ECTROMAGNETISM

cs and engineering which still plays
iew technology. Electromagnetism
ers, material scientists, and applied
se aspects of applied and theoretical
ly important in modern and rapidly
the needs of researchers, students,

ISM

S EDITOR
E PARK, MARYLAND

Magnetic Information Storage Technology

Shan X. Wang
Department of Materials Science and Engineering
(and of Electrical Engineering)
Stanford University
Stanford, California

Alexander M. Taratorin
IBM
Almaden Research Center
San Jose, California

ACADEMIC PRESS
San Diego London Boston
New York Sydney Tokyo Toronto

The cover page illustrates a hard disk drive used in modern computers. Hard disk drives are the most widely used information storage devices. The block diagram shows the principle of the state-of-the-art disk drives. Any information such as text files or images are first translated into binary user data, which are then encoded into binary channel data. The binary data are recorded in magnetic disks by a write head. For example, the magnetization pointing to the right (or left) represents "1" (or "0"). The magnetization pattern generates a voltage waveform when passing underneath a read head, which could be integrated with the write head and located near the tip of the stainless suspension. The voltage waveform is then equalized, i.e., reshaped into a proper form. This equalization step, along with a so-called maximum likelihood detection algorithm, allows us to detect the binary channel data with a high reliability. The trellis diagram shown with 16 circles and 16 arrows is the foundation of maximum likelihood detection. The detected binary channel data are then decoded back to the binary user data, which are finally translated back to the original information.

This book is printed on acid-free paper. ♾

ACADEMIC PRESS
a division of Harcourt Brace & Company
525 B Street, Suite 1900, San Diego, California 92101-4495, USA
http://www.apnet.com

Academic Press
24-28 Oval Road, London NW1 7DX, UK
http://www.hbuk.co.uk/ap/

Library of Congress Cataloging-in-Publication Data

Wang, Shan X.
Magnetic information storage technology / Shan X. Wang, Alex Taratorin.
p. cm. — (Electromagnetism)
Includes index.
ISBN 0-12-734570-1
1. Magnetic recorders and recording. I. Wang, Shan X.
II. Taratorin, A. M. III. Title. IV. Series.
TK7881.6.W26 1998
621.39'7—dc21 98-24797
CIP

Printed in the United States of America
99 00 01 02 03 MB 9 8 7 6 5 4 3 2 1

Contents

CHAPTER 11

Peak Detection Channel

Peak detection channel is a simple and reliable method of data detection. Peak detection was the first detection channel utilized in magnetic disk drives. It was extensively used for several decades and is still found in many disk drive products. In this chapter we will describe the main principles of peak detection channel operation and consider the error rates of this channel.

11.1 PEAK DETECTION CHANNEL MODEL

A block diagram of a typical peak detection channel is shown in Fig. 11.1. It is based on the assumption that each transition results in a relatively sharp peak of voltage. The goal of the peak detection channel is to detect each individual voltage peak.

The input analog signal is passed through two paths. One path qualifies a peak of voltage by rectification and threshold detection. When a voltage level exceeds some threshold, a comparator is turned on and a rectangular pulse appears at the output of the threshold detector.[1,2] The other path consists of a differentiator and a zero-crossing detector. A voltage peak will correspond to a zero-crossing after differentiation. The zero-crossing detector generates a short rectangular pulse for each zero-crossing. If a zero-crossing is detected and it is located within the region where signal amplitude exceeds a specific threshold, a transition is detected and a "qualified" pulse appears at the peak detector output.

The "coincidence" scheme used in the peak detection channel makes it robust to find the positions of magnetic transitions. If only the threshold detector is used, the pulse generated by the rectifier is relatively wide, which cannot give the exact location of the transition. On the other hand, if only the zero-crossing detector is used, a lot of extra zero-crossings

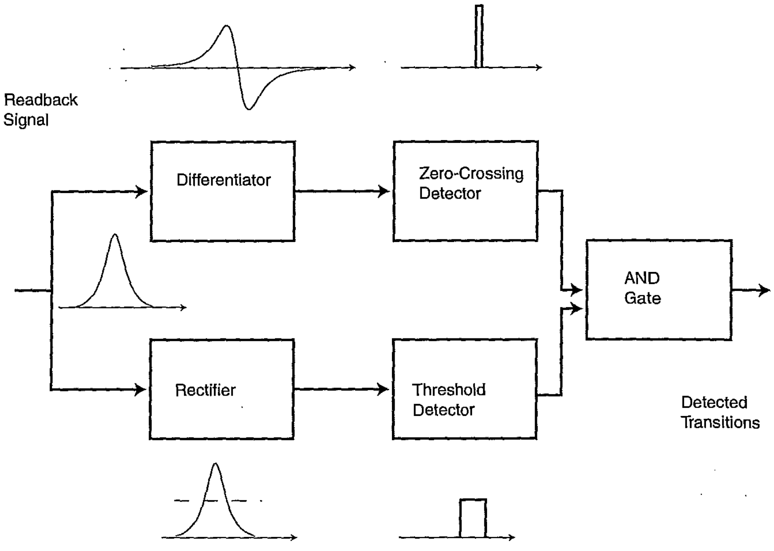
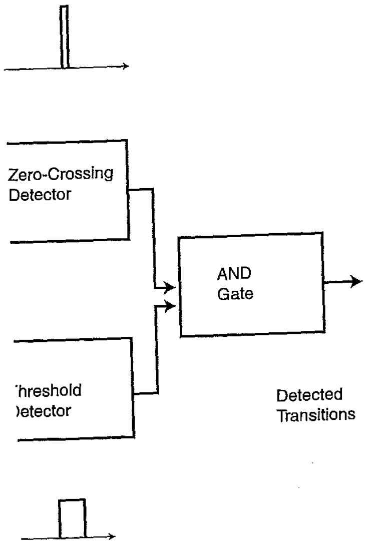


FIGURE 11.1. Block diagram of peak detection channel.

caused by noises will be mistaken as magnetic transitions. Only when both a zero-crossing and a rectified pulse are detected simultaneously, a magnetic transition is found reliably.

To distinguish between adjacent transitions and to combat instabilities of the disk rotational speed, each pulse of voltage is detected inside an appropriate *detection window,* also called a *timing window* and should be equal to the *channel bit period.* A special phase-locked loop (PLL) system is used to provide a detection window for each channel bit. The PLL updates its frequency based on detected pulses. Each incoming transition or voltage pulse is searched inside its detection window. As shown in Fig. 11.2, each pulse should be detected after the previous channel bit and before the next channel bit, so the timing window is equal to a channel bit period or *bit cell.* If a peak detection channel uses (1,7) modulation encoding, the detection window is equal to 50% of the minimum timing distance between two transitions that are written in the magnetic medium.

