|  |  |
| --- | --- |
| **ITU – Telecommunications Standardization Sector**  STUDY GROUP 16 Question 6  **Video Coding Experts Group (VCEG)**  73rd Meeting: 17-26 January 2024, Teleconference | Document VCEG-BU03 |

|  |  |  |  |
| --- | --- | --- | --- |
| Question: | Q.6/SG16 (VCEG) | | |
| Source: | **Jonathan Pfaff, Heiner Kirchhoffer, Christian Rudat, Sophie Pientka, Christian Helmrich, Heiko Schwarz, Detlev Marpe, Thomas Wiegand (Fraunhofer HHI)** | Tel:  Email: | [firstname.lastname @hhi.fraunhofer.de](mailto:thomas.wiegand@hhi.fraunhofer.de) |
| Title: | **Description of Fraunhofer HHI’s response to the Call for Evidence on the compression of biomedical waveform data** | | |
| Purpose: | Information | | |

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

# Introduction

This document describes Fraunhofer HHI’s response to the Call for Evidence on the coding of biomedical waveform data [1]. The codec submitted in the response follows the concept of a block-based hybrid architecture. Thus, the samples are partitioned into blocks of variable size. For each block, a prediction signal is computed out of already reconstructed sample values which either belong to the same channel or to a different channel. The prediction residual is transformed, where the transform is either the identity or a trigonometric transform. If the transform is the identity, an additional sample-wise prediction on the prediction residual is supported. Finally, the resulting transform coefficients are quantized and the quantization indices and the prediction mode information are entropy coded. At the end of the document, average bits per sample (BPS) versus percentage root mean square distortion (PRD) curves are shown that compare the anchor to the presented codec for the three categories of the Call for Evidence.

# Description of the codec architecture

## Block partitioning and sample processing order

Let be a biomedical waveform signal that consists of channels, where each channel consists of samples. For and , let denote the -th sample value of the -th channel.

The samples of are partitioned into a sequence of blocks denoted to . Each block is described by the position of its first sample, denoted by , and by its length in sample direction, denoted by . The length is always an integral power of two. The block is defined as

Furthermore, adjacent blocks and contain adjacent samples. More precisely, one has

(1)

The blocks are sequentially coded. Thus, one starts with coding and, until codes the block after having coded the block .

For coding a block , one partitions it into its channel-wise sub-blocks with . The block is defined as

The block is coded by sequentially coding the blocks . This means that for the coding of , one starts with coding and, until , codes the block after having coded the block .

The block partitioning is encoded into the bit stream as a sequence of lengths in sample-direction which consists of lengths to . The values can be reconstructed from (1), where .

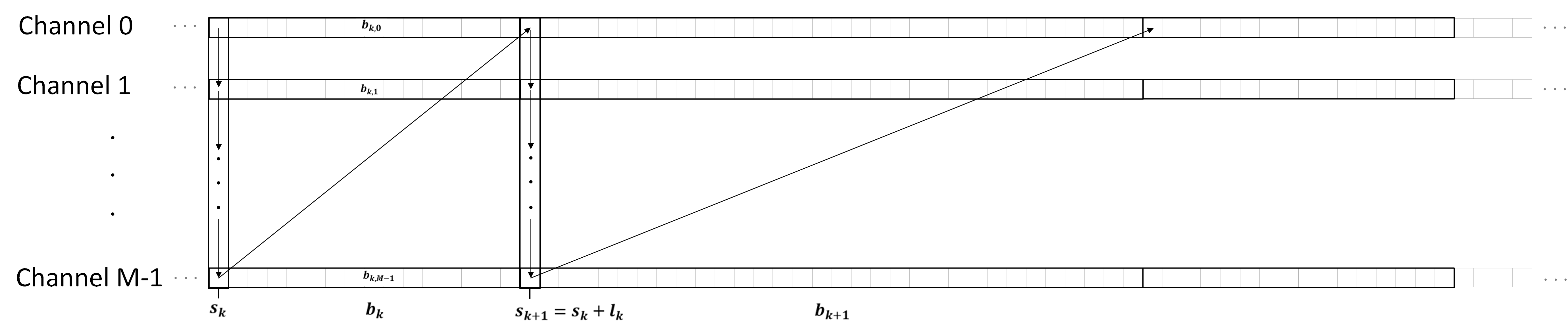


Figure 1: Illustration of the block partitioning and sample processing order

## Prediction

A prediction signal

is generated on each block out of already reconstructed sample values

that belong to the same channel or out of already reconstructed sample values

that belong to a previous channel in channel order.

