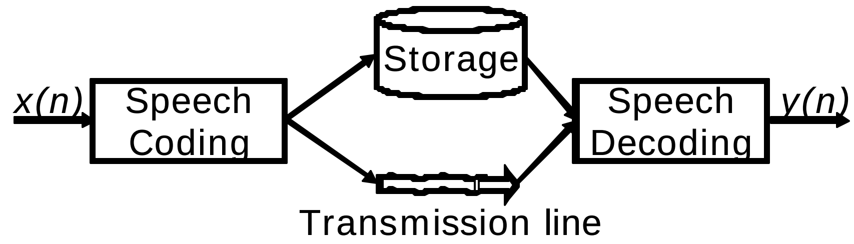
Part III
Speech Coding

Coding



$$e(n) = x(n) - y(n) \qquad SNR(dB) = 10\log(\sigma_x^2 / \sigma_e^2)$$

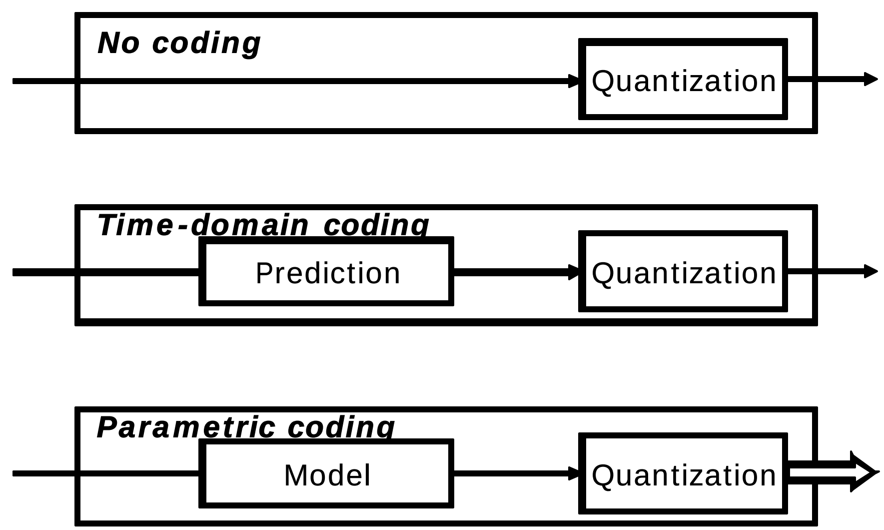
$$SNRSEG(dB) = E[SNR(dB)]$$

- Minimize storage/transmission bit-rate while keeping high SNRSEG
- NB : in practice, subjective tests are also used (MOS: mean opinion scores)

Lower bound?

- Speech = content, speaker, para-linguistic information
- **Hyp**: Speech information ≈ sequence of phonemes
  Bit rate?
  - 10 phonemes/sec
  - 30-50 phonemes for a language ≈ 32 = 2^5
  - Min. bit rate=5*10= **50 bps**!
  - To be compared with the **33.6 Kbps** allowed on an analog telephone line (and **64 Kbps** on a digital line)!

Coding, Quantizing



Contents

- **No coding**
  - Uniform quantization
  - Non-uniform quantization
  - Un. quantization with compression-expansion
  - Adaptive quantization
- Time-domain speech coding
  - DPCM
  - ADPCM
- Parametric coding
  - LPC (scalar quantization/vetor quatization)
  - Hybrid H/N coding (MBE)
  - MP-LPC, CELP

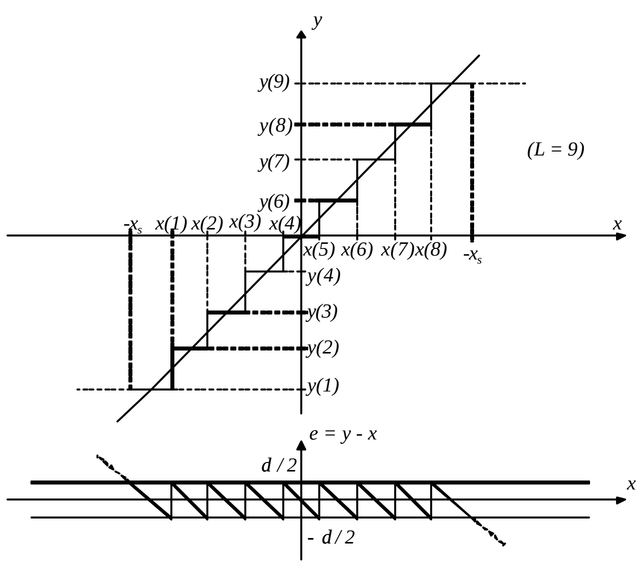
Uniform quantization

- b = number of bits
  x_s = saturation level
  G = load factor = x_s/ s_x (ex: Gaussian G=3)
- If x is guassian and the quantization step is small enough, than in the non-saturated range of x:

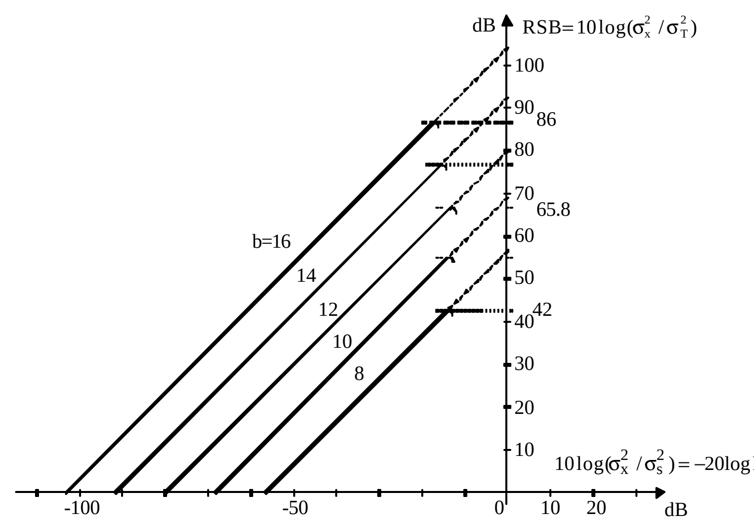
$$RSB = \mathbf{10.log(3}s_x^2 / x_s^2 \mathbf{.2}^{2b}\mathbf{)} \approx \mathbf{6,02}b + \mathbf{4,77} - \mathbf{20.log}\,\Gamma$$

- For a given G, **every bit counts for 6 DBs**

Uniform quantization

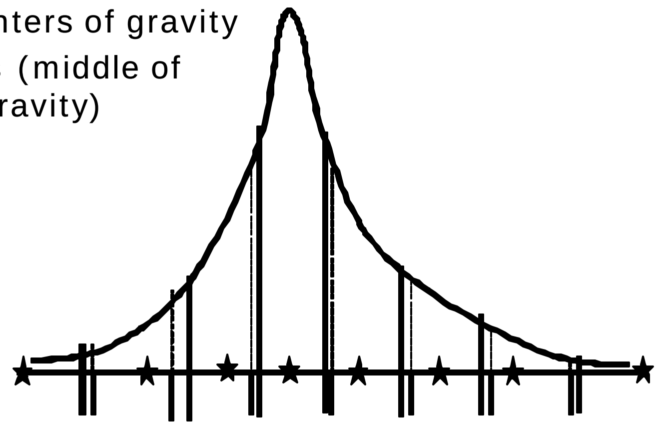


Uniform quantization



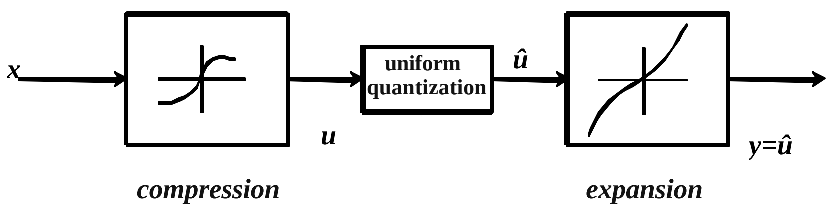
Non-uniform quantization

- Bit assignment should follow the p.d.f. of *x*
- Lloyd-Max optimization:
  - start with uniform quantization
  - compute centers of gravity
  - new borders (middle of centers of gravity)



