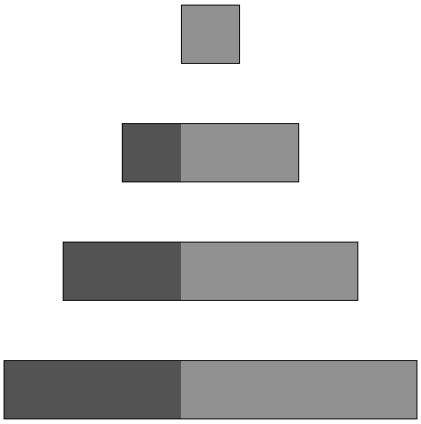
DIGITAL SUBBAND VIDEO

VERSION 2.8

REVISION 2

June, 2025

2024-2025 EMMIR
ENVEL GRAPHICS



Contact Information:

GitHub : https://github.com/LMP88959
YouTube: https://www.youtube.com/@EMMIR_KC/videos
Discord: https://discord.com/invite/hdYctSmyQJ

LICENSING

Specification License

Software License

The accompanying software was designed and written by EMMIR, 2024 of Envel Graphics.
If you release anything with it, a comment in your code/README saying where you got this code would be a nice gesture but it’s not mandatory.

Please note this document is not intended to be formal or comprehensive. It would be virtually impossible to create a fully functional DSV2 decoder explicitly using this document as reference.

CONVENTIONS

Pseudo code is often written in the C or a pseudo-C language.
All data types are assumed to be 32-bit integers unless otherwise specified.

A prefix of 0x for a number means that the number is in hexadecimal format. For example, 0x55 is 85 in decimal.

DEFINITIONS

Block: a rectangular subregion of a plane of a frame.

Chroma/UV: the two planes of a frame that represent the color of the image at every point.

Frame: three two-dimensional grids of pixels representing the video at a certain point in time.

Half-Pixel Filter: a filter used to generate a new image as though the original image had been shifted over by half a pixel.

In-Loop Filter: Refers to using the filtered frame as the reference for the next frame rather than a post-processing filter.

Interleaved Exponential-Golomb Code: A form of variable length coding which can be more efficiently decoded due to its structure.

Inter Frame: a frame of video that relies on another frame to be correctly reconstructed.

Intra Frame: a frame of video that does not rely on another frame to be correctly reconstructed.

Luma/Y: the plane of a frame that represents the brightness of the image at every point within it.

Motion Compensation: applying temporal prediction in order to reconstruct an inter frame.

Plane: a single component of the video, also known as a channel. Luma is its own plane, chroma consists of two planes: U, and V.

Variable Length Code: A way to represent an arbitrarily large integer in a variable number of bits.

TABLE OF CONTENTS

A. INTRODUCTION

A.1 Description

Digital Subband Video 2 (DSV2) is a digital video compression specification designed for efficient software decoding and playback on consumer hardware while offering compression performance comparable to Moving Picture Experts Group's MPEG-4 Part 2 and Part 10 video standards. DSV2 was designed for medium-low to high bit rates.

DSV2 is highly portable and does **not** require use of:

- data types with widths greater than 32 bits
- floating point data types or literals
- a specific operating system, processor
- any 3rd party libraries

A.2 Brief Technology Overview

DSV2 is a lossy/lossless hybrid video codec[1] which utilizes multiple subband filters and block based motion compensation to achieve spatial and temporal compression.

Overview:

- operates on YCC video (luma with two chroma components)
  - supports chroma subsampling of 4:4:4, 4:2:2, 4:2:0, 4:1:1, 4:1:0
  - only 8 bits of depth per component supported
  - no explicit support for interlaced video
- frame sizes must be divisible by two
- subband transform
  - two-dimensional Haar + smoothing filter
  - various hybrids of 3 and/or 5 tap filters with varying coefficients
  - lossy asymmetric subband transform
  - full decomposition (recurse on LL band until 1 pixel remaining)
- block motion compensation
  - half and quarter pixel precision
  - block size can vary but all blocks in one frame must be the same size
  - intra blocks can be split into 4 sub-blocks
  - in-loop filtering
- adaptive quantization
- intra (I) frames and inter (P) frames with variable length closed GOP (group of pictures)

[1] Hybrid refers to the use of image transform coding and motion estimation/compensation. This is in contrast to coding a video as if it were a 3D signal.