Five prediction modes are supported: The DC Prediction mode, the Half-Slope Prediction mode and the Quarter-Slope Prediction mode which are intended to exploit short term correlations between the sample values, the Inter-Channel Prediction mode which is intended to exploit correlations between the sample values of different channels and the Block-Copy Prediction mode which is intended to exploit a periodic behavior of the underlying signal. The prediction mode is selected per block and is signaled in the bit stream.

Additionally, it is also possible to signal in the bit stream that for a given block the prediction is skipped, in which case one sets for all

For the DC Prediction mode, the Half-Slope Prediction mode, the Quarter-Slope Prediction mode and the Inter-Channel Prediction mode, a zero padding at the left signal boundary is used. Thus, one puts

in this case.

### DC Prediction

In the DC Prediction mode, the predictions signal is defined by the mean value of the four preceding already reconstructed sample value. Thus, one puts

and sets

### Half-Slope Prediction

In the Half-Slope Prediction mode, the prediction signal is defined by a straight line whose slope is equal to half the difference between the two reconstructed sample values preceding the current block. Thus, with

one puts

### Quarter-Slope Prediction

Similar to the case of the Half-Slope Prediction mode, for the Quarter-Slope Prediction mode one puts

and sets



Figure 2: Illustration of the DC, Half-Slope and Quarter-Slope Predictions

### Inter-Channel Prediction

In the Inter-Channel Prediction mode, the prediction signal is generated by a linear model using collocated reconstructed samples from a different channel. The model parameters are derived from already reconstructed samples.

For the inter-channel mode, which is only applicable if , a reference channel index with is transmitted.

Then, setting , integral values and are derived from the preceding already reconstructed sample values

in the reference channel and from the preceding already reconstructed sample values

in the current channel by solving a linear equation that is derived from the minimization of the quadratic prediction error

The linear equation is solved in integer arithmetic, where the division operation is replace by a table lookup and a right shift.

After the determination of andthe prediction signal is computed as

where is a fixed integral precision and .

