Chapter 2

QUANTIZATION

2.1 Quantization and the Source Encoder

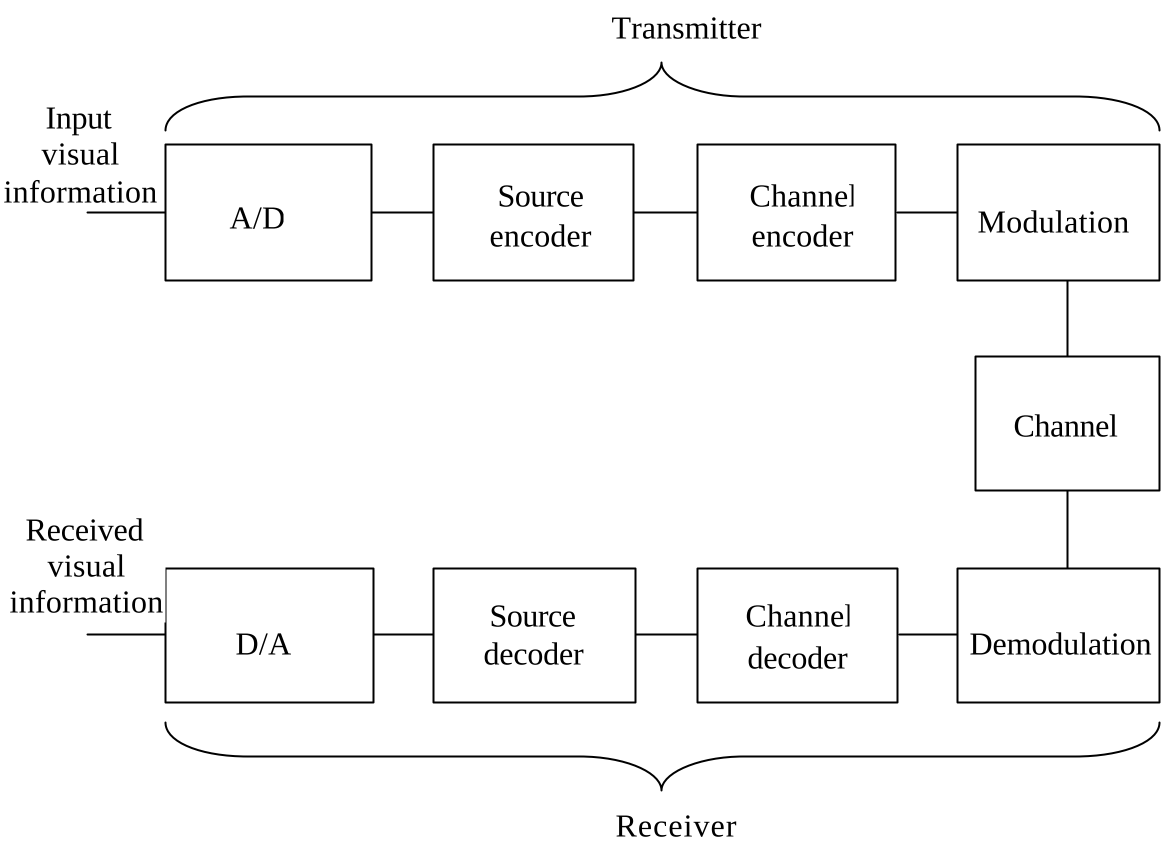


Figure 2.1 Block diagram of a visual communication system

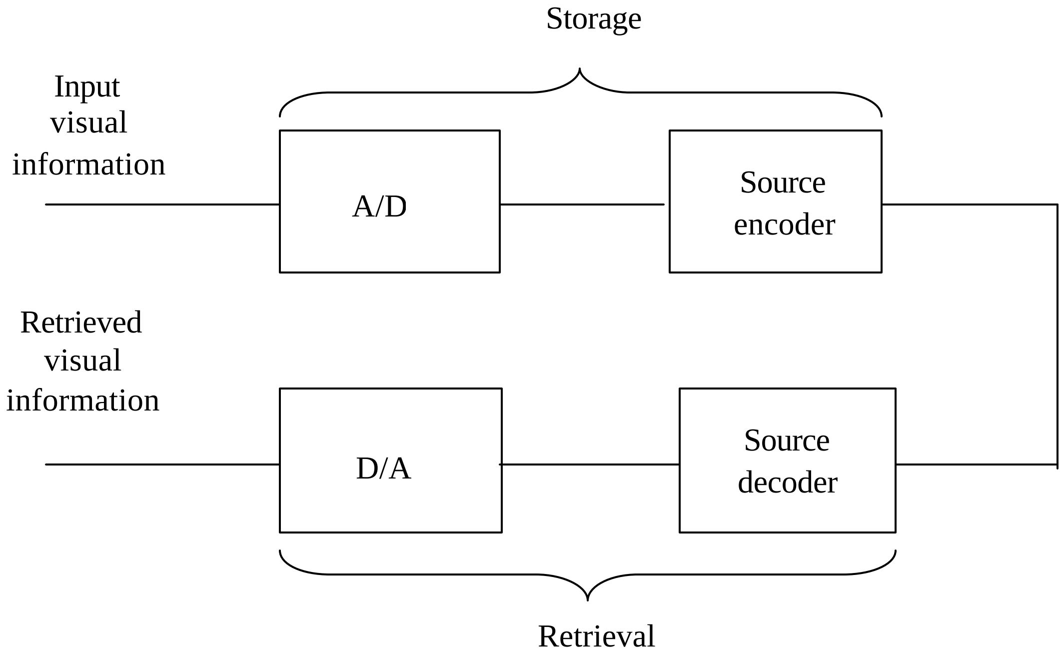
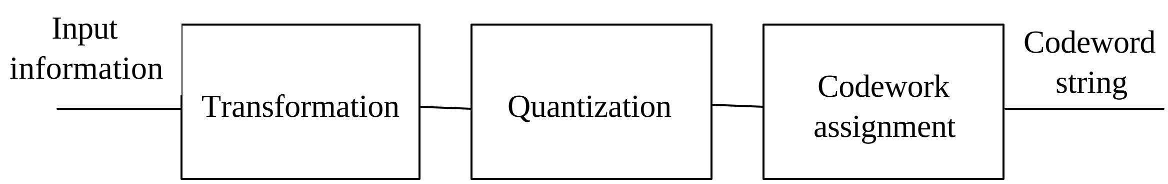
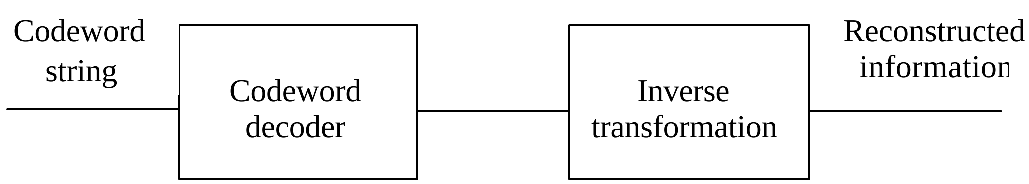


Figure 2.2 Block diagram of a visual storage system



(a) source encoder



(b) source decoder

Figure 2.3 Block diagram of a source encoder and a source decoder

- Quantization: an irreversible process.
- Quantization: a source of information loss.
- Quantization: a critical stage in image and video compression.

  It has significant impact on

  - the distortion of reconstructed image and video
  - the bit rate of the encoder.

2.2 Uniform Quantization

- ♦ Simplest
- ♦ Most popular
- ♦ Conceptually, of great importance.

2.2.1 Basics

2.2.1.1 Definitions

The **input-output characteristic** of the quantizer in Figure 2.4

- ♦ Staircase-like
- ♦ Nonlinear

$$y_i = Q(x) \quad if \quad x \in (d_i, d_{i+1}), \qquad (2.1)$$

where $i = 1, 2, \cdots, 9$ and *Q(x)* is the output of the quantizer with respect to the input *x*.

