Chapter 14

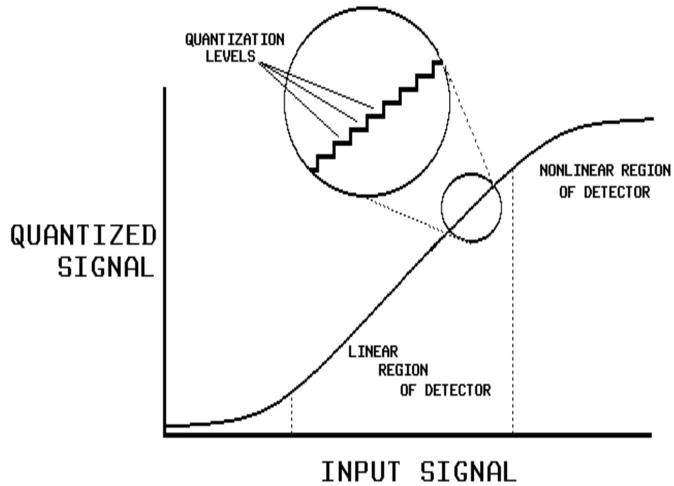
Review of Quantization

14.1 Tone-Transfer Curve

The second operation of the digitization process converts the continuously valued irradiance of each sample at the detector (i.e., the brightness) to an integer, i.e., the sampled image is quantized. The entire process of measuring and quantizing the brightnesses is significantly affected by detector characteristics such as dynamic range and linearity. The dynamic range of a detector image is the range of brightness (irradiance) over which a change in the input signal produces a detectable change in the output. The input and output quantities need not be identical; the input may be measured in $\frac{\text{W}}{\text{mm}^2}$ and the output in optical density. The effect of the detector on the measurement may be described by a transfer characteristic or tone-transfer curve (TTC), i.e., a plot of the output vs. input for the detector. The shape of the transfer characteristic may be used as a figure of merit for the measurement process. A detector is linear if the TTC is a straight line, i.e., if an incremental change in input from any level produces a fixed incremental change in the output. Of course, all real detectors have a limited dynamic range, i.e., they will not respond at all to light intensity below some minimum value and their response will not change for intensities above some maximum. All realistic detectors are therefore nonlinear, but there may be some regions over which they are more-or-less linear, with nonlinear regions at either end. A common such example is photographic film; the TTC is the *H-D* curve which plots recorded optical density of the emulsion vs. the logarithm of the input irradiance $[\frac{\text{W}}{\text{mm}^2}]$. Another very important example in digital imaging is the video camera, whose TTC maps input light intensity to output voltage. The transfer characteristic of a video camera is approximately a power law:

$$V_{out} = c_1 B_{in}^{\gamma} + V_0$$

where V_0 is the threshold voltage for a dark input and γ (gamma) is the exponent of the power law. The value of γ depends on the specific detector: typical values are $\gamma \cong 1.7$ for a vidicon camera and $\gamma \cong 1$ for an image orthicon.



Nonlinear tone-transfer curve of quantizer, showing a linear region.

14.2 Quantization

Quantization converts continuously valued measured irradiance at a sample to a member of a discrete set of gray levels or digital counts, e.g.,the sample $f[x, y]$ e.g., $f[0, 0] = 1.234567890 \cdots \frac{\mathrm{W}}{\mathrm{mm}^2}$, is converted to an integer between 0 and some maximum value (e.g., 255) by an analog-to-digital conversion (A/D converter or ADC). The number of levels is determined by number of bits available for quantization in the ADC. A quantizer with m bits defines $M = 2^m$ levels. The most common quantizers have $m = 8$ *bits* (one byte); such systems can specify 256 different gray levels (usually numbered from $[0, 255]$, where 0 is usually assigned to "black" and 255 to "white". Images digitized to 12 or even 16 bits are becoming more common, and have 4096 and 65536 levels, respectively.

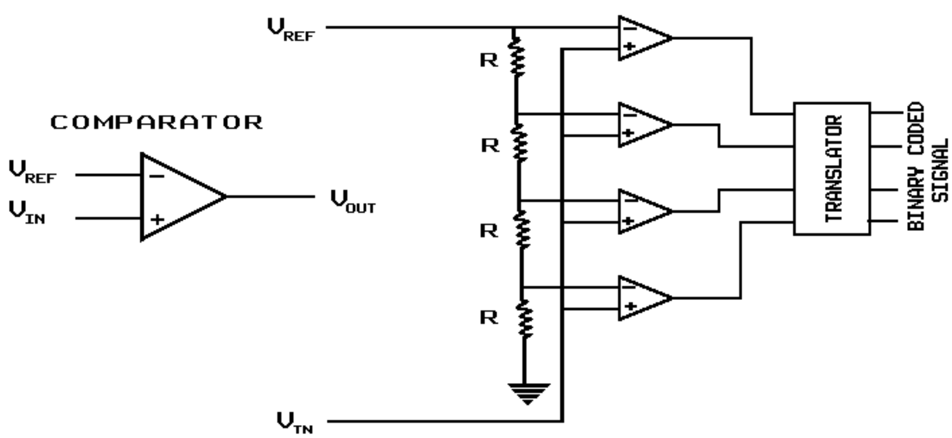
The resolution, or step size b, of the quantizer is the difference in brightness between adjacent gray levels. It makes little sense to quantize with a resolution b which is less than the uncertainty in gray level due to noise in the detector system. Thus the effective number of levels is often less than the maximum possible.

Conversion from a continuous range to discrete levels requires a thresholding operation (e.g.,truncation or rounding). Some range of input brightnesses will map to a single output level, e.g., all measured irradiances between 0.76 and $0.77 \frac{\mathrm{W}}{\mathrm{mm}^2}$ might map to gray level 59. Threshold conversion is a nonlinear operation, i.e., the threshold of a sum of two inputs is not necessarily the sum of the thresholded outputs. The concept of linear operators will be discussed extensively later, but we should say at this point that the nonlinearity due to quantization makes it inappropriate to analyze the complete digital imaging system (digitizer, processor, and display) by common linear methods. This problem is usually ignored, as is appropriate for large numbers of quantized levels that are closely spaced so that the digitized image appears continuous. Because the brightness resolution of the eye-brain is limited, quantizing to only 50 levels is satisfactory for many images; in other words, 6bits of data is often sufficient for images to be viewed by humans.

The quantization operation is performed by digital comparators or sample-and-hold circuits. The simplest quantizer converts an analog input voltage to a 1-bit digital output and can be constructed from an ideal differential amplifier, where the output voltage V_{out} is proportional to the difference of two voltages V_{in} and V_{ref}:

$$V_{out} = \alpha(V_{in} - V_{ref})$$

V_{ref} is a reference voltage provided by a known source. If α is large enought to approximate ∞, then the output voltage will be $+\infty$ if $V_{in} > V_{ref}$ and $-\infty$ if $V_{in} < V_{ref}$. We assign the digital value "1" to a positive output and "0" to a negative output. A quantizer with better resolution can be constructed by cascading several such digital comparators with equally spaced reference voltages. A digital translator converts the comparator signals to the binary code. A 2-bit ADC is shown in the figure:



Comparator and 2-Bit ADC. The comparator is a "thresholder;" its output is "high" if $V_{in} > V_{ref}$ and "low" otherwise. The ADC consists of 4 comparators whose reference voltages are set at different values by the resistor-ladder voltage divider. The translator converts the 4 thresholded levels to a binary-coded signal.

In most systems, the step size between adjacent quantized levels is fixed ("uniform quantization"):

$$b = \frac{f_{\max} - f_{\min}}{2^m - 1}$$

where f_{max} and f_{min} are the extrema of the measured irradiances of the image samples and m is the number of bits of the quantizer.

