# ECS332 2017/1 Part III.3 Dr.Prapun

## 8 PCM

**8.1.** Generally, **analog** signals are continuous in time and in range (amplitude); that is, they have values at every time instant, and their values can be anything within the range. On the other hand, **digital** signals exist only at discrete points of time, and their amplitude can take on only finitely (or countably) many values.

**8.2.** Suppose we want to convey an *analog* message $m(t)$ from a source to our destination. We now have many options.

(a) Use $m(t)$ to modulate a carrier $A\cos(2\pi f_c t)$ using AM (Defn. 4.61), FM (Defn. 5.15), or PM (Defn. 5.3) techniques studied earlier.

(b) **Sample** the continuous-time message $m(t)$ to get a discrete-time message $m[n]$:

$$m(t) \to \boxed{\text{Sampler}} \to m[n]$$

- May LPF $m(t)$ before sampling to eliminate aliasing (and reduce out-of-band noise).
- Need to make sure that the sampling rate $f_s$ is fast enough.

(i) Send $m[n]$ using **analog pulse modulation** techniques (PAM, PWM, PPM) illustrated in Example 7.5.

- Note that $m[n]$ is a sequence of numbers. Even when $m(t)$ (and hence $m[n]$) is bounded, there are uncountably many possibilities for these numbers. They lie in a continuous dynamic range.

- Therefore, information is transmitted basically in analog (not digital) form but the transmission takes place at discrete times.

(ii) In **digital pulse modulation**, $m[n]$ is represented by a (discrete) number (symbol) selected from a finite alphabet set.

i. In **Pulse Code Modulation** (**PCM**), we further **quantize** $m[n]$ into $m_n$ which has finitely many levels. Then, convert $m_n$ into binary sequence represented by two basic pulses.

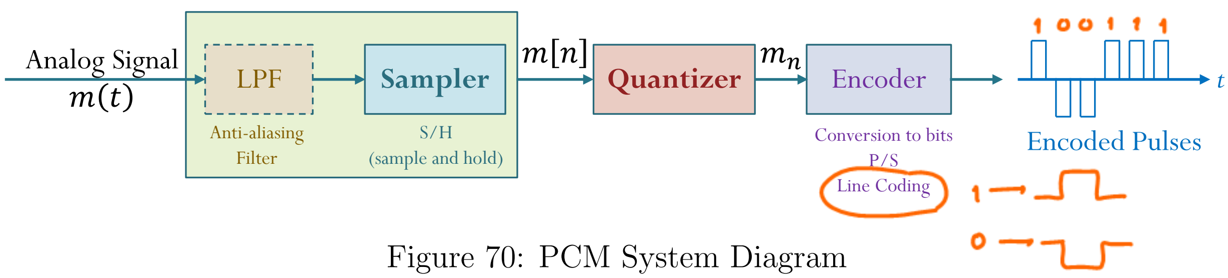



Figure 70: PCM System Diagram

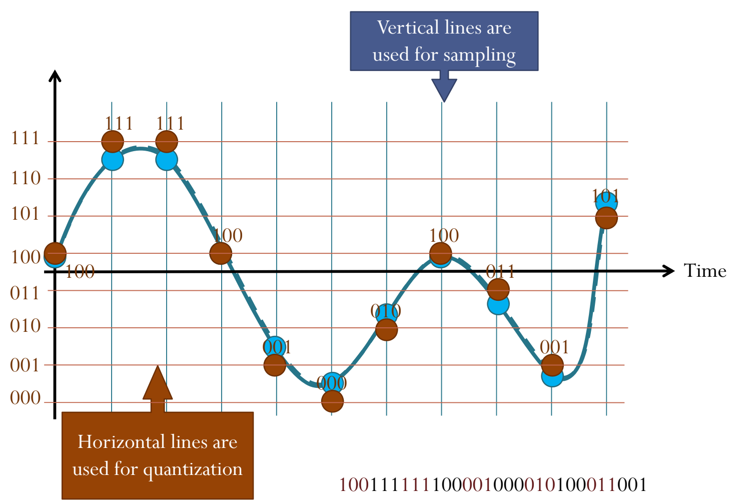



Figure 71: An overview of PCM

ii. There are also other forms of "source coding" such as DPCM and DM.

**Definition 8.3.** **Pulse-code modulation** (**PCM**) is a discrete-time, discrete-amplitude waveform-coding process, by means of which an analog signal is directly represented by a sequence of coded pulses.

