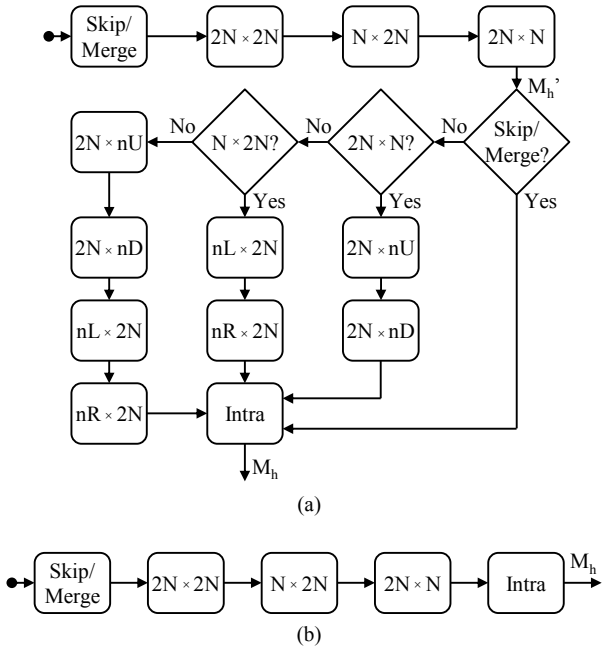

Fig. 2 Mode decision in HM MP (HM 8.0). (a) $N \in \{8, 16\}$. (b) $N \in \{4, 32\}$.

TABLE I
CONDITIONS FOR MERGE MODE ONLY TESTING IN HM MP

Conditions		AMP Merge modes			
M_{h-1}	M_h'	2N×nU	2N×nD	nL×2N	nR×2N
AMP	-	x	x	x	x
Merge	-	x	x	x	x
Intra	2N × N	x	x		
Intra	N × 2N			x	x

TABLE II
COMBINED IMPACT OF ASR, ECU, ESD, AND CFM ON THE RDC VALUES

Reference	Codec	RA		LB	
		BD-rate	Enc. Time	BD-rate	Enc. Time
Bossen et al. [8]	HM 8.0	2.3%	-56%	1.3%	-45%

III. RECENT HEVC ANALYSES

Our previous work [4] presents a comprehensive RDC analysis of HEVC (HM 6.0) and compares the results with AVC reference codec (JM 18.0). Our results show that HEVC reduces the bit rate by around 37% and increases the coding complexity by around 40% over AVC when all essential coding tools of HM 6.0 are utilized.

Bossen *et al.* [8] analyze the optional fast built-in methods of HM 8.0. These methods include *adaptive search range* (*ASR*) [17] for the inter prediction as well as ECU, ESD, and CFM for the mode decision. These four methods are benchmarked in [8] and their combined impact on the RDC characteristics are summarized in Table II.

Rhee *et al.* [12] also examine ECU, ESD, and CFM as well as other fast mode decision algorithms adopted from AVC. The feasibility of these methods is benchmarked in HM 5.0. They preferred ECU, ESD, CFM, and other hierarchical mode decision schemes rather than mode pre-decision based on motion or spatial homogeneity.

Correa *et al.* [21] evaluate different configurations of the HEVC encoder in order to investigate the impact of individual

TABLE III
EXISTING RDC EVALUATIONS AS A FUNCTION OF N_{MAX} AND N_{MIN}

Reference	Codec	RA				LB			
		N_{MAX}	N_{MIN}	BD-rate	Enc. Time	N_{MAX}	N_{MIN}	BD-rate	Enc. Time
Ohm et al. [5]	HM 8.0	16	4	2.2%	-18%	16	4	3.7%	-17%
		8	4	11.0%	-42%	8	4	17.4%	-42%
Kim et al. [6]	HM 6.0	16	4	2.2%	-21%	16	4	2.9%	-22%
		8	4	12.0%	-44%	8	4	14.4%	-43%
		32	8	8.0%	-32%	32	8	7.7%	-31%
		32	16	28.2%	-59%	32	16	29.9%	-59%

TABLE IV
EXISTING RDC EVALUATIONS WITH DISABLED AMP AND SMP

Reference	Codec	RA				LB			
		SMP	AMP	BD-rate	Enc. Time	SMP	AMP	BD-rate	Enc. Time
Ohm et al. [5]	HM 8.0	On	Off	0.9%	-13%	On	Off	1.2%	-12%
Yuan et al. [7]	HM 7.0	On	Off	0.8%	-12%	On	Off	1.1%	-12%
Kim et al. [6]	HM 6.0	On	Off	0.7%	-	On	Off	1.1%	-
		Off	Off	3.6%	-	Off	Off	4.5%	-

TABLE V
AVERAGE AREA SHARES OF THE PREDICTION MODES WITH HM MP

Cfg.	QP	Skip	Intra	Square	SMP	AMP
RA	22	45%	9%	22%	12%	11%
	27	57%	8%	16%	10%	8%
	32	66%	7%	12%	9%	7%
	37	73%	6%	9%	7%	5%
LB	22	40%	4%	30%	14%	13%
	27	52%	3%	23%	12%	10%
	32	62%	3%	18%	10%	8%
	37	70%	2%	14%	8%	5%

coding tools on the RDC characteristics. They define the bit rate changes of ≥ 1.5% and the coding complexity increase of ≥ 5% as significant selection criteria for the tools. The tested configurations cover many tool sets, but they are all based on the default coding structure, where $N_{MAX} = 32$ and $N_{MIN} = 4$.

Ohm *et al.* [5] and Kim *et al.* [6] analyze the RDC characteristics of the HEVC encoder as a function of N_{MAX} and N_{MIN}. Table III summarizes their results, which are here reported over the default configuration of HM. In HM 8.0, setting $N_{MAX} = 16$ has the BD-rate penalties of 2.2% and 3.7% in the RA and LB cases, respectively. Since the complexity reduction of these cases remains under 20%, enabling the fast built-in methods (see Table II) is the preferred approach for complexity reduction. According to the evaluations with HM 6.0, increasing N_{MIN} deteriorates the bit rate even more drastically, so $N_{MAX} = 32$ and $N_{MIN} = 4$ can be concluded to be the recommended settings.

Ohm *et al.* [5] and Kim *et al.* [6] also analyze the effects of disabling the AMP modes in the inter prediction. The analogous analysis has been done by Yuan *et al.* [7] with HM 7.0. The bit rate overheads obtained by these works have been gathered in Table IV. Disabling the AMP modes increases the bit rate by 0.8-1.2% but it does not shorten the encoding time by more than 12-13%, so the AMP modes have been included in HM by default since HM 7.0. The combined effect of the disabled SMP/AMP modes is analyzed with HM 6.0 in [6], according to which the average bit rate overhead is around 4%. However, the gain in encoding time is not reported.

This paper continues the work of [5]-[7] by examining various mode decision schemes of the HEVC inter prediction. Table V tabulates the yielded average area shares of the prediction modes when benchmarking HM MP with 8-bit test sequences adopted from the common test conditions (classes *A*-*E*) [19]. Despite that the Skip and Intra modes contribute only a fraction (≤ 3%) of the entire encoding complexity [4], they account for 55-79% and 43-73% of the PBs under the RA and LB configurations (*cfg.*), respectively. The majority of these PBs are coded in the Skip mode, whose share grows rapidly as a function of the quantization parameter (QP).

The rest of the PBs has been coded in the inter prediction mode (Square, SMP, and AMP modes) whose remaining share is 21-45% of the PBs in the RA case and 27-57% of the PBs in the LB case. These shares also cover the Merge modes. However, coding these PBs accounts for 60-70% of the encoding time [4], so the inter prediction has a crucial impact on the whole encoder complexity.

This work analyzes the decision of the SMP and AMP modes whose combined share is around half of all inter-coded PBs (12-23% in the RA case and 13-27% in the LB case). The outcome of our analysis is then utilized to derive new potential mode decision schemes.

IV. ANALYSIS SETUP

In this work, the optimized mode decision schemes have been developed for HM MP (HM 8.0) under the RA and LB configurations. The optional early termination mechanisms (ECU, ESD, and CFM) have been disabled in all analyzed schemes. The 8-bit test sequence classes (*A*-*E*) listed in Table VI and the QPs of 22, 27, 32, and 37 have been adopted from the common test conditions [19]. In each class, the sequences marked with superscripts *H* and *L* have been identified as the most and least complex ones in our previous analysis [4].

A. Setup for rate-distortion analysis

The RD comparisons are based on the sequence-specific bit rate differences for an identical $PSNR_{AVG}$ value that is a weighted average of luma ($PSNR_Y$) and chroma ($PSNR_U$ and $PSNR_V$) PSNR components [5]. All involved 8-bit test sequences (see Table VI) are in 4:2:0 color format, for which $PSNR_{AVG}$ values per picture are computed as

TABLE VI
TEST SEQUENCES

Class	Format	Sequence	# of frames	Frame rate	RA	LB
A	2560×1600 (1600p)	Traffic^{L}	150	30 fps	x	
		PeopleOnStreet^{H}	150	30 fps	x	
B	1920×1080 (1080p)	Kimono	240	24 fps	x	x
		ParkScene^{L}	240	24 fps	x	x
		Cactus	500	50 fps	x	x
		BQTerrace	600	60 fps	x	x
		BasketballDrive^{H}	500	50 fps	x	x
C	832×480 (WVGA)	RaceHorses^{H}	300	30 fps	x	x
		BQMall^{L}	600	60 fps	x	x
		PartyScene	500	50 fps	x	x
		BasketballDrill	500	50 fps	x	x
D	416×240 (WQVGA)	RaceHorses^{H}	300	30 fps	x	x
		BQSquare^{L}	600	60 fps	x	x
		BlowingBubbles	500	50 fps	x	x
		BasketballPass	500	50 fps	x	x
E	1280×720 (720p)	FourPeople	600	60 fps		x
		Johnny^{L}	600	60 fps		x
		KristenAndSara^{H}	600	60 fps		x

$$\mathrm{PSNR_{AVG}} = (6 \times \mathrm{PSNR}_Y + \mathrm{PSNR}_U + \mathrm{PSNR}_V)/8. \quad (1)$$

The overall $\mathrm{PSNR_{AVG}}$ value for the whole sequence is then obtained by averaging its picture-specific $\mathrm{PSNR_{AVG}}$ values.

