ISO/IEC JTC1/SC29/WG11
MPEG93/403
March 1993

Telecommunication Standardization Sector
Study Group 15
Experts Group for ATM Video Coding
(Rapporteur's Group on Part of Q.2/15)

Document AVC-451
March 1993

SOURCE : Japan
TITLE : Clarification of VBV Operation
PURPOSE : Proposal
Relevant sub-group: Video

1.Introduction

It has been agreed to adopt skipped picture(S-picture) to keep the low buffering delay performance. But we can find several ambiguities on VBV operation in case of using S-picture. This contribution intends to clarify these ambiguities. Section 2 is of normative nature while Section 3 is informative.

2.VBV operation taking into account of S-picture

2.1. Usage of the **vbv_delay** value

The purpose of **vbv_delay** is to initialize the VBV operation. It is assumed that there is a picture header detector at the input of VBV buffer. This detector catches a picture header and reads **picture_coding_type** and **vbv_delay** value. If the **picture_coding_type** is NOT skipped, the VBV starts the decoding operation when '**vbv_delay**' seconds have elapsed, in other words **vbv_delay** sets the VBV decoding timing. In case the **picture_coding_type** is "skipped", the VBV ignores the **vbv_delay** value and does not start the decoding operation.

2.2. VBV operation

At each decoding timing, VBV removes from the VBV buffer the coded data of the picture containing the first available non-skipped picture header and any bits before the second non-skipped picture header. Hence sequence and GOP headers preceding the picture and all S-picture headers following the picture are also removed at the same time.

On the other hand, If the second non-skipped picture header does not reside in the VBV buffer, VBV defers the data removal by the next decoding timing.

3.Encoding operation taking into account of S-Picture

3.1. Model for low delay encoding

For stationary signals, the encoder buffer occupancy should stay between zero and Bl. Hence in the stationary state the encoder controls the bit generation for a picture by referring to the estimated encoder buffer occupancy just before encoding the next picture, filling dummy bits if necessary. Again if the encoder buffer occupancy exceeds Bl, necessary number of pictures are skipped.

Fig.1 denotes an example for M=1 Frame coding with Bl corresponding to 1.5 frame time. In this example, until frame F low delay operation is maintained and G is the first frame of a new scene. Three frames H,J and L are skipped and low delay operation is resumed from M in this case. But encoder can decide skipped frames arbitrarily. Furthermore coded frames after G can also generate many bits to invoke subsequent frame skipping, though it causes slower return to the low delay.

3.2. Setting **vbv_delay** at the encoder

Fig.1 also shows the transition of encoder/decoder buffer occupancy. These buffers are cascaded, and we can see when a picture header arrives at input of the decoder buffer. It is the time when a slant broken line crosses the baseline indicated as "transmission channel". In this example, picture header of F,G and I arrive at t1,t2 and t3 respectively.

The **vbv_delay** value is equivalent to the decoder buffer occupancy immediately before a particular picture is removed from the buffer if it does not invoke subsequent picture skipping, and that decoder buffer occupancy corresponds to the encoder buffer occupancy just before starting to encoder the picture. Hence **vbv_delay** value can be set by the encoder before encoding the picture.

This way of setting **vbv_delay** value is extended to a picture that generates so many bits to invoke subsequent picture skipping, picture G in Fig.1. The vbv_delay of picture G indicates the decoder buffer occupancy indicated as G'. Assuming that the decoder starts to receive the bitstream between t1 and t2, the decoder captures the picture header of G at t2, and prompts the decoding at t2+**vbv_delay**. At that time the second non-skipped picture header, which will be received at t3, does not reside in the decoder buffer. Therefore the decoding of picture G is deferred by the next timing as mentioned in VBV operation (Section2.2.).

The relations between encoder and decoder buffer occupancy are shown in Fig.2. The parameters in Fig.2 imply;

- Be_n: encoder buffer occupancy at the time to start encode a picture n,
- Bd_n: decoder buffer occupancy at the time to start decode the picture n,
- Bl: maximum buffer occupancy intended for low delay operation (VBV buffer size B is much bigger than this to accommodate large pictures),
- R: channel bit rate,
- P: picture rate.

and before encoding a picture n, **vbv_delay** time of the picture n is determined as follows.

$$\textbf{vbv\_delay}_n = 90000*Bd_n/R$$

It should be noted that this setting does not involve how many bits are generated by the picture to be coded. Also the **vbv_delay** of S-picture can be arbitrary because it is ignored by VBV.