**8.4.** Advantages of PCM

(a) **Robustness** to channel noise, distortion, and interference


(i) assuming these corruptions are within limits.

(ii) With analog messages, on the other hand, any distortion or noise, no matter how small, will distort the received signal.

(b) Efficient regeneration of the coded signal along the transmission path by using regenerative **repeaters**.

- For analog communications,
  - A message signal becomes progressively weaker as it travels along the channel, whereas the cumulative channel noise and the signal distortion grow progressively stronger.
  - Ultimately the signal is overwhelmed by noise and distortion.
  - Amplification offers little help because it enhances the signal and the noise by the same proportion.
  - Consequently, the distance over which an analog message can be transmitted is limited by the initial transmission power.
- For digital communications,
  - We can set up repeater stations along the transmission path at distances short enough to be able to detect signal pulses before the noise and distortion have a chance to accumulate sufficiently.
  - At each repeater station the pulses are detected, and new, clean pulses are transmitted to the next repeater station, which, in turn, duplicates the same process.

(c) Digital signals can be **coded** to remove redundancy, protect against channel corruption, and provide privacy.

**8.5.** PCM has emerged as the most favored scheme for the digital transmission of analog information-bearing signals (e.g., voice and video signals). [4, p 267]

- The method of choice for the construction of public switched telephone networks (PSTNs).

**8.6.** Technically, the term "modulation" used in PCM, DPCM, and DM is a misnomer. In reality, PCM, DPCM, and DM are different forms of source coding. [4, p 277]

## 8.1 Uniform Memoryless Quantization

**Definition 8.7.** Through **quantization**, each sample value $m[n]$ is transformed to (, e.g., approximated, or "rounded off," to the nearest) **quantized level** [6, p 320] or **quantum level** [3, p 545] (permissible number) taken from a finite alphabet set $\mathcal{M}$.

**8.8.** Sampling vs. Quantization:

(a) Sampling operates in the time domain. Quantization operates in the amplitude domain.

(b) The sampling process is the link between an analog waveform and its discrete-time representation. The quantization process is the link between an analog waveform and its discrete-amplitude representation.

**Definition 8.9.** Suppose the range of the quantizer is $(-m_p, m_p)$.

- Note that, here, $m_p$ is not necessarily the peak value of $m(t)$. The amplitudes beyond $\pm m_p$ will be simply chopped off.

A simple (memoryless) quantizer partitions the range into $L$ intervals. Each sample value is approximated by the midpoint of the interval in which the sample value falls.

- Each sample is now represented by one of the $L$ numbers.
- Such a signal is known as an $L$**-ary** digital signal.

step size

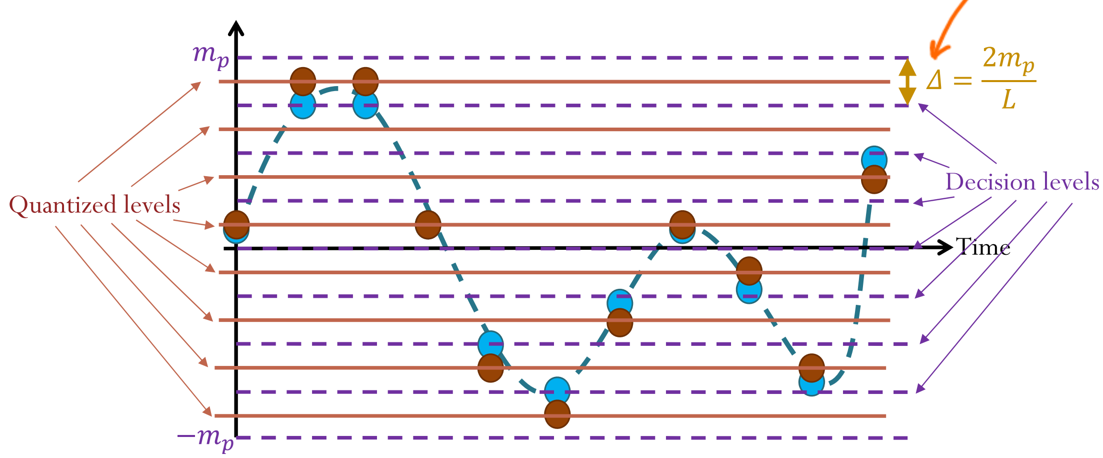



Figure 72: Quantized levels

- The length of each interval is denoted by $\Delta = \frac{2m_p}{L}$.
- Because the quantized levels are uniformly spaced, we say that the quantizer is **uniform**.

