A Flexible Audio Coding Platform Based on the Autoregressive Model

Alexey Ozerov and W. Bastiaan Kleijn

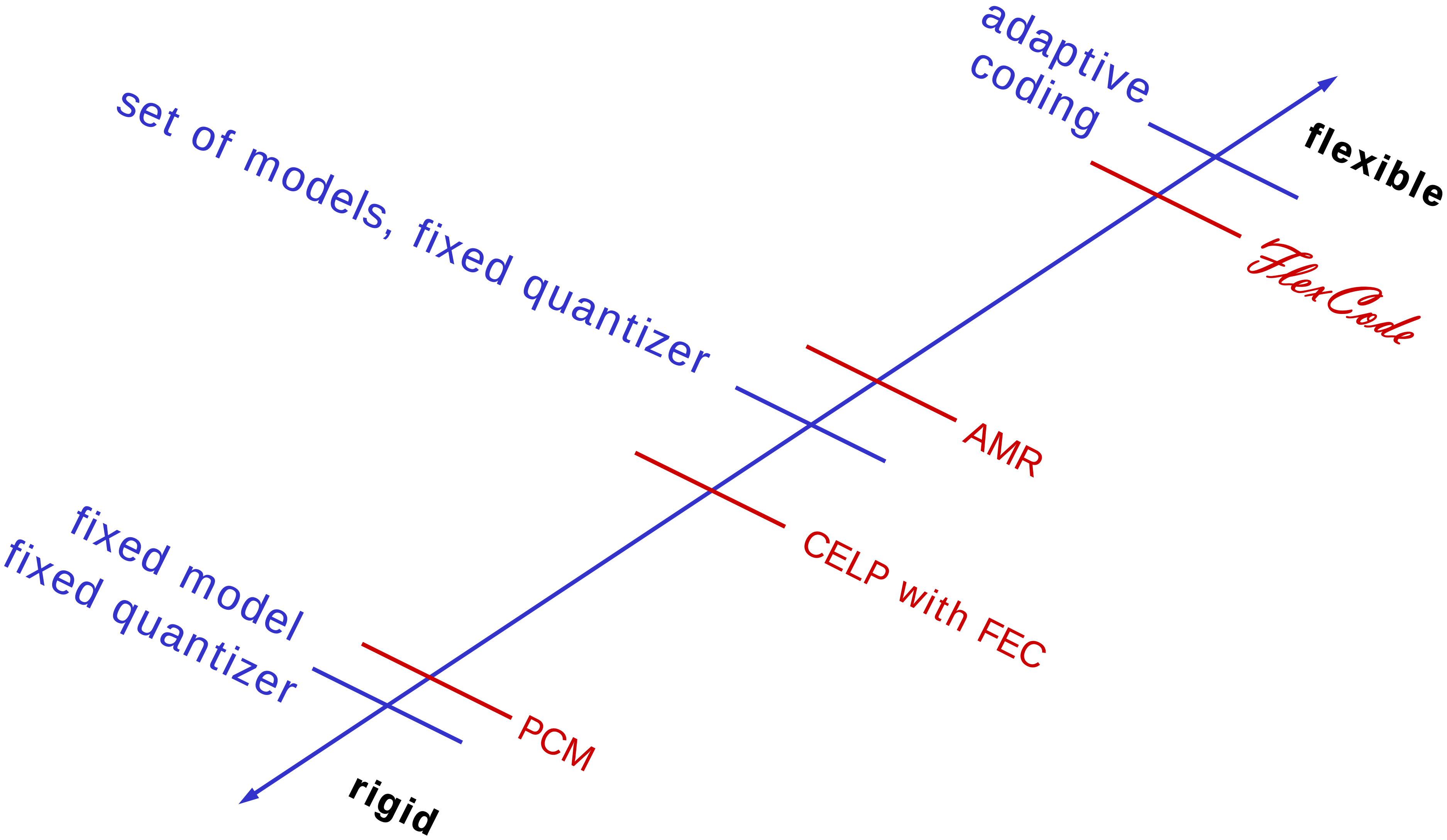
Royal Institute of Technology, Stockholm, Sweden

