Scaling Bandwidth and Complexity of Hybrid MC/DCT Video Codecs

Kristof Denolf, Christophe De Vleeschouwer[1], Gauthier Lafruit and Jan Bormans

IMEC, Kapeldreef 75, 3001 Leuven, BELGIUM
kristof.denolf@imec.be

Abstract. Multi-user video conferencing has high and diverse bandwidth and performance requirements. We analyse the scaling behaviour of a standard hybrid video codec by varying the quantization degree, the temporal and spatial resolution. We define a framework to select the set of coding parameters when the bandwidth and complexity constraints are known. Additionally, the efficiency of simulcast and layered solutions to overcome large bandwidth and/or computational resource variations among the participants are compared. In this context, the temporal resolution turns out to be the best-suited parameter to provide multiple resolutions through independent streams dealing with the network unreliability.

1 Introduction

Video conferencing applications are very demanding, both in terms of data transport and processing performance. In practice, distinctive users can interact, ranging from a mobile, remote worker to a central, powerful node in the company. Consequently, the implementation constraints are characterised by a large diversity of bandwidth and performance. Data streams can flow in different ways to the participants: point-to-point, one to many (broadcast or multicast) and many to one (one is a manager, receiving, processing, and distributing streams to many others). The communication channel can be a fixed line (expensive) guaranteeing minimal bandwidth and maximal delay or, more likely, a best-effort and/or error-prone network, using wired and/or wireless connections. In that case, bandwidth and delay are not guaranteed and are fluctuating. The video conferencing equipment can range from high quality, dedicated appliances over a multimedia-enhanced desktop PC to a mobile power-constrained device (e.g. laptop, PDA).

The main goal of this paper is to analyse how the bandwidth and complexity of a hybrid Motion Compensated/Discrete Cosine Transform (MC/DCT) video encoder/decoder (H.26x/MPEG) scale with the encoding parameters. To overcome the variety of network and terminal resources of a possible conference participant, we argue 1) the selection of its lower cost operating point by taking both bandwidth and complexity constraints into account, 2) the use of the three main coding parameters (QP, spatial and temporal resolution). Additionally, we evaluate the benefit of layering versus simulcast solutions when concurrently distributing the same content to multiple heterogeneous users (broadcast or multicast). In that context, we highlight the additional benefits of simple, non-layered, but multiple-descriptive codecs in an error-prone environment.

In order to evaluate the alternatives to deal with the (simultaneous) user heterogeneity, Section 2 and 3 consider the impact of the coding parameters (the quantisation degree, the frame rate and the resolution) on the bandwidth and codec complexity respectively. Section 4 envisions a simple point-to-point connection and investigates how to set the coding parameters, depending on the complexity and bandwidth constraints on both sides of the connection. Section 5 addresses the problem of a participant/server delivering visual content to a set of heterogeneous receivers. The section evaluates the efficiency of conventional spatial- and SNR-layering when targeting coarse grained encoding points and concludes on the fact that simple simulcast solutions provide better bandwidth/complexity trade-offs in most cases. Moreover, a non-naive temporal simulcast-like solution, consisting of two independent and intertwined streams at a lower frame rate, offers a good alternative to B-frames (= layering) in an error-prone environment.

---

[1] Currently at UCL, Place du Levant 2, 1348 Louvain la Neuve, Belgium, funded by the Belgian National Fund.

2 Bandwidth Scaling

The test-bench used in this paper proposes sequences of different complexities. *Mother and Daughter* is a low complexity sequence, a head and shoulders sequence. *Foreman* has a medium complexity and *Calendar and Mobile* is a high complexity sequence, with a lot of movement including rotation. Only the results of the two extreme cases are included, but the observations and conclusions are inferred from the complete test set. All sequences are compressed using a proprietary MPEG-4 video encoder [1]. The rate control has been disabled to exclude the QP variation along the sequence.

2.1 SNR Scaling: QP

The Quantisation Parameter (QP) controls the accuracy of the texture information. A high QP lowers the number of bits required to code a texture block at the cost of precision. Figure 2.1.a,b display the influence of the QP at different frame rates and spatial resolutions (QCIF and CIF).

On Figure 2.1.a,b the absolute bitrate depends on the sequence, but its relative decrease is quasi independent of both the sequence type (complex or not) and resolution (QCIF or CIF). A major trend to notice on these graphs is the exponential decrease of the bit-rate with QP. This drop is particularly fast and significant at high frame rates. At lower frame rates, the higher relative amount of information contained in the motion vectors reduces the impact of QP on the bitstream size. It is further discussed in Section 5.

