Super-Wideband Fine Spectrum Quantization for Low-rate High-Quality MDCT Coding Mode of The 3GPP EVS Codec

Panasonic NOKIA SAMSUNG ELECTRONICS

Presented by

Srikanth Nagisetty, Hiroyuki Ehara

15th Dec 2015

Topics of this Presentation

- Background
- EVS Encoder Overview
- LR-MDCT Coder
  - Envelope Coding
  - Spectrum Coding
    - Bit allocation
    - Gap Filling
- Evaluation Results
- Summary

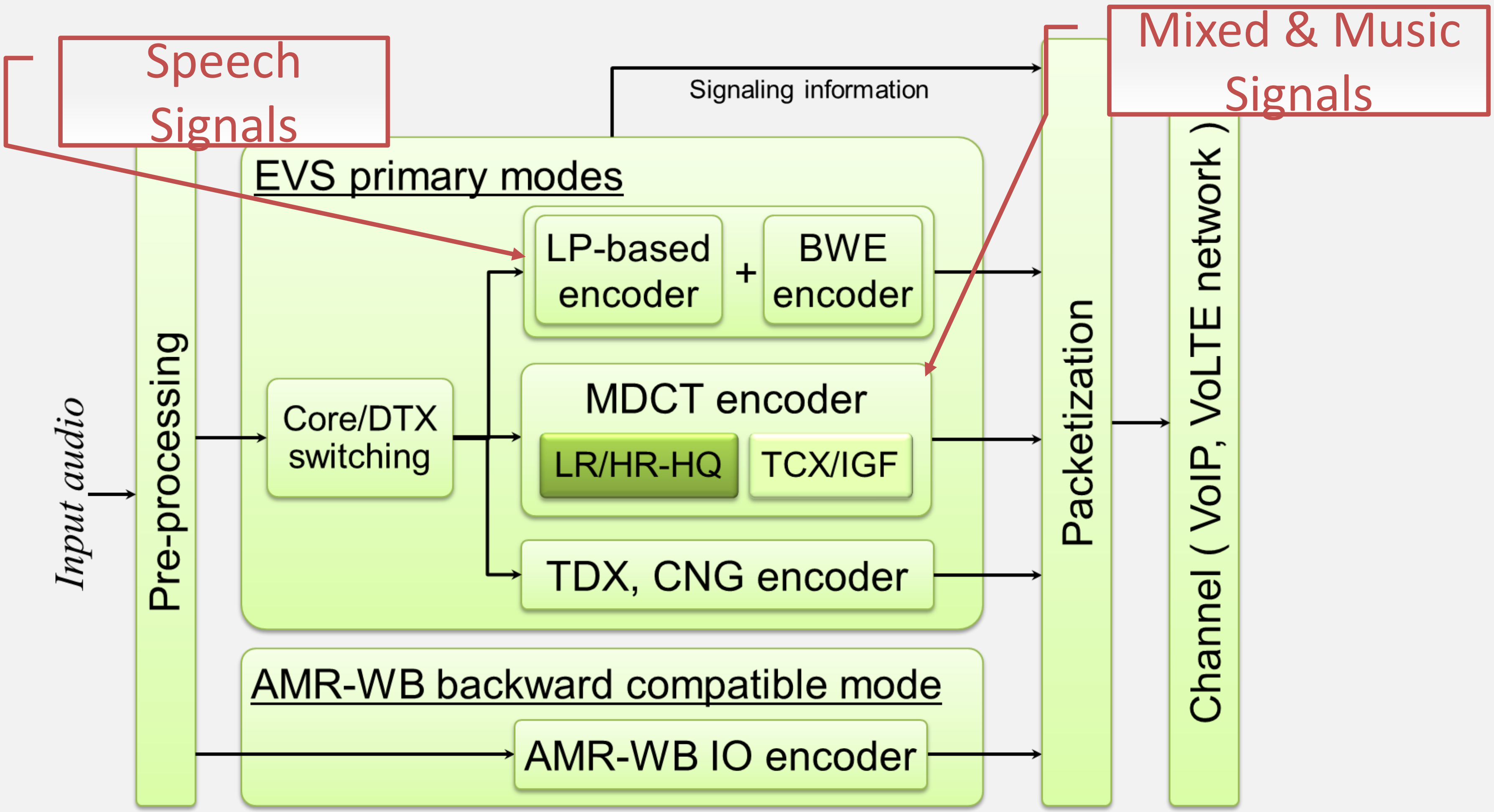
Background

Challenge:

- Encoding SWB band signal at low delay & low bit-rate
  - Mixed contents and music sampled at 32kHz
  - Frame length : 20 ms
  - Too few bits for quantizing SWB spectrum （quantizing 560 bins using around 256 bits）

✓ Highly efficient quantization algorithm is needed

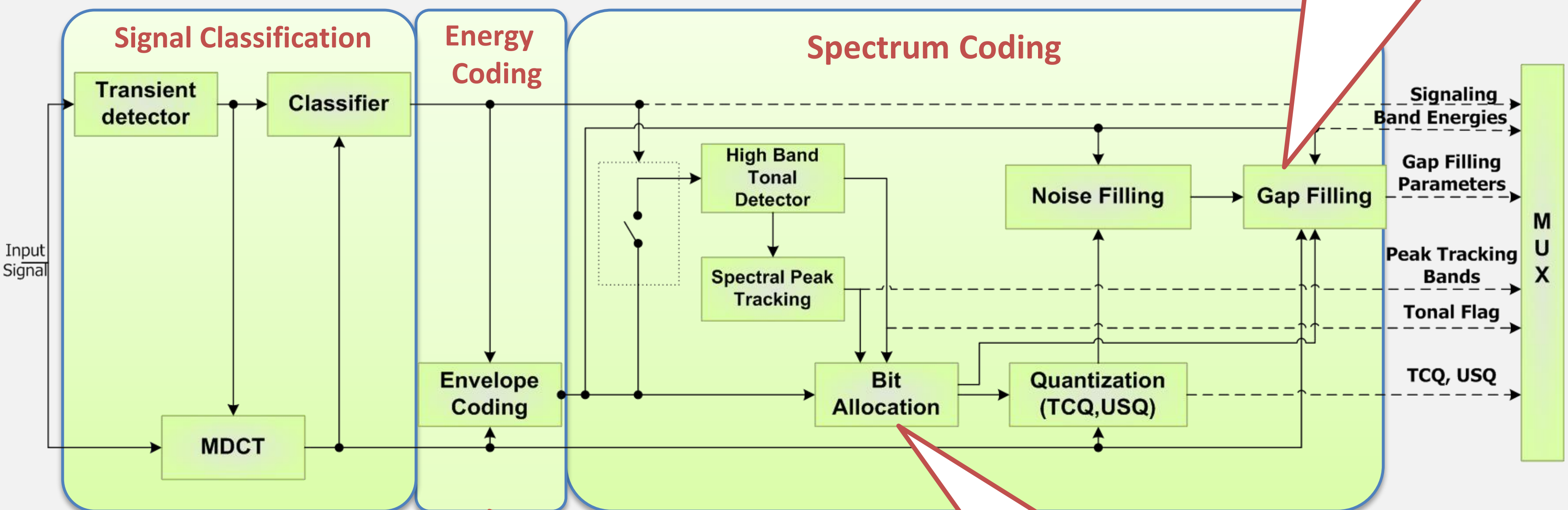
EVS Encoder Overview



- The Low Rate High Quality (**LR-HQ**) **MDCT** coding is one of the mode in the EVS MDCT coder.