Public Seminar at Ericsson
October 17, 2007

Outline

- Motivation
- Basics of Adaptive Quantization under High-Rate Assumptions
- Flexible Audio Coding Scheme
- Experimental Illustration
- Further Work
- Conclusion

Adaptation and Coding



Global Goals

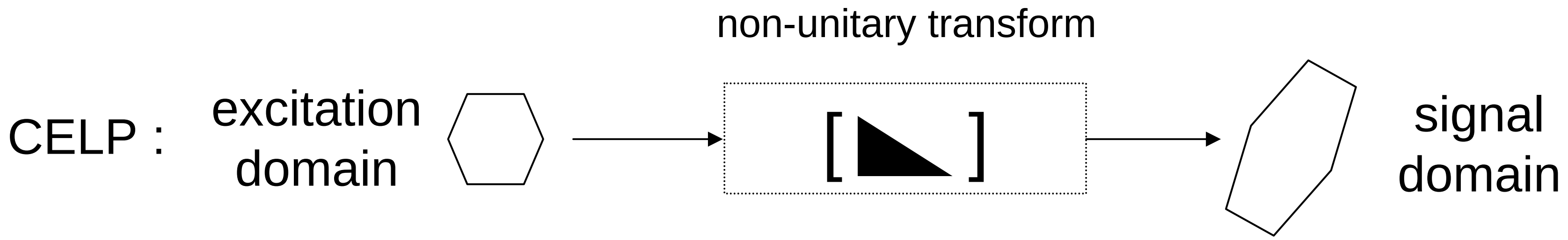
- In *FlexCode* project we want to develop a flexible audio coder that
  - performs as good as the state-of-the-art speech and audio coders
  - is able to run for any rate from the continuum of the rates
  - has a computational complexity which is independent on the rate
  - uses an advanced perceptual measure
  - is able to adapt according to some feedback from the transmission channel

AR Model

- Autoregressive (AR) model is good to model both speech and audio signals
- Thus, we start to develop a flexible coder based on the AR model
- We consider Code Excited Linear Predictive (CELP) coding scheme as a starting state-of-the-art reference

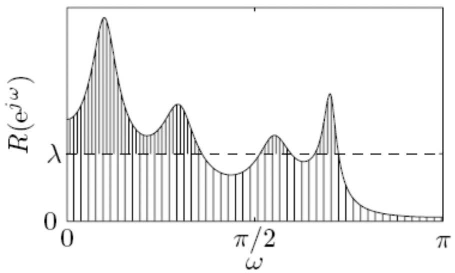
CELP Shortcomings

- A conventional CELP coding scheme
  - is not adaptable for any particular rate (codebook (CB) must be re-trained for every particular rate)
  - has a computational complexity growing linearly with CB-size (or exponentially with rate)
  - does not allow the rate to be varied continuously in time
  - has a fixed “optimal” CB in the excitation domain, which is mapped every time to the signal domain by some non-unitary transform



CELP Shortcomings

- A conventional CELP coding scheme
  - does not account for the so-called reverse water-filling effect



original power spectrum

reconstruction power spectrum

distortion

- In the flexible audio coding scheme that we propose, we overcome all the abovementioned shortcomings of CELP

Adaptive Quantization

- Since we need a coder that is able to run for any rate from the continuum of the rates,
  - we cannot train and store the codebooks
  - we need adaptive codebooks, which can be computed on the fly
- Probabilistic source modeling together with high-rate theory approximation allows that

Outline

- Motivation
- Basics of Adaptive Quantization under High-Rate Assumptions
- Flexible Audio Coding Scheme
- Experimental Illustration
- Further Work
- Conclusion