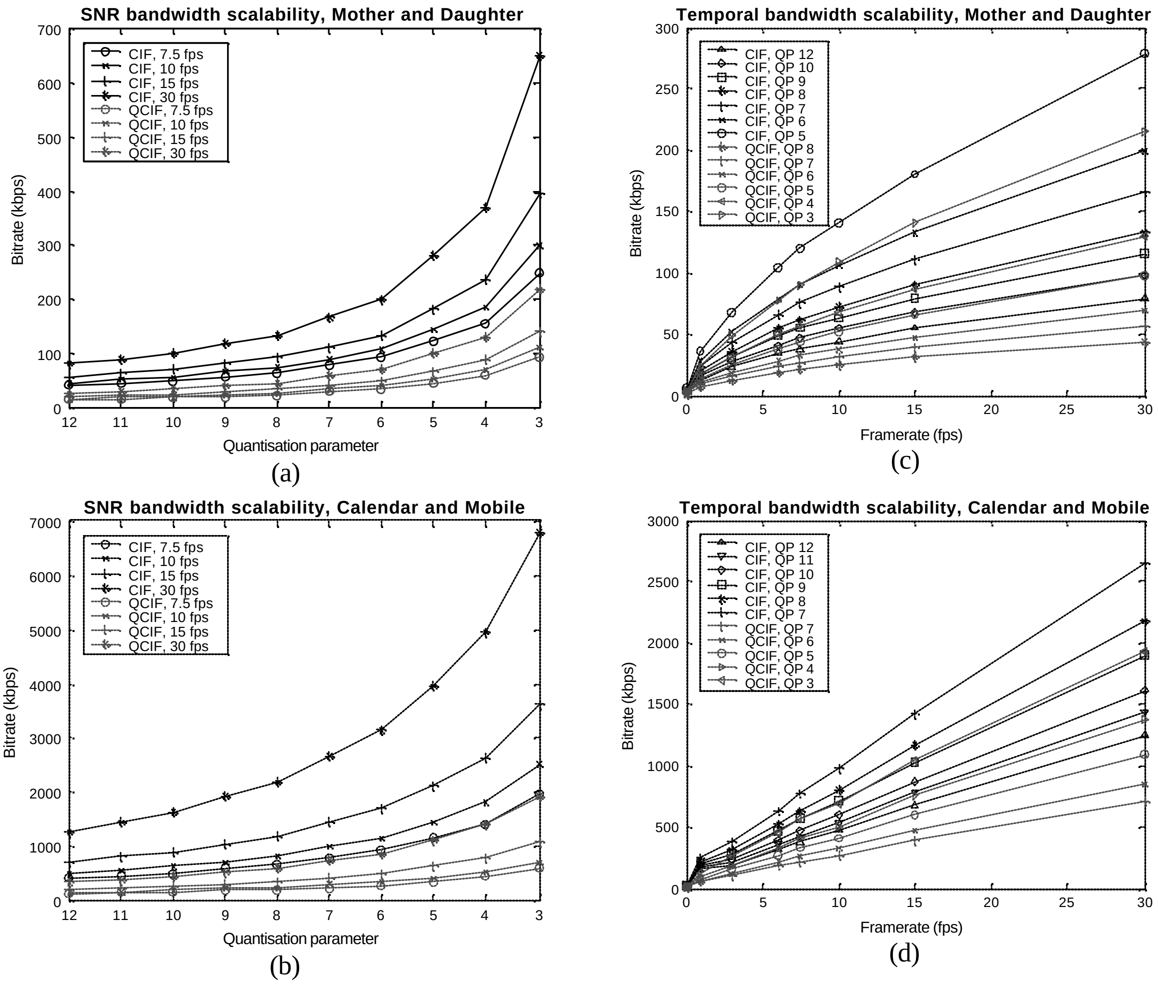

Figure 2.1: Bandwidth scaling by QP increasing (a,b) and by lowering the frame rate (c,d) has respectively an exponential and quasi-linear impact. The results are plotted for the *Mother and Daughter* (a,c) and *Calendar and Mobile* (b,d) video sequences.

2.2 *Temporal Resolution Scaling: Frame rate*

Lowering the frame rate decreases the bit rate. Due to accuracy/efficiency loss of the Motion Estimation (ME) at low frame rates and due to the variable length, predictive motion vector encoding method, the reduction is expected to be non-linear. Figure 2.1.c,d illustrate the bandwidth scaling with temporal resolution for a set of sequences, encoded with different QP and spatial resolution (QCIF or CIF):

- A high quantisation level masks the temporal scaling effect (smaller slope of the curve). In that case, the motion description part dominates the bitstream.
- For complex sequences, increasing the frame rate has a higher relative cost than for simple and temporally correlated sequences. However, this increase is never more than linear.
- The graph is close to a linear at high frame rates.
- The spatial resolution (QCIF and CIF) has no influence on the relative bitrate reduction.

2.3 *Spatial Resolution Scaling: QCIF or CIF*

The effect of the frame size reduction (from CIF to QCIF) on the bitrate is illustrated in Table 1. The bandwidth drop slightly increases with the frame rate and the complexity of the sequence. Actually, QCIF frames are obtained by sub-sampling spatially filtered CIF sequences (to avoid aliasing). Consequently, the relative amount of information per encoded macroblock is increased at lower resolution and hence, the bitrate drop is below the expected factor four. For low complexity sequences containing few high frequency spatial components, the filtering removes few information and the relative bitrate drop is lower. A higher quantisation parameter increases the effect of the frame size reduction on the bitrate as it removes more of this relative increment of high frequency components.

Table 1: Spatial total bandwidth reduction factors *Mother and Daughter* and *Calendar and Mobile*

fps \ QP	3.00	6.00	7.50	10.00	15.00	30.00
12	3.2	3.2	3.3	3.2	3.2	3.3
11	3.1	3.1	3.2	3.2	3.1	3.2
10	3.1	3.1	3.1	3.1	3.1	3.1
9	3.0	3.0	3.0	3.0	3.0	3.1
8	2.9	3.0	2.9	2.9	2.9	3.0
7	2.9	2.9	2.8	2.8	2.8	2.9
6	2.9	2.8	2.8	2.8	2.8	2.9
5	2.8	2.7	2.7	2.7	2.7	2.9
4	2.7	2.7	2.7	2.7	2.7	2.9
3	2.8	2.7	2.7	2.8	2.8	3.0

fps \ QP	3.00	6.00	7.50	10.00	15.00	30.00
12	3.7	3.6	3.6	3.8	3.8	3.9
11	3.6	3.6	3.6	3.7	3.7	3.9
10	3.6	3.6	3.6	3.7	3.7	3.9
9	3.6	3.6	3.5	3.7	3.7	3.8
8	3.6	3.5	3.5	3.6	3.6	3.8
7	3.5	3.5	3.5	3.6	3.6	3.7
6	3.5	3.5	3.5	3.6	3.6	3.7
5	3.5	3.5	3.5	3.6	3.5	3.6
4	3.5	3.5	3.4	3.5	3.5	3.6
3	3.5	3.5	3.5	3.6	3.5	3.5

3 Complexity Scaling

The evaluation of the complexity of the hybrid MC/DCT video coded is based on performance measurements of a proprietary MPEG-4 simple profile codec [1],[2] optimised towards memory [3]. As video coding/decoding are data dominated systems, efficient memory management and data transfer and storage optimisation directly impact their resource requirements. The performances are measured as coding speeds on a Pentium III running at 700 MHz with Windows NT and are expressed in relative times, with the duration of the sequence in seconds as reference. Meeting real-time constraints means having a relative coding time smaller than one.

