Multiple Description Audio Coding for Wireless Low-Frequency Sound Zones

Jan Østergaard[1], Christian S. Pedersen[1], Mo Zhou[1], Niels de Koeijer[2], and Martin B. Møller[2]

[1]Section on Artificial Intelligence and Sound
Department of Electronic Systems
Aalborg University
Denmark

[2]Bang & Olufsen
Bang og Olufsen Allé 1
7600 Struer
Denmark

IEEE Data Compression Conference, 2023

Story line

- Introduction to personal sound zones
- Low frequency (<600 Hz) sound zone generation
- Wireless low frequency sound zones could suffer from packet losses
- Solution
  - Combining multiple descripton audio coding with sound zone control
- Simulation study
- Conclusions

Introduction to personal sound zones

Zone A

Zone B

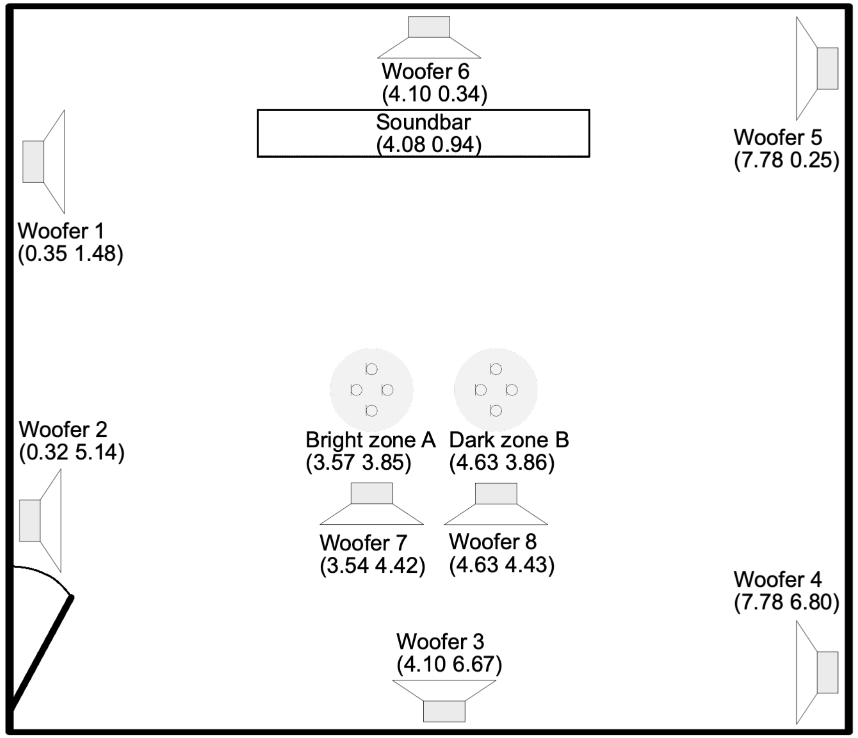
Rock in Zone A

"Leakage"

"Leakage"