4.Proposal of modification

Considering above discussion, we propose to modify the description on the semantics for **vbv_delay**, Video Buffer Verifier and Low Delay Coding in TM and WD. Proposed descriptions are shown in ANNEXES.1~3 respectively.

5.Conclusion

Some modifications are proposed to clarify the VBV operation. The current discussion is for the case of constant bit rate(CBR) operation. A regulation of **vbv_delay** for variable bit rate (VBR) operation is also discussed in companion document, AVC-452.

ANNEX.1 Semantics for vbv_delay

vbv_delay--The **vbv_delay** is a 16-bit unsigned integer. For constant bit rate operation the **vbv_delay** is used to set the initial occupancy of the decoder's buffer at the start of operation so that the decoder's buffer does not overflow or underflow. The **vbv_delay** measures the time needed to fill the VBV buffer from an initially empty state at the target bit rate, R, to the correct level when VBV starts to remove the coded data of the picture owning that **vbv_delay**.
The value of **vbv_delay** is the number of periods of the 90kHz system clock that VBV should wait after receiving the final byte of the picture start code. It may be calculated from the state of the VBV as follows:

$$\mathbf{vbv\_delay}_n = 90000 * B_n^* / R$$

where:

$n > 0$

B_n^* = VBV occupancy, measured in bits, at the time to start removal of picture n from the buffer. Any data such as group of picture layer data and sequence header data and the **picture_start_code** that immediately precedes the data elements of picture n are not included in B_n^*.

R = bit rate

NOTE:
(For non-constant bit rate operation, **vbv_delay** will be also provided to initialize decoder operation. A regulation of **vbv_delay** for non-constant bit rate operation is discussed in the companion document AVC-452.)

ANNEX.2 Video Buffer Verifier

APPENDIX C : VIDEO BUFFER VERIFIER

Constant rate coded video bitstreams shall meet constraints imposed through a Video Buffering Verifier (VBV) defined in Section C.1.

The VBV is a hypothetical decoder that is conceptually connected to the output of an encoder. Coded data is placed in the buffer at the constant bit rate that is being used. Coded data is removed from the buffer as defined in Section C.1.4, below. It is a requirement of the encoder (or editor) that the bitstream it produces will not cause the VBV to either overflow or underflow.

C.1 Video Buffering Verifier

1. The VBV and the video encoder have the same clock frequency as well as the same picture rate, and are operated synchronously.

2. The VBV has a receiving buffer of size B, where B is given in the **vbv_buffer_size** field in the sequence header.

3. The VBV has a picture header detector at the input of the receiving buffer. This detector catches picture headers and reads their **picture_coding_type** and **vbv_delay** value.

4. The receiving buffer is initially empty, and it is filled from the bitstream.

5. If the **picture_coding_type** is not skipped, the VBV starts the decoding operation when **vbv_delay** seconds have elapsed. In case of the **picture_coding_type** is skipped, the VBV ignores the **vbv_delay** value and does not start the decoding operation.

6. At each decoding timing, VBV instantaneously removes from the receiving buffer the coded data of the picture containing the first available non-skipped picture header and any data before the second non-skipped picture header. Hence sequence and GOP headers preceding the picture and all S-picture headers following the picture are also removed at the same time.
But if the second non-skipped picture header does not reside in the receiving buffer, the VBV defers the data removal by the next decoding timing.

The receiving buffer occupancy shall lie between zero bits and B bits immediately before removing any data and immediately after that is removed. To meet this requirement the number of bits for the (n+1)'th non-skipped picture d_{n+1} (including any preceding sequence header and group of picture layer data elements) must satisfy:

$$d_{n+1} > B_n + 2R/P - B$$

$$d_{n+1} < B_n + R/P$$

where:

$n >= 0$
B_n is the buffer occupancy just after time t_n
R = bit rate
P = number of pictures per second