**8.10.** Quantization introduces permanent errors that appear at the receiver as quantization noise in the reconstructed signal.

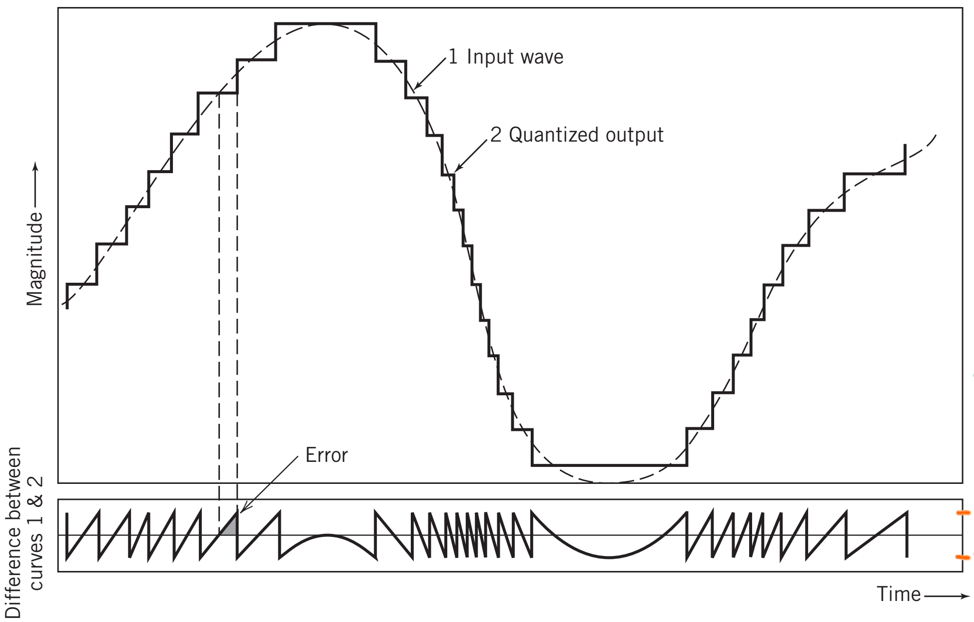



Figure 73: Error in the quantized output [4, Fig. 6.10].

$\Delta/2$

$-\Delta/2$

$-\Delta/2 \quad \Delta/2$

$X \sim \mathcal{U}(a, b)$

$\mathbb{E}X = \frac{a+b}{2} = 0$

$\text{Var } X = \frac{(b-a)^2}{12} = \frac{\Delta^2}{12}$

$= \mathbb{E}[X^2] - \cancel{(\mathbb{E}X)^2}$

Let $q[n] = m[n] - m_n$ be the quantization error for the $n$th sample. If the step size $\Delta$ is sufficiently small (i.e., $L$ is sufficiently large) and $m[n]$ is bounded by $(-m_p, m_p)$, it is reasonable to assume that the quantization error is uniformly distributed on $\left[-\frac{\Delta}{2}, \frac{\Delta}{2}\right]$.

Assuming that the sampling is done with sampling rate $f_s$ that is fast enough ($> 2B$). Then, we know, from the reconstruction equation (77), that

$$m(t) = \sum_{n=-\infty}^{\infty} m[n] \operatorname{sinc}(\pi f_s (t - nT_s)). \tag{81}$$

With quantization, the sequence $m_n$ is transmitted instead of $m[n]$. Hence, at the receiver, the reconstructed signal is

$$\hat{m}(t) = \sum_{n=-\infty}^{\infty} m_n \operatorname{sinc}(\pi f_s (t - nT_s)). \tag{82}$$

The distortion component in the reconstructed signal, which is referred to as the **quantization noise**, is

$$q(t) = m(t) - \hat{m}(t).$$

$\Delta = \frac{2m_p}{L}$

One can then show that the average power of the quantization noise is

$$P_q = \mathbb{E}\left[(m[n] - m_n)^2\right] = \mathbb{E}\left[q^2[n]\right] = \frac{\Delta^2}{12} = \frac{m_p^2}{3L^2}. \tag{83}$$

So, the signal-to-(quantization)-noise power ratio (SNR or SQNR) is

$$\mathrm{SNR} = \frac{P_m}{P_q} = \frac{12P_m}{\Delta^2} = \frac{3L^2P_m}{m_p^2}. \tag{84}$$

This is an indication of the quality of the received signal. We want SNR to be large.