The close relation between the memory load and the speed performance has been evaluated for test cases with different coding parameters in [2], justify assessing the complexity of the system by measuring the relative coding/decoding time fluctuations on a specific platform. The reported resource scaling is expected to be representative for any platform.

3.1 *SNR scaling*

As for the bandwidth, the complexity decrease is exponential with QP and increases with the frame rate (Figure 3.1). The nature of the sequence does not really impact it. Only at very small QP (high quality), a faster decrease of the decoding complexity for simple and stationary video occurs. For those sequences, the increase of QP results in lots of zero texture blocks, free of IDCT computation.

The spatial resolution does not affect the relative decrease of the decoder complexity. However, on the encoder side, the impact of QP on the total QCIF encoder complexity is limited.

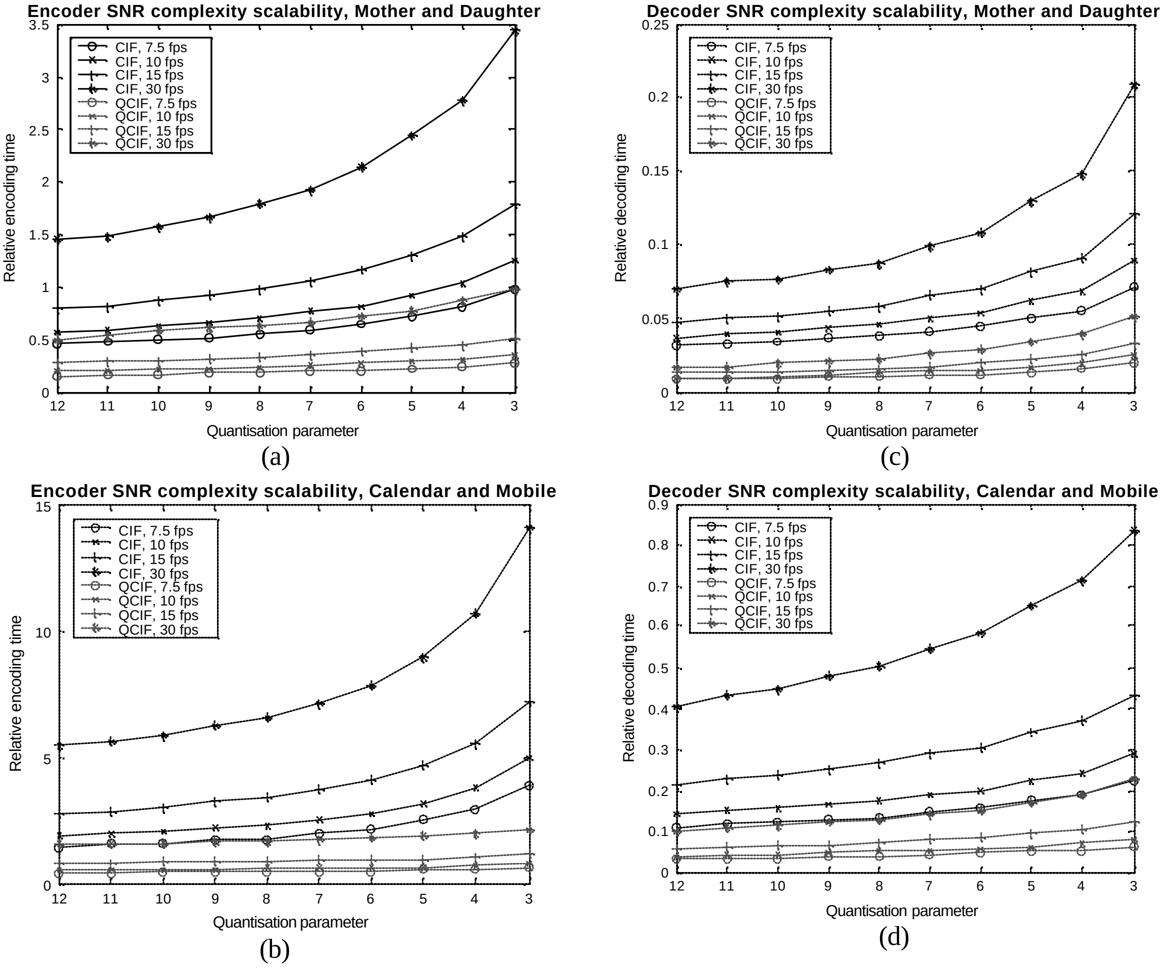

Figure 3.1: Complexity scaling by QP increasing for the encoder (a,b) and the decoder (c,d) has an exponential impact. The results are plotted for the *Mother and Daughter* (a,c) and *Calendar and Mobile* (b,d) video sequences.

3.2 Temporal complexity scaling

The encoder complexity has an almost linear increase with the frame rate (Figure 3.2). Consequently, encoding two independent streams at half the reference frame rate requires about the same resources as encoding a single stream at the reference frame rate[2]. This is further discussed in Section 5.2.

For simpler video sequences, the complexity scaling of the decoder is not linear. The decoder optimisations restrict the number of the texture reconstructions (inverse quantisations and IDCTs) to the non-zero texture blocks. This tempers the decoder complexity increase at higher frame rates (smaller slope) for simple sequences.