A.3 Improvements and New Features

DSV2's fundamentals and architecture are very similar to DSV1. The substantial video coding improvements come mostly from a few key new features.

New Features:

- in-loop filtering (deblock and sharpening) after motion compensation
- intra frame filtering and deringing
- more adaptive quantization potential
- skip blocks for temporal stability
- better motion compensation through Expanded Prediction Range Mode (EPRM)
- new subband filters + support for adaptive subband filtering
- quarter pixel compensation with adaptive filtering
- psychovisual optimizations in the codec and encoder design
  - spatial and temporal masking
  - ringing transform
- adaptive Rice coding

B. BITSTREAM DECODING

The fundamental unit of digital information is a binary digit, or bit, which can be either a 0 or 1. A byte is defined as 8 bits.
DSV operates on bits exclusively and only uses these types of binary encoding:

Encoding type	Description
Raw	The bits stored directly represent the binary digits of a number.
Interleaved Exponential-Golomb Code (UEG)	A method of encoding arbitrarily large positive (including zero) integers in a variable number of bits.
Signed Interleaved Exponential-Golomb Code (SEG)	A method of encoding arbitrarily large integers in a variable number of bits.
Non-zero Interleaved Exponential-Golomb Code (NEG)	A method of encoding arbitrarily large non-zero integers in a variable number of bits.
Adaptive Rice Code (URC)	A method of encoding arbitrarily large positive (including zero) integers in a variable number of bits.
Non-zero Adaptive Rice Code (NRC)	A method of encoding arbitrarily large non-zero integers in a variable number of bits.

Some parts of the DSV bitstream contain a form of data encoding called Zero Bit Run-Length-Encoding (ZBRLE). ZBRLE operates on bits and consists of a concatenated sequence of UEG codes. The value of each decoded UEG code is the number of zero bits that it represents. If no zero bits remain, the decoded bit is a one. Decoding a ZBRLE bit is described in the following pseudo code:

```
if (number_of_zeros_remaining[2] == 0) {
    number_of_zeros_remaining = read_UEG();
    return (number_of_zeros_remaining == 0 ? 1 : 0);
}
number_of_zeros_remaining--;
return (number_of_zeros_remaining == 0 ? 1 : 0);
```

A DSV stream consists of a number of packets. Packets contain a header and a payload. The data in the following tables in each subsection is sequential in the bitstream.

[2] *number_of_zeros_remaining* is a state variable associated with the ZBRLE stream being decoded.

B.1 Packet Header

Data	Contains	Description
8 bits	ASCII character ‘D’ Hex 0x44	First character of the Four Character Code (FourCC)
8 bits	ASCII character ’S’ Hex 0x53	A method of encoding arbitrarily large positive (including zero) integers in a variable number of bits.
8 bits	ASCII character ‘V’ Hex 0x56	A method of encoding arbitrarily large integers in a variable number of bits.
8 bits	ASCII character ‘2’ Hex 0x32	A method of encoding arbitrarily large non-zero integers in a variable number of bits.
8 bits	Minor version number	Minor version number of the DSV specification the encoder adhered to.
8 bits	Packet type	A bit pattern which describes the contents packet.
32 bits	Previous link offset	Length (in bytes) of the previous packet.
32 bits	Next link offset	Length (in bytes) of the current packet.

B.1.1 Packet Type

Packet Type	Description
Equal to 0x00	Packet is metadata.
Equal to 0x10	Packet is an end of stream marker.
Packet Type bit 0x04 is set	Packet is a picture. If Packet Type bit 0x01 is set, it has a reference picture. If Packet Type bit 0x02 is set, it is a reference picture.