LR-HQ MDCT SWB Encoder (1/2)

Quantized LF information for gap filling



Huffman Coding

Advanced bit allocation

Transient

Normal

Harmonic

LR-HQ MDCT SWB Coder

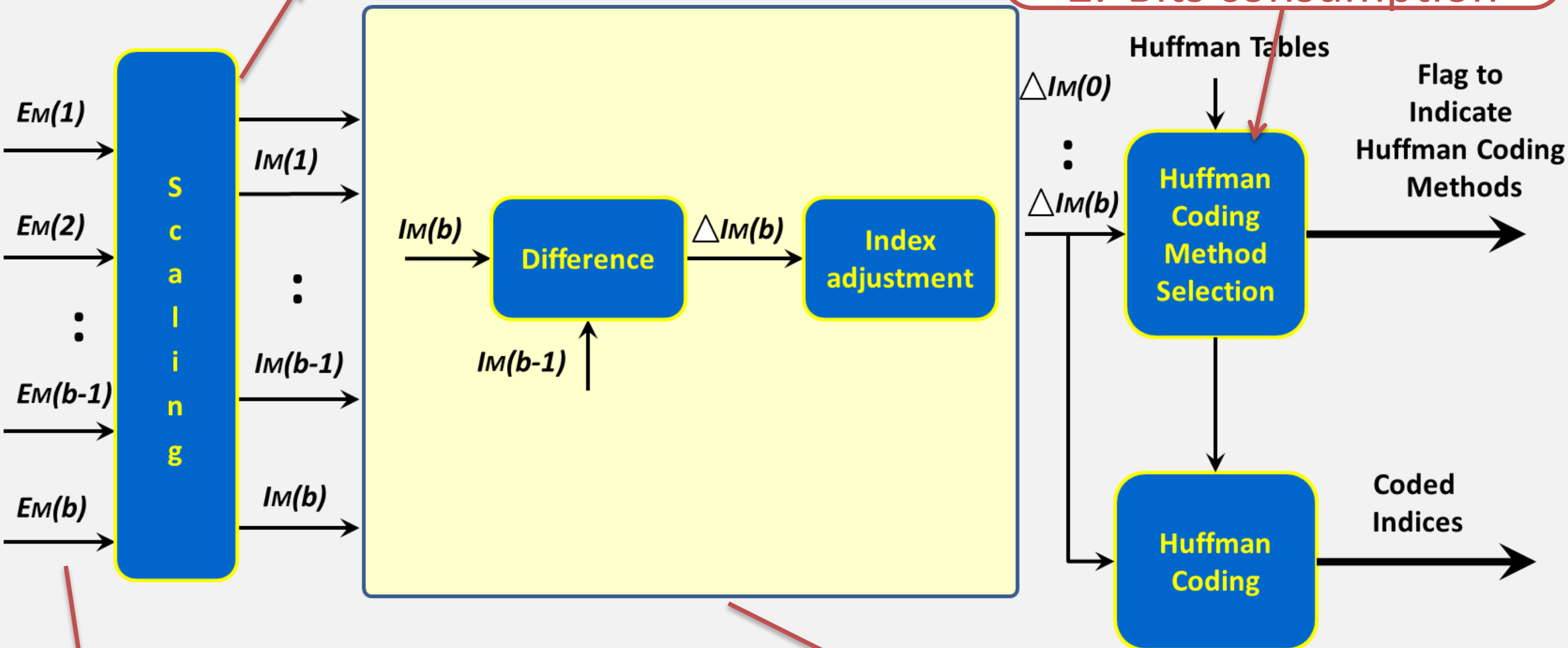
-Envelope coding

Scaled Energies

$$I_M(b) = round(\frac{E_M(b)}{q_{int}}), \quad b = 0,\ldots,N_{bands}-1$$

Selection based on

1. Range of indices
2. Bits consumption



$$E_M(b) = \log_2\left(\sum_{k=k_{start}(b)}^{k=k_{end}(b)-1} X_M(k)^2 + Epsilon\right), \quad b = 0,\ldots,N_{bands}-1$$

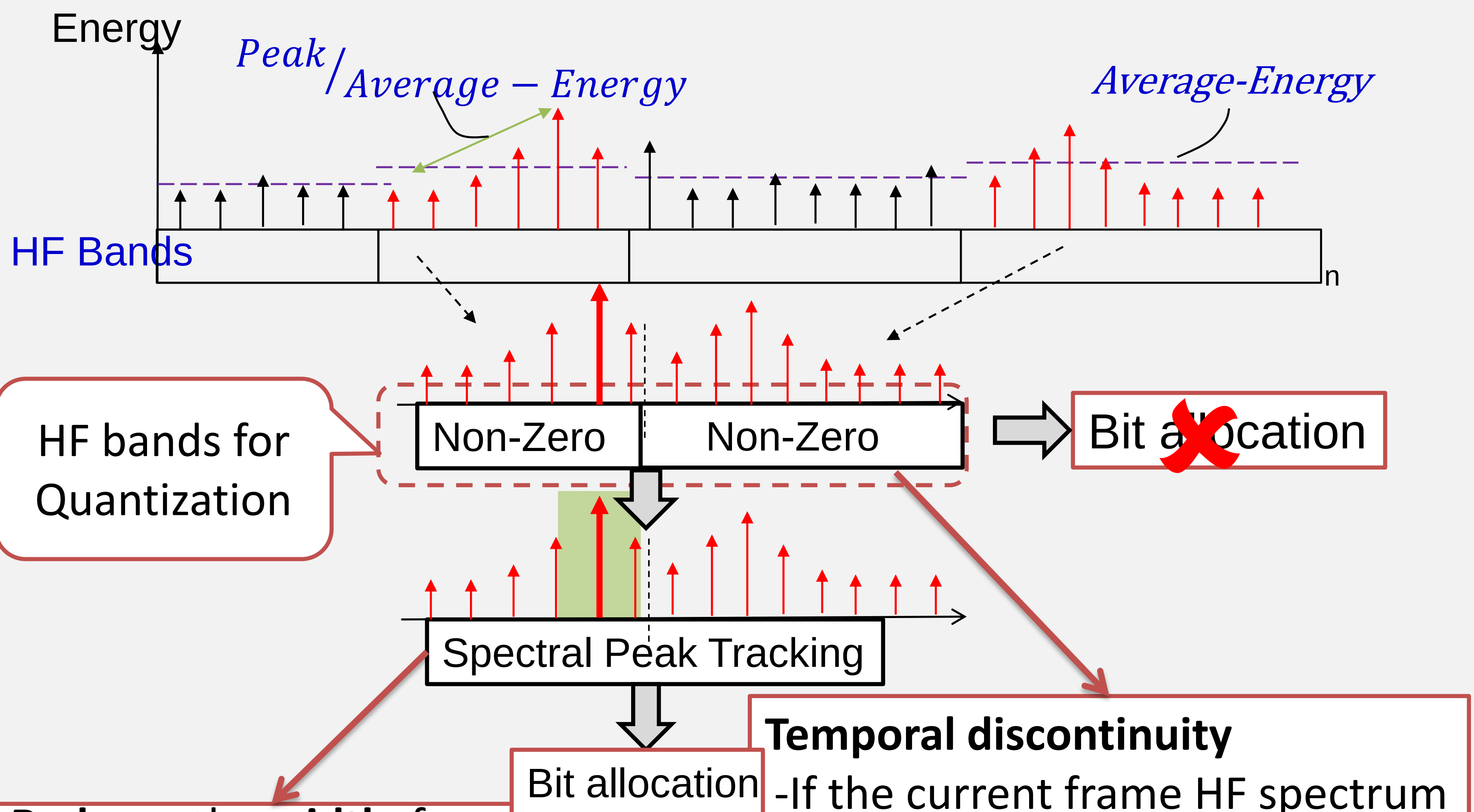
$$\Delta I_M(0) = I_M(0) - round(\frac{I_{ref}}{q_{int}})$$