Uniform quantization with compression-expansion

- Equivalent to non-uniform quantization

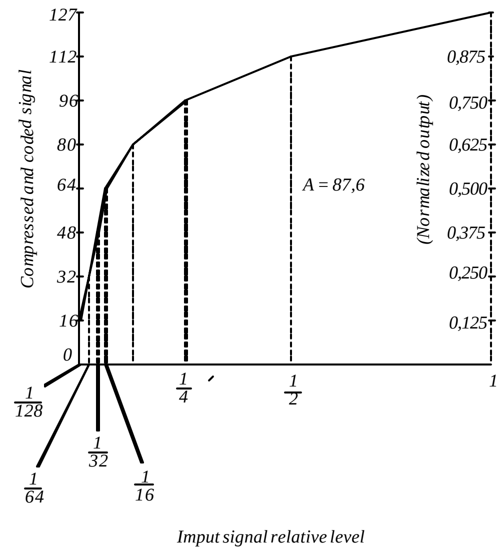


- Europe : A law

$$F(x) = \begin{cases} \dfrac{A|x|}{1+\ln A}.sgn(x) & 0 \le |x|/x_{\max} \le 1/A \\ x_{\max}.\dfrac{1+\ln\left[A.|x|/x_{\max}\right]}{1+\ln A}.sgn(x) & 1/A \le |x|/x_{\max} \le 1 \end{cases}$$

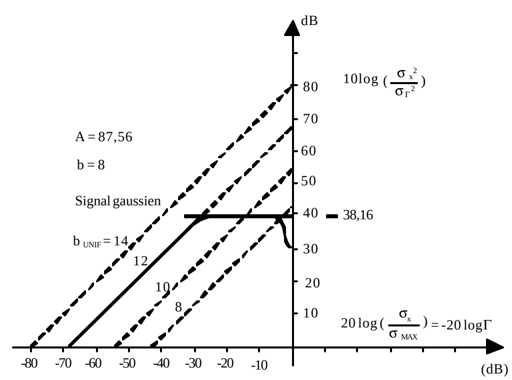
A Law

- $A = 87.56$
  $b = 8$



A Law

- x>> $RSB = 6{,}02.b + 4{,}77 - 20\log(1+\ln A) = 38.15\,dB$
- x<< $RSB = 6{,}02.b + 4{,}77 + 20.\log\left(\dfrac{A}{1+\ln A}\right) - 20.\log\Gamma$

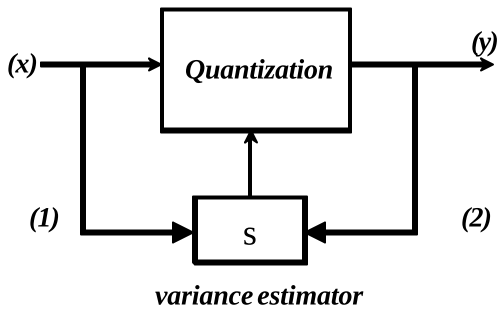


20 log(16)≈24 = 4 bits!

G.711 norm at 64 kbps (8bits at 8kHz)

Adaptive quantization

- Adapt the quantization step to the local variance of the signal (AQF-AQB)



(1) Adaptive quantization feedforward AQF

(2) Adaptive quantization backward AQB

Contents

- No coding
  - Uniform quantization
  - Non-uniform quantization
  - Un. quantization with compression-expansion
  - Adaptive quantization
- **Time-domain speech coding**
  - **DPCM**
  - ADPCM
- Parametric coding
  - LPC (scalar quantization/vetor quatization)
  - Hybrid H/N coding (MBE)
  - MP-LPC, CELP

DPCM: differential pulse modulation coding

- Speech is *redundant* : each sample can be somehow prodicted from previous samples

$$d(n) = x(n) - x(n-1)$$

$$\sigma_d^2 = 2\sigma_x^2 - 2\phi_{xx}(1) = 2\sigma_x^2.[1 - \rho_x(1)]$$

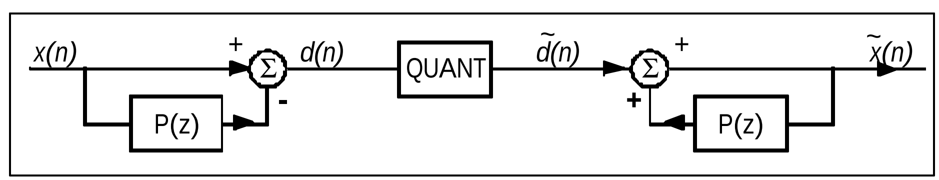
- If the autocorrelation is >O.5, than the variance of *d* is smaller than that of *x*, and:

$$\Delta RSB = 20.\log\left(\frac{\sigma_x}{\sigma_d}\right) = 10.\log\frac{1}{2.[1 - \rho_x(1)]}$$

