

# SNR Scalable Layered Video Coding

*Peter Amon, Klaus Illgner, Jürgen Pandel*

Siemens Corporate Technology  
Information and Communications CT IC 2  
D-81730 Munich, Germany  
E-mail: {peter.amon, klaus.illgner, juergen.pandel}@mchp.siemens.de

## 1 Introduction and Motivation

Scalable video coding gets currently increasing attention in the community and was already a major research area for some time. In the scientific literature a broad variety of different approaches are published and the standardisation bodies included this functionality in some profiles, for instance MPEG-4 FGS [1]. Nevertheless, scalable video coding approaches are rarely used in practical applications. This raises the question why this technology is not used, although the interest and the need of this functionality is frequently mentioned.

Certainly, the lack of performance of current approaches is a significant reason not to use scalability. Today it is more efficient to send independent bitstreams for each quality / resolution layer simultaneously (Simulcast) rather than use a scalable video coder. But, if there is a real need for such a technology, some of the reasonably well performing known scientific approaches would have been adopted, or new ones would have been developed. This leads to the conclusion, that the need of scalable video coding approaches is not clearly understood.

The following paragraphs view the problem from different angles. There are well understood scenarios in which multiple encodings of the same data source at different quality levels or resolutions have to be transmitted or at least prepared for simultaneous transmission. Examples include multicast and broadcast delivery to consumers using different types of terminals (TV, PDA, SmartPhones, ...) and/or connected to different networks (DVB, DSL, ISDN, UMTS, ...). Other applications provide the option of flexible user preferences or different levels of pay-per-view services. Also in streaming applications there is a need to provide numerous bitstreams encoded for several transmission conditions.

Since scalable approaches do not work well enough, a common way to achieve different quality layers is to encode a sequence several times with several configurations for different transmission and reception conditions. If necessary the different streams are sent simultaneously to the different receiver devices (Simulcast). From an efficiency point of view this is a waste of bandwidth, as the low quality layers are somehow included in the high quality / resolution layer.

This paper presents a new coding scheme [2], which is a kind of intermediate technology between Simulcast and fully scalable approaches. The concept requires only slight modifications of a Simulcast system providing several quality levels with lower bit-rate. This is achieved by synchronizing several encoders and differential coding of the transform coefficients. In principle the approach can be applied to any kind of video codec based on the hybrid coding scheme.

## 2 Concept

### 2.1 General Structure

The structure of the system is based on a Simulcast coding scheme, which consists of  $N$  parallel but independent coders. To achieve different quality levels (SNR scalability) each layer employs a dif-

ferent quantizer. Although coded independently, the signals of the individual coders show strong correlation, which can be exploited. The key characteristics of the concept presented is that it does not independently encode each layer, but predicts the higher quality layers from the next lower quality layer. The invention enabling a considerable coding gain and making the concept unique is also to design appropriate quantizers and synchronize all layers in terms of the coding parameters (e.g. motion vectors, prediction modes).

### 2.1.1 Encoder

The encoders of our system are not running independently of each other like in Simulcast encoding. They are synchronized such that each stage uses the same high level decisions (e.g. motion vectors, prediction modes). In combination with an appropriate quantizer design this allows an efficient differential encoding of the coefficients.

Figure 1 shows an example of the concept with three layers. For encoding each coder is basically independent, except that the same configuration is used. The signals being transmitted, however, are the differential signals of the prediction errors.<sup>1</sup>