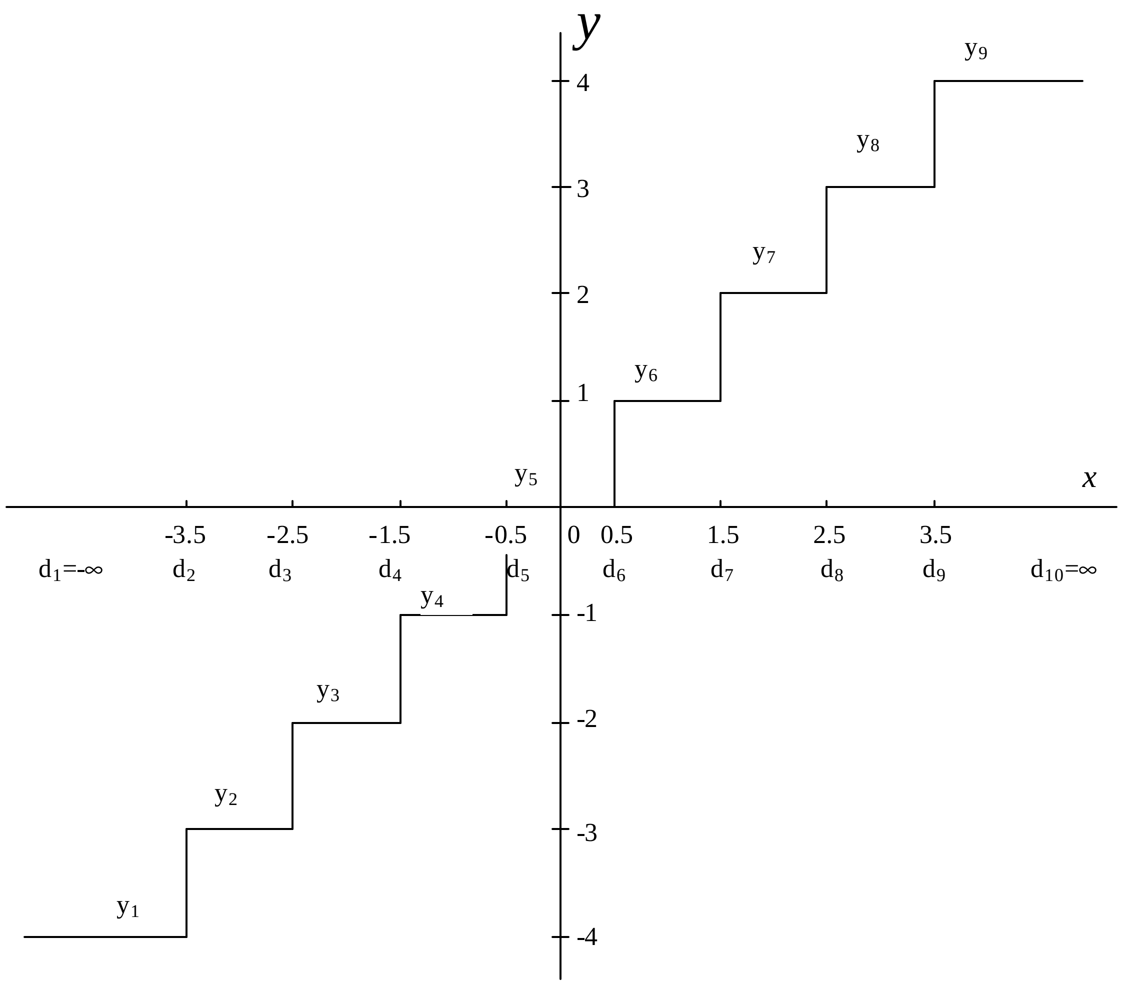


Figure 2. 4 Input-output characteristic of a uniform midtread quantizer

❖ ***Decision levels*** : The end points of the intervals, denoted by d_i

with i being the index of intervals.

❖ ***Reconstruction level*** **(***quantizing level)* :

The output of the quantization, denoted by y_i

❖ ***Step size*** of the qunatizer:

The length of the interval, denoted by Δ

♦ **Two features** of a uniform quantizer:

1. Except possibly the right-most and left-most intervals, all intervals (hence, decision levels) along the x-axis are uniformly spaced.
2. Except possibly the outer intervals, the reconstruction levels of the quantizer are also uniformly spaced.

Furthermore, each inner reconstruction level is the arithmetic average of the two decision levels of the corresponding interval along the x-axis.

- The uniform quantizer depicted in Figure 2.4 is called ***midtread* quantizer**.

  Usually utilized for an odd number of reconstruction levels

- The unifrom quantizer in Figure 2.5 is called ***midrise* quantizer**

  The reconstructed levels do not include the value of zero.

  Uausally utilized for an even number of reconstruction levels

- WLOG, assume: Both input-output characteristics of the midtread and midrise uniform quantizers are odd symmetric with respect to the vertical axis $x=0$.

    Subtraction of statistical mean of input x

    Addition of statistical mean back after quantization

- N: the total number of reconstruction levels of a quantizer.

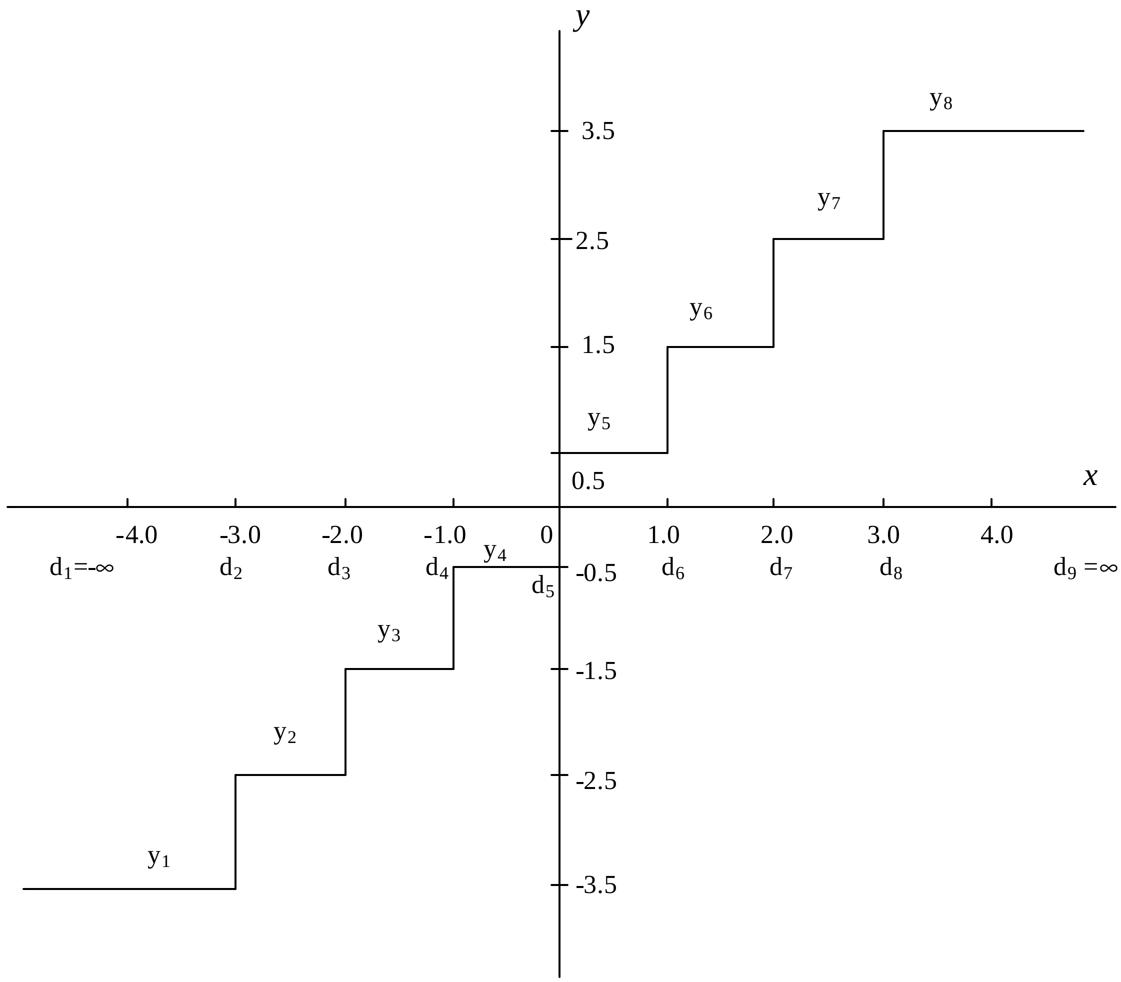


Figure 2. 5 Input-output characteristic of a uniform midrise quantizer

2.2.1.2 Quantization Distortion

- In terms of objective evaluation, we define quantization error, e_q,

$$e_q = x - Q(x), \qquad (2.\ 2)$$

x and *Q(x)* are input and quantized output, respectively.

- Quantization error is often referred to as *quantization noise*.

- Mean square quantization error, MSE_q:

$$MSE_q = \sum_{i=1}^{N} \int_{d_i}^{d_{i+1}} (x - Q(x))^2 f_X(x) dx \qquad (2.\ 3)$$

$f_x(x)$: probability density function (*pdf*)

the outer decision levels may be -∞ or ∞

when the *pdf*, $f_X(x)$, remains unchanged, fewer reconstruction levels (smaller *N*, coarse quantization) result in more distortion.