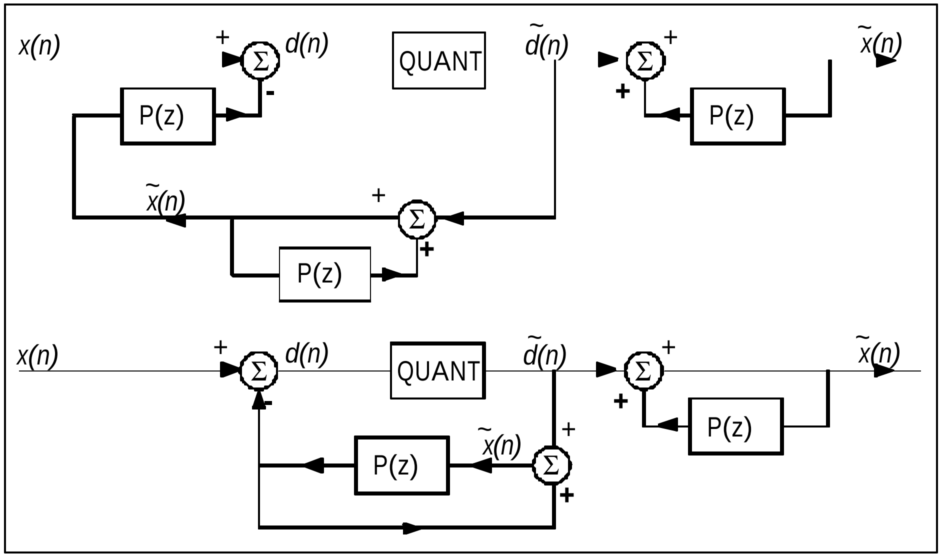
DPCM: more generally

$$d(n) = x(n) - \sum_{i=1}^{p} a_i x(n-i)$$

$$P(z) = \sum_{i=1}^{p} a_i z^{-i}$$

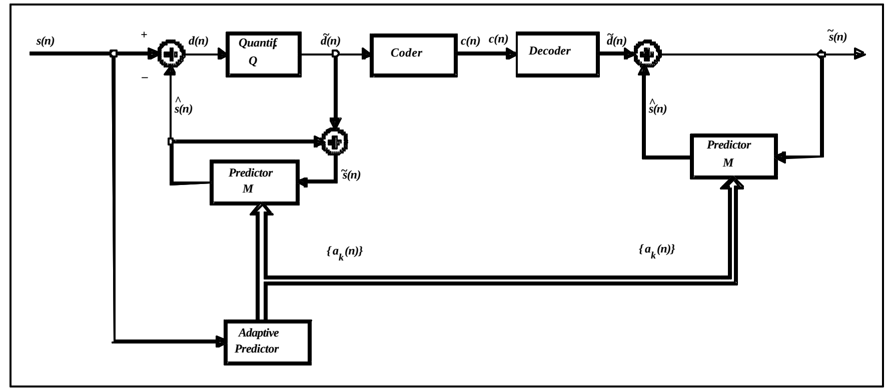


Avoiding the error drift

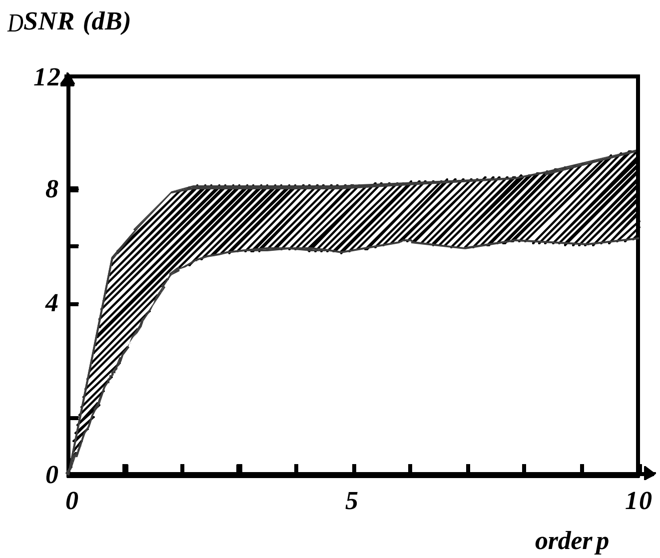


Adaptive DPCM: ADPCM

- DPCM with adaptive prediction (ex: APF)