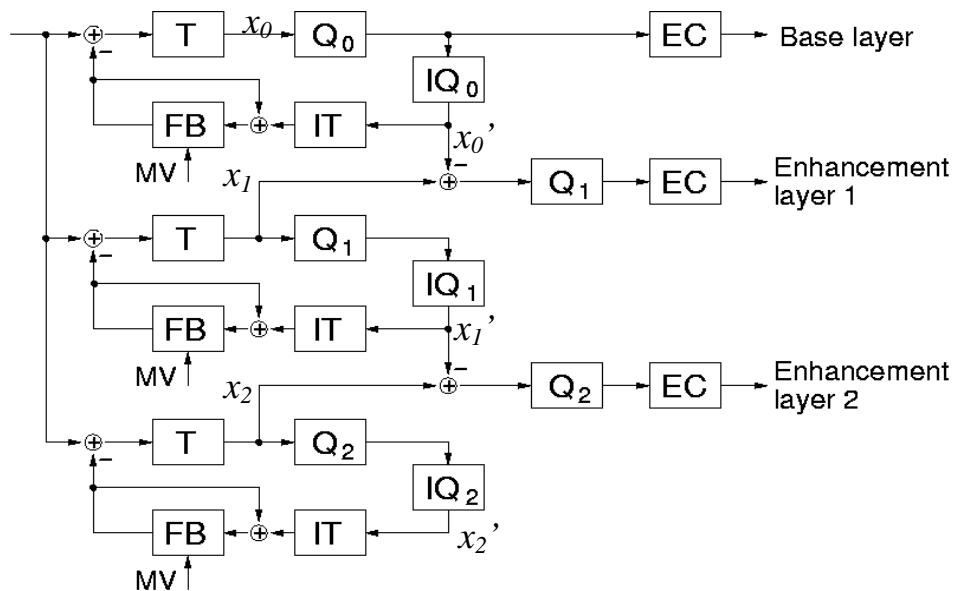


Figure 1: Encoder part

### 2.1.2 Decoder

The decoder part shown in Figure 2 does the entropy decoding and “inverse quantization” of the residual streams first. The transform coefficients for a particular stage are reconstructed by adding all residual streams up to this level.

<sup>1</sup> The abbreviations used in Figure 1 and Figure 2 are

- T: transformation,
- Q: quantization,
- IQ: ‘inverse quantization’,
- IT: inverse transformation,
- FB: frame buffer,
- MV: motion vectors,
- EC: entropy coder,
- ED: entropy decoder.

In case only a lower quality reconstruction is needed, for example in low complexity devices, only the first stage may be used.

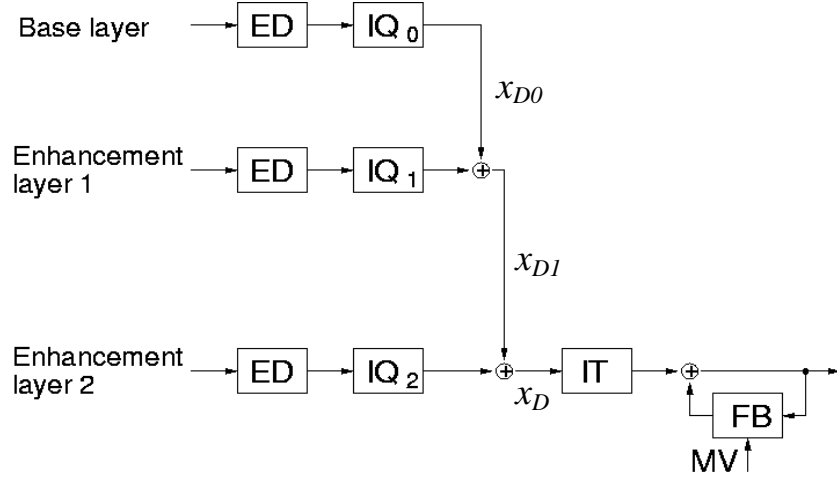


Figure 2: Decoder part

## 2.2 Quantization

As simple as the idea may appear in principle, it is not obvious that the signal reconstructed from the coded and quantized residual signal is indeed identical to the decoded prediction error signal of the synchronized but independent encoder. In other words, it must be proven that there is no mismatch between encoder and decoder, respectively that there is no drift. Key is the design of the quantizers used for the residual signal and the relationship of the quantizers within the encoders. The later constraint is considered by employing so called “embedded quantizers”

$$(1) \quad \frac{Q_{i-1}}{Q_i} = n_i \in \mathbb{IN} \text{ for } i = 1 \dots l-1,$$

where  $Q_i$  is the quantization step size of level  $i$ , and  $l$  is the number of layers. If this equation does not hold drifting effects will occur. For a practical situation where the quality levels are not too much different, a value of  $n_i = 2$  for all consecutive levels is suitable.

The quantization process with rounding parameter  $\alpha_i$  is defined as

$$x'_i = \left\lfloor \frac{x_i}{Q_i} + \alpha_i \right\rfloor \cdot Q_i = \left\lfloor \frac{x_i + \alpha_i \cdot Q_i}{Q_i} \right\rfloor \cdot Q_i,$$

where  $\lfloor x \rfloor$  is the next integer smaller or equal  $x$ .

Note, if the parameter  $\alpha_i$  is set to 0 the generic rounding operation is reduced to the floor operation.