The performance of a detection channel is often characterized by channel bit rate as well as *bit error rate* (BER). Bit error rate P_e is the probability of mistaking a "0" as a "1", or mistaking a "1" as a "0" due to the noises, distortions, or interferences in the channel. The reciprocal of P_e means 1 error per $1/P_e$ bits transferred in the channel. Obviously,



tion channel.

nagetic transitions. Only when
ɜe are detected simultaneously, a

sitions and to combat instabilities
of voltage is detected inside an
l a *timing window* and should be
phase-locked loop (PLL) system
r for each channel bit. The PLL
pulses. Each incoming transition
detection window. As shown in
ɩfter the previous channel bit and
ɩg window is equal to a channel
ɩ channel uses (1,7) modulation
l to 50% of the minimum timing
written in the magnetic medium.
annel is often characterized by
ɛ (BER). Bit error rate P_e is the
or mistaking a "1" as a "0" due
ɜs in the channel. The reciprocal
ɜrred in the channel. Obviously,

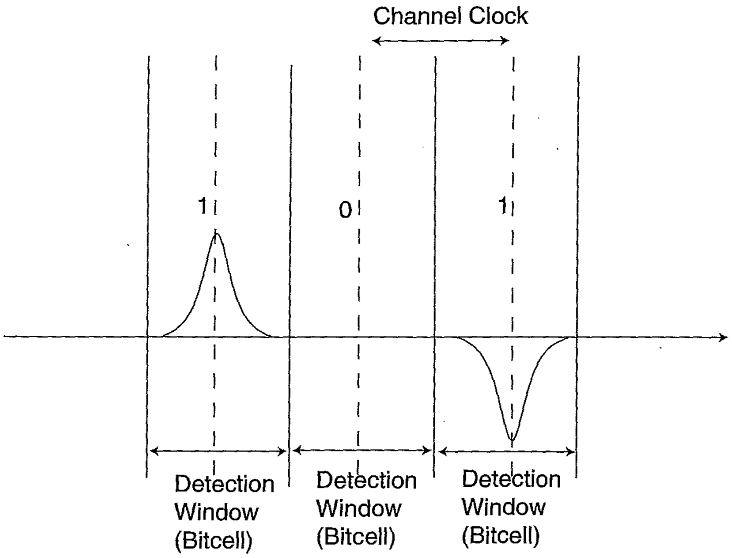


FIGURE 11.2. Clock and detection window in peak detection channel.

we would like P_e to be as small (ideally zero) as possible. The BER can be reduced through error detection and correction. At present, the *corrected* BER is usually in 10^{-12}–10^{-11} range, while the *raw* (uncorrected) BER is typically 10^{-9}–10^{-7}.

The error rate at the threshold detector is determined by the probability of drop-outs when the pulse amplitude falls below the specified threshold or the probability of strong noise outbursts when total media and electronic noise exceeds the specified signal level. The error rate at the zero-crossing detector is determined by random shifts of the zero-crossing position from the correct peak location. Random noises cause fluctuations of the zero-crossing position in the differentiated readback signal. An error occurs when the zero-crossing position falls beyond the detection window, i.e., when zero-crossing is detected earlier or later than the current bit cell. Next we will examine how to calculate BER.

11.2 BER AT THE THRESHOLD DETECTOR

The error rate of the threshold detector may be calculated from the channel SNR. If the zero-to-peak signal voltage amplitude is V_{0-p} and the rms noise voltage is $V_{rms,n}$, then the SNR at the threshold detector is defined as:

$$SNR(\text{dB}) = 20 \log \frac{V_{0-p}}{V_{rms,\,n}}. \tag{11.1}$$

To estimate the BER of the threshold detector in the peak detection channel, we assume that the noise voltage n in the recording channel is approximately Gaussian with a zero mean value and a standard deviation of $\sigma = V_{rms,n}$. This means that the probability density of the noise voltage n is given by the Gaussian distribution:

$$f(n) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{n^2}{2\sigma^2}\right). \tag{11.2}$$

Now we will consider the probability of errors in the threshold detector with a threshold fixed at 50% of the zero-to-peak signal amplitude. If there is no peak of voltage in the channel, we may detect a false transition when the noise outburst exceeds one-half of zero-to-peak signal voltage. The probability of this error event, i.e., mistaking a "0" for "1", is given by the following integral of the Gaussian distribution function:

$$P_{0/1} = \int_{V_{0-p}/2}^{\infty} \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{n^2}{2\sigma^2}\right)dn. \tag{11.3}$$

Define the following integral of the Gaussian distribution as the complementary error function:

$$\text{erfc}(x) = \frac{2}{\sqrt{\pi}}\int_x^{\infty}\exp\left(-y^2\right)dy. \tag{11.4}$$

Alternately, one can define a $Q(x)$ function:

$$Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^{\infty} e^{-y^2/2}dy = \frac{1}{2}\text{erfc}(x/\sqrt{2}),$$

which represents the probability that a unit-variance zero-mean Gaussian noise exceeds x. Then we can easily obtain that the BER is.

$$P_{0/1} = \frac{1}{2}\text{erfc}\left(\frac{V_{0-p}}{2\sqrt{2}\sigma}\right) = \frac{1}{2}\,\text{erfc}\left(\frac{\sqrt{SNR}}{2\sqrt{2}}\right) = Q\left(\frac{\sqrt{SNR}}{2}\right). \tag{11.5}$$

Sim
burst ex

Therefor
is stable
of the ch
channel S
at the thr
for by th
medium
above.

11.3 BER

Zero-cross
tive of vol
both of wh
is equal tc
signal are
These zero
An err
shifts out c
has a Gaus
ity of a bit

$P_e = 2\int_{T_w/2}^{\infty}$

The err
for the all
readback w
recording a
is:

threshold detector is defined

$\frac{V_{0-p}}{\ _{rms,\,n}}$. (11.1)

in the peak detection channel,
ecording channel is approxi-
and a standard deviation of
density of the noise voltage

$\left.\frac{n^2}{2\sigma^2}\right)$. (11.2)

errors in the threshold detec-
-to-peak signal amplitude. If
may detect a false transition
zero-to-peak signal voltage.
king a "0" for "1", is given
stribution function:

$\left.-\frac{n^2}{2\sigma^2}\right)dn.$ (11.3)

distribution as the comple-

$\left.y^2\right)dy.$ (11.4)

rfc$(x/\sqrt{2})$,

riance zero-mean Gaussian
at the BER is.