$$\Delta I_M(b) = I_M(b) - I_M(b-1), \quad b = 1,\ldots,N_{bands}-1$$

Band Energies

Differential Indices

Spectral Peak Tracking



Reduces the **width** for the current frame

- **Bit savings**
- **Inter-frame Continuity**

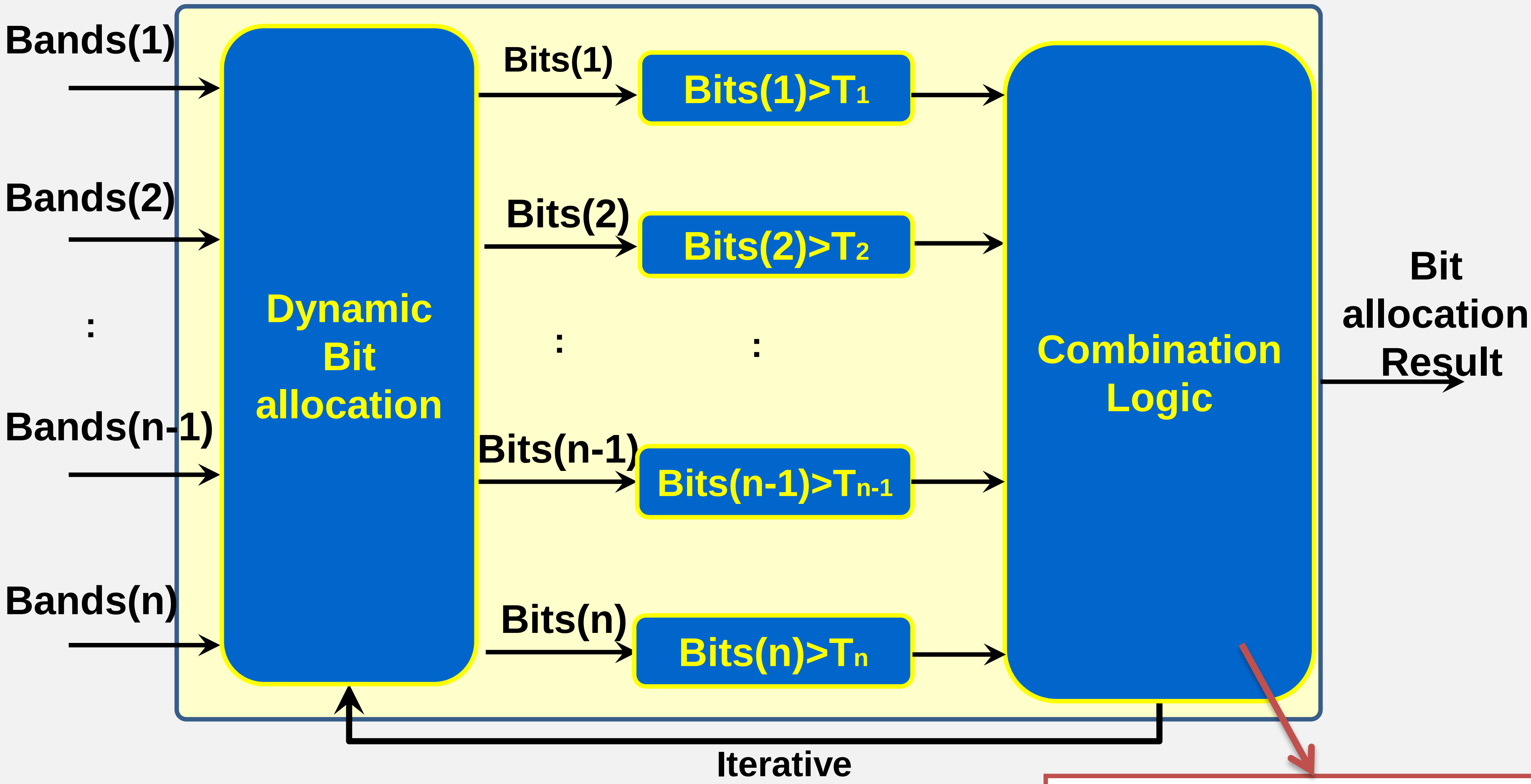
Temporal discontinuity

-If the current frame HF spectrum peak is quantized while the previous frame is not quantized

-Perceivable degradation

LR-HQ MDCT Coder
-Bit allocation (1/2)

Dynamic Bit allocation

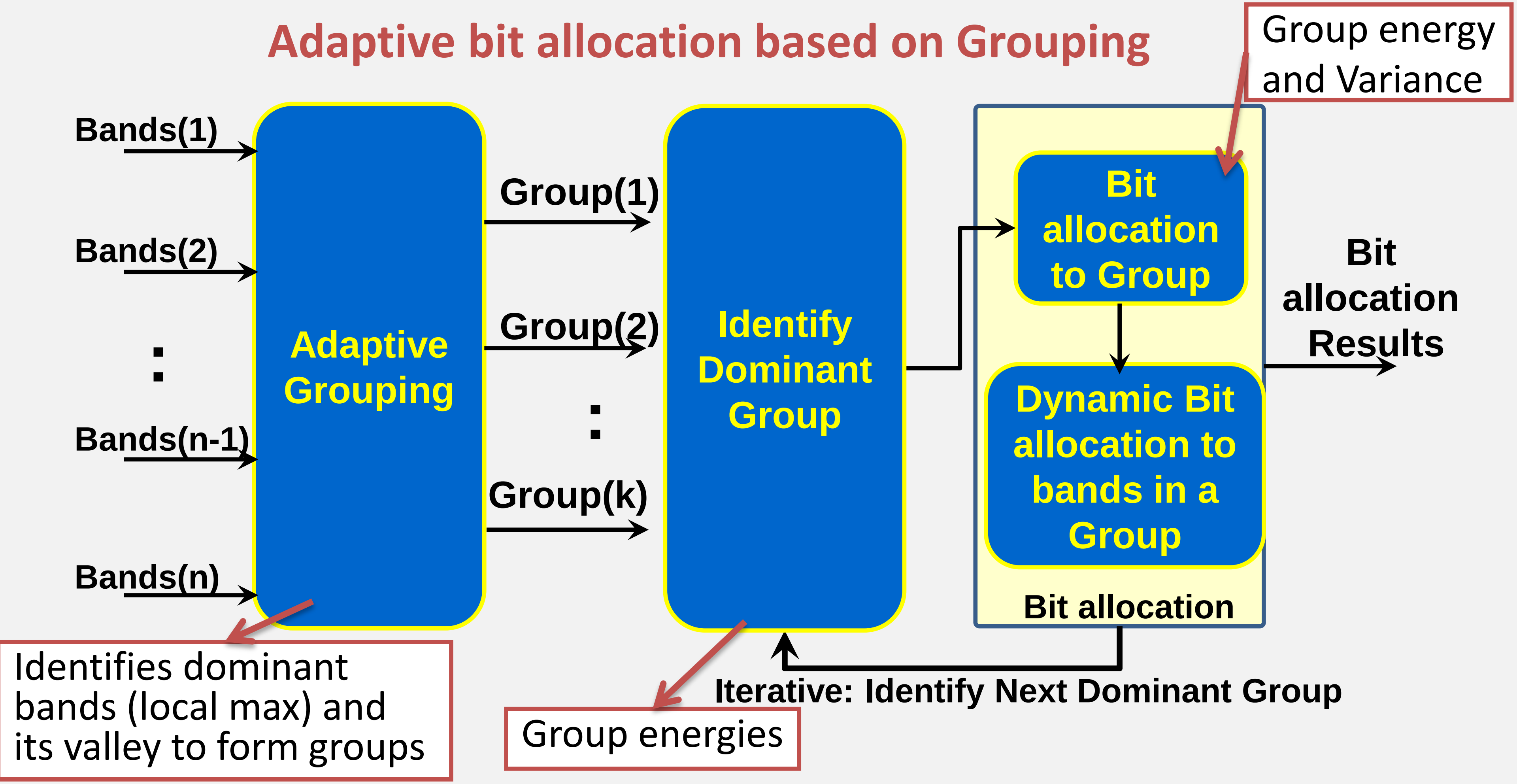


Bands with bits < Threshold identified and relocates bits to other bands

LR-HQ MDCT Coder

-Bit allocation (2/2)