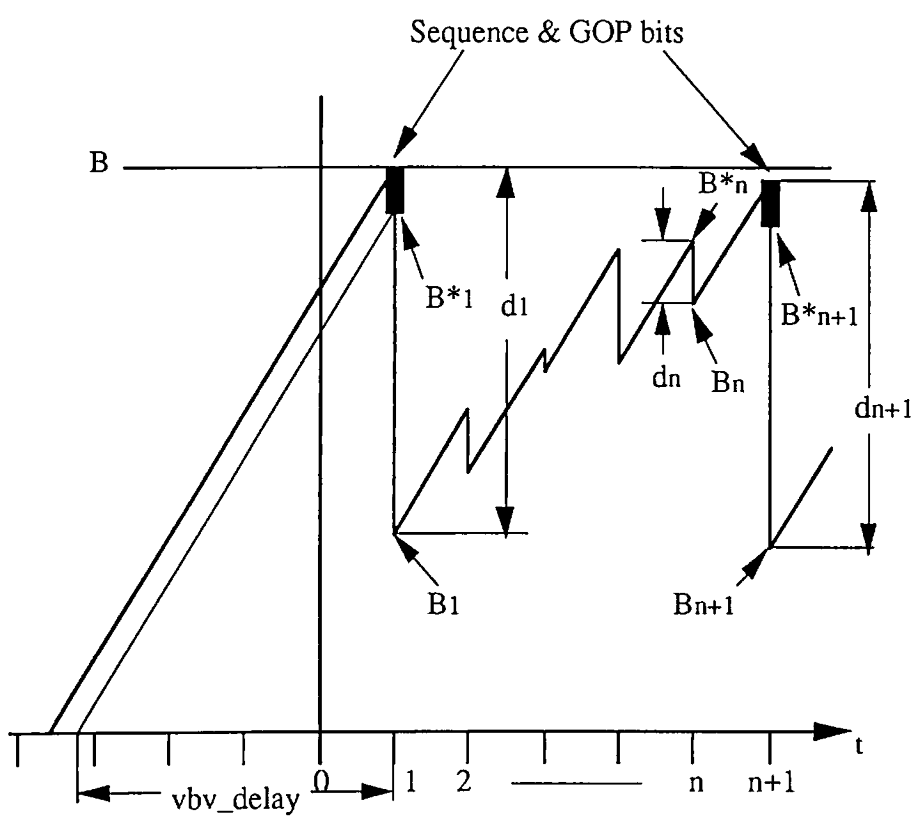


Figure C.1 VBV Buffer Occupancy

ANNEX.3 Low Delay Coding

H.1.2 Handling of scene change to maintain low delay.

A major contribution to the buffer delay is the complete INTRA frames/fields coding. To get around this delay it is necessary for the low delay operational mode to have the possibility of picture skipping. The encoder may decide to skip pictures after a "large" picture due to scene cut.

For the prediction of pictures following skipped one, the re-constructed image of the skipped picture is forced to set to the re-constructed image of the "large" picture.

Model for decoding

It can be found in Annex. C(Video Buffer Verifier).

Model for encoding

Two coding control examples, M=1 frame coding and M=2 field coding, are given to show that there exist methods to meet the VBV requirements.

This encoder operates as follows;

1) It has a transmission buffer of size B. This means that the maximum number of bits per picture should be B. Also a buffer occupancy value Bl (less than B) is set corresponding to the required buffering delay time (Bl/R - 1/P where R is transmission rate, P is picture rate).

2) The buffer is empty when encoding starts.

3) At each encoding timing, the encoder instantaneously encodes the picture and sets **vbv_delay** value in the picture header. **vbv_delay** value of a picture n is determined as follows. The relations between encoder and decoder buffer occupancy are shown in Fig.2. The parameters in Fig.2 imply;

- Be_n: encoder buffer occupancy at the time to start encode a picture n,
- Bd_n: decoder buffer occupancy at the time to start decode the picture n,
- Bl: maximum buffer occupancy intended for low delay operation (VBV buffer size B is much bigger than this to accommodate large pictures),
- R: channel bit rate,
- P: picture rate.

and before encoding a picture n, **vbv_delay** time of the picture n is decided as;

$$\mathbf{vbv\_delay}_n = 90000 * Bd_n / R$$

It should be noted that this setting does not involve how many bits are generated by the picture to be coded. Also the **vbv_delay** of S-picture can be arbitrary because it is ignored by VBV.

4) The buffer content is read out at a constant rate.

5) For stationary signals, the coding operation is controlled so that the buffer occupancy stays between zero and B_l at any time. Hence in the stationary state the encoder controls the bit generation for a picture by referring to the estimated encoder buffer occupancy just before encoding the next picture, filling dummy bits if necessary.

6) If a scene change or forced updating takes place, the buffer occupancy exceeds Bl, and necessary number of pictures are skipped. If there is no picture skipping, the low delay mode encoder controls so that no overflow nor underflow of the buffer with size Bl should take place.