In contrast to the bandwidth case, the QCIF complexity of both encoder and decoder is far below the one for CIF. The distance between QCIF and CIF complexity is larger for sequences with a high degree of motion. The temporal scaling is independent of the resolution and the quantisation degree.

[2] In practice, it requires appropriate memory management as two reference frames have to be dealt with.

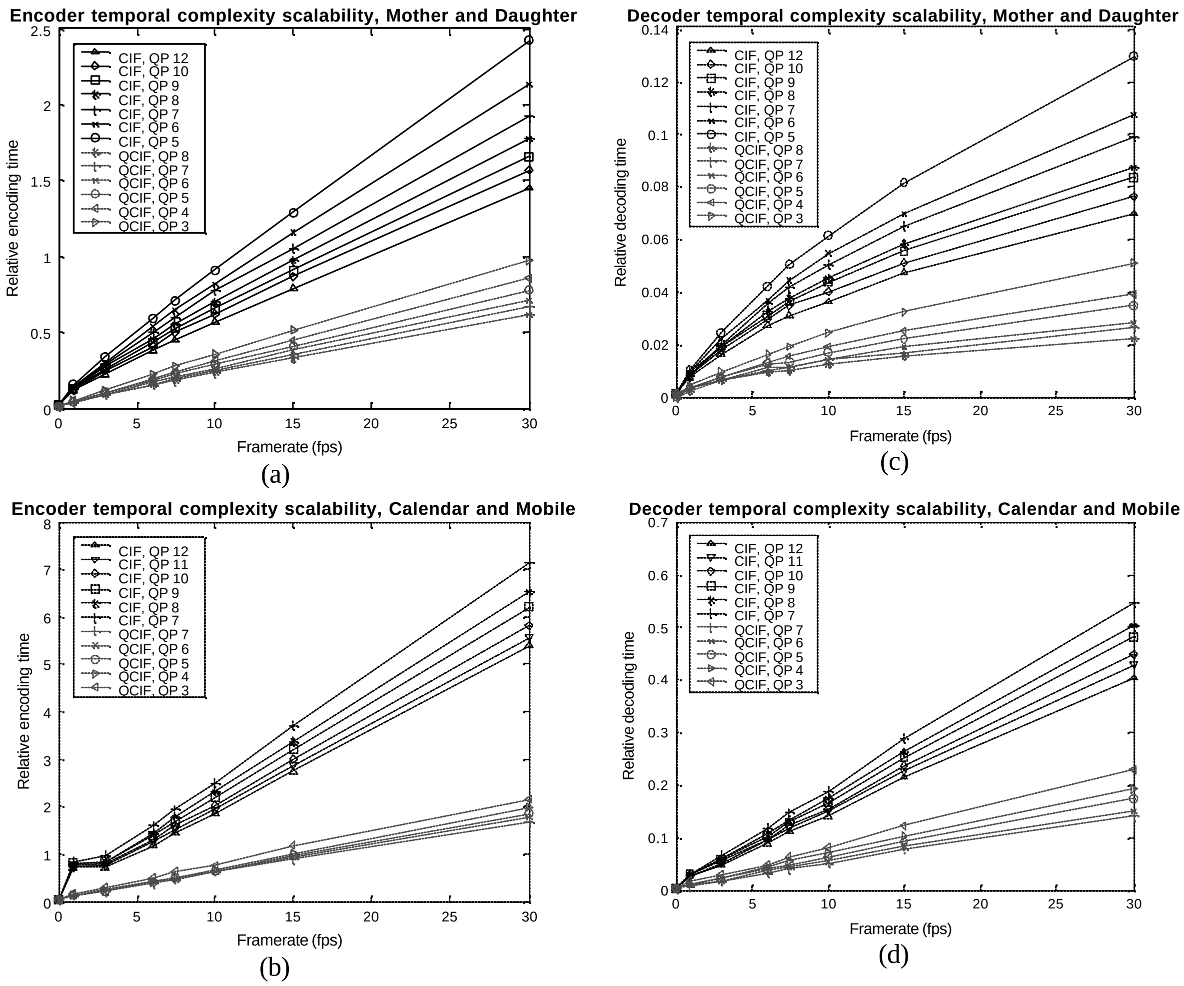

Figure 3.2: Complexity scaling by lowering the frame rate for the encoder (a,b) and the decoder (c,d) has a quasi linear impact. The results are plotted for the *Mother and Daughter* (a,c) and *Calendar and Mobile* (b,d) video sequences.

3.3 Spatial Complexity Scaling

Switching from CIF to QCIF resolution requires approximately four times fewer resources to decode video, independent of the frame rate and the complexity of the sequence (Table 2 to Table 3). The factor is directly proportional to the number of processed blocks.

The relative block type distribution (each type needs a specific amount of operations) is respected when changing the resolution. At the encoder side, next to the number blocks to be encoded, the smaller search are of the ME [1] at QCIF size allows the complexity reduction to slightly augment spatial ratio (four), especially at low QP (Table 3).