If the darkest and brightest samples of a continuous-tone image have measured irradiances f_{min} and f_{max} respectively, and the image is to be quantized using m bits (2^m graylevels), then we may define a set of uniformly spaced levels $f_q$that span the dynamic range via:

$$f_q[x,y] = \mathcal{Q}\left\{\frac{f[x,y]-f_{\min}}{b}\right\} = \mathcal{Q}\left\{\frac{f[x,y]-f_{\min}}{f_{\max}-f_{\min}}\cdot 2^m - 1\right\}$$

where $\mathcal{Q}\{\ \}$ represents the nonlinear truncation or rounding operation, e.g., $\mathcal{Q}\{3.657\} = 3$ if $\mathcal{Q}$ is truncation or 4 if $\mathcal{Q}$ is rounding. The form of $\mathcal{Q}$ determines the location of the decision levels where the quantizer jumps from one level to the next. The image irradiances are reconstructed by assigning all pixels with a particular gray level f_q to the same irradiance value $E[x,y]$, which might be defined by "inverting" the quantization relation. The reconstruction level is often placed between the decision levels by adding a factor $\frac{b}{2}$:

$$\hat{E}[x,y] = \left(f_q[x,y]\cdot\frac{E_{\max}-E_{\min}}{2^m-1}\right) + E_{\min} + \frac{b}{2}$$

Usually (of course), $\hat{E}[x,y] \neq E[x,y]$ due to the quantization, i.e., there will be *quantization error*. The goal of optimum quantization is to adjust the quantization scheme to reconstruct the set of image irradiances which most closely approximates the ensemble of original values. The criterion which defines the goodness of fit and the statistics of the original irradiances will determine the parameters of the quantizer, e.g., the set of thresholds between the levels.

The quantizer just described is *memoryless*, i.e., the quantization level for a pixel is computed independently that for any other pixel. The schematic of a memoryless quantizer is shown below. As will be discussed, a quantizer with memory may have significant advantages.

14.3 Quantization Error ("Noise")

The gray value of the quantized image is an integer value which is related to the input irradiance at that sample. For uniform quantization, where the steps between adjacent levels are the same size, the constant of proportionality is the difference in irradiance between adjacent quantized levels. The difference between the true input irradiance (or brightness) and the corresponding irradiance of the digital level is the quantization error at that pixel:

$$\epsilon[n\cdot\Delta x, m\cdot\Delta y] \equiv f[n\cdot\Delta x, m\cdot\Delta y] - f_q[n\cdot\Delta x, m\cdot\Delta y].$$

Note that the quantization error is bipolar in general, i.e., it may take on positive or negative values. It often is useful to describe the statistical properties of the quantization error, which will be a function of both the type of quantizer and the input image. However, if the difference between quantization steps (i.e., the width of a quantization level) is b, is constant, the quantization error for most images may be approximated as a uniform distribution with mean value $\langle\epsilon[n]\rangle = 0$ and variance $\langle(\epsilon_1[n])^2\rangle = \frac{b^2}{12}$. The error distribution will be demonstrated for two 1-D 256-sample images. The first is a section of a cosine sampled at 256 points and quantized to 64 levels separated by $b = 1$:

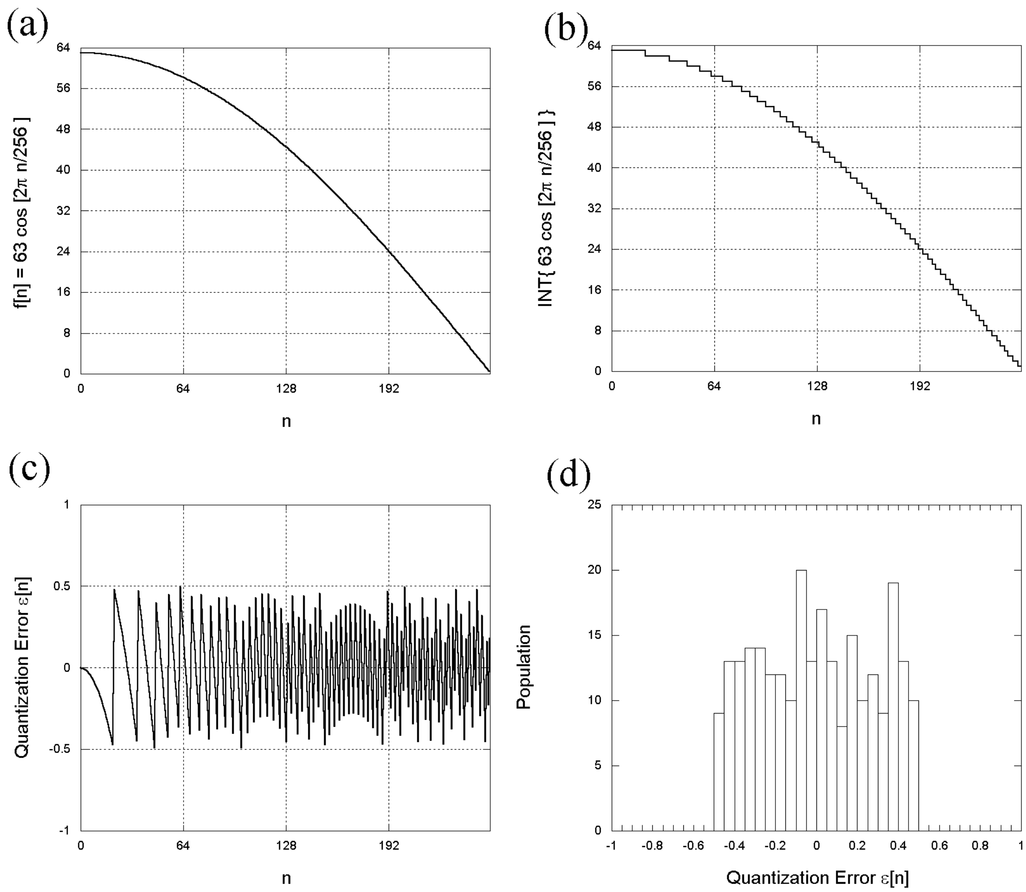


Illustration of the statistics of quantization noise: (a) $f[n] = 63\cos\left[2\pi\frac{n}{256}\right]$ *for* $0 \leq n \leq 255$*; (b) after quantization by rounding to nearest integer; (c) quantization error* $\varepsilon[n] \equiv f[n] - f_q[n]$*, showing that* $-\frac{1}{2} \leq \varepsilon \leq +\frac{1}{2}$*; (d) histogram of 256 samples of quantization error, showing that the statistics are approximately uniform.*

The histogram of the error $\epsilon_1[n] = f_1[n] - Q\{f_1[n]\}$ is approximately uniform over the interval $-\frac{1}{2} \leq \epsilon_1 < +\frac{1}{2}$. The computed statistics of the error are $\langle\epsilon_1[n]\rangle = -5.1 \cdot 10^{-4} \cong 0$ and variance is $\langle\epsilon_1^2[n]\rangle = 0.08 \cong \frac{1}{12}$.

The second image is comprised of 256 samples of Gaussian-distributed random noise in the interval $[0, 63]$ that again is quantized to 64 levels. The histogram of the error $\epsilon_2[n]$ again is approximately uniformly distributed in the interval $[-0.5, +0.5]$ with mean $4.09 \cdot 10^{-2} \cong 0$ and variance $\sigma^2 = \langle\epsilon_2^2[n]\rangle \cong 0.09 \cong \frac{1}{12}$.

