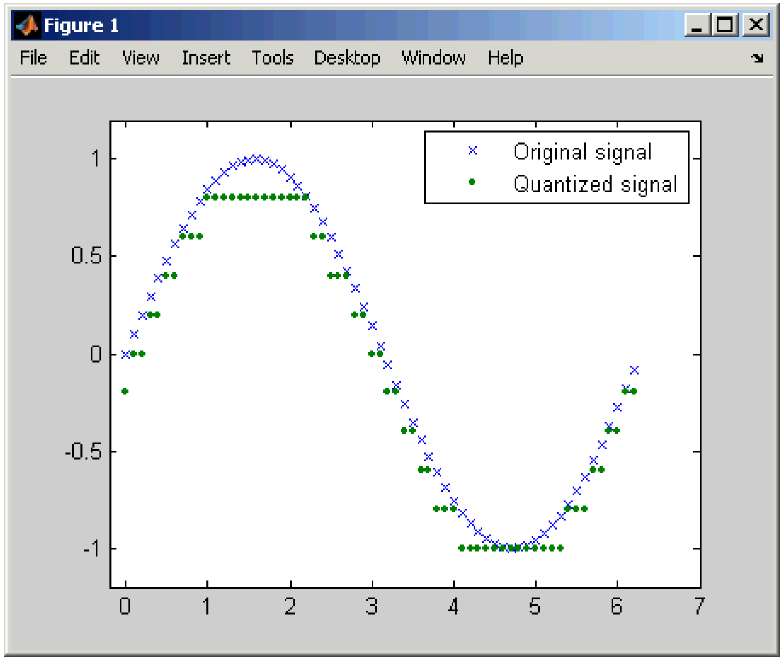
The Secrets of Quantization

Nimrod Peleg
Update: Sept. 2009

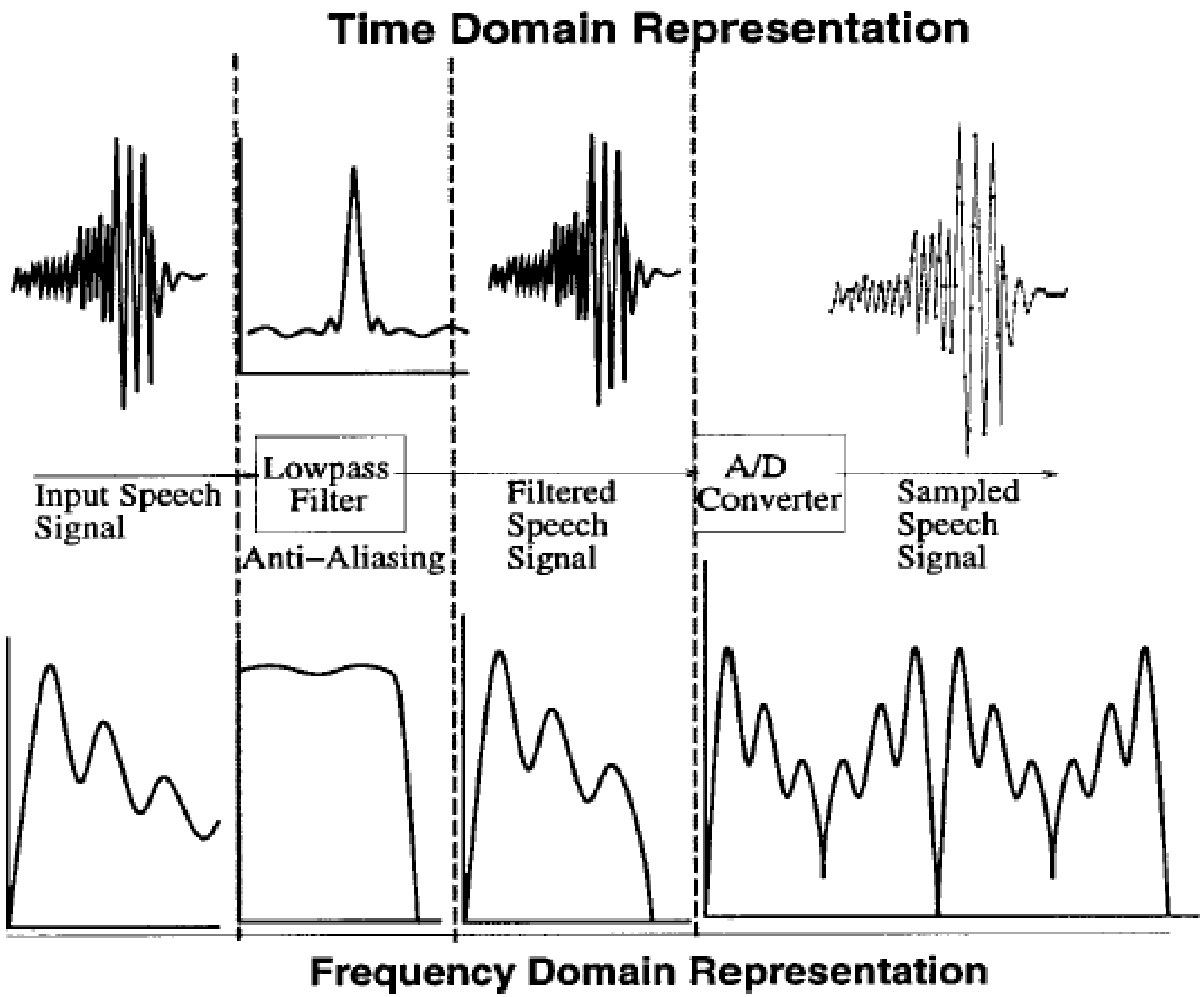


What is Quantization

- Representation of a large set of elements with a much smaller set is called quantization.
- The number of elements in the original set in many practical situations is infinite (like the set of real numbers.)
- In speech coding, prior to storage or transmission of a given parameter, it must be quantized in order to reduce storage space or transmission bandwidth for a cost-effective solution.
- In the process, some quality loss is introduced, which is undesirable.