B.2 Packet Payload

B.2.1 Metadata Packet

Data	Contains	Description
UEG encoded	Width of video frame	Horizontal resolution in pixels.
UEG encoded	Height of video frame	Vertical resolution in pixels.
UEG encoded	Chroma subsampling	4:4:4 = 0x00 4:2:2 = 0x04 4:2:0 = 0x05 4:1:1 = 0x08 4:1:0 = 0x0a
UEG encoded	Frames per second (FPS) numerator	Used in combination with the FPS denominator to define the (potentially fractional) frame rate of the video. FPS = numerator / denominator
UEG encoded	Frames per second (FPS) denominator	Used in combination with the FPS numerator to define the (potentially fractional) frame rate of the video. FPS = numerator / denominator
UEG encoded	Aspect ratio numerator	Used in combination with the aspect ratio denominator to define the aspect ratio of the video. Aspect ratio = numerator / denominator
UEG encoded	Aspect ratio denominator	Used in combination with the aspect ratio numerator to define the aspect ratio of the video. Aspect ratio = numerator / denominator
UEG encoded	Inter sharpen flag	Flag stating whether the video uses inter frame sharpening

B.2.2 End of Stream Packet

Data	Contains	Description
None	Nothing	End of stream packet has no payload.

B.2.3 Picture Packet

Data	Contains	Valid Range	Description
32 bits	Frame number	All positive integers, including zero.	ID of the frame. Ideally sequential/ chronological.
UEG encoded	log2(block_width) - 4	[0, 1]	Motion block width can be calculated by multiplying 16 by two to the power of the decoded number. In C: `block_width = 16 << decoded`
UEG encoded	log2(block_height) - 4	[0, 1]	Motion block height can be calculated by multiplying 16 by two to the power of the decoded number. In C: `block_height = 16 << decoded`
1 bit	Stable statistic	[0, 1]	1 if zeros are used as the end-of-run symbol 0 if ones are used as the end-of-run symbol
1 bit	**If intra frame,** maintain statistic. **If inter frame**, mode statistic.	[0, 1]	1 if zeros are used as the end-of-run symbol 0 if ones are used as the end-of-run symbol
1 bit	**If intra frame,** ringing statistic. **If inter frame**, EPRM statistic.	[0, 1]	1 if zeros are used as the end-of-run symbol 0 if ones are used as the end-of-run symbol
1 bit	Intra / inter frame filter flag	[0, 1]	1 if filtering is enabled for the frame 0 if filtering is disabled for the frame
12 bits	Quantization parameter ***Q***	[0, 4095]	Frame quantization parameter. Note that a ***Q*** of zero indicates lossless mode

The number of motion blocks in the horizontal and vertical direction can be derived like so:

```
horiz_blocks = (frame_width + block_width - 1) / block_width
vert_blocks = (frame_height + block_height - 1) / block_height
```

Where *frame_width* and *frame_height* come from the metadata packet.

B.2.3.1 Stability Blocks

Every picture has a stability block subsection, in the case of P frames, it represents skipped blocks. Note that the ZBRLE bits should be modified appropriately based off of the corresponding statistic read in the **B.2.3 Picture Packet** table.

Data	Contains	Valid Range	Description
UEG encoded Align to next byte afterwards.	Length of this subsection in bytes.	All positive integers, including zero.	Number of bytes to skip to reach the next subsection of the picture packet.
ZBRLE encoded	One bit for each motion block in the frame. In I frames: used for adaptive quantization. In P frames: used to flag a block as a skip block.	N/A	Compute (horiz_blocks * vert_blocks) to determine how many ZBRLE bits must be read.

B.2.3.2 Ringing Blocks

Only I frames have a ringing blocks subsection. Note that the ZBRLE bits should be modified appropriately based off of the corresponding statistic read in the **B.2.3 Picture Packet** table.