$\left.\right) = Q\left(\frac{\sqrt{SNR}}{2}\right)$. (11.5)

Similarly, a "1" signal will be mistaken as "0" if a negative noise outburst exceeding a value of $-V_{0-p}/2$ occurs. Such a BER can be derived as:

$$P_{1/0} = \int_{-\infty}^{-V_{0-p}/2} \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{n^2}{2\sigma^2}\right\}dn = \frac{1}{2}\,\text{erfc}\left(\frac{\sqrt{SNR}}{2\sqrt{2}}\right) = Q\left(\frac{\sqrt{SNR}}{2}\right). \qquad (11.6)$$

Therefore, the error rate at any bit is a constant if the signal amplitude is stable and the channel noise is Gaussian. The BER is a strong function of the channel SNR. The expected BER at the threshold detector for a channel SNR of 20 dB is 3×10^{-7}, while that for a channel SNR of 24 dB at the threshold detector is $<10^{-15}$. However, the factors not accounted for by the above model, such as the signal amplitude instability and medium defects, may greatly increase the threshold errors predicted above.

11.3 BER AT THE ZERO-CROSSING DETECTOR

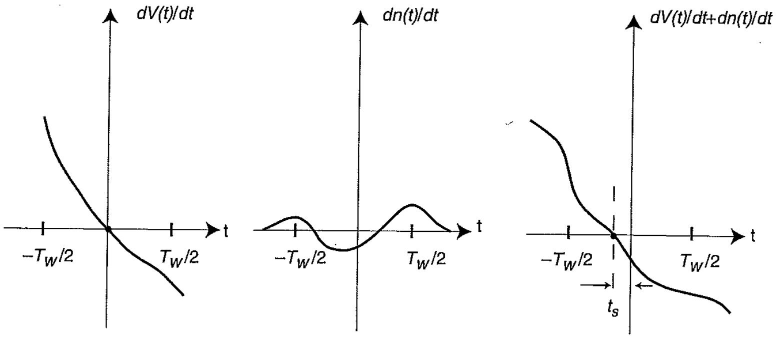
Zero-crossing detector locates the transition by looking at the time derivative of voltage signal. The readback signal $V(t)$ is mixed with noise $n(t)$, both of which are differentiated, so the output signal of the differentiator is equal to $V'(t) + n'(t)$. As a result, the zero-crossing locations in the signal are shifted from the transition locations, as shown in Fig. 11.3. These zero-crossing shifts are also called the *peak-shifts* or *bit-shifts*.[1,2]

An error will occur at the zero-crossing detector if the zero-crossing shifts out of the detection window. Assume that the zero-crossing shift t_s has a Gaussian distribution with a standard deviation σ_t, then the probability of a bit error due to the zero-crossing shift can be calculated as:

$$P_e = 2\int_{T_w/2}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_t}\exp\left\{-\frac{t_s^2}{2\sigma_t^2}\right\}dt_s = \text{erfc}\left(\frac{T_w}{2\sqrt{2}\sigma_t}\right) = 2Q\left(\frac{T_w/2}{\sigma_t}\right). \qquad (11.7)$$

The error rate at the output of the zero-crossing detector can be derived for the all "1"s NRZI pattern, which has an approximately sinusoidal readback waveform: $V(t) = V_0 \sin(\omega t)$, where $\omega = 2\pi f = \pi/T_w$ is the recording angular frequency. The signal at the output of the differentiator is:

$$V^*(t) = V'(t) = \frac{dV(t)}{dt} = \omega V_0 \cos(\omega t). \qquad (11.8)$$



Differentiated Signal

Differentiated Noise

Differentiated Signal + Noise

FIGURE 11.3. Zero-crossing and noise (after differentiation) in peak detection channel. T_w is the detection window or channel bit period.

Therefore, the zero-crossings occur when

$$V^*(t_n) = 0,\ t_n = (2n + 1)T_w/2,\ n = 0, \pm 1, \pm 2, \ldots$$

If the random noise after the differentiator is n^*, then the zero-crossing is shifted by t_s, and they are related by the following:

$$V^*(t_n + t_s) + n^* \approx V^*(t_n) + \left.\frac{dV^*}{dt}\right|_{t_n} t_s + n^* = 0,$$

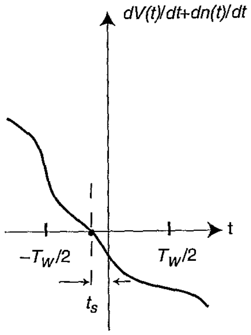
$$n^* \approx -t_s \left.\frac{dV^*}{dt}\right|_{t_n} = -t_s \left.\frac{d^2V}{dt^2}\right|_{t_n}.$$

Calculating the rms value of noise at the differentiator output, we obtain that

$$V^*_{rms,\,n} = \sqrt{\langle (n^*)^2 \rangle} = \left|\frac{d^2V}{dt^2}\right|_{t_n} \sigma_t. \tag{11.9}$$

Combining Equations (11.8), (11.9), and (11.7), the BER at the zero-crossing detector becomes

$$P_e = \mathrm{erfc}\left(\frac{T_w \left|\frac{d^2V}{dt^2}\right|_{t_n}}{2\sqrt{2}V^*_{rms,\,n}} \right) = \mathrm{erfc}\left(\frac{T_w \omega V^*_{0-p}}{2\sqrt{2}V^*_{rms,\,n}} \right) = \mathrm{erfc}\left(\omega T_w \frac{\sqrt{SNR^*}}{2\sqrt{2}} \right). \tag{11.10}$$



Differentiated
Signal + Noise

:rentiation) in peak detection
period.

), ±1, ±2, . . .

n^*, then the zero-crossing
lowing:

$t_s + n^* = 0,$

$\Big|_{t_n}.$

:ntiator output, we obtain

$\sigma_t.$ (11.9)
t_n

e BER at the zero-crossing

$\left(\omega T_w \frac{\sqrt{SNR^*}}{2\sqrt{2}}\right).$ (11.10)

where $V^*_{0-p} = \omega V_0$ is the zero-to-peak signal amplitude after differentiation, and $SNR^* = (V^*_{0-p}/V^*_{rms,n})^2$ is the SNR at the output of the differentiator.

Assume that the noise power spectral density from the readback head is approximately constant (white noise):

$$\eta(\omega) \approx \eta,$$

which is valid within the system bandwidth as determined by the cutoff angular frequency ω_c. The ideal differentiator has the following frequency response:

$$H(i\omega) = i\omega, \qquad \omega < \omega_c.$$

Therefore, the noise power after the differentiation is given by the integral:

$$N^* = (V^*_{rms,\,n})^2 = \int_0^{\omega_c} \omega^2 \eta d\omega = \frac{\omega_c^3}{3}\eta = \frac{\omega_c^2}{3}N, \qquad (11.11)$$

where $N = \omega_c \eta$ is the noise power before differentiation. Since the differentiated signal has an amplitude of ωV_0, the SNR after differentiation is

$$SNR^* = \frac{\omega^2 V_0^2}{\omega_c^2 N/3} = 3\frac{\omega^2}{\omega_c^2} SNR, \qquad (11.12)$$

where $SNR = V_0^2/N$ is the SNR before differentiation. Most of the signal energy is concentrated in the frequency range where $\omega < \omega_c$, so SNR^* (after differentiation) may be smaller than SNR (before differentiation). For example, if $\omega = \omega_c/2.25$, then the SNR is reduced by 2.3 dB due to differentiation. For the all "1"s pattern without modulation encoding, $\omega = \pi/T_w$, so the bit-shift error rate becomes

$$P_e = \text{erfc}\left(\pi\frac{\sqrt{SNR^*}}{2\sqrt{2}}\right) = \text{erfc}\left(2.42\frac{\sqrt{SNR}}{2\sqrt{2}}\right).$$