Table 2: Spatial total encoder and decoder complexity reduction factors Mother and Daughter

fps \ QP	3.00	6.00	7.50	10.00	15.00	30.00
12	3.3	3.1	3.1	3.0	3.0	2.9
11	3.3	3.1	3.1	3.0	2.9	2.8
10	3.3	3.1	3.1	3.0	2.9	2.7
9	3.3	3.2	3.1	3.0	2.9	2.8
8	3.2	3.2	3.1	3.0	3.0	2.9
7	3.3	3.2	3.1	3.2	3.0	2.9
6	3.4	3.2	3.2	3.1	3.1	3.0
5	3.5	3.3	3.3	3.2	3.2	3.1
4	3.6	3.4	3.4	3.3	3.3	3.2
3	3.7	3.6	3.6	3.6	3.6	3.5

fps \ QP	3.00	6.00	7.50	10.00	15.00	30.00
12	3.2	4.5	3.9	4.0	3.9	4.3
11	4.3	4.0	4.0	4.3	4.1	4.3
10	3.4	4.2	3.8	4.0	3.9	4.0
9	3.8	3.9	3.6	3.9	3.9	4.2
8	3.2	3.6	3.7	3.7	3.9	3.9
7	3.3	3.5	3.7	3.6	3.8	3.8
6	3.7	3.3	4.0	3.9	3.6	3.8
5	3.4	3.5	3.8	3.6	3.7	3.8
4	3.9	3.5	3.7	3.6	3.6	3.8
3	3.7	3.7	3.7	3.7	3.8	4.1

Table 3: Spatial total encoder and decoder complexity reduction factors Calendar and Mobile

fps \ QP	3.00	6.00	7.50	10.00	15.00	30.00
12	4.1	3.7	3.7	3.7	3.6	3.5
11	4.2	3.9	3.8	3.7	3.7	3.7
10	4.3	4.0	3.9	3.9	3.7	3.7
9	4.4	4.0	4.0	4.0	3.9	3.9
8	4.4	4.2	4.1	4.1	4.0	3.9
7	4.8	4.4	4.3	4.1	4.2	4.2
6	5.0	4.2	4.6	4.5	4.5	4.5
5	5.6	5.2	5.1	5.0	4.9	4.9
4	6.2	5.8	5.8	5.6	5.6	5.4
3	6.9	6.5	6.5	6.5	6.3	6.6

QP \ fps	3.00	6.00	7.50	10.00	15.00	30.00
12	3.9	3.8	3.6	3.8	3.9	4.1
11	3.6	3.7	3.6	3.7	3.8	4.1
10	3.7	3.7	3.7	3.7	3.8	4.0
9	3.5	3.5	3.6	3.7	3.8	4.0
8	3.9	3.6	3.6	3.6	3.7	3.9
7	3.7	3.6	3.6	3.7	3.7	3.9
6	3.7	3.6	3.6	3.7	3.7	3.9
5	3.8	3.7	3.6	3.7	3.7	3.8
4	3.7	3.6	3.6	3.6	3.6	3.7
3	3.6	3.7	3.6	3.7	3.6	3.7

4 Point-to-point, one-to-one connection

On a point-to-point connection, the sender can know its own and the receivers' computational capabilities and can asses tolerated transmission bitrates, even on packet switched networks (e.g. TCP like protocols on the Internet) [4],[5]. Elaborated scalable and error-resilient source coding schemes have been proposed to cope with the instantaneously and randomly changing network conditions [6]. Our study however, restricted to standardised hybrid MC/DCT video codecs, considers the impact of a complete set of coding parameters on both bandwidth and complexity. The goal of this paper is not to achieve fine rate-control or error-resilience regulation but rather understanding how to select the optimal codec general settings. Conventional rate control algorithms and even fine-grained scalable encoding methods [7] can then be used for fine parameter adaptation around the coarse set of coding parameters, while adaptive channel coding methods (e.g. adaptive forward error correcting codes) further tune the application bandwidth and error-resilience to the instantaneous channel conditions.

4.1 Bitrate Reduction

A logical and well-known way to reduce the number of bits required to encode a sequence is to increase the quantisation parameter. At a certain point, further QP increase becomes inadequate. Figure 2.1.c,d show plots where a CIF sequence with QP x requires the same bandwidth as QCIF with QP y ($x > y$). At some point, encoding the scene in QCIF format, i.e. at a lower spatial resolution, with a smaller QP provides a better quality than further increasing QP at CIF resolution. Observations indicate a QP of 10 as practical upper bound for the CIF format. At that point, it is good to shift to a QCIF encoding scheme and halve the QP. Decreasing the QP again at QCIF size allows the continuation of the bitrate drop.

Table 4: The bitrate reduction (using SNR, spatial and temporal scalability; QP 5 to 10; frame rate factor of 3) varies from 15.8 to 24.3 and increases with the complexity of the sequence.

Scalability type	Mother and Daughter (kbps)	Reduction Factor	Foreman (kbps)	Reduction Factor	Calendar and Mobile (kbps)	Reduction Factor
	279.3		1004.3		3948.3	
QP (5 to 10)	97.9	2.9	374.8	2.7	1610.7	2.5
Spatial (1/4)	31.3	3.1	106.8	3.5	418.1	3.9
Subtotal		*8.9*		*9.4*		*9.4*
Temporal (1/3)	17.6	1.8	56.7	1.9	162.2	2.6
Total		**15.8**		**17.7**		**24.3**

Eventually, temporal scaling can be used to achieve even lower bitrates. Table 4 lists the reduction by applying the different scalability types in the suggested order. The total cutback increases with the complexity of the sequence. Typically, an increment of QP from 5 to 10 gives a 2.5 decrease factor. Going from CIF to QCIF yields a factor 3 and reducing the frame rate to one third provides a factor 2. In total, a factor 15 is within reach with acceptable quality loss.

A standardised hybrid MC/DCT codec can efficiently and easily provide scalability and error robustness through the generation of multiple independent bitstreams at different frame rates (see Section 5). This motivates to exploit the temporal resolution parameter after the other coding parameters have been adjusted. Typically, respectively halving, taking one third and one fourth of the original frame rate can yield a bandwidth decrease factor of 1.5, 2 and 2.5 (see Figure 2.1.c,d).

4.2 Complexity Reduction

In analogy, changing the coding parameters affects the complexity of both the encoder and decoder. Table 5 and Table 6 list the reduction factors when adapting them in the order proposed in the previous subsection. Typically, incrementing the QP from 5 to 10 gives a decrease factor of 1.5 for encoder and decoder. Going from CIF to QCIF yields a factor 3 (4) and reducing the frame to one third provides a factor 2.5 (2) for the encoder (and decoder respectively). In total, a factor 11 (12) is within reach.

