FlexCode project - flexible source coder and some theoretical results on model-based quantization under high-rate theory assumptions

Alexey Ozerov

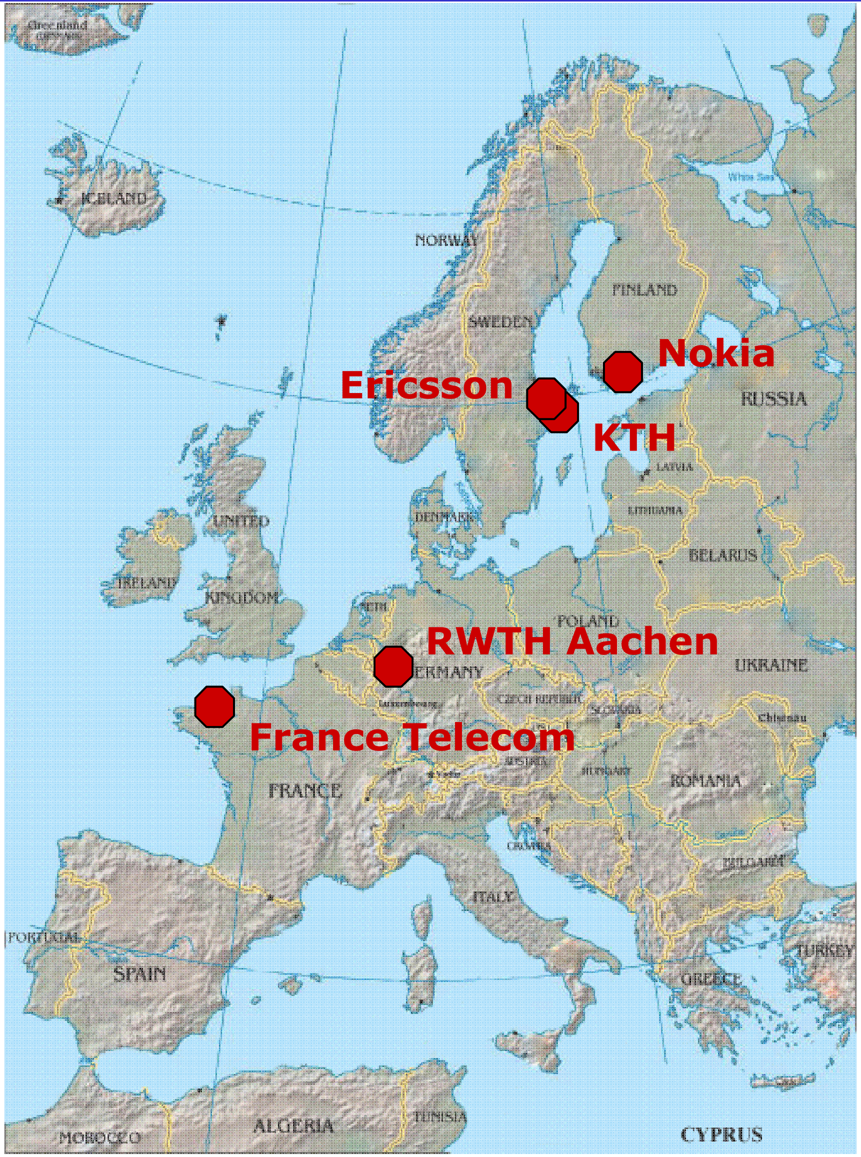
Institut TELECOM, TELECOM ParisTech, CNRS LTCI

Seminar at TELECOM ParisTech
December 16, 2008

Outline

- Flexcode in a Nutshell
- Introduction
- Basics of Adaptive Quantization under High-Rate Assumptions
- Flexible Coding Scheme
- Theoretical Results
  - Optimal Bit Allocation between Signal and Model
  - Optimal Bit Allocation between Signal and Model of Perception
  - Optimal Criteria for Signal Model Estimation
- Conclusion

Who?

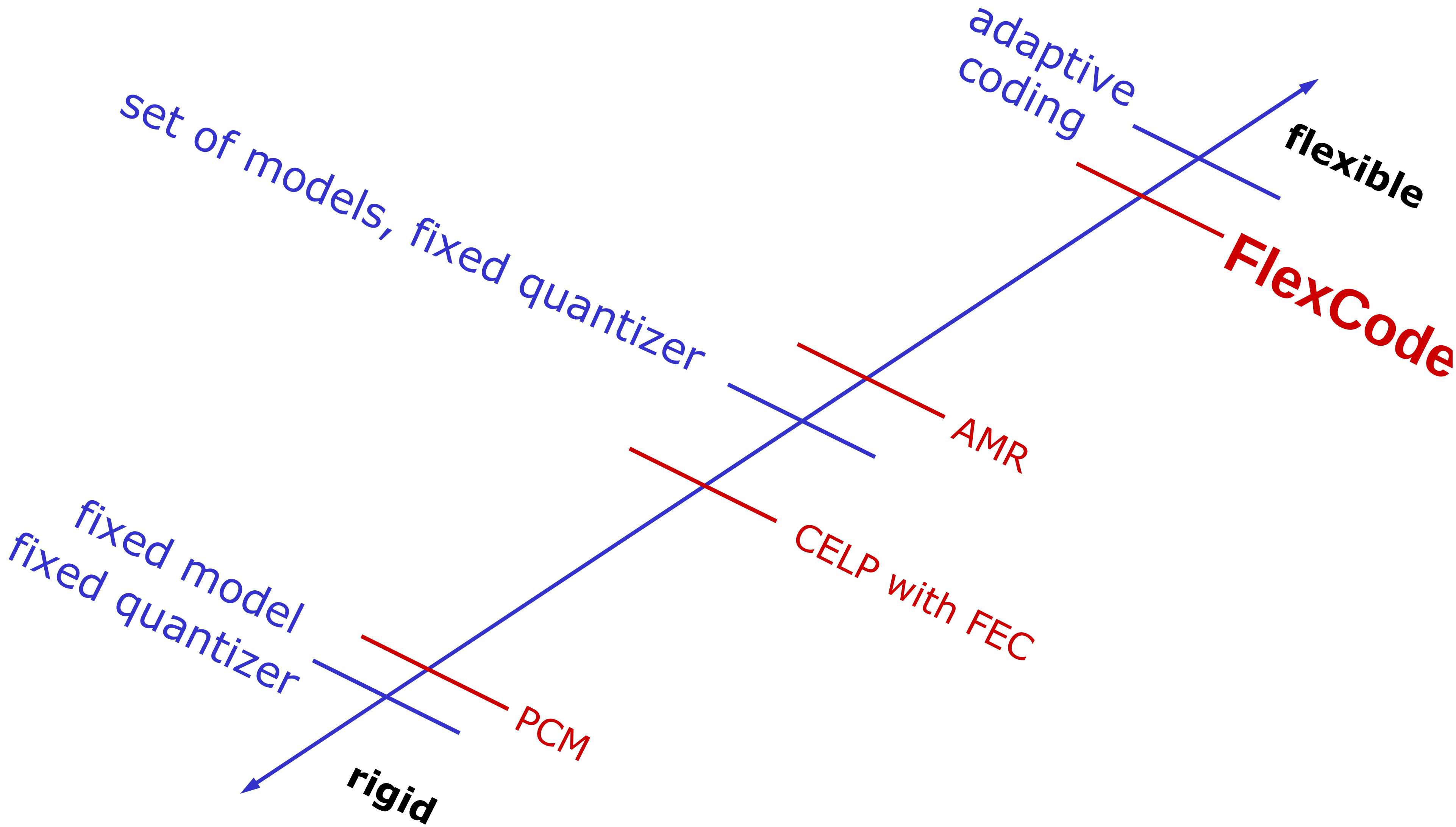

Nokia
Ericsson
KTH
RWTH Aachen
France Telecom
ICELAND
NORWAY
SWEDEN
FINLAND
RUSSIA
DENMARK
UNITED
KINGDOM
IRELAND
POLAND
BELARUS
UKRAINE
FRANCE
ROMANIA
ITALY
SPAIN
GREECE
TURKEY
ALGERIA
TUNISIA
MOROCCO
CYPRUS