Two Types of Quantization

- Constrained Resolution (CR) quantization
  - Fixed number of bits per vector
  - R bits per vector = 2^R codewords in the codebook
- Constrained Entropy (CE) quantization
  - Any number of bits per vector (variable rate)
  - The average rate (or the entropy of the codeword indices) is constrained

$$H(I) = -\sum_{i \in \mathcal{I}} p_I(i) \log(p_I(i)) = R$$

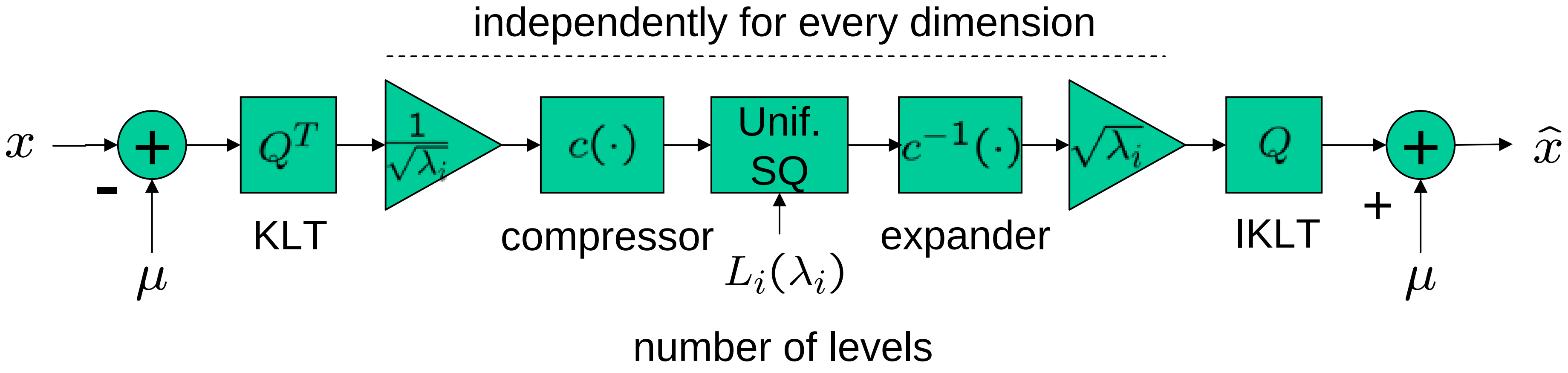
- CE performs better than CR, but needs a more flexible transmission channel

Single Gaussian case

- CR quantization (with companded scalar quantizers)

$X \in \mathbb{R}^k, \qquad X \sim \mathcal{N}(\mu, \Sigma)$

EVD $\quad \Sigma = Q\Lambda Q^T \qquad Q^TQ = I \qquad \Lambda = \mathrm{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_k\}$

