|  |  |
| --- | --- |
| **Joint Collaborative Team on Video Coding (JCT-VC)**  **of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11**  7th Meeting: Geneva, CH, 21-30 November, 2011 | Document: JCTVC-G336 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **Unified NAL unit header design for HEVC and its extensions** | | |
| *Status:* | Input Document to JCT-VC | | |
| *Purpose:* | Proposal | | |
| *Author(s) or Contact(s):* | Ying Chen Ye-Kui Wang Marta Karczewicz  5775 Morehouse Dr San Diego, CA 92121 USA | Tel: Email: | 1-858-845-6589 [cheny@qualcomm.com](mailto:cheny@qualcomm.com)  1-858-651-8345 [yekuiw@qualcomm.com](mailto:yekuiw@qualcomm.com)  1-858-658-5673 [martak@qualcomm.com](mailto:martak@qualcomm.com) |
| *Source:* | Qualcomm Incorporated | | |

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Abstract

In AVC extensions, such as SVC and MVC, NAL unit header extensions have to be added. This makes AVC, SVC and MVC substantially different from each other in terms of NAL unit headers. In HEVC, the current NAL unit header contains basic syntax elements similar to AVC NAL unit header, expect for the temporal\_id that indicates the temporal layer of the NAL units when the bitstream is temporally scalable. However, if we follow the design principles in AVC scalable or multiview extension, the HEVC extensions may have similar NAL unit header extensions.

In this proposal, a unified NAL unit header is proposed which can be used for both the HEVC non-scalable bitstreams as well as the scalable bitstreams conforming to the potential scalable or multiview extensions of HEVC. A unified NAL unit header differs from the current HEVC NAL unit header in the following aspects: 1) Fixed NAL unit header length for one whole coded video sequence while the length can vary across different coded video sequences; 2) Efficient coding of the scalability syntax elements in the NAL unit header, and when a particular syntax element it is not needed it is not present.

# Introduction

In AVC, a non-scalable bitstream has an 8-bit NAL unit header. In SVC, a 3-byte NAL unit header extension was introduced to signal the scalability IDs, and in MVC, a different 3-byte NAL unit header extension was introduced to signal the view ID and temporal scalability ID, among others.

The NAL unit header design in AVC and its extensions have the following drawbacks:

* In each AVC extension, a different prefix NAL unit has to be used to carry information that is in the SVC\MVC NAL unit header, but not in the AVC NAL unit header, for the base layer\view.
* In various situations, a scalable bitstream doesn’t have any variation in one of the scalability or view dimensions. However, the syntax element related to that dimension still has to be signaled with a fixed maximum possible length. For example, quality\_id has to be 3-bit even the bitstream just uses 2-layer spatial scalability. SVC and MVC might be jointly supported in one HEVC amendment, wherein a NAL unit header might cost 6 bytes.

# Proposal

To provide high level syntax hooks in the “base” HEVC design for its possible extensions and also to provide efficient NAL unit header design in a unified fashion, it is proposed that in the NAL unit header, a set of syntax elements are signaled, each of which may have a variable length in different coded video sequence and the length can be zero.

This is realized by A NAL unit header map that specifies the layout of NAL unit header fields, including the presence and the length (when present) of the scalability and view identifiers. The NAL unit header map is included in a NAL unit header map parameter set (NPS). The design also enables backward compatible future extensions.

## Concept

In the NPS, each scalability or view dimension, such as spatial scalability, temporal scalability, quality scalability or view, corresponding to a syntax element in the NAL unit header has its length specified.

If a specific dimension doesn’t change in the whole coded video sequence, then the length of the corresponding syntax element is 0, meaning that the syntax element is not present in the NAL unit header, while a default value is derived for all NAL units in the coded video sequence.

