ISO-IEC/JTC1/SC29/WG11/
MPEG 92/

CCITT SGXV
W.P. XV/1
Experts Group for ATM Video Coding

Doc. AVC - 255
Stockholm, 18-20 March 1992

Title: **About the constraints on variable bit rate coding**

Source: DBP-Telekom

Purpose: For discussion

Introduction

This paper tries to give a survey on the major problems that appear when variable bitrate coding is used in ATM networks. Some possible solutions are presented. Finally the applicatoation of variable bitrate coding for distribution purposes is discussed.

Some basic problems of variable bitrate coding in ATM networks

It has always been stated, that one of the major advantages of a future ATM network will be it's ability to enable the transmission of bitstreams with variable rates (VBR), a feature that is most tempting for applications like video coding. VBR faciliates to code with (near) constant quality, because in phases of high scene activity the datarate is allowed to rise according to the demanded values. Additionaly, more static phases of a picture sequence do not cause the insertion of any stuffing bits, whereby the momentary bitrate might almost be reduced to zero. Because of that, the *mean* rate of a VBR bitstream could fall significantly below the rate of a constant bit rate (CBR) coder of the same quality. This should not only reduce the connection costs, but leads to an other profit from VBR coding. With the so called statistical multiplexing, i.e. the allocation of a channel bandwidth smaller than the peak rate to every single of multiple VBR sources, it is in principle possible to accept more sources on a certain channel than it would be with CBR sources of the same quality.

For a simple reason the application of VBR is restricted significantly under realistic operating conditions. A basic funktion of any ATM network is the 'connection admission control' (CAC). It has to decide whether a new call can be admitted or not. An admission will only be possible, if the 'quality of service' (QOS) can still be guaranteed for all present connections. In particular, the 'cell loss ratio' (CLR) which tends to rise with growing network occupation has to be kept below a certain value (10^{-10}). For this task the CAC has to be provided with some information about the (statistical) behaviour of the admitted VBR sources. No such informations are available about unconstraind VBR sources. For this reason a contract between network and source is necessary, that sets values for some of the basic parameters of the connection. It is commonly exepted, that the peakrate is the most important of those parameters. Others could in principle be the maximum duration of a peak, the minimum time between two peaks or the avarage rate of the connection.

Once these parameters are negotiated, they have to be controled during the whole connection time, to prevent faulty sources or sources that do not keep the contract from affecting the functionality of the net. If the 'usage parameter control' (UPC) functions detect a contract violation, the concerned cells are deleted or marked for deletion (marked as low priority). Inside the codec the same function has to be implemented as *preventive* UPC. This seems to be the only way to exploit the assigned bandwidth without the danger of violating the contract unintentionally. UPC and CAC are closely related. The CAC always has to assume the worst case of what the UPC will accept to be in line with the negotiations.

The selection of suitabel parameters for the UPC does not only influence the coding results, but also the whole network perfomance. In particular, parameters like the mean rate that can only be measured with sufficient accuracy when the measurement intervall is long enough are very problematic. It almost seems to be impossible to make an exact determination of these parameters compatible with an accurate and fast recognition of parameter violations. Further, the shape of bitstreams that are in line with the negotiations may take a very unfavourable progress when disadvantageous parameters are choøsen. A peak/mean UPC for instance faciliates on/off sources, that transmit with peak rate for the maximum permitted period and are than silent until the mean is reached again. For a certain QOS the multiplexing gain for sources like this would be minimal. For this reason sophisticated UPC mechanisms have to be investigated, that are based on parameters that the network is easily able to measure, and that force the sources to a suitable shape.

Anyway, UPC is not a topic video coding should by-pass. UPC will always be present in the codecs as *preventive UPC*, and different methods of UPC could be due to very distinct coding results. To make investigations on VBR coding schemes comparable, a preventive UPC reference model is needed.

The different potentialities of video coding in ATM networks

As described above, a basic conflict is araising from the goal to have the VBR bitstreams as unconstrained as possible, the necessety of reliable CAC and UPC functions and a most efficient exploitation of the network resources. This chapter tries to give a survey on all the (known) methods to deal with this conflict and comprises their pro's and con's.