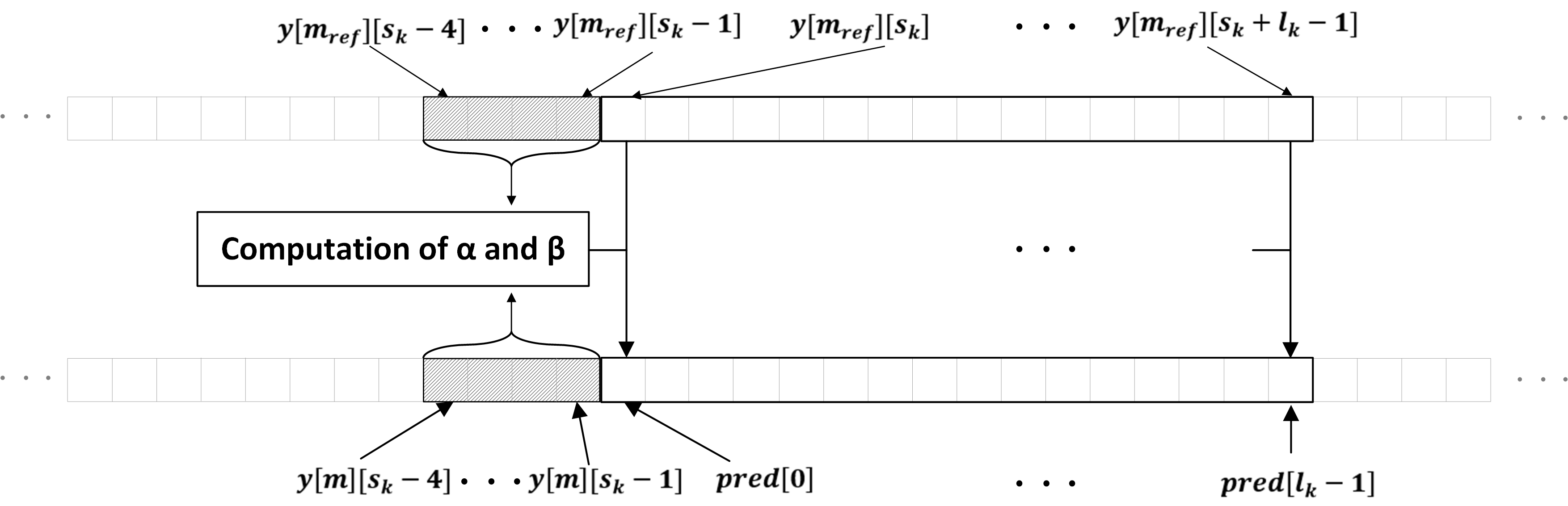


Figure 3: Ilustration of the Inter-Channel Prediction

### Block-Copy Prediction

In the Block-Copy Prediction mode, the prediction signal is generated by copying the sample values of an already reconstructed reference block which has the same length as the current block and whose location is transmitted. In order to model cases of an underlying periodic analog signal with non-integral period, a half-pel accurate Block-Copy Prediction is supported.

The Block-Copy Prediction mode is only applicable if . The reference block location is transmitted by coding an integral offset value

Moreover, an additional half-pel flagis transmitted. If this flag is false, one puts

If this flag is true, one puts

where

are coefficients of a half-pel interpolation filter. In (2), a constant boundary extension to the right is assume for the reconstructed sample values if and a constant boundary extension to the left is assumed if .

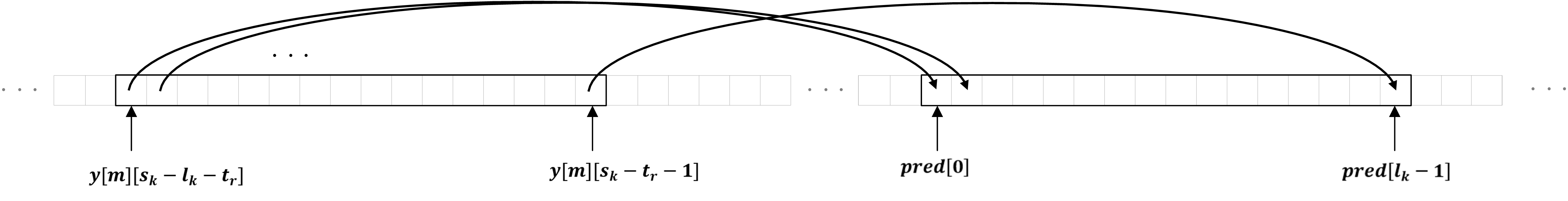


Figure 4: Illustration of the Block-Copy Prediction (with value of half-pel flag equal to 0)

## Transform and quantization

### Trigonometric transforms

For the block , if the prediction residual

can be transform coded by using a trigonometric transform. It is signaled in the bit stream whether a trigonometric transform or the identity transform is used.

For , both the discrete sine transform DST-VII and the discrete cosine transform DCT-II are supported, where the selection of the transform is signaled in the bit-stream. For , only the DCT-II is supported. The trigonometric transforms are used in a non-overlapping way, i.e., all basis functions of the inverse transform are zero outside .

At the encoder, the resulting transform coefficients are quantized using a scalar uniform reconstruction quantizer to obtain the transform coefficient levels. The stepsize of the quantizer does not depend on the frequency position and is constant for all blocks . The stepsize is signaled once for the whole sequence . The levels are transmitted in the bit-stream as described below.

At the decoder, the reconstructed transform coefficients are obtained by multiplying the levels with the quantizer stepsize. Afterwards, the inverse trigonometric transform is applied to the to obtain the reconstructed residual sample values . The reconstructed sample values are then obtained as

Both the forward and the inverse trigonometric transforms are computed as full matrix multiplications using fixed-point arithmetic.

### Identity transform and secondary prediction

If the identity transform is used on the current block , a secondary prediction for the residual samples is supported for the case that the prediction mode is either the Inter-Channel Prediction or the Block-Copy Prediction mode or that the Zero Prediction mode is used.

Two types of secondary prediction are supported: Secondary prediction with signal-adaptive filter coefficients and secondary prediction with predetermined filter coefficients. It is signaled in the bit stream whether the secondary prediction is to be used or not and which type of the secondary prediction is to be used.