Data	Contains	Valid Range	Description
UEG encoded Align to next byte afterwards.	Length of this subsection in bytes.	All positive integers, including zero.	Number of bytes to skip to reach the next subsection of the picture packet.
ZBRLE encoded	One bit for each motion block in the frame.	N/A	Compute (horiz_blocks * vert_blocks) to determine how many ZBRLE bits must be read.

B.2.3.3 Maintain Blocks

Only I frames have a maintain blocks subsection. Note that the ZBRLE bits should be modified appropriately based off of the corresponding statistic read in the **B.2.3 Picture Packet** table.

Data	Contains	Valid Range	Description
UEG encoded Align to next byte afterwards.	Length of this subsection in bytes.	All positive integers, including zero.	Number of bytes to skip to reach the next subsection of the picture packet.
ZBRLE encoded	One bit for each motion block in the frame.	N/A	Compute (horiz_blocks * vert_blocks) to determine how many ZBRLE bits must be read.

B.2.3.4 Motion Data

Only exists for pictures that have a reference (inter frames). Align to next byte before starting the decoding process. The subsections occur in order, data for each motion block is not interleaved. First, a UEG encoded integer is read before each subsection. This integer is the length in bytes of the subsection that follows it. The stream is aligned to the next byte after each UEG decoded. Note that the ZBRLE bits should be modified appropriately based off of the corresponding statistic read in the **B.2.3 Picture Packet** table.

Subsection	Data	Contains	Description
Mode	ZBRLE encoded	Block inter/ intra modes.	0x00 = block is INTER 0x01 = block is INTRA
EPRM	ZBRLE encoded	Expanded Prediction Range Mode flags.	0x00 = block is in limited prediction range mode 0x01 = block is in expanded prediction range mode
Motion vector X components	SEG encoded	X component of each block's motion vector.	If SKIP then set component to zero and do not read any data. If NOT SKIP, read the data and reconstruct the actual component by adding its prediction to the decoded value.
Motion vector Y components	SEG encoded	Y component of each block's motion vector.	If SKIP then set component to zero and do not read any data. If NOT SKIP, read the data and reconstruct the actual component by adding its prediction to the decoded value.
Intra sub-block masks	Raw - Described below	Bit masks of the sub-blocks.	Exists only if block is INTRA.

<u>Intra Block Motion Vectors</u>

Intra blocks have motion vectors as well but they do not have any subpixel precision. The X and Y components need to be scaled up by 4 before they can be used. This same logic applies to the vector prediction for the intra block, the predicted components need to be scaled down by 4 before adding it to the decoded values.

Motion Vector Prediction

Motion vector components for a given block are predicted from the vector components of its left, upper, and upper left neighbors in the block grid. If the current block is on the left edge, top edge, or top left, the unavailable neighbors' component used for prediction will be zero.

The prediction using the three neighbors is obtained by:

```
motion_vector_prediction(LEFT, TOP, TOP_LEFT)
{
    dif = LEFT + TOP - TOP_LEFT;
    lft = absolute_value(dif - LEFT);
    top = absolute_value(dif - TOP);
    if (lft < top) {
        return LEFT;
    }
    return TOP;
}
```

Intra Sub-Block Mask Decoding

Intra blocks are split into four sub-blocks. Each sub-block can then either be coded as intra or inter.

1	2
4	8

Mask value of each sub-block

Decoding the mask is done by first reading a bit. If the bit was a 1 then all the sub-blocks are marked as intra, otherwise 4 bits are read and assigned to the mask. Pseudo code for the described behavior:

```
if (read_bit() == 1) {
    submask = ALL_INTRA;
} else {
    submask = read_4_bits();
}
```

Sub-blocks can be optionally predicted with a DC value transmitted in the bitstream. They are decoded in a similar fashion.

```
if (read_bit() == 1) {
    dc = read_8_bits() | 0x100;
} else {
    dc = 0;
}
```

B.2.3.5 Image Data

Every picture has an image data subsection.