If we want to proof the absence of drift effects we have to show that  $x_{D_i} = x'_i$  for  $i = 0 \dots l-1$ , where the case  $i = 0$  is obvious. The enhancement layer signal before quantization can be calculated as

$$\Delta x_i = x_i - x'_{i-1} = x_i - \left\lfloor \frac{x_{i-1} + \alpha_{i-1} Q_{i-1}}{Q_{i-1}} \right\rfloor \cdot Q_{i-1}.$$

After quantization the enhancement layer signal is

$$\begin{aligned}\Delta x'_i &= \left[ \frac{\Delta x_i + \alpha_i Q_i}{Q_i} \right] \cdot Q_i = \left[ \frac{x_i - \left[ \frac{x_{i-1} + \alpha_{i-1} Q_{i-1}}{Q_{i-1}} \right] \cdot Q_{i-1} + \alpha_i Q_i}{Q_i} \right] \cdot Q_i \\ &= \left[ \frac{x_i + \alpha_i Q_i}{Q_i} - \left[ \frac{\Delta x_{i-1} + \alpha_{i-1} Q_{i-1}}{Q_{i-1}} \right] \cdot \frac{Q_{i-1}}{Q_i} \right] \cdot Q_i.\end{aligned}$$

If the condition

$$\frac{Q_{i-1}}{Q_i} = n_i \in \mathbb{N}$$

holds,

$$\Delta x'_i = \left[ \frac{\Delta x_i + \alpha_i Q_i}{Q_i} \right] \cdot Q_i - \left[ \frac{\Delta x_{i-1} + \alpha_{i-1} Q_{i-1}}{Q_{i-1}} \right] \cdot Q_{i-1} = x'_i - x'_{i-1}.$$

Hence

$$x_{Di} = \Delta x'_i + x_{Di-1} = x'_i \text{ for } i = 1 \dots l-1,$$

what we wanted to show.

### 2.3 Synchronization

The concept relies on the assumption that all encoders are running synchronously, or in other words high level coding decisions are taken only once. For instance, motion vectors will be calculated only at a single layer, which also saves considerably computational resources. In principle, the computation of the motion vectors can be done at any layer. Not only the base layer coder can be chosen as a master coder, but any of the enhancement layer coders can be optimized regarding motion estimation. All other coders (slaves) do their motion compensation based on the decisions of the master. As a consequence the motion vectors have to be transmitted only once, saving additional data rate.

In more sophisticated coders, a variety of other parameters have to be synchronized as well. Examples derived from the H.26L codec are [2]

- the decision between intra and inter prediction of certain blocks,
- the decision between different prediction modes,
- the decision between different block sizes (e.g. for motion search or intra prediction),
- deblocking and interpolation filter parameters.

These high level decisions also have to be transmitted only once. They are transmitted in the base layer, although the values may be the decision of any enhancement layer. Every other stage of the encoder delivers the coded differential coefficients (enhancement layers).

As previously mentioned, the determination of these high level decisions can be made at any stage, so the most important level can be optimized. For instance, in the case of three stages, the synchronization parameters can be optimized for the second quality level in order to avoid great divergence between the optimal and the synchronized high level decisions. The advantage of this strategy has been shown in simulations.

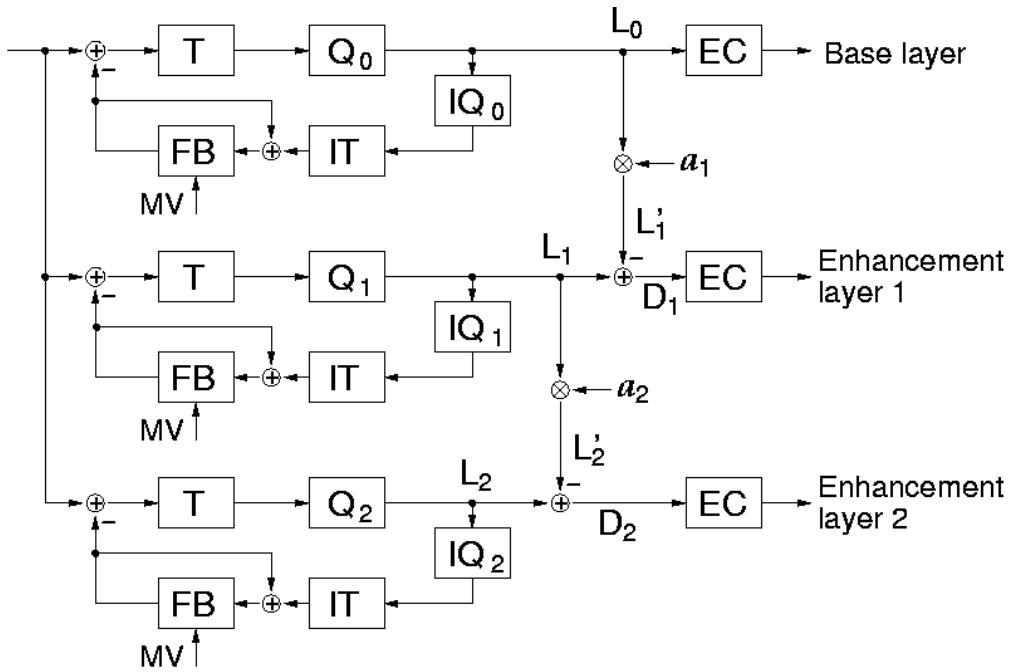
### 3 Modified Approach: Level Prediction