The argument of the complimentary function is 2.42 times that in Equation (11.6). Therefore, in this case, the bit-shift error rate at the zero-crossing detector is much smaller than the threshold error rate at the threshold detector. In contrast, for the all "1"s pattern with (d, k) encoding, $\omega = \pi/T_w(d + 1)$, i.e., the detection window is now smaller than transition period. As a result, the bit-shift error rate will dominate if $d \geq 2$. In general, the total BER in a peak detection channel can be expressed as

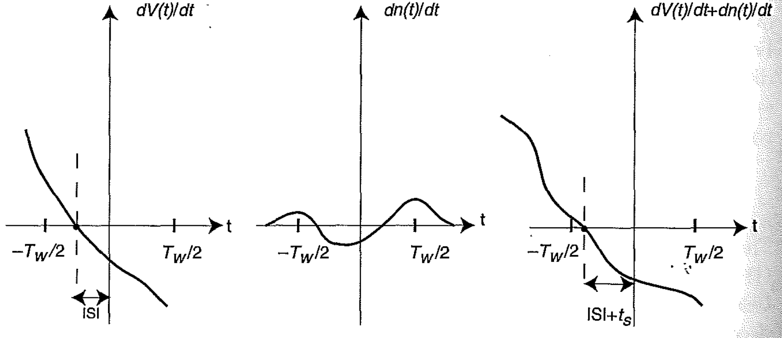
$$P_{e,\text{tot}} = P_{e,\text{bs}} + P_{e,\text{th}} - P_{e,\text{bs}}P_{e,\text{th}} \approx P_{e,\text{bs}} + P_{e,\text{th}},$$

where $P_{e,\text{bs}}$ is the bit-shift error rate at the zero-crossing detector, and $P_{e,\text{th}}$ is the threshold error rate at the threshold detector, both of which must be much smaller than 1.

It must be cautioned that SNR is not the only factor that affects the error rate at the zero-crossing detector. Both liner intersymbol interference (ISI) and nonlinear transition shift (NLTS) can cause peak-shifts. In this case, the peaks and the zero-crossings in the readback signal are shifted from the *desired* transition locations, which should be at the center of the detection window. Furthermore, noise is mixed to the distorted signal, so the zero-crossings may be shifted even more from the center of the detection window, as illustrated in Fig. 11.4. Consequently, the BER at the zero-crossing detector increases. Based on Equation (11.7), the bit-shift BER taking ISI and NLTS into consideration can be expressed as follows:

$$P_{e,\text{bs}} = \frac{1}{2}\operatorname{erfc}\left(\frac{T_w/2-\Delta}{\sqrt{2}\sigma_t}\right) + \frac{1}{2}\operatorname{erfc}\left(\frac{T_w/2+\Delta}{\sqrt{2}\sigma_t}\right),$$

where $\pm\Delta$ is the net bit-shift, and we assumed that the peak has equal probabilities to shift early or late. Note that peak detection channels are usually used with (1,7) or (2,7) RLL codes, so the transition separations are relatively long. In other words, for peak detection channels linear ISI is more significant than NLTS. The latter is a critical factor in PRML channels (Chapter 12).



Differentiated Signal

Differentiated Noise

Differentiated Signal + Noise

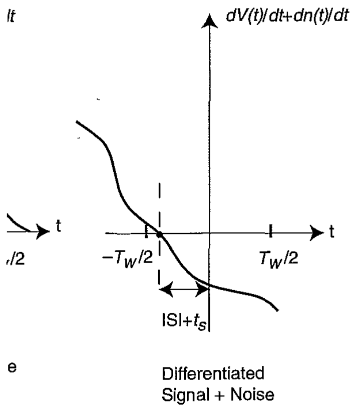
FIGURE 11.4. Bit-shifts caused by intersymbol interference (ISI) and noise in peak detection channel.

the zero-crossing detector, and reshold detector, both of which

t the only factor that affects the th liner intersymbol interference S) can cause peak-shifts. In this the readback signal are shifted ch should be at the center of the mixed to the distorted signal, so ore from the center of the detec- onsequently, the BER at the zero- quation (11.7), the bit-shift BER can be expressed as follows:

$$\frac{1}{2}\mathrm{erfc}\left(\frac{T_w/2+\Delta}{\sqrt{2}\sigma_t}\right),$$

ssumed that the peak has equal that peak detection channels are es, so the transition separations eak detection channels linear ISI ter is a critical factor in PRML



ıbol interference (ISI) and noise in

11.4 WINDOW MARGIN AND BIT-SHIFT DISTRIBUTION

The output of the zero-crossing detector of a peak detection channel produces sharp pulses at the locations where the zero-crossings are detected, as shown in Fig. 11.5. As we have discussed in the previous section,

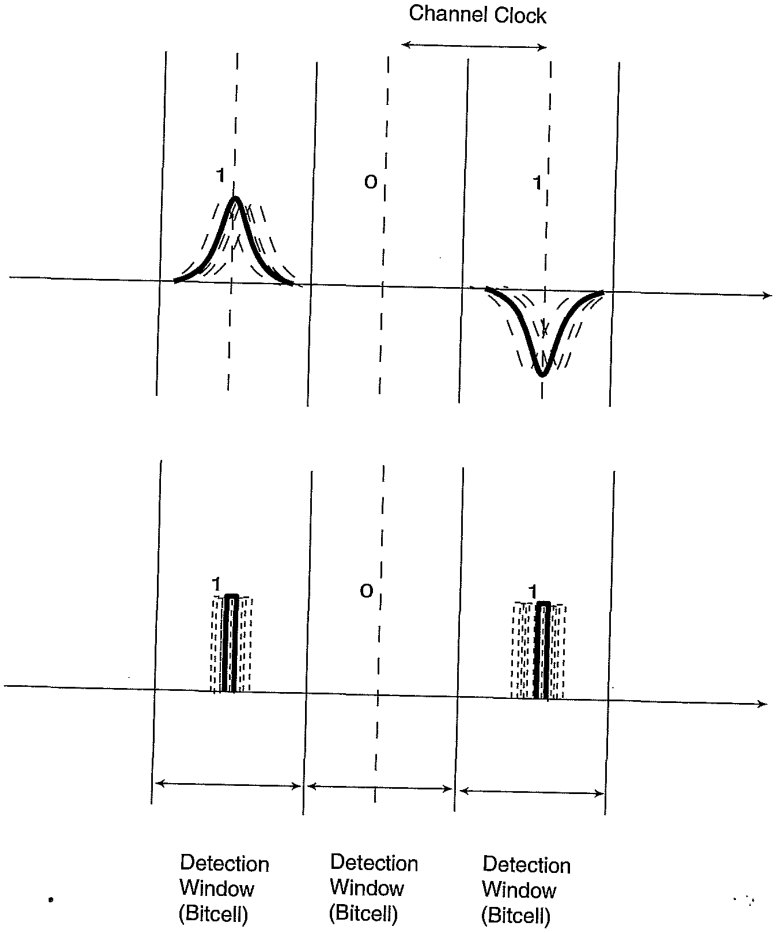

Channel Clock
1
0
1
1
0
1
Detection Window (Bitcell)
Detection Window (Bitcell)
Detection Window (Bitcell)