The sequence-specific bit rate differences have been compared in terms of the *Bjøntegaard delta bit rate* (*BD-rate*) [22]. For each scheme, the RD curves for the BD-rate computations have been interpolated through experimentally specified RD points that represent the QPs of 22, 27, 32, and 37. These four QPs have also been selected when reporting the QP-specific delta bit rates for the optimized schemes. In practice, these QP-specific delta bit rates have been obtained by matching the QP-specific $\mathrm{PSNR_{AVG}}$ values of the optimized schemes to the RD curve of the default scheme. Our analysis relies on the averages of the sequence-specific results. The sequences involved in the RA and LB cases are identified with *x* in Table VI.

B. Setup for complexity analysis

Our profiling environment [4] is composed of two identical processor platforms whose characteristics are listed in Table VII. Only a single core per processor has been used. SIMD extensions (MMX/SSE) have not been exploited in order to maintain platform independency.

The analysis relies on Intel VTune Amplifier XE profiler, which is able to report estimated cycle counts for each encoder function. This cycle-level profiling also considers internal complexities of the functions so it is more reliable than monitoring the function calls only. To save profiling time, all complexity results are averages of the corner case sequences marked with *H* and *L* in Table VI.

V. PROPOSED MODE DECISION SCHEMES

The existing works [5]-[7] rely on a hypothesis that a feasible block partition scheme in the HEVC inter prediction includes either 1) the unconditional Square modes, 2) the unconditional Square/SMP modes, or 3) the unconditional Square/SMP modes and the conditional AMP modes. However, these three mode decision schemes cover only a fraction of all available schemes in the HEVC inter prediction.

TABLE VII
PROFILING PLATFORM

Processor	Intel Core 2 Duo E8400 (2 × 3.0 GHz)
Memory	8 GB
L1 cache	2 × 32 KB (instruction) + 2 × 32 KB (data)
L2 cache	6 MB
Compiler	Microsoft Visual C++ 2010
Operating system	64-bit Microsoft Windows 7 Enterprise SP 1

This work develops mode decision schemes that yield the best possible trade-off between RD performance and encoding complexity. We will particularly focus on the schemes whose BD-rate increment over HM MP is close to 1% at maximum. Hence, the RDC characteristics of the potential schemes are examined under the following conditions:

1) N_{MAX} = 32 and N_{MIN} = 4 to avoid a bit rate penalty of at least 2% with a limited range of N [5].
2) The Intra and Skip modes are evaluated according to HM MP specification since they contribute only 3% (or less) of the whole HM MP encoder complexity [4].
3) PART_2N×2N is evaluated according to HM MP specification due to its fundamental role in the picture partitioning and HM encoder control.
4) PART_N×N is entirely disabled in the inter prediction as specified by HEVC MP for the case of N_{MIN} = 4.
5) The AMP modes are disabled for N = 4 as specified by HEVC MP.

These five conditions justify us to focus on investigating the most feasible composition of the SMP and AMP modes when $N_{\mathrm{SMP}} \in \{4, 8, 16, 32\}$ and $N_{\mathrm{AMP}} \in \{8, 16, 32\}$. The other encoder settings are equal to those of the default HM 8.0 configuration, which is used as an anchor for our optimizations. Here, the optimized mode decision schemes are identified as S_x, where x represents a consecutive numbering of these schemes.

A. Mode decision without SMP and AMP modes

The HM encoder is first configured by disabling the SMP and AMP modes entirely (N_{SMP} = Ø and N_{AMP} = Ø) to see their impact on the RDC characteristics. Fig. 3 depicts this highly pruned mode decision process (S_0) that is uniform for each N.

Table VIII reports the RDC figures of S_0 over the default scheme (see Fig. 2) in the RA and LB cases. The RD results include the bit rate differences per four individual QPs (*Δ bit rate*) and the BD-rates. The QP-specific complexity differences are given as *delta million cycles per frame* (*Δ Mcpf*), which reports the combined cycle counts of the mode decision and prediction operations. The rationale for including the prediction operations in *Δ Mcpf* value is that the complexity shares of the prediction modes (Intra, Inter, Skip, and Merge) are highly dependent on the mode decision process itself. The average complexity differences between the compared schemes is characterized by the remaining two

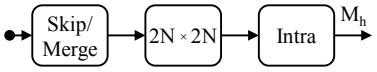

Fig. 3. Mode decision without the SMP/AMP modes (S_0), $N \in \{4, 8, 16, 32\}$.

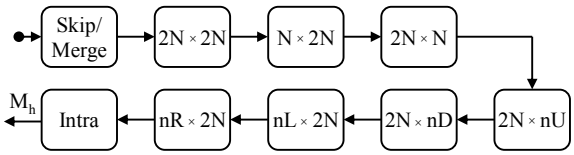

Fig. 4. Mode decision with the unconditional AMP modes (S_1), $N \in \{8, 16, 32\}$.

TABLE VIII
THE IMPACT OF DISABLING THE SMP AND AMP MODES ON THE RDC CHARACTERISTICS OF HEVC

Cfg.	ID	N_{SMP}	N_{AMP}	QP=22		QP=27		QP=32		QP=37		BD-rate	Average	Average
				Δ bit rate	Δ Mcpf	Δ bit rate	Δ Mcpf	Δ bit rate	Δ Mcpf	Δ bit rate	Δ Mcpf		Δ Mcpf	Δ $Mcpf_{enc}$
RA	S_0	Ø	Ø	3.3%	-68%	4.0%	-68%	3.9%	-68%	3.3%	-67%	3.8%	-68%	-59%
LB	S_0	Ø	Ø	3.8%	-68%	4.7%	-68%	4.7%	-68%	4.1%	-67%	4.5%	-68%	-61%

TABLE IX
THE IMPACT OF THE UNCONDITIONAL AMP MODES ON THE RDC CHARACTERISTICS OF HEVC

Cfg.	ID	N_{SMP}	N_{AMP}	QP=22		QP=27		QP=32		QP=37		BD-rate	Average	Average
				Δ bit rate	Δ Mcpf	Δ bit rate	Δ Mcpf	Δ bit rate	Δ Mcpf	Δ bit rate	Δ Mcpf		Δ Mcpf	Δ $Mcpf_{enc}$
RA	S_1	{8,16,32, 64}	{8,16,32}*	-0.3%	54%	-0.4%	61%	-0.6%	69%	-0.7%	76%	-0.5%	65%	51%
LB	S_1	{8,16,32, 64}	{8,16,32}*	-0.5%	57%	-0.6%	65%	-0.8%	72%	-1.2%	80%	-0.8%	69%	56%

* = Unconditional AMP mode evaluation

values. The *Average Δ Mcpf* value reports an arithmetic mean of the QP-specific Δ Mcpf values, whereas the *Average Δ* $Mcpf_{enc}$ value shows the respective mean at the encoder level by counting all encoding operations.

S_0 reduces the average prediction complexity by almost 70% but at a cost of 3.8% and 4.5% BD-rate increase in the RA and LB cases, respectively. At the encoder level, the achieved complexity reduction of around 60% would make S_0 an attractive solution, but its bit rate penalty limits its usage to the cases where the complexity reduction is much more important than the RD performance.

B. Mode decision with unconditional AMP modes

S_0 gives the minimum attainable complexity among the schemes evaluated in this work. Fig. 4 visualizes the other corner case (S_1), where the AMP modes are unconditionally evaluated for $N \in \{8, 16, 32\}$. In this case, the Merge mode only testing can be disabled, since the respective functionality already belongs to the AMP mode evaluation. For $N = 4$, the AMP modes are disabled according to HEVC standard, so the mode decision with $N = 4$ equals that of Fig. 2(b).

The RDC characteristics tabulated in Table IX illustrate that S_1 is able to improve the BD-rate by 0.5% in the RA case and 0.8% in the LB case. However, the respective complexity overheads of the encoding are 51% and 56%. Here, the cost of the obtained RD gain is considered too high for practical implementations, so the default AMP mode scheme is selected as a starting point for our optimizations.

C. Mode decision with limited SMP and AMP depths

S_0 and S_1 demonstrate that the coding options selected for the SMP and AMP mode evaluation can change the encoding complexity from -61% up to 56% over the default configuration. Correspondingly, the BD-rate varies from +3.8% to -0.5% in the RA case and from +4.5% to -0.8% in the LB case. In order to find the best operating point between these two corner cases as a function of SMP and AMP depths, different encoder configurations have been tested by limiting the ranges of N_{SMP} and N_{AMP}.

TABLE X
COMBINED SHARES OF THE SMP AND AMP MODES WITH HM MP

Cfg.	QP	N=32	N=16	N=8	N=4
RA	22	3%	8%	9%	3%
	27	4%	7%	6%	2%
	32	4%	6%	4%	1%
	37	4%	5%	2%	0%
LB	22	3%	9%	11%	3%
	27	4%	9%	7%	2%
	32	5%	8%	4%	1%
	37	5%	6%	2%	0%

Table X shows the combined shares of the SMP and AMP modes distributed over the range of N. The distribution is centered on $N \in \{8, 16\}$, so the cases when $N \in \{4, 32\}$ are the most potential ones to be excluded. In addition, the center of the distribution moves towards the larger PBs as a function of QP.