While the concept described above combines quantized and non-quantized transformation coefficients, we investigated a direct prediction of the higher layer coefficient levels from the lower layer coefficient levels. To achieve optimal prediction of the levels of the higher layer  $i + 1$ , the levels at the lower layer  $i$  have to be multiplied by a factor  $a_{i+1} > 1$ . This factor  $a_{i+1}$  depends on the ratio of the quantization step sizes, hence

$$(2) \quad a_{i+1} = \frac{Q_i}{Q_{i+1}} \text{ for } i = 0..l-1.$$

Note that the factor  $a_{i+1}$  does not necessarily have to be an integer to avoid drift effects. So the advantage of this level prediction scheme is that there is no longer a restriction on the quantization step size, which allows a finer granularity of the quality levels.

A simplified block diagram of the encoder for level prediction is shown in Figure 3.



**Figure 3: Encoder part for level prediction scheme**

Instead of multiplying a level  $L_i$  with a floating point number  $a_{i+1}$  we use an integer multiplication and a subsequent right shift to assure that the multiplication is performed exactly in the same way in the encoder and any decoder. The prediction level  $L'_{i+1}$  is then calculated by

$$(3) \quad L'_{i+1} = \text{sign}(L_i) \cdot \left\lfloor \frac{|L_i| \cdot M_{i+1}}{2^n} \right\rfloor \text{ for } i = 0..l-1,$$

where

$$(4) \quad M_{i+1} = \left\lfloor 2^n \cdot \frac{Q_i}{Q_{i+1}} \right\rfloor.$$

$n$  is an arbitrary positive integer to achieve an accurate approximation of the factor  $a_{i+1}$ . The base layer levels  $L_0$  and the prediction errors  $D_i = L_i - L'_i$  will be entropy coded and transmitted to the decoder.

In Figure 4 the block diagram of the decoder is shown. The prediction level again is given by Equation 3. There is no need for the transmission of the factor  $a_{i+1}$ , because it can be calculated from the quantization step size. Depending on how many layers are transmitted to the receiver, the decoder must switch to the respective output of the ‘inverse quantizer’ in order to get the right coefficient level. In any case only one frame buffer is needed. When data of a higher layer is lost due to bad transmission conditions, the decoder shall switch to the next uncorrupted lower layer.