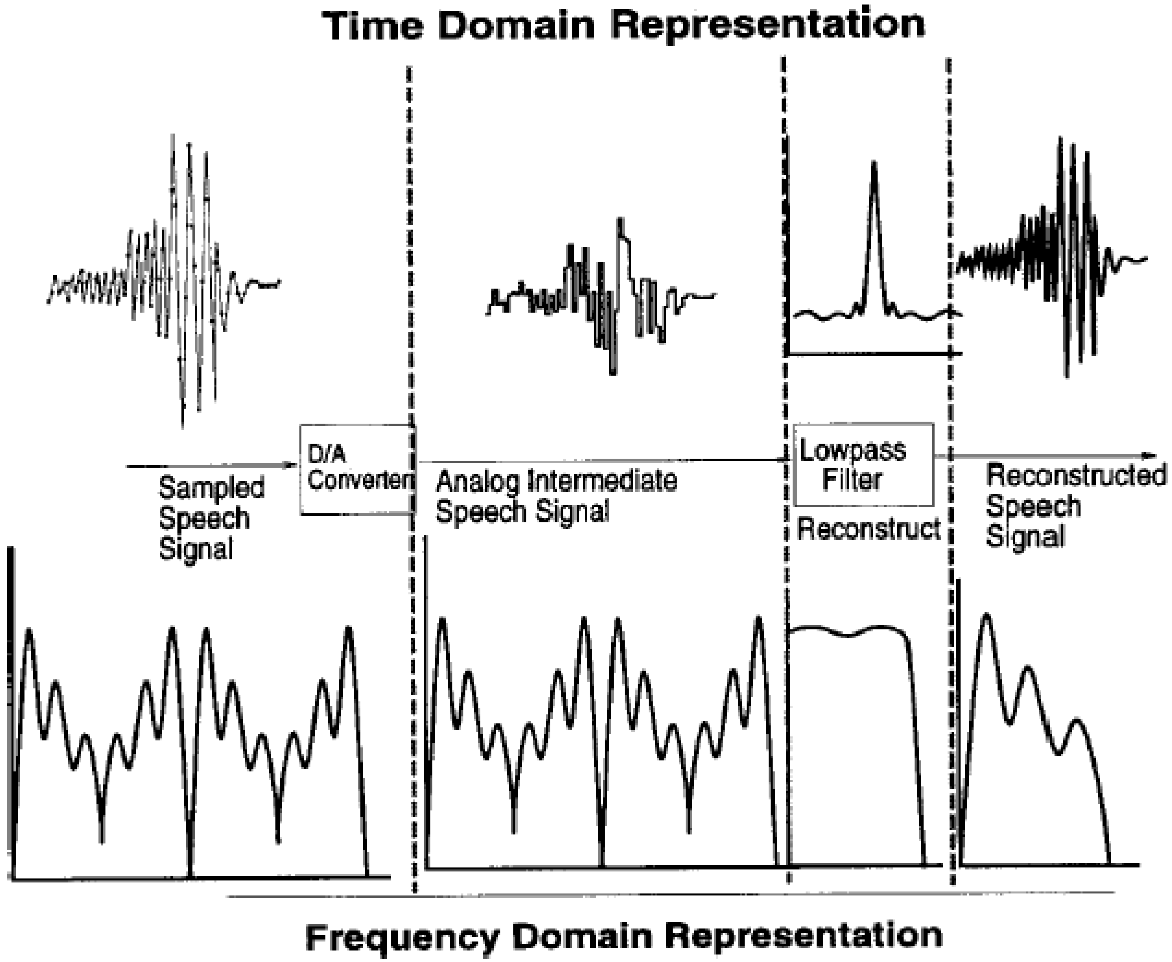
- How to minimize loss for a given amount of available resources is the **central problem of quantization.**

First…Sampling !



Time and frequency domains representations of signals at different stages during pulse code modulation (PCM) analysis.

And …Reconstruction



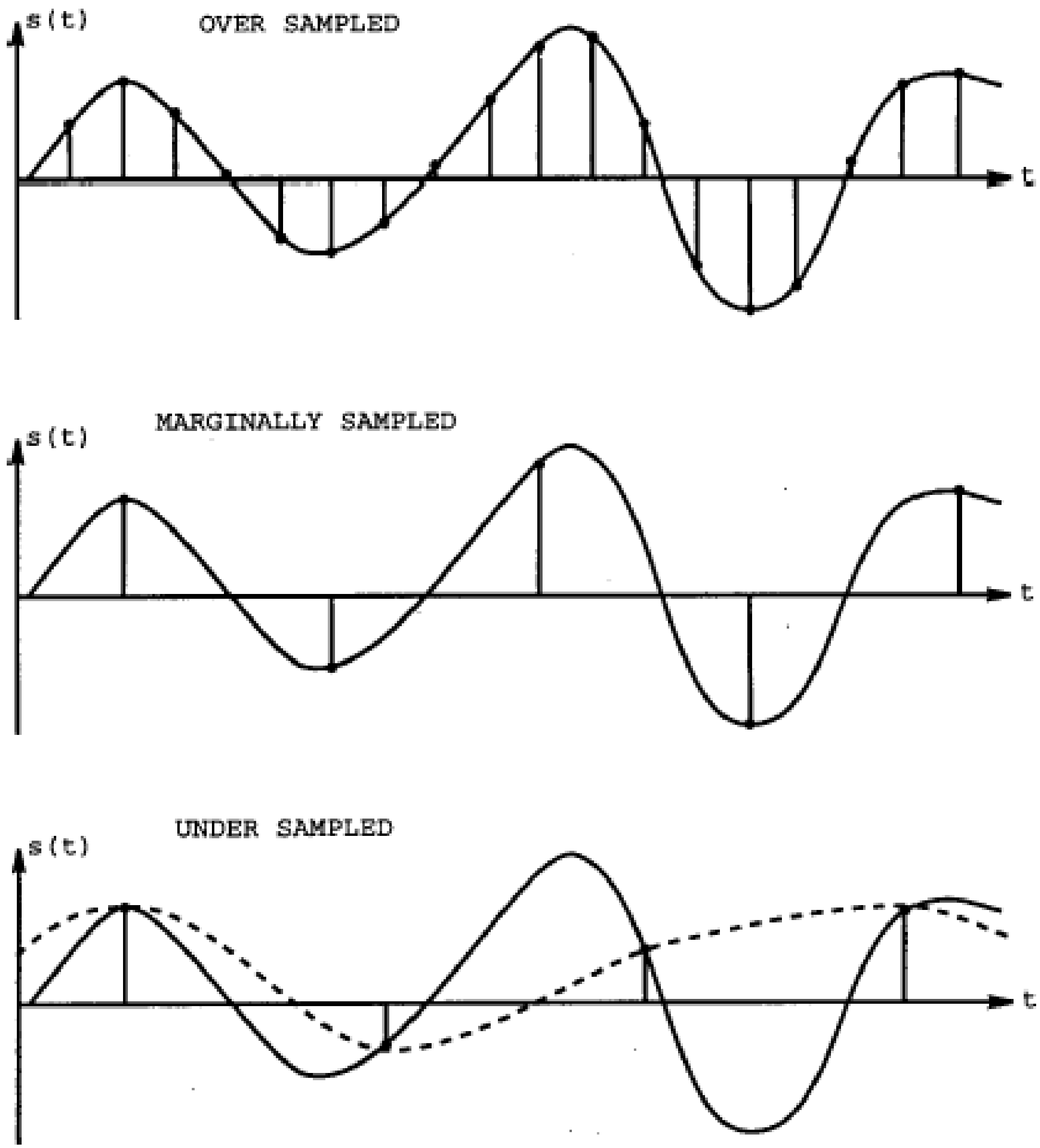
From: A Practical Handbook of Speech Coders, Goldberg, R. G.