If the secondary prediction with signal adaptive filter coefficients is to be used, a prediction-filter length with as well as integral coefficients are transmitted.   
For the secondary prediction with predetermined filter coefficients, the filter length is set to and three modes with different filter coefficients are supported. The mode is signaled in the bit stream. One has for the first mode for the second mode and for the third mode.

From a decoder perspective, if the identity transform is used, first, the residual coefficients are reconstructed out of the transmitted quantization indices by multiplication with a stepsize that does not depend on the position .   
Then, if a secondary prediction is used, the final reconstructed residual coefficients are sequentially computed starting from until and incrementing in each step. Thus, for , one puts

For , setting for the secondary prediction with signal adaptive coefficients and for the secondary prediction with predetermined coefficients, one puts

and sets

If no secondary prediction is used, one puts for all . Finally, the reconstructed sample values are generated as

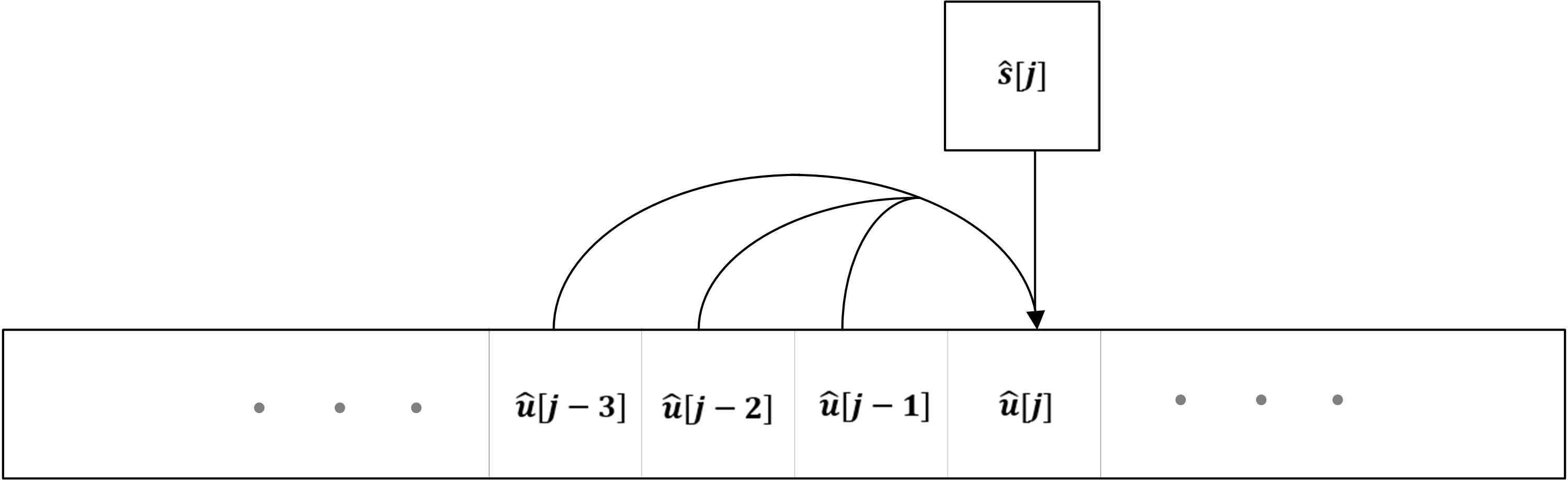


Figure 5: Illustration of the generation of a final reconstructed residual sample out of the previous final reconstructed residual samples and the coded and inverse quantized intermediate residual sample ]. This corresponds to equation (3) for the case K=3.

## Entropy coding of transform coefficient levels

The transform coefficient levels are entropy coded using a variant of context-based adaptive binary arithmetic coding (CABAC). Here, the implementation of the probability estimator and the arithmetic coding engine is based on the MPEG NNC standard [3]. Two separate paths of entropy coding are used depending on whether a trigonometric transform or whether the identity transform is used for the current block .

### Trigonometric transforms

If a trigonometric transform is used, a value with is transmitted first. Here, for all with one has

and, if , one always has

(4)

Then, starting with , the coefficients are sequentially coded in a backwards scan and in a single pass. Thus, one first codes and then, until , codes after having coded .