Plane Decoding

Data	Contains	Description
32 bits	Length in bytes of encoded plane.	Amount of bytes to skip to get to the next color plane.
Complex - Coefficient Decoding	Encoded coefficients of plane.	Subband coefficients of the color plane.

The above is performed for each of the three color frames, making sure to byte align the stream after each plane.

Coefficient Decoding

Data	Contains	Description
SEG encoded	LL coefficient of the entire frame.	All the low pass energy of the image. It is not quantized and often has a very large magnitude.
Complex - HZCC	Encoded coefficients of plane.	Subband coefficients of the color plane.
8 bits	End of Plane Symbol Hex 0x55	Check decoded value matches 0x55 after HZCC decoding to see if there was an error in the plane's bit stream.

Decoding Hierarchical Zero Coefficient Coding (HZCC) Data

Align the stream to the next byte before beginning.

Data	Contains	Description
32 bits	Number of runs.	Total number of zero runs in the coefficients.
UEG encoded	Length of first zero run.	How many of the first coefficients are zero.

Begin loop through the 'LL' subband. If there are no more zeros left in the current run, read the following data to get the next run and then continue the loop.

Data	Contains	Description
UEG encoded	Length of next zero run.	How many of the next set of coefficients are zero.
NEG encoded	Non-zero coefficient.	Quantized, needs to be dequantized unless in lossless mode.

After decoding the 'LL' band, begin decoding the 3 highest frequency subbands in order from lowest to highest:

Data	Contains	Description
UEG encoded	Length of next zero run.	How many of the next set of coefficients are zero.
NRC encoded	Non-zero coefficient.	Quantized, needs to be dequantized unless in lossless mode.

Once all subbands are decoded, align to the next byte and continue with the error check in the **Coefficient Decoding** table. Set the first coefficient of the plane to be the LL coefficient decoded in the **Coefficient Decoding** table.

Repeat for each plane until image data is fully decoded.

NOTE: **all** subband coefficients must be stored as signed 32-bit integers.

At this point, no more data needs to be read from the stream for this frame.