Sampling Rate: Shannon-Nyquist rule

$$T_s \leq 1/2f_{\max}$$

$$= T_{\min}/2$$

$$f_s = 2 * f_{\max}$$



Historical background

- The **sampling theorem** was implied by the work of Harry Nyquist in 1928 ("Certain topics in telegraph transmission theory"), in which he showed that up to 2B independent pulse samples could be sent through a system of bandwidth B; but he did not explicitly consider the problem of sampling and reconstruction of continuous signals.
- The sampling theorem, essentially a dual of Nyquist's result, was proved by Claude E. Shannon in 1949 ("Communication in the presence of noise").

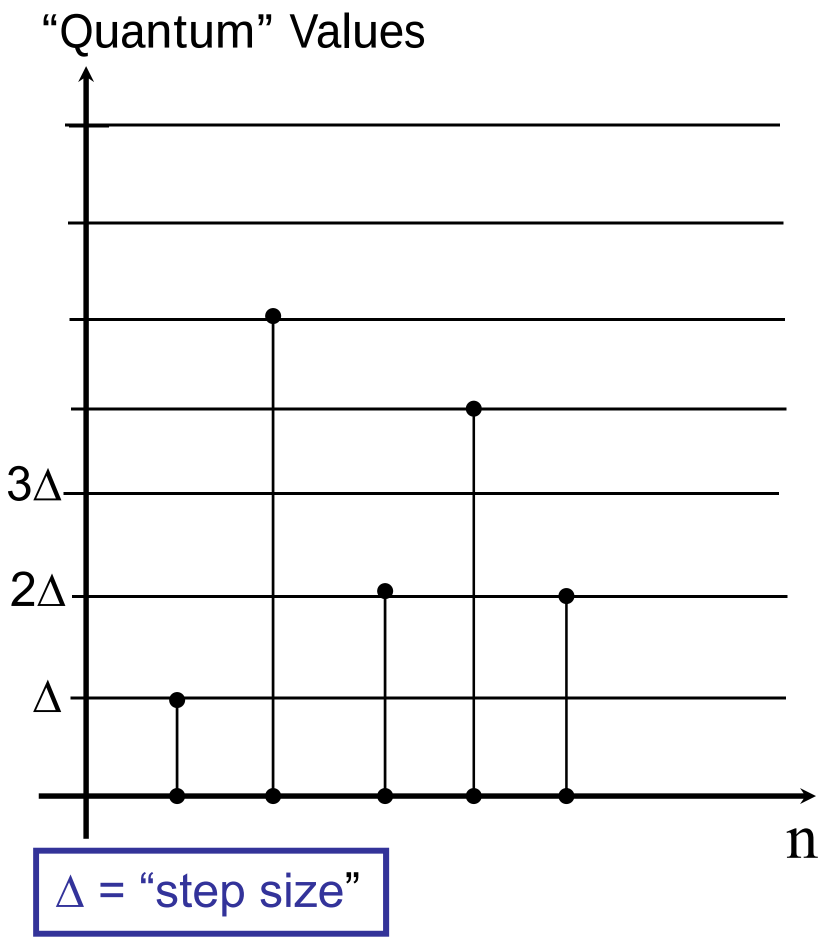
- V. A. Kotelnikov published similar results in 1933 ("On the transmission capacity of the 'ether' and of cables in electrical communications", translation from the Russian), as did the mathematician E. T. Whittaker in 1915 ("Expansions of the Interpolation-Theory", "Theorie der Kardinalfunktionen"), J. M. Whittaker in 1935 ("Interpolatory function theory"), and Gabor in 1946 ("Theory of communication").

Wikipedia, “**Nyquist–Shannon sampling theorem**”

Quantization Process

- Sampling process does not imply any limit on the values of the samples
- We can't represent a continuous range with a finite number of bits.
- The solution:

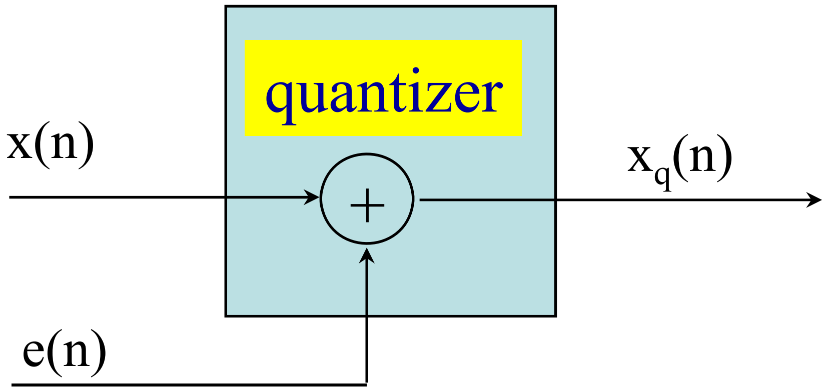
  we impose a grid on the vertical axis



Decisions to Make…

- Resolution: how many bits should we use ?
- Step Size: how should we spread the resulting quantization levels ?
- Quantization noise: how efficient can this process be ?
  - How much noise we insert to the quantized signal ?
  - SNR, MSE ….

Quantization Noise



x(n) : Original signal

$e(n) = x_q(n) - x(n)$: Quantization noise

$x_q(n)$: Quantized signal

Note that:

- For random input signal and some simple assumptions, the variance of the noise:

$$\sigma_e^2 = \frac{\Delta^2}{12}$$

- Less levels = more noise

Quantization - a deeper look ....