**8.11. Sinusoidal Modulating Signal**: For sinusoidal signal $m(t)$ with amplitude $A$, its power is $P_m = \frac{1}{2}A^2$ and therefore,

$$\mathrm{SNR} = \frac{12P_m}{\Delta^2} = \frac{12\left(\frac{1}{2}A^2\right)}{\Delta^2} = \frac{6A^2}{\Delta^2} = 10\log_{10}\left(\frac{6A^2}{\Delta^2}\right)\ [\mathrm{dB}]. \tag{85}$$

**8.12. Full-Load Sinusoidal Modulating Signal**: If the sinusoidal $m(t)$ fully covers the whole range of the quantizer, we have $A = m_p$, $P_m = \frac{1}{2}m_p^2$ and

$$\mathrm{SNR} = \frac{3L^2P_m}{m_p^2} = \frac{3L^2\left(\frac{1}{2}m_p^2\right)}{m_p^2} = 1.5L^2 = 10\log_{10}\left(1.5L^2\right)\ [\mathrm{dB}]. \tag{86}$$

**8.13.** As mentioned earlier, in PCM, the quantized samples are coded and transmitted as binary pulses. At the receiver, some pulses may be detected incorrectly. Hence, there are two sources of error in this scheme:

(a) quantization error

(b) pulse detection error

In almost all practical schemes, the pulse detection error is quite small compared to the quantization error and can be ignored. [6, p 322]

**8.14.** The quantization error can be reduced as much as desired by in creasing the number of quantizing levels, the price of which is paid in an increased required transmission bandwidth. See (88).

## 8.2 Pulse Coding

**8.15.** From practical viewpoint, a **binary** digital signal (a signal that can take on only two values) is very desirable because of its simplicity, economy, and ease of engineering. We can convert an $L$-ary signal into a binary signal by using **pulse coding**.

- A **bi**nary dig**it** is called a bit.
- $L = 2^\ell$ levels can be mapped into (represented by) $\ell$ bits. $\log_2 L$

**Example 8.16.** In Figure 71, $L = 8$. The binary code can be formed by the binary representation of the 8 decimal digits from 0 to 7; that is we assign the word 000 to the lowest level and progresses upward to 111 in the natural order of binary counting. This "natural" code is also known as "offset binary code".

**Example 8.17.** Telephone (speech) signal:

- The components above 3.4 kHz are eliminated by a low-pass filter.
  - For speech, subjective tests show that signal intelligibility is not affected if all the components above 3.4 kHz are suppressed.
- The resulting signal is then sampled at a rate of 8,000 samples per second (8 kHz).
  - This rate is intentionally kept higher than the Nyquist sampling rate of 6.8 kHz so that realizable filters can be applied for signal reconstruction.
- Each sample is finally quantized into 256 levels ($L = 256$), which requires eight bits to encode each sample ($2^8 = 256$).

$$R_b = 8 \frac{\text{bits}}{\text{sample}} \times 8000\ \text{sa}/\text{s} = 64\ \text{kbps}$$

[6, p 320]

**Example 8.18.** Compact disc (CD) audio signal:

- High-fidelity: Require the audio signal bandwidth to be 20 kHz.
- The sampling rate of 44.1 kHz is used.
- The signal is quantized into $L = 65,536$ of quantization levels, each of which is represented by 16 bits (16-bit two's complement integer) to reduce the quantizing error.

$\log_2$

$\ell$

left/right ch

$2 \times 44.1\ \mathrm{k}\,\frac{\mathrm{Sa}}{\mathrm{s}} \times 16\ \frac{\mathrm{bits}}{\mathrm{Sa}} \times 80\ \frac{\mathrm{min}}{\mathrm{CD}} \times 60\ \frac{\mathrm{s}}{\mathrm{min}} \times \frac{1}{8}\ \frac{\mathrm{B}}{\mathrm{bits}} \approx 800\ \frac{\mathrm{MB}}{\mathrm{CD}}$

1.4112 Mbps (audio bit rate)

[6, p 321]

**8.19.** The SNR for the full-load sinusoidal modulating signal discussed in 8.12 is then

$L^2 = (2^\ell)^2$

$$\mathrm{SNR} = 1.5\left(2^{2\ell}\right) = 10\log_{10}1.5 + 2\ell\underbrace{10\log_{10}2}_{\approx 3}\ [\mathrm{dB}] \approx 1.76 + 6\ell\ \ [\mathrm{dB}]. \qquad (87)$$