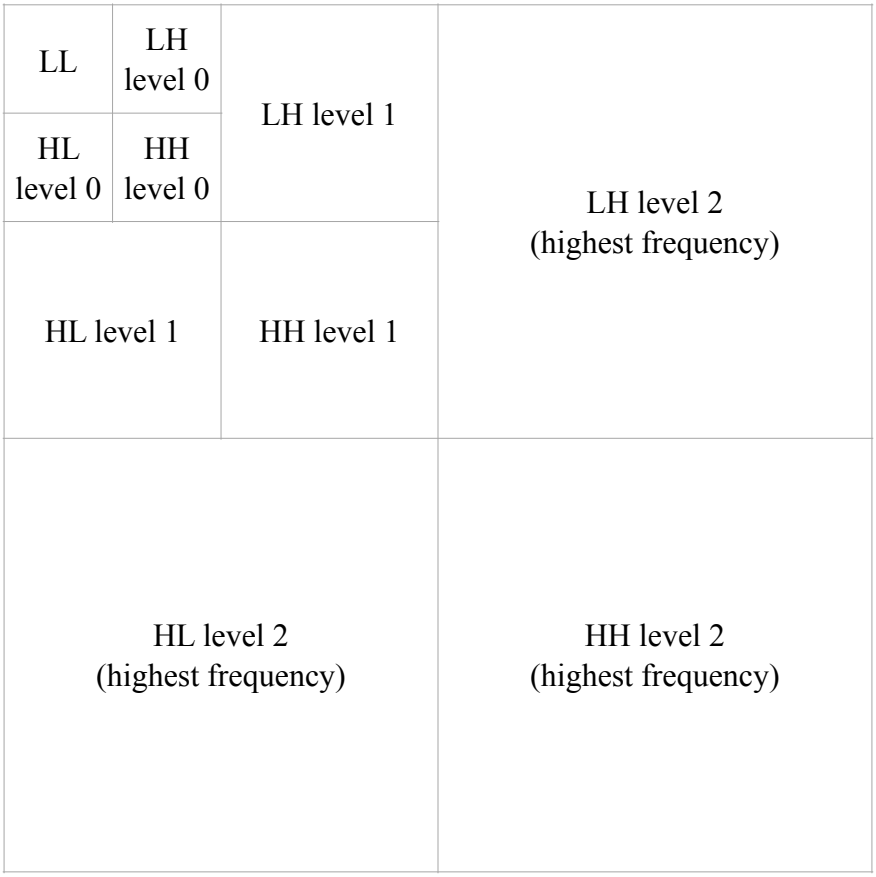
C. IMAGE RECONSTRUCTION

C.1 Subband Order and Traversal

Only the three highest frequency subbands are treated specially in DSV. The other lower frequency subbands get grouped together into the LL region in the context of HZCC and quantization. The subbands are traversed in raster scan order. Raster scan order is left to right, top to bottom. Pseudo code for such a traversal is given below:

```
for (y = top; y < bottom; y++) {
    for (x = left; x < right; x++) {
        /* data is at [x, y] */
    }
}
```

X increases horizontally, Y increases vertically



DSV subband structure in 2D

NOTE: when computing subband dimensions through subdivision, make sure to round up to the next integer.

C.2 Dequantization

In DSV, each level of the structure is quantized in a different fashion and each block is quantized in a potentially different fashion.
The quantization parameter ***Q*** is passed into the HZCC decoder and scaled by 3 / 2. The minimum quantization parameter ***MINQUANT*** is **8** for all levels.

Please refer to reference implementation for more information.

NOTE the fixed point precision for determining what block a pixel (x, y) lies in is **14**.

C.2.1 Dequantization Functions

Please refer to reference implementation for more information.

C.2.2 Quantization Parameter Derivation

Please refer to reference implementation for more information.

C.2.3 Subband Dequantization

Lossless mode does not perform dequantization.

Please refer to reference implementation for more information.

C.3 Subband Transforms

DSV performs a full decomposition. A full decomposition is a recursive subband transform on the LL subband of the structure until the LL subband is a single pixel.

DSV uses a few different transforms, the *Haar Transform* (HT), 3 or 5 tap filters, and the ASF93 (asymmetric subband filter).

The ***MIN*** function is defined as:

```
MIN(a, b) ((a) < (b) ? (a) : (b))
```

The ***MAX*** function is defined as:

```
MAX(a, b) ((a) > (b) ? (a) : (b))
```

The functions ***ROUND2*** and ***ROUND4*** simply divide the argument by the number in the function's name while also rounding away from zero (properly accounting for negatives).
For example:

```
ROUND4(V)
    if (V < 0) {
        return -(((-V) + 2) / 4);
    }
    return (V + 2) / 4;
```

C.3.1 Haar Transform

C.3.1.1 Simple Inverse Transform

The HT is the simplest transform, it operates on 2x2 blocks of pixels.

Given a 2x2 block of pixels in the form:
{ x0, x1,
  x2, x3 }
The simple inverse HT in DSV is defined as follows:

```
x0 = (LL + LH + HL + HH) / 4;
x1 = (LL - LH + HL - HH) / 4;
x2 = (LL + LH - HL - HH) / 4;
x3 = (LL - LH - HL + HH) / 4;
```