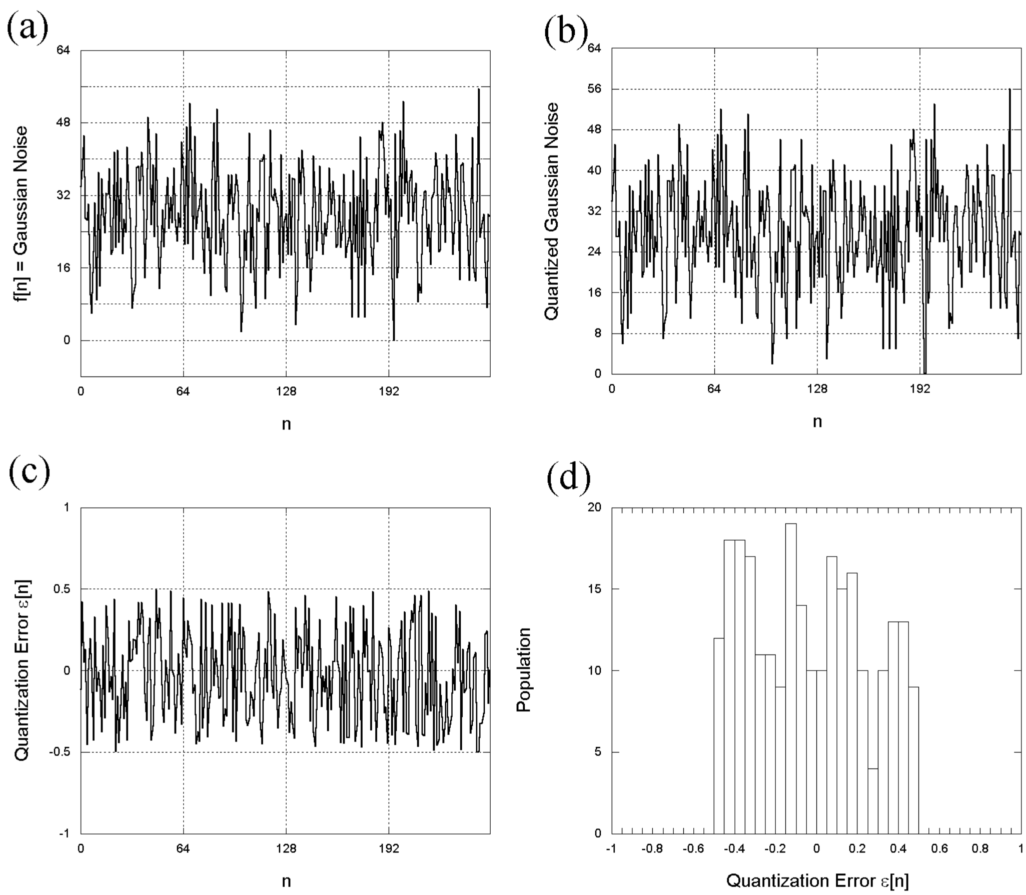


Illustration of the statistics of quantization noise: (a) $f[n]$ is Gaussian noise with measured $\mu = 27.7$, $\sigma = 10.9$ for $0 \leq n \leq 255$; (b) after quantization by rounding to nearest integer; (c) quantization error $\varepsilon[n] \equiv f[n] - f_q[n]$, showing that $-\frac{1}{2} \leq \varepsilon \leq +\frac{1}{2}$; (d) histogram of 256 samples of quantization error, showing that the statistics are STILL approximately uniform.

The total quantization error is the sum of the quantization error over all pixels in the image:

$$\epsilon = \sum_i \sum_j \epsilon\,[n \cdot \Delta x, m \cdot \Delta y]\,.$$

An image with large bipolar error values thus *may* have a small total error. The mean-squared error (average of the squared error) is a better descriptor of the fidelity of the quantization:

$$\epsilon^2 = \frac{1}{N} \sum_i \sum_j \left(\epsilon^2\,[n \cdot \Delta x, m \cdot \Delta y]\right),$$

where N is the number pixels in the image. If the irradiance is measured in $\frac{\text{W}}{\text{mm}^2}$, ϵ^2 will have units of $\left(\frac{\text{W}}{\text{mm}^2}\right)^2$. The root-mean-squared (RMS) error has the same dimensions as the error:

$$\text{RMS Error} \equiv \sqrt{\epsilon^2} = \sqrt{\frac{1}{N}\sum_i \sum_j \epsilon^2 \left[n \cdot \Delta x, m \cdot \Delta y\right]}.$$

It should be obvious that the RMS error for one image is a function of the quantizer used, and that the RMS error from one quantizer will differ for different images. It should also be obvious that it is desirable to minimize the RMS error in an image. The brute-force method for minimizing quantization error is to add more bits to the ADC, which increases the cost of the quantizer and the memory required to store the image.

We now extend the discussion to consider the concepts of signal bandwidth and digital data rate, which in turn require an understanding of signal-to-noise ratio (*SNR*) and its relationship to quantization. Recall that the variance σ^2 of a signal is a measure of the spread of its amplitude about the mean value.

$$\begin{aligned}\sigma_f^2 &= \int_{-\infty}^{+\infty} \left[f\left[x\right] - \langle f\left[x\right]\rangle\right]^2 dx \\ &\Longrightarrow \frac{1}{X_0}\int_{-\frac{X_0}{2}}^{+\frac{X_0}{2}} \left[f\left[x\right] - \langle f\left[x\right]\rangle\right]^2 dx\end{aligned}$$

The signal-to-noise power ratio of an analog signal is most rigorously defined as the dimensionless ratio of the variances of the signal and noise:

$$\boxed{SNR \equiv \frac{\sigma_f^2}{\sigma_n^2}}$$

Thus a large *SNR* means that there is a larger variation of the signal amplitude than of the noise amplitude. This definition of *SNR* as the ratio of variances may vary over a large range – easily several orders of magnitude – so that the numerical values may become unwieldy. The range of *SNR* may be compressed by expressing it on a logarithmic scale with dimensionless units of *bels*:

$$SNR = \log_{10}\left[\frac{\sigma_f^2}{\sigma_n^2}\right] = 2\,\log_{10}\left[\frac{\sigma_f}{\sigma_n}\right] \quad [bels]$$

This definition of *SNR* is even more commonly expressed in units of tenths of a bel so that the integer value is more precise. The resulting metric is in terms of *decibels*:

$$SNR = 10\ \log_{10}\left[\frac{\sigma_f^2}{\sigma_n^2}\right] = 10\ \log_{10}\left[\left(\frac{\sigma_f}{\sigma_n}\right)^2\right]$$
$$= 20\ \log_{10}\left[\frac{\sigma_f}{\sigma_n}\right]\quad [decibels]$$

Under this definition,$SNR = 10\ dB$ if the signal variance is ten times larger than the noise variance and 20 dB if the standard deviation is ten times larger than that of the noise.

The variances obviously depend on the statistics (the histograms) of the signal and noise. The variances depend only on the range of gray values and not on their "arrangement" (i.e., numerical "order" or "pictorial" appearance in the image. Since the noise often is determined by the measurement equipment, a single measurement of the noise variance often is used for many signal amplitudes. However, the signal variance must be measured each time. Consider the variances of some common 1-D signals.

