COMP ENG 4TL4:

Digital Signal Processing

Notes for Lecture #30

Friday, November 21, 2003

7. SIGNAL COMPRESSION

7.1 Introduction to Signal Compression

Compression involves the representation of N bits of information with N_c bits, where $N_c < N$.

The ratio $N{:}N_c$ is referred to as the compression ratio.

Two forms of compression exist:

1. Lossless, in which the exact original signal can be retrieved (without error/distortion). Mathematically this corresponds to an *invertible* operation.
2. Lossy, in which the original signal cannot be exactly retrieved. Mathematically this corresponds to a *noninvertible* operation.

Much higher compression ratios can be obtain for lossy compression, but this comes at the expense of distortion/error in the retrieved signal.

7.2 Parametric Signal Compression

Parametric signal compression is based on (parametric) modeling of certain signal types.

For example, if a signal $y[n]$ is known to be well described by a slowly-varying autoregressive (AR) process:

$$y\left[n\right] = \sum_{k=1}^{N} a_k y\left[n-k\right] + u\left[n\right],$$

then the AR filter coefficients a_k can be estimated at regular time intervals to describe the slowly-varying properties of the signal.

Note that the input signal to the AR process $u[n]$ must also be characterized.

Linear prediction coding (LPC) of speech signals:

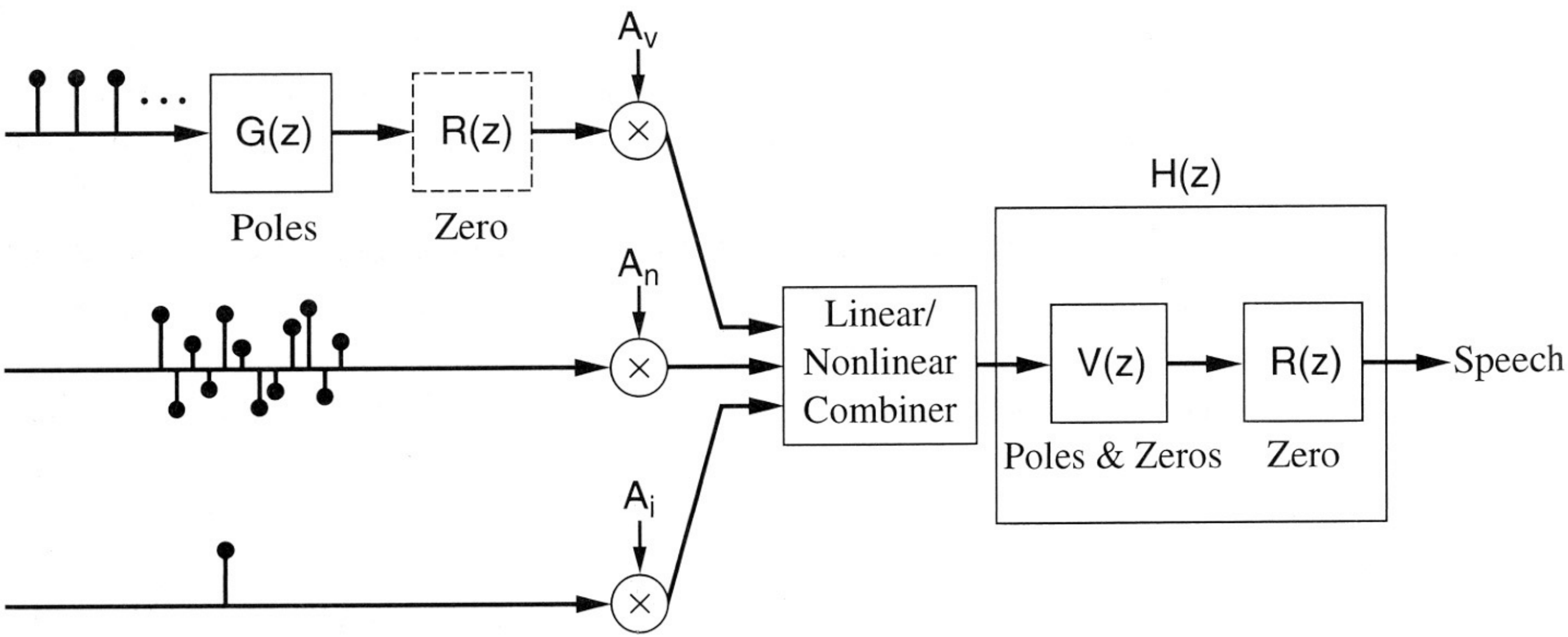
We recall that an AR system corresponds to an all-pole filter:

$$V(z) = \frac{Y(z)}{U(z)} = \frac{1}{1 - \sum_{k=1}^{N} a_k z^{-k}} = \frac{1}{A(z)}.$$

An all-pole filter does a pretty good job of modelling the vocal tract frequency response caused by the vocal tract resonances (formants).