Scalar Quantization

- A scalar quantizer Q of size N is a mapping from the real number $x \in R$ into a finite set Y containing N output values ("codewords").
- Y is known as the codebook of the quantizer.
- The mapping action is written as

$$Q(x) = y_i \; ; \; x \in \boldsymbol{R} \quad ; \quad i = 1, \ldots, N$$

- In all cases of practical interest, N is finite so that a finite number of binary digits is sufficient to specify the output value.
- We further assume that the indexing of output values is chosen so that $y_1 < y_2 < \ldots < y_N$

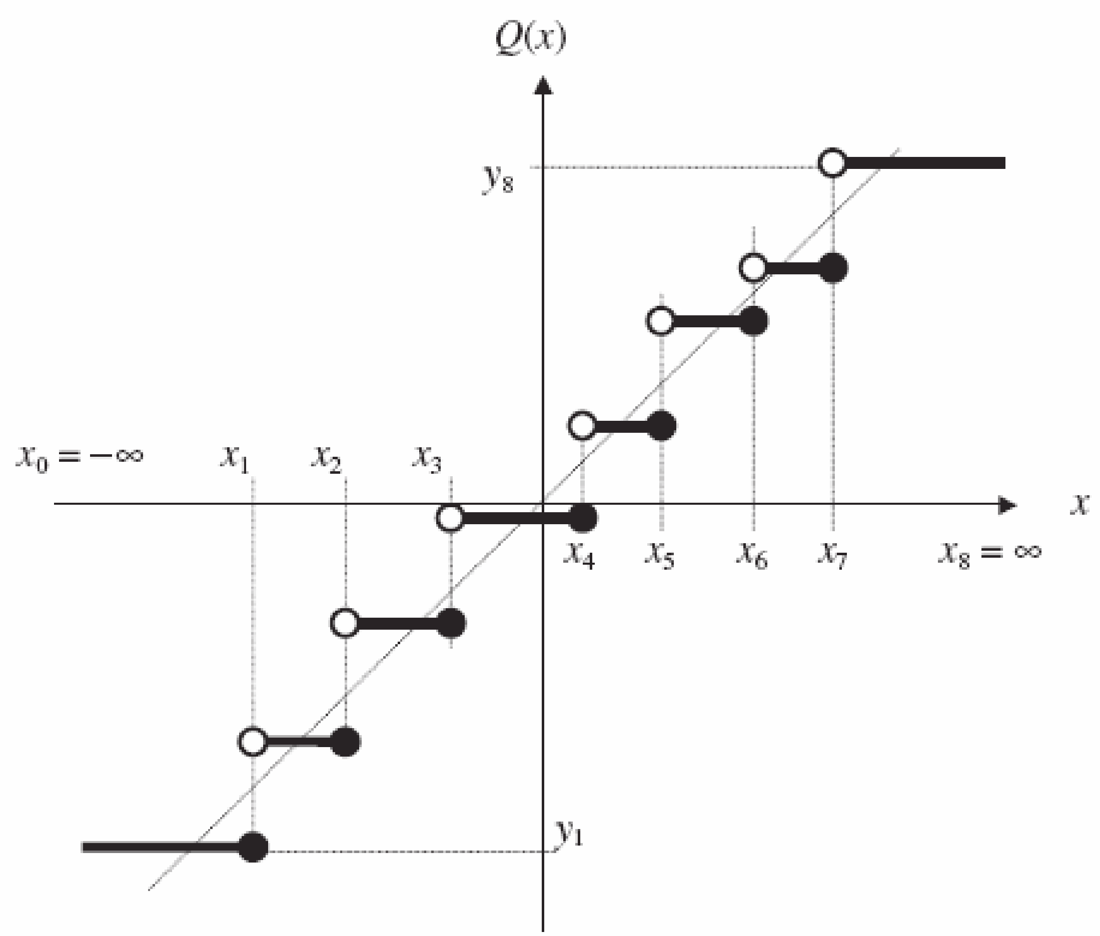
Some Definitions

- Resolution: We define the resolution r of a scalar quantizer as $r = log_2 N$, which measures the number of bits needed to uniquely specify the quantized value.

- Cell: Associated with every N point quantizer is a partition of the real line R into N cells R_i

- Regular Quantizer. A quantizer is defined to be regular if each cell Ri is an interval such that $y_i \in (x_{i-1}, x_i)$.
  - Since most quantizers for coding applications are regular, only regular quantizers are considered in this book.

A Regular Quantizer



Example of the transfer characteristic for a regular quantizer with eight output levels

Distance or Distortion Measure.

- A non-negative cost *d(x, Q(x))* measure associated with quantizing any input value x with a reproduction point *Q(x)*:

$$d(x, Q(x)) = \begin{cases} 0 & x = Q(x) \\ > 0 & \text{Otherwise} \end{cases}$$

- Given a distortion measure we can quantify the performance of a system by the expected value of d.
- The performance of a quantizer is often specified in terms of a Signal-to-Noise Ratio (SNR) , given by:

$$SNR = 10 \log_{10} \frac{\sigma_x^2}{\sigma_d^2}$$

Mean-Squared Error Criterion

- Due to its simplicity and analytical elegance, the Mean-Squared Error (MSE) is widely used in many practical situations.

- Consider the distortion measure defined by the squared error:

$$d(x,\hat{x}) = (x-\hat{x})^2$$

- Then , the expected value of the distortion, or MSE is given by:

$$D = E\left\{(x - Q(x))^2\right\}$$

Uniform Quantizer

- Simple to design and widely used.
- For a uniform quantizer, the transfer $Q(x)$ is:

  $y_{i+1} - y_i = \Delta \;\; ; i = 1, 2, \ldots, N-1$

  $x_{i+1} - x_i = \Delta \;\; ; x_{i+1}, \; x_i \; : finite$

  – Δ : constant, known as the step size.

