# NETWORK AND DEVICE DRIVEN MOTION-COMPENSATED SCALABLE VIDEO FOR WIRELESS SYSTEMS

M. van der Schaar* and H. Radha*&

*Philips Research USA &Michigan State University

## ABSTRACT

Transmission of video over wireless and mobile networks requires a scalable solution that is capable of adapting to the varying channel conditions in real-time (bit-rate scalability). Furthermore, video content needs to be coded in a scalable fashion to match the capabilities of a variety of devices (complexity scalability). These two properties – bit-rate and complexity scalability – provide the flexibility that is necessary to satisfy the "Anywhere, Anytime and Anyone" network paradigm of wireless systems. Meanwhile, the MPEG-4 Fine-Granular-Scalability (FGS) has been introduced as a flexible low-complexity solution for video streaming over heterogeneous networks (e.g., the Internet and wireless networks). FGS is also highly resilient to packet losses. However, the flexibility and packet-loss resilience associated with the FGS framework come at the expense of decreased coding efficiency compared with non-scalable coding under lossless conditions. In this paper, a novel scalable video-coding framework and corresponding compression methods for wireless video streaming is introduced. Building upon the FGS approach, the proposed framework, which we refer to as Adaptive Motion-Compensation FGS (AMC-FGS), provides improved video quality of up to 2 dB. Furthermore, the new scalability structures provide the FGS framework with the flexibility to provide trade-offs between resilience, higher coding efficiency and terminal complexity for more efficient wireless transmission.

## 1. INTRODUCTION

Real-time streaming of audiovisual content over wireless networks is becoming an important technology area in multimedia communications (see for example [1] [2]). In general, emerging wireless systems can be divided into two categories:

- Cellular mobile networks such 2.5G, 3G, and 4G systems;
- Wireless LANs including IEEE 802.11, HiperLAN 2, Bluetooth etc.

Both of the aforementioned wireless systems are characterized by a large data rates range (from several kbps up to tens of Mbps). Furthermore, the prevailing channel conditions, the mobility of the terminal, the transmission conditions external to the wireless network (for instance, congestion on the Internet) will have a major impact on the efficacy of future wireless multimedia networks. Meanwhile, an important concept for wireless systems is the Universal Multimedia Access (UMA) that refers to the way in which multimedia data can be accessed by a large number of users/clients to view any desired video stream anytime and from anywhere. In the UMA framework, multimedia information is accessed from the network depending on the following three parameters: user preference, channel characteristics and device capabilities. Thus, video content needs to be coded in a scalable fashion to match the capabilities of a variety of devices (complexity scalability) and for fast and easy adaptability to a particular bandwidth (bit-rate scalability). This will provide the flexibility and mobility that are necessary to satisfy the "Anywhere, Anytime and Anyone" network paradigm of wireless systems.

Recently, several scalable coding methods have been successfully proposed for video transmission through heterogeneous networks (see for example [1]- [8]). One of these techniques is the MPEG-4 Fine-Granular Scalability (FGS) scheme [6][7][8], that can adapt in real-time (i.e., at transmission time) to the bandwidth variations over heterogeneous networks and to the terminal capabilities, while using the same pre-encoded stream. Some key advantages of the MPEG-4 FGS framework are its packet-loss resilience and flexibility in supporting streaming applications [9]. Naturally, these properties come, in general, at the expense of video quality. In [8], the FGS performance was compared with that of non-scalable streams coded at discrete bit-rates covering the same bandwidth range. The results obtained indicated around 2- 3 dB reduction in video quality (when comparing FGS with non-scalable) for certain sequences that exhibit high temporal correlations among successive frames. This is mainly due to the lack of motion compensation at the FGS enhancement layer. Hence, for sequences that have a high-degree of motion and large number of scene cuts (e.g., "MTV" like sequences or high-action scenes), FGS performance is comparable to the performance of non-scalable coding. Moreover, in [8], FGS has also been compared with "traditional" SNR coding with multiple layers (without motion-compensation within the enhancement-layer) [13] as employed in MPEG-2 or MPEG-4, and the results indicated that FGS outperforms multi-layer (discrete) SNR coding with several dBs over a wide range of bit-rates due to its adaptive and effective bitplane coding technique and the lack of overhead associated with introducing a new enhancement-layer.

To (a) improve the performance of the MPEG-4 FGS coding method compared with MPEG-4 non-scalable coding and (b) address the bandwidth- and complexity-scalability requirements of wireless systems, there is a need for an *adaptive* solution that addresses both the quality and scalability issues in a joint manner. In this paper, a scalable video-coding framework and a corresponding compression method for wireless video streaming is presented. Building upon the FGS approach, the proposed framework provides improved coding efficiency of up to 2 dB. This improvement is based on incorporating some prediction within the original FGS structure in an adaptive manner. Our proposed solution is based on new FGS-based scalability structures. These new structures provide the FGS framework with the flexibility to provide easy trade-offs between higher coding efficiency, bandwidth scalability, and terminal complexity for more efficient wireless transmission.

Note that several mechanisms for improving "traditional" SNR scalability in predictive coding by exploiting enhancement-layer information have already been proposed. For instance, in [14] the current enhancement-layer frame is predicted from the motion-compensated reconstruction of the previous enhancement-layer frames. However, in this case, the enhancement-layer does not exploit the information in the current base-layer residual. An improved technique is proposed in [15] that effectively switches between transform coefficients from the reconstructed base-layer and the predicted enhancement-layer frame for the prediction of the next enhancement-layer frame, leading to an improvement of up to several dBs. However, the disadvantage of this technique is that it does not enable fine-granular adaptation to the bandwidth variations without introducing a relatively high overhead for each additional enhancement-layer as can be seen from the results in [15]. Nevertheless, the drift analysis provided in [15] is very useful and can be directly applied in the proposed scheme. Several alternative techniques have also been proposed to exploit further temporal redundancies within the FGS framework [11][11]. One such technique is the Progressive FGS (PFGS) method introduced in [11]. PFGS employs additional motion-compensation loop(s) for the P and B enhancement-layer frames in order to improve the performance of the FGS framework. Our solution differs from PFGS in two ways. First, we present two simplified scalability structures that are suitable for low-complexity wireless devices, and second, we propose an adaptive framework that switches between the original FGS and the newly proposed structures based on the sequence characteristics, channel conditions and/or allowed device complexity. Although one of our scalability structures can be considered as a simplified version of the PFGS method proposed in [11], we believe that the combination of complexity-scalability/complexity-reduction with quality improvement, which characterizes our proposed solution, is more flexible for new and emerging wireless and mobile networks. Hence, the novelty of our solution resides mainly in its adaptability to the sequence characteristics, channel conditions and/or allowed device complexity. The scalability structures proposed here can considerably improve the performance of the FGS framework while preserving most of the flexibility and attractive characteristics typical to the "basic" FGS scheme. We refer to our proposed scalability structures as Motion-Compensation FGS (MC-FGS). Moreover, our MC-FGS based structures can serve two classes of wireless receivers. One is tailored for relatively powerful devices (e.g., laptops) connected to wireless LANs. The second MC-FGS solution is for a lower complexity solution that is more suitable for "thin" devices. Each of these solutions has their own advantages and disadvantages as described below.