In the diagram at the end of the document everything important is collected. Besides coding with constant bitrates it contains four approaches to VBR coding, from which the ones labeled with 2-4 are mainly different from their methods of UPC and network management.

method 1:

From the coding point of view CBR does not add any new aspects, exept the freedom to transmit with virtually every desired bitrate. On the network side a restriction to CBR will shurely result in the least cost intensive way of network arcitecture.

method 2:

Once VBR is available, it is most tempting for the network to restrict itself on peakrate

UPC, because this could be done without much problems and makes a simple CAC feasible. Unfortunately, to guarantee a certain QOS, the CAC has to allocate the peak bandwidth for every VBR service in this case and charge accordingly. A statistical multiplexing will than be impossible and a VBR codec would transmit as CBR source at peakrate to optimize it's quality within the rented bandwidth.

A modified network strategy may solve this problem. A "real VBR" with very simple UPC could be made attractive by letting the charge be based on counting cells (or similar methods). This would result in a "leftover bandwidth" that may be used by non timedependent services like those for data transfer. The great advantage of this idea is, that it combines an almost unconstraind VBR with a simple UPC mechanism. On the other hand it creates the need for huge buffers in which the non timedependent cells have to wait for sufficient leftover capacity. A special protocol has to be introduced, that requests the non timedependent services to refill these buffers. Further this method might be a bad deal for the network, when no data cells are available for longer periods of time, and it is not yet clear if the QOS can be guaranteed for the non timedependent services (parameters that deal with the 'maximum transmission time' might be violated). It has to be studied whether this idea works in practical operation or not, but if it does it's a very good solution.

method 3:

The "conventional" policed VBR algorithms suffer from the fundamental drawback that only short peaks are coded without quality degradation. Long peaks or general fluctuations of the mean bitrate are not covered by these algorithms. This statement is more or less valid for all UPC algorithms that control more than only the peakrate. Always some kind of 'effective datarate' significantly below the peakrate has to be met (with differing definitions for different UPC's). Whether this 'effective datarate' is sufficient or not is only dependent on the average activity of the scene. Similar to CBR coding, the *average* picture quality does consequently always correspond to this *effective* datarate, only the time constraints are more relaxed. In CBR coding the average bitrate has to be met within a period of time, that is dependent on the buffer length; in the VBR case this time might be much longer (for instance 10-30 sec.) but this does not change the situation in principle. Moreover, compared to openloop VBR the multiplex gain will drop considerably when preventive UPC is utilized.

method 4:

A special kind of network management may partly overcome the problems of method 3. With the so called "Fast Reservation Protocol" it might be possible to renegotiate unfavourable UPC-parameters during the connection. In this way longer peaks or rate fluctuations may be compensated by new parameters. It's not quite clear until now how long such a renegotiation would take and how often it could happen. Further it has to be studied if this procedure is practicable without intermediate decay of picture quality.

method 5:

With 2-layer coding a fundamentally different approach to VBR coding was developed.

The data is splitted into two bitstreams. The part of the base layer contains all information that is basic for the reconstraction of a frame. This layer is coded with CBR, has a high priority and a guaranteed low cell loss ratio of lets say 10^{-10}. Decoding only the base layer would result in a complete sequence with reduced quality. The enhancement layer contains additional data to lift the decoded sequence to it's final quality level. This second layer is coded with VBR and has low priority. In overload situations the net may discard any of these cells without violation of the contract. Cell loss ratios of up to 10^{-2} or even 10^{-1} are possible.

Coding schemes based on this principles have some advantages. The UPC is not critical for these algorithms, because the base layer is CBR ("peak" rate UPC only) and the enhancement layer has only very little constraints. Almost no UPC should be necessary for it.

When, due to fluctuations of the average activity of the sequence the quality of the base layer is decaying too much, a combination with method 4 should be easy, and most important not critical in timing. Due to the fact that high priority cells in the network are constant rate cells, it is much easier for the network to perform the CAC and guarantee a certain QOS. An other profit from 2-layer coding is the theoretically higher network utilisation. This is due to the fact, that the network may be driven much closer to overload. In an actual overload situation only low priority cells are thrown away. Unfortunately in many cases the higher network utilisation is compensated or even reversed by the smaller efficienty of the 2-layer coding schemes. Further work has to be done to optimize their efficiency.