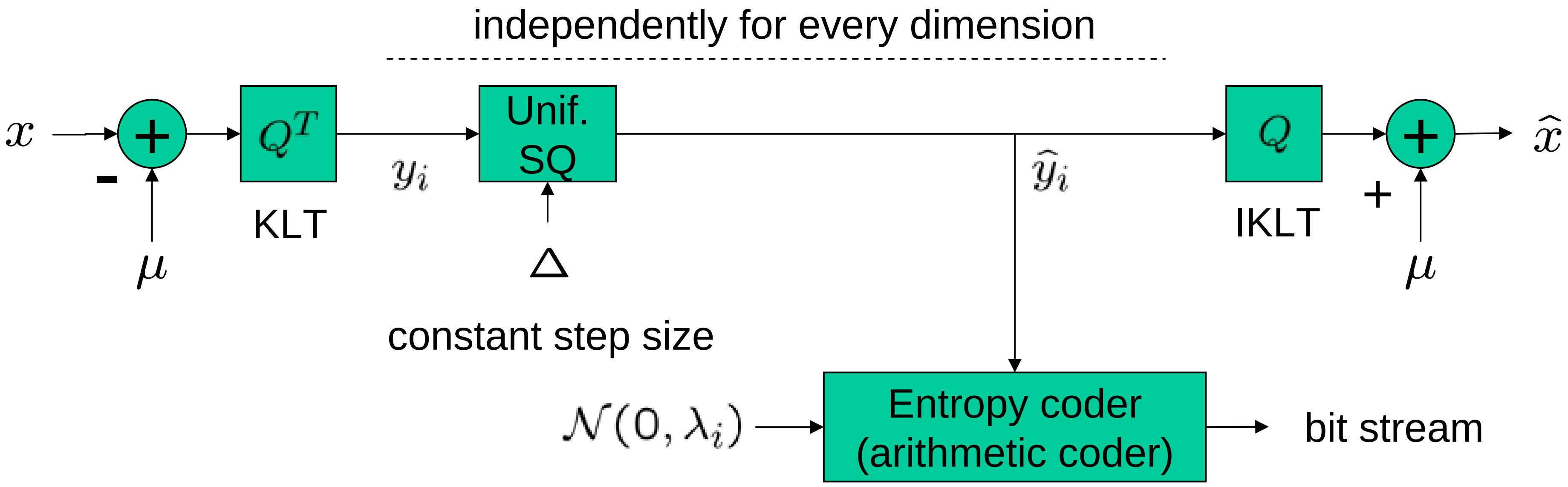


Single Gaussian case

- CE quantization (with scalar quantizers)

$$X \in \mathbb{R}^k, \qquad X \sim \mathcal{N}(\mu, \Sigma)$$

EVD $\quad \Sigma = Q \Lambda Q^T \qquad Q^T Q = I \qquad \Lambda_i = \mathrm{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_k\}$



Single Gaussian case

- We see that with this approach
  - we can quantize with any rate
  - the computational complexity is independent on the particular rate

- We can do better using vector lattice quantizers instead of scalar quantizers
  - we can gain up to 0.25 bits per sample in rate
  - which is equivalent to 1.5 dB in distortion

Gaussian Mixture Model (GMM) case

- With GMM the quantization consists in the following steps:
  - For each input vector *x*, choose the component (state) maximizing the *a posteriori* probability *p(i|x)*
  - Quantize using selected Gaussian component (in CR or CE case), as described before

- With this approach we loose in optimality, when the Gaussian components are not well separated

Outline

- Motivation
- Basics of Adaptive Quantization under High-Rate Assumptions
- Flexible Audio Coding Scheme
- Experimental Illustration
- Further Work
- Conclusion

Signal Model

- Input signal is segmented into frames: s

- To account for within frame redundancy
  - – we use AR model

- To account for between frames redundancy
  - – we use “ringing” (or zero impulse response) subtraction

Signal Model

s frame to be quantized

"ringing" AR model: LPC and excitation variance

r

$$\frac{\sigma_e}{A(z)} = \frac{\sigma_e}{1+a_1 z^{-1}+...+a_p z^{-p}}$$

then $\boxed{s \sim \mathcal{N}(r, \Sigma)}$

with $\Sigma = A^{-1} A^{-T}$

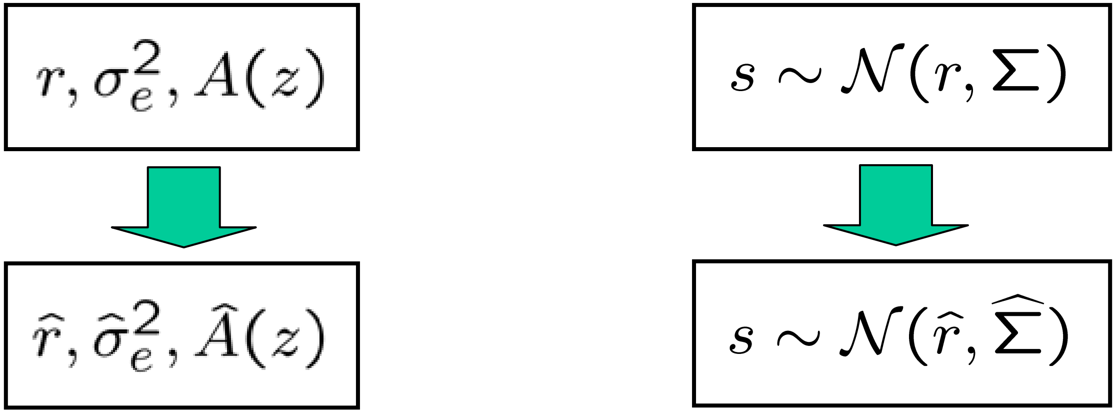
A is a lower triangular Toeplitz (k x k) matrix with as first column

$$\sigma_e^{-1}[1, a_1, a_2, \ldots, a_p, 0, \ldots, 0]^T$$

Model Quantization

- AR model parameters are quantized and transmitted as well (forward adaptation)