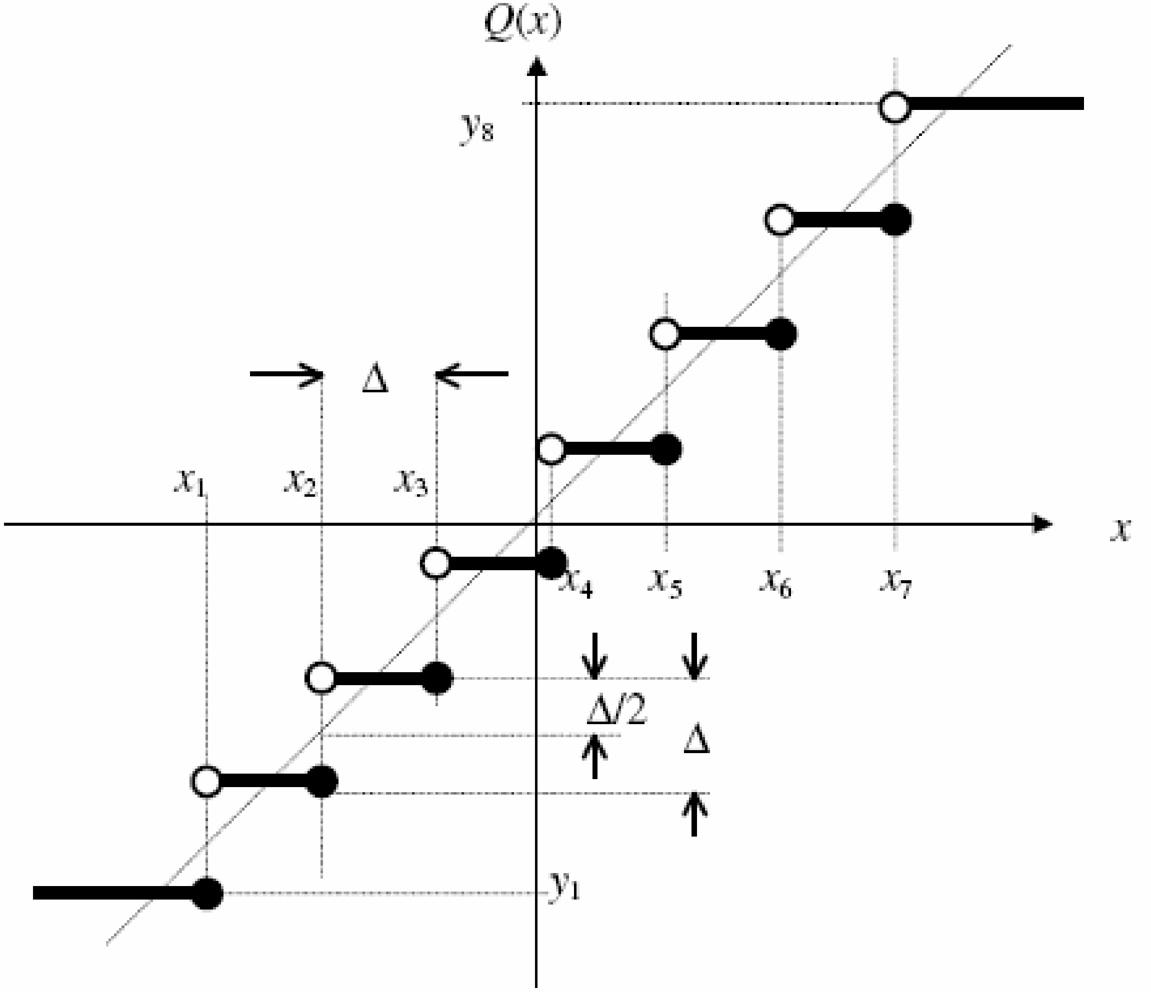
- The output levels for a uniform quantizer are

  $y_i = x_i^- \Delta/2 \;\; ; i = 1, 2, \ldots, N-1$

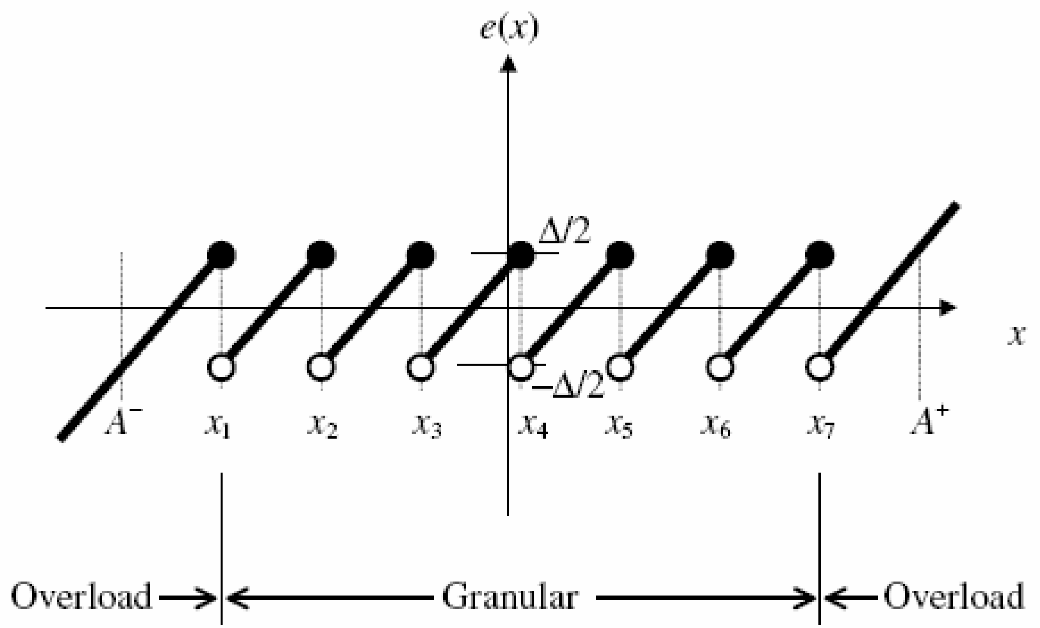
  $y_N = x_{N-1}^+ \Delta/2$

- The quantzation error is defined as: $e(x)=x-Q(x)$

Uniform Quantizer Example



Uniform Quantizer Quantization Error



Note that:

$$|e(x)| \le \Delta / 2 \qquad A^{-} \le x \le A^{+}$$

$$A^{+} = x_{N-1} + \Delta \qquad A^{-} = x_1 - \Delta$$

Uniform Quantizer Design

- One design technique for uniform quantizers is to assign A^+ and A^- to be equal to the maximum and minimum of the input value, respectively.

- Hence, excessive overload error is eliminated.

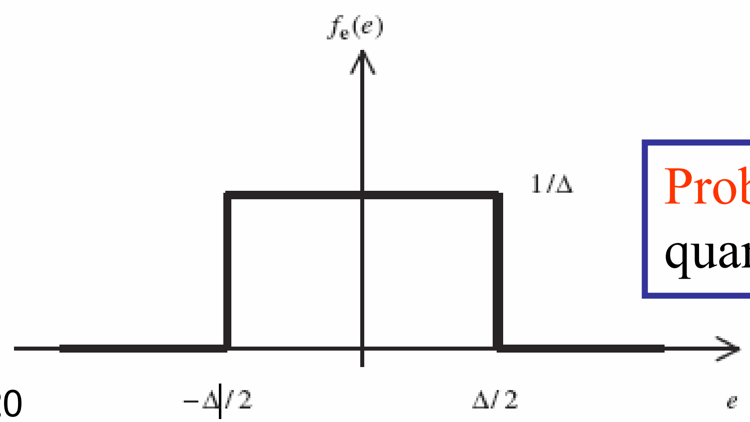
- Once the values of A^+ and A^- are known, the step size can be found by:

$$\Delta = \frac{A^+ - A^-}{N}$$

Uniform Quantizer with Uniform Input

- Consider the following case:
  - Uniform quantizer
  - The input is bounded in the range $[A^{-} .. A^{+}]$
  - The input is uniformly distributed within that range
- The quantizer error considered as a continuous random variable has a uniform distribution:

$$[-\Delta/2 , \Delta/2]$$



Probability density function (PDF) of the quantization error for uniform input distribution

Uniform Input Case Cont'd

- The variance of this error is: $Var(e)=E(e^2)= \Delta^2/12$
- This is equal to the expected value of the distortion if the MSE criterion is adopted.