FIGURE 11.5. The distribution of the readback pulses (top) and the output of zero-crossing detector (bottom).

an error occurs when the corresponding pulse falls outside the detection window. The error rate at the output of the zero-crossing detector can be measured by counting the pulses falling outside the prescribed detection window. However, the error rate of an actual magnetic recording channel tends to be very low, so the pulse counting measurement may take a long time. In fact, the raw BER of a magnetic disk drive is typically $\sim 10^{-9}$. At such a low probability level, more than 10^{10} bits of data should be collected and analyzed to obtain reliable statistics. This would require $>10^4$ disk revolutions and at least several minutes of measurement time for each experimental condition, just to capture the data bits.

An effective and fast method for evaluating error performance is based on the window margin analysis.[3] To understand the principle of window margin, let us imagine that we are able to measure the exact position of each pulse at the output of the zero-crossing detector and to accumulate the histogram of such positions, as shown in Fig. 11.6. The height of the histogram $H(t_k)$ at each timing position $t = t_k$ corresponds to the total number of pulses having a timing shift of t_k. The sum of $H(t_k)$ equals to the total number of detected pulses.

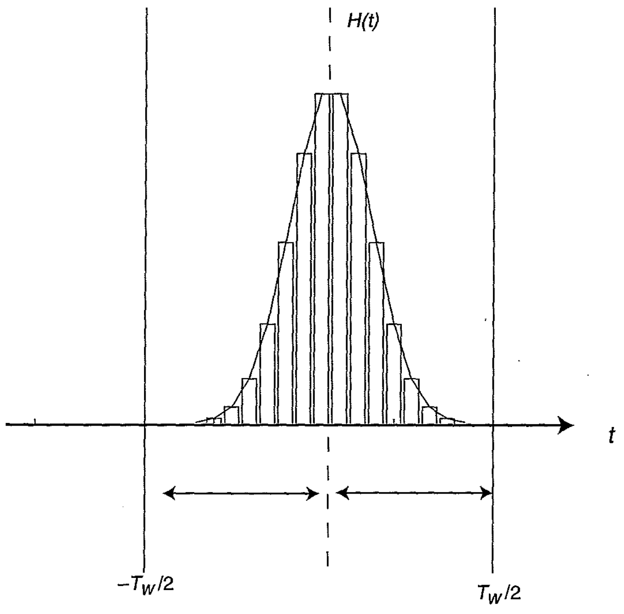
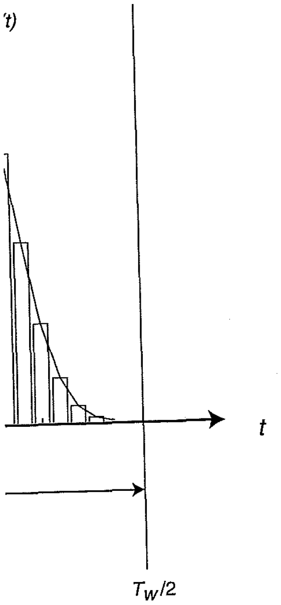


FIGURE 11.6. Histogram of peak shifts at the output of zero-crossing detector.

ulse falls outside the detection
e zero-crossing detector can be
utside the prescribed detection
al magnetic recording channel
measurement may take a long
sk drive is typically ~10^{-9}. At
bits of data should be collected
This would require >10^4 disk
of measurement time for each
e data bits.

aluating error performance is
To understand the principle of
are able to measure the exact
e zero-crossing detector and to
ons, as shown in Fig. 11.6. The
ng position $t = t_k$ corresponds
ing shift of t_k. The sum of $H(t_k)$
ulses.

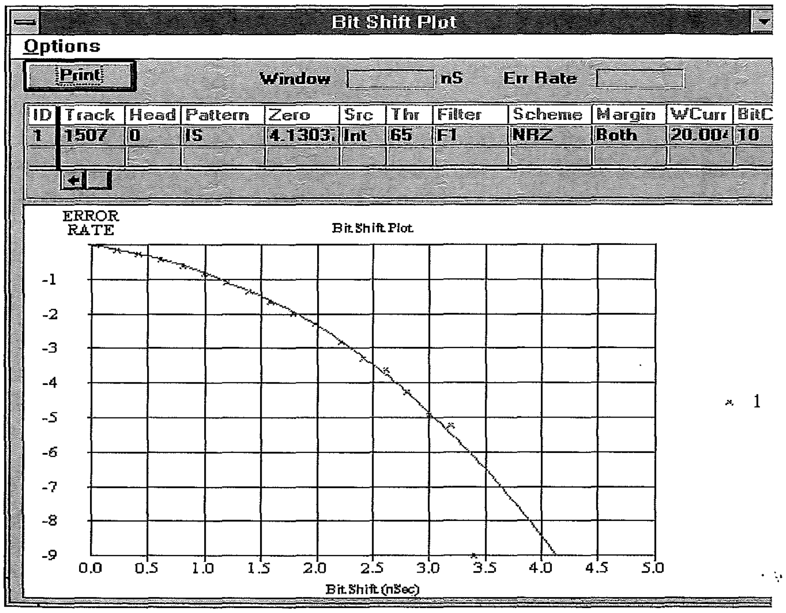


he output of zero-crossing detector.

Once we know the histogram, the number of pulses with a bit-shift value larger than t_s, $N(t_s)$, can be expressed as the sum of $H(t_k)$ for $|t_k| > t_s$:

$$N(t_s) = \sum_{t_k>t_s} H(t_k) + \sum_{t_k<-t_s} H(t_k),$$

Obviously, the total number of pulses in the histogram is $N_{\text{tot}} \approx N(t_s = 0)$. The larger the t_s value, the smaller the $N(t_s)$ value. If we plot $N(t_s)/N_{\text{tot}}$ vs. t_s in a logarithmic scale, we get a *bit-shift plot* as shown in Fig. 11.7. The horizontal axis of this plot represents the bit-shift value (in nanoseconds), and the *half detection window* (or *half timing window*) equals to 5 ns. When the pulse from the output of the zero-crossing detector deviates more than 5 ns from its nominal position, i.e., the center of the detection window, the peak detection channel makes an error. If the half detection window is chosen to be $t_s < 5$ ns, then $N(t_s)/N_{\text{tot}}$ is the corresponding bit error rate. In this sense, the bit-shift plot shows the logarithmic BER as a function of the half timing window.