Classical in Zone B

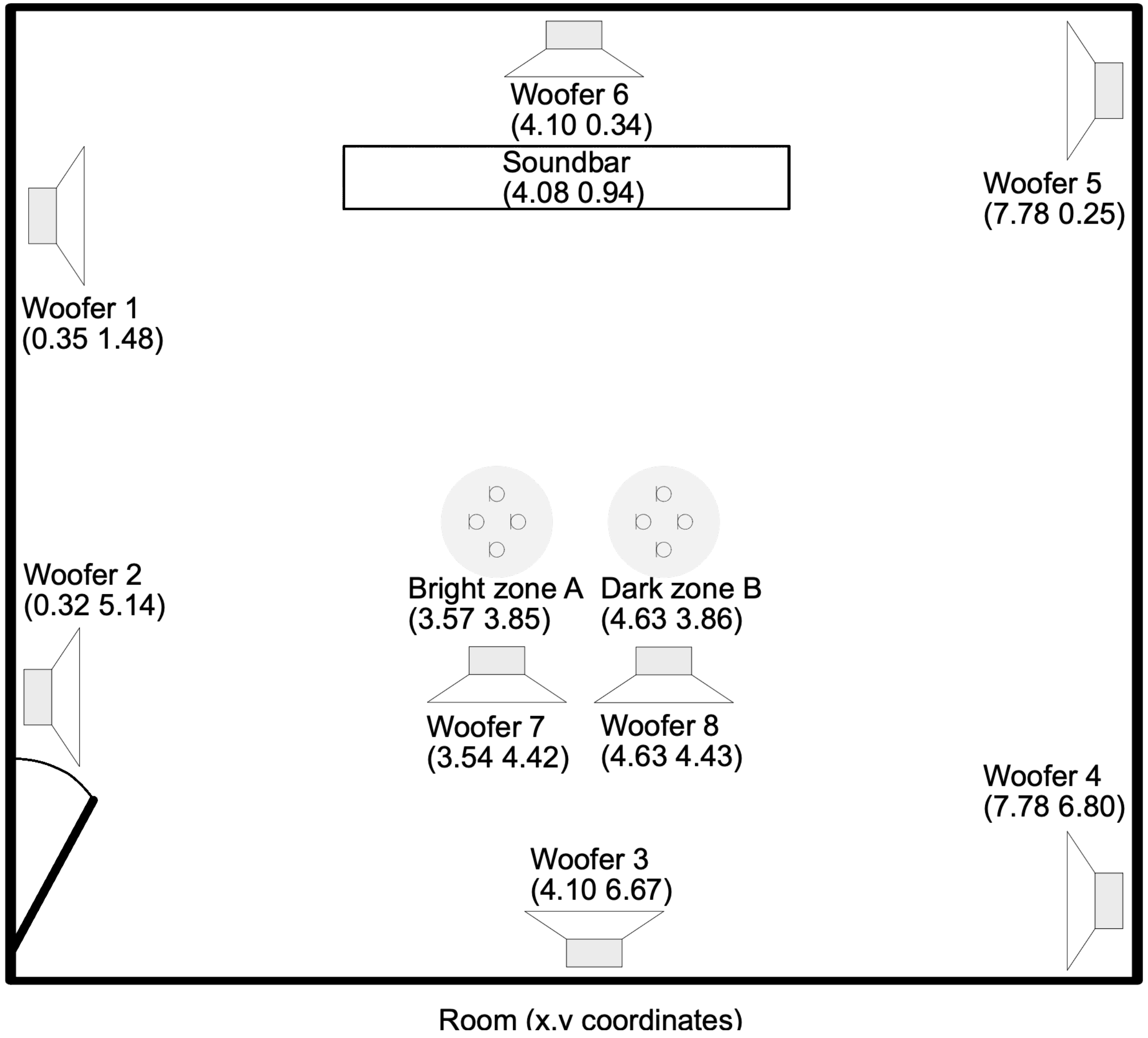
Both Zones playing



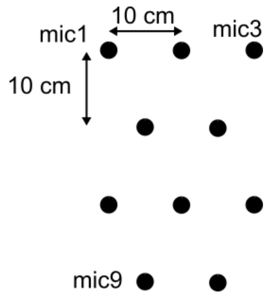
Room (x,y coordinates)

Room photo lab prototype

Personal sound zones- low frequencies – below 600 Hz: Feedforward sound field control based on transfer functions