- Odd symmetry of the input-output characteristic respect to the $x=0$ axis

implies that : $E(x)=0$

$$MSE_q = s_q^{\ 2}$$

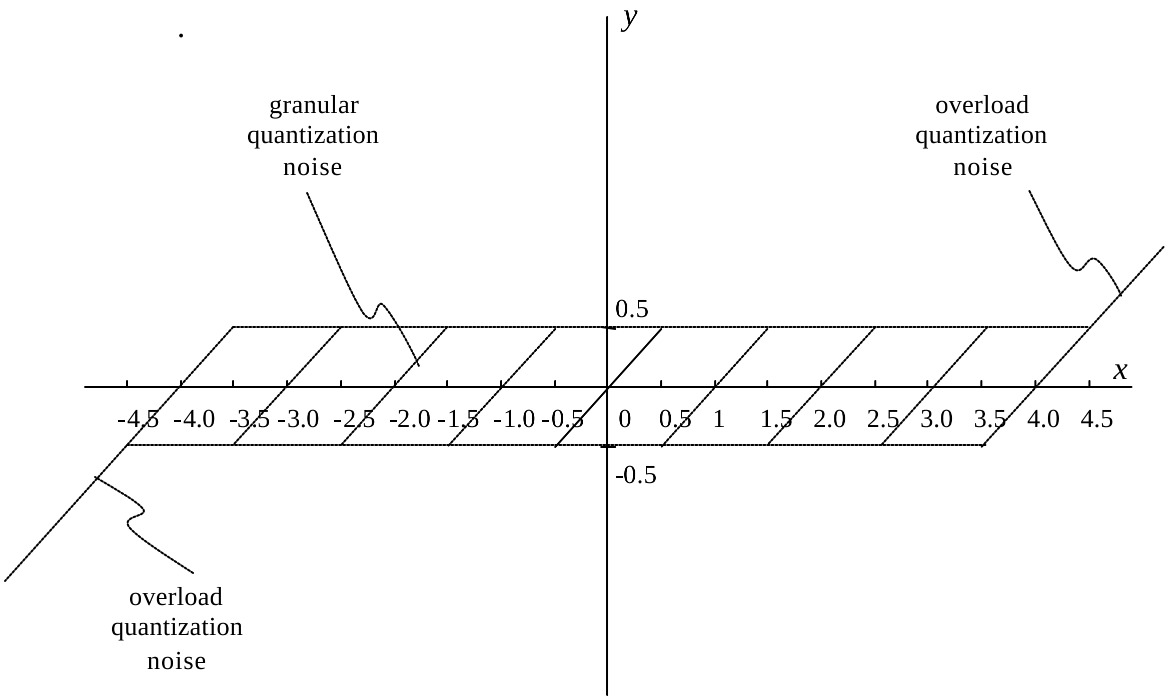


Figure 2. 6 Quantization noise of the uniform midtread quantizer shown in Figure 2.4

2.2.1.3 Quantizer Design

- The design of a quantizer (either uniform or nonuniform):
  - choosing the number of reconstruction levels, N

  - selecting the values of decision levels and reconstruction levels

- The design of a quantizer is equivalent to specifying its input-output characteristic.

- *Optimum* quantizer design:

  For a given probability density function of the input random variable, $f_X(x)$, design a quantizer such that the mean square quantization error, MSE_q, is minimized.

- In the uniform quantizer design:

  - N is usually given.

  - According to the two features of uniform quanitzers,

    Only one parameter that needs to decide: the step size Δ.

- As to the optimum uniform quantizer design, a different *pdf* leads to a different step size.

2.2.2 Optimum Uniform Quantizer

2.2.2.1 Uniform Quantizer with Uniformly Distributed Input

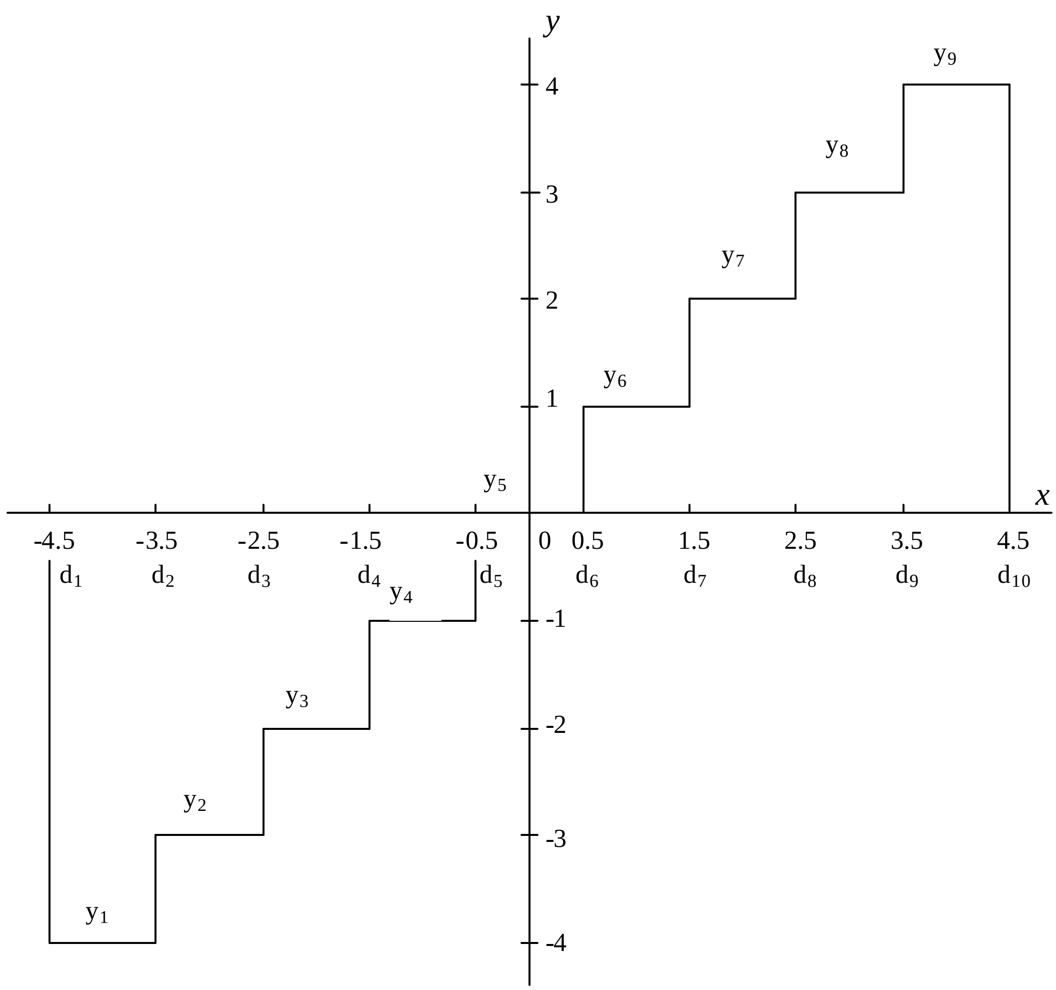


Figure 2. 7 Input-output characteristic of a uniform midtread quantizer with input x uniformly distribution in [-4.5, 4.5], N=9