14.3.1 Example: Variance of a Sinusoid

The variance of a sinusoid with amplitude A_0 is easily computed by direct integration:

$$f[x] = A_0 \cos\left[2\pi\frac{x}{X_0}\right]$$
$$\sigma_f^2 = \frac{1}{X_0}\int_{-\frac{X_0}{2}}^{+\frac{X_0}{2}} (f[x] - \langle f[x]\rangle)^2\, dx = \frac{1}{X_0}\int_{-\frac{X_0}{2}}^{+\frac{X_0}{2}} \left(A_0\cos\left[2\pi\frac{x}{X_0}\right]\right)^2 dx$$
$$= \frac{A_0^2}{X_0}\int_{-\frac{X_0}{2}}^{+\frac{X_0}{2}} \frac{1}{2}\left(1+\cos\left[4\pi\frac{x}{X_0}\right]\right) dx = \frac{A_0^2}{2X_0}(X_0+0)$$
$$= \boxed{\sigma_f^2 = \frac{A_0^2}{2}\ \textit{for sinusoid with amplitude}\ A_0}$$

Note that the variance does not depend on the period (i.e., on the spatial frequency) or on the initial phase – it is a function of the histogram of the values in a period and not of the "ordered" values. It also does not depend on any "bias" (additive constant) in the signal. The standard deviation of the sinusoid is just the square root of the variance:

$$\sigma_f = \frac{A_0}{\sqrt{2}}\ \textit{for sinusoid with amplitude}\ A_0$$

14.3.2 Example: Variance of a Square Wave:

The variance of a square wave with the same amplitude also is easily evaluated by integration of the thresholded sinusoid:

$$f[x] = A_0 \; SGN\left[\cos\left[2\pi\frac{x}{X_0}\right]\right]$$

$$\sigma_f^2 = \frac{1}{X_0}\int_{-\frac{X_0}{2}}^{+\frac{X_0}{2}} [f[x] - \langle f[x]\rangle]^2 \, dx = \frac{1}{X_0}\left(\int_{-\frac{X_0}{4}}^{+\frac{X_0}{4}} [-A_0]^2 \, dx + \int_{+\frac{X_0}{4}}^{+\frac{3X_0}{4}} [+A_0]^2 \, dx\right)$$

$$= \frac{1}{X_0}\left(A_0^2\frac{X_0}{2} + A_0^2\frac{X_0}{2}\right) = A_0^2$$

$$\boxed{\sigma_f^2 = A_0^2 \textit{ for square wave with amplitude } A_0}$$

$$\boxed{\sigma_f = A_0 \textit{ for square wave with amplitude } A_0}$$

Note that the variance of the square wave is larger than that of the sine wave with the same amplitude:

$$\boxed{\sigma_f \textit{ for square wave with amplitude } A_0 > \sigma_f \textit{ for sinusoid with amplitude } A_0}$$

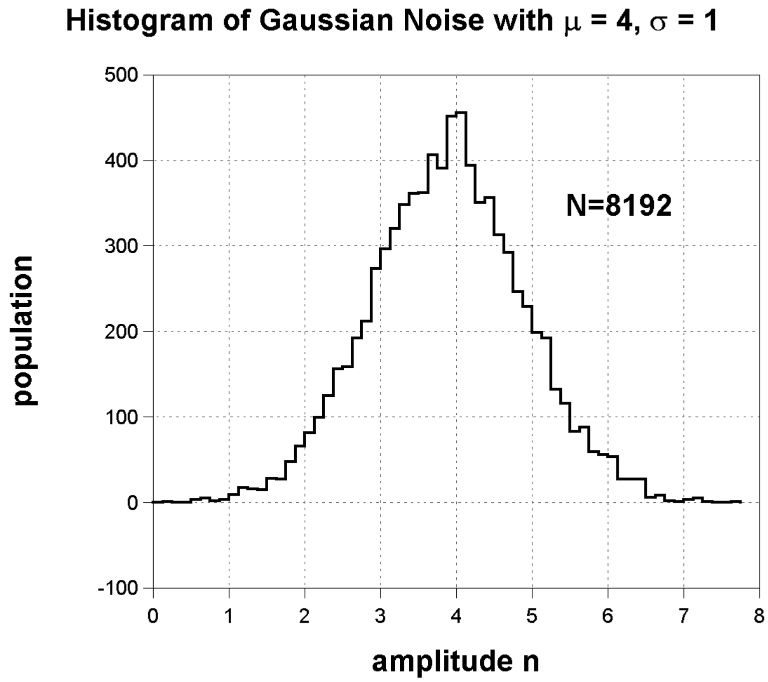
which makes intuitive sense, because the amplitude of the square wave is more often "distant" from its mean than the sinusoid is.

14.3.3 Variance of "Noise" from a Gaussian Distribution

A set of amplitudes selected at random from a Gaussain probability distribution is called (conveniently enough) "Gaussian noise." The most common definition of the statistical distribution is:

$$p[n] = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

This probability distribution function has unit area, as required. The Gaussian distribution is specified by the two parameters μ, the mean value of the distribution, and σ^2, its variance. The standard deviation σ is a measure of the "width" of the distribution and so influences the range of output amplitudes.



Histogram of 8192 samples taken from the Gaussian distribution
$p[n] = \frac{1}{\sqrt{2\pi}} \exp\left[-\left(\frac{n-4}{2}\right)^2\right]$

14.3.4 Approximations to *SNR*

Since the variance depends on the statistics of the signal, it is common (though less rigorous) to approximate the variance by the square of the *dynamic range*, which is the "peak-to-peak signal amplitude" $f_{\max} - f_{\min} \equiv \Delta f$. In most cases, $(\Delta f)^2$ is larger (and often much larger) than σ_f^2. In the examples of the sinusoid and the square wave already considered, the approximations are:

$$\textit{Sinusoid with amplitude } A_0 \Longrightarrow \sigma_f^2 = \frac{A_0^2}{2},\ \ (\Delta f)^2 = (2A_0)^2 = 4A_0^2 = 8\ \sigma_f^2$$
$$\textit{Square wave with amplitude } A_0 \Longrightarrow \sigma_f^2 = A_0^2,\ \ (\Delta f)^2 = (2A_0)^2 = 4A_0^2 = 4\ \sigma_f^2$$

For the example of Gaussian noise with variance $\sigma^2 = 1$ and mean μ, the dynamic range Δf of the noise technically is infinite, but its extrema often be approximated based on the observation that few amplitudes exist outside of four standard deviations, so that $f_{\max} \cong \mu+4\sigma$, $f_{\min} \cong \mu-4\sigma$, leading to $\Delta f \cong 8\sigma$. The estimate of the variance of the signal is then $(\Delta f)^2 \cong 64\sigma_f^2$, which is (obviously) 64 times larger than the actual variance. Because this estimate of the signal variance is too large, the estimates of the *SNR* thus obtained will be too optimistic.

Often, the signal and noise of images are measured by photoelectric detectors as differences in electrical potential in volts; the signal dynamic range is $V_f = V_{\max} - V_{\min}$, the average noise voltage is V_n, and the signal-to-noise ratio is:

$$SNR = 10\ \log_{10}\left(\frac{V_f^2}{V_n^2}\right) = 20\ \log_{10}\left(\frac{V_f}{V}\right) \quad [dB]$$

As an aside, we mention that the *signal amplitude* (or *level*) of analog electrical signals often is described in terms of dB measured relative to some fixed reference. If the reference level is 1 Volt, the signal level is measured in units of dBV:

$$level = 10\ \log_{10}\left(V_f^2\right)\ dBV = 20\ \log_{10}\left(V_f\right) \quad dBV$$

The level is measured relative to 1 mV is in units of dBm:

$$level = 10\ \log_{10}\left(\frac{V_f^2}{10^{-3}V^2}\right)\ dBV = 10\ \log_{10}\left(\frac{V_f^2}{V^2}\right) \quad dBm$$