The Problem

- Heterogeneity of networks increasing
- Networks inherently variable (mobile users)

- But:
  - Coders not designed for specific environment
  - Coders inflexible (codebooks and FEC)
  - Feedback channel underutilized

Adaptation and Coding



The Tools

- Tools include
  - Models of source, channel, receiver
  - High-rate quantization theory
  - Multiple description coding (MDC)
  - Iterative source-channel decoding
  - Distortion measures using the sensitivity matrix

Outline

- Flexcode in a Nutshell
- Introduction
- Basics of Adaptive Quantization under High-Rate Assumptions
- Flexible Coding Scheme
- Theoretical Results
  - Optimal Bit Allocation between Signal and Model
  - Optimal Bit Allocation between Signal and Model of Perception
  - Optimal Criteria for Signal Model Estimation
- Conclusion

How Can we Code an Audio Signal?

- Irrelevance
  - Parts of the signal that we cannot perceive
  - *FlexCode* : use advanced auditory model expressed in terms of sensitivity matrix
- Redundancy
  - Statistical dependencies that allows the information to be expressed with fewer bits
  - *FlexCode* : use a probabilistic model of the signal

Motivation

- Redundancy reduction
- Conventional codebook-based approaches
  - Train a codebook for a particular rate
  - Or train a set of codebooks for a set of the rates
- In *FlexCode* we want
  - Coder that is able to run for any rate from the continuum of the rates
  - Computational complexity to be independent on the rate
- Thus, we cannot train codebooks, we need to compute them on the fly
- Probabilistic source modeling together with high-rate theory approximation allows that

Outline

- Flexcode in a Nutshell
- Introduction
- Basics of Adaptive Quantization under High-Rate Assumptions
- Flexible Coding Scheme
- Theoretical Results
  - Optimal Bit Allocation between Signal and Model
  - Optimal Bit Allocation between Signal and Model of Perception
  - Optimal Criteria for Signal Model Estimation
- Conclusion

High-Rate Theory

$x^k \in \mathbb{R}^k$ source vector

$f_{X^k}(x^k)$ its pdf

Distortion in a quantization cell :

$$D_i = \frac{1}{k}\frac{\int_{\mathbf{V_i}} f_{X^k}(x^k)\|x^k - Q(x^k)\|^2 dx^k}{\int_{\mathbf{V_i}} f_{X^k}(x^k) dx^k}$$

$$\text{HR approx.} \approx \frac{1}{k}\frac{f_{X^k}(c_i^k)\int_{\mathbf{V_i}} \|x^k - c_i^k\|^2 dx^k}{f_{X^k}(c_i^k)\int_{\mathbf{V_i}} dx^k} = \frac{1}{kV_i}\int_{\mathbf{V_i}} \|x^k - c_i^k\|^2 dx^k$$

- optimal for high-rate

- still work well for low-rate

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

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

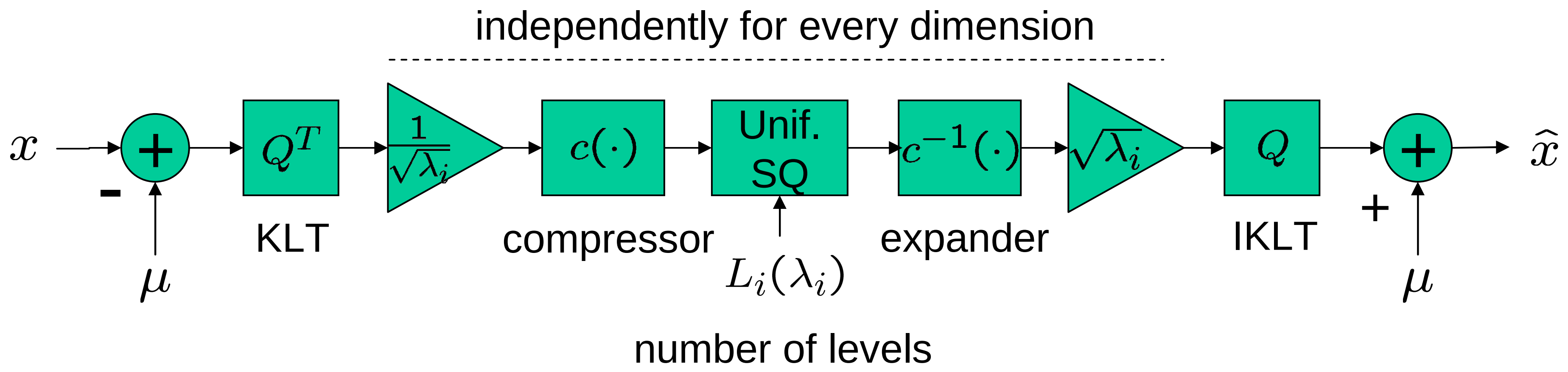
- CE performs better than CR, but needs a more flexible transmission channel

Single Gaussian case

- CR quantization (with companded scalar quantizers)