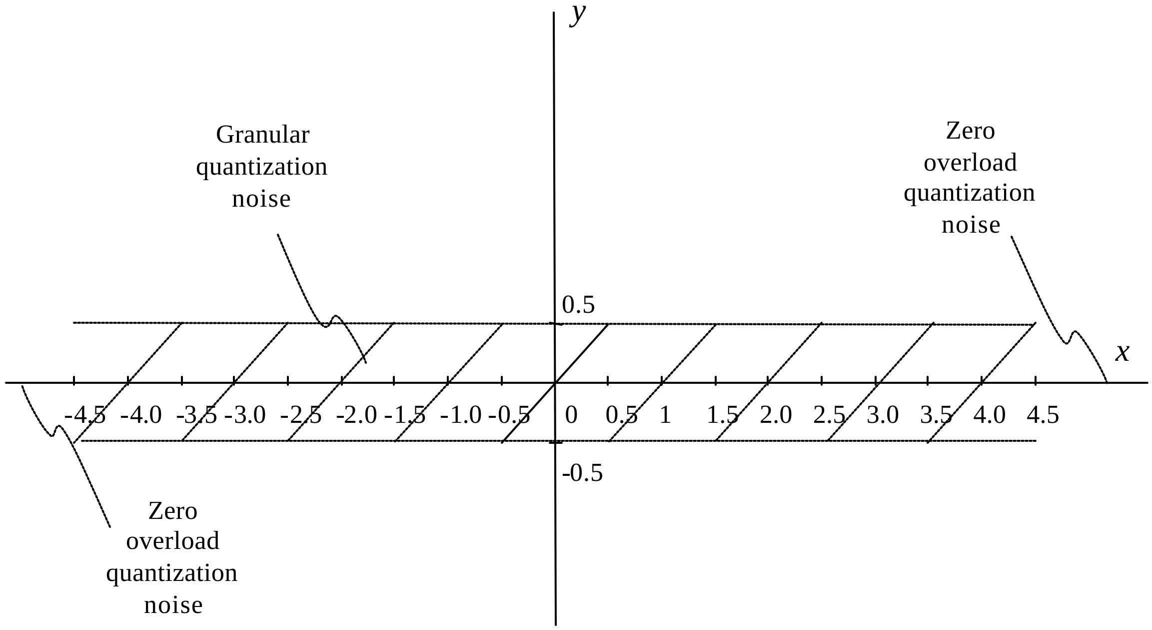


Figure 2. 8 Quantization noise of the quantizer shown in Figure 2.7

The mean square quantization error:

$$MSE_q = N \int_{d_1}^{d_2} (x - Q(x))^2 \frac{1}{N\Delta} dx$$

$$MSE_q = \frac{\Delta^2}{12}. \tag{2. 4}$$

$$SNR_{ms} = 10\log_{10}\frac{s_x^2}{s_q^2} = 10\log_{10} N^2. \quad (2.5)$$

If we assume $N = 2^n$, we then have

$$SNR_{ms} = 20\log_{10} 2^n = 6.02n \quad dB.$$

- The interpretation of the above result:
  - If use natural binary code to code the reconstruction levels of a uniform quantizer with a uniformly distributed input source, then every increased bit in the coding brings out a 6.02 dB increase in the SNR_{ms}.
  - Equivalently from Equation 2.7, whenever the step size of the uniform quantizer decreases by a half, the MSE_q decreases four times.

2.2.2.2 <u>Conditions of Optimum Quantization</u>

- Derived by: [lloyd 1957, 1982; max 1960] for a given pdf $f_X(x)$.

- Sufficient conditions:

1. $x_1 = -\infty$ and $x_{N+1} = +\infty$ (2. 6)

2. $\int_{d_i}^{d_{i+1}} (x - y_i) f_X(x) dx = 0 \quad i = 1, 2, \cdots, N$ (2. 7)

3. $d_i = \frac{1}{2}(y_{i-1} + y_i) \quad i = 2, \cdots, N$ (2. 8)

  - First condition: for an input x whose range is $-\infty < x < \infty$.
  - Second: each reconstruction level is the centroid of the area under the pdf $f_X(x)$ and between the two adjacent decision levels.
  - Third: each decision level (except for the outer intervals) is the arithmetic average of the two neighboring reconstruction levels

- These conditions are general in the sense that there is no restriction imposed on the *pdf*.

2.2.2.3 Optimum Uniform Quantizer with Different Input Distributions

Table 2. 1 Optimal symmetric uniform quantizer for Gaussian, Laplacian and Gamma distributions (having zero mean and unit variance). Dutch[max 1960] [paez 1972]. The numbers enclosed in rectangles are the step sizes.

	Uniform			Gaussian			Laplacian			Gamma		
N	**d_i**	**y_i**	**MSE**	**d_i**	**y_i**	**MSE**	**d_i**	**y_i**	**MSE**	**d_i**	**y_i**	**MSE**
2	-1.000 0.000 $\boxed{1.000}$	-0.500 0.500	8.33 $\times 10^{-2}$	-1.596 0.000 $\boxed{1.596}$	-0.798 0.798	0.363	-1.414 0.000 $\boxed{1.414}$	-0.707 0.707	0.500	-1.154 0.000 $\boxed{1.154}$	-0.577 0.577	0.668
4	-1.000 -0.500 0.000 $\boxed{0.500}$ 1.000	-0.750 -0.250 0.250 0.750	2.08 $\times 10^{-2}$	-1.991 -0.996 0.000 $\boxed{0.996}$ 1.991	-1.494 -0.498 0.498 1.494	0.119	-2.174 -1.087 0.000 $\boxed{1.087}$ 2.174	-1.631 -0.544 0.544 1.631	1.963 $\times 10^{-1}$	-2.120 -1.060 0.000 $\boxed{1.060}$ 2.120	-1.590 -0.530 0.500 1.590	0.320
8	-1.000 -0.750 -0.500 -0.250 0.000 $\boxed{0.250}$ 0.500 0.750 1.000	-0.875 -0.625 -0.375 -0.125 0.125 0.375 0.625 0.875	5.21 $\times 10^{-3}$	-2.344 -1.758 -1.172 -0.586 0.000 $\boxed{0.586}$ 1.172 1.758 2.344	-2.051 -1.465 -0.879 -0.293 0.293 0.879 1.465 2.051	3.74 $\times 10^{-2}$	-2.924 -2.193 -1.462 -0.731 0.000 $\boxed{0.731}$ 1.462 2.193 2.924	-2.559 -1.828 -1.097 -0.366 0.366 1.097 1.828 2.559	7.17 $\times 10^{-2}$	-3.184 -2.388 -1.592 -0.796 0.000 $\boxed{0.796}$ 1.592 2.388 3.184	-2.786 -1.990 -1.194 -0.398 0.398 1.194 1.990 2.786	0.132
16	-1.000 -0.875 -0.750 -0.625 -0.500 -0.375 -0.250 -0.125 0.000 $\boxed{0.125}$ 0.250 0.375 0.500 0.625 0.750 0.875 1.000	-0.938 -0.813 -0.688 -0.563 -0.438 -0.313 -0.188 -0.063 0.063 0.188 0.313 0.438 0.563 0.688 0.813 0.938	1.30 $\times 10^{-3}$	-2.680 -2.345 -2.010 -1.675 -1.340 -1.005 -0.670 -0.335 0.000 $\boxed{0.335}$ 0.670 1.005 1.340 1.675 2.010 2.345 2.680	-2.513 -2.178 -1.843 -1.508 -1.173 -0.838 -0.503 -0.168 0.168 0.503 0.838 1.173 1.508 1.843 2.178 2.513	1.15 $\times 10^{-2}$	-3.648 -3.192 -2.736 -2.280 -1.824 -1.368 -0.912 -0.456 0.000 $\boxed{0.456}$ 0.912 1.368 1.824 2.280 2.736 3.192 3.648	-3.420 -2.964 -2.508 -2.052 -1.596 -1.140 -0.684 -0.228 0.228 0.684 1.140 1.596 2.052 2.508 2.964 3.420	2.54 $\times 10^{-2}$	-4.320 -3.780 -3.240 -2.700 -2.160 -1.620 -1.080 -0.540 0.000 $\boxed{0.540}$ 1.080 1.620 2.160 2.700 3.240 3.780 4.320	-4.050 -3.510 -2.970 -2.430 -1.890 -1.350 -0.810 -0.270 0.270 0.810 1.350 1.890 2.430 2.970 3.510 4.050	5.01 $\times 10^{-2}$

- A uniform quantizer is optimum when the input has uniform distribution.
- Normally, if the pdf is not uniform, the optimum quantizer is not a uniform quantizer.
- Due to the simplicity of uniform quantization, however, it may sometimes be desirable to design an optimum uniform quantizer for an input with a nonuniform distribution.
- Under these circumsatances, however, Equations 2.13, 2.14 and 2.15 are not a set of simultaneous equations one can hope to solve with any ease. Numerical procedures were suggested to solve for design of optimum uniform quantizers.
- Max derived uniform quantization step size Δ for an input with a Gaussian distribution [max 1960].
- Paez and Glisson found step size Δ for Laplacian and Gamma distributed input signals [paez 1972].

- Zero mean:

  In Table 2.1, all distributions: a zero mean.

  If the mean is not zero, only a shift in input is needed when applying these results.

- Unite variance:

  In Table 2.1, all distributions: a unit variance.

  If the standard deviation is not unit, the tabulated step size needs to be multiplied by the standard deviation.