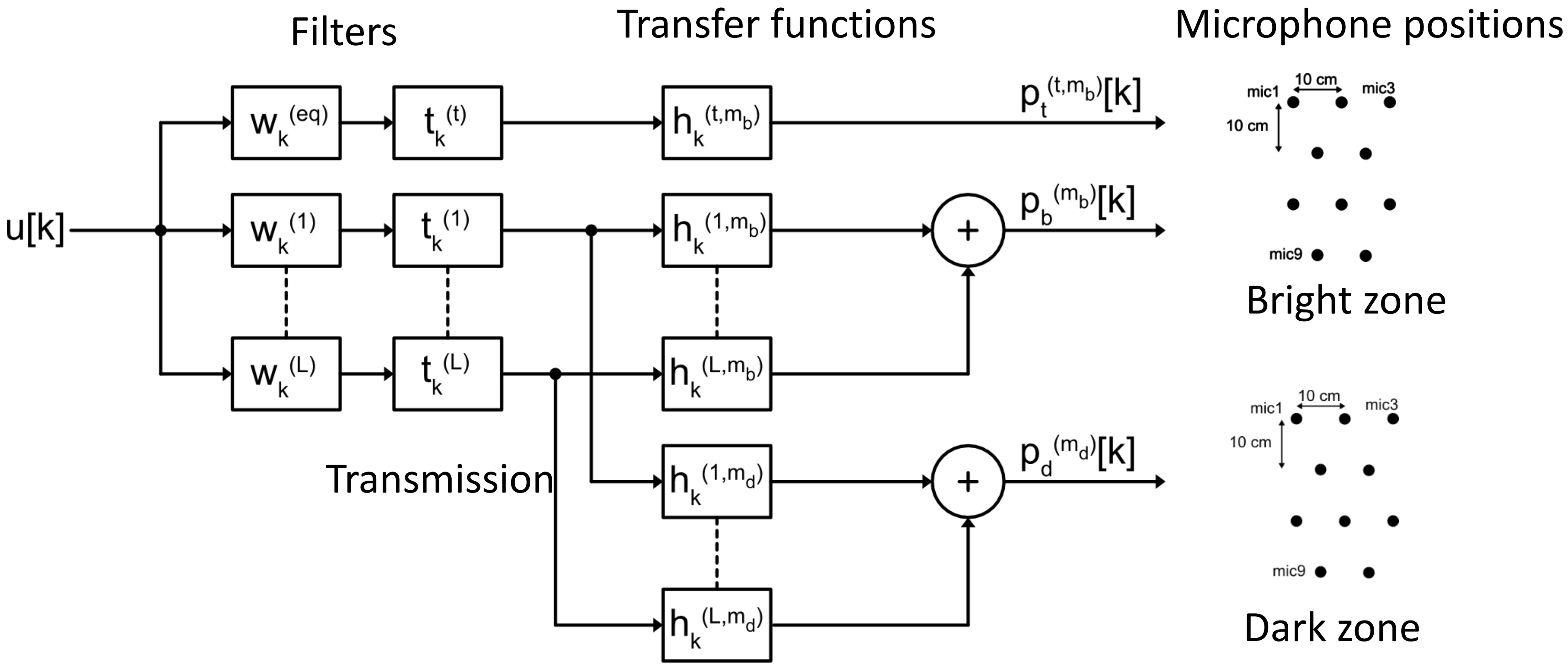


Two heights
of 10 mics
=20 mic positions

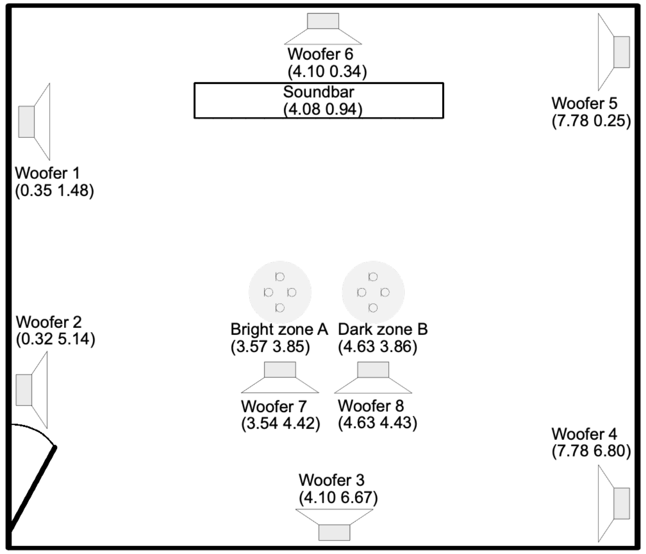


Sound zone generation at low frequencies

- Pressure matching in time domain by minimizing error between desired and actual pressure in the microphone positions in the bright and dark zone. [Galvez et al.’15]
- By superposition several sound zones can be made



Sound field control



- When there are L woofers, the sound pressure level in the bright zone at the position of the m-th microphone is given by the super position of the signals from all L woofers:

Sound pressure:

$$p_b^{(m)}(n) = \sum_{\ell=1}^{L} \sum_{j=0}^{N_w-1} \sum_{i=0}^{N_h-1} \bar{h}_b^{(m,\ell)}(i)\bar{w}^{(l)}(j)u(n-i-j).$$

Average energy:

$$P_{\text{bright}} \triangleq \frac{1}{N_u M} \sum_{m=1}^{M} \sum_{n=0}^{N_u-1} |p_b^{(m)}(n)|^2, \quad P_{\text{dark}} \triangleq \frac{1}{N_u M} \sum_{m=1}^{M} \sum_{n=0}^{N_u-1} |p_d^{(m)}(n)|^2.$$

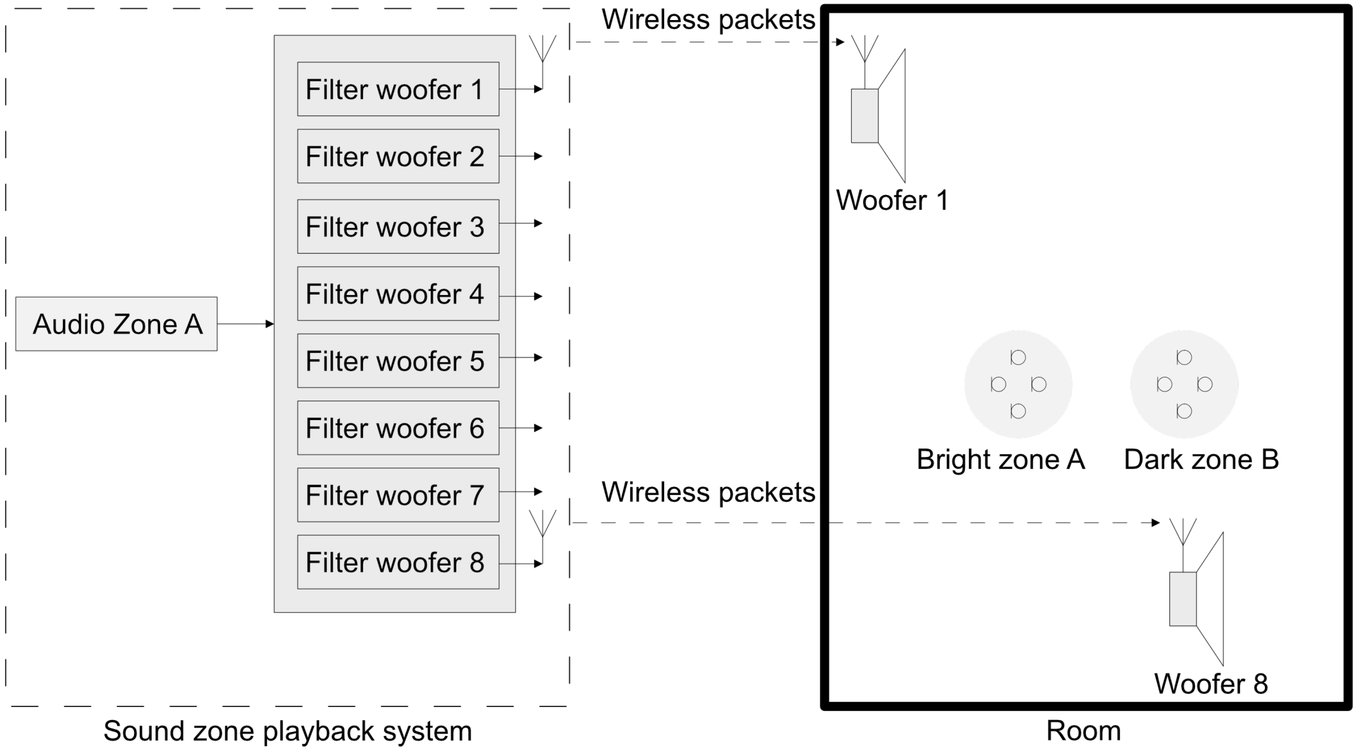
Acoustic contrast ratio:

$$C = 10 \log_{10} \left(\frac{P_{\text{bright}}}{P_{\text{dark}}} \right) [\text{dB}].$$

Wireless transmission to woofers

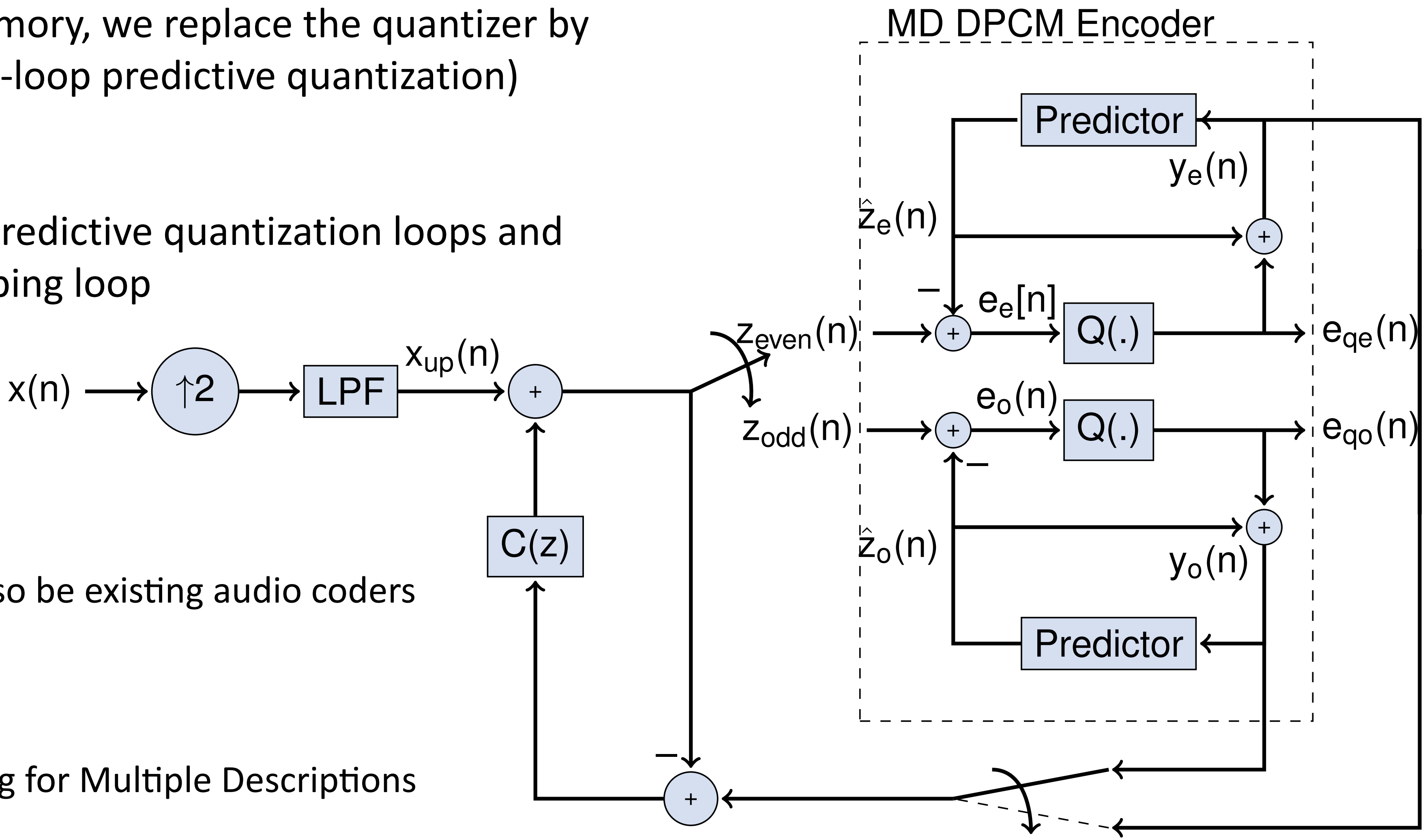
- Distributed woofers makes wireless transmission much more practical – e.g., no wires needed from the sound zone system.
- Packet losses could potentially occur and degrade the resulting performance.