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

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

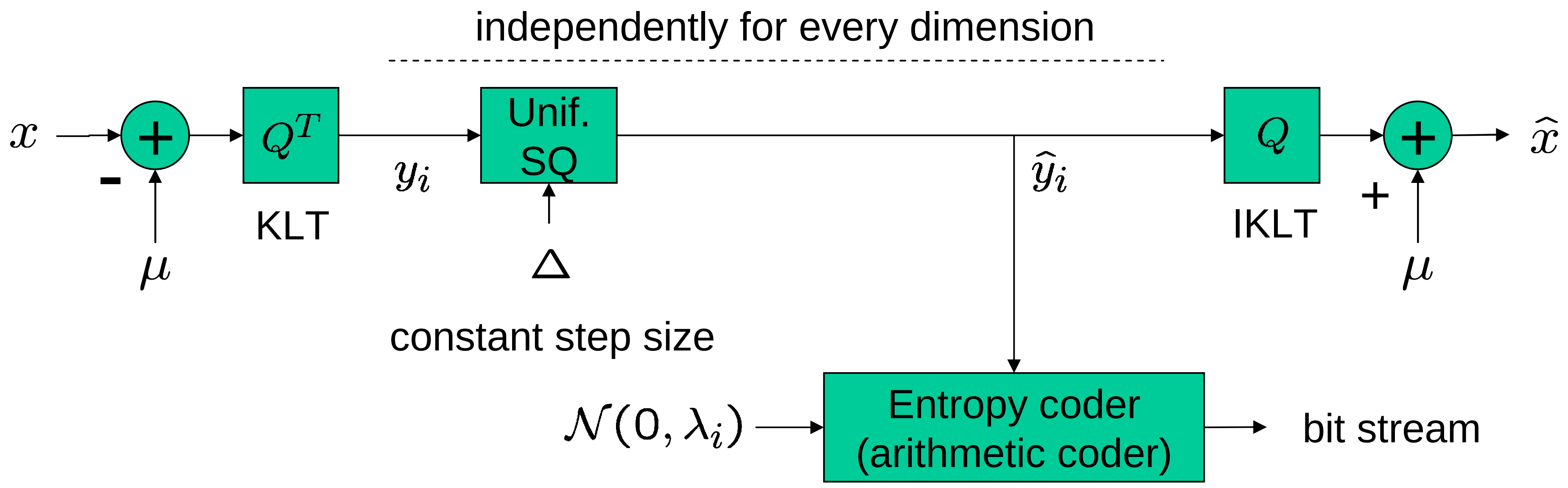


Single Gaussian case

- 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
- In the CE case 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

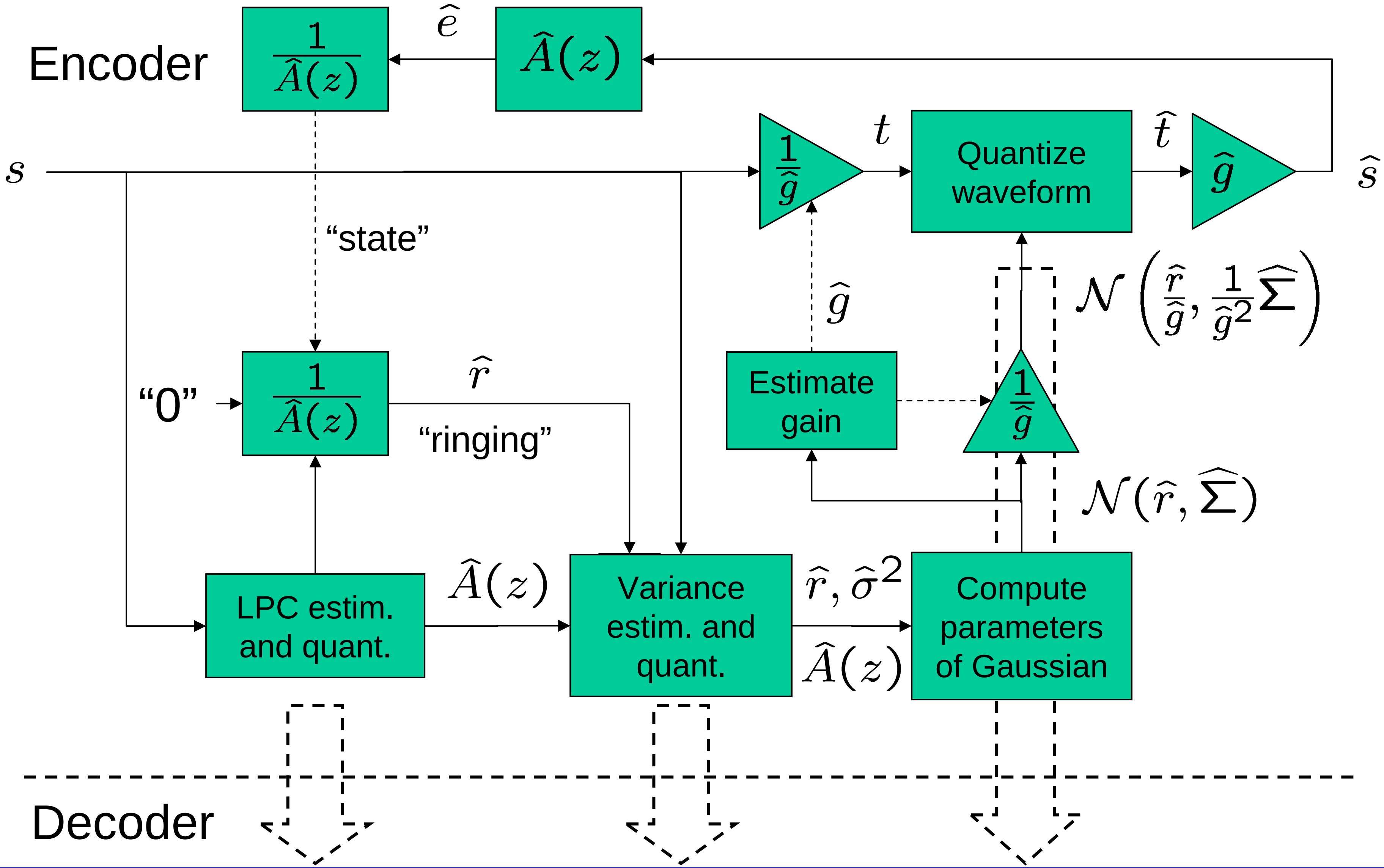
- With GMM the quantization consists in the following steps:
  - For each input vector *x*, choose the component (state) maximizing the *a posteriori* probability $p(i|x)$
  - Quantize using selected Gaussian component (in CR or CE case), as described before

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

Outline

- Flexcode in a Nutshell
- Introduction
- Basics of Adaptive Quantization under High-Rate Assumptions
- Flexible Coding Scheme
- Theoretical Results
  - Optimal Bit Allocation between Signal and Model
  - Optimal Bit Allocation between Signal and Model of Perception
  - Optimal Criteria for Signal Model Estimation
- Conclusion