2.3 Nonuniform Quantization

2.3.1 Optimum (Nonuniform) Quantization

Table 2. 2 Optimal symmetric quantizer for uniform, Gaussian, Laplacian and Gamma distributions (The uniform distribution is between [-1, 1], the other three distributions have zero mean and unit variance.) [lloyed 1957, 1982] [max 1990] [paez 1972]

	Uniform			Gaussian			Laplacian			Gamma		
N	d_i	y_i	**MSE**	d_i	y_i	**MSE**	d_i	y_i	**MSE**	d_i	y_i	**MSE**
2	-1.000 0.000 1.000	-0.500 0.500	8.33×10^{-2}	$-\infty$ 0.000 ∞	-0.799 0.799	0.363	$-\infty$ 0.000 ∞	-0.707 0.707	0.500	$-\infty$ 0.000 ∞	-0.577 0.577	0.668
4	-1.000 -0.500 0.000 0.500 1.000	-0.750 -0.250 0.250 0.750	2.08×10^{-2}	$-\infty$ -0.982 0.000 -0.982 ∞	-1.510 -0.453 0.453 1.510	0.118	$-\infty$ -1.127 0.000 1.127 ∞	-1.834 -0.420 0.420 1.834	1.765×10^{-1}	$-\infty$ -1.205 0.000 1.205 ∞	-2.108 -0.302 0.302 2.108	0.233
8	-1.000 -0.750 -0.500 -0.250 0.000 0.250 0.500 0.750 1.000	-0.875 -0.625 -0.375 -0.125 0.125 0.375 0.625 0.875	5.21×10^{-3}	$-\infty$ -1.748 -1.050 -0.501 0.000 0.501 1.050 1.748 ∞	-2.152 -1.344 -0.756 -0.245 0.245 0.756 1.344 2.152	3.45×10^{-2}	$-\infty$ -2.377 -1.253 -0.533 0.000 0.533 1.253 2.377 ∞	-3.087 -1.673 -0.833 -0.233 0.233 0.833 1.673 3.087	5.48×10^{-2}	$-\infty$ -2.872 -1.401 -0.504 0.000 0.504 1.401 2.872 ∞	-3.799 -1.944 -0.859 -0.149 0.149 0.859 1.944 3.799	7.12×10^{-2}
16	-1.000 -0.875 -0.750 -0.625 -0.500 -0.375 -0.250 -0.125 0.000 0.125 0.250 0.375 0.500 0.625 0.750 0.875 1.000	-0.938 -0.813 -0.688 -0.563 -0.438 -0.313 -0.188 -0.063 0.063 0.188 0.313 0.438 0.563 0.688 0.813 0.938	1.30×10^{-3}	$-\infty$ -2.401 -1.844 -1.437 -1.099 -0.800 -0.522 -0.258 0.000 0.258 0.522 0.800 1.099 1.437 1.844 2.401 ∞	-2.733 -2.069 -1.618 -1.256 -0.942 -0.657 -0.388 -0.128 0.128 0.388 0.657 0.942 1.256 1.618 2.069 2.733	9.50×10^{-3}	$-\infty$ -3.605 -2.499 -1.821 -1.317 -0.910 -0.566 -0.266 0.000 0.266 0.566 0.910 1.317 1.821 2.499 3.605 ∞	-4.316 -2.895 -2.103 -1.540 -1.095 -0.726 -0.407 -0.126 0.126 0.407 0.726 1.095 1.540 2.103 2.895 4.316	1.54×10^{-2}	$-\infty$ -5.050 -3.407 -2.372 -1.623 -1.045 -0.588 -0.229 0.000 0.229 0.588 1.045 1.623 2.372 3.407 5.050 ∞	-6.085 -4.015 -2.798 -1.945 -1.300 -0.791 -0.386 -0.072 0.072 0.386 0.791 1.300 1.945 2.798 4.015 6.085	1.96×10^{-2}

- The solution to optimum quantizer design for finitely many reconstruction levels N when input x obeys Gaussian distribution was obtained numerically [lloyd 1957, 1982, max 1960].
- Lloyd-Max quantizers.
- The design for Laplacian and Gamma distribution were tabulated in [paez 1972].
- Performance comparison

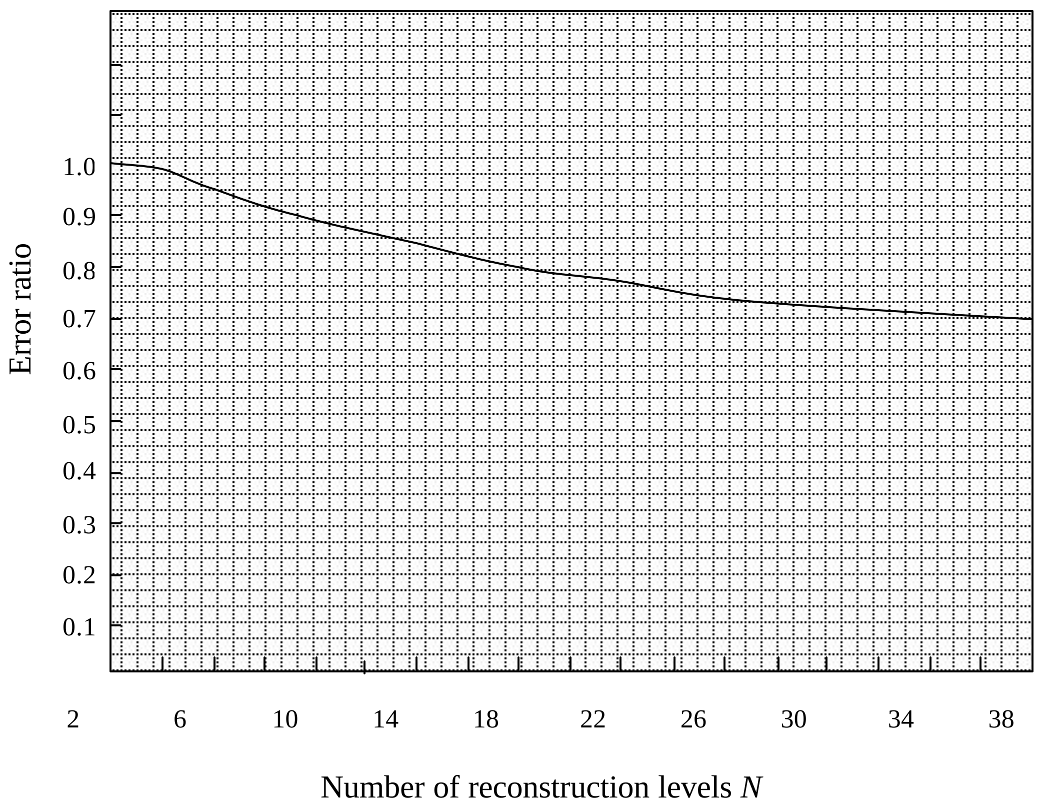


Figure 2. 9 Ratio of error for optimal quantizer to error for optimum uniform quantizer vs. number of reconstruction levels *N*. (Minimum mean square error for Gaussian distributed input with a zero mean and unit variance). Data from [max 1960].

2.4 Adaptive Quantization

- Consider an optimum quantizer for a Gaussian distributed input, N=8. Figure 2.13.