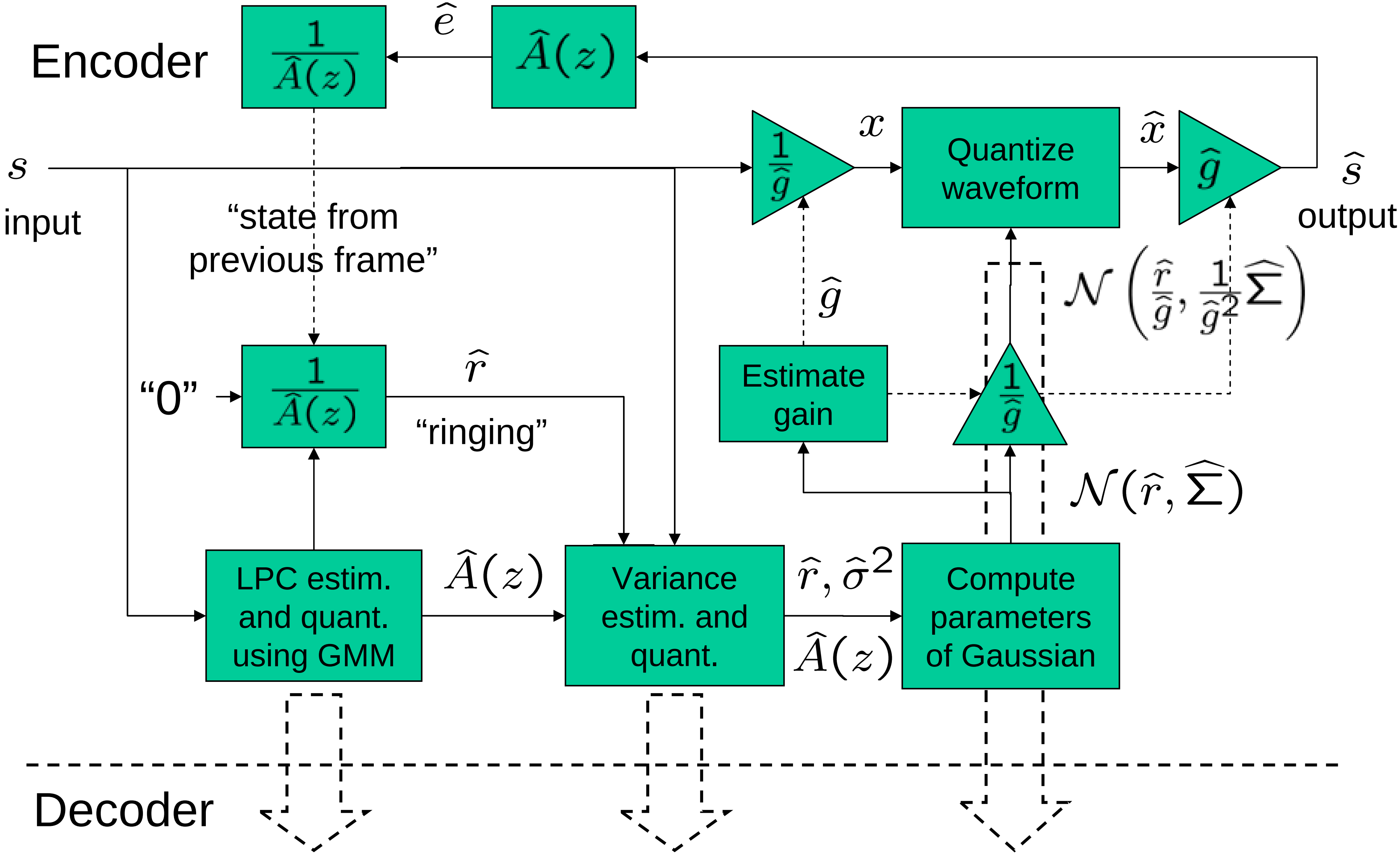
- Thus we are using quantized model parameters rather than non-quantized



$$r, \sigma_e^2, A(z) \rightarrow \hat{r}, \hat{\sigma}_e^2, \hat{A}(z)$$

$$s \sim \mathcal{N}(r, \Sigma) \rightarrow s \sim \mathcal{N}(\hat{r}, \widehat{\Sigma})$$