Adaptive bit allocation based on Grouping

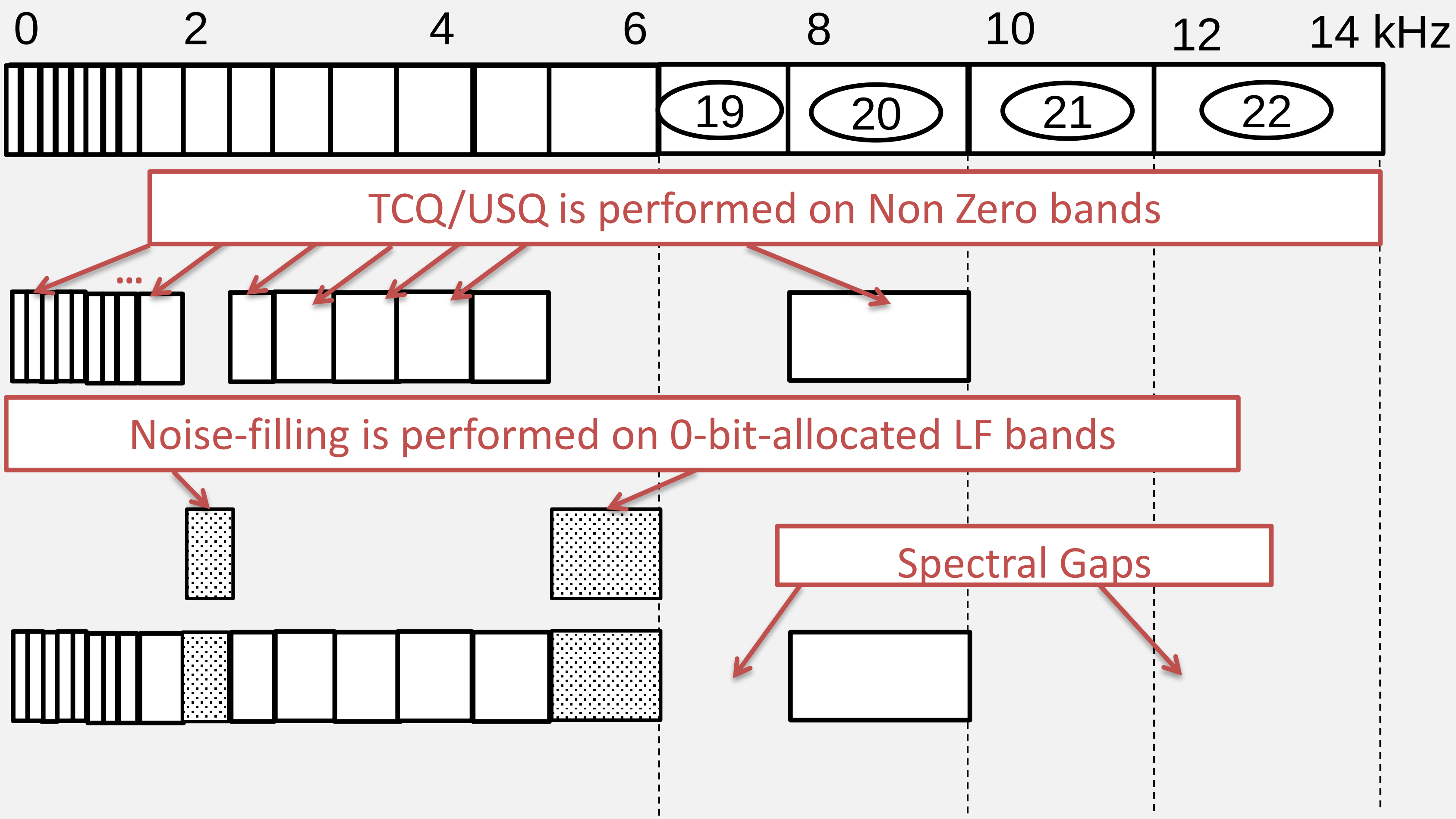


Bits allocated 1) Adaptively grouping the bands and 2) By exploiting the relationship between the groups.

This approach is more suitable for tonal (Harmonic) like signals as the energy of the bands is mainly concentrated at discrete tones.

LR-HQ MDCT Coder

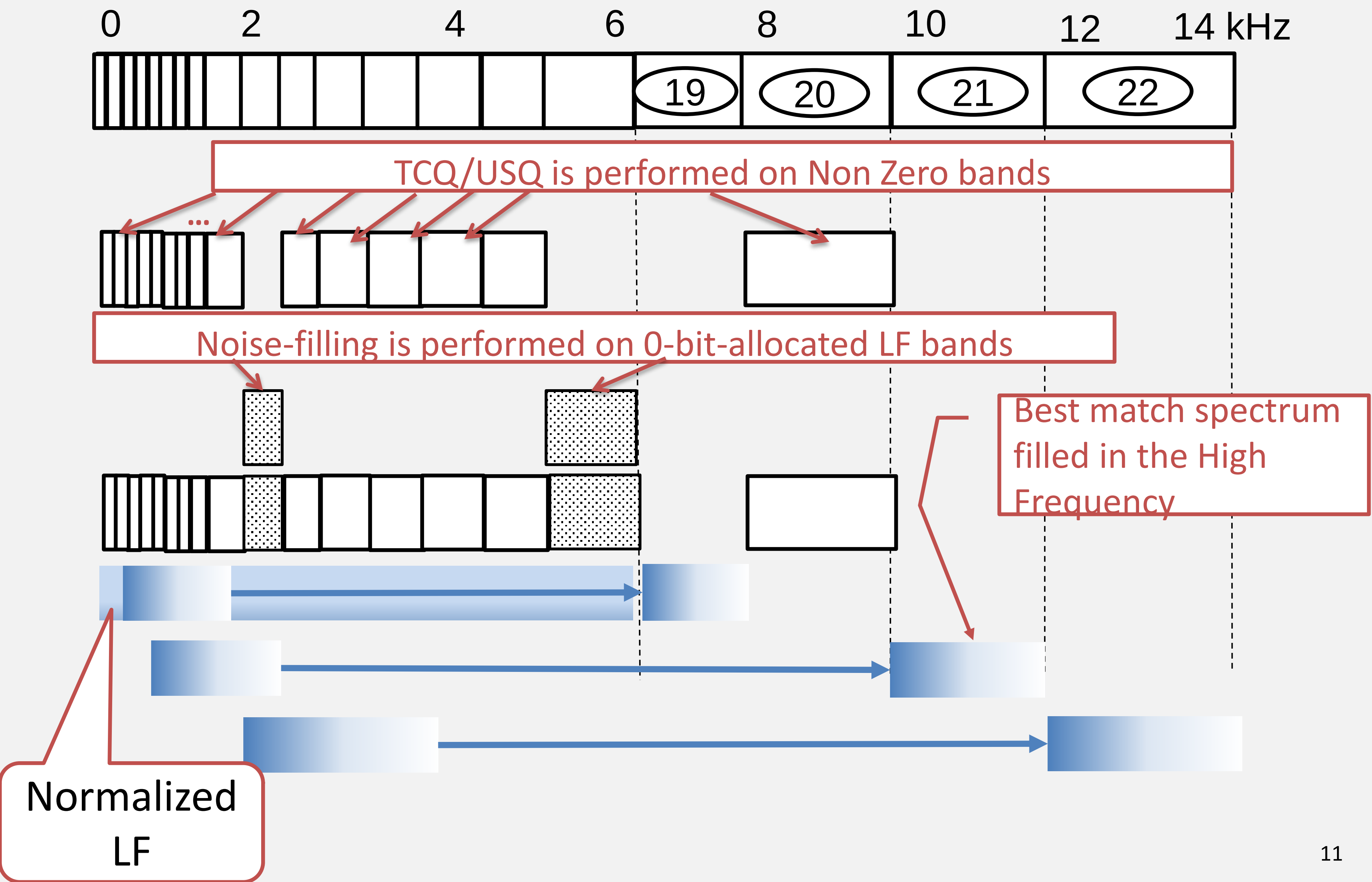
-Gap Filling



- Zero-bit bands cause spectral gaps, which lead to audible artifacts if left alone. Gap filling technique is used.