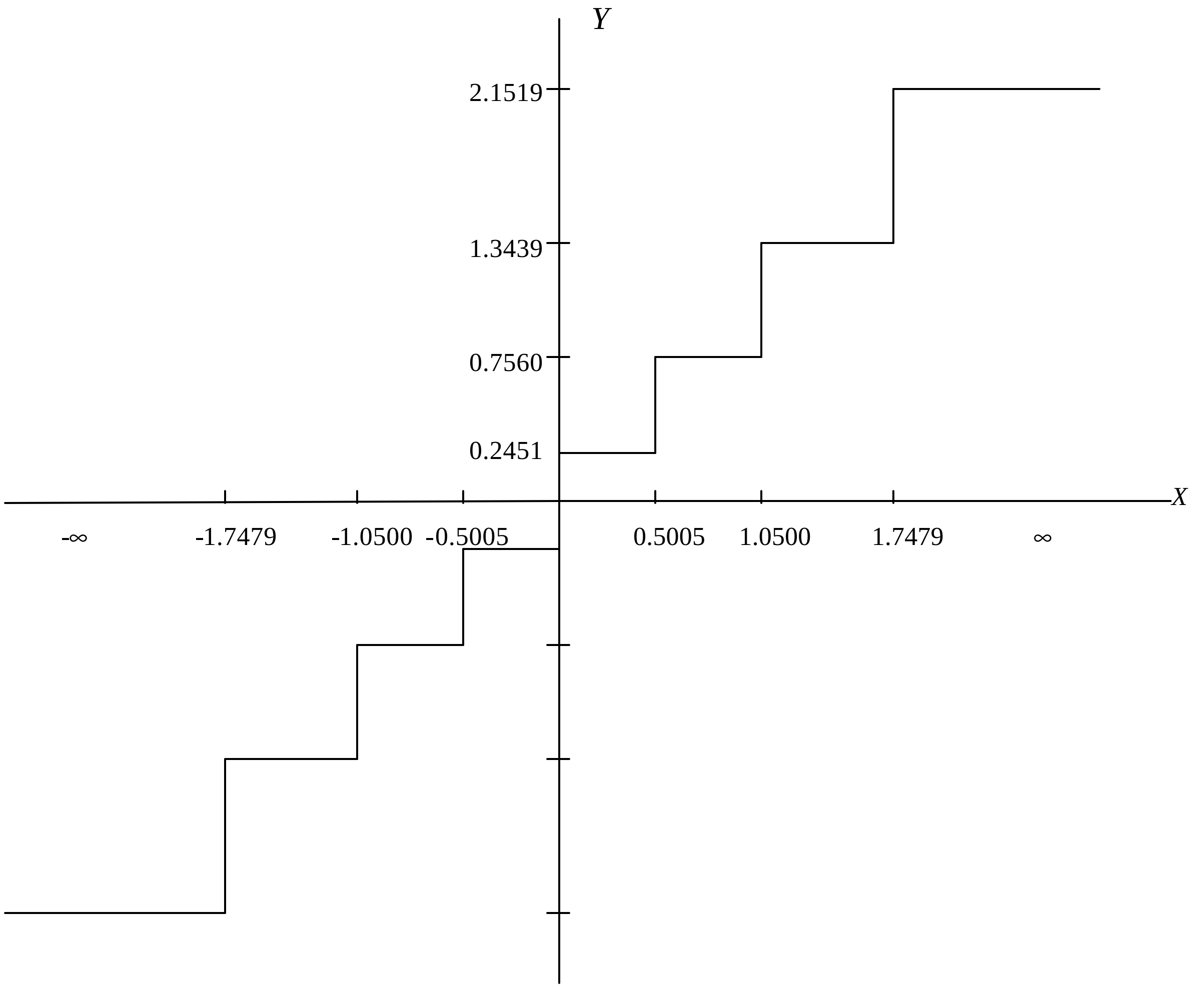


Figure 2.10 Input-output characteristic of the optimal quantizer for Gaussian distribution with zero mean, unit variance, and N=8

  - This curve reveals that the decision levels are densely located in the central region of the *x*-axis and coarsely elsewhere.

- Input-output characteristic: time-invariant.
  - Not designed for nonstationary input signals.
  - Even for a stationary input signal, if its *pdf* deviates from that with which the optimum quantizer is designed, then what is called *mismatch* will take place and the performance of the quantizer will deteriorate.
  - Two main types of mismatch
    - One is called variance mismatch.
    - Another type is *pdf* mismatch.

❖ Adaptive quantization attempts to make the quantizer design adapt to the varying input statistics in order to achieve better performance.

- By statistics, we mean the statistic mean, variance (or the dynamic range), and type of input *pdf*.
  - When the mean of the input changes, differential coding (discussed in the next chapter) is a suitable method to handle the variation.

- For other types of cases, adaptive quantization is found to be effective. The price paid in adaptive quantization is processing delay and an extra storage requirement as seen below.

- There are two different types of adaptive quantization:
  - forward adaptation and
  - backward adaptation.

❖ An alternative way to define quantization [jayant 1984].

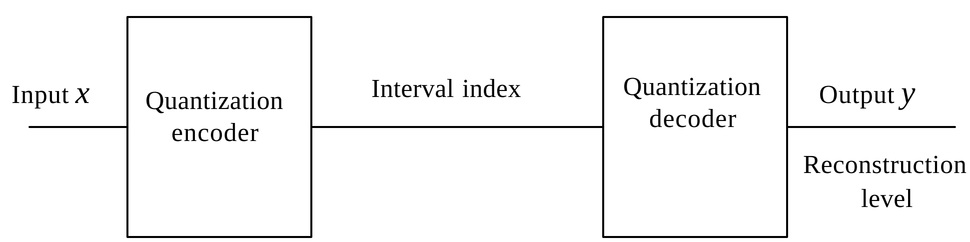


Figure 2. 11 A two-stage model of quantization

➢ In the quatization encoder:

the input to quantization is converted to the index of an interval into which the input x falls.

➢ In the quantization decoder:

the index is mapped to (the codeword that represents) the econstruction level corresponding to the interval in the decoder.

2.4.1 Forward Adaptive Quantization

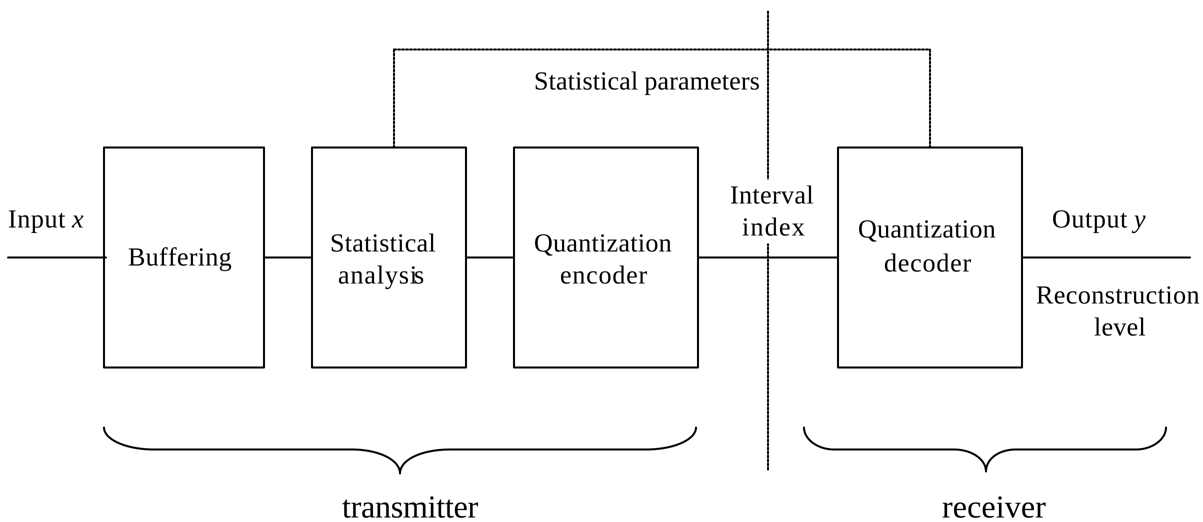


Figure 2. 12 Forward adaptive quantization

- The encoder setting parameters:
  ***side* information**.

- <u>The selection of block size is a critical issue</u>.
  - ✓If the size is small, the adaptation to the local statistics will be effective, but the side information needs to be sent frequently. (more bits used for sending the side information)
  - ✓If the size is large, the bits used for side information decrease. The adaptation becomes less sensitive to changing statistics, and both processing delay and storage required increase.
  - ✓In practice, a proper compromise between quantity of side information and effectiveness of adaptation produces a good selection of the block size.