For the remainder of this paper, we assume that our proposed adaptive solution can be supported by (a) an FGS-based encoder, which has access to some information regarding the bandwidth variation and complexity of devices served by a particular wireless network (e.g., a wireless LAN), or (b) a server residing at the interface between the wired and wireless segment of an end-to-end streaming service and which has access to a "basic" FGS stream. Consequently, our proposed solution could be realized by modifying a standard FGS stream at the "interface" server. This server can either real-time stream or store the modified (still scalable) stream locally for further viewing by the different devices that are capable of accessing the server over the wireless network.

The remainder of the paper is organized as follows. In Section 2, the MC-FGS scalability structures are described with the corresponding encoder algorithms and architectures. We also describe a high-level (heuristic) adaptive MC-FGS based algorithm that employs the MC-FGS scalability structures. In Section 3, simulation results for the improvements in video quality obtained with the newly proposed algorithms are presented. In Section 4, we outline the conclusions and discuss the suitability of adaptive MC-FGS for wireless networks. Finally, in Appendix I, an analysis of the FGS coding penalty associated with the high flexibility in adapting to the bandwidth variations is provided.

## 2. MOTION-COMPENSATION FGS (MC-FGS) STRUCTURE

In this section, two extensions to the "basic" FGS scheme are presented (see Fig. 1) that introduce motion-compensation within the FGS enhancement-layer.

### 2.1 Two-loop MC-FGS for B-frames

***Algorithmic description.*** The proposed MC-FGS framework portrayed in Fig. 1a introduces a motion-compensation (MC) loop within the FGS enhancement-layer to exploit the remaining temporal correlation within this layer. The MC within the FGS-layer re-uses the base-layer motion-vectors, while the resulting enhancement-layer residual is compressed using the same embedded codec as in the "basic" FGS scheme. As illustrated in Fig. 1a, not all FGS bitplanes (up to $R_{max}$) are part of the enhancement-layer MC-loop. The number of bitplanes, M, included in the FGS MC-loop is chosen by trading off

- coding gain (i.e. a large M value potentially leads to an improved temporal decorrelation within the FGS-layer), and

- prediction drift[1] occurring within the FGS-layer at low bit-rates, when fewer bitplanes than M are transmitted/received, (i.e. a small value of M limits prediction drift).

Hence, to limit the drift incurred at low bit-rates, preferably only a few FGS bitplanes are included in the enhancement-layer MC-loop, e.g. M=2-3 bitplanes. (For an improved performance, a different value of M can be chosen for each frame.) The MC-prediction within the FGS-layer is restricted to B-frames because a relatively high coding gain is obtained while preserving many of the "basic" FGS structure benefits. For instance, under this framework, if a number of bitplanes lower than M is transmitted for the I and/or P enhancement-layer frames, the prediction drift is confined to the B enhancement-frames.

Furthermore, this MC-FGS structure achieves a higher coding gain (see Section 3) than the "basic" FGS scheme due to its superior temporal decorrelation for the B enhancement-frames. Since B-frames account for 66% of the total enhancement-layer bit-rate[2] in an IBBP GOP structure, the loss in quality associated with restricting the enhancement-layer MC to the B-frames is limited for most sequences. Another reason why eliminating the MC between the P enhancement-frames has only a limited effect on coding efficiency, resides in the less accurate MC prediction of the P enhancement-frames, which have a larger distance to their reference frames than the B-frames. Hence, we will include in the two-loop MC-FGS only the B enhancement-frames.

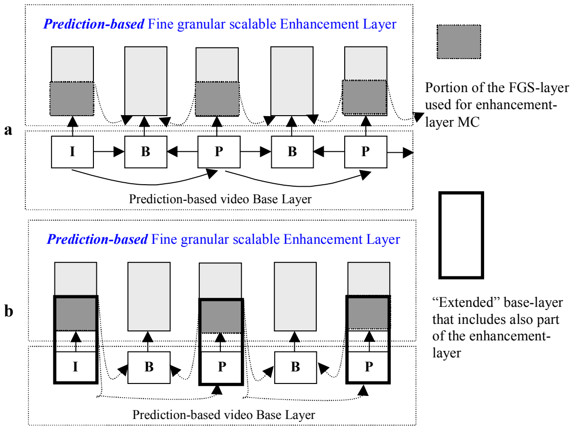

*Fig. 1. MC-FGS scalability structures: two-loop MC-FGS (a) and one-loop MC-FGS (b).*

Since the two-loop MC-FGS is restricted to the B enhancement-frames, a packet-loss occurring in an I or P enhancement-frame will not propagate beyond the enhancement-layer B-frames using this frame as prediction reference. Thus, subsequent enhancement-layer P-frames within the GOP remain unaffected. This represents one of the advantages of this simplified MC-FGS structure when compared with a more elaborate scheme such as the one proposed in [11].