Table XI tabulates the RDC results for different ranges of N_{SMP} and N_{AMP}. Despite that the share of the small SMP/AMP sizes remains at 0-3%, their importance is manifested when the lower bounds of N_{SMP} and N_{AMP} are raised (S_2-S_4). Disabling $N_{SMP} = 4$ in S_2 decreases the encoding complexity by around 15% but at the BD-rate cost of 1.1% in the RA case and 1.3% in the LB case. Limiting the lower bound further (S_3 and S_4) increase the BD-rate by almost 3% at minimum.

Lowering the upper bounds of N_{SMP} and N_{AMP} (S_5-S_7) increases the BD-rate more moderately. Disabling $N_{SMP} = N_{AMP} = 32$ in S_5 raises the BD-rate by 0.1% in the RA case and 0.3% in the LB case. Although the respective complexity decrement is not more than 10-14%, the overall RDC characteristics of S_5 outperform the ones obtained with S_2. S_5 is particularly effective at lower QPs. Disabling also $N_{SMP} = N_{AMP} = 16$ in S_6 ends up with the BD-rate loss close to that of S_2 that disables NSMP = 4 only. Hence, larger PBs are of lower importance.

TABLE XI
THE IMPACT OF SMP AND AMP SIZES ON THE RDC CHARACTERISTICS OF HEVC

Cfg.	ID	N_{SMP}	N_{AMP}	QP=22		QP=27		QP=32		QP=37		BD-rate	Average	Average
				Δ bit rate	Δ Mcpf	Δ bit rate	Δ Mcpf	Δ bit rate	Δ Mcpf	Δ bit rate	Δ Mcpf		Δ Mcpf	Δ $Mcpf_{enc}$
RA	S_2	{8,16,32}	{8,16,32}	1.3%	-15%	1.4%	-15%	1.0%	-15%	0.4%	-15%	1.1%	-15%	-15%
	S_3	{16,32}	{16,32}	2.9%	-38%	3.3%	-37%	2.8%	-36%	1.9%	-36%	2.9%	-37%	-34%
	S_4	{32}	{32}	3.2%	-57%	3.9%	-57%	3.8%	-56%	3.1%	-54%	3.7%	-56%	-49%
	S_5	{4,8,16}	{8,16}	0.0%	-11%	0.1%	-11%	0.2%	-12%	0.3%	-13%	0.1%	-12%	-10%
	S_6	{4,8}	{8}	0.4%	-26%	0.7%	-27%	1.2%	-27%	1.5%	-28%	0.9%	-27%	-21%
	S_7	{4}	Ø	1.7%	-52%	2.2%	-53%	2.7%	-57%	2.7%	-51%	2.4%	-53%	-45%
LB	S_2	{8,16,32}	{8,16,32}	1.4%	-14%	1.6%	-14%	1.2%	-14%	0.5%	-19%	1.3%	-15%	-16%
	S_3	{16,32}	{16,32}	3.1%	-38%	3.7%	-37%	3.4%	-36%	2.2%	-35%	3.3%	-36%	-34%
	S_4	{32}	{32}	3.7%	-58%	4.7%	-57%	4.6%	-55%	3.7%	-54%	4.4%	-56%	-51%
	S_5	{4,8,16}	{8,16}	0.1%	-11%	0.2%	-17%	0.4%	-20%	0.5%	-13%	0.3%	-15%	-14%
	S_6	{4,8}	{8}	0.7%	-27%	1.2%	-27%	1.7%	-28%	2.2%	-28%	1.4%	-28%	-23%
	S_7	{4}	Ø	2.1%	-54%	2.8%	-54%	3.3%	-54%	3.5%	-53%	3.0%	-53%	-46%

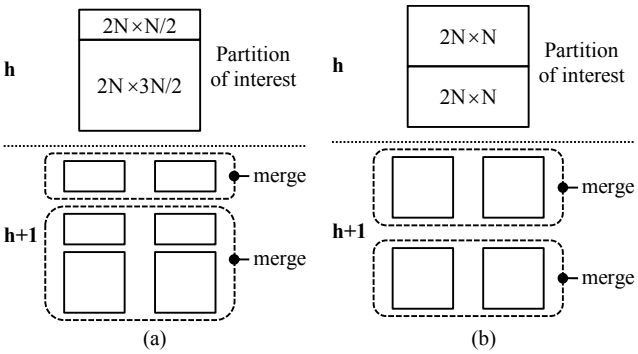

Fig. 5. Composing the partition of interest from the smaller blocks with the Merge mode. (a) PART_2N×nU, AMP off. (b) PART_2N×N, SMP off.

D. SMP/AMP mode only decision

The existing approaches [5]-[7] include the Square and SMP modes as an integral part of the HEVC inter prediction, whereas the AMP modes are conditionally executed, optional, or completely disabled. A mode decision scheme without the AMP modes can be adopted from Fig. 2(b) for each N. This scheme was used by a default configuration of HM 6.0 and its predecessors. Since HM 7.0, the AMP modes have been incorporated in the default configuration of HM MP because they have been shown to improve the BD-rate by around 1% at a complexity overhead less than 15% (see Table IV).

The default mode decision of HM MP prioritizes the SMP modes by evaluating them unconditionally and limits the AMP mode evaluation with preconditions (see Fig. 2). Despite this unequal prioritization, the share of the AMP modes is still only 2% smaller than that of the SMP modes (see Table V). In addition, the other partition modes can be more easily used to replace the missing SMP modes than the missing AMP modes.

Fig. 5 (a) illustrates an example how an AMP of interest can be compensated by merging the smaller square and symmetric PBs when the AMP modes are disabled. In this case, AMP at level h can be composed of two Squares and four SMPs at level h + 1. Fig. 5 (b) shows the respective circumstances for SMP of interest when the SMP modes are disabled. Now, the compensation is possible by merging only four smaller Squares. This latter case can also utilize higher block granularity which increases the probability that the prediction result still converges with the partition of interest. Therefore, it is worth considering the coding schemes where the SMP modes are unconditional/disabled but the AMP modes are on.

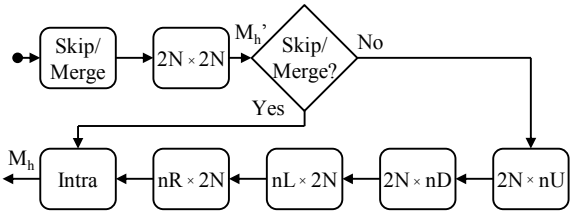

Fig. 6. Mode decision with the disabled SMP modes (S_8), $N \in \{8, 16\}$.

Fig. 6 depicts a mode decision scheme without the SMP modes (S_8) for $N \in \{8, 16\}$. The respective functionality for $N \in \{4, 32\}$ reduces to that of Fig. 3. The RDC results for S_8 and for the scheme without the AMP modes (S_9) are tabulated in Table XII. The reported impact of disabling the AMP modes is in line with [5] (see Table IV), but here more comprehensive RDC results are presented for it.

Disabling the SMP modes increments the BD-rate by at least 2%, so it doubles the BD-rate penalty over the case where the AMP modes are off. However, turning the SMP modes off lowers the encoding complexity in the RA case by 44%. This is almost 3.5 times the complexity reduction obtained by disabling the AMP modes despite that the number of the AMP modes is twice that of the SMP modes (see Fig. 1). In the LB case, the respective complexity drop is 47% due to which the gain factor grows close to 4. This complexity inconsistency is mainly due to the conditional AMP mode evaluation. Hence, HM MP provides more room for SMP optimization making conditional SMP mode evaluation an attractive option to investigate.

E. Proposed SMP mode decision

Fig. 7 depicts our first candidate for an optimized mode decision scheme (S_{10}), in which the precondition used for the SMP modes is adopted from the AMP modes. In S_{10}, the *best interim mode before the SMP modes* (M_h'') is selected among the Skip, Merge, and Square modes. If $M_h'' \in$ {Skip, Merge}, all SMP/AMP modes are skipped and the process proceeds

TABLE XII
THE IMPACT OF THE SMP AND AMP MODES ON THE RDC CHARACTERISTICS OF HEVC

Cfg.	ID	N_{SMP}	N_{AMP}	QP=22		QP=27		QP=32		QP=37		BD-rate	Average	Average
				Δ bit rate	Δ Mcpf	Δ bit rate	Δ Mcpf	Δ bit rate	Δ Mcpf	Δ bit rate	Δ Mcpf		Δ Mcpf	Δ $Mcpf_{enc}$
RA	S_8	Ø	{8,16,32}	1.9%	-47%	2.2%	-51%	2.0%	-54%	1.6%	-57%	2.0%	-52%	-44%
	S_9	{4,8,16,32}	Ø	0.8%	-17%	0.9%	-14%	1.0%	-12%	1.1%	-9%	0.9%	-13%	-13%
LB	S_8	Ø	{8,16,32}	2.1%	-49%	2.6%	-52%	2.4%	-55%	1.9%	-58%	2.4%	-53%	-47%
	S_9	{4,8,16,32}	Ø	0.9%	-17%	1.1%	-14%	1.4%	-11%	1.3%	-8%	1.2%	-13%	-12%

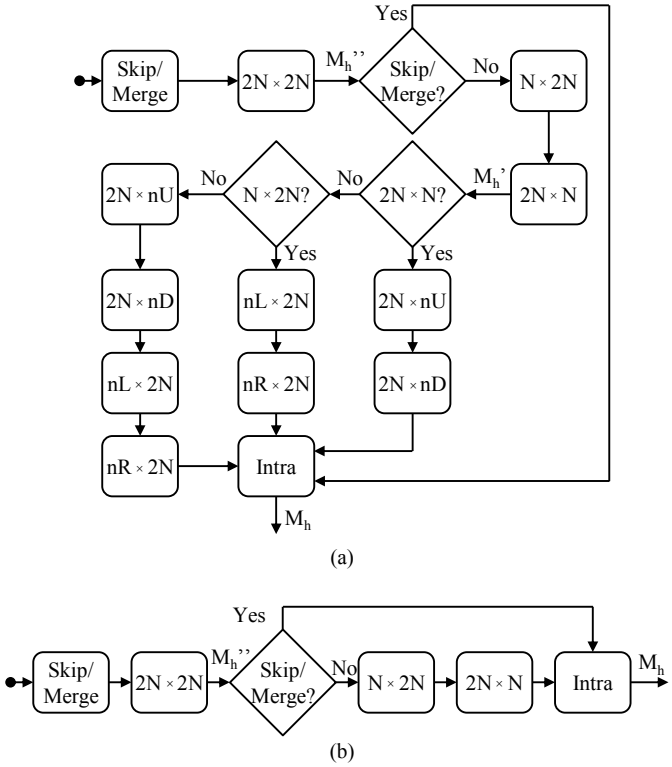

Fig. 7. Optimized mode decision with joint (*j*) precondition for the SMP/AMP modes (S_{10}). (a) $N \in \{8, 16\}$. (b) $N \in \{4, 32\}$.