- Therefore, to reduce the expected distortion, the step size must be decreased, which is accomplished by increasing the quantizer size N.

- An excessively high N, however, requires a large amount of bits, translating directly to higher **coding cost…**

What about Non-Uniform Quantizer ?

- In the specific case that the samples have a certain well defined distribution – identical to the Laplace distribution,

An optimal quantizer can be designed to exploit it:

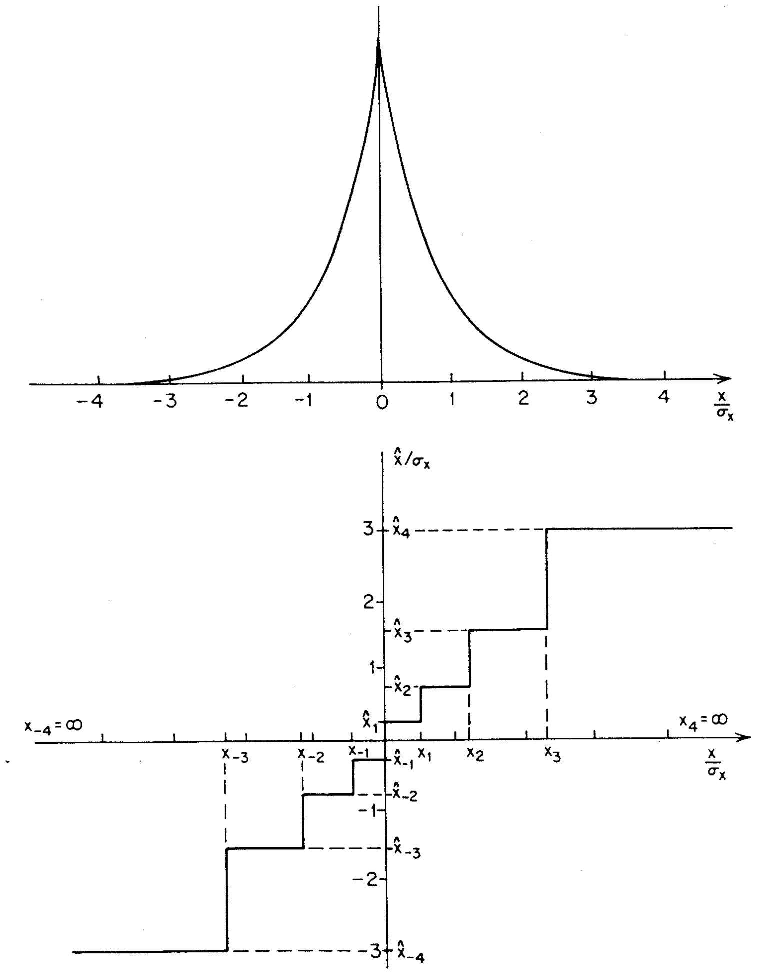


Fig. 5.20 Density function and quantizer characteristic for Laplace density function and a 3-bit quantizer.

Optimal Quantizer

- The primary goal of quantizer design is to select the reproduction levels and the partition regions or cells so as to provide the minimum possible average distortion for:
  - a fixed number of levels N

  or

  - equivalently a fixed resolution r.

- These conditions will serve as references to develop the optimization procedure .

Optimal Quantizer Definition

- A quantizer Q of size N is said to be optimum if it minimizes the expected value of the distortion:

$$D = E\{d(x, Q(x))\} = \sum_{i=1}^{N} \int_{Ri} d(x, y_i) f_x(x) dx$$

  - ***Ri*** is the cells of the quantizer and ***fx(x)*** the PDF of the input random variable *x*.

- Therefore, for optimal operation, it is necessary to specify the output points $\boldsymbol{y_i}$ and partition cells ***Ri*** for a given PDF of ***x*** so as to minimize ***D***.

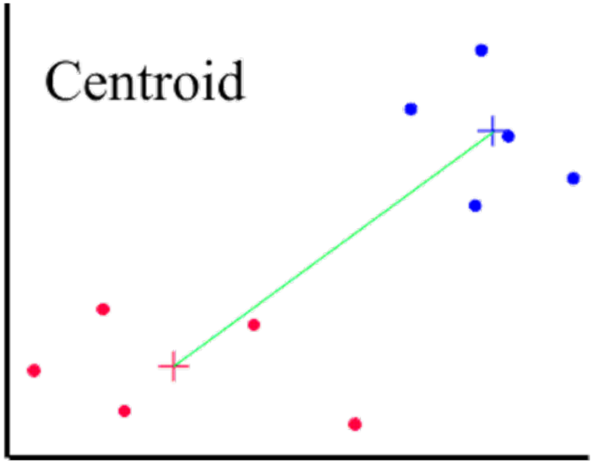
The Nearest-Neighbor Condition for Optimality

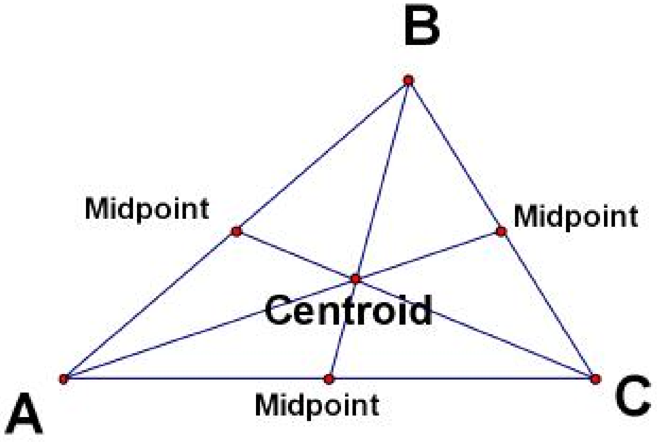
- For a given codebook Y of size N, the optimal partition cells satisfy: $Ri=\{x: d(x,y_i) \leq d(x,y_j))$
  - for all $i \neq j$.

- That is, $Q(x)= y_i$ only if $d(x,y_i) \leq d(x,y_j)$. Hence, $d(x,Q(x)\ ) = min_i(d(x,y_i)$

The Centroid Condition for Optimality

- Definition: We define the centroid cent(Ro), of any nonempty set Ro $\in$ R, as the value y_o (if it exists) that minimizes the expected distortion between x and yo, given that x lies in Ro.