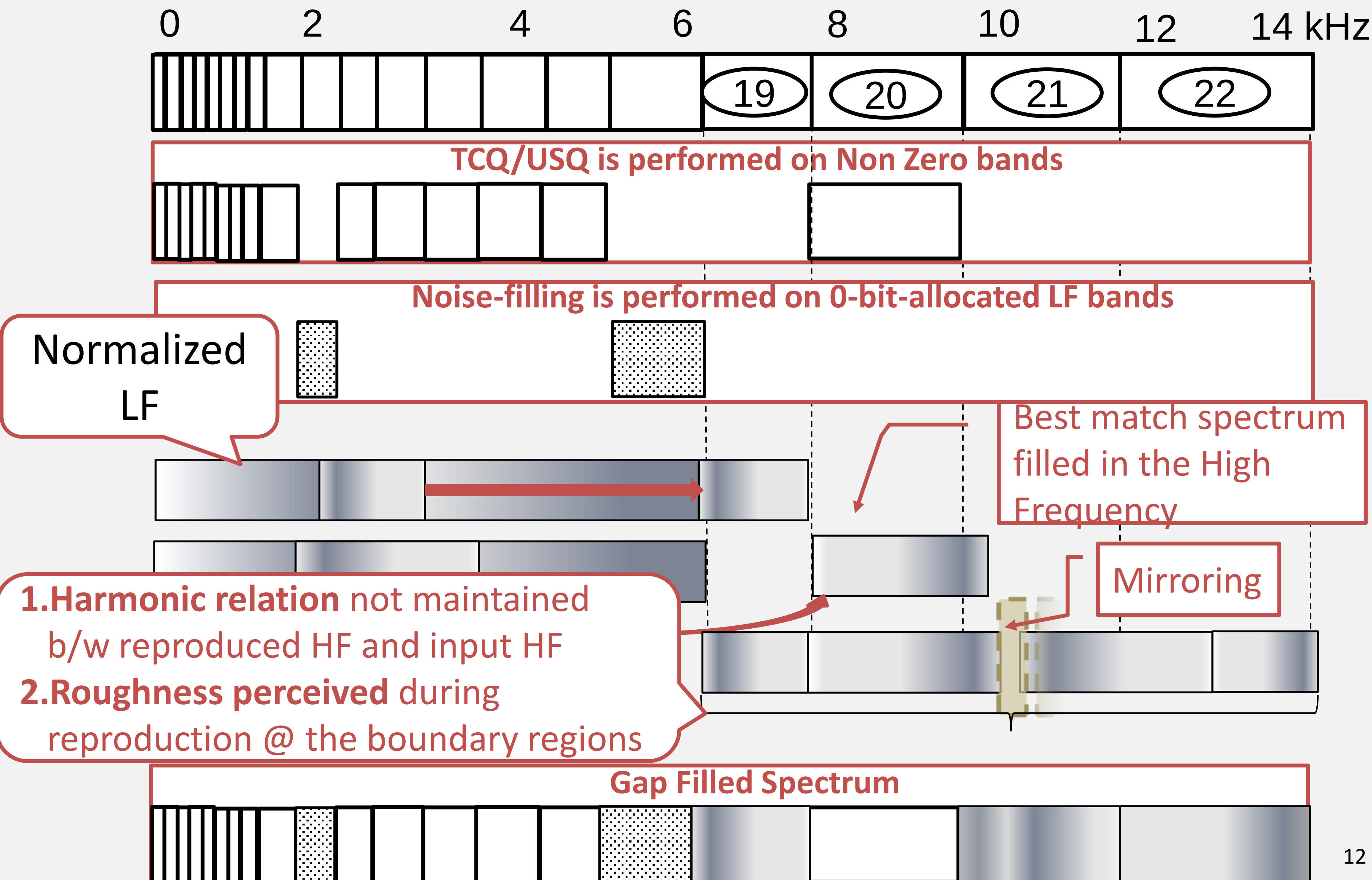
LR-HQ MDCT Coder

-Gap Filling Normal Mode



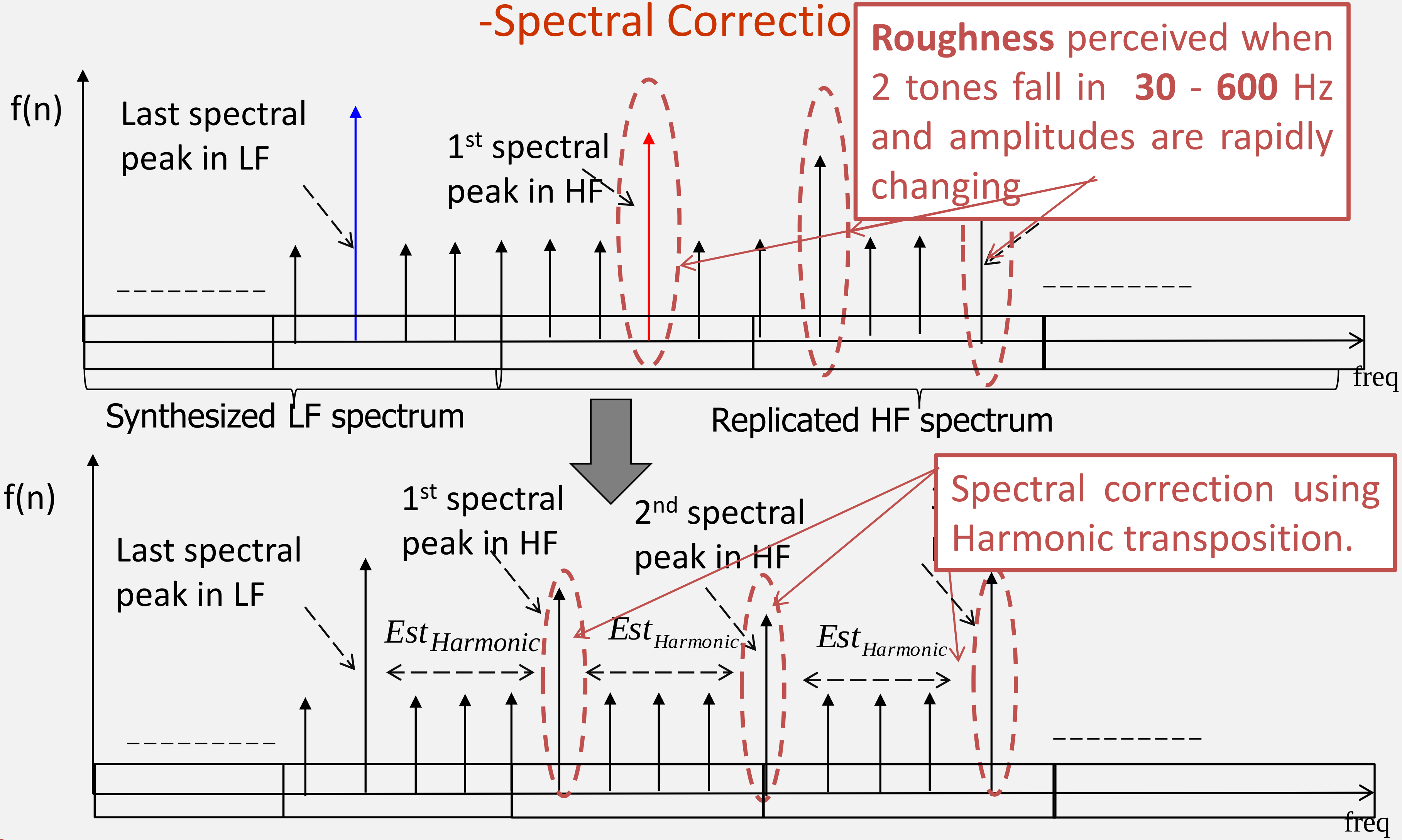
LR-HQ MDCT Coder

-Gap Filling Harmonic Mode



LR-HQ MDCT Coder

-Spectral Correction



- This method retains both the fine structure of the spectrum and the harmonic relationships between the low frequency tones and the replicated high frequency tones