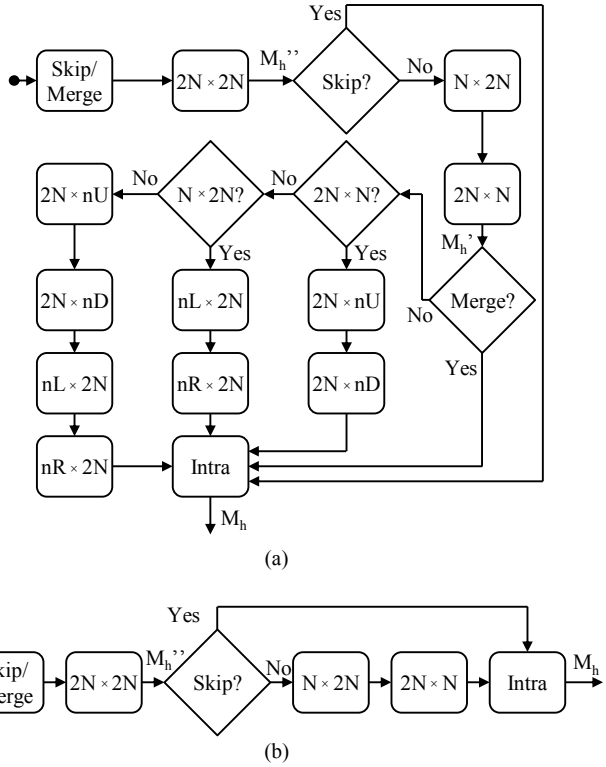

Fig. 8. Optimized mode decision with distributed (*d*) preconditions for the SMP/AMP modes (S_{14}). (a) $N \in \{8, 16\}$. (b) $N \in \{4, 32\}$.

directly to Intra mode. Otherwise, the operation continues as in the default scheme (see Fig. 2) but without any preconditions before the AMP modes. According to our experiments, adopting this *joint* (*j*) precondition for the SMP and AMP modes reduces the average encoding complexity by as much as 37% in the RA case and 40% in the LB case. However, the respective BD-rate increments over the default scheme are 1.2% and 1.7%, so the precondition *j* is considered to be too severe criterion for the SMP modes.

Fig. 8 illustrates our second candidate scheme (S_{14}), in which the encountered BD-rate loss is restrained through the loosened precondition for the SMP modes. S_{14} tests only whether M_h'' = Skip before the SMP modes, whereas the Merge mode comparison is left as a precondition for the AMP modes. Using these *distributed* (*d*) preconditions for the SMP and AMP modes diminishes the BD-rate loss down to 0.2% in the RA case and 0.3% in the LB case. The respective complexity savings in encoding are still as high as 31% and 32% although the only modification over the default scheme is the repositioned Skip mode comparison.

Table XIII tabulates the RDC characteristics of S_{10} (see Fig. 7), S_{14} (see Fig. 8), and three potential hybrid solutions (S_{11}-S_{13}) between them. S_{11}-S_{13} are used to benchmark the impact of shifting the boundary of the preconditions *d* and *j* gradually towards the larger N_{SMP} values. The rationale for preferring S_{11}-S_{13} to other possible hybrid schemes is that the relative BD-rate gain of the SMP modes decreases as a function of N_{SMP} (see Table XI), i.e., the smaller the SMP the higher its importance. Hence, adopting the looser precondition *d* in increasing order of N_{SMP} as in S_{11}-S_{13} yields the most favorable RDC trade-off among the available hybrid solutions.

Before evaluating S_{11}-S_{13} more accurately, they are compared with the schemes having limited N_{SMP} range (S_{15}-S_{17}). The N_{SMP} range limitations for S_{15}, S_{16}, and S_{17} have been derived respectively from S_{13}, S_{12}, and S_{11} by disabling the SMP sizes that are dependent on the precondition *j*. As a result, the functionalities of S_{15}-S_{17} equal that of S_{14} (see Fig. 8) when the block size of interest is in the N_{SMP} range. Otherwise, these schemes are reduced to those with the disabled SMP modes: Fig. 6 for $N \in \{8, 16\}$ and Fig. 3 for $N = 32$. Hence, S_{15}-S_{17} can be seen as reduced derivatives of S_{14}. The RDC values reported in Table XIV show that S_{14} and its reduced derivatives obtain better RDC trade-offs than S_{10} and hybrid schemes. On average, limiting N_{SMP} range by disabling

TABLE XIII
THE IMPACT OF SMP PRECONDITIONS ON THE RDC CHARACTERISTICS OF HEVC

Cfg.	ID	N_{SMP}	N_{AMP}	QP=22		QP=27		QP=32		QP=37		BD-rate	Average	Average
				Δ bit rate	Δ Mcpf	Δ bit rate	Δ Mcpf	Δ bit rate	Δ Mcpf	Δ bit rate	Δ Mcpf		Δ Mcpf	Δ $Mcpf_{enc}$
RA	S_{14}	{4[d],8[d],16[d],32[d]}	{8,16,32}	0.1%	-23%	0.2%	-32%	0.3%	-39%	0.3%	-45%	0.2%	-35%	-31%
	S_{13}	{4[d],8[d],16[d],32[j]}	{8,16,32}	0.1%	-26%	0.2%	-34%	0.3%	-41%	0.3%	-46%	0.2%	-37%	-32%
	S_{12}	{4[d],8[d],16[j],32[j]}	{8,16,32}	0.2%	-29%	0.3%	-36%	0.5%	-43%	0.5%	-48%	0.4%	-39%	-34%
	S_{11}	{4[d],8[j],16[j],32[j]}	{8,16,32}	0.4%	-33%	0.6%	-39%	0.8%	-44%	0.8%	-50%	0.6%	-42%	-36%
	S_{10}	{4[j],8[j],16[j],32[j]}	{8,16,32}	1.1%	-37%	1.3%	-41%	1.3%	-45%	1.1%	-50%	1.2%	-43%	-37%
LB	S_{14}	{4[d],8[d],16[d],32[d]}	{8,16,32}	0.2%	-23%	0.3%	-32%	0.4%	-39%	0.5%	-45%	0.3%	-35%	-32%
	S_{13}	{4[d],8[d],16[d],32[j]}	{8,16,32}	0.2%	-26%	0.3%	-34%	0.5%	-41%	0.5%	-47%	0.4%	-37%	-33%
	S_{12}	{4[d],8[d],16[j],32[j]}	{8,16,32}	0.3%	-30%	0.5%	-37%	0.6%	-43%	0.8%	-49%	0.5%	-40%	-36%
	S_{11}	{4[d],8[j],16[j],32[j]}	{8,16,32}	0.7%	-35%	0.8%	-41%	1.1%	-46%	1.1%	-51%	0.9%	-43%	-38%
	S_{10}	{4[j],8[j],16[j],32[j]}	{8,16,32}	1.5%	-39%	1.7%	-42%	1.7%	-46%	1.5%	-51%	1.7%	-45%	-40%

d = Disable SMPs if "M_h" = Skip holds (see Fig. 8)
j = Disable SMPs if "M_h" ∈ {Skip, Merge} holds (see Fig. 7)

TABLE XIV
THE IMPACT OF LIMITED SMP RANGE ON THE RDC CHARACTERISTICS OF HEVC

Cfg.	ID	N_{SMP}	N_{AMP}	QP=22		QP=27		QP=32		QP=37		BD-rate	Average	Average
				Δ bit rate	Δ Mcpf	Δ bit rate	Δ Mcpf	Δ bit rate	Δ Mcpf	Δ bit rate	Δ Mcpf		Δ Mcpf	Δ $Mcpf_{enc}$
RA	S_{14}	{4[d],8[d],16[d],32[d]}	{8,16,32}	0.1%	-23%	0.2%	-32%	0.3%	-39%	0.3%	-45%	0.2%	-35%	-31%
	S_{15}	{4[d],8[d],16[d]}	{8,16,32}	0.1%	-30%	0.2%	-38%	0.4%	-45%	0.4%	-50%	0.3%	-41%	-35%
	S_{16}	{4[d],8[d]}	{8,16,32}	0.2%	-35%	0.4%	-42%	0.6%	-48%	0.8%	-53%	0.5%	-44%	-37%
	S_{17}	{4[d]}	{8,16,32}	0.6%	-42%	0.8%	-47%	1.1%	-52%	1.2%	-56%	0.9%	-49%	-41%
LB	S_{14}	{4[d],8[d],16[d],32[d]}	{8,16,32}	0.2%	-23%	0.3%	-32%	0.4%	-39%	0.5%	-45%	0.3%	-35%	-32%
	S_{15}	{4[d],8[d],16[d]}	{8,16,32}	0.2%	-30%	0.3%	-38%	0.5%	-45%	0.7%	-51%	0.4%	-41%	-36%
	S_{16}	{4[d],8[d]}	{8,16,32}	0.4%	-36%	0.6%	-43%	0.8%	-49%	1.1%	-54%	0.7%	-45%	-40%
	S_{17}	{4[d]}	{8,16,32}	0.8%	-44%	1.0%	-48%	1.3%	-53%	1.4%	-57%	1.1%	-50%	-44%

d = Disable SMPs if "M_h" = Skip holds (see Fig. 8)