For the coding of , one decomposes

,

where

Then, the value is transmitted by a Rice code with Rice parameter 2. If , the value of is transmitted using truncate binary coding.

For the coding of a coefficient , its significance flag , defined as

is transmitted first except for the case that and , where the value of can be inferred to be at the decoder by (4).

Next, if , the sign of and, afterwards, the absolute value are transmitted. For the transmission of , one decomposes it as

(5)

where

, .

The value is transmitted first using truncated unary coding. Then, if , the value is transmitted using Exponential Golomb coding.

The significance flag as well as the sign flag and all bins of the truncated unary code for are context coded while the bins in the Exponential Golomb coding of are coded in the bypass mode. For each channel, a separate set of context models is used.   
For the significance flag, 45 different context models are supported per channel. The selection of the context model index depends on the value and on the sum

of the at most three previously reconstructed absolute values inside the same block. More precisely, for two predefined monotonically increasing functions

one sets

For the sign flag, a single context model is used per channel.   
Finally, for all bins appearing in the truncated unary binarization of from (5), a single context model is used which is selected from a set of 15 different context models per channel. Here, the selection of the context model index is defined as

and thus only depends on the position .

### Identity transform

If the identity transform is used, the coding of the coefficients levels is similar to the case of the trigonometric transforms with the following changes.   
First, the coding of a value is bypassed and consequently, the coefficient levels are always sequentially coded in decreasing order starting from . Moreover, if , the binarization (5) is replaced by a binarization of the form

where

, , .

Here, is again transmitted using a truncated unary code. Moreover, if , the value is transmitted using a Rice code with Rice parameter . Here, the Rice parameter depends on the sum

of all previously coded absolute values of the same block and can be computed as

for a fixed monotonically increasing function

Finally, if , the value is transmitted using an Exponential Golomb code.

For the identity transform, the significance flag as well as the bins of the truncated unary code for and the bins of the truncated unary part of the Rice code for are context coded while all other bins are coded in the bypass mode. A set of 10 separate context models per channel for the coding of the values is supported. The context model selection for each bin is always independent of and depends only on the position of the bin within the used binarization.

# Experimental Results

Subsections 3.1 to 3.3 show graphs with PRD values over BPS values for the three datasets averaged per working point.   
Here, the anchor points for the Extended HE-AAC audio codec [2] have been taken from document VCEG-BU01 that has been sent around on the VCEG email reflector on December 1st 2023. Moreover, for informational purposes, lossless compression results have been generated for the MPEG-4 Audio Lossless Coding (ALS) standard, using the ALS-reference codec RM23 [4] with preset -7 (“parameters set for optimum compression”).

According to [1], if is the mean of the i-th channel, i.e.

the percentage root mean square distortion (PRD) is defined as

Rate-distortion optimization has been employed in the encoder using a fixed Lagrange multiplier per bitstream. Note that this corresponds to an optimization to PRD values. Optimizing to CPRD values might require adjusting the Lagrange multipliers per channel according to the variance of the corresponding channels original signal variance.

Subsections 3.1 to 3.3 show graphs with PRD values over BPS values for the three datasets averaged per working point.

Full per-sequence results as well as a decoder executable and all bit streams have been uploaded to the FTP server specified in [1] as files VCEG-BT07-v1-Fraunhofer-HHI-response-results.zip and VCEG-BT07-v1-Fraunhofer-HHI-response.zip.

## EEG



Figure 6: PRD values over bits per sample (BPS) averaged per working point for the EEG dataset. Vertical grid lines correspond to the working points of [1].