C.3.1.2 Filtered Inverse Transform

The filtered inverse HT in DSV is defined as follows:

```
if (NOT THE FIRST OR LAST COEFFICIENT IN ROW) {
    lp = p_LL; /* previous LL on X axis (x - 1) */
    ln = n_LL; /* next LL on X axis (x + 1) */
    mx = LL - ln; /* find difference between LL values */
    mn = lp - LL;
    if (mn > mx) SWAP(mn, mx)
    mx = MIN(mx, 0); /* must be negative or zero */
    mn = MAX(mn, 0); /* must be positive or zero */
    /* if they are not zero, then there is a potential
     * consistent smooth gradient between them */
    if (mx != mn) {
        n = ROUND2(CLAMP(ROUND4(lp - ln), mx, mn) - (LH * 2));
        LH += CLAMP(n, -HQP, HQP); /* nudge LH to smooth it */
    }
}
/* do the same as above but in the Y direction, but nudging HL
 * instead of LH */
x0 = (LL + LH + HL + HH) / 4;
x1 = (LL - LH + HL - HH) / 4;
x2 = (LL + LH - HL - HH) / 4;
x3 = (LL - LH - HL + HH) / 4;
```

The ***HQP*** variable is approximately half the quantization value used for the subband. Due to the adaptive quantization, we cater to the regions that are stable and smooth those properly since they are more likely to be noticed when improperly filtered.

HQP for given ***Q*** can be derived as follows:

```
get_HQP(Q, L)
{
    if (LUMA) {
        if (INTER) {
            return Q / 14;
        }
        if (L > 4) {
            return Q / 2;
        }
        return Q / 8;
    }
    return Q / 2;
}
```

NOTE: the Haar subband transforms must properly account for odd dimensions by treating any pixel or coefficient in the 2x2 block that lies outside of the image as though it has a value of zero.

C.3.2 The 3 and 5-Tap Filters

Refer to the code in **sbt.c** for potentially more clarification.

NOTE the fixed point precision for determining what block a pixel (x, y) lands in is **14**.

The function ***reflect*** is defined as follows:

```
reflect(i, n)
{
    if (i < 0) {
        i = -i;
    }
    if (i >= n) {
        i = n + n - i;
    }
    return i;
}
```

C.3.3 ASF93

This transform is ONLY done to the 1st level luma decomposition and ONLY done to INTRA frames. The inverse transform is identical to the inverse transform of the 3-tap filter. The forward transform is up to the encoder to define. Refer to the code in **sbt.c** for potentially more clarification.

D. MOTION COMPENSATION

Motion compensation is done per plane. Do not expand the bit depth of the image at any point to do motion compensation.

D.1 Compensating Inter Blocks

Inter blocks are compensated using half or quarter-pixel precision. The luma plane has a special filter for computing the subpixels and the chroma planes use simple bilinear filtering.
Motion vectors for chroma are derived (in the case of chroma subsampled formats) by dividing the x and/or y component of the motion vector by the ratio of chroma subsampling in that direction. For example, in a 4:2:0 scheme, both the x and y components are divided by two for the chroma planes.

Luma Half-Pixel Filter A (Bicubic)

```
(19 * (b + c) - 3 * (a + d) + 16) / 32
```

Luma Half-Pixel Filter B (Bicubic)

```
(20 * (b + c) - 4 * (a + d) + 16) / 32
```

Luma Quarter-Pixel Filter and Chroma Half-Pixel Filter (Bilinear)

```
(a + b + 1) / 2
```

The two filters are chosen independently for each direction (x,y) based on a few factors. Filter A is selected when the vector component is >= 8 units, is NOT quarter-pel, or the temporal motion compensation flag is 1. Filter B is selected otherwise.

Compute the filtered block and add it to the current block. Compensate according to **D.4**.

D.2 Compensating Intra Blocks

For each of the four intra sub-blocks, calculate the average value of the sub-block in the reference picture at the full-pel motion vector location and add it to the pixels in the current sub-block. If the block is not intra, add the reference sub-block directly.
Compensate according to **D.4**.