**8.20.** Recall, from 6.37, that a maximum of $2B$ independent pieces (symbols) of information per second can be transmitted, errorfree, over a noiseless channel of bandwidth $B$ Hz. In other words, one can send at most 2 [Sa/s/Hz]. This is achieved by using the sinc pulse train. Here, because we use binary coding, one symbol is the same as one bit. Therefore, for PCM, one can send at most 2 [b/s/Hz]. Equivalently, if the bit rate is $R_b$ [bps], the minimum required baseband transmission bandwidth is

$$\frac{R_b}{2} = \frac{f_s\ell}{2} \geq \frac{2B\ell}{2} = \ell B. \qquad (88)$$

## 8.3 Line Coding

**8.21.** The last signal-processing operation in the transmitter is that of line coding, the purpose of which is to represent/convert sequence of bits by/into a sequence of (electrical) pulses.

**8.22.** Figure 74 depicts various **line codes** for the binary message 10110100, taking rectangular pulses for clarity.

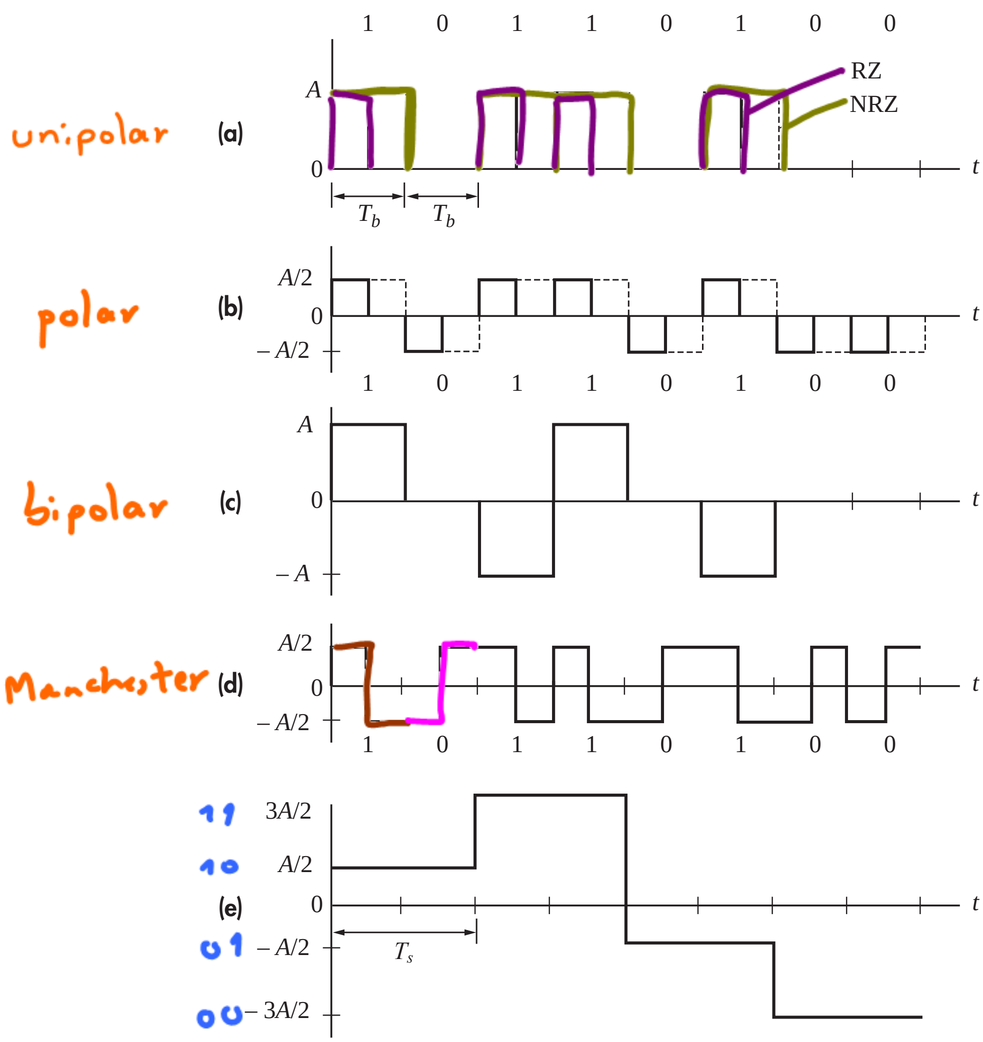



Figure 74: Line codes with rectangular pulses: (a) unipolar RZ and NRZ; (b) polar RZ and NRZ; (c) bipolar NRZ; (d) split-phase Manchester; (e) polar quaternary NRZ. [3, Fig 11.1-1 p 483]

**Definition 8.23.** The simple **on-off waveform** in Figure 74a represents each 0 by an "off" pulse and each 1 by an "on" pulse.