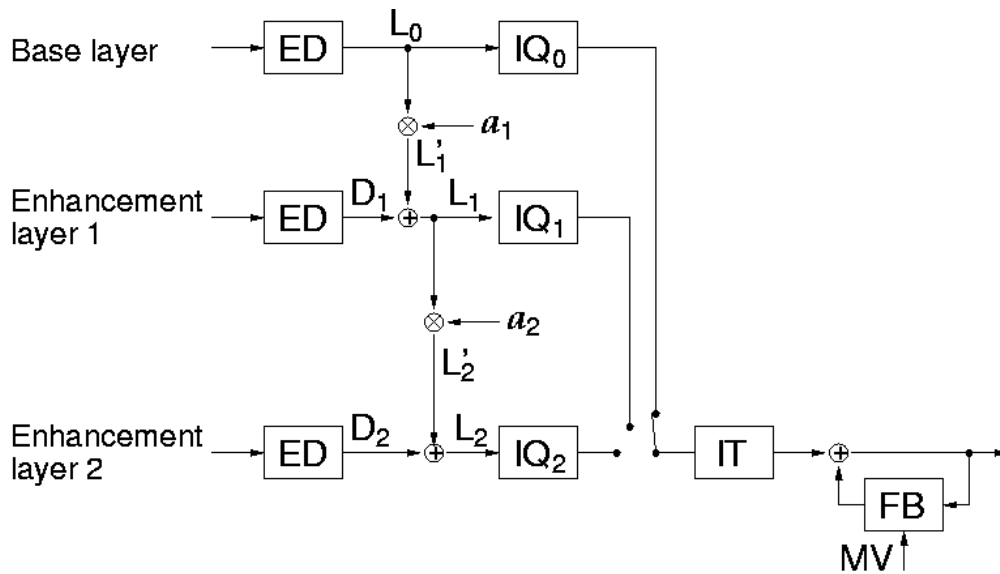


Figure 4: Decoder part for level prediction scheme

## 4 Results

Simulations were performed using a coder based on the current ITU-T H.26L technology. This coder was modified, so that several layers of the coder can be synchronized and the levels of higher layers can be predicted from the coefficients of lower layers. For the simulations, we used the sequence ‘Foreman’ in QCIF. 100 frames were coded with a frame rate of 10 frames per second.

In Table 1 and Figure 5 the results of our new layered schemes compared to Unicast coding and Simulcast coding are shown.

	Simulcast	Advanced Simulcast	Saving
Base layer	62.8 kbit/s	62.8 kbit/s	-
Enhancement layer	133.7 kbit/s	106.4 kbit/s	20.4 %
Total	196.5 kbit/s	169.2 kbit/s	13.9 %

Table 1: Results for two layers

Simulcast coding and our Advanced Simulcast coding scheme consist of two layers. The first layer (base layer) was chosen at the same quality. The second layer was also chosen with a similar quality. Due to the differential encoding of the levels the second layer is reached earlier by our Advanced Simulcast coding scheme, saving data rate. Its quality is slightly below the quality of the Simulcast

coding scheme, because the high-level decisions were optimized for the base layer. (Note: The optimization can also be done at the enhancement layer.)

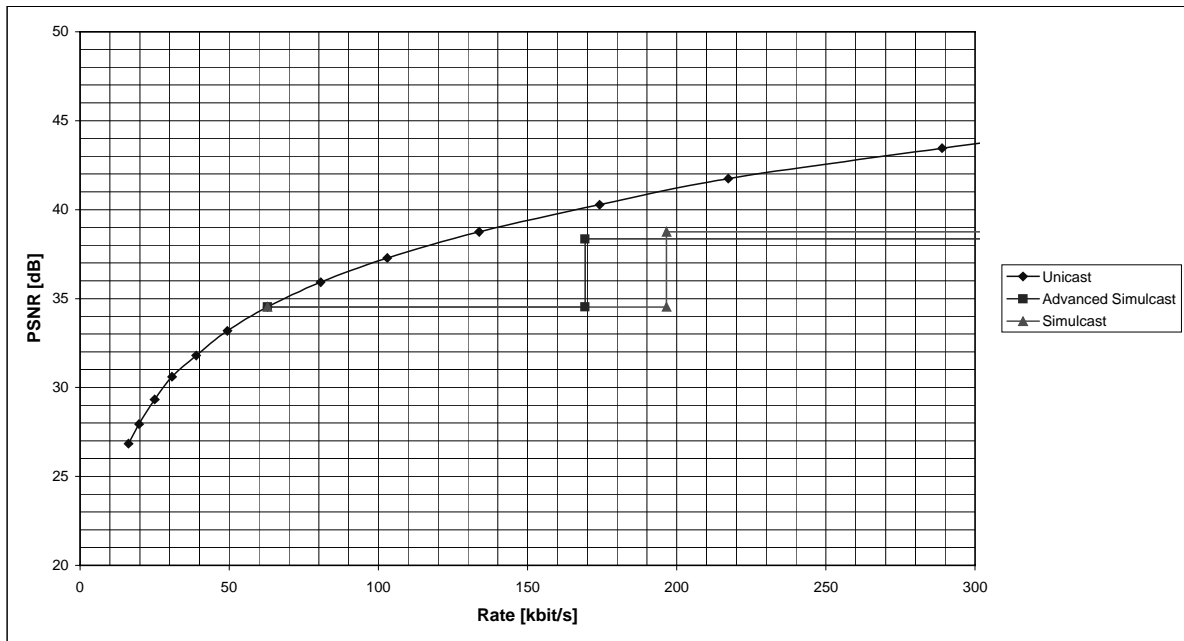


Figure 5: Results for two layers

In Table 2 and Figure 6 the comparison using three quality layers is shown. Here the Advanced Simulcast coding scheme is optimized for the first enhancement layer. Again, a significant reduction of the rate can be observed with only a slight decrease in quality of the base layer an the second enhancement layer. As expected the saving increases the more layers are used.