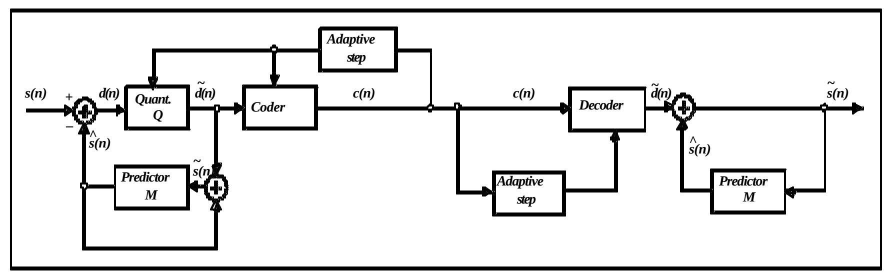


Prediction gain



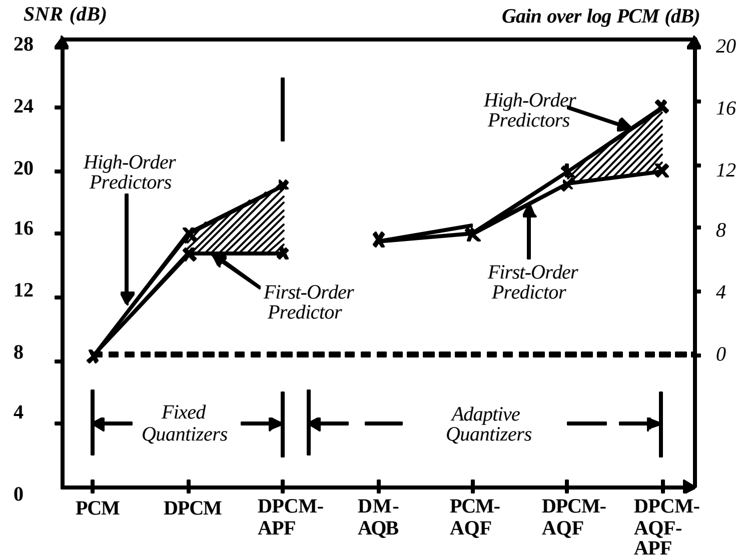
Adaptive DPCM: ADPCM

- DPCM with adaptive quantization (ex: AQB)