Encoder structure



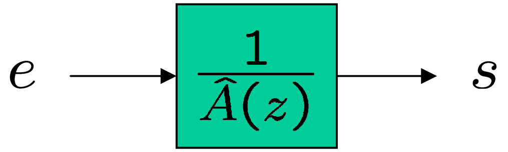
LPC estimation and quantization

- LPCs are estimated as in AMR-WB coder
  - Estimate LPC for every 5 ms frame

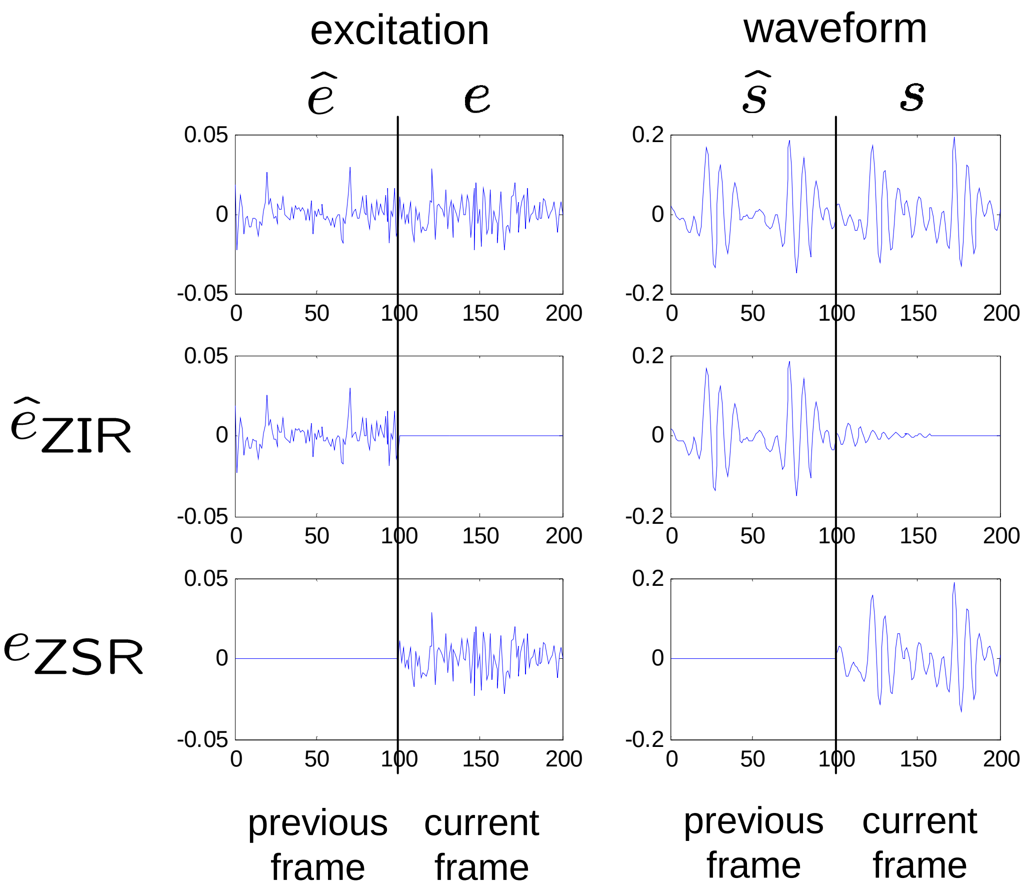
- LPCs are quantized in LSP domain using a GMM

- Quantized LPC are interpolated in LSF domain for every 1.25 ms subframe (as in AMR-WB coder)

“Ringing” (or ZIR) computation



$$s = \hat{s}_{\text{ZIR}} + s_{\text{ZSR}}$$



“Ringing” or

Zero Input Response (ZIR)

$$\hat{r} = \hat{s}_{\text{ZIR}}$$

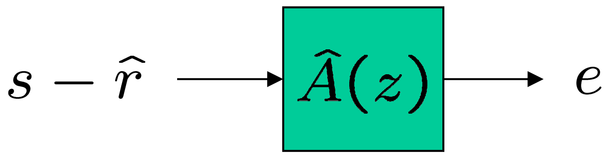
Zero State Response (ZSR)

$$s_{\text{ZSR}}$$

$$t = s_{\text{ZSR}} \rightarrow \hat{t}$$

Variance estimation and quantization

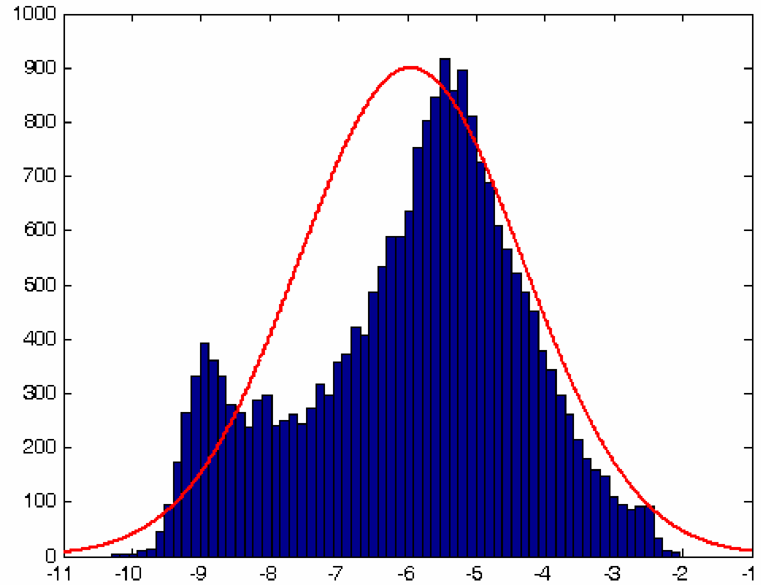
- Variance estimation
  - in ML sense (**???**)



compute variance

$$\sigma^2 = \mathrm{var}(e)$$

- Variance quantization
  - Modeled by a single Gaussian in log-domain
  - Quantized using this distribution



Compute parameters of Gaussian

"ringing" LPC and variance

$\hat{r}$ $$\frac{\hat{\sigma}_e}{\hat{A}(z)} = \frac{\hat{\sigma}_e}{1+\hat{a}_1 z^{-1}+\ldots+\hat{a}_p z^{-p}}$$

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

where $\widehat{\Sigma} = \hat{A}^{-1}\hat{A}^{-T}$

$\hat{A}$ is a lower triangular Toeplitz (k x k) matrix with as first column

$$\hat{\sigma}^{-1}[1, \hat{a}_1, \hat{a}_2, \ldots, \hat{a}_p, 0, \ldots, 0]^T$$

Comparison with a CELP scheme

- We compare with a CELP (analysis by synthesis scheme) with a CB trained minimizing MSE in speech 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
Gain rate	3	2.7	5
Speech rate	9	9.3	7
Av. SSNR	15.8	17.85	17.54

  - This is with scalar quantizers, and for quite rate

Listening tests

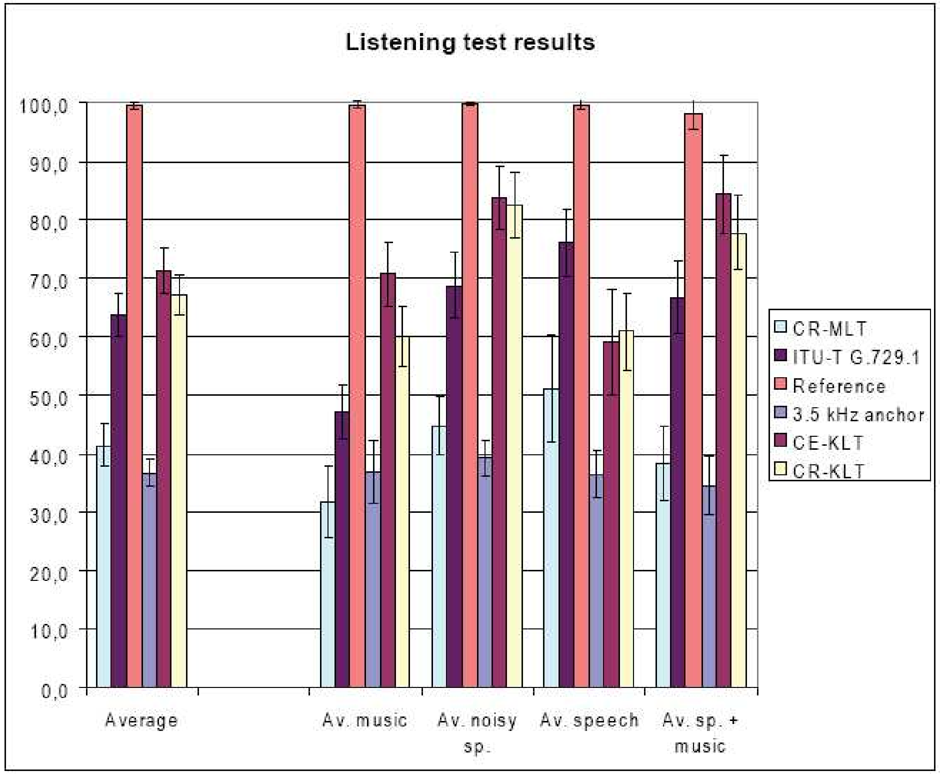


Figure 3.5: Averaged MUSHRA test results (respectively: all clips, 4 music clips, 2 noisy speech clips, 4 speech clips, 2 speech + music clips). The coders appear in the following order: *FlexCode* CR-MLT, ITU-T G.729.1, Reference, 3.5kHz Anchor, *FlexCode* CE-KLT, *FlexCode* CR-KLT. All the results are averaged for 12 listeners. All the coders operate at 24 kbps.