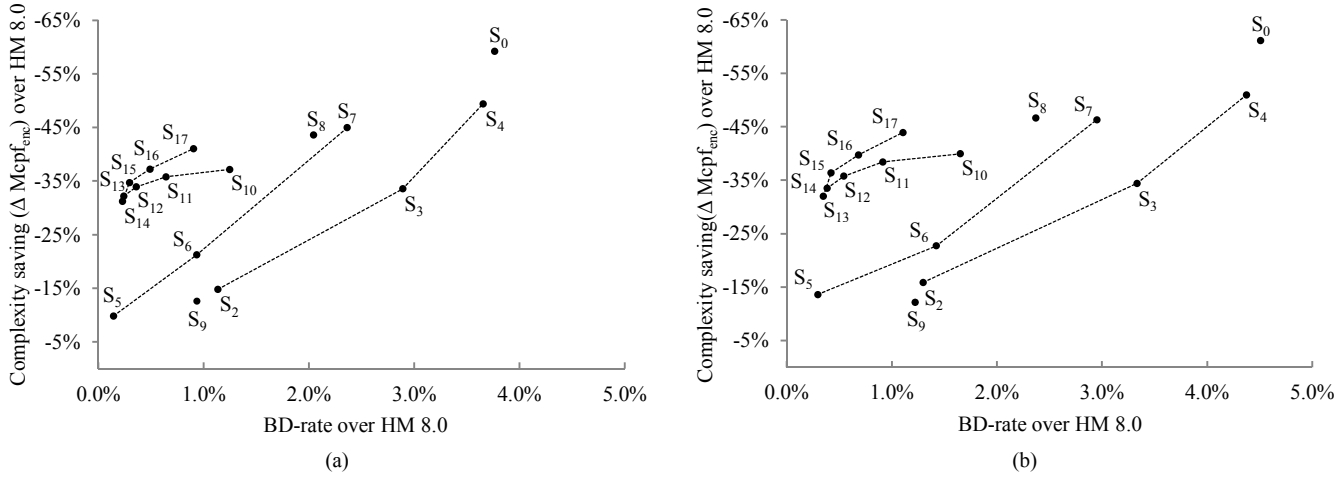

Fig. 9. Comparison of the optimized mode decision schemes in terms of the RDC characteristics. (a) RA case. (b) LB case.

the precondition *j* lowers the encoding complexity by around 4% with the BD-rate penalty of under 0.2%.

Fig. 9 illustrates graphically the RDC characteristics of S_{14}-S_{17} over the other mode decision schemes introduced so far in this paper. Only S_1 is outside the plot due to its increased complexity over HM MP. The schemes belonging to the different groups (S_2-S_4, S_5-S_7, S_{10}-S_{13}, and S_{14}-S_{17}) are connected with dashed lines to improve their visual comparison. The common trend between RA and LB cases is that S_{14}-S_{17} achieve the best RDC trade-off among the scheme groups, so their N_{SMP} ranges are used as a baseline when evaluating the most suitable composition of the AMP modes.

F. Proposed AMP mode decision

Our experiments with the SMP modes revealed the importance of selecting the best possible preconditions and N_{SMP} ranges for them. HM MP already contains the preconditions for the AMP modes, so our focus here is to optimize the default AMP range, i.e., $N_{AMP} \in \{8, 16, 32\}$.

Table XV tabulates the RDC values for the eight candidate configurations (S_{18}-S_{25}) that combine the favored four N_{SMP} ranges (see Table XIV) with the limited N_{AMP} ranges.

TABLE XV
THE IMPACT OF LIMITED AMP RANGE ON THE RDC CHARACTERISTICS OF HEVC

Cfg.	ID	N_{SMP}	N_{AMP}	QP=22		QP=27		QP=32		QP=37		BD-rate	Average	Average
				Δ bit rate	Δ Mcpf	Δ bit rate	Δ Mcpf	Δ bit rate	Δ Mcpf	Δ bit rate	Δ Mcpf		Δ Mcpf	Δ $Mcpf_{enc}$
RA	S_{18}	{4[d],8[d],16[d],32[d]}	{8,16}	0.1%	-24%	0.2%	-33%	0.4%	-40%	0.4%	-46%	0.3%	-36%	-33%
	S_{19}	{4[d],8[d],16[d]}	{8,16}	0.1%	-32%	0.3%	-40%	0.4%	-46%	0.5%	-51%	0.3%	-42%	-38%
	S_{20}	{4[d],8[d]}	{8,16}	0.2%	-37%	0.4%	-44%	0.7%	-49%	0.9%	-54%	0.5%	-46%	-40%
	S_{21}	{4[d]}	{8,16}	0.6%	-43%	0.8%	-48%	1.1%	-53%	1.3%	-57%	0.9%	-50%	-43%
	S_{22}	{4[d],8[d],16[d],32[d]}	{8}	0.3%	-33%	0.5%	-40%	0.7%	-51%	0.8%	-51%	0.6%	-44%	-38%
	S_{23}	{4[d],8[d],16[d]}	{8}	0.3%	-40%	0.6%	-47%	0.9%	-52%	1.0%	-56%	0.7%	-49%	-43%
	S_{24}	{4[d],8[d]}	{8}	0.5%	-47%	0.9%	-53%	1.3%	-57%	1.6%	-60%	1.1%	-54%	-48%
	S_{25}	{4[d]}	{8}	0.9%	-53%	1.3%	-57%	1.8%	-61%	2.1%	-62%	1.5%	-58%	-51%
LB	S_{18}	{4[d],8[d],16[d],32[d]}	{8,16}	0.2%	-24%	0.3%	-32%	0.5%	-40%	0.5%	-46%	0.4%	-35%	-33%
	S_{19}	{4[d],8[d],16[d]}	{8,16}	0.2%	-31%	0.4%	-40%	0.7%	-46%	0.9%	-52%	0.6%	-42%	-39%
	S_{20}	{4[d],8[d]}	{8,16}	0.4%	-37%	0.7%	-44%	1.0%	-50%	1.3%	-55%	0.8%	-46%	-42%
	S_{21}	{4[d]}	{8,16}	0.8%	-45%	1.1%	-49%	1.4%	-54%	1.6%	-57%	1.2%	-51%	-46%
	S_{22}	{4[d],8[d],16[d],32[d]}	{8}	0.5%	-32%	0.8%	-39%	1.1%	-46%	1.3%	-50%	0.9%	-42%	-39%
	S_{23}	{4[d],8[d],16[d]}	{8}	0.5%	-39%	0.9%	-47%	1.3%	-52%	1.7%	-56%	1.1%	-49%	-45%
	S_{24}	{4[d],8[d]}	{8}	0.8%	-47%	1.4%	-53%	1.8%	-57%	2.4%	-60%	1.6%	-54%	-49%
	S_{25}	{4[d]}	{8}	1.3%	-57%	1.8%	-58%	2.4%	-61%	2.7%	-63%	2.0%	-60%	-54%

d = Disable SMPs if M_h" = Skip holds (see Fig. 8)

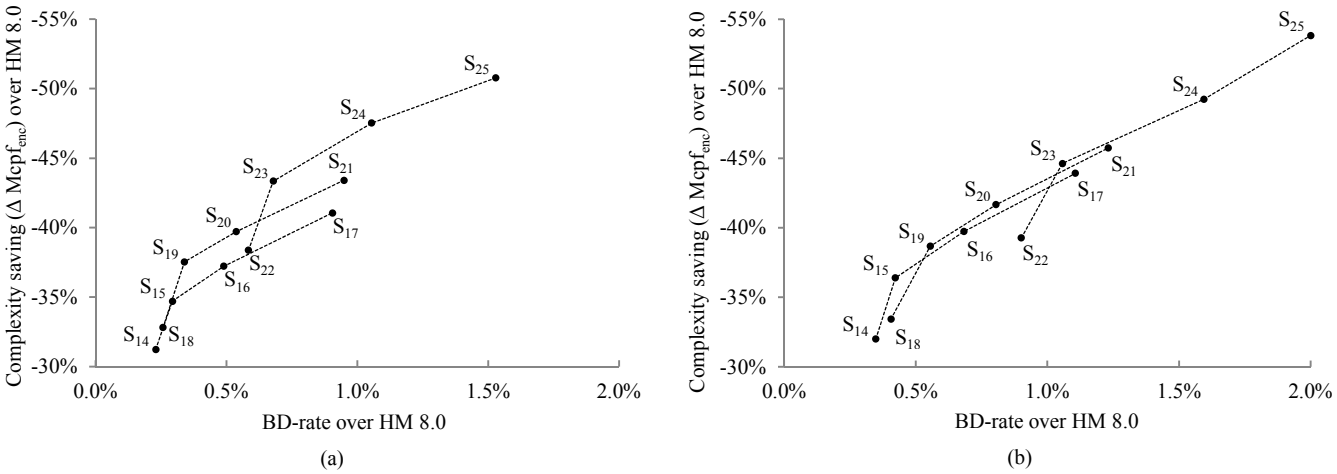

Fig. 10. Comparison of the most potential QP-independent mode decision schemes in terms of the RDC characteristics. (a) RA case. (b) LB case.