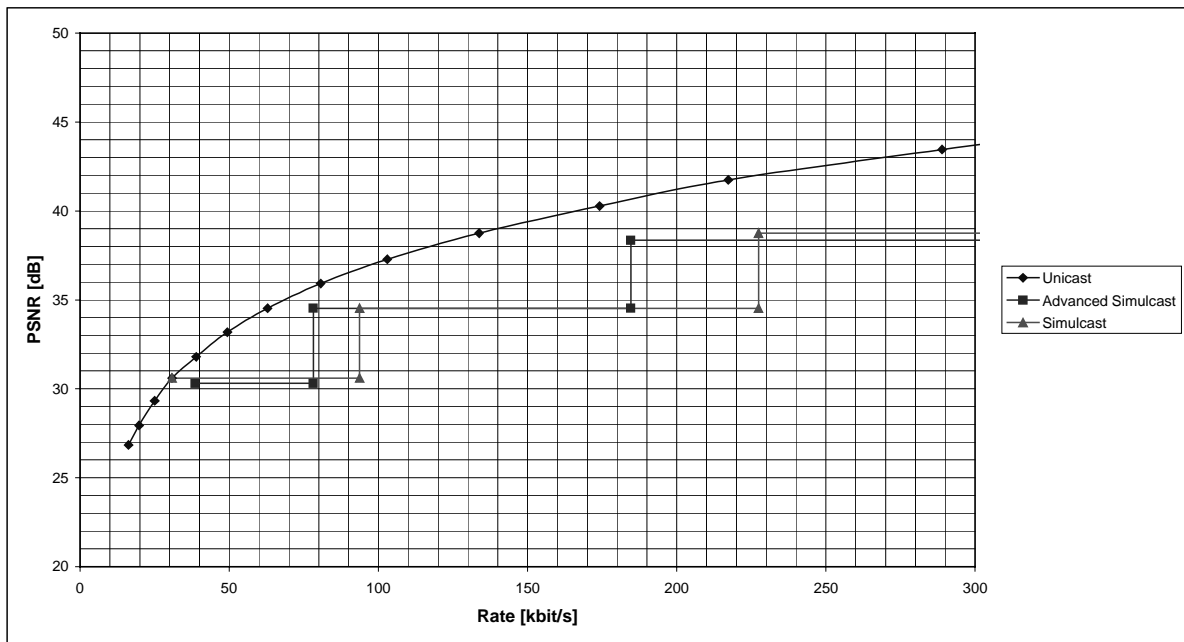


Figure 6: Results for three layers

	Simulcast	Advanced Simulcast	Saving
Base layer (BL)	30.8 kbit/s	38,6 kbit/s	-25.3 %
Enhancement layer 1 (EL1)	62.8 kbit/s	39.6 kbit/s	36,9 %
Enhancement layer 2 (EL 2)	133.7 kbit/s	106.4 kbit/s	20.4 %
Total (BL + EL1)	93.6 kbit/s	78.2 kbit/s	16.5 %
Total (BL + EL1 + EL2)	227.3 kbit/s	184.6 kbit/s	18.8 %

**Table 2: Results for three layers**

## 5 Conclusions

We presented a new layered video coding scheme for SNR scalability. Several layers are synchronized and the coefficients of higher layers are predicted by the coefficients of the respective lower layer. We pointed out the restrictions of the quantizer step sizes that are necessary to have the property of embedded quantizers.

We modified the scheme in order to directly predict the coefficient levels that are transmitted instead of the transform coefficients. This eliminates the restrictions that are necessary for the prediction of the coefficients. Therefore, we can achieve finer granularity between neighboring layers.

The proposed Advanced Simulcast scheme shows a significant better performance than Simulcast coding, to which it has to be compared to. The more layers are encoded and the more similar the rates of the individual quality layers of a Simulcast coder are the higher is the gain achieved by the Advanced Simulcast coding scheme presented here. In addition, the gap between our scheme and Unicast coding, where the quality is optimized for a single coding rate, is remarkably small.

## 6 References

- [1] W. Li, *Overview of fine granularity scalability in MPEG-4 video standard*, In IEEE Transactions on Circuits and Systems for Video Technology, Vol. 11, No. 3, pp. 301-317, March 2001.
- [2] P. Amon, G. Bäse, K. Illgner, J. Pandel, *Efficient coding of synchronized H.26L streams*, document VCEG-N35, ITU-T Video Coding Experts Group (VCEG) Meeting, Santa Barbara, CA, USA, September 2001.