- Plus pitch model (long-term predictor)

$$\frac{\sigma_a}{A(z)} \Rightarrow \frac{\sigma_a}{A(z)} \frac{\sigma_p}{P(z)}$$

- Plus model of perception (derived from the source model)

Demo

External demo

CE vs. CR quantization

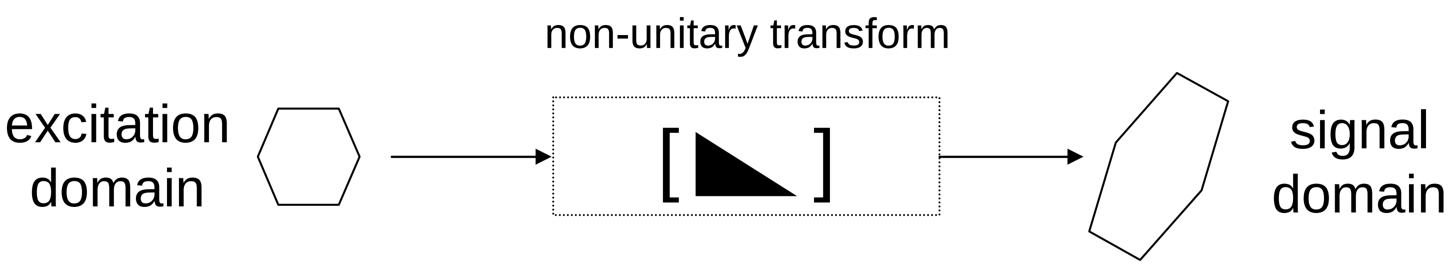
- Disadvantage
  - Need a more flexible channel, since rate varies

- Advantages
  - Constant distortion is very good for perception
  - Constant distortion allows to avoid many instability problems (coder design becomes much easier than in CR case)

FlexCode coder vs. CELP

- Disadvantage
  - Scalar quantizes in the KLT domain
- Advantages
  - Flexibility (can run for any rate)
  - Constant computational complexity (independent on the rate)
  - Varying rate (CE quantization) is allowed
  - Codebook is optimized in the right domain

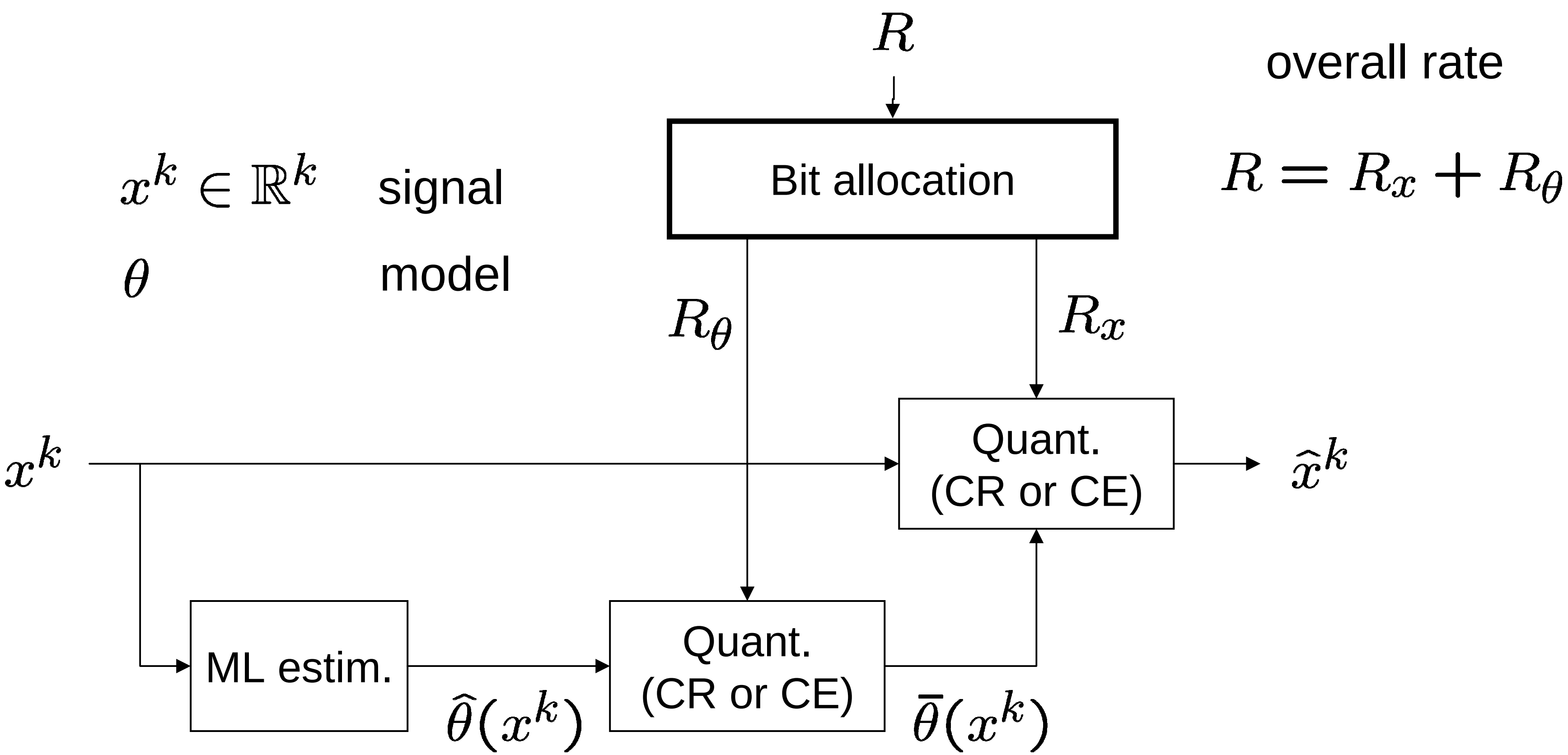
CELP :



Outline

- Flexcode in a Nutshell
- Introduction
- Basics of Adaptive Quantization under High-Rate Assumptions
- Flexible Coding Scheme
- Theoretical Results
  - Optimal Bit Allocation between Signal and Model
  - Optimal Bit Allocation between Signal and Model of Perception
  - Optimal Criteria for Signal Model Estimation
- Conclusion

Optimal Bit Allocation between Signal and Model



Problem: given the overall rate R, what is the optimal bit allocation between signal and model?