(a) In the (unipolar) **return-to-zero** (**RZ**) format, the pulse duration is smaller than $T_b$ after which the signal return to the zero level.

(b) A (unipolar) **nonreturn-to-zero** (**NRZ**) format has "on" pulses for full bit duration $T_b$.

**Definition 8.24.** The **polar signal** in Figure 74b has opposite polarity pulses

- Its DC component will be zero if the message contains 1s and 0s in equal proportion.

**Definition 8.25.** Figure 74c, we have **bipolar signal** where successive 1s are represented by pulses with alternating polarity.

- Use three amplitude levels
- Also known as pseudo-ternary or alternate mark inversion (AMI)

**Definition 8.26.** The split-phase or **Manchester** format in Figure 74d represents 1s with a positive half-interval pulse followed by a negative half-interval pulse, and vice versa for the representation of 0s.

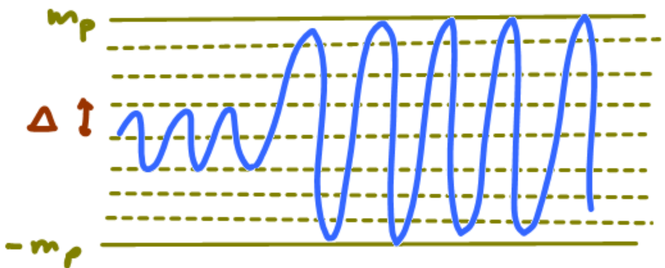
- Also called twinned binary.
- Guarantee zero DC component regardless of the message sequence.

**Definition 8.27.** Figure 74e shows a **quaternary signal** derived by grouping the message bits in blocks of two and using four amplitude levels to prepresent the four possible combinations 00, 01, 10, and 11.

- Quaternary coding can be generalized to **M-ary coding** in which blocks of $n$ message bits are represented by an M-level waveform with $M = 2^n$.

## 8.4 Companding: Nonuniform Quantization

**8.28.** Recall, from (84) that SNR $= \frac{3L^2 P_m}{m_p^2}$. So, the SNR is directly proportional to the signal power which implies that the quality of the received signal will deteriorate markedly when the person speaks softly. Statistically, it is found that smaller amplitudes predominate in speech and larger amplitudes are much less frequent. This means the SNR will be low most of the time.




**8.29.** The problem can be solved by using smaller steps for smaller amplitudes (nonuniform quantizing). In other words,

(a) the weak (soft) passages needing more protection are favored at the expense of the loud passages,

(b) the loud talkers and stronger signals are penalized with higher step sizes to compensate the soft talkers and weaker signals.

This is illustrated in Figure 75.

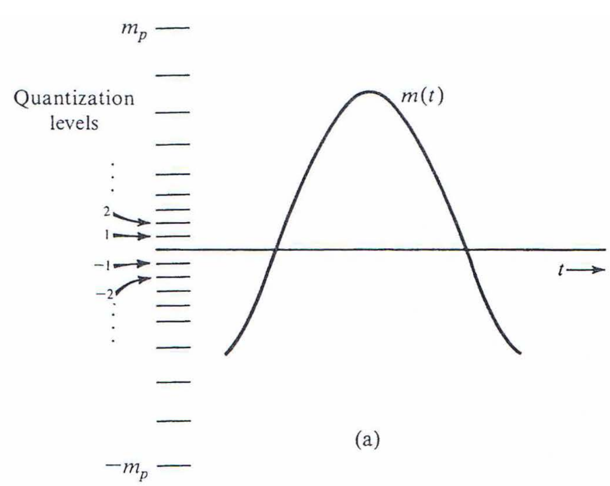



Figure 75: Nonuniform quantization. [6, Fig 6.15a p. 326]

**8.30.** The same result is obtained by first compressing (with a compressor) signal samples and then using a uniform quantization.

$$m(t) \rightarrow \boxed{\text{Compressor}} \rightarrow y(t) \rightarrow \boxed{\begin{array}{c}\text{Uniform}\\ \text{Quantizer}\end{array}} \rightarrow \begin{array}{c}\text{Compressed}\\ \text{Output Signal}\end{array}$$

An example of the input-output characteristic of the compressor is illustrated in Figure 76.