Reducing $N_{AMP} \in \{8, 16, 32\}$ to $N_{AMP} \in \{8, 16\}$ in S_{18}-S_{21} means in practice that the Merge mode only testing (see Table I) is disabled for $N_{AMP} = 32$. This optimization comes almost without the BD-rate overhead and it saves the encoding complexity further by over 2% in the RA case. The respective values in the LB case are around 0.1% and 2%. Limiting $N_{AMP} \in \{8, 16\}$ to $N_{AMP} = 8$ in S_{22}-S_{25} has much higher impact on the RDC characteristics. This limitation reduces the encoding complexity further by around 5-8% at a cost of the supplementary BD-rate overhead of 0.3-0.6% in the RA case. The respective values in the LB case are 6-8% and 0.5-0.8%.

Fig. 10 visualizes the impact of the limited N_{AMP} range by plotting the RDC characteristics of S_{14}-S_{17}, S_{18}-S_{21}, and S_{22}-S_{25}. These three scheme groups are indicated with the dashed lines. In addition, the coordinates of the plot have been zoomed in from Fig. 9 to better illustrate the differences between the closely distributed RDC values. These scheme groups obtain the top RDC characteristics among the examined QP-independent approaches, but the crossing lines indicate that none of them provides the best gain alone. Instead, the QP-specific values in Tables XIV and XV let us assume that composing a single scheme from the variable ranges of N_{SMP} and N_{AMP} can improve the outcome further.

G. *Proposed QP-specific mode decision*

Table XVI gathers the most favored QP-specific ranges of N_{SMP} and N_{AMP} from Tables XIV and XV. Here, the step size of QPs is maintained in five, i.e., $QP \in \{22, 27, 32, 37\}$. Shortening the step size would improve the accuracy of the QP-specific specifications but at a cost of more complicated parameterization of N_{SMP} and N_{AMP}.

Our recommendations are based on the RDC characteristics according to which five different ranges of N_{SMP} and N_{AMP} are listed per a tested QP. At each QP of interest, these ranges have been adopted from the schemes whose RDC characteristics meet a local maximum point, i.e., the highest prediction complexity drop among the schemes at a certain Δ bit rate value. If more than five local maximum points exist, the best ones have been selected under the following RDC

TABLE XVI
THE LIST OF RECOMMENDED QP-SPECIFIC RANGES OF N_{SMP} AND N_{AMP}

Cfg.	N_{SMP}				N_{AMP}				Δ bit rate / Δ Mcpf			
	QP=22	QP=27	QP=32	QP=37	QP=22	QP=27	QP=32	QP=37	QP=22	QP=27	QP=32	QP=37
RA	{4d,8d,16d}			{4d,8d,16d,32d}	{8,16}			{8,16,32}	0.1% / -32%	0.3% / -40%	0.4% / -46%	0.3% / -45%
	{4d,8d}			{4d,8d,16d}	{8,16}			{8,16,32}	0.2% / -37%	0.4% / -44%	0.7% / -49%	0.4% / -50%
	{4d,8d,16d}			{4d,8d}	{8}			{8,16,32}	0.3% / -40%	0.6% / -47%	0.9% / -52%	0.8% / -53%
	{4d,8d}			{4d,8d,16d}	{8}				0.5% / -47%	0.9% / -53%	1.3% / -57%	1.0% / -56%
	{4d}			{4d,8d}	{8}				0.9% / -53%	1.3% / -57%	1.8% / -61%	1.6% / -60%
LB	{4d,8d,16d}		{4d,8d,16d,32d}		{8,16}	{8,16,32}			0.2% / -31%	0.3% / -38%	0.4% / -39%	0.5% / -45%
	{4d,8d}		{4d,8d,16d}		{8,16}	{8,16,32}			0.4% / -37%	0.6% / -43%	0.5% / -45%	0.7% / -51%
	{4d,8d,16d}		{4d,8d}		{8}		{8,16,32}		0.5% / -39%	0.9% / -47%	0.8% / -49%	1.1% / -54%
	{4d,8d}		{4d}		{8}		{8,16,32}		0.8% / -47%	1.4% / -53%	1.3% / -53%	1.4% / -57%
	{4d}		{4d,8d}		{8}				1.3% / -57%	1.8% / -58%	1.8% / -57%	2.4% / -60%

d = Disable SMPs if "M_h" = Skip holds (see Fig. 8)

TABLE XVII
THE PROPOSED QP-INDEPENDENT RANGES OF THE SMP AND AMP MODES

Cfg.	ID	N_{SMP}				N_{AMP}				Δ bit rate			Δ $Mcpf_{enc}$		
		QP=22	QP=27	QP=32	QP=37	QP=22	QP=27	QP=32	QP=37	min	max	BD-rate	min	max	avg
RA	S_{14}	{4d,8d,16d,32d}				{8,16,32}				0.1%	0.3%	0.2%	-19%	-42%	-31%
	S_{15}	{4d,8d,16d}				{8,16,32}				0.1%	0.4%	0.3%	-22%	-45%	-35%
	S_{16}	{4d,8d}				{8,16,32}				0.2%	0.8%	0.5%	-26%	-47%	-37%
	S_{17}	{4d}				{8,16,32}				0.6%	1.2%	0.9%	-31%	-50%	-41%
LB	S_{14}	{4d,8d,16d,32d}				{8,16,32}				0.2%	0.5%	0.3%	-20%	-43%	-32%
	S_{15}	{4d,8d,16d}				{8,16,32}				0.2%	0.7%	0.4%	-24%	-47%	-36%
	S_{16}	{4d,8d}				{8,16,32}				0.4%	1.1%	0.7%	-28%	-50%	-40%
	S_{17}	{4d}				{8,16,32}				0.8%	1.4%	1.1%	-34%	-52%	-44%

d = Disable SMPs if "M_h" = Skip holds (see Fig. 8)

conditions: additional prediction complexity decrement has to be at least 2% if Δ bit rate increases by 0.1%. In the RA case, the ranges of N_{SMP} and N_{AMP} are adopted for QP ∈ {22, 27, 32} from S_{19}, S_{20}, S_{23}, S_{24}, S_{25} and for QP = 37 from S_{14}, S_{15}, S_{16}, S_{23}, S_{24}. In the LB case, the respective source schemes are S_{19}, S_{20}, S_{23}, S_{24}, S_{25} for QP = 22, S_{15}, S_{16}, S_{23}, S_{24}, S_{25} for QP = 27, and S_{14}, S_{15}, S_{16}, S_{17}, S_{24} for QP ∈ {32, 37}.

VI. GUIDELINES AND COMPARISON

Our analysis results can be generalized to guidelines for HEVC encoder optimization. The guidelines given can be divided into QP-independent and QP-specific ones.

A. Guidelines for QP-independent optimizations

When the default configuration of HM MP is used as a baseline, the recommended order of steps in the QP-independent optimization is

1) Use the Skip mode testing as a precondition for the SMP modes, i.e., set $N_{SMP} \in \{4^d, 8^d, 16^d, 32^d\}$
2) Disable the SMP modes with $N = 32$, i.e., set $N_{SMP} \in \{4^d, 8^d, 16^d\}$
3) Disable the SMP modes with $N = 16$, i.e., set $N_{SMP} \in \{4^d, 8^d\}$
4) Disable the SMP modes with $N = 8$, i.e., set $N_{SMP} \in \{4^d\}$

This four-step optimization decrements the complexity of HM MP encoding gradually up to 41% in the RA case and 44% in the LB case. The respective BD-rate increments of the encoder are 0.9% and 1.1%. Almost an identical BD-rate penalty is paid when the AMP modes are completely disabled (see Table XII), but in that case the complexity saving is only under one third of that obtained by our four-step proposal. This difference is due to the much higher optimization potential related to the evaluation of the SMP modes.

Table XVII summarizes the RDC characteristics of these four steps which, in fact, equal the schemes recommended after our SMP analysis (S_{14}-S_{17}). The most significant step is the first one, which alone reduces the average (*avg*) encoding complexity by over 30%. After that, the complexity decreases by 2-4% per step whereas the relative step size of the BD-rate overhead grows from 0.1% to 0.4% as a function of these steps. In Table XVII, the ranges of the Δ bit rate and encoding complexity are reported through the QP-specific minimum (*min*) and maximum (*max*) bounds which in each step are met with the QPs of 22 and 37, respectively. The QP-specific Δ bit rates and prediction complexities for these steps can be obtained from Table XIV.

B. Guidelines for QP-specific optimizations

The RDC values obtained by our QP-independent four-step proposal can be further improved through QP-specific parameterization of N_{SMP} and N_{AMP}. This more fine-grained parameterization reuses the same N_{SMP} ranges but assigns them separately for each QP and combines them with a default or a limited range of N_{AMP}. The most eligible combinations of N_{SMP} and N_{AMP} ranges are the ones listed in Table XVI.