Optimal bit allocation between signal and model

- 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 independent on the overall rate**
  - Very useful in practice, since that means that model should be always quantized with a constant rate
  - This result is true for any model and in both CR and CE cases
- More-over, for AR model quantization it is shown that minimization of mean Square Log Spectral Distortion (SLSD) is near optimal (this result is obtained theoretically without using any perceptual knowledge)

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)*, pages 243-246, Mohonk, NY, Oct. 2007.

Optimal Bit Allocation between Signal and Model

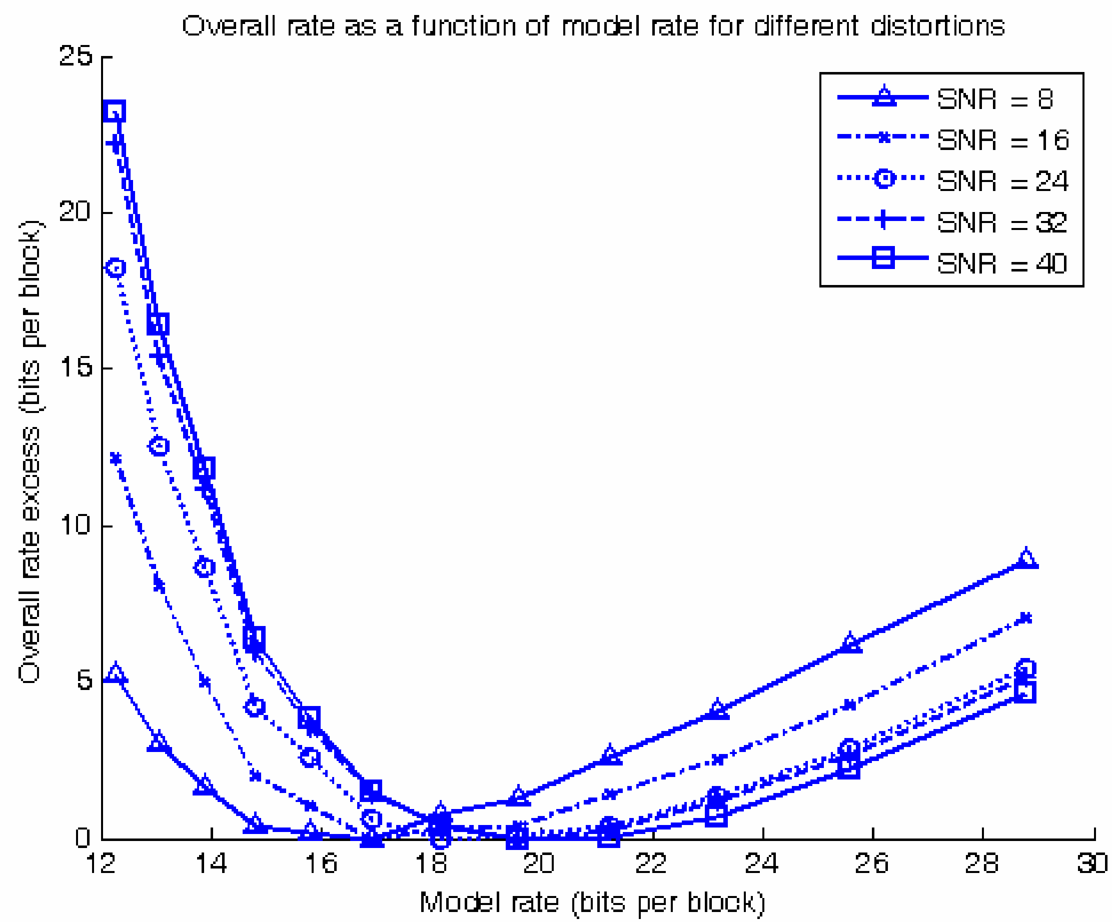
Confirmation

Table 1: Bit rates of the AMR-WB coder [7].

Rate	6.6	8.85	12.65	14.25	15.85	18.25	19.85	23.05
AR model parameters	36	46	46	46	46	46	46	46
pitch-model parameter	23	26	30	30	30	30	30	30
excitation	48	80	144	176	208	256	288	352

Optimal Bit Allocation between Signal and Model

Experimental verification



Theoretically predicted rate =

19.0 bits per block

Optimal Bit Allocation between Signal and Model

Number of sentences = 10,

Fs = 8000 Hz, Frame length = 2.5 ms (20 samples),

Total bits per frame

Bits for gain

	20 bits	40 bits	60 bits	80 bits	100 bits
2 bits	7.21	13.17	18.30	23.66	28.90
3 bits	8.97	15.45	21.02	26.85	32.76
4 bits	**9.37**	**16.06**	**21.63**	**27.38**	**33.46**
5 bits	9.19	15.97	21.53	27.31	33.38
6 bits	8.83	15.70	21.29	26.98	33.10
7 bits	8.41	15.43	21.00	26.71	32.80

Outline

- Flexcode in a Nutshell
- Introduction
- Basics of Adaptive Quantization under High-Rate Assumptions
- Flexible Coding Scheme
- Theoretical Results
  - Optimal Bit Allocation between Signal and Model
  - Optimal Bit Allocation between Signal and Model of Perception
  - Optimal Criteria for Signal Model Estimation
- Conclusion

Optimal Bit Allocation between Signal and Model of Perception

- What is the optimal bit allocation between signal and model of perception

- Answer (from HR theory): **0 bits for model of perception**, i.e., the model of perception should not be transmitted at all

R. Heusdens, W. B. Kleijn, and A. Ozerov, "Entropy-constrained high-resolution lattice vector quantization using a perceptually relevant distortion measure," In *IEEE Asilomar Conference on Signals, Systems, and Computers (Asilomar CSSC'07)*, Pacific Grove, CA, Nov. 2007.

Outline

- Flexcode in a Nutshell
- Introduction
- Basics of Adaptive Quantization under High-Rate Assumptions
- Flexible Coding Scheme
- Theoretical Results
  - Optimal Bit Allocation between Signal and Model
  - Optimal Bit Allocation between Signal and Model of Perception
  - Optimal Criteria for Signal Model Estimation
- Conclusion