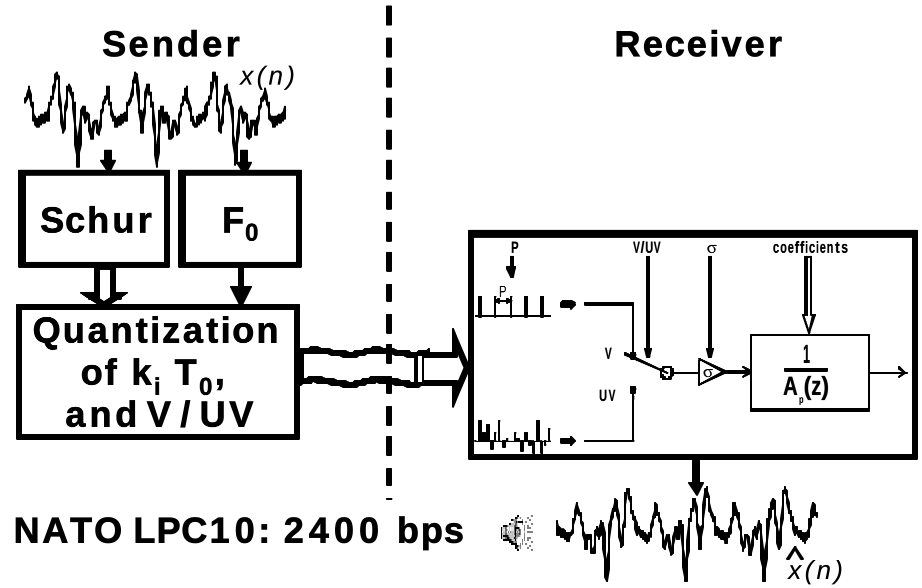


- **CCITT G721-723-726-727 (32 kb/s) are AQB-APB**
- **ex: 16 kbps**

A comparison at 24 kbps



Linear Prediction Coding



NATO LPC10: 2400 bps

Contents

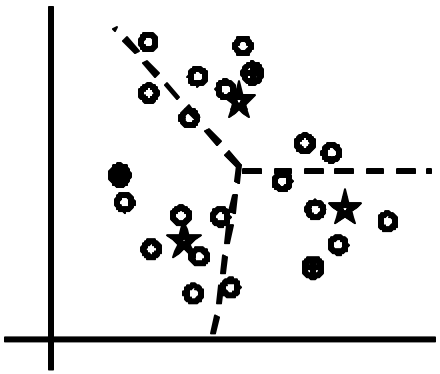
- No coding
  - Uniform quantization
  - Non-uniform quantization
  - Un. quantization with compression-expansion
  - Adaptive quantization
- Time-domain speech coding
  - DPCM
  - ADPCM
- **Parametric coding**
  - LPC (scalar quantization/vetor quatization)
  - Hybrid H/N coding (MBE)
  - MP-LPC, CELP

LPC + Vector Quantization

- Idea: pre-establish a (finite) list of $\{k_i\}$ so as to transmit the index of a $\{k_i\}$ in this list, instead of the $\{k_i\}$ itself
- More generally: separate the continuum of $\mathbf{x} = [x_1, x_1, .., x_p]^T$ into a finite set of classes, represented by a *centroid* $\mathbf{C_i}$
- For $\mathbf{x} = [x_1, x_1, .., x_p]^T$ send to index of the closest centroid
- *Distance?* Euclidian, Itakura, etc.
- *Centroids?* Generalized Lloyd-Max

LPC + Vector Quantization

- Genralized Lloyd-Max optimization:
  - start with random choice for centroids
  - assign each vector to its closest centroid
  - new centroids = centers of gravity of classes



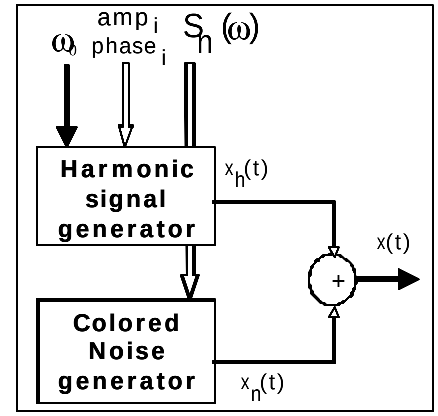
Typ: 800 bps
Used in spacial communications

Contents

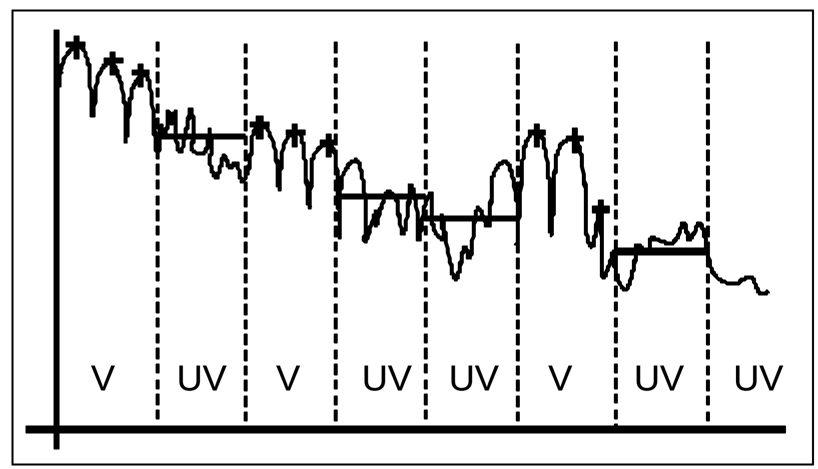
- No coding
  - Uniform quantization
  - Non-uniform quantization
  - Un. quantization with compression-expansion
  - Adaptive quantization
- Time-domain speech coding
  - DPCM
  - ADPCM
- Parametric coding
  - LPC (scalar quantization/vetor quatization)
  - **Hybrid H/N coding (MBE)**
  - MP-LPC, CELP

Multi-Band Excited (MBE)coder

- Use a hybrid H/N model, and send the H and N components with the least possible number of bits
- Send amplitude of harmonics? No: send spectral envelope (LPC)
- Phases? No assume min. phase
- Noise? Send noise levels / freq band