Window [ ] nS Err Rate [ ]

ID	Track	Head	Pattern	Zero	Src	Thr	Filter	Scheme	Margin	WCurr	BitC
1	1507	0	IS	4.1303	Int	65	F1	NRZ	Both	20.00	10

FIGURE 11.7. An example of a bit-shift plot.

The great advantage of a bit-shift plot is that it can be acquired very fast. The bit-shift plot is usually a smooth curve. Therefore, even though the actual measurement points stop at a BER of $\sim 10^{-5}$, these points can be curve-fitted and then extrapolated to a much lower BER as shown by a solid line in Fig. 11.7. It is predicted from the curve fitting that the BER is 10^{-9} if the half timing window is 4.1 ns. In other words, if we specify a BER of 10^{-9} and a half timing window of 5 ns, there is still 0.9 ns left to accommodate any additional error sources, which is often called the *timing window margin*. The timing window margin is often normalized by the half detection window, so it becomes 0.9/5 = 18% in this case. In short, the timing window margin is defined as the percentage of the half timing window (or half detection window) which is left at a specified BER level.

A parabolic curve fitting method is often used in bit-shift plots. This is based on the following mathematical approximation (for $x \gg 1$):

$$\mathrm{erfc}(x) \approx e^{-x^2}/(\sqrt{\pi}\,x),$$
$$\ln[\mathrm{erfc}(x)] \approx -x^2 - \ln(\sqrt{\pi}\,x) \approx -x^2.$$

It follows from Equation (11.7) that

$$\ln\left[\frac{N(t_s)}{N_{\mathrm{tot}}}\right] = \ln\left[\mathrm{erfc}\left(\frac{t_s}{\sqrt{2}\sigma_t}\right)\right] = -\frac{t_s^2}{2\sigma_t^2},$$

which is valid if only Gaussian noise exists at the zero-crossing detector. Obviously, the larger the standard deviation (σ_t), which measures the random peak shifts, the slower the BER drops. A small noise will result in a sharply descending bit-shift plot.

Figure 11.8 demonstrates a useful representation of a bit-shift plot: the logarithmic BER vs. the percentage of the half timing window for different SNR ratios. If SNR = 23 dB, the 50% timing window margin is reached at an error rate of 10^{-6}. In comparison, if *SNR* = 25 dB, the 50% margin is reached at an error rate of 10^{-10}. Both curves predict error rates well below 10^{-12} at the full half detection window (0% timing window margin). Generally speaking, if the same timing window margin is required, a higher SNR tends to give a much lower BER.

A simple and practical way to measure the bit-shift distribution of a peak detection channel is to vary the detection window of its zero-crossing detector and to count the actual number of pulses detected outside a given detection window. The number of error bits divided by the total number of bits counted is the BER. In a real disk drive, the bit-shift distribution reflects complicated interactions among noises, linear ISI, and nonlinear distortions.[2,3] Since the generation of bit-shift plot is fast, it is

t is that it can be acquired very
ı curve. Therefore, even though
BER of ~10^{-5}, these points can
ı much lower BER as shown by
n the curve fitting that the BER
s. In other words, if we specify
of 5 ns, there is still 0.9 ns left
ırces, which is often called the
v margin is often normalized by
).9/5 = 18% in this case. In short,
he percentage of the half timing
h is left at a specified BER level.
ften used in bit-shift plots. This
ıpproximation (for $x \gg 1$):

$$\sqrt{\pi}\, x) \approx -x^2.$$

$$\frac{s}{2\sigma_t}\Big)\Big] = -\frac{t_s^2}{2\sigma_t^2},$$

sts at the zero-crossing detector.
ıation (σ_t), which measures the
drops. A small noise will result

epresentation of a bit-shift plot:
of the half timing window for
e 50% timing window margin is
arison, if SNR = 25 dB, the 50%
^{0}. Both curves predict error rates
on window (0% timing window
ıe timing window margin is re-
ıch lower BER.
ure the bit-shift distribution of a
ction window of its zero-crossing
ıer of pulses detected outside a
f error bits divided by the total
ı a real disk drive, the bit-shift
ions among noises, linear ISI, and
ation of bit-shift plot is fast, it is

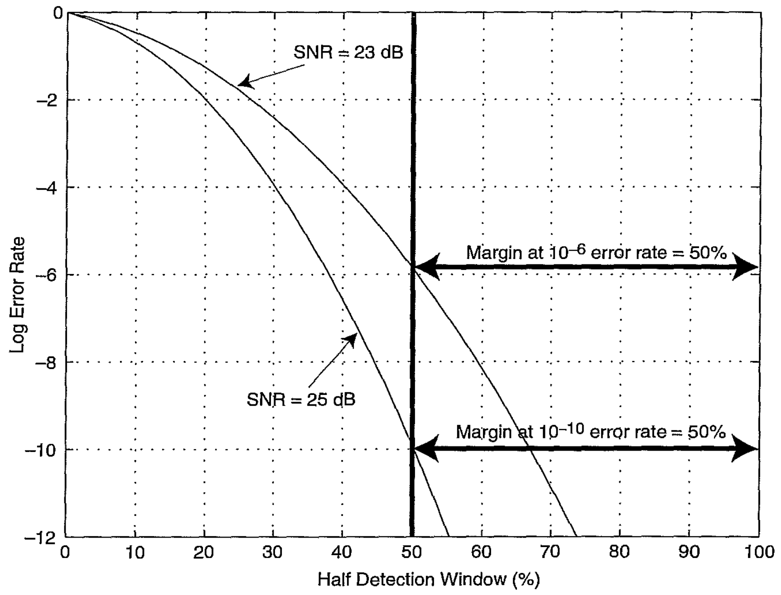


FIGURE 11.8. The dependence of the 50% window margin on system SNR.

an efficient tool for testing magnetic recording components and channel performance.