Optimal Criteria for Signal Model Estimation

- Practical Rate-Distortion (RD) relations under HR theory assumptions (Gaussian model case)

$$R = -\frac{k}{2}\log_2 D + \psi(\mathrm{s}, \theta)$$

$\psi(\mathrm{s}, \theta)$ depends only on data s and model parameters θ

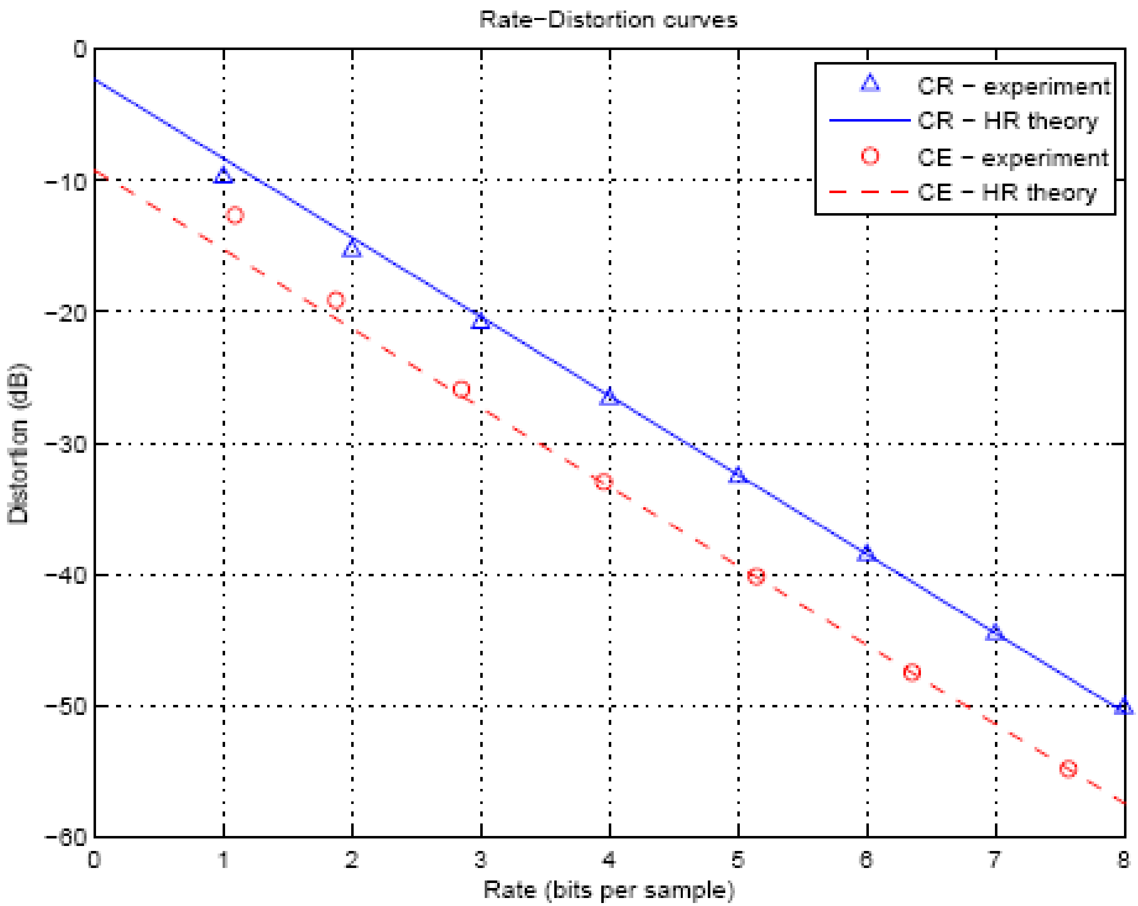
$$\psi_{\mathrm{CE}}(\mathrm{s}, \theta) = \frac{k}{2}\log_2 C - \frac{1}{N}\log_2 \prod_{n=1}^{N} p_{S_n}(s^n|\theta_n),$$

$$\psi_{\mathrm{CR}}(\mathrm{s}, \theta) = \frac{k}{2}\log_2\left(\frac{3(2\pi)^{2/3}C}{k}\right) + \frac{k}{2}\log_2 \frac{1}{N}\sum_{n=1}^{N}\left[\prod_{l=1}^{k}\lambda_{n,l}^{\frac{1}{k}}\sum_{i=1}^{k}\lambda_{n,i}^{-\frac{1}{3}}N(y_i^n; 0, \lambda_{n,i})^{-\frac{2}{3}}\right],$$

A. Ozerov, and W. B. Kleijn, "Optimal parameter estimation for model-based quantization," In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'09)*, 2009, accepted.

Optimal Criteria for Signal Model Estimation

- Practical confirmation



Optimal Criteria for Signal Model Estimation

- What is the best model for quantization of a particular data sequence?

$$R = -\frac{k}{2}\log_2 D + \psi(\mathrm{s},\theta)$$

$$\theta^* = \arg\min_{\theta} \psi(\mathrm{s},\theta)$$

$$\psi_{\mathrm{CE}}(\mathrm{s},\theta) = \frac{k}{2}\log_2 C - \frac{1}{N}\log_2 \prod_{n=1}^{N} p_{S_n}(s^n|\theta_n),$$

equivalent to maximum likelihood (ML) criterion (minimum description length (MDL) principle)

$$\psi_{\mathrm{CR}}(\mathrm{s},\theta) = \frac{k}{2}\log_2\left(\frac{3(2\pi)^{2/3}C}{k}\right) + \frac{k}{2}\log_2 \frac{1}{N}\sum_{n=1}^{N}\left[\prod_{l=1}^{k}\lambda_{n,l}^{\frac{1}{k}}\sum_{i=1}^{k}\lambda_{n,i}^{-\frac{1}{3}}N(y_i^n;0,\lambda_{n,i})^{-\frac{2}{3}}\right],$$

is not equivalent to maximum likelihood (ML) criterion !!!

Usually ML criterion is always used

Optimal Criteria for Signal Model Estimation

- New criterion for model estimation in the CR case

$$\boxed{\theta_{\mathsf{CR\_MDL}} = \arg\min_{\theta} \phi(\mathrm{s}, \theta)}$$

$$\phi(\mathrm{s},\theta) = \log \sum_{n=1}^{N} \left[\prod_{l=1}^{k} \lambda_{n,l}^{\frac{1}{k}} \sum_{i=1}^{k} \exp\left\{ \frac{1}{3} \frac{(y_i^n)^2}{\lambda_{n,i}} \right\} \right],$$

$$\phi(\mathrm{s},\theta) = A\psi_{\mathsf{CR}}(\mathrm{s},\theta) + B$$

Optimal Criteria for Signal Model Estimation

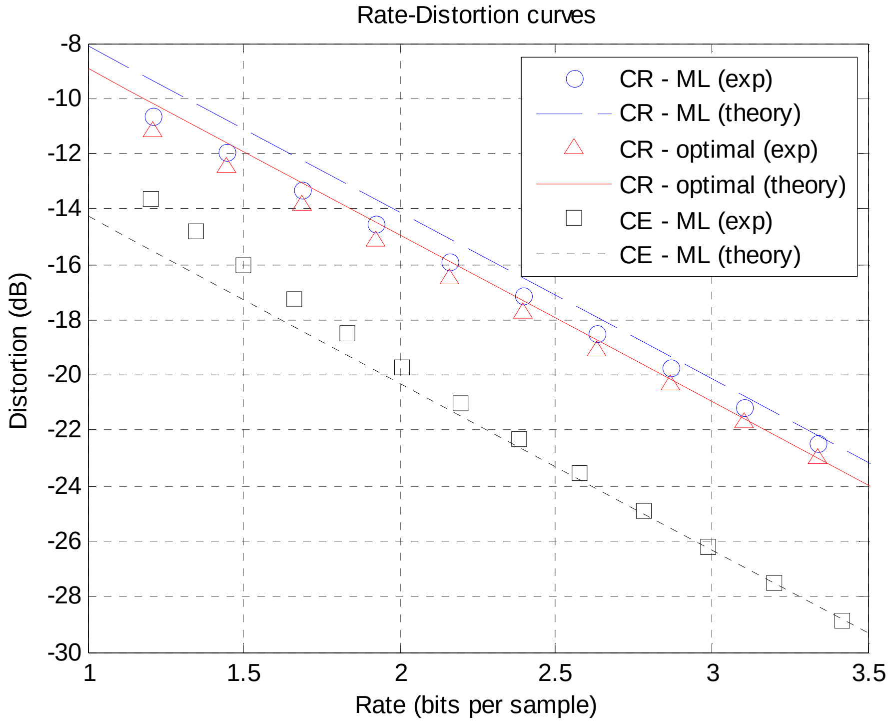
- Practical optimization issues
  - Proposed CR-MDL is difficult to solve in a closed form
  - We use Newton's method