Encoder structure



Encoder
$\frac{1}{\widehat{A}(z)}$
$\widehat{e}$
$\widehat{A}(z)$
s
input
$\frac{1}{\widehat{g}}$
x
Quantize waveform
$\widehat{x}$
$\widehat{g}$
$\widehat{s}$
output
"state from previous frame"
$\widehat{g}$
$\mathcal{N}\left(\frac{\widehat{r}}{\widehat{g}}, \frac{1}{\widehat{g}^2}\widehat{\Sigma}\right)$
"0"
$\frac{1}{\widehat{A}(z)}$
$\widehat{r}$
"ringing"
Estimate gain
$\frac{1}{\widehat{g}}$
$\mathcal{N}(\widehat{r}, \widehat{\Sigma})$
LPC estim. and quant. using GMM
$\widehat{A}(z)$
Variance estim. and quant.
$\widehat{r}, \widehat{\sigma}^2$
$\widehat{A}(z)$
Compute parameters of Gaussian
Decoder

Gain estimation

- This gain should be considered as a part of the model of perception
- Have a sense for CE case only

$$\boxed{\hat{g} = \sqrt{\frac{1}{k}E[s^T s]} = \sqrt{\frac{1}{k}(\hat{r}^2 + \mathrm{tr}(\widehat{\Sigma}))}}$$

- More details about this coding scheme:
  - A. Ozerov, and W. B. Kleijn, "Flexible quantization of audio and speech based on the autoregressive model," In *IEEE Asilomar Conference on Signals, Systems, and Computers*, Nov. 2007.

Rate distribution between model and signal

- General result:
  - When the signal is quantized based on some already quantized model and the HR assumptions are verified, the optimal rate for the model is constant, i.e., it is independent on the overall rate
  - This result is true for any model and in both CR and CE cases
- More details about this result:
  - W. B. Kleijn, and A. Ozerov, "Rate distribution between model and signal," In *IEEE Worksh. on Apps. of Signal Processing to Audio and Acoustics (WASPAA'07)*, Mohonk, NY, Oct. 2007.

Rate distribution between model and signal

- This result significantly simplifies design of the presented coding scheme, and in particular
  - first, it means that when the total rate changes, only the rate spent for signal quantization should be adjusted, and the rate for model should be kept constant,
  - second, one do not need the quantizer for model (but not for signal) to be flexible, any quantizer can be used in principle.

Computational complexity and storage requirements

- Computational complexity
  - – is quite low (except EVD computation)
  - – is independent on the rate
- Storage requirements
  - – are very low (only GMM model parameters used for LPC quantization must be stored)

Outline

- Motivation
- Basics of Adaptive Quantization under High-Rate Assumptions
- Flexible Audio Coding Scheme
- Experimental Illustration
- Further Work
- Conclusion