2.4.2 Backward Adaptive Quantization

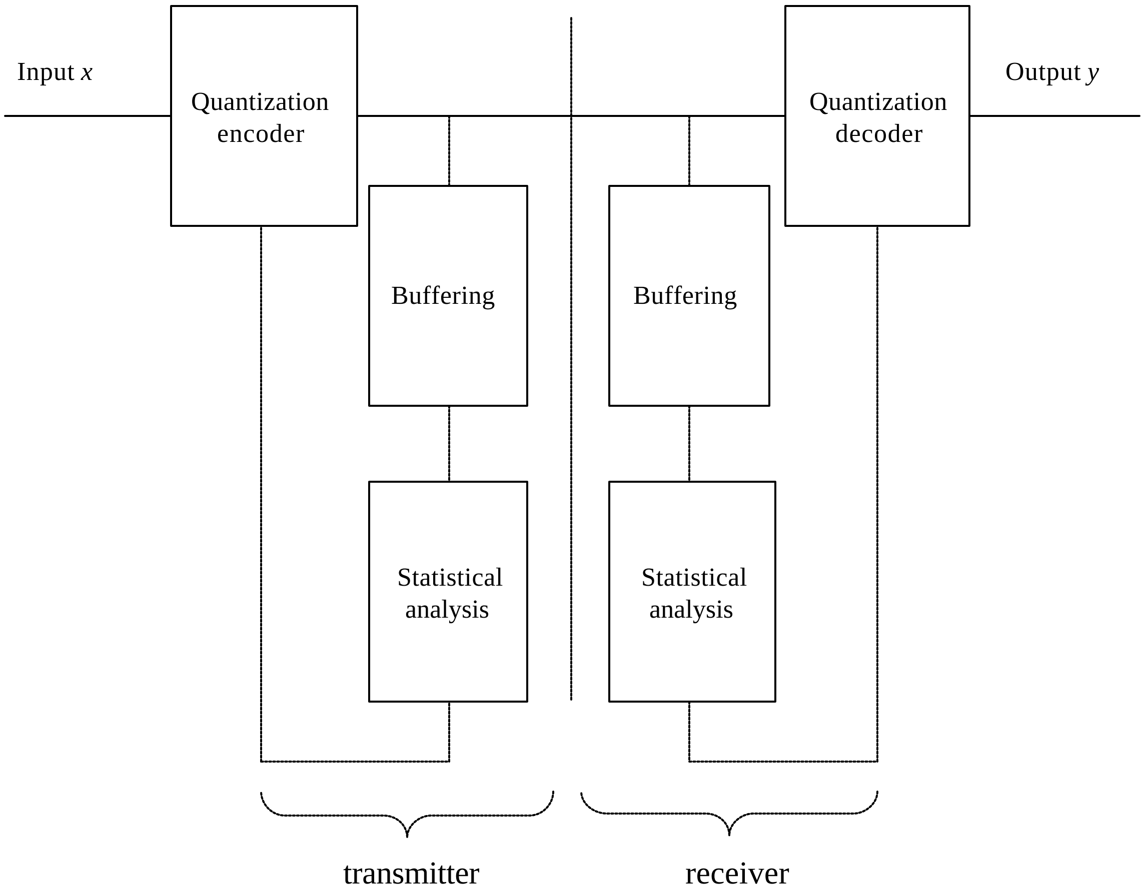


Figure 2. 13 Backward adaptive quantization

- There is no need to send side information.
- The sensitivity of adaptation to the changing statistics will be degraded, however, since, instead of the original

input, only is the output of the quantization encoder used in the statistical analysis. That is, the quantization noise is involved in the statistical analysis.

2.4.3 Adaptive Quantization with a One-Word Memory

- Intuitively, it is expected that observing a sufficient large number of input or output (quantized) data is necessary in order to track the changing statistics and then adapt the quantizer setting in adaptive quantization.
- Jayant showed that effective adaptations can be realized with an explicit memory of only one word.

  That is, either one input sample, *x*, in forward adaptive quantization or a quantized output, *y*, in backward adaptive quantization is sufficient [jayant 1973].
- The idea is as follows.

- If at moment t_i the input sample x_i falls into the outer interval, then the step size at the next moment t_{i+1} will be enlarged by a factor of m_i ($m_i > 1$).
- If the input x_i falls into an inner interval close to *x=0* then, the multiplier $m_i < 1$.
- In this way, the quantizer adapts itself to the input to avoid *overload* as well as *underload* to achieve better performance.

2.4.4 Switched Quantization

- Another adaptive quantization scheme. Figure 2.14.
- It is reported that this scheme has shown improved performance even when the number of quantizers in the bank, *L*, is two [jayant 1984].

- Interestingly, it is noted that as $L \rightarrow \infty$, the switched quantization converges to the adaptive quantizer discussed above.

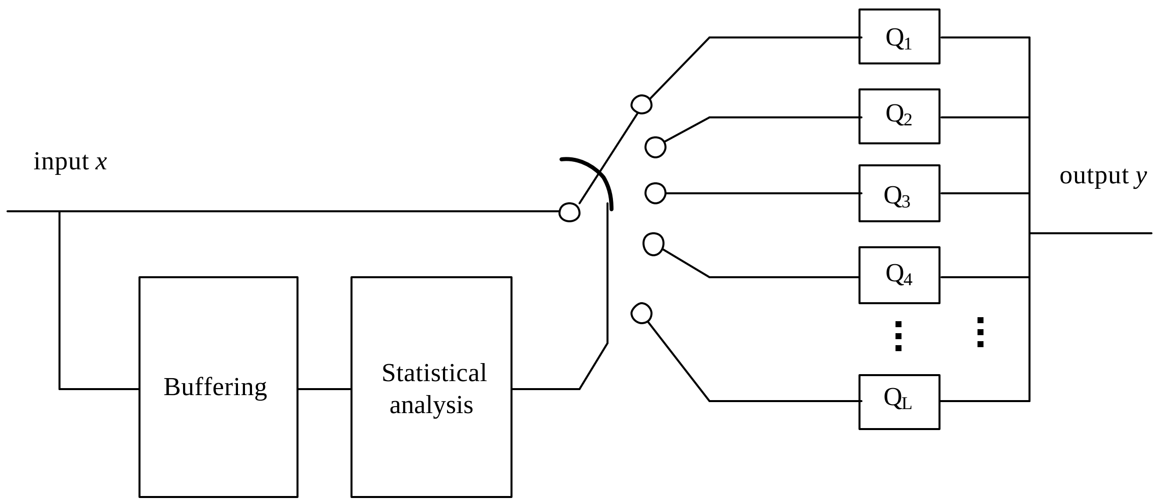


Figure 2. 14 Switched quantization

2.5 PCM

- Pulse code modulation (PCM) is closely related to quantization.
- PCM is the earliest, best established, and most frequently applied coding system

  despite the fact that

  - the most bit-consuming digitizing system (since it encodes each pixel independently)
  - a very demanding system in terms of bit error rate on the digital channel.
- PCM is now the most important form of pulse modulation.
- Pulse modulation links an analog signal to a pulse train in the following way.
  - The analog signal is first sampled
  - The sampled values are used to modulate a pulse train.

- If the modulation is through the amplitude of the pulse train: PAM.
- If the modified parameter of the pulse train is the pulse width: PWM.
- If the pulse width and magnitude are constant -- only the position of pulses is modulated by the sample values – then: PPM.