As discussed in Section 4.1, changing the coding format (CIF/QCIF) while adjusting the QP appropriately allows for graceful and smooth bandwidth adaptation. Figure 3.1 and Figure 3.2 show this is not the case for complexity: changing the coding format always results in a significant step in complexity evolution (two well separated graph sets). Consequently, too frequent format shiftings should be avoided in a practical implementation. Accepting larger variations of QP or of the frame rate until the format-shifting threshold is regularly passed prevents oscillations.

Table 5: The encoder complexity reduction (using SNR, spatial and temporal scalability; QP 5 to 10; frame rate factor of 3) varies from 11.9 to 17.0 and increases with the complexity of the sequence.

Scalability type	Mother and Daughter	Reduction Factor	Foreman (kbps)	Reduction Factor	Calendar and Mobile (kbps)	Reduction Factor
	2.4		3.8		8.9	
QP (5 to 10)	1.6	1.5	2.6	1.5	5.8	1.5
Spatial (1/4)	0.6	2.7	0.8	3.4	1.6	3.7
Subtotal		*4.2*		*5.0*		*5.7*
Temporal (1/3)	0.2	2.8	0.3	2.6	0.5	3.0
Total		**11.9**		**13.0**		**17.0**

Table 6: The decoder complexity reduction (using SNR, spatial and temporal scalability; QP 5 to 10; frame rate factor of 3) varies from 12.9 to 15.5 and increases with the complexity of the sequence.

Scalability type	Mother and Daughter (kbps)	Reduction Factor	Foreman (kbps)	Reduction Factor	Calendar and Mobile (kbps)	Reduction Factor
	0.13		0.31		0.65	
QP (5 to 10)	0.08	1.7	0.20	1.5	0.44	1.5
Spatial (1/4)	0.02	4.0	0.05	4.0	0.11	4.0
Subtotal		*6.8*		*6.0*		*5.9*
Temporal (1/3)	0.01	1.9	0.02	2.3	0.04	2.6
Total		**12.9**		**13.7**		**15.5**

4.3 Bandwidth Complexity Relation

In a real life application, like a virtual meeting set-up, both the available bandwidth and the resources are limited and vary over time. Figure 4.1 plots the required bitrate as function of the consumed processing time by varying QP (*Calendar and Mobile* sequence). Conventional rate control schemes try to follow the bandwidth constraint by playing on QP, the frame rate and spatial resolution being fixed. The operating point of these codecs moves along a particular curve in Figure 4.1. As can be concluded from the steep slopes of these curves, such QP adaptation poorly impacts the encoding or decoding complexity. To really enable complexity tuning, the operating point should be permitted to shift from one curve to another, i.e. to adapt the spatial resolution and the frame rate parameter. Varying all three coding parameters results in a denser and more uniform coverage of the bandwidth/complexity space. Figure 4.2 plots the bitrate as a function of complexity by varying the frame rate for *Calendar and Mobile* sequence. Changing the frame rate has greater impact on the complexity then modifying the QP.

In practice, adding a quality measure as a third dimension to these graphs allows choosing the best solution, i.e. the one providing the lowest distortion, among a set of acceptable operating points (e.g complexity and bandwidth below some threshold). The surface connecting these points would make it a 3D Pareto plot [9], [10]. An interesting research topic would be to incorporate channel-coding performances into the quality measurement in order to optimise the end-to-end quality and not just the quality on the encoder side. In a more advanced and complete design, the dynamics of the bitrate and complexity restrictions have to be taken into account. This requires navigating along the optimal trade-offs. Moreover, just as for conventional rate control algorithms, possible operating points can not be exhaustively investigated. Appropriate models are required to estimate the optimal trade-off. This is another important subject open to future research.

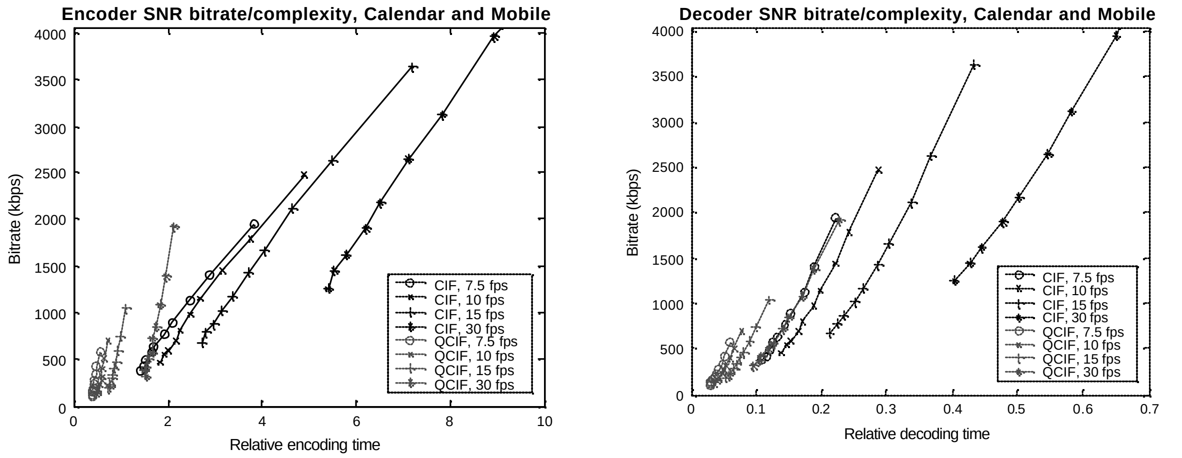

Figure 4.1: Bandwidth complexity relation using SNR scaling.