	Zone A	Zone B
No packet losses		
5% iid packet losses only on woofer 7		



Noise-shaping & source prediction

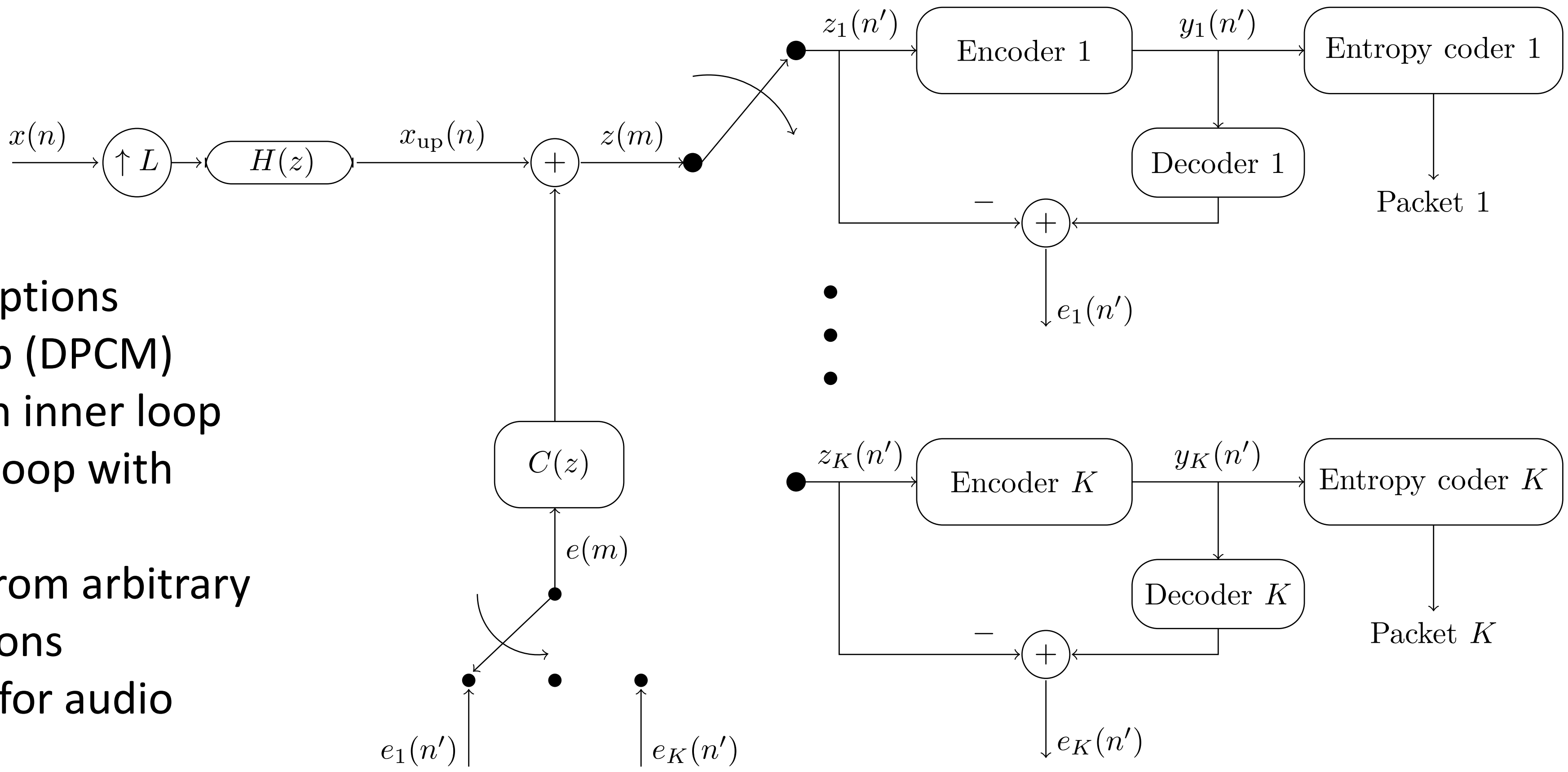
- For sources with memory, we replace the quantizer by a DPCM loop (closed-loop predictive quantization)
- We have two inner predictive quantization loops and one outer noise-shaping loop



The DPCM loop can actually also be existing audio coders

Noise-Shaped Predictive Coding for Multiple Descriptions of a Colored Gaussian Source.
[Y. Kochman, J. Østergaard, R. Zamir. IEEE Data Compression Conference, 2008.]