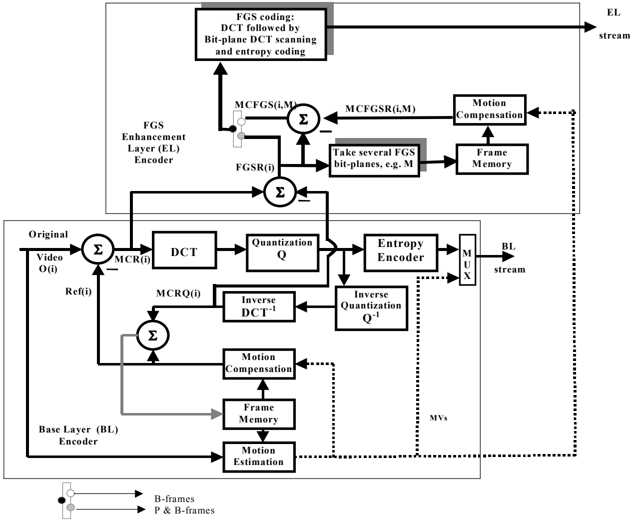

*Fig. 2. Block-diagram of two-loop MC-FGS encoder.*

---

[1] Prediction drift occurs within the FGS enhancement-layer if K<M bitplanes are transmitted/received because the references used for motion-compensation prediction within the FGS-layer are different at the encoder and decoder: at the encoder side prediction is based on M-bitplanes of the reference, while at the decoder it is based on only K-bitplanes.

[2] In the "basic" FGS framework, a simple, yet efficient rate-control can be performed by allocating an equal number of bits to all enhancement-layer frames (I/P/B), since there is no motion-compensation in this layer [7].

***Complexity discussion.*** The encoder block-diagram of the two-loop MC-FGS is depicted in Fig. 2. From a complexity perspective, it is important to notice that the base-layer structure remains unchanged. For the I and P-frames, there is no MC within the enhancement-layer and hence, for these frames, the complexity is the same as for the "basic" FGS framework. An additional enhancement-layer motion-compensation loop is introduced for the B-frames, at both the encoder and decoder. For the enhancement-layer B-frames, the signal coded in the FGS-layer of the $i^{th}$-frame equals:

MCFGS(i,M) = FGSR(i) – MCFGSR(i,M) =

= MCR(i)-MCRQ(i) – MCFGSR(i,M), where

*FGSR(i)* is equal to MCR(i)-MCRQ(i), as in the "basic" FGS,

*MCFGSR(i,M)* is the $i^{th}$-frame MC-prediction based on the M most significant bitplanes of the FGS-layer reference frame(s),

*MCR(i)* is the MC-residual of the $i^{th}$-frame,

*MCRQ(i)* is the reconstructed motion-compensation residual of the $i^{th}$-frame (after quantization and dequantization).

The enhancement-layer MC loop for the B-frames re-uses the base-layer motion-vectors and prediction modes. Furthermore, the additional complexity associated with the motion-compensation within the enhancement-layer is lower than for the base-layer decoding, since the motion-vectors are already decoded by the base-layer and the motion-compensation is reduced to fetching the reference data and adding the residual data.

The proposed two-loop MC-FGS is less complex than the PFGS-scheme in [11], since in MC-FGS the motion-compensation is inserted within the enhancement-layer only for the B-frames. Furthermore, if the decoder does not have enough processing power to decode the additional enhancement-layer motion-compensation loop, with MC-FGS, the enhancement-layer B-frames or even the entire B-frames can be discarded without affecting the consecutive frames. Hence, the proposed two-loop MC-FGS allows for graceful complexity scalability besides bit-rate-scalability.

Summarizing, the two-loop MC-FGS characteristics are:

*Advantages*

- improved coding efficiency compared with the "basic" FGS scheme;
- unmodified base-layer structure and complexity;
- prediction drift occurring at low bit-rates is confined to the enhancement-layer B-frames;
- packet-losses within the enhancement-layer do not propagate beyond the enhancement-layer B-frames;

*Disadvantages*

- the coding gain is limited since the base-layer remains unchanged independent of the transmission bit-rate and thus it does not take full advantage of the temporal correlation among successive frames;
- higher complexity, since an additional MC-loop is added in the enhancement-layer for the B-frames. However, this disadvantage is compensated for by providing a device the option of resorting to a single-loop decoding when needed .

### 2.2 Single-loop MC-FGS

***Algorithmic description.*** In the single loop MC-FGS depicted in Fig. 1b, both the base- *and* enhancement-layer are used for the *base-layer* prediction. Thus, unlike the two-loop MC-FGS, this new structure *does modify* the base-layer performance. While the base-layer coding process remains unaltered, the coding parameters (e.g. quantization-step) change due to the improved reference frames resulting from the introduction of enhancement-layer data in the motion-compensation prediction loop. The fine-granular method employed by the "basic" FGS coding scheme is also used for coding the enhancement-layer residual.

For the single-loop MC-FGS, we introduce the notion of "extended base-layer" which includes an integrated base-/enhancement-layer data. Hence, if the transmission bandwidth drops below the rate necessary for transmitting this "extended" base-layer, the truncated "extended base-layer" data will suffer from drift until the next I-frame (for the case MC-FGS is employed for all frames). Consequently, with the proposed approach, even though prediction drift occurs at low transmission bit-rates, the fine-granular scalability property is still preserved, and a decodable stream can be generated at all bit-rates between the base-layer bitrate $R_{BL}$ and the maximum bitrate $R_{max}$. Two implementations of the proposed one-loop MC FGS method can be envisaged:

- the MC-FGS is applied for all frames, thereby achieving improved temporal decorrelation for all frames and hence, higher coding efficiency;
- the MC-FGS is restricted to the B-frames, to ensure that even at low bit-rates, the prediction drift does not propagate beyond subsequent base-layer P-frames.

It is important to notice that for the one-loop MC-FGS, there is significant coding gain that can be obtained by including all-frames (i.e. not only the B-frames) in the enhancement-layer MC, unlike the two-loop MC-FGS frames.

From a packet-loss resilience perspective, if the one loop MC-FGS is applied only upon the B-frames, then losses occurring within an I or P enhancement-layer frame are confined to the B-frames for which this frame is used as a reference. However, compared with the two-loop MC-FGS presented above, it is important to notice that in this case, the base-layer of the B-frame will also be affected. Alternatively, if all frames are included in the one-loop MC-FGS, a loss within an I or P enhancement-layer frame will propagate to all base- and enhancement-layer frames until the end of the GOP. This represents one of the drawbacks of the one-loop approach when compared with the two-loop framework.