## ECG

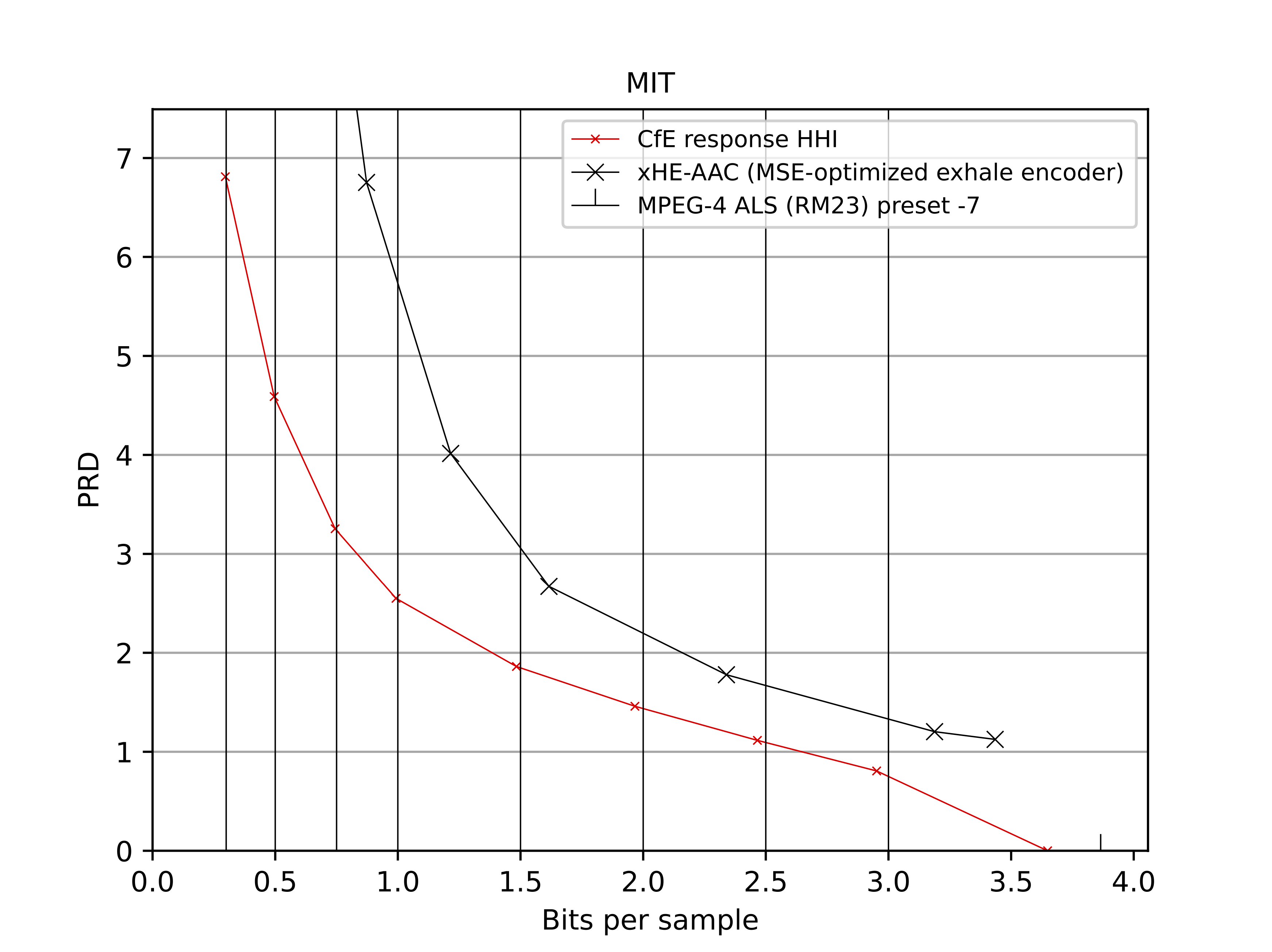


Figure 7: PRD values over bits per sample (BPS) averaged per working point for the ECG dataset. Vertical grid lines correspond to the working points of [1].