Each syntax element in the NAL unit header is to be signaled in a more compact fashion, such that if there are only M possible values of a syntax element, but the specific values can take N bits, which is much larger than 1 << Ceil(log2(M+1)), the signaling can be further optimized by only signal an index to the instances. For example, the view\_id in multiview extension costs 10 bits (as in AVC multiview extension), however if a selected set of views have instances of view\_id values as e.g., 45, 50, 55, 60, then only a 2-bit view\_idx is signaled in the NAL unit header.

An example of the set of scalability or view dimension identifiers included in the NAL unit header is shown in Figure 1, wherein priority\_id, temporal\_id, dependency\_id, quanlity\_id and view\_idx are all present in the NAL unit header.



**Figure 1: Scalability or view dimension IDs present in the NAL unit header**

The following are a few examples of such a NAL unit header design for different scenarios, all without priority\_id signaled:

* A scalable bitstream with QVGA to VGA spatial scalability while each dependency layer has 3 temporal layers. In such a case, only 3 bits are needed to signal the scalability and/or view dimensions in the NAL unit header:
  + priority\_id: not present
  + temporal\_id: 2 bits
  + dependency\_id: 1 bit
  + quality\_id: not present
  + view\_idx: not present
* A stereoscopic bitstream with multiple (2) spatial layers for each view, each of which has up to 3 temporal layers. In such a case, only 4 bits are needed to signal the scalability and/or view dimensions in the NAL unit header:
  + priority\_id: not present
  + temparal\_id: 2 bits
  + dependency\_id: 1 bit
  + quality\_id: not present
  + view\_idx: 1 bit
* A multiview bitstream contains 8 views, each has two quality layers. The bitstream is also coded with hierarchical B prediction structure with a GOP Size 16 (thus 4 temporal layers) In such a case, 7 bits are need to signal the scalability and/or view dimensions in the NAL unit header
  + priority\_id: not present
  + temparal\_id: 3 bits
  + dependency\_id: not present
  + quality\_id: 1 bit
  + view\_idx: 3 bits

## Syntax

### NAL unit header map parameter set RBSP syntax

|  |  |
| --- | --- |
| nal\_unit\_header\_map( ) { | Descriptor |
| **nal\_unit\_header\_map\_id** | u(8) |
| **priority\_id\_len** | u(3) |
| **temporal\_id\_len** | u(3) |
| **dependency\_id\_len** | u(3) |
| **quality\_id\_len** | u(3) |
| **view\_idx\_len** | u(4) |
| **reserved\_flags\_len** | u(4) |
| view\_idx2id\_table( ) |  |
| **nps\_extension\_flag** | u(1) |
| if( nps\_extension\_flag ) |  |
| while( more\_rbsp\_data( ) ) |  |
| **nps\_extension\_data\_flag** | u(1) |
| rbsp\_trailing\_bits( ) |  |
| **}** |  |

### View index to ID table syntax

|  |  |
| --- | --- |
| view\_idx2id\_table( ) { | Descriptor |
| **view\_cnt** | u(v) |
| if( view\_cnt ) |  |
| for( i=0; i< view\_cnt; i++ ) |  |
| **view\_id[** i **]** | u(v) |
| } |  |

### Sequence parameter set RBSP syntax

|  |  |
| --- | --- |
| seq\_parameter\_set\_rbsp( ) { | Descriptor |
| **profile\_idc** | u(8) |
| **reserved\_zero\_8bits** /\* equal to 0 \***/** | u(8) |
| **level\_idc** | u(8) |
| **nal\_unit\_header\_map\_id** | u(8) |
| **seq\_parameter\_set\_id** | ue(v) |
| **…** |  |
| rbsp\_trailing\_bits( ) |  |
| } |  |