Therefore, a reasonable model of speech over a short time window is an all-pole filter (with filter coefficients a_k describing the vocal tract frequency response) being driven by an excitation source $u[n]$, which can be switched between a *noisy signal* for *unvoiced speech*, a *periodic pulse-train* for *voiced speech*, and a *single pulse* for *plosives*.

Discrete-time model for speech production:



(Quatieri)

Figure 4.20 Overview of the complete discrete-time speech production model.

A standard method for obtaining the filter coefficients a_k and an estimate of the excitation source signal $u[n]$ is referred to as linear prediction coding (LPC).

In this method, a windowed segment of speech $s[n]$ is passed through the inverse of the all-pole model, $A(z)$, to obtain an estimate of the excitation source signal $u[n]$.

Note that the filter $A(z)$ can be broken into two parallel subsystems, in which a difference is taken between $s[n]$ and $s[n]$ filtered by $P(z)$, where:

$$P(z) = \sum_{k=1}^{N} a_k z^{-k}.$$

Filtering view of linear prediction:

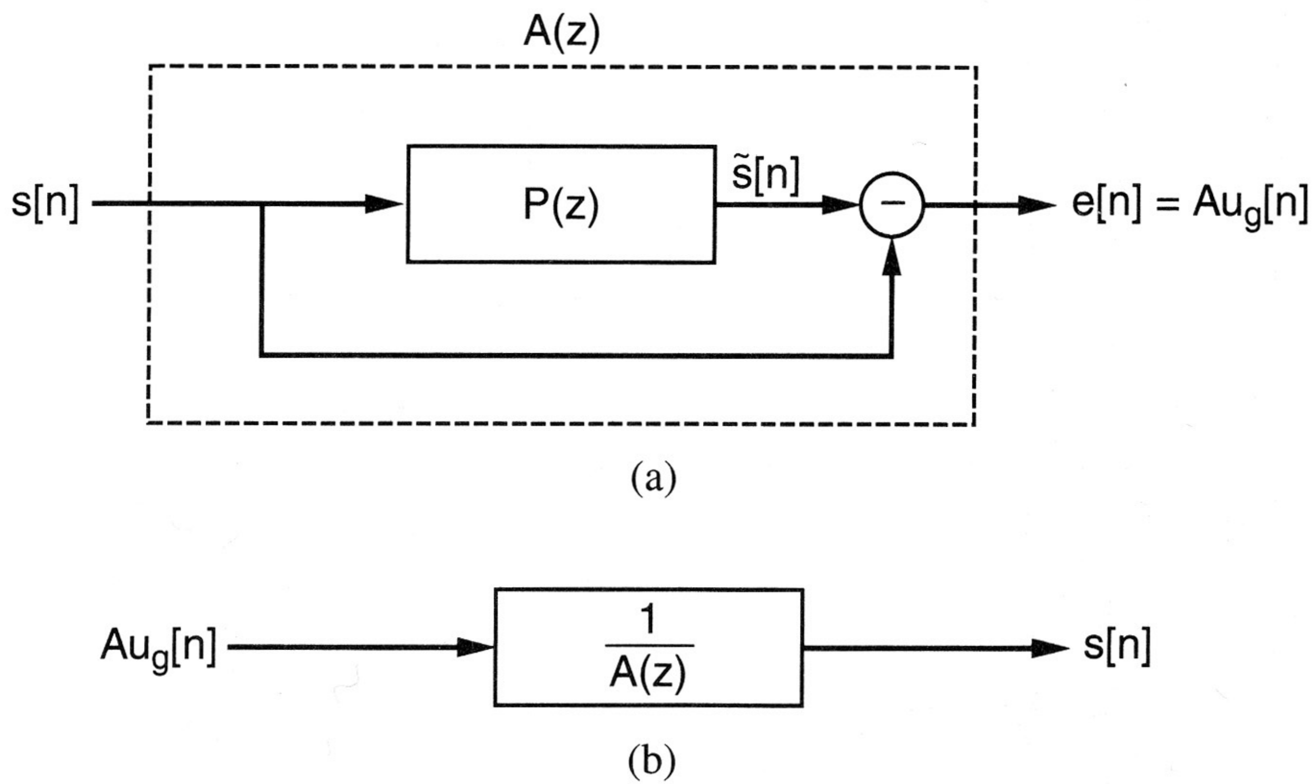


Figure 5.1 Filtering view of linear prediction: (a) prediction-error filter $A(z) = 1 - P(z)$; (b) recovery of $s[n]$ with $\frac{1}{A(z)}$ for $\alpha_k = a_k$. $A(z)$ is also considered the inverse filter because it can yield the input to an all-pole transfer function.

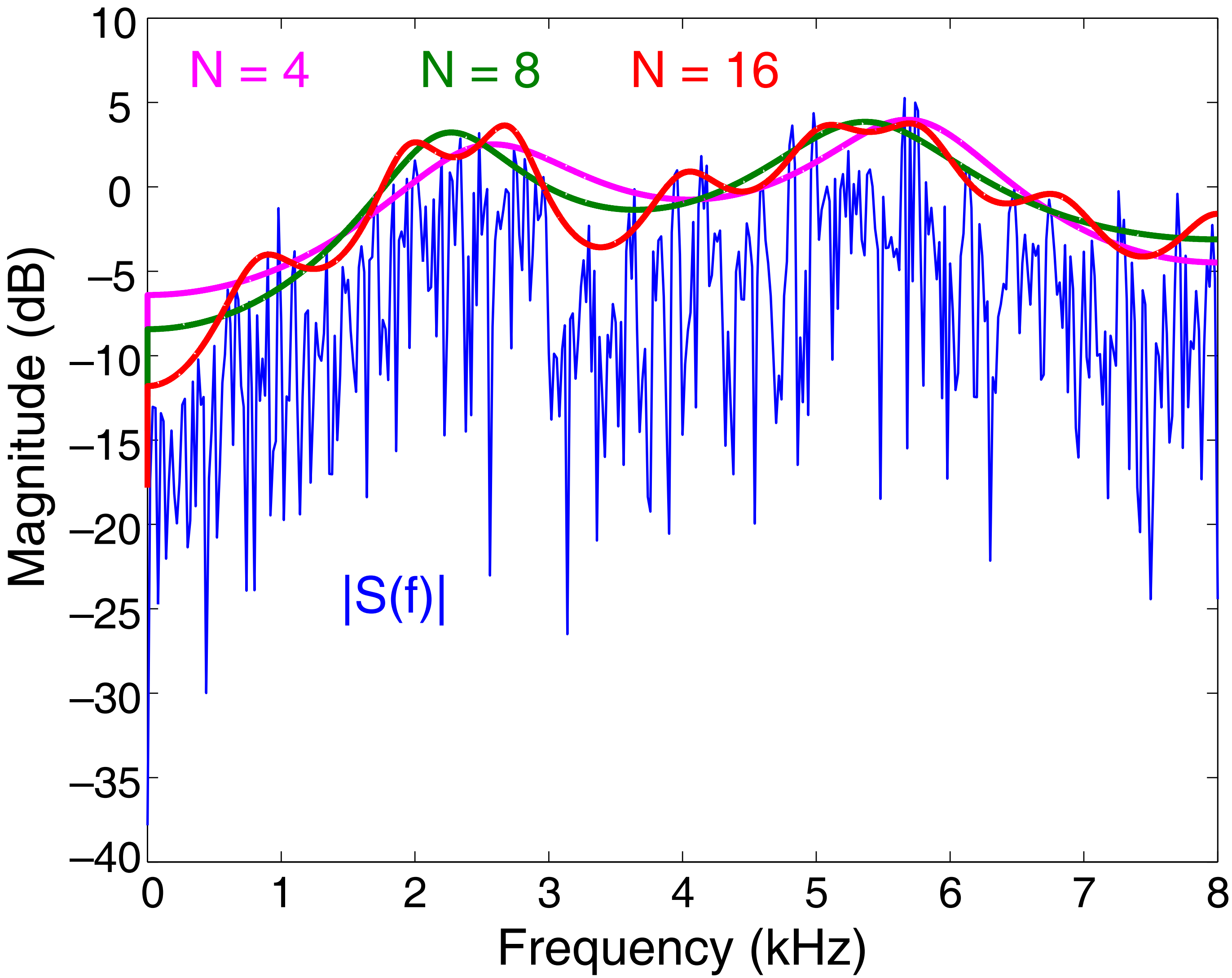
(Quatieri)