Table XVIII contains three examples of these QP-specific configurations (S_{26}-S_{28}) whose BD-rate overheads over HM MP have been preset to 0.5%, 0.9%, and 1.3%, respectively. Since the QP-specific Δ bit rates of all examined schemes increase as a function of QP, the preset BD-rate limits are met safely, if each intermediate QP between the tested ones

TABLE XVIII
THE PROPOSED QP-SPECIFIC RANGES OF THE SMP AND AMP MODES

Cfg.	ID	N_{SMP} QP=22	QP=27	QP=32	QP=37	N_{AMP} QP=22	QP=27	QP=32	QP=37	Δ bit rate min	max	BD-rate	Δ $Mcpf_{enc}$ min	max	avg
RA	S_{26}	{4^d,8^d}	{4^d,8^d,16^d}			{8}		{8,16}	{8,16,32}	0.4%	0.6%	0.5%	-38%	-45%	-42%
	S_{27}	{4^d}	{4^d,8^d}	{4^d,8^d,16^d}		{8}				0.9%	1.0%	0.9%	-43%	-52%	-47%
	S_{28}	{4^d}		{4^d,8^d}		{8}				0.9%	1.6%	1.3%	-43%	-55%	-50%
LB	S_{26}	{4^d,8^d,16^d}	{4^d,8^d}	{4^d,8^d,16^d}	{4^d,8^d,16^d,32^d}	{8}	{8,16,32}			0.5%	0.6%	0.5%	-34%	-43%	-39%
	S_{27}	{4^d,8^d}	{4^d,8^d,16^d}	{4^d,8^d}		{8}	{8,16,32}			0.8%	1.1%	0.9%	-40%	-50%	-44%
	S_{28}	{4^d}	{4^d,8^d}	{4^d}		{8}	{8,16,32}			1.3%	1.4%	1.3%	-47%	-52%	-49%

d = Disable SMPs if "M_h" = Skip holds (see Fig. 8)

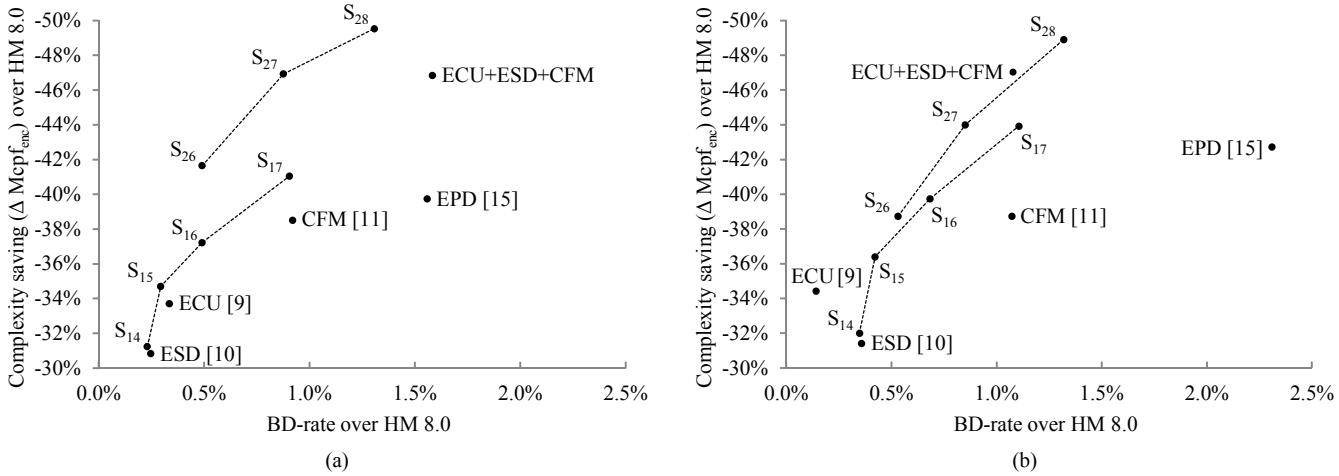

Fig. 11. Comparison of the proposed and contemporary mode decision schemes in terms of the RDC characteristics (HM 8.0). (a) RA case. (b) LB case.

parameterizes the scheme as the higher one of the tested QPs. That is, the QP-specific parameterizations tabulated for S_{26}-S_{28} are recommended to apply to QP = 22, QP ∈ [23, 27], QP ∈ [28, 32], and QP ∈ [33, 37], respectively. On average, this approach increases the reported complexity by around 2% but without the BD-rate compromises. The reparameterization of S_{26}-S_{28} can be implemented at CTU-level, which eases their adaptation to the different QPs.

The QP-specific schemes offer additional flexibility to meet the desired RDC characteristics. For example, the further complexity drop of S_{28} over S_{17} is 9% in the RA case and 5% in the LB case. The respective additional BD-rate overheads are 0.4% and 0.2%, so the RDC impact of this optimization is close to that of one step in our four-step proposal. These schemes also lessen the gap between the minimum (*min*) and maximum (*max*) QP-specific values.

Besides the RDC characteristics, the easiness of configuration has been one of the main design principles when developing our optimization techniques. The precondition *d* (see Fig. 8) and the QP-specific parameterization of N_{SMP} and N_{AMP} can be seamlessly incorporated in the existing control structures of the HEVC encoder. Furthermore, none of these schemes limits parallel processing of different CTUs or adjacent coding tree depths of a single CTU. Conditional SMP/AMP evaluations are effective optimization techniques when PBs at each coding tree depth *h* are processed sequentially, but their benefit is easily revoked when PBs are processed in parallel. However, limiting SMP/AMP ranges also simplifies parallel processing at level *h* since fewer PB types need to be supported.

C. Comparison with existing techniques

The existing speed-up methods of HM (ECU, ESD, and CFM) have been benchmarked individually in [9]-[11], but all those evaluations have been performed on obsolete HM versions (HM 5.0 or earlier). With HM 8.0, only the combination of the several speed-up methods has been tested (see Table II). In addition, the presented results have been derived from a limited subset of test sequences relying on runtime as a complexity measure. In our experiments, ECU, ESD, and CFM have been individually benchmarked in HM 8.0 under the equal test conditions and with the same test set as the proposed schemes so that a fair comparison is obtained.

Fig. 11 visualizes the RDC characteristics of ECU, ESD, and CFM against the proposed QP-independent (S_{14}-S_{17}) and QP-specific (S_{26}-S_{28}) schemes (see Tables XVII and XVIII). In the RA case, the proposed schemes outperform these built-in methods without compromises. For example, S_{14}, S_{15}, and S_{17} achieve higher complexity reduction with the smaller RD overhead than ESD, ECU, and CFM, respectively. The RDC characteristics of our optimization techniques are also better than those obtained when ECU, EDS, and CFM are all enabled (ECU+ESD+CFM). For example, S_{27} achieves equal complexity loss than ECU+ESD+CFM but it almost halves the BD-rate overhead from 1.6% to 0.9%.

In the LB case, the proposed schemes outperform ESD and CFM as in the RA case. However, ECU is able to provide the best RDC trade-off among the schemes whose complexity drop is up to 34%. In addition, ECU+ESD+CFM is competitive to S_{27} and S_{28} if its BD-rate overhead of 1.1% is

TABLE XIX
RDC COMPARISON OF THE PROPOSED QP-INDEPENDENT (S_{14}-S_{17}) AND QP-SPECIFIC (S_{26}-S_{28}) SCHEMES IN HM 8.0 AND HM 11.0

Cfg.	ID	N_{SMP} QP=22	N_{SMP} QP=27	N_{SMP} QP=32	N_{SMP} QP=37	N_{AMP} QP=22	N_{AMP} QP=27	N_{AMP} QP=32	N_{AMP} QP=37	BD-rate HM 8.0/HM 11.0	Δ Mcpf$_{enc}$ HM 8.0	Δ Mcpf$_{enc}$ HM 11.0
RA	S_{14}	{4^d,8^d,16^d,32^d}				{8,16,32}				0.2%	-31%	-34%
	S_{15}	{4^d,8^d,16^d}				{8,16,32}				0.3%	-35%	-37%
	S_{16}	{4^d,8^d}				{8,16,32}				0.5%	-37%	-40%
	S_{17}	{4^d}				{8,16,32}				0.9%	-41%	-43%
	S_{26}	{4^d,8^d}	{4^d,8^d,16^d}			{8}	{8,16}	{8,16,32}		0.5%	-42%	-44%
	S_{27}	{4^d}	{4^d,8^d}	{4^d,8^d,16^d}		{8}				0.9%	-47%	-48%
	S_{28}	{4^d}		{4^d,8^d}		{8}				1.3%	-50%	-51%
LB	S_{14}	{4^d,8^d,16^d,32^d}				{8,16,32}				0.3%	-32%	-36%
	S_{15}	{4^d,8^d,16^d}				{8,16,32}				0.4%	-36%	-40%
	S_{16}	{4^d,8^d}				{8,16,32}				0.7%	-40%	-43%
	S_{17}	{4^d}				{8,16,32}				1.1%	-44%	-46%
	S_{26}	{4^d,8^d,16^d}	{4^d,8^d}	{4^d,8^d,16^d}	{4^d,8^d,16^d,32^d}	{8}	{8,16,32}			0.5%	-39%	-42%
	S_{27}	{4^d,8^d}	{4^d,8^d,16^d}	{4^d,8^d}		{8}	{8,16,32}			0.9%	-44%	-47%
	S_{28}	{4^d}	{4^d,8^d}	{4^d}		{8}	{8,16,32}			1.3%	-49%	-50%

d = Disable SMPs if M_h" = Skip holds (see Fig. 8)

allowed. However, the QP-specific complexity drop of ECU+ESD+CFM varies between -29% and -68%, so its complexity gap is much wider than that of S_{27} or S_{28} (see Table XVIII). Anyway, the potential future work is to find synergy gain between these built-in methods and our techniques in order to optimize the HEVC mode decision further.

The other speed-up techniques [13]-[15] introduced for the HEVC mode decision operate on different threshold values that are able to reduce the depth of a coding tree in advance and/or terminating the ongoing mode decision process. The first approach [13] uses Bayesian decision theory to evaluate probability functions according to which a decision for further quadtree partition is made at each quadtree depth *h*. The second approach [14] pre-selects the depth range of a coding tree as a function of the spatially and temporally related CUs and exploits similar content-adaptive thresholds to speed-up the pre-limited mode decision processes. The third approach [15] introduces *early partition decision* (*EPD*) algorithm, which terminates the entire operation at depth *h* and further depths ($h + 1 \ldots h_{max}$) if the RD cost of the evaluated Square, SMP, or AMP modes is smaller than the average RD cost of the previously coded skipped CUs of same size. The RDC values of EPD have been enhanced by combining ECU with it.