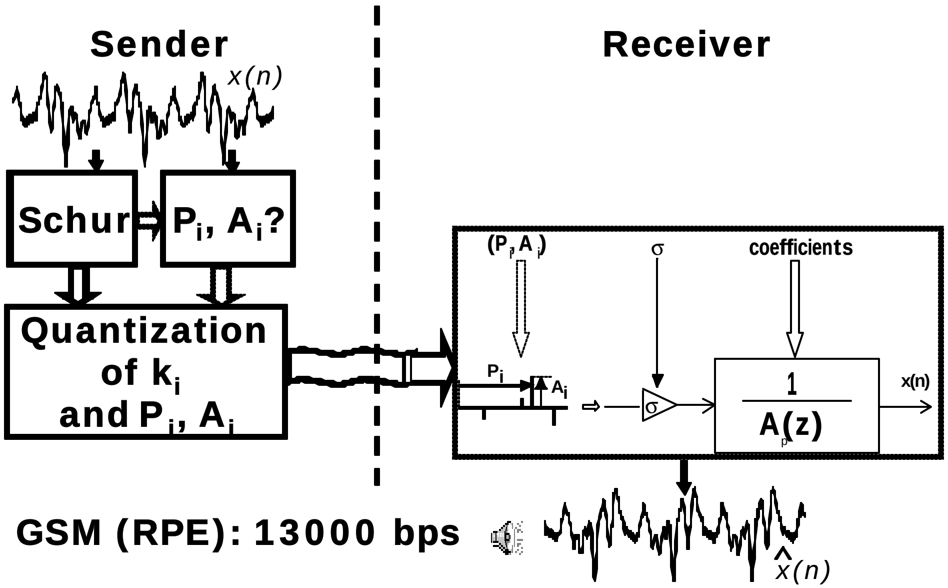


Multi-Band Excited (MBE)coder



Used in marine communications (4.8 kbps)

MP-LPC coder



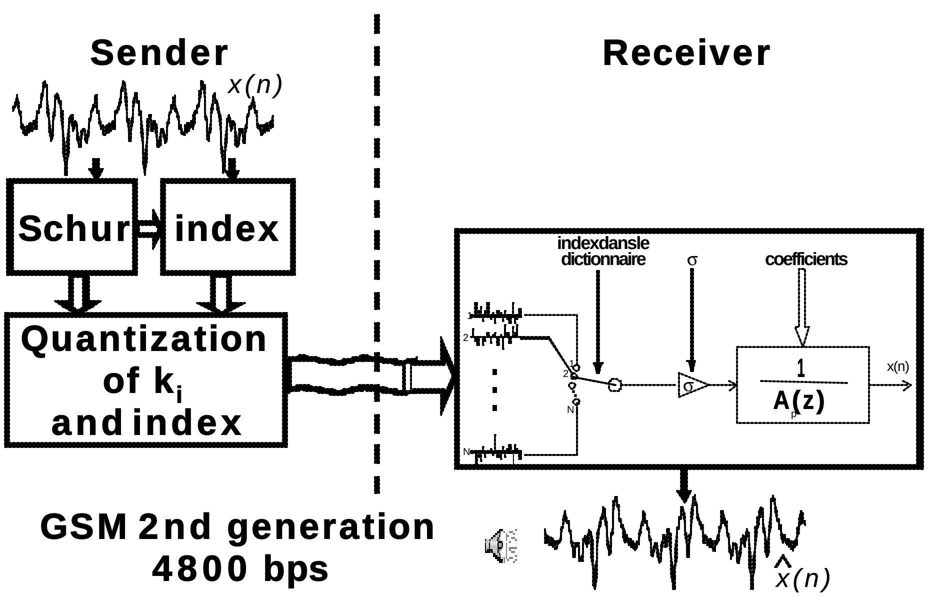
Issues in CELP coding

- Optimize the excitation dictionary?
- Computational load >>
  ⇒ use specially designed (tree-based) dictionaries
- For very low bit rate, use a predictor on the excitation itself:

$$\hat{e}(n) = \hat{e}_c(n) + b.\hat{e}(n-P)$$

- Latest versions reach the ADPCM 32 kbps quality with 4.8 kbps only.

CELP coder



Conclusion