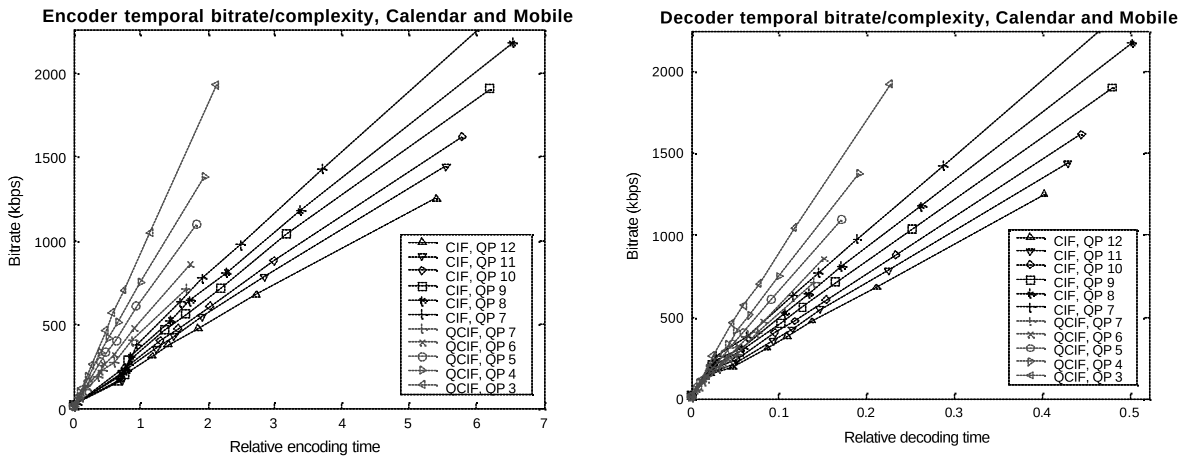

Figure 4.2: Bandwidth complexity relation using temporal scaling.

5 Multicast or broadcast, one-to-many connections: How to cope with heterogeneous receivers

This section focuses on coarse grain scalability as it assumes that users are only affected by significant changes in quality. A close match between the user's available resources and the processing requirements of the stream is not desired. This hypothesis is motivated by the increasing popularity of video communications where a participant receives multiple streams from different sources at the same time (e.g. distributed conferencing system, watching simultaneous football games, etc.). In that case, ensuring co-operation of the senders to target the requirements of users is more complex than allowing a receiver to choose between a few coarse operation points for each active connection. This second receiver-driven alternative appears a more simple and appealing solution: the user shares its resources between active connections according to its needs and capabilities.

When increased flexibility is required, fine-grained scalability (FGS) can be built on top of a coarse grained system. Moreover, FGS performances largely depend on the quality of the reference base layer [7]. Until sufficient quality has been obtained for the base layer, FGS provides worse coding efficiency than the non-scalable simple hybrid MC/DCT scheme. A system providing a number of coarse granular streams and using FGS when a sufficient good quality has been achieved for the reference stream is perfectly imaginable.

5.1 SNR and Spatial Scalability

As noted in section 2.1, the bandwidth drops rapidly when QP increases, especially at high frame rates. Moreover, Section 2.3 shows the QCIF bandwidth is less than thirty percent of CIF bandwidth (with the other coding parameters fixed). When considering coarse SNR or spatial scalability, only few, very distinct coding points have to be simultaneously provided to the set of receivers. In that case, the bandwidth overload (or cost) due to the simultaneous transmission of low quality or small resolution streams is limited (below 30%). This raises questions to the value of layered codecs in this specific case. Table 7 provides the bandwidth cost of simulcast compared to the MPEG-4 simple layered scalable solution. As expected, the bandwidth benefit of layering appears to be small. In [8], these results are confirmed and even strengthened for a large range of bit-rates and frame-rates. There, the efficiency of the H.263 SNR scalable codec has been measured and compared to simulcast. Rather surprisingly, the layered scheme appears to spend even more bits than the simulcast version when coding points correspond to largely different qualities, i.e. for coarse scalability.

Table 7: Bandwidth benefit of layering for MPEG-4 SNR scalability.

	Frame rate (fps)	Base layer quality	High layer quality	Bitrate low quality (kbps)	Bitrate high quality (kbps)	Bitrate two layers (kbps)	Bandwidth cost simulcast
Silent	10	31 dB	34 dB	25	48	69	6 %
Coastguard	10	30 dB	32 dB	56	96	141	8 %
Foreman	10	31 dB	33 dB	49	78	107	19 %

From the complexity point of view, layering is worse than simulcast. Without loss of generality, consider a two-layer scenario. At the encoder side, the enhancement layers of scalable schemes (simple scalable MPEG-4 and H.263) consider two references to predict each macroblock (MB), one is the base layer co-located MB, the other is a search window in the previous enhancement layer frame. Typically, this increases the reference memory size and transfer rate. At the decoder side, decoding the layered high quality stream is roughly as complex as decoding both non-layered streams. Figure 3.1 shows that the additional cost of decoding the lower layer is most of the time significant.

5.2 Temporal Scalability

Obviously, simulcast (e.g. one stream at 15 fps and one stream at 30 fps) makes no sense for temporal scalability. It spoils bandwidth to redundant information and does not provide a significant complexity gain on the decoder side. Two streams at half the highest frame-rate (e.g. two streams at 15 fps) are more attractive, both for bandwidth and resources.

Figure 3.2 indicates the encoding effort for a 30 fps stream is nearly the same as for two at 15 fps (curves are almost linear and originated in the origin). From the decoder point of view, the trend is similar, especially for high complexity sequences. Indeed, processing two independent streams doubles the reference frame memories, but these memories are never active simultaneously. The same frame memory is accessed for all the MBs of a given frame. An intelligent implementation shifts from one reference to the other each time a new frame is encoded/decoded. Consequently, the number and cost of data transfers is not strongly affected.