14.3.5 *SNR* of Quantization

We can use these definitions to evaluate the signal-to-noise ratio of the quantization process. Though the input signal and the type of quantizer determine the probability density function of the quantization error in a strict sense, the quantization error for the two examples of quantized sinusoidal and Gaussian-distributed signals both exhibited quantization errors that were approximately uniformly distributed. We will continue this assumption that the probability density function is a rectangle. In the case of an m-bit uniform quantizer (2^m gray levels) where the levels are spaced by intervals of width b over the full analog dynamic range of the signal, the error due to quantization will be (approximately) uniformly distributed over this interval b. If the nonlinearity of the quantizer is *rounding*, the mean value of the error is 0; if *truncation* to the next lower integer, the mean value is $-\frac{b}{2}$. It is quite easy to evaluate the variance of uniformly distributed noise:

$$\sigma_n^2 = \frac{b^2}{12}$$

For an m-bit quantizer and a signal with with maximum and minimum amplitudes $f_{\max}$ and $f_{\min}$, the width of a quantization level is:

$$b = \frac{f_{\max} - f_{\min}}{2^m} \equiv \frac{\Delta f}{2^m}$$

and by assuming that the quantization noise is uniformly distributed, the variance of the quantization noise is:

$$\sigma_n^2 = \frac{b^2}{12} = \frac{(\Delta f)^2}{12 \cdot (2^m)^2} = (\Delta f)^2 \cdot \left(12 \cdot 2^{2m}\right)^{-1}$$

The resulting *SNR* is the ratio of the variance of the signal to that of the quantization noise:

$$SNR \equiv \frac{\sigma_f^2}{\sigma_n^2} = \sigma_f^2 \cdot \frac{12 \cdot 2^{2m}}{(\Delta f)^2}$$

which, when expressed on a logarithm scale, becomes:

$$\begin{aligned} SNR &= 10\ \log_{10}\left[\sigma_f^2 \cdot 12 \cdot 2^{2m}\right] - 10\ \log_{10}\left[(\Delta f)^2\right] \\ &= 10\ \ \log_{10}\left[\sigma_f^2\right] + 10\ \ \log_{10}[12] + 20m\ \ \log_{10}[2] - 10\log_{10}\left[(\Delta f)^2\right] \\ &\cong 10\log_{10}\left[\sigma_f^2\right] + 10 \cdot 1.079 + 20m \cdot 0.301 - 10\log_{10}\left[(\Delta f)^2\right] \\ &\cong 6.02\ m + 10.8 + 10\log_{10}\left[\left(\frac{\sigma_f^2}{(\Delta f)^2}\right)\right]\quad [dB] \end{aligned}$$

The third term obviously depends on both the signal and the quantizer. This equation certainly demonstrates that the *SNR* of quantization increases by $\gtrapprox 6\ dB$ for every bit added to the quantizer. If using the (poor) estimate that $\sigma_f^2 = (\Delta f)^2$, then the third term evaluates to zero and the approximate *SNR* is:

$$\boxed{SNR \text{ for quantization to } m \text{ bits} \cong 6.02\ m + 10.8 + 10\ \ \log_{10}[1]) = 6.02\ m + 10.8\quad [dB]}$$

The statistics of the signal (and thus its variance σ_f^2) may be approximated for many types of signals (*e.g.*, music, speech, realistic images) as resulting from a random process. The histograms of these signals usually are peaked at or near the mean value μ and the probability of a gray level decreases for values away from the mean; the signal approximately is the output of a Gaussian random process with variance σ_f^2. By selecting the dynamic range of the quantizer Δf to be sufficiently larger than σ_f, few (if any) levels should be saturated at and clipped by the quantizer. As already stated, we assume that virtually no values are clipped if the the maximum and minimum levels of the quantizer are four standard deviations from the mean level:

$$\mu_f - f_{min} = f_{max} - \mu_f = \frac{\Delta f}{2} = 4\ \sigma_f$$

In other words, we may choose the step size between levels of the quantizer to satisfy the criterion:

$$\Delta f = 8\ \sigma_f \Longrightarrow \frac{\sigma_f^2}{(\Delta f)^2} = \frac{1}{64}$$

The *SNR* of the quantization process becomes:

$$\begin{aligned} SNR &= 6.02\ m + 10.8 + 10\ \ \log_{10}\left[\frac{1}{64}\right] \\ &= 6.02\ m + 10.8 + 10\ \ (-1.806) \\ &= 6.02\ m - 7.26\ [dB] \end{aligned}$$

which is 18 dB less than the estimate obtained by assuming that $\sigma_f^2 \cong (\Delta f)^2$. This again demonstrates that the original estimate of *SNR* was optimistic.

This expression for the *SNR* of quantizing a Gaussian-distributed random signal with measured variance σ_f^2 may be demonstrated by quantizing that signal to m bits over the range $f_{min} = \mu - 4\sigma_f$ to $f_{max} = \mu + 4\sigma_f$, and computing the variance of the quantization error σ_n^2. The resulting *SNR* should satisfy the relation:

$$SNR = 10\ \log_{10}\left[\frac{\sigma_f^2}{\sigma_n^2}\right] = (6.02\ m - 7.26)\ dB$$

The *SNR* of a noise-free analog signal after quantizing to 8 bits is $SNR_8 \cong 41\ dB$; if quantized to 16 bits (common in CD players), $SNR_{16} \cong 89\ dB$. The best *SNR* that can be obtained from analog recording (such as on magnetic tape) is about 65 dB, which is equivalent to that from a signal digitized to 12 bits per sample or 4096 gray levels.

The flip side of this problem is to determine the effective number of quantization bits after digitizing a noisy analog signal. This problem was investigated by Shannon in 1948. The analog signal is partly characterized by its bandwidth $\Delta\nu$ [Hz], which is the analog analogue of the concept of digital data rate [bits per second]. The bandwidth is the width of the region of support of the signal spectrum (its Fourier transform).

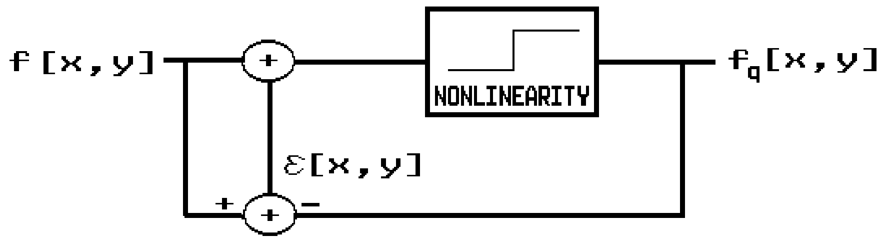
When sampling and quantizing a noisy analog signal, the *bit rate* is determined by the signal-to-noise ratio of the analog signal. According to Shannon, the bandwidth $\Delta\nu$ of a transmission channel is related to the maximum digital data rate R_{max} and the dimensionless signal-to-noise power ratio *SNR* via:

$$R_{max}\left(\frac{\text{bits}}{\text{sec}}\right) = (2\cdot\Delta\nu)\log_2\left[1+SNR\right]$$

where Shannon defined the *SNR* to be the ratio of the peak signal power to the average white noise power. It is very important to note that the *SNR* in this equation is a dimensionless ratio; it is NOT compressed via a logarithm and is not measured in dB. The factor of 2 is needed to account for the negative frequencies in the signal. The quantity $\log_2[1+SNR]$ is the number of effective quantization bits, and may be seen intuitively in the following way: if the total dynamic range of the signal amplitude is S, the dynamic range of the signal power is S^2. If the variance of the noise power is σ^2, then the effective number of quantization *transitions* is the power *SNR*, or $\frac{S^2}{\sigma^2}$. The number of quantization levels is $1+SNR$, and the effective number of quantization bits is $\log_2[1+SNR]$.