These three approaches are all based on HM 4.0 or earlier versions, so their RDC values are not directly comparable to ours. In general, all of them are able to lower the encoding time of HM by around 40%. However, the respective BD-rate overhead of the first approach [13] is close to 2% and the second approach [14] is assumed to suffer from the similar overhead if the impact of the associated PSNR deviations is added to the reported bit rate differences. The third approach [15] featuring around 1% BD-rate penalty over HM 4.0 is the most potential one of these works. Therefore, EPD has been implemented in HM 8.0 and benchmarked against the proposed schemes with the same test setup. The RDC comparison in Fig. 11 shows that our approach outperforms also EPD in the RA and LB cases. In addition, more complex control together with an unlimited set of coding modes makes EPD a less hardware-oriented approach.

D. Validation of RDC results in HM 11

The robustness of the obtained RDC results have been confirmed by benchmarking the proposed and contemporary schemes in HM 11.0 which was the most up-to-date version at the time our development work was finished.

Table XIX gathers our QP-independent (S_{14}-S_{17}) and QP-specific (S_{26}-S_{28}) schemes together and compares their RDC characteristics in the context of HM 8.0 and HM 11.0. The BD-rate penalties met by incorporating our schemes in HM 11.0 are consistent with the results of HM 8.0. On average, the respective complexity gain is 2 percentage points in the RA case and 3 percentage points in the LB case. That is, S_{14}-S_{17} reduce the complexity of the default HM 11.0 configuration by 34-43% and S_{26}-S_{28} by 44-51% in the RA case. The respective values in the LB case are 36-46% and 42-50%. The RDC characteristics of the previous schemes behave similarly than those of our schemes, so the gap between the proposed and previous approaches remains approximately the same as in HM 8.0.

VII. CONCLUSION

This paper examined the RDC characteristics of the HEVC inter prediction and used the obtained RDC results to optimize the mode decision and associated block partition structures of the HEVC reference encoder (HM 8.0). The resolutions of the test sequences varied from WQVGA to 1600p and the evaluated mode decision schemes were benchmarked with the QPs of 22, 27, 32, and 37 under RA and LB cases. The analysis relied on PSNR as an objective quality measure, whereas the complexities were obtained through cycle-level profiling with Intel VTune.

This work focused on the mode decision schemes whose bit rate overheads over HM MP are close to 1% at maximum. Therefore, the structure of a coding tree and the square motion partitions were adopted from the HM MP without optimizations due to their key role in RD performance gain. In addition, no optimizations were performed for the Intra and Skip modes because of their diminutive impact on complexity.

Hence, all our efforts were directed to investigating the most feasible composition of the SMP and AMP modes whose combined impact on the complexity of the HM MP encoder was shown to be 60%. This additional complexity was evaluated to give the bit rate gain of 3.8% in the RA case and 4.5% in the LB case. Hence, obtaining most of this RD gain with a diminutive complexity overhead was set our target.

This work tackles the complexity overhead of the SMP and AMP mode evaluation with three new techniques: 1) the conditional evaluation of the SMP modes, 2) the range limitations primarily in the SMP sizes and secondarily in the AMP sizes, and 3) the selection of the SMP and AMP ranges as a function of the QP. The first two of these techniques are utilized by the proposed QP-independent schemes (S_{14}-S_{17}) whose RDC characteristics are further enhanced by the proposed QP-specific schemes (S_{26}-S_{28}) which exploit all these three techniques. In the RA case, the proposed schemes reduce the complexity of the HM MP encoder by 31-51% at a cost of 0.2-1.3% bit rate increase. The respective values in the LB case are 32-50% and 0.3-1.3%. The QP-independent schemes lower the complexity by up to 43% in the RA case and by up to 46% in the LB case after which the remaining complexity reduction is obtained through QP-specific schemes.

The proposed optimizations are hardware oriented and they can be seamlessly incorporated in the existing control structures of the HEVC encoder. In the future, our purpose is to combine the proposed techniques with the other existing approaches without compromising the RD performance or losing the ability for parallelization and hardware acceleration.

REFERENCES

[1] ITU-T Recommendation H.265, "High efficiency video coding," International Telecommunication Union, Apr. 2013.

[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, Dec. 2012, pp. 1649-1668.

[3] ITU-T Recommendation H.264, "Advanced video coding for generic audiovisual services," *International Telecommunication Union*, Mar. 2009.

[4] J. Vanne, M. Viitanen, T. D. Hämäläinen, and A. Hallapuro, "Comparative rate-distortion-complexity analysis of HEVC and AVC video codecs," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, Dec. 2012, pp. 1885-1898.

[5] J. R. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand," Comparison of the coding efficiency of video coding standards – including High Efficiency Video Coding (HEVC)," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, Dec. 2012, pp. 1669-1684.

[6] I. K. Kim, J. Min, T. Lee, W. J. Han, J. Park, "Block partitioning structure in the HEVC Standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, Dec. 2012, pp. 1697-1706.

[7] Y. Yuan, I. K. Kim, X. Zheng, L. Liu, X. Cao, S. Lee, M. S. Cheon, T. Lee, Y. He, and J. H. Park, "Quadtree based non-square block structure for inter frame coding in HEVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, Dec. 2012, pp. 1707-1719.

[8] F. Bossen, B. Bross, K. Sühring, and D. Flynn, "HEVC complexity and implementation analysis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, Dec. 2012, pp. 1685-1696.

[9] K. Choi, S. H. Park, and E. S. Jang, "Coding tree pruning based CU early termination," *document JCTVC-F092*, Torino, Italy, Jul. 2011.

[10] J. Yang, J. Kim, K. Won, H. Lee, and B. Jeon, "Early SKIP detection for HEVC," *document JCTVC-G543*, Geneva, Switzerland, Nov. 2011.

[11] R. H. Gweon and Y.-L. Lee, "Early termination of CU encoding to reduce HEVC complexity," *document JCTVC-F045*, Torino, Italy, Jul. 2011.

[12] C. E. Rhee, K. Lee, T. S. Kim, and H. J. Lee, "A survey of fast mode decision algorithms for inter-prediction and their applications to high efficiency video coding," *IEEE Trans. Consumer Electron.*, vol. 58, no. 4, Nov. 2012, pp. 1375-1383.

[13] X. Shen, L. Yu, and J. Chen, "Fast coding unit size selection for HEVC based on Bayesian decision rule," in *Proc. Picture Coding Symp.*, May 2012, pp. 453-456.

[14] L. Shen, Z. Liu, X. Zhang, W. Zhao, and Z. Zhang, "An effective CU size decision method for HEVC encoders," *IEEE Trans. Multimedia*, vol. 15, no. 2, Feb. 2013, pp. 465-470.

[15] H. L. Tan, F. Liu, Y. H. Tan, and C. Yeo, "On fast coding tree block and mode decision for high-Efficiency Video Coding (HEVC)," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 2012, pp. 825-828.

[16] I. K. Kim, K. McCann, K. Sugimoto, B. Bross, and W. J. Han, "HM8: High Efficiency Video Coding (HEVC) Test Model 8 encoder description," *document JCTVC-J1002*, Stockholm, Sweden, Jul. 2012.

[17] Joint Collaborative Team on Video Coding Reference Software, ver. HM 8.0, Available online: http://hevc.hhi.fraunhofer.de/.

[18] I. K. Kim, K. McCann, K. Sugimoto, B. Bross, and W. J. Han, "High Efficiency Video Coding (HEVC) Test Model 11 (HM11) encoder description," *document JCTVC-M1002*, Incheon, South Korea, Apr. 2013.

[19] F. Bossen, "Common test conditions and software reference configurations," *document JCTVC-J1100*, Stockholm, Sweden, Jul. 2012.

[20] P. Helle, S. Oudin, B. Bross, D. Marpe, M. O. Bici, K. Ugur, J. Jung, G. Clare, and T. Wiegand, "Block merging for quadtree-based partitioning in HEVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, Dec. 2012, pp. 1720-1731.

[21] G. Corrêa, P. Assuncao, L. Agostini, and L. A. da Silva Cruz, "Performance and Computational Complexity Assessment of High Efficiency Video Encoders," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, Dec. 2012, pp. 1899-1909.

[22] G. Bjøntegaard, "Calculation of average PSNR differences between RD curves," *document VCEG-M33*, Austin, TX, USA, Apr. 2001, pp. 1-4.

Jarno Vanne (M'02) received the M.Sc. degree in information technology and the Ph.D. degree in computing and electrical engineering from the Tampere University of Technology (TUT), Tampere, Finland, in 2002 and 2011, respectively.

He is currently a Research Fellow with the Department of Pervasive Computing, TUT. His current research interests include video and image processing systems, video coding standards, motion estimation, parallel memories, and computer arithmetic.

Marko Viitanen (M'09) is currently pursuing the M.Sc. degree in computer systems with the Tampere University of Technology (TUT), Tampere, Finland.

He is currently a Research Assistant with the Department of Pervasive Computing, TUT. His current research interests include video compression, video coding standards, and performance analysis.

Timo D. Hämäläinen (M'95) received the M.Sc. and Ph.D. degrees from the Tampere University of Technology (TUT), Tampere, Finland, in 1993 and 1997, respectively.

He has been a Full Professor at TUT since 2001. He is an author of more than 70 journals and 210 conference publications. He holds several patents. His current research interests include design methods and tools for multiprocessor systems-on-a-chip and parallel video codec implementations.