The Lloyd-Max Algorithm

- Step 1. Begin with an initial codebook Y. Set j=1.
  - Decision levels: $\{x_k, k=2,3,\ldots N \; ; x_1 = -\infty)$
  - Representation levels: $\{y_k, k=1,2,\ldots.,N)$
- Step 2. Find y_j such that it is the centroid of (x_j, x_{j+1})
- Step 3. Find x_{j+1} that lies in the middle of $[y_j, y_{j+1}]$

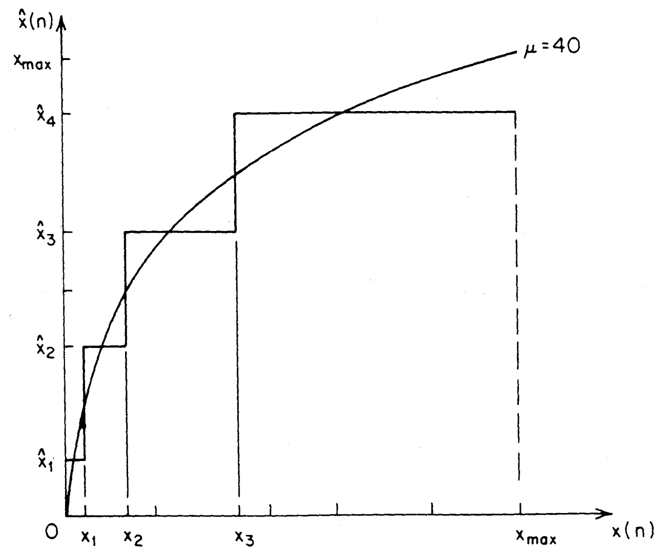
  $$x_{j+1} = (y_j + y_{j+1})/2$$
- Step 4. If j=N, go to Step 5, **otherwise**: $j \leftarrow j+1$, go back to Step 2
- Step 5. Calculate C: the centroid of the region (x_N, ∞).
  - If $|Y_N - C| < \varepsilon$ Than **STOP** , otherwise go to Step 6.
- Step 6. Perform: $y_N \leftarrow y_N - \alpha(Y_N - C)$ and set j=1, go back to Step 2.

$\varepsilon > 0$: A "small" number, chosen according to system demands ; $0 < \alpha < 1$

Constant Quality Quantizers

- For constant quality (fixed SNR), the ratio between step-size and level is constant :logarithmic step!
  - Exact logarithmic quantization is impossible.

- Approximate schemes, called: μ-law and A-law are widespread used in telephony systems.
- Achieve 12 bit quality with 8 bits,



'ig. 5.16 Distribution of quantization levels for a μ-law 3-bit quantizer /ith $\mu = 40$.

Example: constant SNR

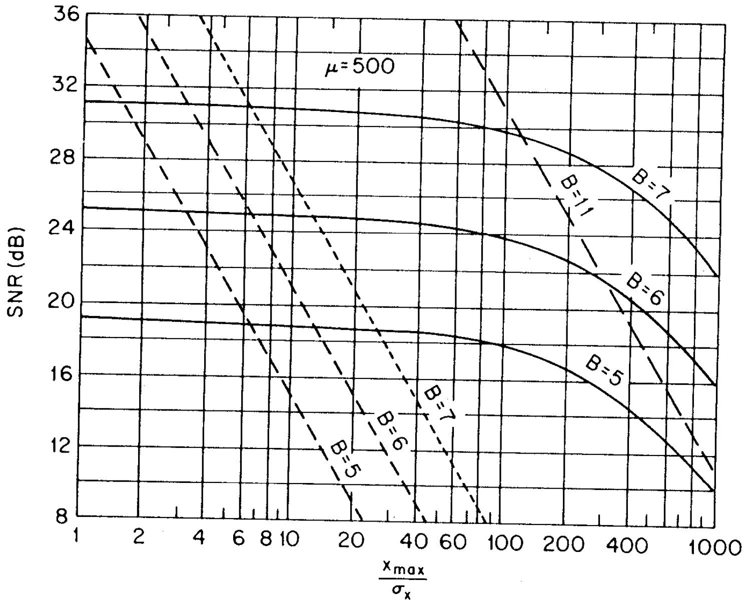
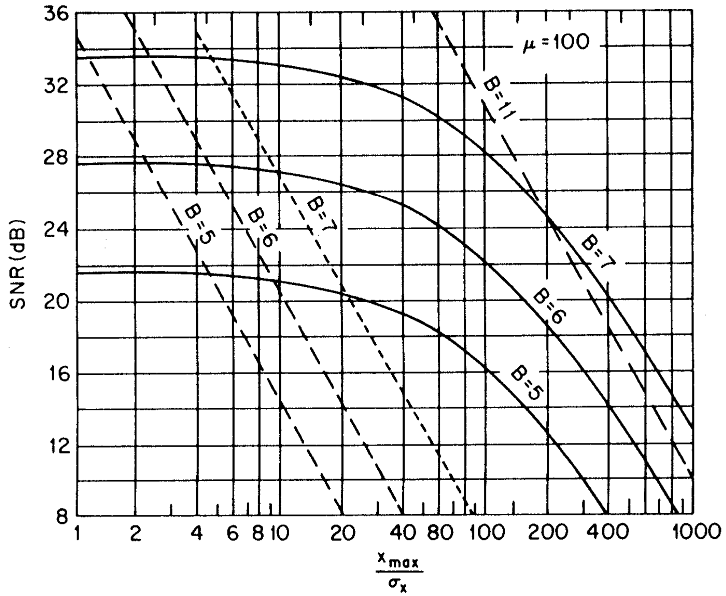


Fig. 5.18 *SNR* for μ-law and uniform quantizers for $\mu = 500$, $B = 5, 6, 7, 11$ bits. (After Smith [10].)



ig. 5.17 *SNR* for μ-law and uniform quantizers as a function of $_{max}/\sigma_x$ for $\mu = 100$ and different numbers of bits (B) of the quantizer. After Smith [10].)

- Almost constant over a wide range of inputs
- Nonlinear - Can't be processed…

Adaptive Quantizers

- The most sophisticated quantizers are adaptive: they change the step size according to the changes in the input signal or the dependancy between adjacent samples.

- The receiver must be able to follow the adaptation !