As mentioned previously, the slope of the bit-shift plot reflects the noise level. A small noise leads to a steep bit-shift plot. Conversely, a large noise level results in a relatively flat bit-shift plot. Peak-shifts due to ISI typically cause a flat part in bit-shift plot. The flat part arises from the fact the bit error rate increases as a result of peak-shifts, as shown in Fig. 11.4. In other words, a larger detection window is required to achieve the same BER, so the bit-shift plot is moved to the right with respect to that without ISI, as shown in Fig. 11.9. The leftmost curve is obtained from a data pattern consisting of isolated transitions with the read head at the nominal on-track position. In this case, the time margin at an error rate of 10^{-9} is 50%. The middle curve is generated for the same pattern but with the read head in an off-track position. When the head is shifted away from the track center, the readback signal amplitude drops, and at the same time more medium noise is read from adjacent medium regions. Therefore, the off-track signal has a lower SNR (Chapter 14), and this curve is less steep. Now the time margin at an error rate of 10^{-9} becomes

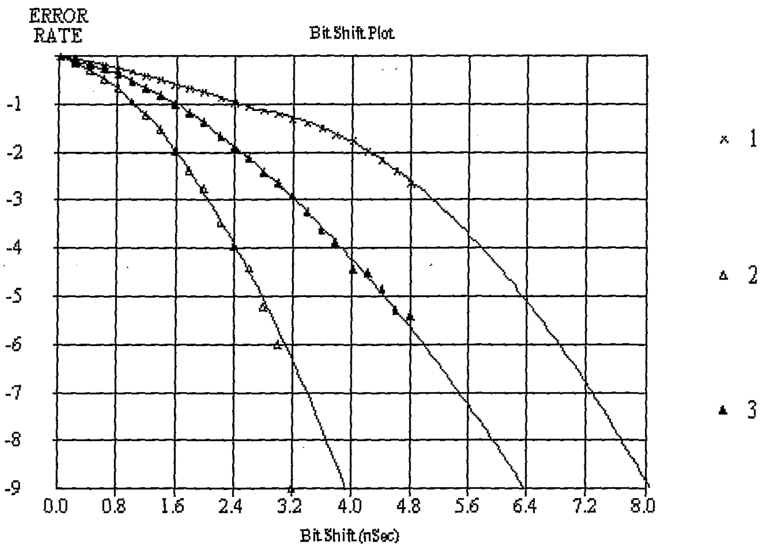
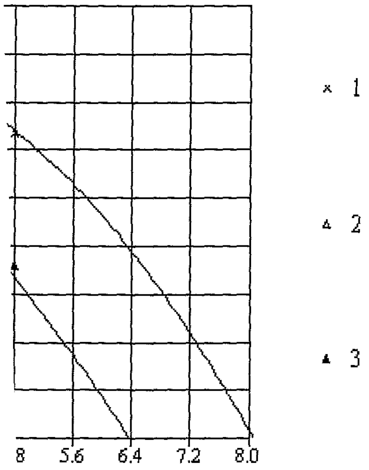


FIGURE 11.9. Examples of bit-shift plots. Left: isolated transition pattern and on-track reading; middle: same pattern in off-track position; right: data pattern with adjacent transitions.

(8.0 ns − 6.4 ns)/8.0 ns = 20%. The rightmost curve is obtained for a data pattern consisting of both isolated and adjacent transitions. A certain number of pulses are now shifted from their nominal positions, resulting in a flat part in the curve. This is the least desirable case among the three, in which there is no time margin left if the required BER is 10^{-9}, i.e., the actual error rate of the peak detector will be $\sim 10^{-9}$.

The last curve in Fig. 11.9 can be further illuminated by Fig. 11.10. The peak-shift due to ISI interacts with the random noise and the histogram of the bit-shift distributions for adjacent transitions are shifted ±2.4 ns from that of isolated transitions. The distribution of isolated transitions has an amplitude of ~10 times higher, meaning that approximately one out of every 10 transitions in this pattern is shifted. The initial part of curve 1 in Fig. 11.9 (until about 2.4 ns bit-shift) is slowly decreasing as the error events are dominated by those of isolated transitions. When the error rate level of $\sim 10^{-1}$ is reached at a bit-shift of about 2.4 ns, the error events are caused by both isolated and adjacent transitions. Consequently, the curve has a nearly flat part just beyond 2.4 ns as the peak shifts of adjacent transitions slow down the decrease of the BER.



ft: isolated transition pattern and on-
ıck position; right: data pattern with

nost curve is obtained for a data
adjacent transitions. A certain
ıeir nominal positions, resulting
: desirable case among the three,
ıe required BER is 10^{-9}, i.e., the
l be $\sim 10^{-9}$.
er illuminated by Fig. 11.10. The
dom noise and the histogram of
ısitions are shifted ±2.4 ns from
ɔn of isolated transitions has an
that approximately one out of
ted. The initial part of curve 1
slowly decreasing as the error
transitions. When the error rate
about 2.4 ns, the error events
transitions. Consequently, the
ns as the peak shifts of adjacent
BER.

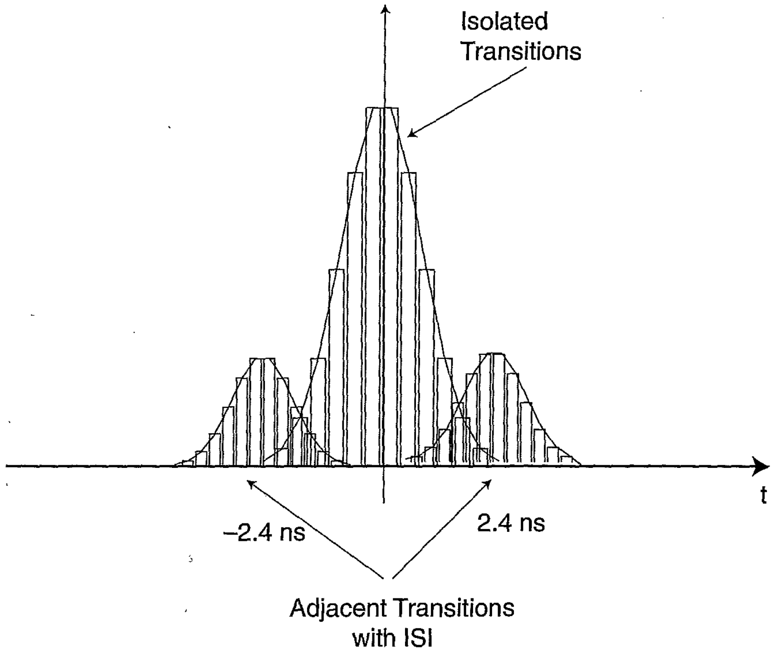


FIGURE 11.10. Bit-shift histogram for the data pattern corresponding to curve 1 (rightmost) shown in Fig. 11.9.

References

1. A. S. Hoagland and J. E. Monson, *Digital Magnetic Recording,* (New York, J. Wiley & Sons, 1991).
2. P. H. Siegel and J. K. Wolf, "Modulation and coding for information storage," *IEEE Communications Magazine,* December 1991, 12, 68–86.
3. E. Katz and T. Campbell, "Effect of bitshift distribution on error rate in magnetic recording," *IEEE Trans. Magn.,* **15**, 1050, 1979.

CHAPTER 12

PRML Channels

When recording densities are low, each transition written on a magnetic medium results in a relatively isolated voltage peak and a peak detection channel works well to recover written information. At high channel densities, however, the peak detection channel can not provide reliable data detection. Superposition of pulses (linear ISI) shifts peaks of readback signal and increases the probability of errors in the zero-crossing detector. At the same time, the signal amplitude is lowered at high densities (roll-off) and the errors in the threshold detector also increase.