Multiple descriptions by fractional sampling

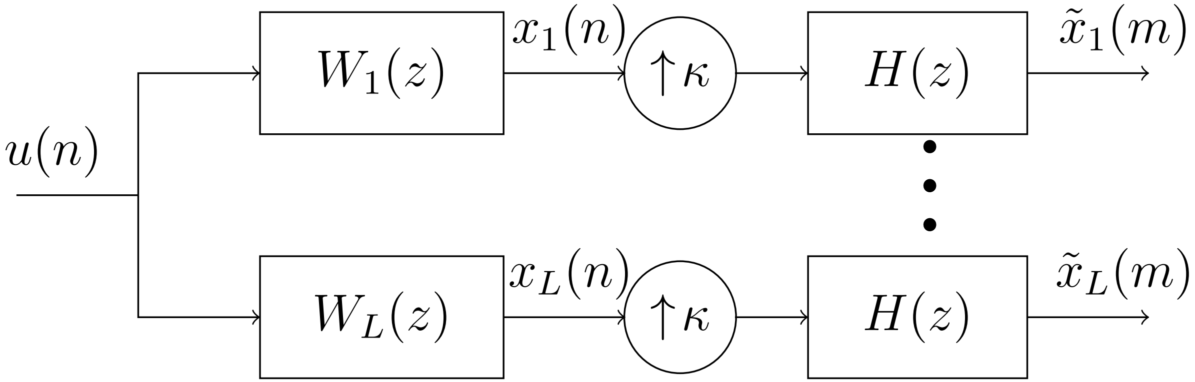


- Upsample by L ≥ 2
- Create K >= L descriptions
- Perform closed-loop (DPCM) quantization in each inner loop
- Perform one outer loop with noise shaping
- Perform decoding from arbitrary subsets of descriptions
- Decoder optimized for audio

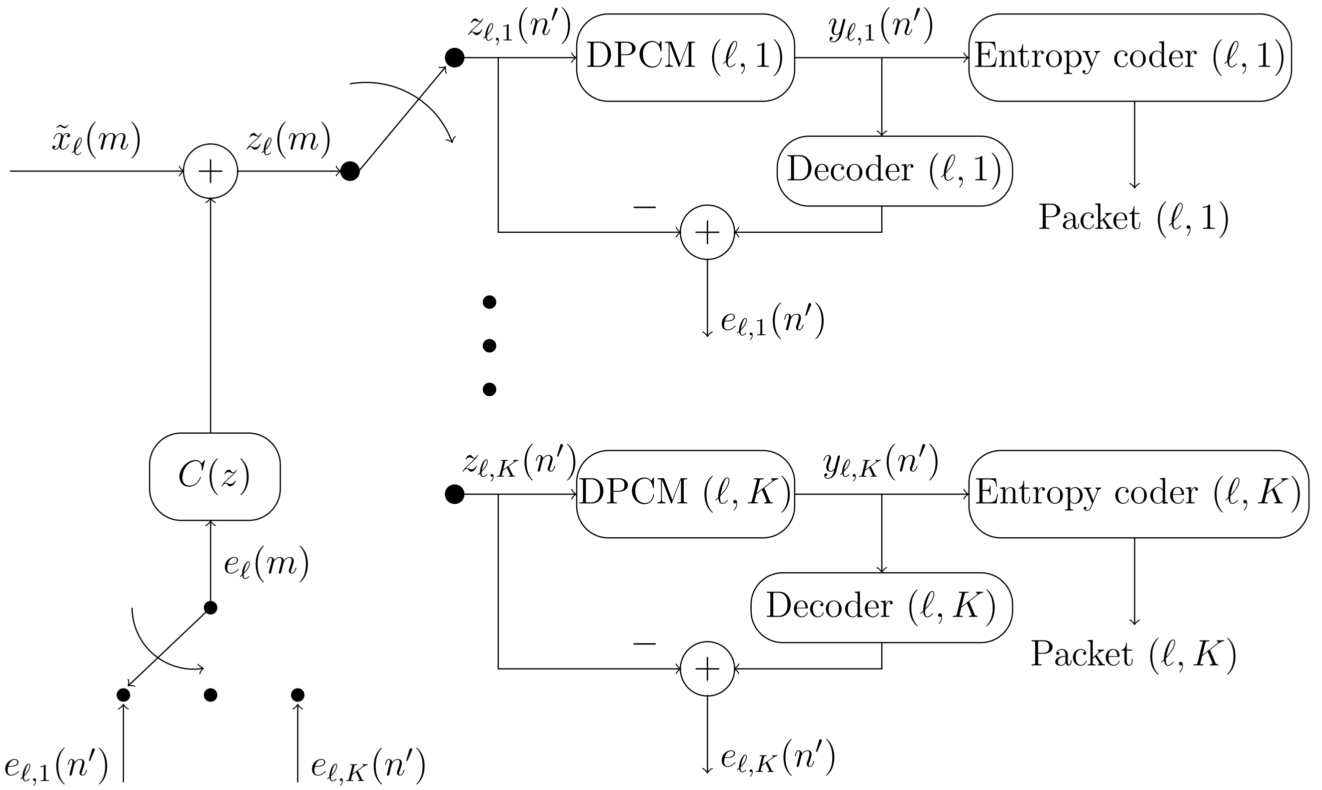
[Østergaard, DCC 2021]