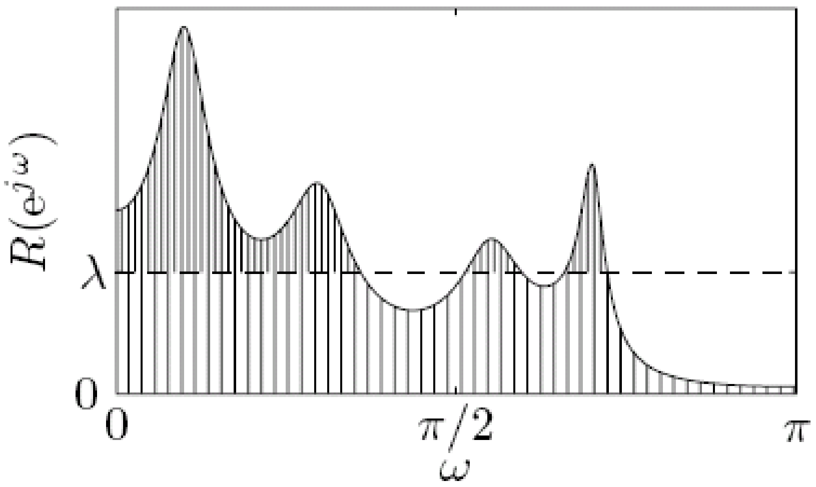
Comparison with a CELP scheme

- We compare with some CELP scheme, with a CB trained minimizing MSE in signal domain
  - 8 kHz speech, frame length = 5 samples,
  - Rate = 19.2 kbps (12 bits per frame)

	AR coder (CR case)	AR coder (CE case)	CELP
Variance rate (bpf)	3	2.7	5
Signal rate (bpf)	9	9.2	7
Av. SSNR (dB)	16.66	17.96	17.82

  - This is with scalar quantizers, and for quite low rate
  - By increasing frame length up to 10 samples, we have
    - 18.84 dB in CR case for the same rate, and
    - 20.43 dB in CE case for the same rate