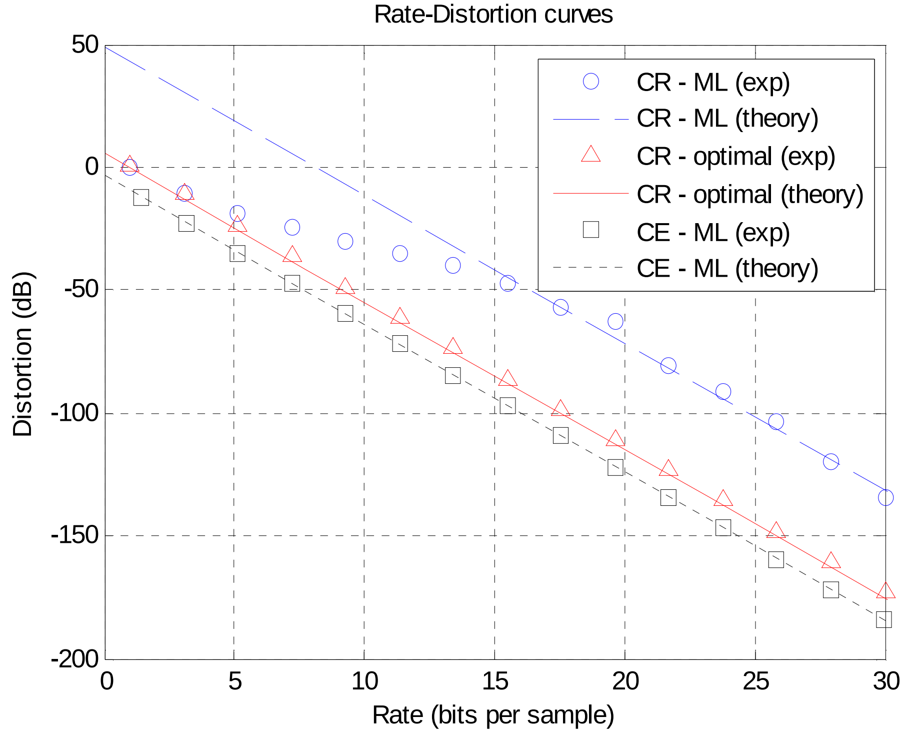
$$\theta^{m+1} = \theta^m - \gamma \left[H_\theta \phi(\mathrm{s}, \theta^m) \right]^{-1} \nabla_\theta \phi(\mathrm{s}, \theta^m),$$

  - We apply proposed CR-MDL criterion for gain (variance) estimation

Optimal Criteria for Signal Model Estimation

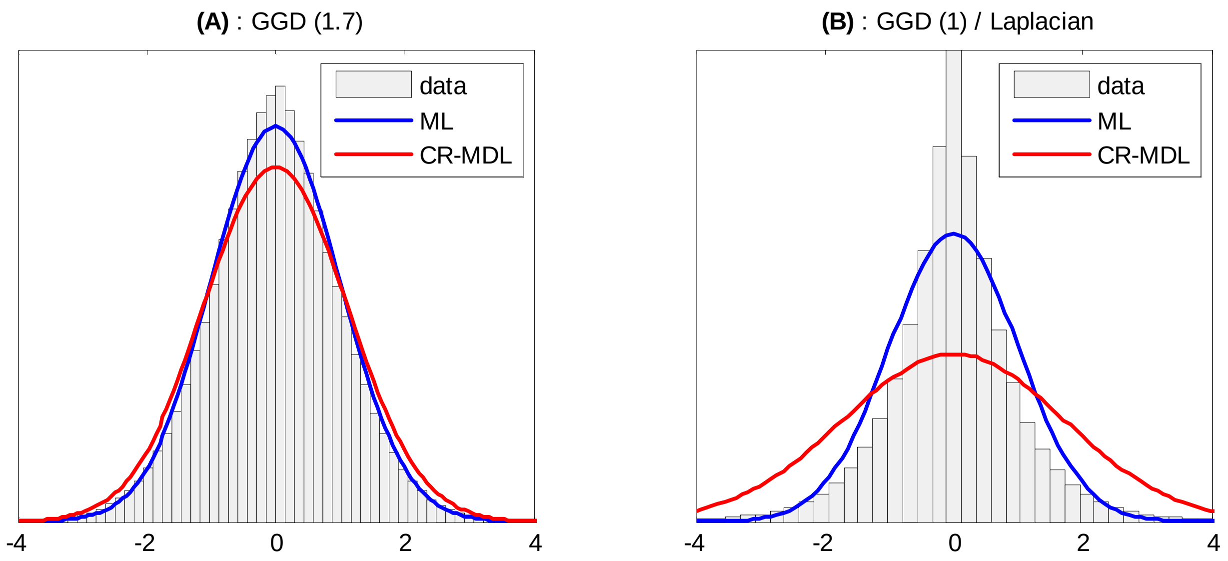
- Experiments





AR model-based scheme with KLT (data distribution is rather Gaussian)

Fixed Frequency Weighted MLT-based scheme (data distribution is rather Laplacian)



- Conclusion
  - There is an **advantage** in using new (CR-MDL) criterion, as compared to ML criterion, when there is a **mismatch between data and model distributions**, which is often the case for practical applications

Outline

- Flexcode in a Nutshell
- Introduction
- Basics of Adaptive Quantization under High-Rate Assumptions
- Flexible Coding Scheme
- Theoretical Results
  - Optimal Bit Allocation between Signal and Model
  - Optimal Bit Allocation between Signal and Model of Perception
  - Optimal Criteria for Signal Model Estimation
- Conclusion

Conclusion

- Rate
  - This scheme can run for any rate from the continuum of the rates
  - Computational complexity is independent on the rate
- CE quantization has a lot of advantages compared to CR quantization
- Clarity and simplicity of the scheme
  - Source and perception models are well separated
  - No tweaking (at least as I know)
- New theoretical results on model-based quantization

References

- 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 (Asilomar CSSC'07)*, Pacific Grove, CA, Nov. 2007.

- 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)*, pages 243-246, Mohonk, NY, Oct. 2007.

- R. Heusdens, W. B. Kleijn, and A. Ozerov, "Entropy-constrained high-resolution lattice vector quantization using a perceptually relevant distortion measure," In *IEEE Asilomar Conference on Signals, Systems, and Computers (Asilomar CSSC'07)*, Pacific Grove, CA, Nov. 2007.

- A. Ozerov, and W. B. Kleijn, "Optimal parameter estimation for model-based quantization," In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'09)*, 2009, accepted.