***Complexity discussion.*** The one-loop MC-FGS encoder is depicted in Fig. 3 for the case MC-prediction is limited to B-frames. An important advantage of the one-loop MC-FGS framework is its low implementation complexity: only a set of logical “and” operations are added to the “basic” FGS encoder and decoder. If the “extended” base-layer is used only for the prediction of B-frames, an additional frame memory is necessary for both the encoder and decoder, since the references for the P and B-frames prediction are different.

Summarizing, the one-loop MC-FGS characteristics are:

*Advantages*

- high coding efficiency since the base-layer performance is modified to take advantage of the improved temporal decorrelation at transmission bit-rates higher than the “extended” base-layer bitrate;
- low-complexity due to the single-loop structure;

*Disadvantages*

- prediction drift can result at bitrates lower than the “extended base-layer” bitrate
- packet-losses occurring in the enhancement-layer may affect the base-layer performance.

As mentioned above, to limit prediction drift or packet-loss propagation, the one-loop MC-FGS can be restricted to B-frames at the expense of lower coding gain. Consequently, the one-loop MC–FGS is especially suitable for efficient transmission through channels with few packet-losses and for applications requiring low complexity decoders.

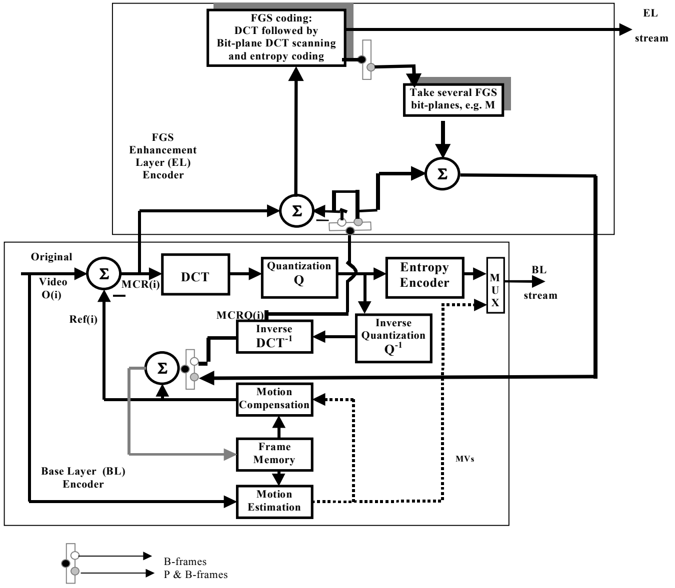

*Fig. 3. Block-diagram of the MC-FGS encoder with one MC-loop applied only for B-frames.*

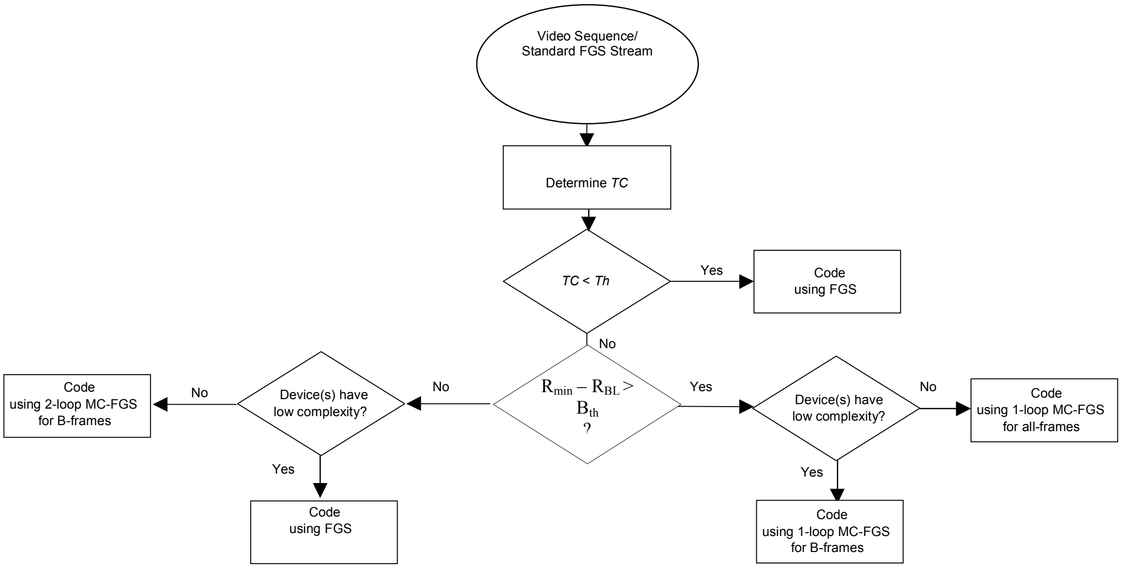

*Fig. 4. Adaptive MC-FGS decision algorithm for switching between the various FGS structures.*

### 2.3 Adaptive MC-FGS (AMC-FGS)

The performance of the motion-compensation structures described in the previous two sections depends on the sequence characteristics as well as the network characteristics or device capabilities. Hence, for an optimal streaming performance over wireless networks, an adaptive streaming system should be employed that chooses the most suitable FGS structure based on the bandwidth variations or device capabilities. The adaptation could take place either at encoding time or at a proxy within the network (e.g. at the base-station). For instance, in the case of a live broadcast, the bandwidth range of the listening receivers can be determined (e.g. based on RTP reports). Then, the bit-rate of the "extended" base-layer can be set to be lower than that of most clients, while the base-layer bit-rate can be set to equal the lowest receiver bit-rate. If the content is coded off-line, the switch between the various structures can take place at the proxy by transcoding between the various FGS formats. Alternatively, if multiple versions of the same content are coded using the various FGS structures, the stream that has the best performance for a particular network condition or device capability can be transmitted. Switching between the various streams can be performed for instance at an I-frame.