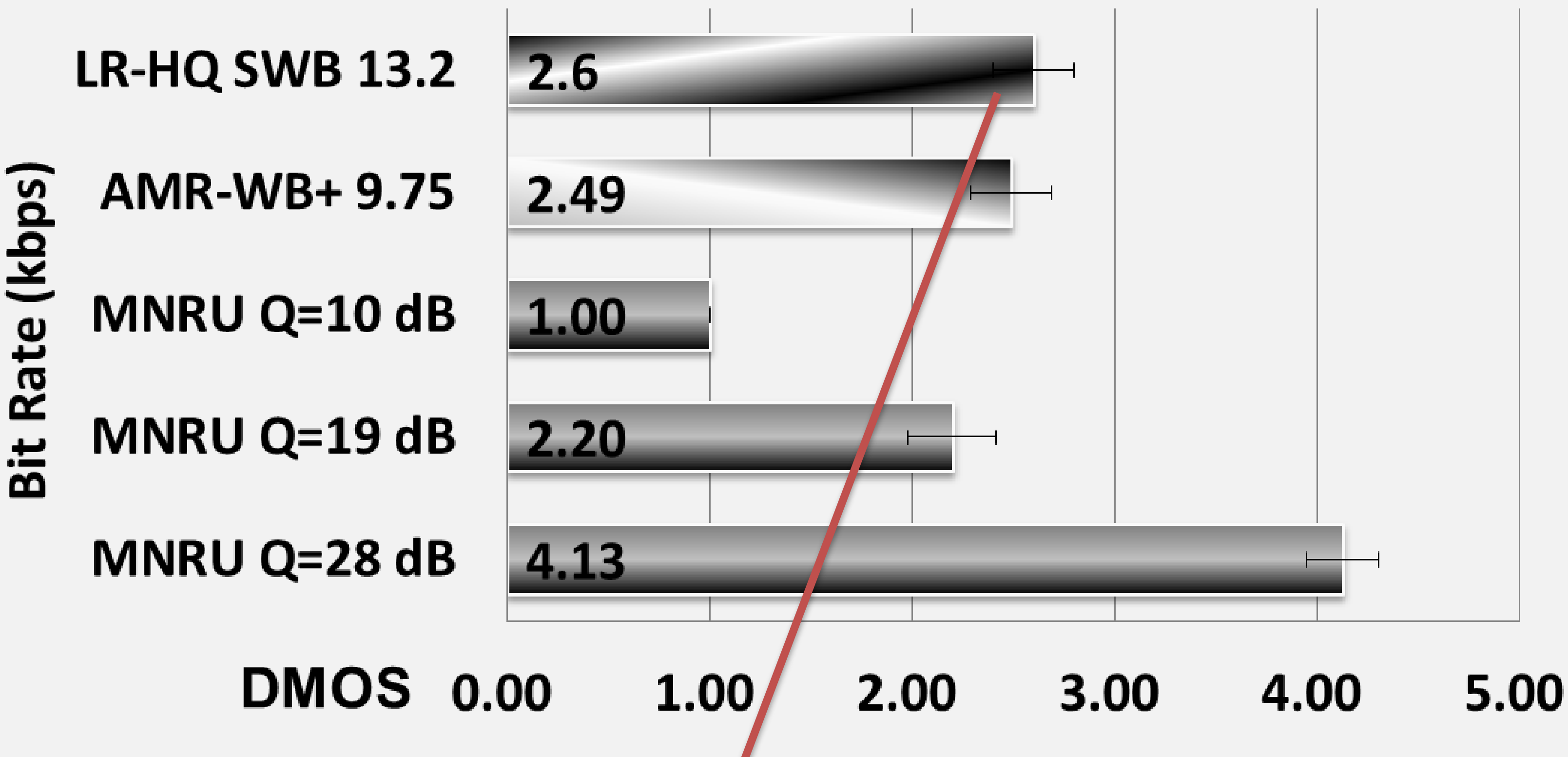
LR-HQ MDCT Coder

- Listening Test Setup

- Degradation Category Rating (DCR) methodology (ITU-T P.800) :
  - 24 Mixed and Music samples recorded in Japanese language
  - 16 Japanese naïve listeners
  - Codecs
    - EVS SWB LR-HQ mode
    - Reference Codec: AMR-WB+

Degradation	Scale
Degradation is inaudible	5
Degradation is audible but not annoying	4
Degradation is slightly annoying	3
Degradation is annoying	2
Degradation is very annoying	1

LR-HQ MDCT Coder
- Evaluation Results



- LR-HQ SWB Performance is equal or greater than AMR-WB+ whose algorithmic delay is longer than twice of EVS (32ms)

Summary

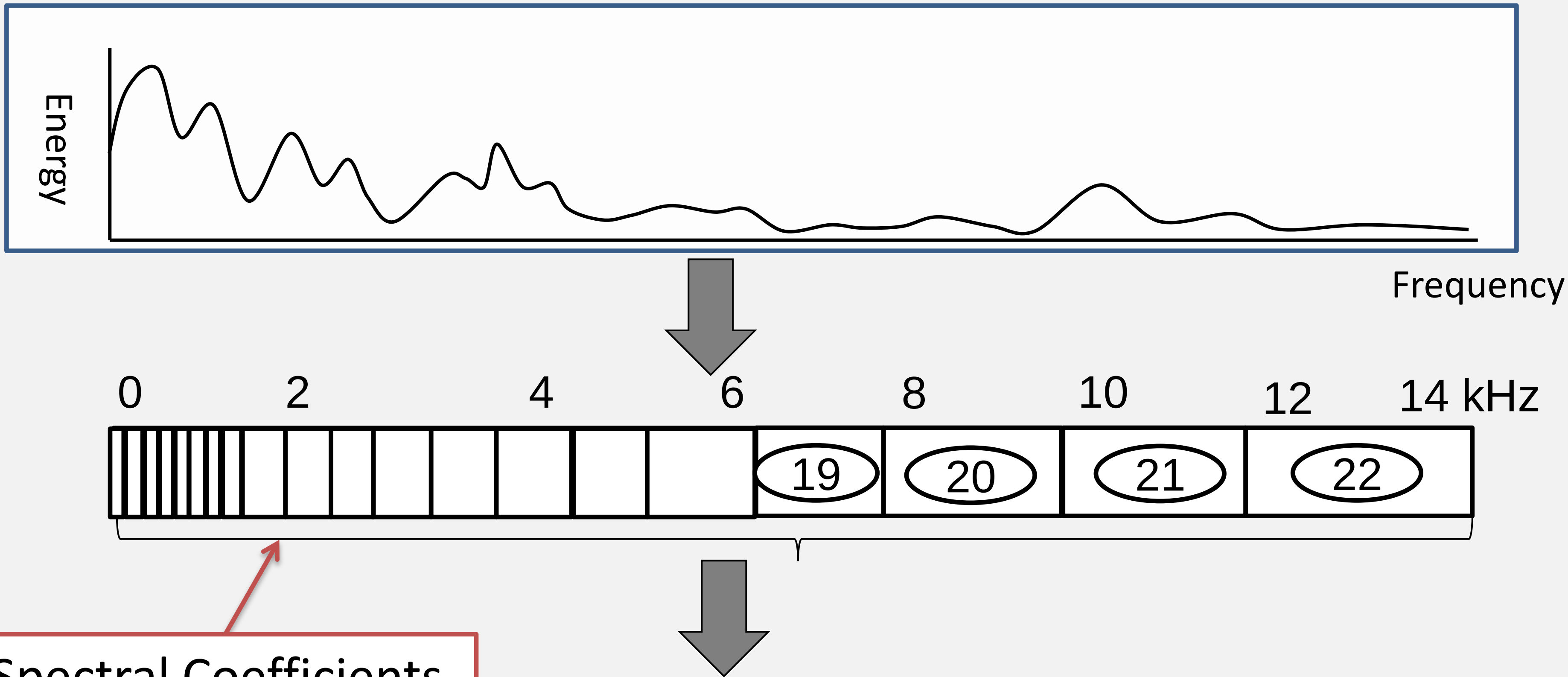
- For encoding the SWB spectral coefficients at low bit budget
  - Spectral band energies are quantized using an efficient Huffman coding methods
  - Advanced bit allocation methods are used for efficient representation of spectrum.
  - Spectral holes in the full spectrum coding is filled using gap filling techniques
    - Gap-filling techniques are improved by introducing a fine spectrum normalization and adaptive sparse BWE coding
- Conclusion: EVS LR-HQ SWB coder meets the performance requirements and is adopted as a part of multi-mode MDCT coding in the EVS codec.