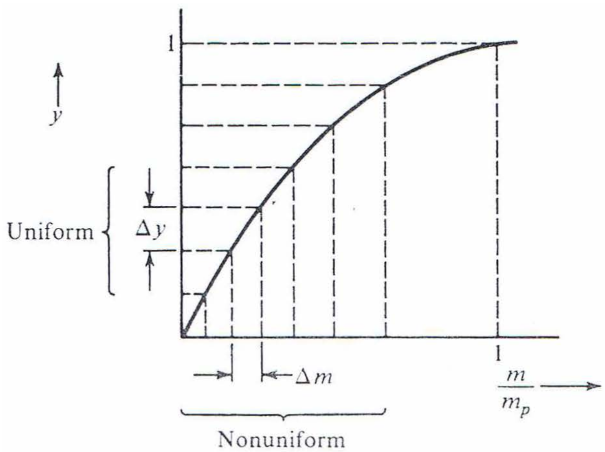



Figure 76: An example of a compressor characteristic curve [6, Fig 6.15b p. 326]

**8.31.** Among several choices, two compression laws have been accepted as desirable standards by the ITU-T:

(a) the $\mu$-law used in North America and Japan

(b) the $A$-law used in Europe and the rest of the world and on international routes

Example of their characteristic curves are plotted in Figure 77.

**Example 8.32.** In the standard audio file format used by Sun, Unix and Java, the audio in "au" files can be pulse-code-modulated or compressed with the ITU-T G.711 standard through either the $\mu$-law or the $A$-law. In both cases, the sampling is performed at 8000 [Sa/s] and the compressor converts the linear PCM samples to 8-bit samples. Therefore, the resulting audio bit stream is at

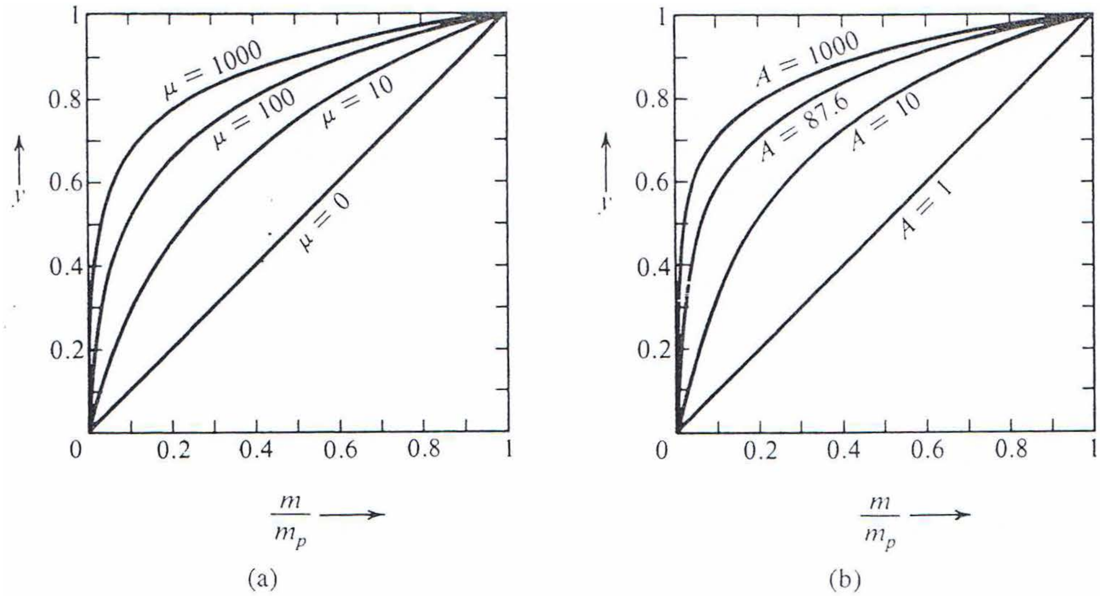



Figure 77: (a) $\mu$-law characteristic. (b) $A$-law characteristic. [6, Fig 6.16 p. 326]

**Example 8.33.** Microsoft WAV audio format also has compression options that use $\mu$-law and $A$-law.

**8.34.** At the receiver, to restore the signal samples to their correct relative level, we must, of course, use a device in the receiver with a characteristic complementary to the compressor. Such a device is called an **expander**. Ideally, the compression and expansion laws are exactly the inverse of each other. The compressor and the expander together are called the **compandor**.