Oversample control signals

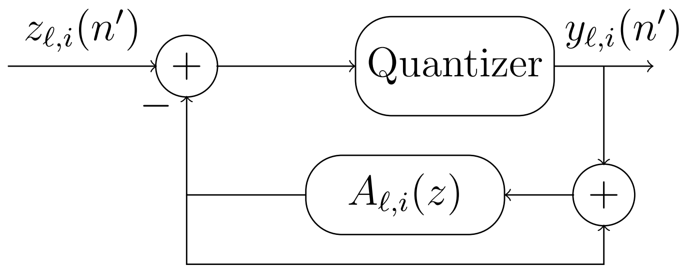
We first oversample the control signals to the woofers

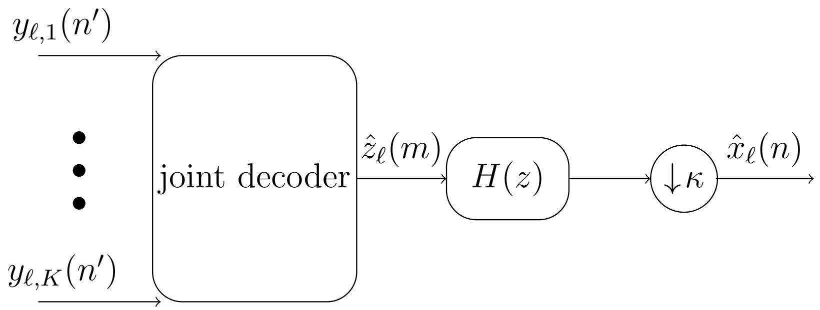


The oversampled control signal from woofer l is sent through the MDC encoder resulting in K descriptions.



DPCM encoder and joint decoder





- Each of the K fractionally sampled signals are sent through closed-loop DPCM encoder following by entropy coding.
- The subset of received descriptions are first indivially ”inverse” DPCM coded, and finally jointly combined and and resampled to the original sampling frequency.
- The reconstructed signal is played out by the l-th woofer
- The above is performed in parallel for all L woofers, and we assume all woofers are synchronized in time.

Design of optimal control filters

Cost function:

$$J_{\mathrm{mdc}}(\{\bar{w}^{(\ell)}\}_{\ell=1}^{L}) = (1-\beta)\sum_{m,n}|\hat{p}_b^{(m)}(n) - \tilde{p}^{(m)}(n)|^2 + \beta\sum_{m,n}|\hat{p}_d^{(m)}(n)|^2 + \gamma\sum_{\ell=1}^{L}\|\bar{w}^{(\ell)}\|^2,$$

Expected packet losses are taken into Account:

$$\hat{p}_b^{(m)}(n) = \mathbb{E}\Big[\sum_{\ell=1}^{L}\sum_{i=0}^{N_h-1}\bar{h}_b^{(m,\ell)}(i)\hat{x}_\ell(n-i)\Big]$$
$$= \sum_{\mathcal{I}_n\subseteq\{1,\ldots,K\}^L} P(\mathcal{I}_n|\mathcal{I}^{n-1})\sum_{\ell=1}^{L}\sum_{i=0}^{N_h-1}\bar{h}_b^{(m,\ell)}(i)\mathbb{E}[\hat{x}_\ell^{\mathcal{I}_n^\ell}(n-i)|\mathcal{I}^n].$$

Closed-form least-squares solution:

$$W = \Big(\sum_{\mathcal{I}_n\subseteq\{1,\ldots,K\}^L} P(\mathcal{I}_n|\mathcal{I}^{n-1})((1-\beta)(\phi(\mathcal{I}_n)H_b)^T\phi(\mathcal{I}_n)H_b$$
$$+\beta(\phi(\mathcal{I}_n)H_d)^T\phi(\mathcal{I}_n)H_d) + \gamma I\Big)^{-1} \times \sum_{\mathcal{I}_n\subseteq\{1,\ldots,K\}^L}(P(\mathcal{I}_n|\mathcal{I}^{n-1})\phi(\mathcal{I}_n)H_b\tilde{P}_r)$$