Example for M=1 frame coding

Fig.1 shows an example for M=1 Frame coding with Bl corresponding to 1.5 frame time. In this example, until frame F low delay operation is maintained and G is the first frame of a new scene. Three frames H,J and L are skipped and low delay operation is resumed from M in this case. But encoder can decide skipped frames arbitrarily. Furthermore coded frames after G can also generate many bits to invoke subsequent frame skipping, though it causes slower return to the low delay.

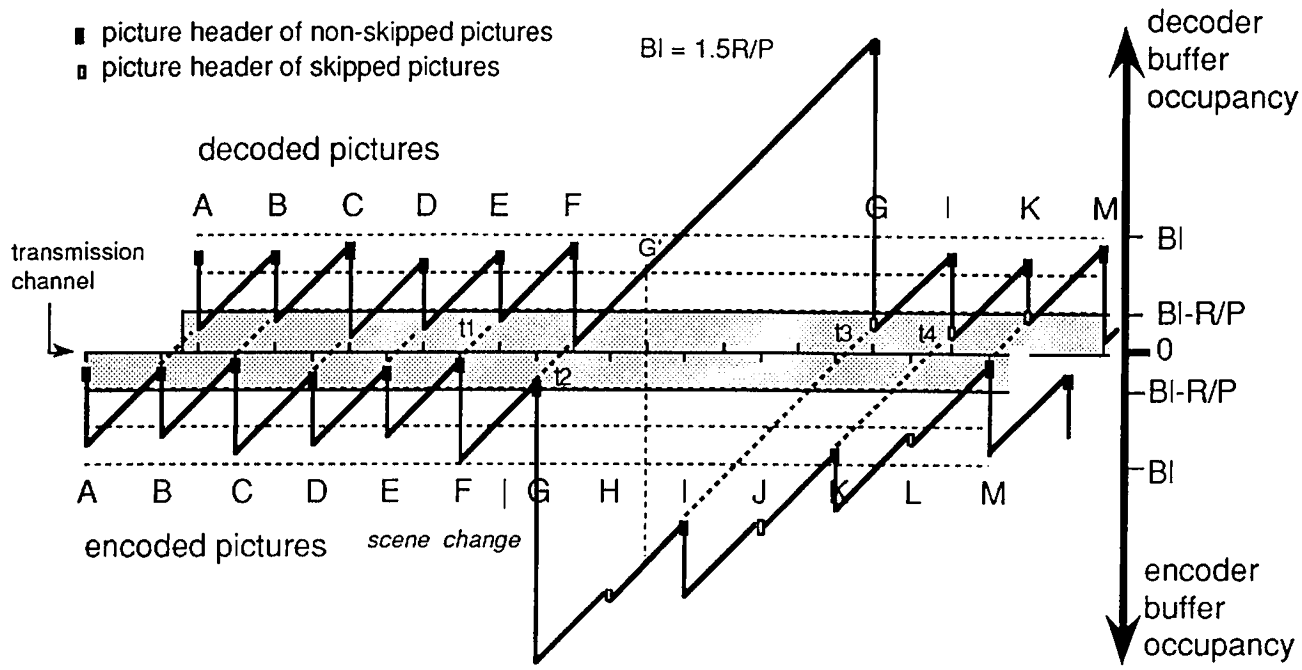
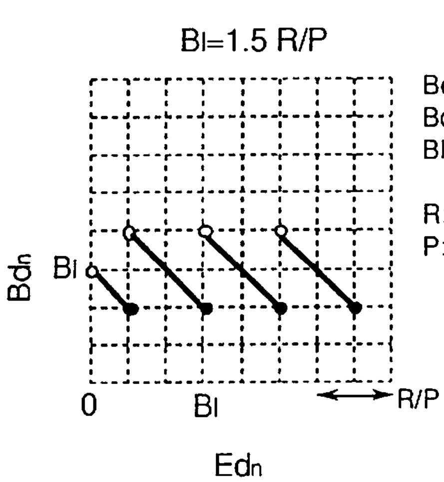


Fig.1 Encoder and decoder buffers for M=1 frame coding



Ben: encoder buffer occupancy at the time to start encode a picture n
Bdn: decoder buffer occupancy at the time to start decode the picture n
BI: maximum buffer occupancy intended for low delay operation
(VBV buffer size B is much bigger than this to accommodate large pictures)
R: channel bit rate
P: picture rate

Fig.2 Relations between encoder and decoder buffer occupancy