D.3 Caveat For Encoder

From DSV1: ~~Due to the restriction of 8 bits per channel, the compensation cannot account for intra blocks that have large changes. Therefore, the encoder must simulate the reduced range intra block motion compensation to see if the block can be successfully reconstructed before it decides to label the block as intra.~~

In DSV2, this limitation was removed by introducing EPRM motion compensation. The encoder, however, should still do its own analysis to decide whether the block needs to be marked as EPRM to effectively predict and compensate the difference.

D.4 Reconstruction

If block is not flagged to use Expanded Prediction Range Mode (EPRM) or the block is a skipped inter block, reconstruction of a pixel is as follows:

```
CLAMP(prediction + residual - 128, 0, 255)
```

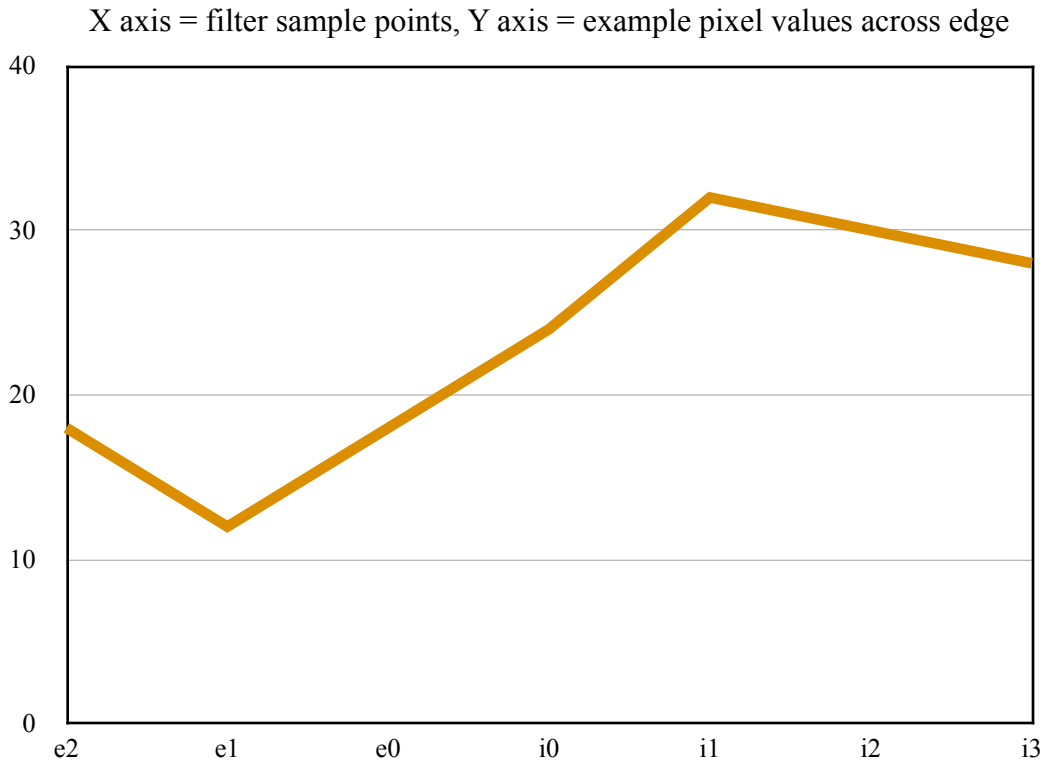
Otherwise, if the block is flagged to use EPRM, this is how its pixels are reconstructed:

```
CLAMP(prediction + (residual - 128) * 2, 0, 255)
```

D.5 In-Loop Filtering

A brief overview of the filtering done to the motion compensated frame. All planes are filtered, not just luma. Refer to the code in **bmc.c** for more clarification.

Filtering is done horizontally (top to bottom) across the vertical edges (left and right) first, then it is performed vertically (left to right) across the horizontal edges (top and bottom). Pixels on the outer edge of the video frame are not filtered.



Sample points for the filtering are prefixed with ‘e’ (external) or ‘i’ (internal). External points are outside of the current block while internal points are inside or on the edge of the current block.