Figure 4 portrays a proposed decision mechanism for choosing between the FGS various structures. In the figure, *TC* is a measure of the temporal correlation among successive FGS frames and *Th* is some (TC) threshold. (See Appendix I for more details.) These parameters (*TC* and *Th*) are used to determine if the standard FGS structure is sufficient (from quality perspective). If that is the case, then FGS is the clear choice due to all its advantages (low complexity, high-scalability, and packet-loss resilience). Otherwise, one of the MC-FGS structures can be employed if there is a high-temporal correlation among successive frames. The decision for using the single-loop versus double-loop MC-FGS can be based on the bandwidth variation. In the figure we show a very simple example of making this decision by evaluating the difference between the minimum bitrate $R_{min}$ (for the different devices) and the base-layer bitrate $R_{BL}$ (e.g., of an already coded FGS stream). If the difference in bitrate is higher than some threshold $B_{th}$, then single-loop MC-FGS is a viable option. Otherwise, to improve quality for higher-bitrate receivers while avoiding the disadvantages of single-loop MC-FGS, one needs to consider double-loop MC FGS. It is important to note that the proposed algorithm can be made adaptive on a GOP-by-GOP basis (or even on a frame-by-frame basis). It should also be noted that for more complex receivers, the two-loop MC-FGS could even be replaced with the PFGS structure proposed in [11].

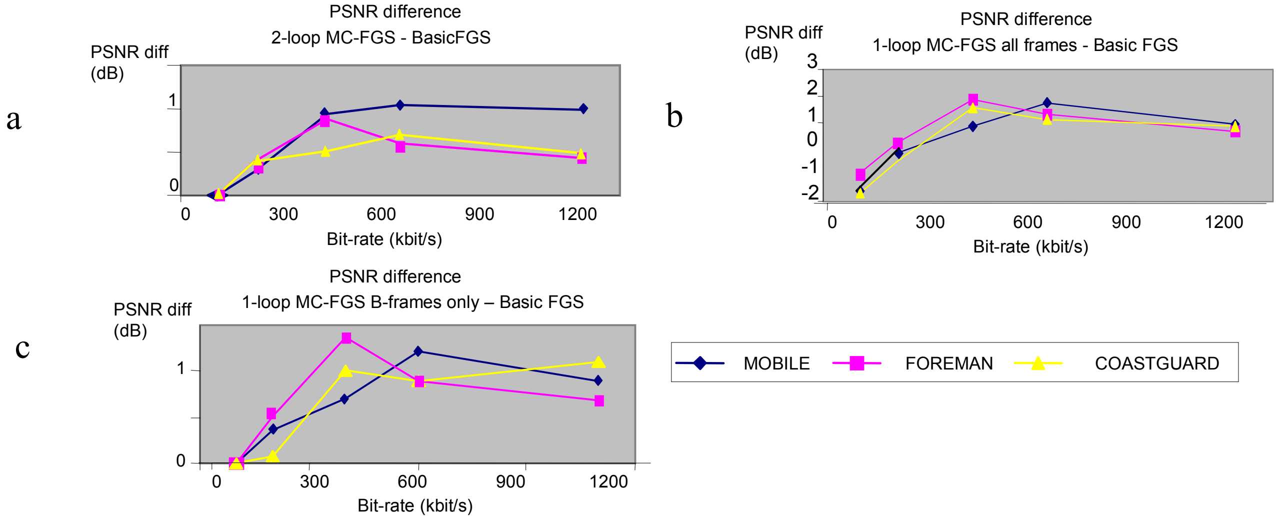

*Fig. 5. Comparison between "Basic" FGS framework and proposed MC FGS structures.*

## 3. SIMULATION RESULTS

In this section, the results for both proposed MC-FGS methods are given[3] for the three MPEG-4 sequences - *Coastguard*, *Foreman* and *Mobile* at CIF-resolution, 10 Hz. It is important to note that rate-control plays an important role in achieving a good performance with these new MC-FGS structures. However, even very simple rate-control mechanisms can result in a good rate-distortion performance. For example, in this section, a simple, yet efficient approach is used, that allocates the total bit-budget $B_{tot}$ of a GOP according to:

$$Btot = bI*No_{I\_frames} + bP*No_{P\_frames} + bB*No_{B\_frames},$$

with $No_{I\_frames}$, $No_{P\_frames}$, $No_{B\_frames}$ being the number of respectively I, P and B-frames within a GOP, and bI, bP, bB being the bit-budgets for the various frame types with bI>bP>bB. As mentioned above, the overall quality can be varied by employing a different number of enhancement-layer bitplanes within the motion-compensation loop.

In Fig. 5a, the PSNR differences between the proposed two-loop MC-FGS and the "basic" FGS scheme are plotted.

---

[3] It is important to note that a comparison between the two-loop MC-FGS and PFGS could not be performed, since neither the implementation details, nor the PFGS software are available.

Similarly, Fig. 5b and Fig. 5c portray the PSNR difference between the one-loop MC-FGS structures for all frames and B-frames, respectively. At bitrates higher than the "extended" base-layer rates, the results of the one-loop MC-FGS structure outperform that of the two-loop MC-FGS. This is mainly because in the one-loop MC-FGS, the reference frames are also improved and thus the base-layer frames take advantage of the improved temporal decorrelation. However, it is important to notice that at very low bit-rates (up to 1.5 $R_{BL}$ dependent on the sequence), the one-loop MC-FGS for all frames results in a relatively poor quality due to prediction drift. For the one-loop MC-FGS restricted to the B-frames, the drift is limited since in this case only the B-frames are affected. To limit the drift occurring at low bit-rates, fewer bitplanes can be inserted in the MC-loop. However, this comes at the expense of reduced coding efficiency for the higher bit-rates.

From the these plots, it can be noticed that the PSNR gain obtained by the various MC-FGS methods depends on the sequence characteristics: for sequences with a high temporal correlation between frames (e.g. "Mobile") the gain is higher than for those with less temporal correlation (e.g. "Coastguard"). Fortunately, the sequences for which the proposed MC-FGS schemes have a high coding efficiency are precisely those sequences for which the "basic" FGS had a poor R-D performance (see Table 1 in Appendix I). Consequently, the proposed MC-FGS structures reduce the FGS coding gain penalty compared to the non-scalable codecs to less than 1 dB for most sequences. Furthermore, to limit the drift of the one-loop MC-FGS for all frames while preserving its gain at higher bit-rates, a different number of enhancement-layer bit-planes should be incorporated in the motion-compensation prediction loop of the one-loop MC-FGS coding based on the temporal correlation within the sequence (see Section 2). For sequences with a high temporal correlation, the one-loop MC-FGS scheme should used to exploit this redundancy, while for sequences with limited temporal correlation, the "basic" FGS scheme with no enhancement-layer motion-compensation should be employed. This mechanism has been employed for the "Foreman" sequence that exhibits moderate temporal correlation in the beginning and high temporal correlation towards the end of the sequence and the results are portrayed in Figure 6. It can be seen that by switching between the one-loop MC-FGS scheme for all frames and the "basic" FGS scheme, the best trade-off between a high performance gain at high bit-rates and a reduced drift at low bit-rates can be achieved.