$P(z)$ is referred to as a linear “prediction filter”—hence the name linear prediction coding—because the present value of the output of the filter is based on a linear combination of the N past values of the input to the filter:

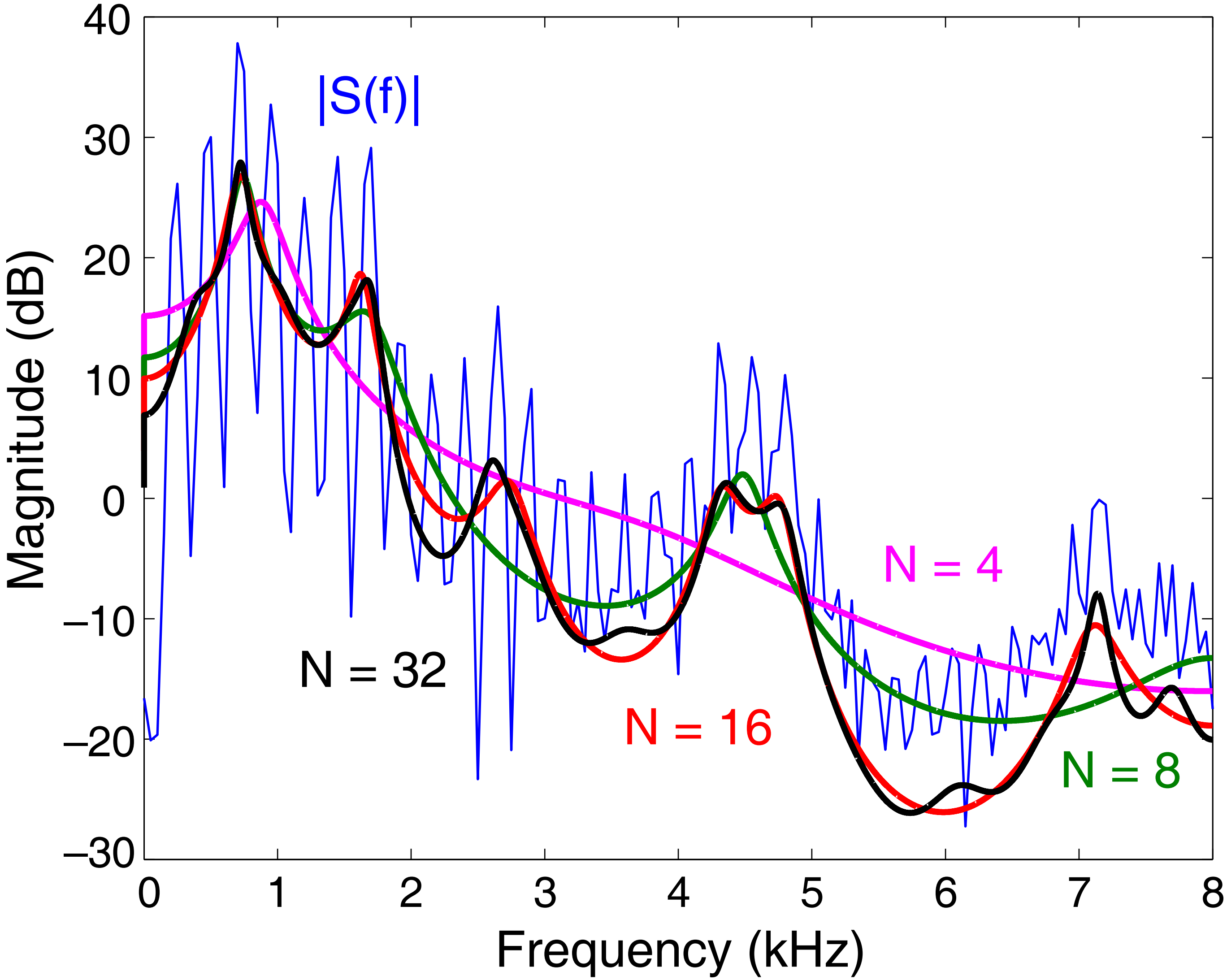
$$\tilde{s}[n] = \sum_{k=1}^{N} a_k s[n-k]\,.$$

We can view the output of the filter $A(z)$ as an error signal $e[n]$. If this error signal is minimized, we obtain the best fit between the actual speech segment $s[n]$ and the predicted speech segment $\tilde{s}[n]$ that is possible with an all-pole filter.