5 7 10 9 8

5 2 3 -1 -1

## 8.5 DPCM and DM

**8.35.** PCM is not a very efficient system because it generates so many bits and requires so much bandwidth to transmit. DPCM and DM are two other useful forms of digital pulse modulation.

**8.36.** In **differential pulse-code modulation (DPCM)**, the main idea is that instead of transmitting the sample values, we transmit the difference between the successive sample values.

- The difference between successive samples is generally much smaller than the sample values. Thus, the peak amplitude of the transmitted values is reduced considerably. Therefore, our quantizer can consider smaller range (smaller $m_p$).
- For a given $L$, from (84), we see that the SNR is improved.
- Alternatively, for a given SNR, we can reduce $L$ which in turn reduce $\ell$ (or transmission bandwidth).
- In general, we try to predict (estimate) $m[n]$ from several previous sample values and transmit the difference (prediction error): $d[n] = m[n] - \hat{m}[n]$

**8.37.** Figure 78 illustrates the tradeoffs between standard PCM, DPCM, and DM.

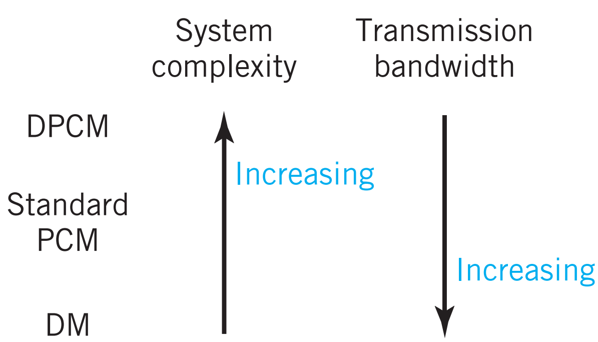



Figure 78: Tradeoffs between PCM, DPCM, and DM [4, Fig. 305]

**8.38.** In **delta modulation (DM)**,

- $m(t)$ is oversampled (typically using four times the Nyquist rate [6, p 346]), and
- system complexity is reduced to the minimum possible by using a single-bit quantizer with only two levels: $\pm\Delta$.
  - DM is basically a 1-bit DPCM

**8.39.** Two types of error in DM:

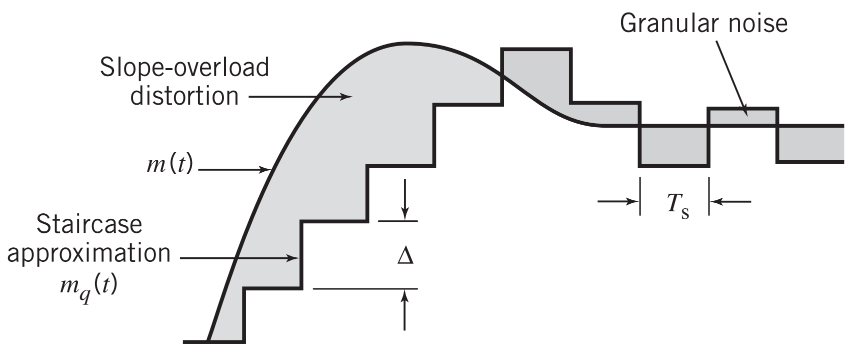



Figure 79: Two different forms of error in DM

(a) Slope overload distortion

- Occur when $m_q(t)$ cannot follow a steep segment of $m(t)$.
- During the sampling interval $T_s$, $m_q(t)$ is capable of changing by the step size $\Delta$. Hence, the maximum slope that $m_q(t)$ can follow is $\Delta/T_s = \Delta f_s$. Hence, no overload occurs if

$$\max|\dot{m}(t)| < \Delta f_s.$$

- The slope overload noise can be reduced by increasing $\Delta$ (the step size). This unfortunately increases the granular noise.

(b) Granular noise

- Occur when the step size $\Delta$ is too large relative to the local slope characteristics of the message signal $m(t)$, thereby causing the staircase approximation $m_q(t)$ to hunt around a relatively flat segment of $m(t)$.