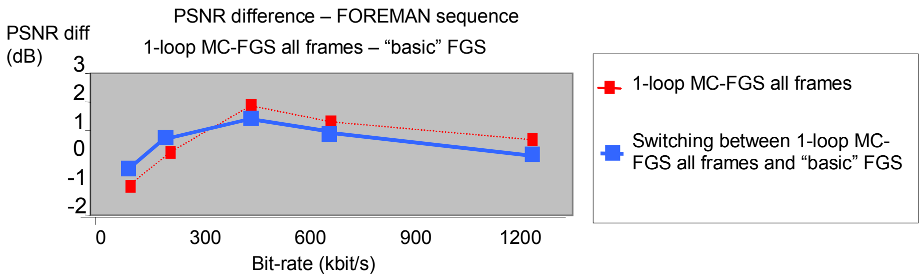

*Fig. 6. Drift reduction by alternating the "basic" FGS and the one-loop MC FGS for all frames based on TC.*

Based on the previously described analysis, it can be determined that MC-FGS should be employed only for sequences with a high temporal correlation. For these sequences, the coding penalty for FGS is high, and employing MC can significantly improve the coding efficiency. Moreover, since the sequences exhibit a large temporal correlation, temporal error concealment techniques become more efficient and thus, the reduced resilience of the MC-FGS is of less importance, since it can be compensated by efficient concealment. Furthermore, for these sequences, the prediction drift introduced at lower bit-rates is also less visible. For sequences with low temporal correlation, MC-FGS is not necessary because the temporal correlation has already been exploited at the base-layer level and the FGS coding penalty is very limited. Hence, the trade-off between FGS and MC-FGS structures can be made dependent on the temporal correlation of the sequence to be coded.

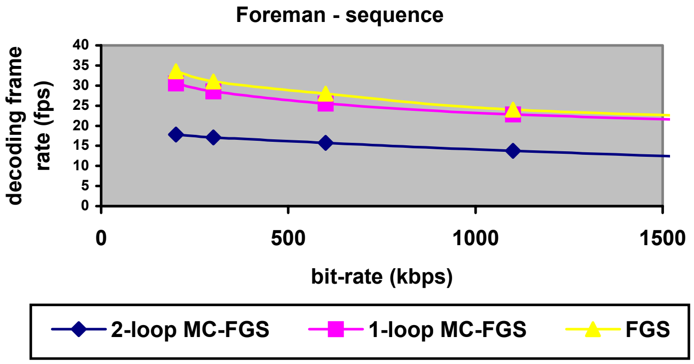

*Fig. 7. The complexity of the various FGS structures.*

From a complexity perspective, the two-loop MC-FGS is the most complex, followed by the one-loop MC-FGS as portrayed by Figure 7, where the complexity is expressed as the number of frames that can be decoded using the various structures on a 300Mhz Pentium PC [4]. Thus, depending on the mobile-terminals capabilities, a different stream can be streamed for optimal performance. Also, from Figure 7 it can be established that the FGS structures exhibit complexity scalability as well as bit-rate scalability, i.e. the decoding complexity decreases with the transmitted bit-rate.

## 4. CONCLUSIONS AND DISCUSSION

In this paper, we presented a new adaptive MC-FGS coding scheme that is able to fulfill the “Anywhere, Anytime and Anyone” wireless transmission paradigm, due to its ability to adapt to large variations in network characteristics and to the resource heterogeneity between devices. The proposed scheme adapts between the basic FGS structure and two novel MC-FGS approaches based on the sequence characteristics, network characteristics and device capabilities and is thereby able to achieve a large gain in coding efficiency (up to 2dB improvement in quality) compared with the basic FGS scheme. The proposed MC-FGS coding structures obtain an improved coding efficiency by introducing motion-compensation within the FGS enhancement-layer. Furthermore, it should be mentioned that the performances of the proposed MC-FGS schemes could be further improved by employing more sophisticated rate-control mechanisms.

The adaptive MC-FGS scheme provides an increased flexibility in customizing the FGS framework for a particular application. For example, with the addition of these MC-FGS structures, trade-offs can be easily made between coding efficiency, robustness to packet-losses and computational complexity. For improved coding gain, the one-loop MC-FGS for all frames can be used for higher bandwidth wireless transmission to “thin” clients. Alternatively, for low bandwidth wireless channels, either the basic FGS or the two-loop MC-FGS scheme for B-frames can be employed depending on the sequence characteristics and device capabilities.

## References

[1] M.-T. Sun and A. Reibmen, Editors, *Compressed Video over Networks*, Marcel Dekker, Inc., 2000.

[2] B. Girod and N. Farber, “Wireless Video,” Chapter in *Compressed Video over Networks*, Marcel Dekker, Inc., 2000.

[3] H. Radha, et al. “Multimedia over Wireless,” Chapter in *Advances in Multimedia: Systems, Standards, and Networks*, Marcel Dekker, Inc., 2000.

[4] M.R. Civanlar, “Internet Video,” Chapter in *Advances in Multimedia: Systems, Standards, and Networks*, Marcel Dekker, Inc., 2000.

[5] W. Tan, 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.

[6] H. Radha et al, “Scalable Internet Video Using MPEG-4,” Signal Processing: *Image Communication*, Sept. 1999.

[7] Study of ISO/IEC 14496-2:1999/FPDAM4, N3670, La Baule, France, USA, Oct. 2000.

[8] H. Radha, M. van der Schaar, and Y. Chen, “The MPEG-4 Fine-Grained Scalable Video Coding method for Multimedia Streaming over IP”, *IEEE Transactions on Multimedia*, Vol. 3, No. 1, March 2001.

[9] M. van der Schaar and H. Radha, “Unequal Resilience for Fine-Granular-Scalability Video,” Accepted for publications in *IEEE Transactions on Multimedia*.

[10] S.J. Choi, J. W. Woods, “Motion-Compensated 3-D Subband Coding of Video,” *IEEE Trans. on Image Proc.*, Feb. 1999.