### NAL unit syntax

|  |  |
| --- | --- |
| nal\_unit( NumBytesInNALunit ) { | Descriptor |
| **forbidden\_zero\_bit** | f(1) |
| **nal\_ref\_flag** | u(1) |
| **nal\_unit\_type** | u(6) |
| NumBytesInRBSP = 0 |  |
| nalUnitHeaderBytes = 1 |  |
| if( nal\_unit\_type != 10 && nal\_unit\_type != 5 ) {  // not NAL unit header map NAL unit or SPS NAL unit, the NAL unit types of // nal\_unit\_header\_map and SPS NAL unit are 10 and 5, respectively |  |
| if( priority\_id\_len ) |  |
| **priority\_id** | u(v) |
| if( temporal\_id\_len ) |  |
| **temporal\_id** | u(v) |
| **reserved\_one\_bit** | u(1) |
| if( dependency\_id\_len ) |  |
| **dependency\_id** | u(v) |
| if( quality\_id\_len ) |  |
| **quality\_id** | u(v) |
| **reserved\_one\_bit** | u(1) |
| if( view\_idx\_len ) |  |
| **view\_idx** | u(v) |
| if( reserved\_flags\_len ) |  |
| **reserved\_flags** | u(v) |
| m = priority\_id\_len + temporal\_id\_len + dependency\_id\_len +  quality\_id\_len + view\_idx\_len + reserved\_flags\_len + 2 |  |
| if( ( ( m + 7 >> 3) << 3 ) – m ) |  |
| **reserved\_bits** | u(v) |
| nalUnitHeaderBytes += ( ( m + 7 ) >> 3 ) |  |
| } |  |
| for( i = nalUnitHeaderBytes; i < NumBytesInNALunit; i++ ) { |  |
| if( i + 2 < NumBytesInNALunit && next\_bits( 24 ) = = 0x000003 ) { |  |
| **rbsp\_byte[** NumBytesInRBSP++ **]** | b(8) |
| **rbsp\_byte[** NumBytesInRBSP++ **]** | b(8) |
| i += 2 |  |
| **emulation\_prevention\_three\_byte** /\* equal to 0x03 \*/ | f(8) |
| } else |  |
| **rbsp\_byte[** NumBytesInRBSP++ **]** | b(8) |
| } |  |
| } |  |

## Semantics

### NAL unit header map parameter set RBSP semantics

A NAL unit header map parameter set (NPS) specifies a NAL unit header map. In each coded video sequence, only and only one NAL unit header map parameter shall be active. The NAL unit type for this RBSP is set to 10 based on the NAL unit type allocation as described in JCTVC-G331.

**nal\_unit\_header\_map\_id** specifies the identification of the NAL unit header map parameter set. In each coded video sequence, one and only one NAL unit header map parameter shall be active.

**priority\_id\_len** specifies the number of bits used to represent the priority\_id syntax element. When prioriy\_id\_len is equal to 0, each VCL NAL unit is inferred to have priority\_id equal to 0. The number of priority layers supported in a coded video sequence referring to the NAL unit header map parameter set shall be in the range of 1 to 2 << priority\_id\_len, inclusive.

**temporal\_id\_len** specifies the number of bits used to represent the temporal\_id syntax element. When temporal\_id\_len is equal to 0, no temporal scalability is supported and each VCL NAL unit is inferred to have temporal\_id equal to 0. The number of temporal layers supported in a coded video sequence referring to the NAL unit header map parameter set shall be in the range of 1 to 2 << temporal\_id\_len, inclusive.

**dependency\_id\_len** specifies the number of bits used to represent the dependency\_id syntax element. When dependency\_id\_len is equal to 0, no spatial scalability or coarse-grain scalability is supported and each VCL NAL unit is inferred to have dependency\_id equal to 0. The number of dependency layers supported in a coded video sequence referring to the NAL unit header map parameter set shall be in the range of 1 to 2 << dependency\_id\_len, inclusive.

**quality\_id\_len** specifies the number of bits used to represent the quality\_id syntax element. When quality\_id\_len is equal to 0, no quality/SNR scalability is supported and each VCL NAL unit is inferred to have quality\_id equal to 0. The number of quality layers supported in a coded video sequence referring to the NAL unit header map parameter set shall be in the range of 1 to 2 << quality\_id\_len, inclusive.