$$W = [\bar{w}^{(1)T}, \ldots, \bar{w}^{(L)T}]^T$$

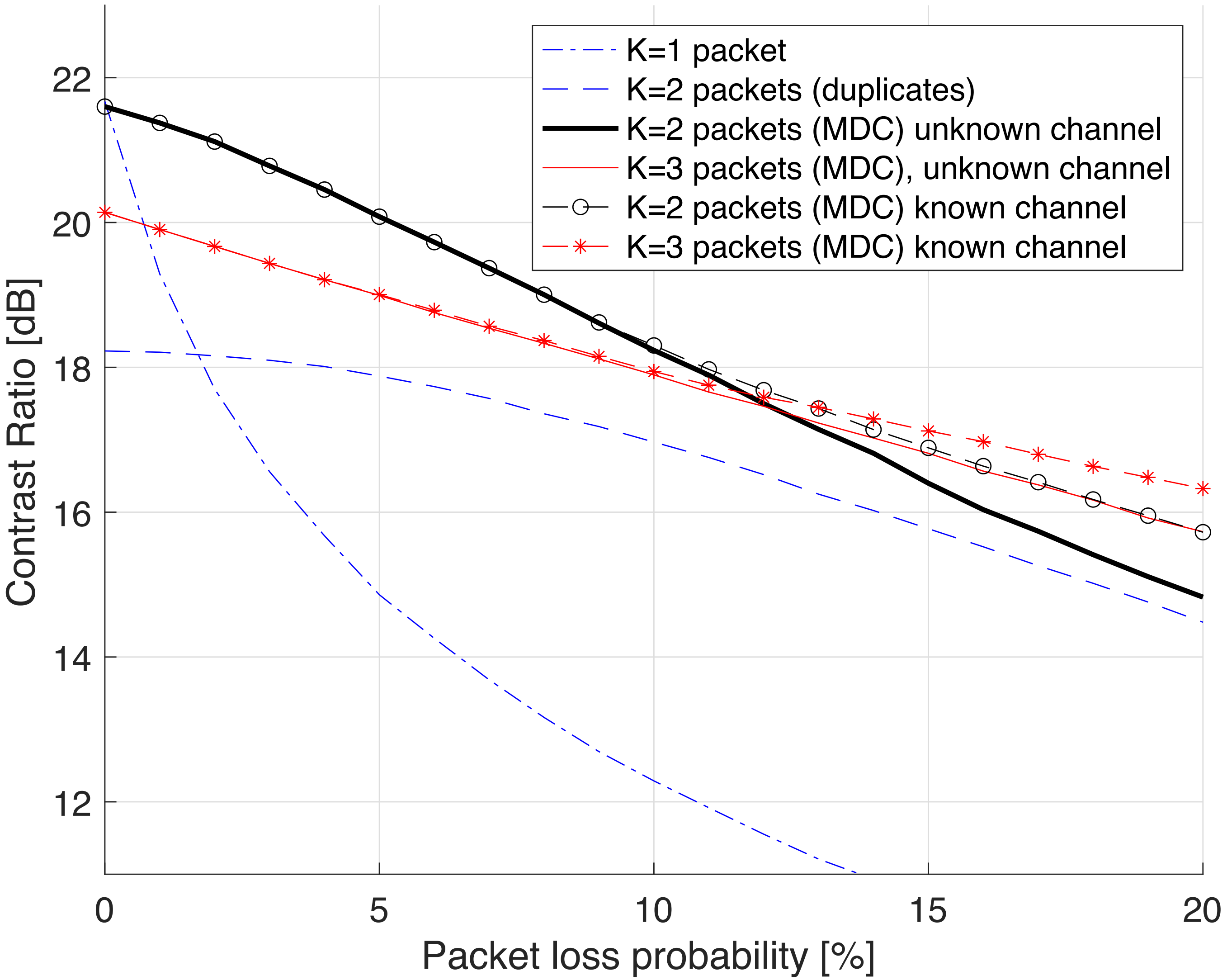
Simulation study

- We used 60 seconds of pop music sampled at 1200 Hz.
- We use block sizes of 20 ms corresponding to 24 samples.
- We used simulated room impulse response functions from a room of size 5.5 x 8.65 x 2.7 meters.
- We used L = 8 woofers.
- Note that L*K packets needs to be transmitted

- The coding rate includes the bits required for coding the quantized AR predictor coefficients and the quantized residual.
- For a K=2 description system, it was sufficient to use a predictor order of 2, while keeping a desired acoustic constrast ratio.

Predictor order	Performance	Rate residual	Rate LSFs	Total rate
1	22.273 dB	65.06 kbps	2.89 kbps	67.95 kbps
2	22.275 dB	53.85 kbps	6.00 kbps	59.85 kbps
3	22.271 dB	53.13 kbps	7.94 kbps	61.07 kbps

Simulation results for 60 kbps total rate



For the case unknown channel we designed for 0% packet loss rates.

Conclusions

- We proposed a joint framework that combined audio coding, multiple descriptions, and sound zone control.
- We used FIR filtering for the sound zone control, and predictive (DPCM) coding combined with oversampling and noise-shaping in order to obtain an audio coder, which is robust to packet losses.
- A closed-form expression for the optimal sound zone control filters were provided, which takes into the account the packet loss rate.
- Significant gain in acoustic contrast was demonstrated even at small to moderate packet losses.

Thanks to all project partners

Acknowledgements:
This work is partly funded by the Innovation Fund Denmark (IFD) under File No 9069-00038B in the project: Interactive Sound Zones for Better Living (ISOBEL)
www.isobel.dk