14.4 Quantizers with Memory – Error Diffusion

Another way to change the quantization error is to use a quantizer with memory, which means that the quantized value at a pixel is determined in part by the quantization error at nearby pixels. A schematic diagram of the quantizer with memory is:



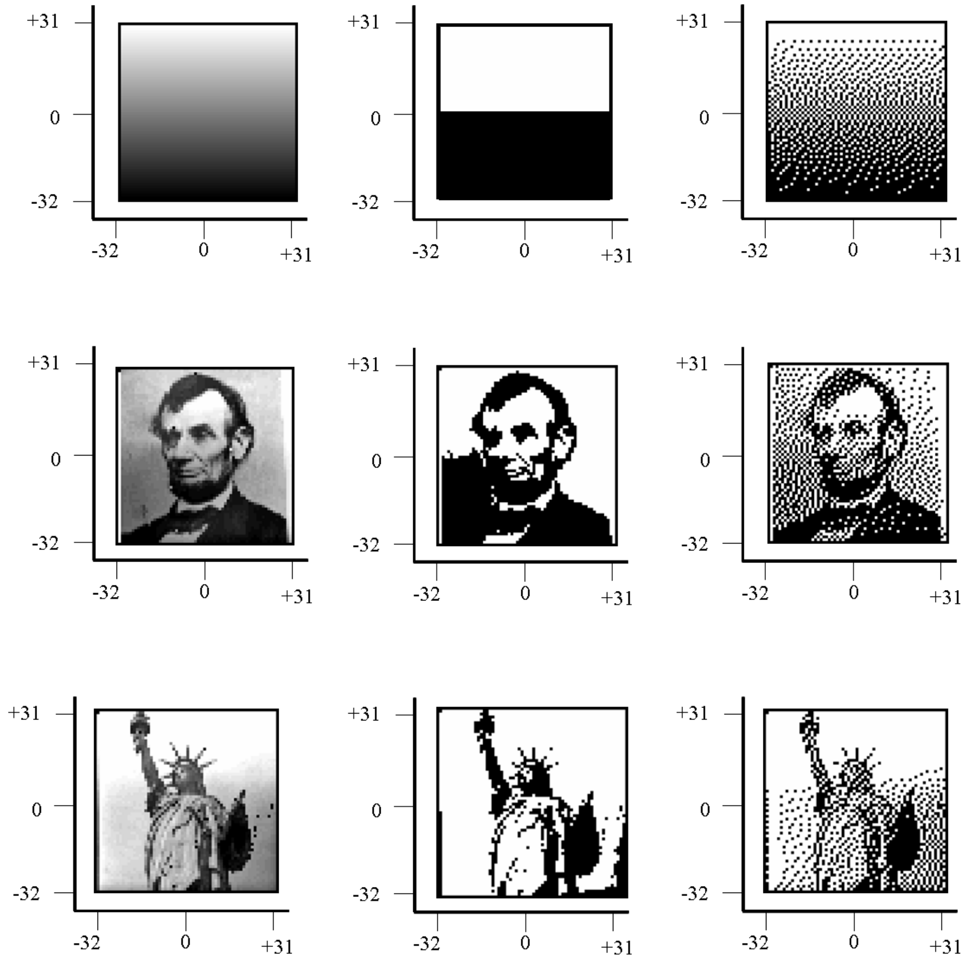
Flow chart for quantizer with memory

A simple method for quantizing with memory that generally results in reduced total error without *a priori* knowledge of the statistics of the input image and without adding much additional complexity of computation was introduced by Floyd and Steinberg (**Proc. SID**, **17**, pp.75-77, 1975) as a means to simulate gray level images on binary image displays and is known as error diffusion. It is easily adapted to multilevel image quantization. As indicated by the name, in error diffusion the quantization error is from one pixel is used to in the computation of the levels of succeeding pixels. In its simplest form, all quantization error at one pixel is added to the gray level of the next pixel before quantization. In the 1-D case, the quantization level at sample location x is the gray level of the sample minus the error $\epsilon[x-1]$ at the preceding pixel:

$$\begin{aligned} f_q[x] &= Q\{f[x] - \epsilon[x-1]\} \\ \epsilon[x] &= f[x] - f_q[x] \\ &= f[x] - Q\{f[x] - \epsilon[x-1]\} \end{aligned}$$

In the 2-D case, the error may be weighted and propagated in different directions. A discussion of the use of error diffusion in ADC was given by Anastassiou (**IEEE Trans. Circuits and Systems, 36**, 1175, 1989).

The examples on the following pages demonstrate the effects of binary quantization on gray-level images. The images of the ramp demonstrate that why the binarizer with memory is often called pulse-density modulation. Note that the error-diffused images convey more information about fine detail than the images from the memoryless quantizer. This is accomplished by possibly enhancing the local binarization error.



2-D error-diffused quantization for three different gray-scale images: (a) linear ramp image, after quantizing at the midgray level, and after Floyd-Steinberg error diffusion at the midgray level; (b) same sequence for "Lincoln"; (c) same sequence for "Liberty." The error-diffused images convey more information about the larger spatial frequencies

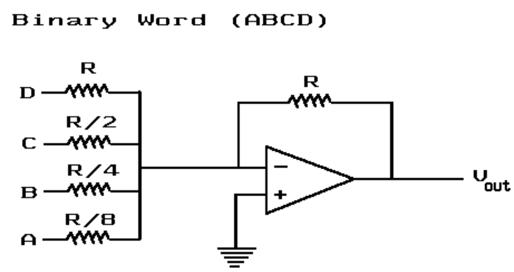
14.5 Image Display Systems – Digital - to - Analog Conversion

A complete image processing system must regenerate a viewable signal from the quantized samples. This requires that the digital signal be converted back to a continuously varying brightness distribution; analog estimates of the samples of the original signal are derived by a digital-to-analog converter (DAC) and the brightness is spread over the viewing area by the interpolation of the display. Each of these processes will be discussed in turn, beginning with the DAC.

The principle of the DAC is very intuitive; each bit of the digital signal represents a piece of the desired output voltage that is generated by a voltage divider ladder network and a summing amplifer. For example, if a 4-bit digital signal is represented by the binary word ABCD, the desired output voltage is:

$$V_{out} = V(8A + 4B + 2C + D)$$

where V is the desired voltage for a signal represented by the binary word 0001. The appropriate DAC signal is shown below:



Digital-to-analog converter circuit for 4-bit binary input with bit values ABCD. The circuit generates an analog output voltage $V = D + 2C + 4B + 8A$.

Variations of the circuit shown are more practical for long binary words, but the principle remains the same. Note that the output voltage is analog, but it is still quantized, i.e., only a finite set of output voltages is possible (ignoring any noise).

14.6 Image Interpolation

The image display generates a continuously varying function $g[x, y]$ from the processed image samples $g_q[n, m]$. This is accomplished by defining an interpolator that is placed at each sample with the same amplitude as the sample. The continuously varying reconstructed image is the sum of the scaled interpolation functions. This is analogous to the connect-the-dots puzzle for children to fill in the contours of a picture. Mathematically, interpolation may be expressed as a convolution of the output sampled image with an interpolation function (the postfilter) h_2. In 1-D:

$$g[x] = \sum_{n=-\infty}^{\infty} g_q[n \cdot \Delta x] \cdot h_2[x - n \cdot \Delta x] = g_q[x] * h_2[x]$$

In an image display, the form of the interpolation function is determined by the hardware and may have very significant effects on the character of the displayed image. For common cathode-ray tubes (CRTs – the television tube), the interpolation function is approximately a gaussian function, but is often further approximated by a circle (or cylinder) function.

The effect of the interpolator on the output is illustrated by a few simple examples. In the 1-D case, the input is a sinusoid with period $X_0 = 64$ sampled at intervals $\Delta x = 8$. The interpolators are a rect function (nearest-neighbor interpolator), triangle function (linear interpolator), cubic b-spline, and a Gaussian. Examples for 2-D images are shown on following pages.