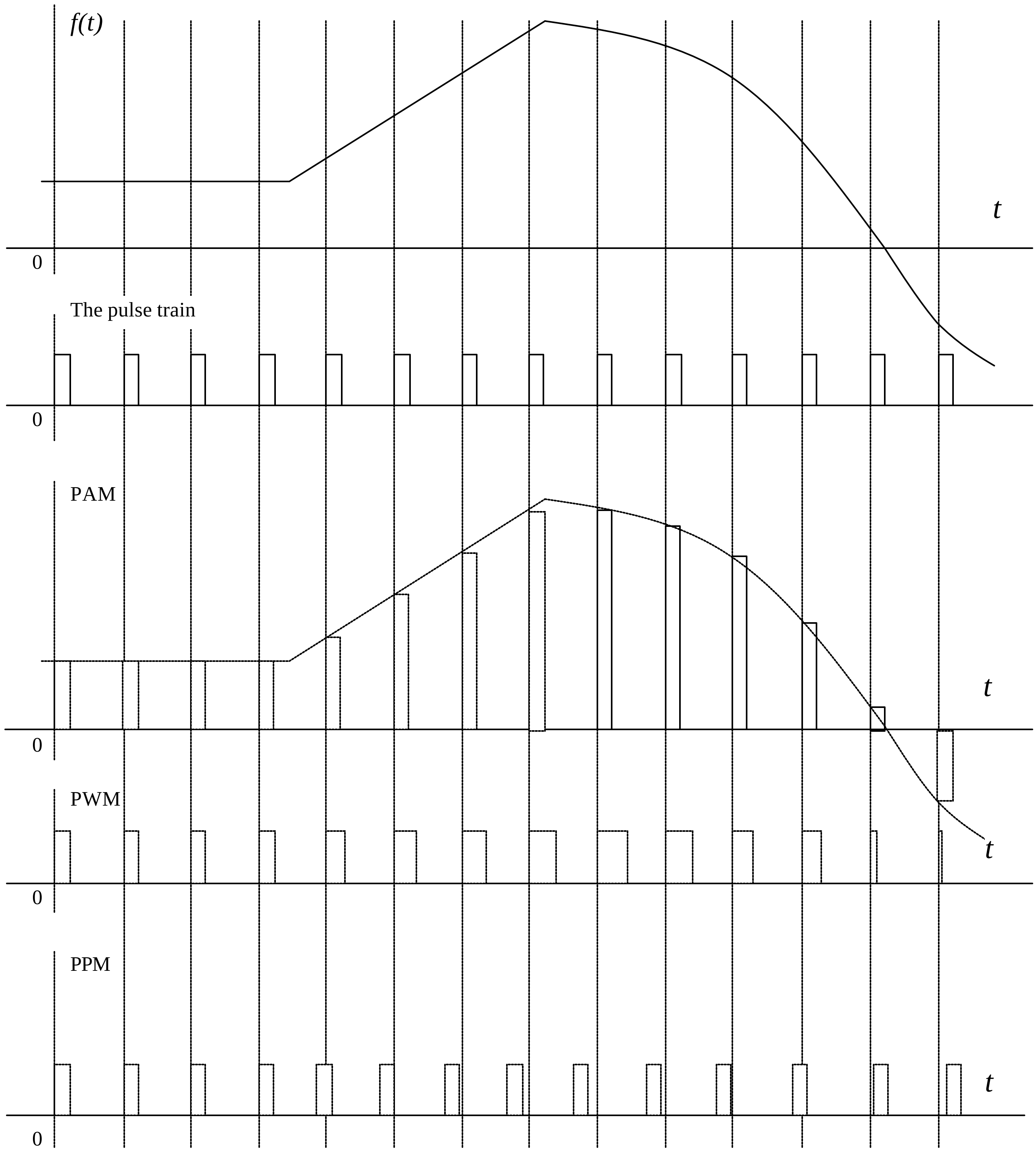
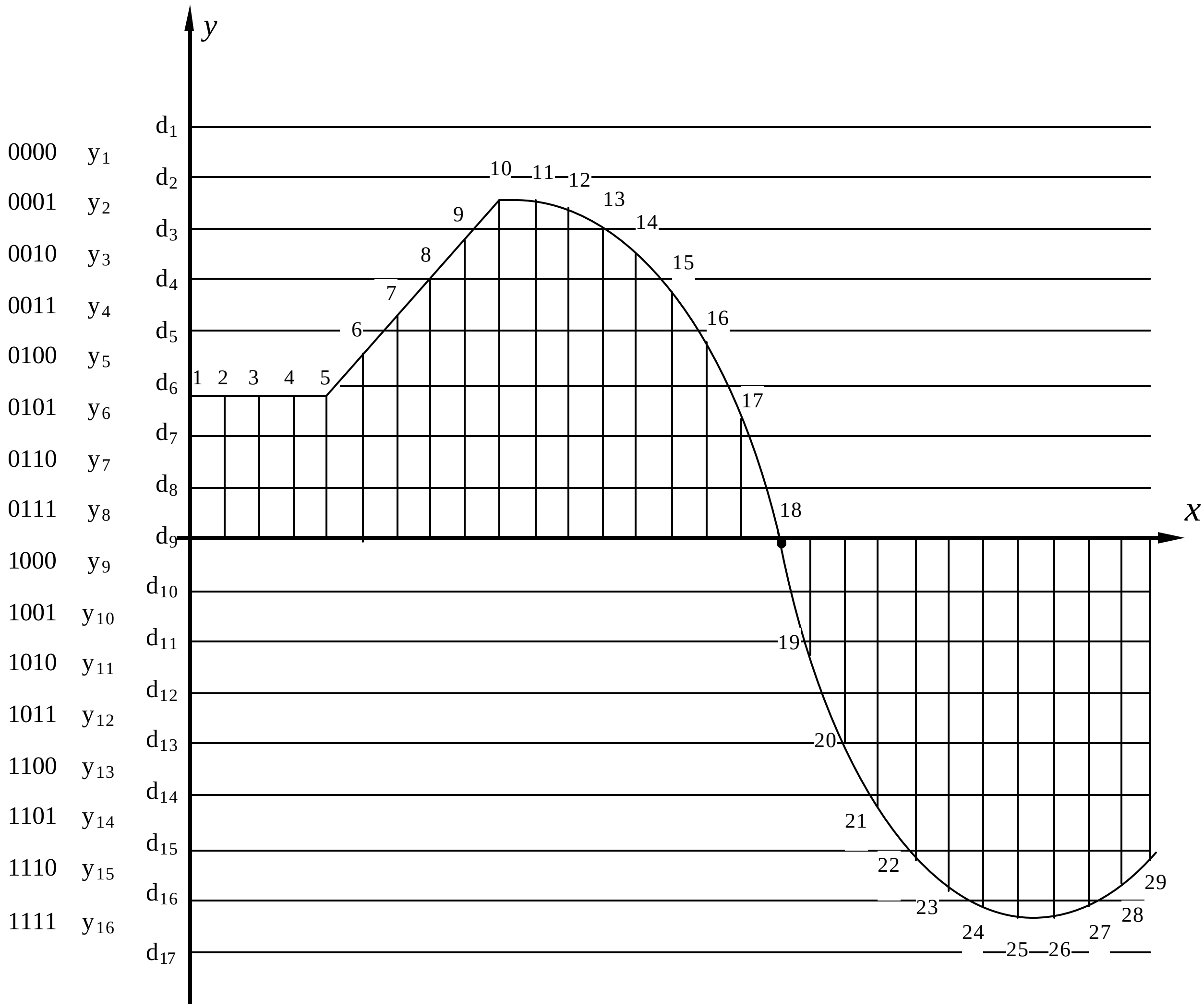


Figure 2.15 Pulse modulation



Output code (from left to right, from top to bottom):

0101	0101	0101	0101	0101	0100	0011	0011	0010	0001	0001	0001	0010	0010	0011
0100	0101	1000	1010	1100	1101	1110	1110	1111	1111	1111	1111	1110	1110	

Figure 2. 16 Pulse code modulation (PCM)

- In PCM, a sampling, a uniform quantization, and a natural binary code converts the input **analog** signal into a **digital** signal.
- In this way, an analog signal modulates a pulse train with the natural binary code.
- By far, PCM is more popular than other types of pulse modulation

  since the *code* modulation is much more robust against various noises than amplitude modulation, width modulation and position modulation.
- In fact, almost all coding techniques include a PCM component.
- In digital image processing, given digital images usually appear in PCM format.
- It is known that an acceptable PCM representation of monochrome picture requires 6 to 8 bits per pixel [huang 1975].
- It is used so commonly in practice that its performance normally serves as a standard against which other coding techniques are compared.

References

[gersho 1977] A, Gersho, ‘‘Quantization,’’ *IEEE Communications Magazine*, pp. 6-29, September 1977.

[fleischer 1964] P. E. Fleischer, ‘‘Sufficient conditions for achieving minimum distortion in quantizer,’’ *IEEE Int. Convention Records*, part I, vol. 12, pp. 104-111, 1964.

[gonzalez 1992] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Addison-Wesley Publishing Company, Reading, Massachusetts, 1992.

[goodall 1951] W. M. Goodall, ‘‘Television by pulse code modulation,’’ Bell System Technical Journal, pp. 33-49, January 1951.

[huang 1975] T. S. Huang, “PCM picture transmission,” *IEEE Spectrum*, vol. 2, pp. 57-63, Dec. 1965.

[jayant 1970] N. S. Jayant, ‘‘Adaptive delta modulation with one-bit memory,’’ *Bell System Technical Journal*, 49, 321-342, March 1970.

[jayant 1973] N. S. Jayant, ‘‘Adaptive quantization with one word memory,’’ *Bell System Technical Journal*, 52, 1119-1144, September 1973.

[jayant 1984] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*, Prentice Hall, 1984.

[li 1995] W. Li and Y.-Q. Zhang, ‘‘Vector-based signal processing and qunatization for image and video compression,’’ *Proceedings of the IEEE*, vol. 83, no. 2, pp. 317-335, February 1995.

[lloyd 1982] S. P. Lloyd, ‘‘Least squares quantization in PCM,’’ Institute of Mathematical Statistics Meeting, Atlantic City, NJ, September 1957; *IEEE Transactions on Information Theory*, pp. 129-136, March 1982.

[max 1960] J. Max, ‘‘Quantizing for minimum distortion,’’ *IRE Trans. Information Theory*, it-6, pp. 7-12, 1960.

[musmann 1979] H. G. Musmann, “Predictive Image Coding,’’ in *Image Transmission Techniques*, W. K. Pratt (Ed.), Academic Press, New York, 1979.

[paez 1972] M. D. Paez and T. H. Glisson, ‘‘Minimum mean squared error qunatization in speech PCM and DPCM Systems,’’ *IEEE Trans. on Communications*, pp. 225-230, April 1972.

[panter 1951] P. F. Panter and W. Dite, "Quantization distortion in pulse count modulation with nonuniform spacing of levels," *Proc. IRE*, 39, 44-48, January 1951.

[sklar 1988] B. Sklar, *Digital Communications: Fundamentals and Applications*, PTR Pretice Hall, Englewood Cliffs, NJ, 1988.

[smith 1957] B. Smith, "Instantaneous companding of quantized signals," *Bell System Technical Journal*, vol. 36, pp. 653-709, May 1957.

[sayood 1996] K. Sayood, *Introduction to Data Compression*, Morgan Kaufmann Publishers, San Francisco, CA, 1996.