References

1. 3GPP TR 22.813 - "Study of Use Cases and requirements for enhanced voice codecs for the EPS", v.10.0.0, Mar 2010.
2. 3GPP TS 26.090 - "Mandatory speech codec speech processing functions; Adaptive Multi-Rate (AMR) speech codec; General description Speech", v.11.0.0, Sep.2012.
3. 3GPP TS 26.190 - "Speech codec speech processing functions; Adaptive Multi-Rate-Wide-band (AMR-WB) speech codec; General description", v.11.0.0, Oct.2012.
4. 3GPP TS 26.445 - "Codec for Enhanced Voice Services (EVS); Detailed Algorithmic Description", v.12.0.0), Sep. 2014.
5. S. Bruhn, et al., "Standardization of the new EVS Codec", IEEE ICASSP, April 2015.
6. S. Nagisetty, et al., "Low Bitrate High Quality MDCT Audio coding of the 3GPP EVS", IEEE ICASSP 2015.
7. ITU-T G.719, "Low complexity, full band audio coding for high quality, conversational applications", ITU-T Recommendation G.719, June 2008.
8. T. R. Fischer, H. Sung, J. Zhan, E. Oh, "High-quality audio transform coded excitation using trellis codes," ICASSP, 2008, pp.197-200.
9. ITU-T G.718 Annex B, "Superwideband scalable extension for ITU-T G.718", ITU-T Recommendation G.718 Amendment 2, March, 2010.
10. L. Laaksonen et al., "Superwideband extension of G.718 and G.729.1 speech codecs", Interspeech 2010.
11. M. Oshikiri, H. Ehara, K. Yoshida, "Efficient spectrum coding for super-wideband speech and its application to 7/10/15 kHz bandwidth scalable coders," IEEE ICASSP, 2004, pp.I-481-484.
12. H. Fastl, E. Zwicker, Psychoacoustics: Facts and Models. Springer series in information sciences. Springer, 3rd edition, 2007.
13. Audio codec processing functions; Extended Adaptive Multi-Rate Wide-band (AMR-WB+) codec", 3GPP TS 26.290.
14. 3GPP, T-doc S4-130522, EVS performance requirements, April 2013.
15. ITU-T P.800, Methods for Subjective Determination of Transmission Quality. International Telecommunication Union (ITU), Series P., August 1996.
16. 3GPP, Tdoc S4-141065, Report of the Global Analysis Lab for the EVS Selection Phase, Aug. 2014.

APPENDIX

LR-HQ MDCT SWB Encoder (2/2)



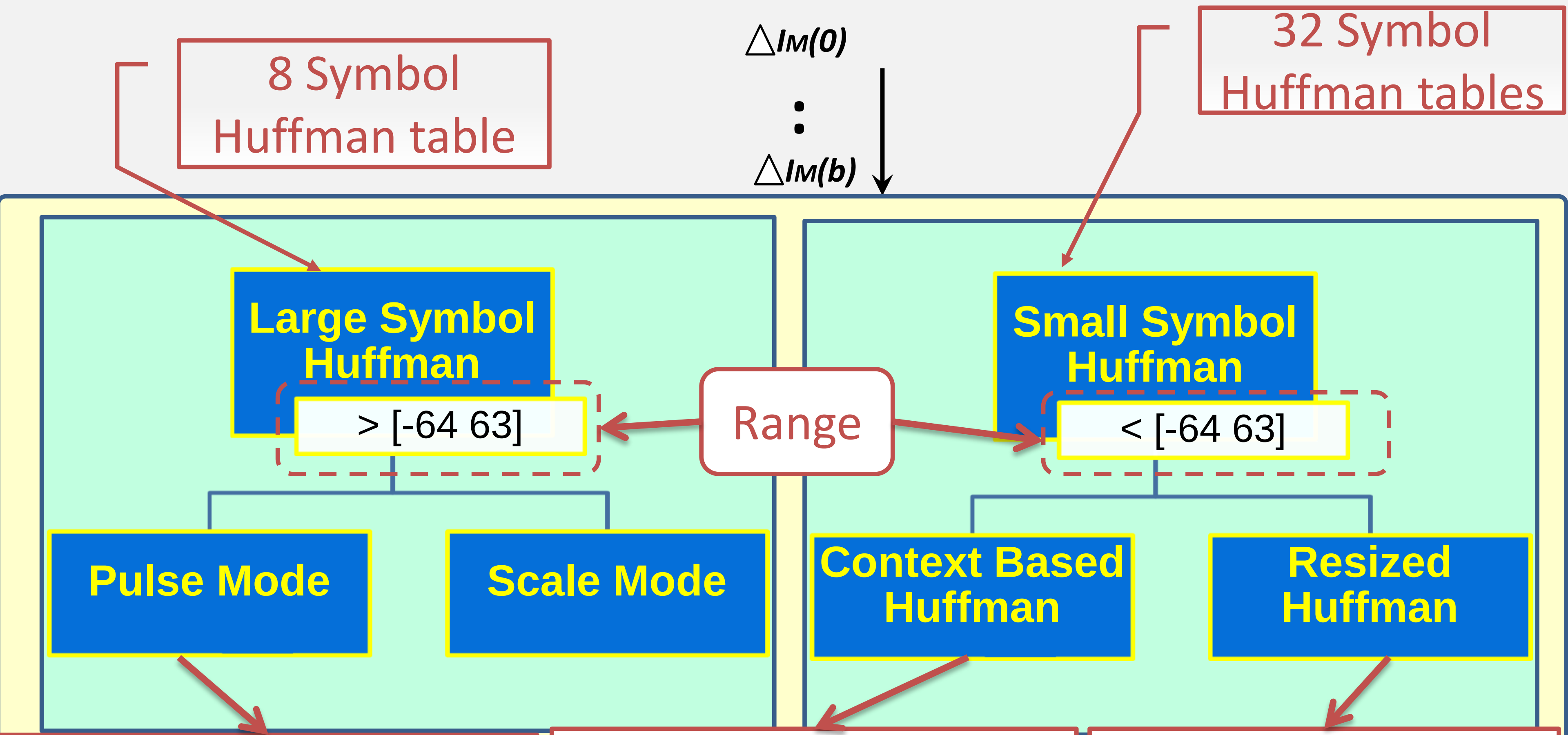
Spectral Coefficients are grouped Spectral bands

$E_M(1)$ $E_M(2)$ •• $E_M(b-1)$ $E_M(b)$

Spectral band energies

LR-HQ MDCT Coder

-Envelope coding (2/2)



If $\triangle IM(b)$ exceeds [-4 3]

1. Position
2. Amplitude of $\triangle IM(b)$ coded directly

$\triangle IM(b-1)$ determines **best** Huffman table for encoding the current band $\triangle IM(b)$

$\triangle IM(b)$ Span is narrowed down to fewer code words (21 symbols)