VBR coding for distribution purposes

An other application of VBR coding could be TV distribution over satellite or cable. Here, the situation is quite different. The number of sources to multiplex is known and constant. Further, the statistical behaviour of each of the sources is completely under control of the channel multiplexing device. Non of the ATM-like UPC functions are necessary, but the codecs might influence each other directly to prevent a channel overflow (fig. 1). This could be done without much additional cost. In that way even a small multiplexing gain could be very advantageous. For instance, a gain of 1.25 would enable 20 instead of 16 sources on a certain channel. If the channel multiplexer has no access to the codecs, a 2-layer like concept might be more benefitial. In case of

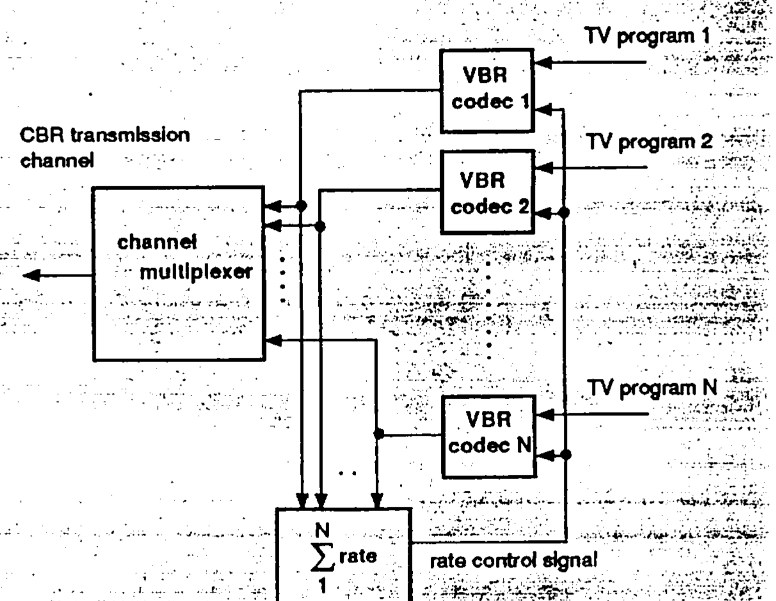
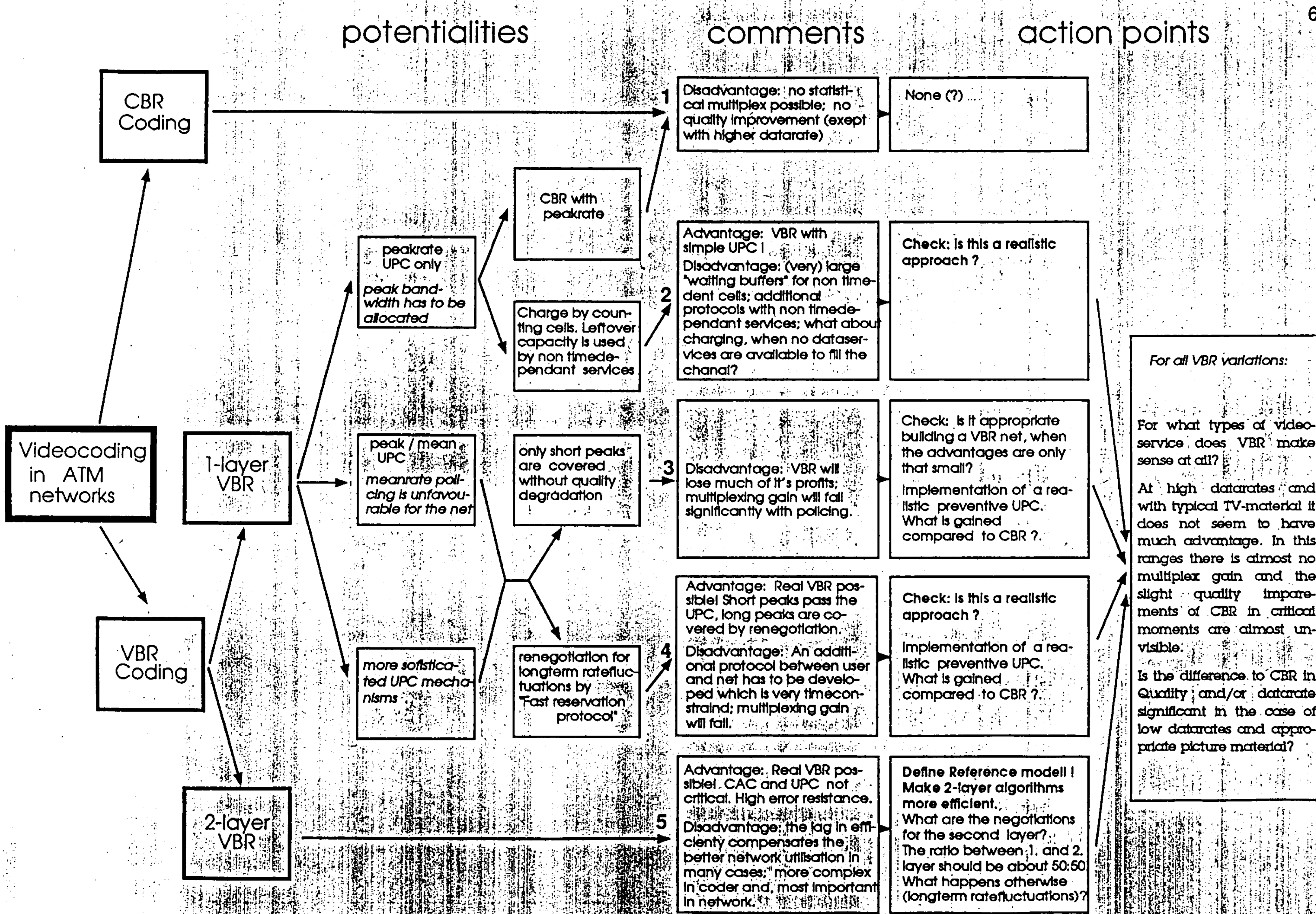