Reverse water-filling



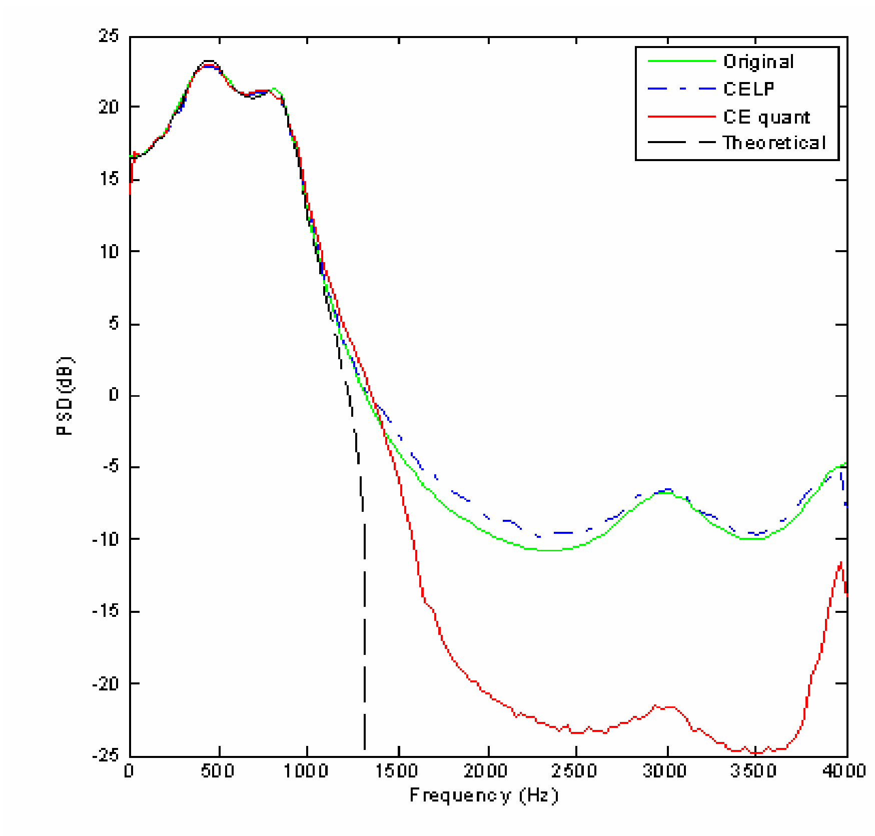
original power spectrum

reconstruction power spectrum

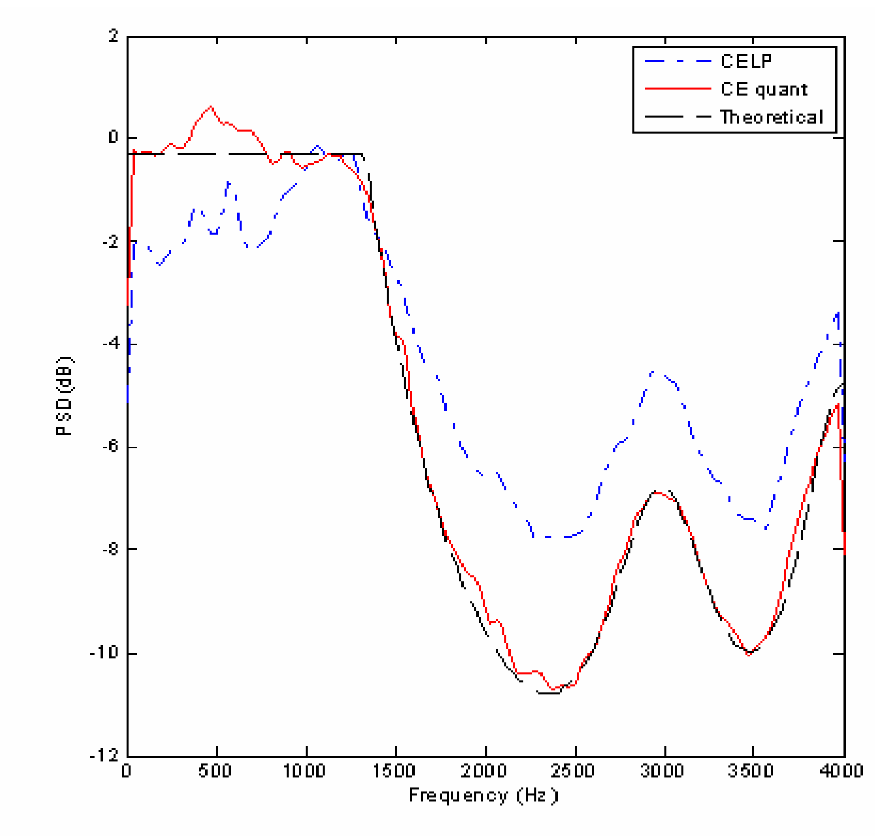
distortion

Reverse water-filling

reconstruction
power spectrum



distortion power
spectrum



Demo

External demo

Back to CELP Shortcomings

- A conventional CELP coding scheme
  - is not adaptable for any particular rate (codebook (CB) must be re-trained for every particular rate)
    - proposed scheme can run for any rate
  - has a computational complexity growing linearly with CB-size (or exponentially with rate)
    - proposed scheme has complexity independent in the rate
  - does not allow the rate to be varied continuously in time
    - proposed scheme allows the rate to be varied in time (CE quantization)
  - has a fixed “optimal” CB in the excitation domain, which is mapped every time to the signal domain by some non-unitary transform
    - in proposed scheme the CB can be constructed in any desired domain
  - does not account for the so-called reverse water-filling effect
    - CE quantization accounts for the reverse water-filling effect

Outline

- Motivation
- Basics of Adaptive Quantization under High-Rate Assumptions
- Flexible Audio Coding Scheme
- Experimental Illustration
- Further Work
- Conclusion

Further Work

- Integration of pitch model (long term prediction)
- Integration of a model of perception
- Using vector lattice quantizers rather than scalar quantizers (Z-lattices)
- Addressing computational complexity issues
  - EVD has a comp. complex. of order $O(N^3)$ ($N = 80$)
    - Can we accelerate EVD computation?
    - Can we replace KLT by some fixed transform (e.g. MDCT)?

Outline

- Motivation
- Basics of Adaptive Quantization under High-Rate Assumptions
- Flexible Audio Coding Scheme
- Experimental Illustration
- Further Work
- Conclusion

Conclusion

- Rate
  - This scheme can run for any rate from the continuum of the rates
  - Computational complexity is independent on the rate
- Compared to CELP, the proposed scheme has other advantages
- Clarity, transparency and simplicity of the scheme
  - Source and perception models are well separated
  - No tweaking (at least at the current stage of development)

Thank you !!!