- Adaptive quantizers: CODECS - next chapter…

Vector Quantization

- Vector quantizer Q is the mapping of:

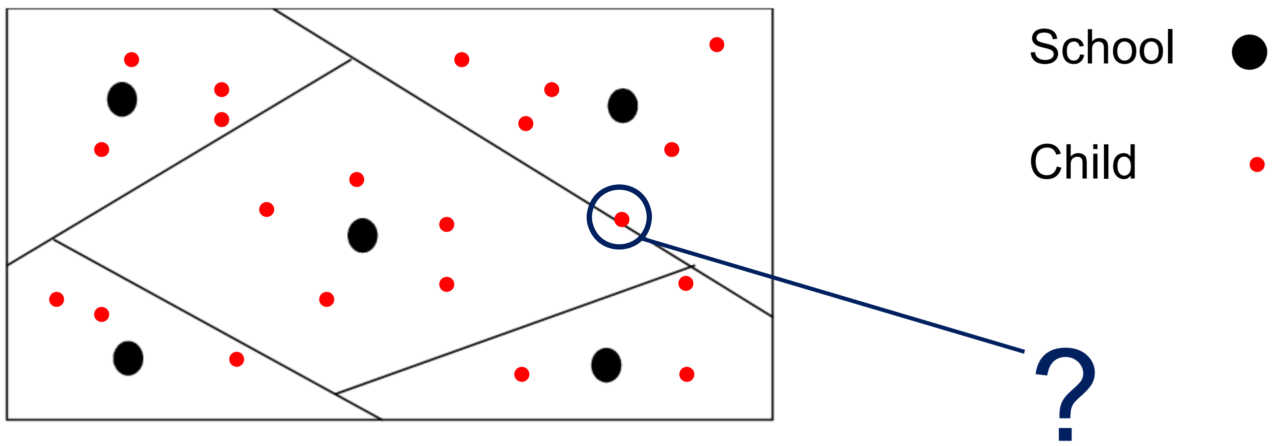
  k-dimensional vectors $\in R^k$

  into a finite set of vectors $Y = \{ y_i : \ i = 1, 2, ..., N \}$

- Each vector is called a CODEWORD
- The set of Codewords is called a CODEBOOK
- Each Codeword is set in a nearest neighbor region called VORONOI Region

A simple case

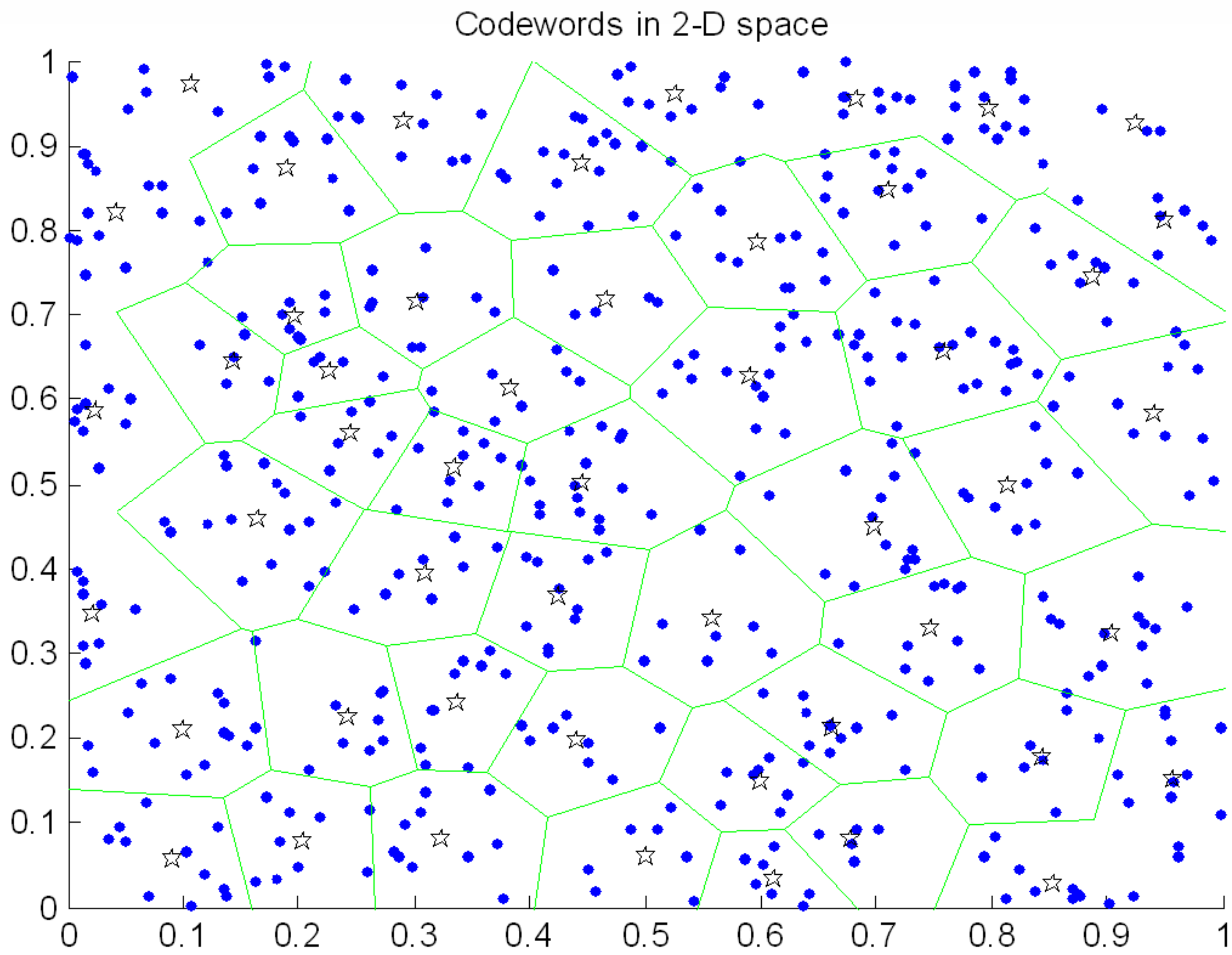
- E.g. "mapping" every child to the closest school, etc.



LBG Algorithm (Linde, Buzu, Gray)

- STEPS
  - Determine the size of codebook ,N.
  - Randomly select N codewords –The **initial codebook**.
  - Classify, according to the Euclidian distance-measure, The input Vectors to the nearest codeword cluster.
  - Compute the New set of Codewords to be the vectors average in each cluster accordingly.
  - Repeat steps 3 And 4 until either The codewords don't change (or the change in the codewords is "small").

Codewords in 2-D space



Speech and Audio Processing Learning Tool

Test Quantization

- Desirable Codebook

Codebook Creation

- Algorithm : LBG / GLloyd
- Vector Dimesion
- Max GLloyd iterations
- Distortion Rate threshold
- Codebook Size / BitRate
- Number Of LBG Splittings
- Distortion Measure: MSE / MAE