fig. 1: Simplified block diagramm showing the multiplexing of multiple VBR sources to a CBR channel.

excess of the channel capacity low priority information is thrown away.

Although this application seems to be quite advantageous, one should be aware of it's higher complexity on the the decoder side, which is due to the much higher bandwidth of the signal and the necessity of demultiplexing.

Independently of the particular method of VBR-coding attention has to be payed to a common question. From several investigations we know, that at high datarates even uncontrolled VBR does not seem to have much advantage compared to CBR. Further, typical TV (film) sequences with lots of cuts, zooms and pans do not profit very much from VBR. A high gain may be possible in video conference or - telefone scenes. So, an upper bound and/or a restriction to certain kinds of video services might be appropriate for VBR coding (at least in ATM networks).


potentialities
comments
action points
Videocoding in ATM networks
CBR Coding
VBR Coding
1-layer VBR
2-layer VBR
peakrate UPC only
peak bandwidth has to be allocated
peak / mean UPC
meanrate policing is unfavourable for the net
more sofisticated UPC mechanisms
CBR with peakrate
Charge by counting cells. Leftover capacity is used by non timedependant services
only short peaks are covered without quality degradation
renegotiation for longterm ratefluctuations by "Fast reservation protocol"
1
Disadvantage: no statistical multiplex possible; no quality improvement (exept with higher datarate)
None (?) ...
2
Advantage: VBR with simple UPC !
Disadvantage: (very) large "waiting buffers" for non timedent cells; additional protocols with non timedependant services; what abou charging, when no dataservices are available to fill the chanal?
Check: Is this a realistic approach ?
3
Disadvantage: VBR will lose much of it's profits; multiplexing gain will fall significantly with policing.
Check: Is it appropriate building a VBR net, when the advantages are only that small?
Implementation of a realistic preventive UPC.
What is gained compared to CBR ?.
4
Advantage: Real VBR possible! Short peaks pass the UPC, long peaks are covered by renegotiation.
Disadvantage: An additional protocol between user and net has to be developed which is very timeconstraind; multiplexing gain will fall.
Check: Is this a realistic approach ?
Implementation of a realistic preventive UPC.
What is gained compared to CBR ?.
5
Advantage: Real VBR possible! CAC and UPC not critical. High error resistance.
Disadvantage: the lag in efficienty compensates the better network utilisation in many cases;" more complex in coder and, most important in network.
Define Reference modell !
Make 2-layer algorithms more efficient.
What are the negotiations for the second layer?
The ratio between 1. and 2. layer should be about 50:50
What happens otherwise (longterm ratefluctuations)?
For all VBR variations:
For what types of videoservice does VBR make sense at all?
At high datarates and with typical TV-material it does not seem to have much advantage. In this ranges there is almost no multiplex gain and the slight quality imparements of CBR in critical moments are almost unvisible.
Is the difference to CBR in Quality and/or datarate significant in the case of low datarates and appropriate picture material?