Example #1: Unvoiced speech; LPC with orders $N = 4, 8 \text{ and } 16$



Example #2: Voiced speech; LPC with orders $N = 4,\ 8,\ 16 \text{ and } 32$

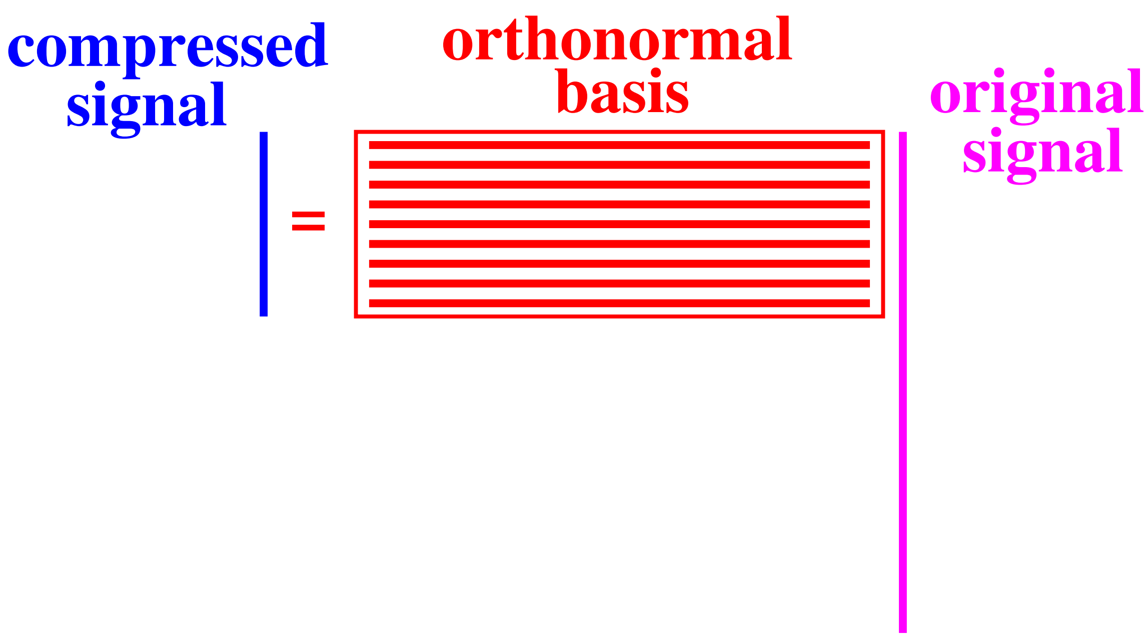


7.3 Nonparametric Signal compression

LPC is an example of coding that allows compression of a signal via *parametric* modelling.

Nonparametric methods also exist for signal compression, i.e., methods for which a model of the signal is not required.

Lossy nonparametric compression can be obtain via an orthonormal transform:



Original signal:

$$\mathbf{x} = \begin{bmatrix} x[0] & x[1] & \dots & x[M-1] \end{bmatrix}^T.$$

Compressed signal:

$$\mathbf{X} = \begin{bmatrix} X[0] & X[1] & \dots & X[L-1] \end{bmatrix}^T = \mathbf{T}^H \mathbf{x},$$

where:

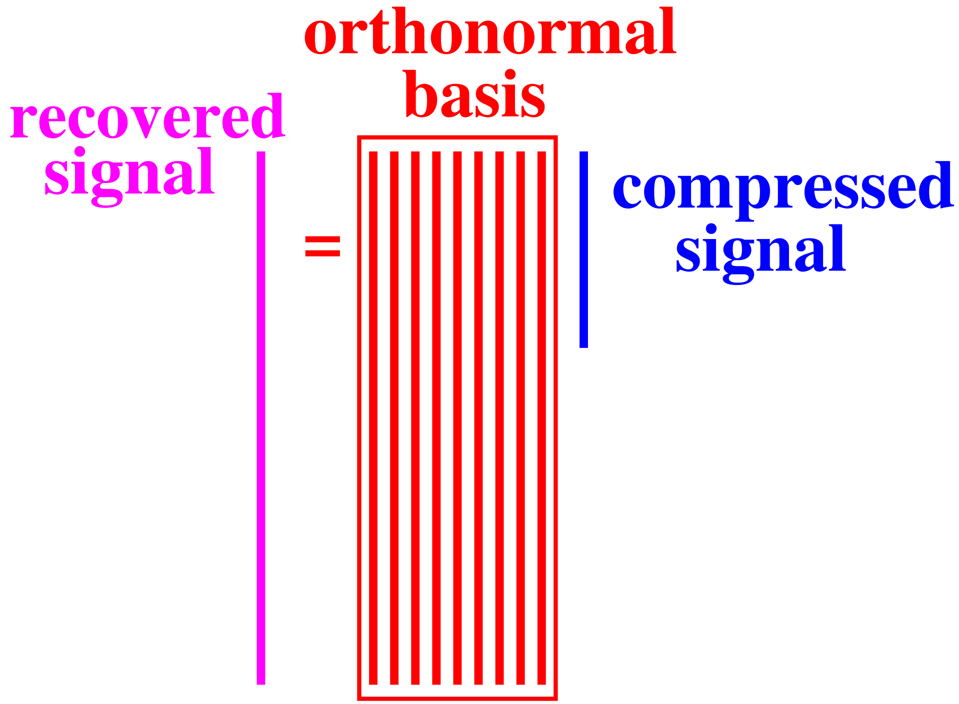
$$\frac{M}{L} = \frac{N}{N_c},$$

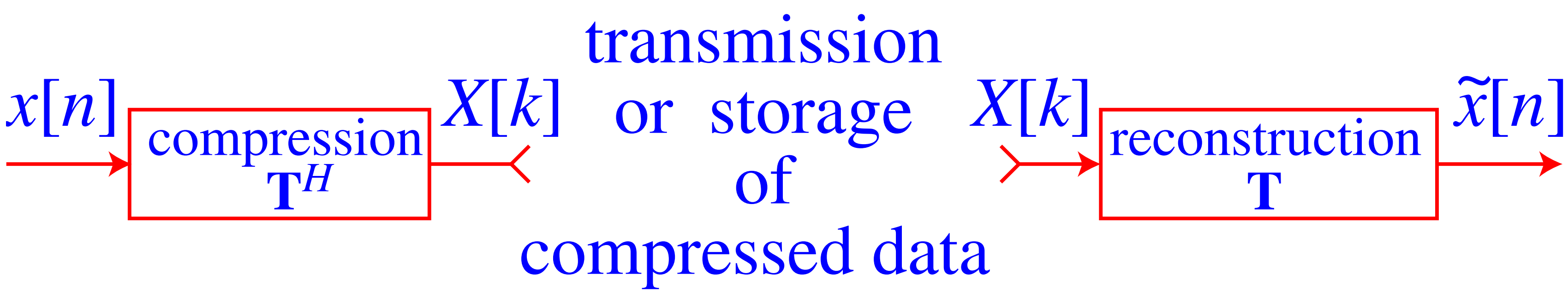
and $\mathbf{T}$ is the $M \times L$ *transformation matrix*.

The signal $\mathbf{X}$ can be interpreted as a vector of coefficients of expansion using the orthonormal basis $\mathbf{T}$:

$$\tilde{\mathbf{x}} = \mathbf{T}\mathbf{X}.$$

Signal reconstruction:





DFT-based compression:

Recall that the matrix formulation of the DFT is:

$$\mathbf{X} = \mathbf{W}\mathbf{x},$$

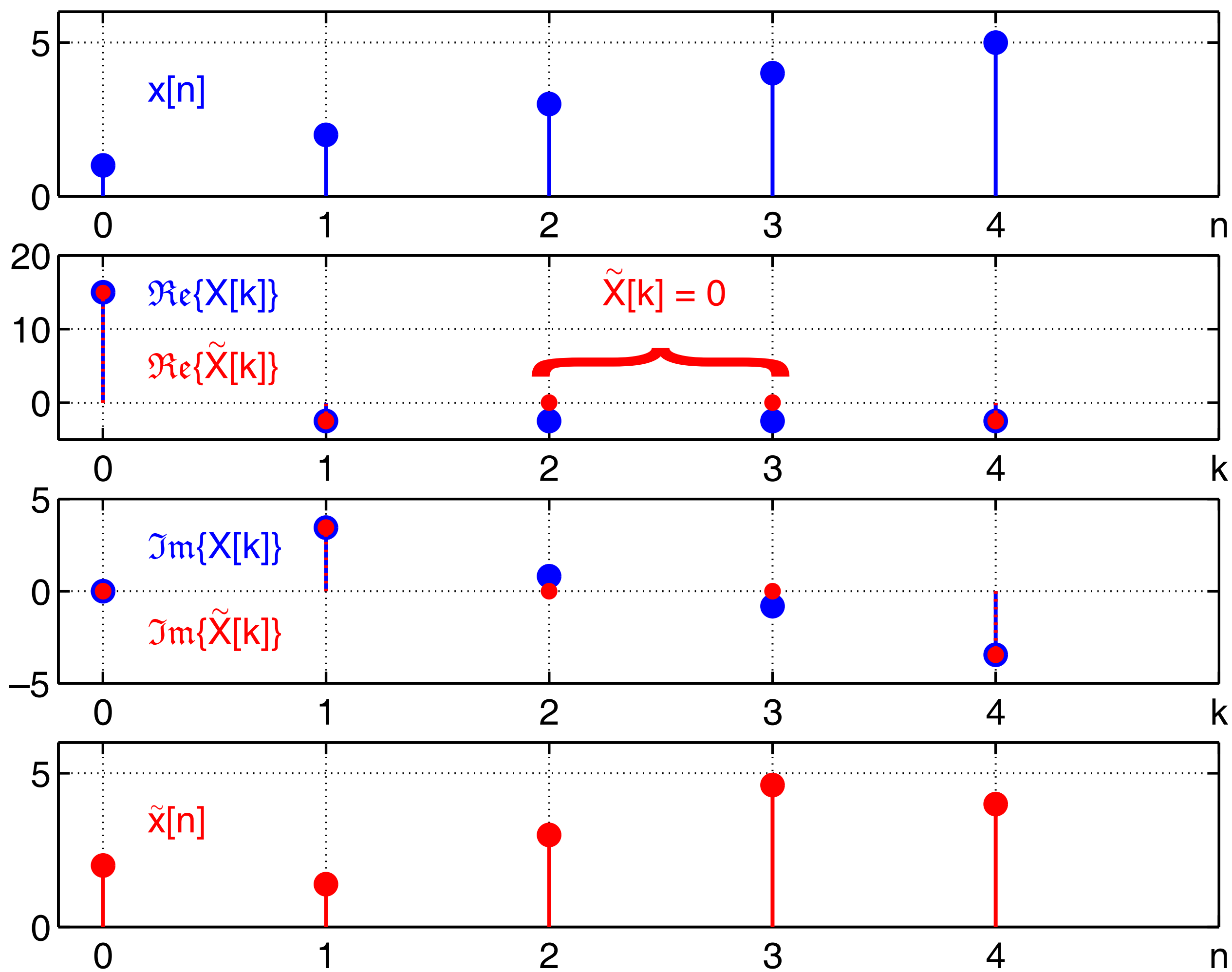
and the inverse DFT is:

$$\mathbf{x} = \frac{1}{N}\mathbf{W}^H\mathbf{X}.$$

Compression of the DFT vector $\mathbf{X}$ can be obtained by setting values of $\mathbf{X}$ at specific frequency indices k *to zero* to give the vector $\tilde{\mathbf{X}}$—these values need not be stored or transmitted, because they are known to be zero. The signal can be reconstructed using the inverse DFT:

$$\tilde{\mathbf{x}} = \frac{1}{N}\mathbf{W}^H\tilde{\mathbf{X}}.$$

Example #3: $M = 5; L = 3 \Rightarrow$ compression ratio = 5:3



Note that the compression scheme has introduced some high-frequency error/distortion into the reconstructed signal.

Example #4: Compression ratio $\simeq$ 16:1

Original image

Reconstructed image