14.6.1 Ideal Interpolation

In the discussion of the Whittaker-Shannon sampling theorem, we have stated that an unaliased function can be perfectly reconstructed from its unaliased ideal samples. Actually, as stated the theorem is true but a bit misleading. To be clearer, we could say the following:

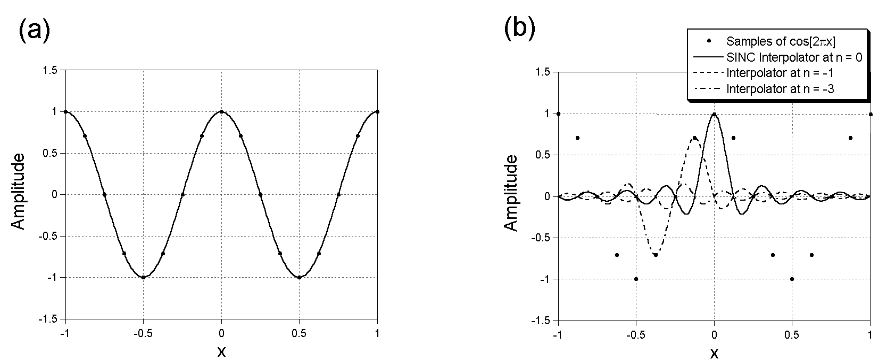
Any function can be perfectly reconstructed from an infinite number of unaliased samples, i.e., samples obtained at a rate greater than two times per period of the highest frequency component in the original function.

In reality, of course, we always have a finite number of samples, and thus we cannot perfectly reconstruct an arbitrary function. Periodic functions may be reconstructed, however, because the samples of a single period will be sufficient to recover the entire function.

In the example just presented, the ideal interpolation function must be something other than a rectangle or gaussian function. We will again assert without proof that the ideal interpolator for samples separated by a distance Δx is:

$$h_2[x] = SINC\left[\frac{x}{\Delta x}\right]$$

Note that the *SINC* function has infinite support and is bipolar; thus it is not obvious how to implement such a display. However, we can illustrate the result by using the example of the sampled cosine already considered. Note that the cosine is periodic.



Ideal interpolation of the function $f[x] = \cos[2\pi x]$ sampled with $\Delta x = \frac{1}{16}$ unit. The weighted Dirac delta functions at each sample are replaced by weighted SINC functions (three shown, for $n = 0, -1, -3$), which are summed to reconstruct the original cosine function.

14.6.2 Modulation Transfer Function of Sampling

We have just demonstrated that images may be perfectly reconstructed from unaliased and unquantized ideal samples obtained at intervals Δx by interpolating with $SINC\left[\frac{x}{\Delta x}\right]$. Of course, reconstructed images obtained from a finite number of samples systems obtained from a system with averaging and quantization will not be perfect. We now digress to illustrate a common metric for imaging system quality by applying it to realistically sampled systems. Though it is not strictly appropriate, the illustration is still instructive.

Averaging by the detector ensures that the modulation of a reconstructed sinusoid $g[x]$ will generally be less than that of the continuous input function $f[x]$, i.e., image modulation is imperfectly transferred from the input to the reconstructed output. The transfer of modulation can be quantified for sinusoids of each frequency; because the averaging effect of the digitizer is fixed, higher-frequency sinusoids will be more affected than lower frequencies. A plot of the modulation transfer vs. spatial frequency is the modulation transfer function or MTF. Note that MTF describes a characteristic of the system, not the input or output.

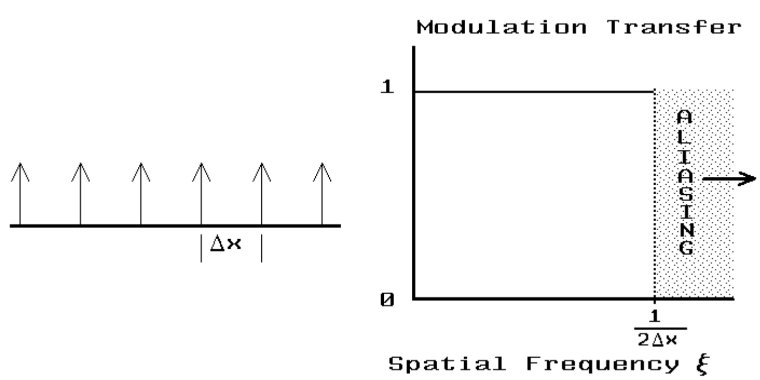
For ideal sampling (and ideal reconstruction) at all frequencies less than Nyquist, the input function $f[x]$ is perfectly reconstructed from the sample values $f_s[n \cdot \Delta x]$, and therefore the modulation transfer function is unity for spatial frequencies less than $\frac{1}{2}$ cycle per pixel.

Sinusoids with frequencies $\xi >$ the Nyquist frequency are aliased by ideal sampling.

The "new" frequency is less than the Nyquist frequency.

Because the output frequency is different from the input frequency,

it is not sensible to talk about the transfer of modulation for frequencies above Nyquist.



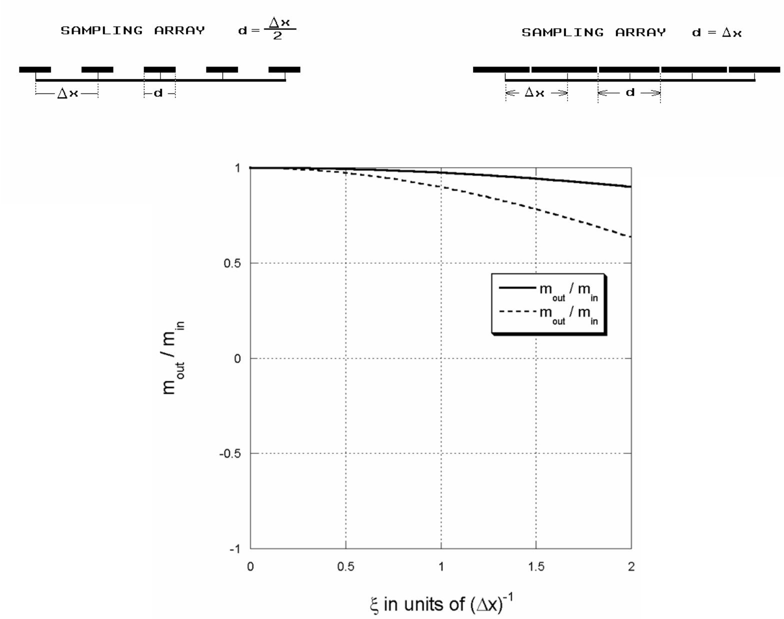
Schematic of the modulation transfer function of the cascade of ideal sampling and ideal interpolation; the MTF is unit at all spatial frequencies out to the Nyquist frequency.

14.6.3 MTF of Realistic Sampling (Finite Detectors)

We have already demonstrated that the modulation due to uniform averaging depends on the detector width d and the spatial frequency ξ of the function as $SINC(d\xi)$. If the detector size is half the sampling interval ($d = \frac{\Delta x}{2}$), the MTF is:

$$SINC\,[d\xi] = SINC\left[\frac{\Delta x}{2}\cdot\frac{1}{2\cdot\Delta x}\right] = SINC\left[\frac{1}{4}\right]$$
$$= \frac{\sin\left[\frac{\pi}{4}\right]}{\frac{\pi}{4}} = \frac{4}{\pi}\cdot\sqrt{\frac{2}{2}} \cong 0.9 \text{ at the Nyquist frequency.}$$

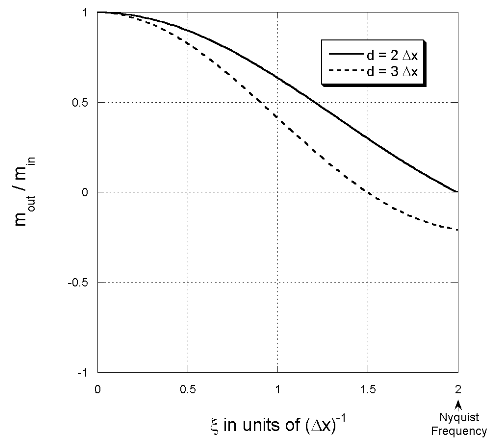
i.e., can still be reconstructed perfectly by appropriately amplifying the attenuated sinusoidal components, a process known as inverse filtering that will be considered later. In the common case of detector size equal to sampling interval ($d = \Delta x$), the minimum MTF is $SINC\,[0.5] = 0.637$ at the Nyquist frequency.