Small Symbol coding method

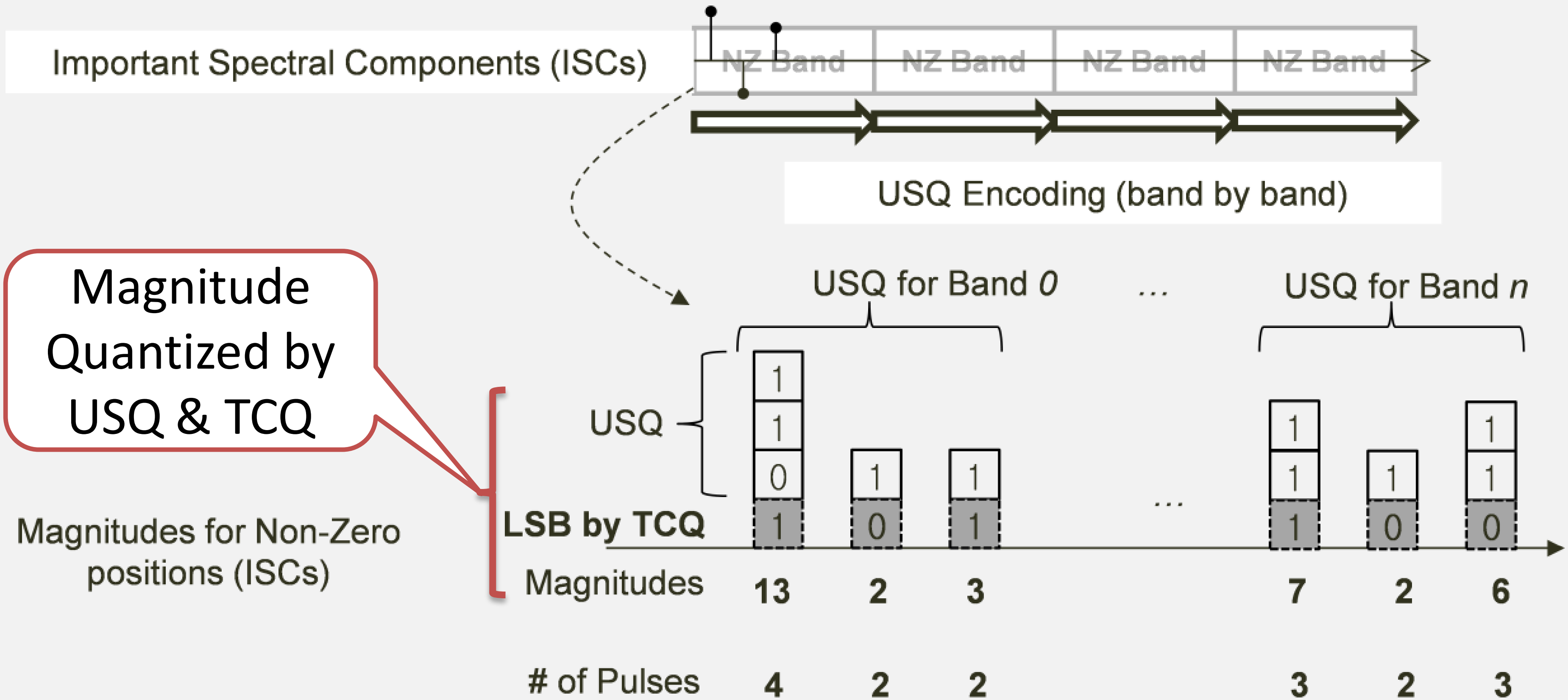
- **Context based Huffman:** $\triangle_{IM(b-1)}$ determines **best** Huffman table for encoding the current band $\triangle_{IM(b)}$
- **Resized Huffman :** $\triangle_{IM(b)}$ narrowed to a smaller range for using Huffman table with fewer symbols (21 symbols).

$$\Delta I'_M(b) = \begin{cases} \Delta I_M(b) + \min(\Delta I_M(b\text{-}1) - T, 3), & \Delta I_M(b\text{-}1) > T \\ \Delta I_M(b) + \max(\Delta I_M(b\text{-}1) - T^1, -3), & \Delta I_M(b\text{-}1) < T \end{cases}$$

$\Delta I'_M(b)$ is the the new differenti al index for band b

$T = 15 + thr, T^1 = 15 - thr$

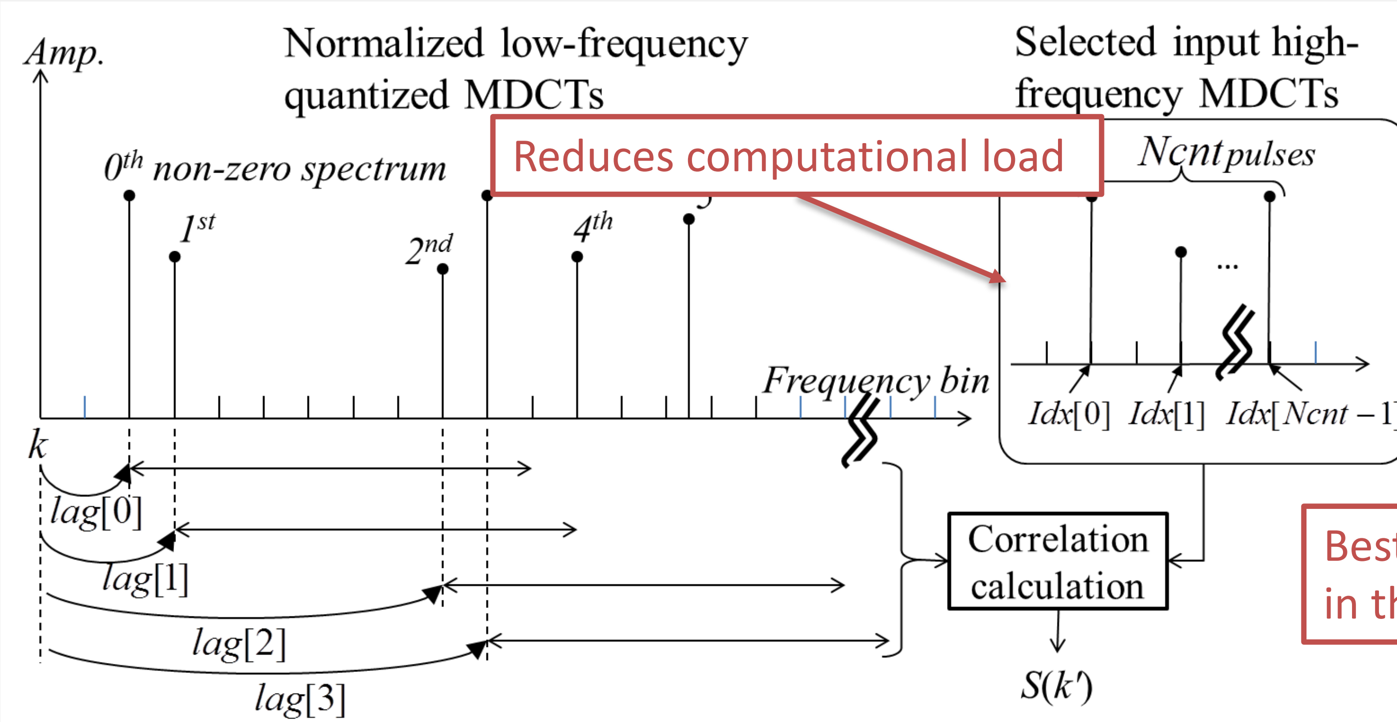
LR-HQ MDCT Coder
-Quantization (TCQ and USQ)



Magnitude Quantized by USQ & TCQ

Position, Number and sign coded by Arithmetic coding

LR-HQ MDCT Coder
-Sparse Band Search



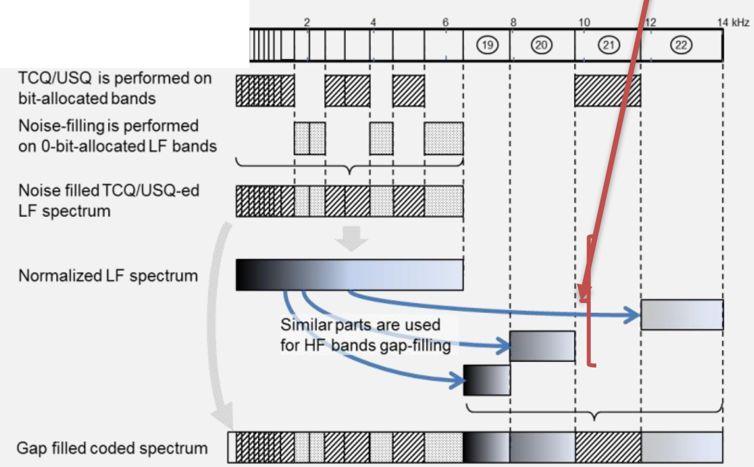
Best match spectrum filled in the High Frequency Region

Best match index is identified by finding the k' which maximizes the correlation measure, $S(k')$ according to

$$S(k') = corr(k')^2 / Ene(k'), \quad k' = 0, \ldots, Nlag - 1$$

$$corr(k') = \sum_{k=0}^{Ncnt-1} X(Idx[k]) \tilde{X}(k + lag[k'] + Idx[k])$$

$$Ene(k') = \sum_{k=0}^{Ncnt-1} \tilde{X}(k + lag[k'] + Idx[k])^2$$



Gap filling for Normal mode