**view\_idx\_len** specifies the number of bits used to represent the view\_idx syntax element. When view\_cnt\_len is equal to 0, only one view is supported and each VCL NAL unit is inferred to have view\_id and view order index both equal to 0. The number of views supported in a coded video sequence referring to the NAL unit header map parameter set shall in the range of 1 to 2 << view\_idx\_len, inclusive.

**reserved\_flags\_len** specifies the number of bits used to represent the reserved\_flags syntax element.

**nps\_extension\_flag** equal to 0 specifies that no nps\_extension\_data\_flag syntax elements are present in the NAL unit header map parameter set RBSP syntax structure. nps\_extension\_flag shall be equal to 0 in bitstreams conforming to this Recommendation | International Standard. The value of 1 for nps\_extension\_flag is reserved for future use by ITU-T | ISO/IEC. Decoders shall ignore all data that follow the value 1 for nps\_extension\_flag in a NAL unit header map parameter set NAL unit.

**nps\_extension\_data\_flag** may have any value. It shall not affect the conformance to profiles specified in this Recommendation | International Standard.

### View index to ID table semantics

This table specifies the map of each view index value to a view identifier value. A view index value is signaled in the NAL unit header and the corresponding view identifier is specified in this table.

**view\_cnt** specifies the maximum number of views included in a coded video sequence referring to the NAL unit header map parameter set.

**view\_id[** i **]** specifies the view identifier of a NAL unit with view index equal to i.

### Sequence parameter set RBSP semantics

The NAL unit type for this RBSP is set to 5 based on the NAL unit type allocation as described in JCTVC-G331.

**nal\_unit\_header\_map\_id** specifies the identifier of the NAL unit header map parameter set referred to by the sequence parameter set.

### NAL unit semantics

The NAL unit header shall contain no consecutive 3 bytes that are equal to 0x000000, 0x000001, 0x000002, or 0x000003.

The semantics of **priority\_id** is similar as the same syntax element in SVC. The number of bits used to represent priority\_id shall be priority\_id\_len.

The semantics of **temporal\_id** is the same as in HEVC WD4d1. The number of bits used to represent temporal\_id shall be temporal\_id\_len.

**reserved\_one\_bit** shall be equal to 1. The value 0 for reserved\_one\_bit may be specified by future extension of this Recommendation | International Standard. Decoders shall ignore the value of reserved\_one\_bit.

The semantics of **dependency\_id** is the same the same syntax element as in H.264/SVC. The number of bits used to represent dependency\_id shall be dependency\_id\_len.

The semantics of **quality\_id** is the same as the same syntax element in H.264/SVC. The number of bits used to represent quality\_id shall be quality\_id\_len.

**view\_idx** specifies the view order index for a view. The number of bits used to represent view\_idx shall be view\_idx\_len.

**reserved\_flags** consists of zero or more bits, wherein each bit shall be equal to 1. Other values for reserved\_flags may be specified by future extension of this Recommendation | International Standard. Decoders shall ignore the value of reserved\_flags. The number of bits used to represent reserved\_flags shall be reserved\_flags\_len.

**reserved\_bits** consists of zero more bits, wherein each bit shall be equal to 1. Other values for reserved\_bits may be specified by future extension of this Recommendation | International Standard. Decoders shall ignore the value of reserved\_bits. The number of bits used to represent reserved\_bits shall be ( ( m + 7 >> 3) << 3 ) – m, where m is equal to the sum of priority\_id\_len, temporal\_id\_len, dependency\_id\_len, quality\_id\_len, view\_idx\_len, reserved\_flags\_len and 2.

# Patent rights declaration(s)

**Qualcomm Incorporated may have current or pending patent rights relating to the technology described in this contribution and, conditioned on reciprocity, is prepared to grant licenses under reasonable and non-discriminatory terms as necessary for implementation of the resulting ITU-T Recommendation | ISO/IEC International Standard (per box 2 of the ITU-T/ITU-R/ISO/IEC patent statement and licensing declaration form).**