[11] U. Benzler et al, ”Result of Core Experiment on Fine Granularity Scalability for Video (MC with drift)”, m4847, July 1999.

[12] F. Wu, S. Li, Y.Q. Zhang, “DCT-Prediction based Progressive Fine Granularity Scalability”, Proc. of ICIP, Oct. 2000.

[13] D. Wilson, M. Ghanbari, “Transmission of SNR scalable two layer MPEG-2 coded video through ATM networks”, Proc. on 7th International Workshop on Packet Video, March 1996.

[14] U. Horn, K. W. Stuhlmüller, M. Link, B. Girod, “Robust Internet Video Transmission Based on Scalable Coding and Unequal Error Protection”, Signal Processing: *Image Communication*, September 1999.

[15] K. Rose, P. Wu, S.L. Regunathan, “Efficient SNR scalability in predictive video coding”, Proc. of ISCASSP, 1998.

## Appendix I: The FGS Coding Method and its performance

While providing high flexibility in adapting to the bandwidth variations and high robustness to packet-losses, the FGS scheme is less efficient than a non-scalable coder functioning at the same transmission bit-rate. An extensive evaluation of the FGS coder performance compared with that of an MPEG-4 non-scalable coder could be found in [8]. In Table 1, the coding penalty of MPEG-4 FGS compared with an MPEG-4 non-scalable codec at various transmission bit-rates is portrayed for several sequences, for which the temporal correlation has also been determined.

---

[4] While the performance analysis was obtained using a high complexity PC, the results obtained are indicative for both the complexity scalability as well as for the complexity levels of the FGS structures.

**Table 1.** *The coding penalty (ΔPSNR) of MPEG-4 FGS compared with an MPEG-4 non-scalable codec at various transmission bit-rates. The temporal correlation (TC) between each frame and its motion-compensated reference (determined by the first-order correlation coefficient) is also given as a reference.*

| Sequence | TC | ΔPSNR @ 300 kbit/s [dB] | ΔPSNR @ 500 kbit/s [dB] | ΔPSNR @ 1 Mbit/s [dB] |
|---|---|---|---|---|
| Coastguard | 0.30 | 1.69 | 1.99 | 1.61 |
| Foreman | 0.45 | 2.45 | 2.66 | 2.05 |
| Mobile | 0.50 | 2.35 | 3.53 | 4.30 |

From the results in Table 1, it seems that the FGS has a lower coding penalty for sequences with low temporal correlation. The explanation for this resides in the intrinsic characteristics of the FGS framework: if there is only limited temporal correlation within the sequence, this can be captured within the base-layer. However, in sequences with high temporal correlation and high frequency textures, the base-layer can not fully exploit this correlation.

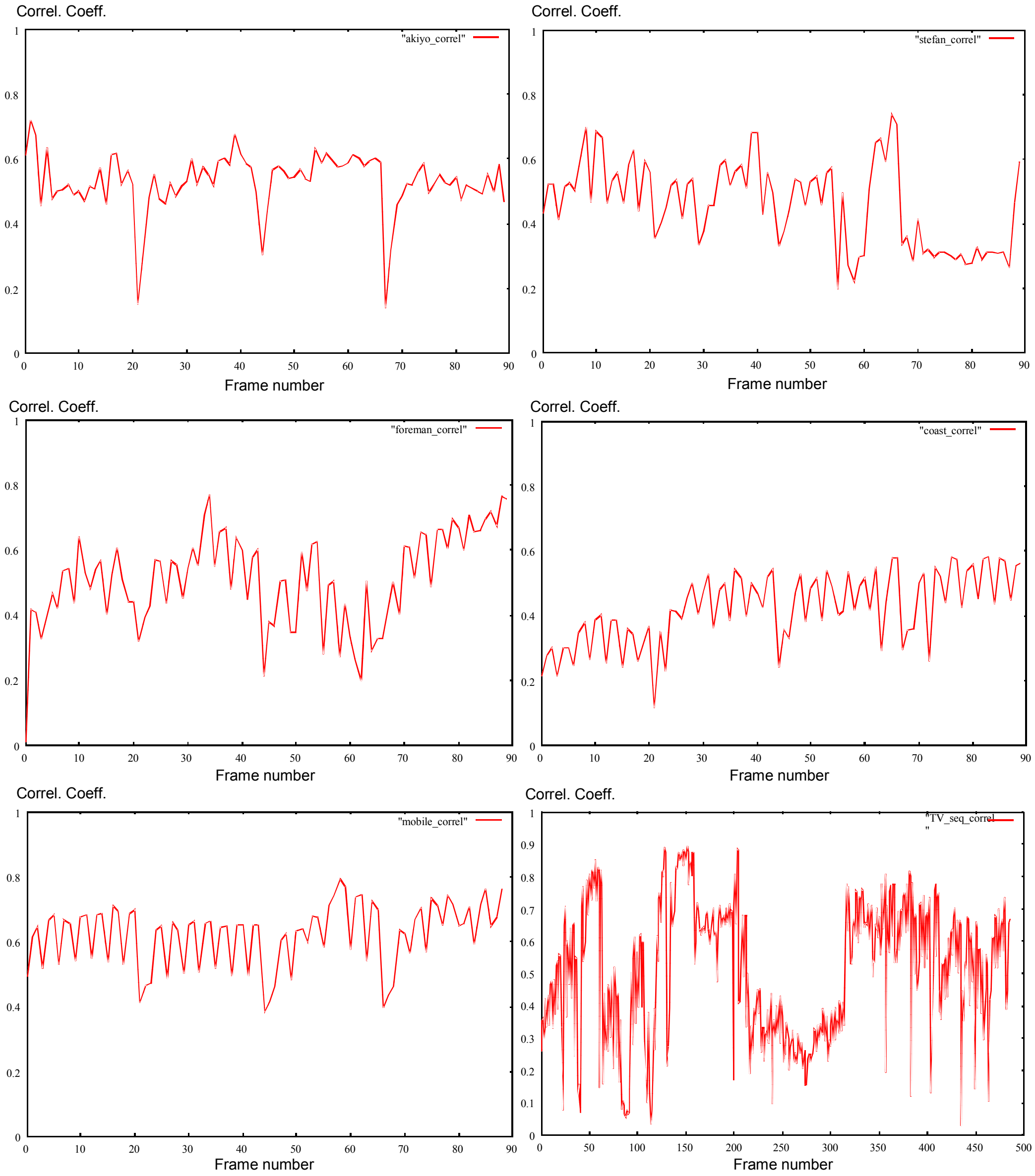

*Fig. 8. Temporal correlation for various sequences.*

In the sequel, we study the correspondence between the FGS coding efficiency penalty and the temporal correlation of the sequence in more detail. In Figure 8, the remaining temporal correlation between enhancement-layer frames is portrayed for several MPEG-4 video test sequences – *Foreman, Coastguard, Mobile, Akiyo, Stefan* - at CIF-resolution and 10Hz. Also, a longer sequence from a television broadcast – *TV_seq* - that contains several scene cuts has been added to the test set to represent "typical" content.