MTF of sampling for $d = \frac{\Delta x}{2}$ and $d = \Delta x$.

By scanning, we can sample the input sequentially, and it is thus possible to a detector size larger than the sampling interval. If $d = 2\cdot\Delta x$, then the detector integrates over a full period of a sinusoid at the Nyquist frequency; the averaged signal at this frequency is constant (usually zero, i.e., no modulation).

For larger scanned detectors, the modulation can invert, i.e., the contrast of sinusoids over a range of frequencies can actually reverse. This has already been shown for the case $\frac{d}{X} = 1.5 \Longrightarrow d = 3\cdot\Delta x$ at the Nyquist rate.



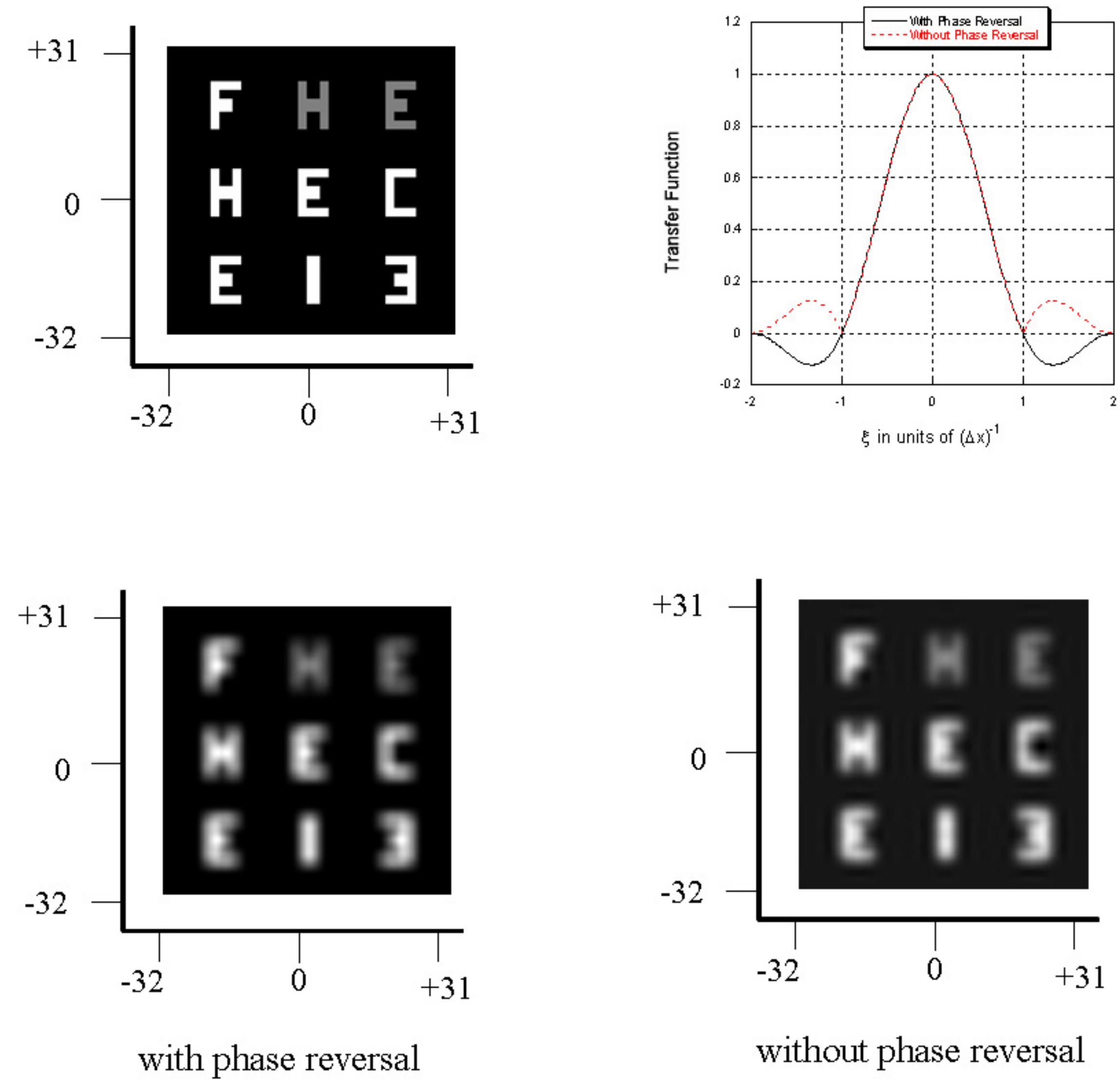
MTF of scanning systems with $d = 2 \cdot \Delta x$ and $d = 3 \cdot \Delta x$, showing that the $MTF = 0$ at one frequency and is negative for larger spatial frequencies approaching the Nyquist frequency in the second case. This leads to a phase shift of the reconstructed sinusoids.

If the inputs are square waves, the analogous figure of merit is the contrast transfer function or CTF.

14.7 Effect of Phase Reversal on Image Quality

To illustrate the effect on the image of contrast reversal due to detector size, consider the examples shown below.

The input was imaged with two different systems: the MTF of the first system reversed the phase of sinusoids with higher frequencies, while the second did not. Note the sharper edges of the letters in the second image:



Effect of phase reversal on image quality. The edges are arguably "sharper" with the phase reversal.

14.8 Summary of Effects of Sampling and Quantization

ideal sampling $\Longrightarrow$ aliasing if undersampled

realistic sampling $\Longrightarrow$ aliasing if undersampled $\Longrightarrow$ modulation reduced at all nonzero spatial frequencies

quantization $\Longrightarrow$ error is inherent in the nonlinear operation

morebits, less noise $\Longrightarrow$ less error

14.9 Spatial Resolution

Photographic resolution is typically measured by some figure of merit like $\frac{\text{cycles}}{\text{mm}}$or line pairs per mm, which are the maximum visible spatial frequency of a recorded sine wave or square wave, respectively. Visibility is typically defined by a specific value of the emulsion's modulation transfer function (MTF, for sinusoids) or contrast transfer function (CTF, for square waves). The specific point of the modulation curve that is used as the resolution criterion may be different in different applications. For example, the resolution of imagery in highly critical applications might be measured as the spatial frequency where the modulation transfer is 0.9, while the frequency where the MTF is 0 may be used for noncritical applications. The spatial resolution of digital images may be measured in similar fashion from the MTF curve due to sampling, which we have just determined to be a function of the sampling interval Δx and the detector width d. The maximum frequency that can be reconstructed is the Nyquist limit $\xi_{max} = \frac{1}{2 \cdot \Delta x}$, and the modulation at spatial frequency ξ varies with the detector size as $SINC[d\xi]$. In remote sensing, it is common to use the instantaneous field of view (IFOV) and ground instantaneous field of view (GIFOV). The IFOV is the full-angle subtended by the detector size d at the entrance pupil of the optical system. The term GIFOV is inappropriate for the definition; spot size would be better. The GIFOV of a digital imaging system is the spatial size of the detector projected onto the object, e.g.,the GIFOV of the French SPOT satellite is 10m.