Figure 2.1.c,d give a measure of the cost in terms of bandwidth. Again, especially at high frame rates and for high complexity sequences, the cost in bandwidth is low, which makes the solution attractive. Note that the temporal layered alternative corresponds to the use of B-frames. The use of B-frames improves the coding efficiency (especially for simple sequences with high temporal correlation) but it requires simultaneous access to both reference frame memories for each MB, which significantly increases the complexity. Moreover, as other layered scalable codecs, B-frames are more sensitive to errors (see next subsection).

5.3 Scalability and Unreliable Networks

Over best-effort packet networks such as the Internet, video communication is hampered by limited bandwidth and packet loss. Layered approaches, either SNR, spatial or temporal, prioritise data and are able to better protect crucial information. However, when an error or a loss occurs in the base layer, the rest of the information becomes useless. To cope with this problem, Multiple Description Coding has been proposed [11] to describe a signal into multiple bitstreams such that any bitstream can be used to reconstruct a baseline signal, and any additional stream improves the reconstruction. The independence of the streams, retains the accurateness of one stream if the other is corrupted. Moreover, by correlating these streams, the correct stream(s) can help concealing erroneous one(s). As recently highlighted in the literature [12],[13], simple standardised hybrid video codecs allow for multiple description (MD) coding. In [13], the input video is partitioned into two subsets of frames (even and odd) which are encoded into independent streams, both at half the frame rate. This approach could be generalized in the spatial domain by sub-sampling a CIF format sequence to generate four QCIF sequences. The cost in bandwidth and complexity of such an approach has been discussed in Section 5.1 and 5.2, using Figure 2.1 to Figure 3.2. The benefit in terms of error robustness depends on the envisioned application and is beyond the scope of this paper. However, in the conferencing application envisioned in the introduction, two protagonists have a central position and become natural relays for other users. In that case, a robust solution consists in the transmission of independent streams to the central relays. Each relay receives one stream and broadcast or multicast it to other users. The resulting path diversity is expected to increase robustness and is another argument to choose multiple independent streams rather than B-frames.

6 Conclusions and Open Questions

Multi-user video conferencing applications have high bandwidth and resource requirements. The diversity of participants' multimedia equipment and connection properties imposes a large variation to the operating constraints. We have analysed the scaling behaviour of both bandwidth and complexity with the encoding parameters of a standardised hybrid MC/DCT video codec by varying the quantisation parameter (SNR scaling), the frame rate (temporal scaling) and the size (spatial scaling).

By presenting possible operating points in a complexity versus bandwidth graph, a valid set can be selected when bandwidth and computational resources are specified. The combination of the three scaling effects allows a reduction of at least one order of magnitude for bandwidth and complexity constraints. A natural and graceful way to control the rate and the performance of the codec is obtained by first varying the QP, then reducing the frame size (with appropriate QP resetting) and finally decreasing the temporal resolution. The study also indicates the need for a multi-parameter rate-control to allow for significant complexity scaling.

In real-life applications, coarse scalability is required to overcome the large differences in possible bandwidths. For those situations, layering is not always a good alternative when considering both bitrate and complexity issues. The simulcast of hybrid MC/DCT streams provides a simple but efficient coarse grained scalability. Moreover, in the temporal case, sending two independent, intertwined video streams, is a promising alternative to deal with the network unreliability.

Acknowledgement

This work was realised in the IMEC MPEG-4 IIAP, partially funded by Xilinx Labs, Xilinx Inc. R&D group.

References

[1] C. De Vleeschouwer and T. Nilsson, "Motion estimation for low power video devices", Proceedings of the IEEE International Conference on Image Processing (ICIP01), Thessaloniki, Greece, October 2001, pp. 953-957.

[2] K. Denolf, et al., "Cost-efficient C-level design of an MPEG-4 video decoder", Proceedings of the IEEE Workshop on Power and Timing Modeling, Optimization and Simulation, pp.233-242, Goettingen, Germany, September 2000.

[3] F. Catthoor, et al., "Custom Memory Management Methodology", ISBN 0-7923-8288-9, Kluwer Academic Pub., 1998.

[4] S. Floyd, et al., "Equation-based congestion control for unicast applications", SIGCOMM 2000, August 2000.
[5] F. Toutain, "TCP-friendly point-to-point video-like source rate control", PacketVideo99, New York, April 1999.
[6] W.-T. Tan and A. Zakhor, "Real-time Internet video using error resilient scalable compression and TCP-friendly transport protocol", IEEE Trans. On Multimedia, vol. 1, No. 2, June 1999, pp. 172-186.
[7] H.M. Radha, M. van der Schaar, and Y. Chen, "The MPEG-4 fine-grained scalable video coding method for multimedia streaming over IP", IEEE Trans. on Multimedia, vol. 3, No. 1, March 2001.
[8] M. Walker and M. Nilsson, "A study of the efficiency of layered video coding using H.263", PacketVideo99, New York, April 1999.
[9] V. Pareto, "Cours D'Economie Politique", Volume I-II, Lausanne, 1896.
[10] H. Eschenauer, et al., "Multicriteria Design Optimization: Procedures and Applications", Berlin, Springer-Verlag, 1990.
[11] A. Reibman, et al., "Multiple description coding of video using motion compensated prediction ", ICIP99, Kobe, Japan.
[12] J.G. Apostolopoulos and S.J. Wee, "Unbalanced multiple description video communication using path diversity", ICIP01, Thessaloniki, Greece, October 2001, pp. 966-969.
[13] M. Gallant, S. Shirani, and F. Kossentini, "Standard-compliant multiple description video coding", ICIP01, Thessaloniki, Greece, October 2001, pp. 946-949.