The temporal correlation can be determined at encoding time either on-the-fly or off-line. Subsequently, the portrayed enhancement-layer temporal correlation for a certain frame has been compared with the coding efficiency penalty for that particular frame between FGS and the non-scalable coder, that is displayed in Figure 9. Based on this comparison, we can conclude that FGS suffers most (when compared with the ``ideal" non-scalable coding case) for sequences with slow motion and high frequency-detail that exhibit a high temporal correlation between successive frames. From Figure 9 it can be concluded that:

- The temporal correlation is a good indication of the coding penalty gap between the non-scalable coder and FGS.
- If the temporal correlation is below a threshold, then the FGS performance is comparable to that of a non-scalable coder (see for instance the scene change in the "Foreman" sequences, the high-moving scene in the "*TV_seq*" sequence etc.). This is because there is no advantage associated with employing motion-compensation for the higher frequencies and because the FGS entropy coder is quite efficient in compressing the high frequent SNR residual signal.

Hence, to determine the FGS coding efficiency penalty compared to the non-scalable coder, the temporal correlation (TC) between enhancement-layer frames needs to be estimated. Then if TC is larger than a preset threshold Th (in our experiments Th=0.4), then the coding penalty compared to the non-scalable coder is mostly below 1dB. Otherwise, the non-scalable coder outperforms FGS at the same bit-rate with more than 1 dB and hence the difference in quality performance between the two codecs becomes visible.

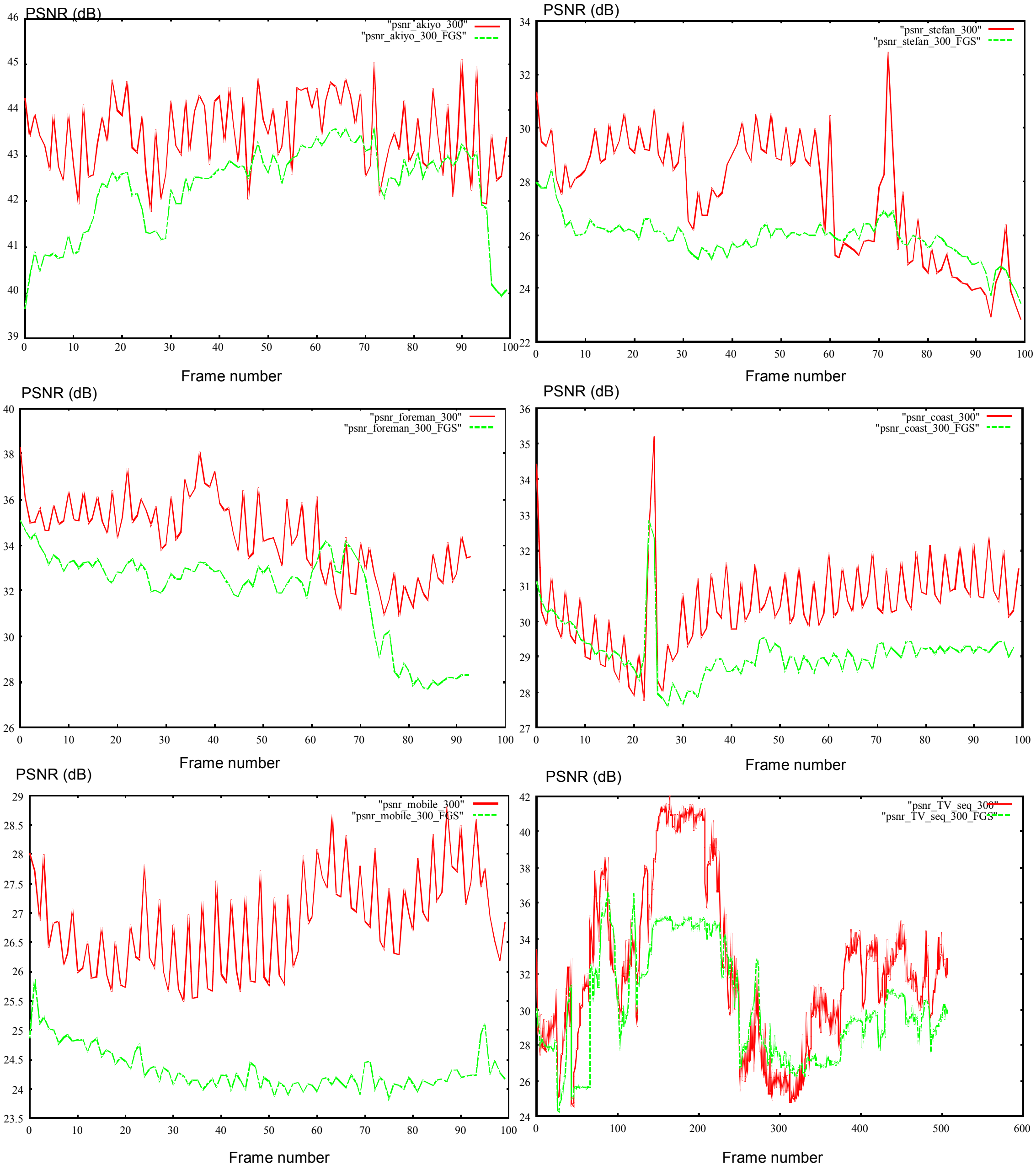

*Fig. 9 PSNR values per frame for various sequences coded with a non-scalable coder at 300 kbit/s and coded with FGS with a 100 kbit/s base-layer and a 200 kbit/s enhancement-layer.*