## EMG

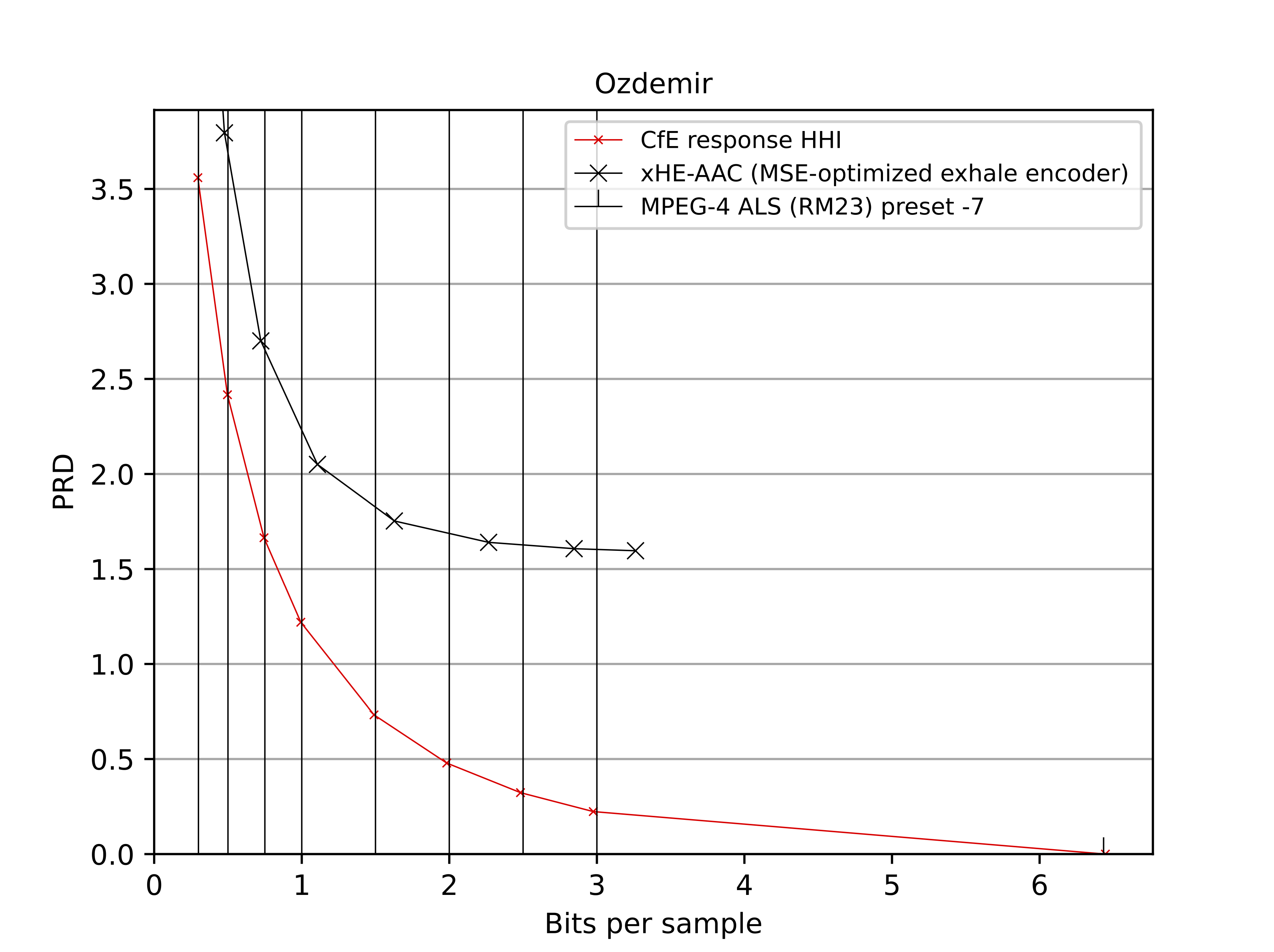


Figure 8: PRD values over bits per sample (BPS) averaged per working point for the EMG dataset. Vertical grid lines correspond to the working points of [1].

# Patent rights declarations(s)

**Fraunhofer HHI 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 (per box 2 of the ITU-T/ITU-R/ISO/IEC patent statement and licensing declaration form).**

# References

1. J. Pfaff and J. Halford, “Call for Evidence on the coding of biomedical waveform data,” Q.6/SG16, *doc.VCEG-BT07*, Hannover, Nov. 2023.
2. ISO/IEC IS 23003-3, “Information technology – MPEG audio technologies – Part 3: Unified speech and audio coding,” Geneva, Jun. 2020.
3. ISO/IEC 15938-17:2024 Information technology Multimedia content description interface Part 17: Compression of neural networks for multimedia content description and analysis
4. ISO / IEC, International Standard 14496-3:2019, “Coding of audio-visual objects,” part 3, subpart 11, “Audio Lossless Coding – ALS,” 2006.
5. ISO/IEC 14496-5:2001/AMD 10:2007/DCOR 5 ALS Floating point data corrections

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