The partial-response maximum likelihood (PRML) channels were proposed to overcome the problem of linear ISI and are now the dominant detection schemes in commercial magnetic disk drives.[1–7] The PRML detection method is not based on voltage peaks, but rather takes into account the fact that the signals from adjacent transitions interfere. PRML channels consist of two relatively independent parts: partial response (PR) equalization and maximum likelihood (ML) detector, which will be discussed in this chapter.

12.1 PRINCIPLE OF PARTIAL RESPONSE AND MAXIMUM LIKELIHOOD

The basic idea of *partial response* is to introduce some controlled amount of ISI into the data pattern rather than trying to eliminate it. The idea was first introduced in the field of digital communication and has proven to be a very powerful concept for magnetic information storage as well. It turns out that magnetic recording channels can be transformed into PR channels which satisfy two fundamental properties: (1) the superposition of voltage pulses from adjacent transitions is linear; (2) the shape of the readback signal from an isolated transition is exactly known and determined. A block diagram of a typical PRML channel is shown in Fig. 12.1.[7] It consists of a variable-gain amplifier (VGA), an analog equalizer, an analog-to-digital converter (ADC), a digital equalizer, an ML detector, and a clock/gain recovery circuit. The circuit blocks (except the ML detector) transform the readback signal into the partial response signal as required.

The analog readback signal from the magnetic head should have a certain and constant level of amplification. Any variation in isolated readback peaks is compensated with the VGA, which gets a control signal from the clock and gain recovery loop.

A PR channel operates within a certain bandwidth, meaning that the spectral components beyond the bandwidth have to be cut off. This is done with the continuous time filter or analog equalizer. The other function sometimes performed by the analog equalizer is to modify the frequency response of the channel. The modification of the frequency response is sometimes required to adjust the shape of the readback signal from the head. For example, it may be necessary to adjust the pulse width to make it proportional to the distance between transitions. The analog equalizer is implemented as a linear filter with a programmable frequency response including a variable cutoff frequency and boost. The analog signal at the equalizer output generally has a slightly different shape than the unmodified signal directly from the head.

The signal from the analog equalizer is sampled (or digitized) with the ADC. The sampling is initiated by a clock signal at the rate of exactly one sample per channel bit period. The frequency and phase of the clock

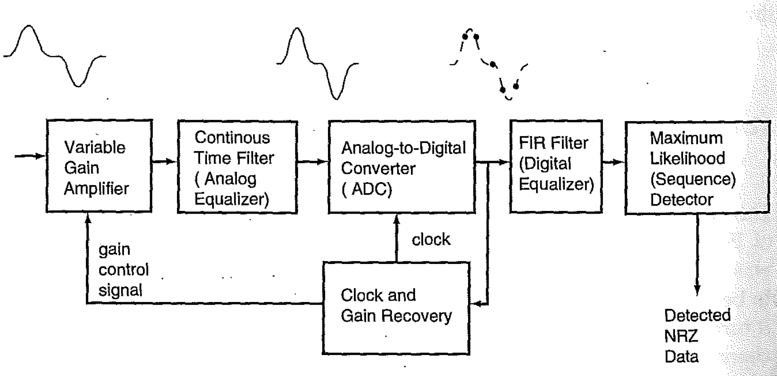
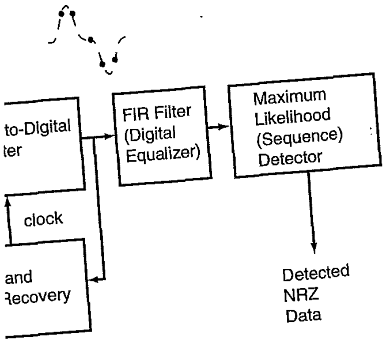

Variable Gain Amplifier
Continous Time Filter (Analog Equalizer)
Analog-to-Digital Converter (ADC)
FIR Filter (Digital Equalizer)
Maximum Likelihood (Sequence) Detector
gain control signal
clock
Clock and Gain Recovery
Detected NRZ Data

FIGURE 12.1. Block diagram of typical PRML channel.

RML channel is shown in Fig.
r (VGA), an analog equalizer,
ital equalizer, an ML detector,
t blocks (except the ML detec-
he partial response signal as

magnetic head should have a
Any variation in isolated read-
A, which gets a control signal

n bandwidth, meaning that the
th have to be cut off. This is done
g equalizer. The other function
lizer is to modify the frequency
on of the frequency response is
of the readback signal from the
o adjust the pulse width to make
ransitions. The analog equalizer
rogrammable frequency response
and boost. The analog signal at
lightly different shape than the
ad.
er is sampled (or digitized) with
clock signal at the rate of exactly
frequency and phase of the clock



l PRML channel.

signal is adjusted by the clock recovery loop. The signal at the ADC output constitutes a stream of digital samples or discrete-time data. Digital samples are often processed (filtered) by an additional digital filter. This digital filtering operation can improve the quality of analog equalization. The principle of the finite impulse response (FIR) filter will be discussed in Section 12.2.

The samples at the ADC output are used to detect the presence of transitions in the readback signal. If the signal quality is good, a simple threshold detector can be used to compare sample values to a threshold. However, much more reliable detection at high recording densities can be achieved with an ML detector. Unlike the peak detection channel, the PRML channels do not assume that the readback signal should contain relatively narrow peaks. The decisions of PRML channels are based on a sequence of the ADC samples of the signal, which are not necessarily taken at the signal peak level.

The partial response equalization and the ML detector are at the heart of a PRML channel. We will introduce their principles here, but defer the details to Sections 12.2 and 12.4.

12.1.1 PR4 channel

A special class of partial-response channel, Class IV partial response (PR4) system, is the first widely used PR channel. The isolated pulse shape in a PR4 system is shown in Fig. 12.2, where T is the channel bit period, and the transition is written at time instant $t = 0$. The sample values at integer number of bit periods before the transition are exactly zeroes. However at $t = 0$ and at $t = T$, the sample values of the pulse are equal to 1. The pulse of voltage reaches its peak amplitude of 1.273 at one-half of the bit period.

The samples of the isolated PR4 pulse shown in Fig. 12.2 at the output of ADC will be . . . .00011000. . . . Of course, value "1" is used for convenience and in reality it corresponds to a certain ADC level, which may be a number between 0 and 2^n-1, where n is the number of bits in ADC. The fact that the isolated transition has two nonzero samples, one at the transition location and the other at the next transition location is very important. If the next transition is written, the pulses will interfere. However, the other sample values are zero, so the interference are easily predictable.

A dibit is formed when the second transition is written immediately after the first one (i.e., one channel period later), which results in a dipulse

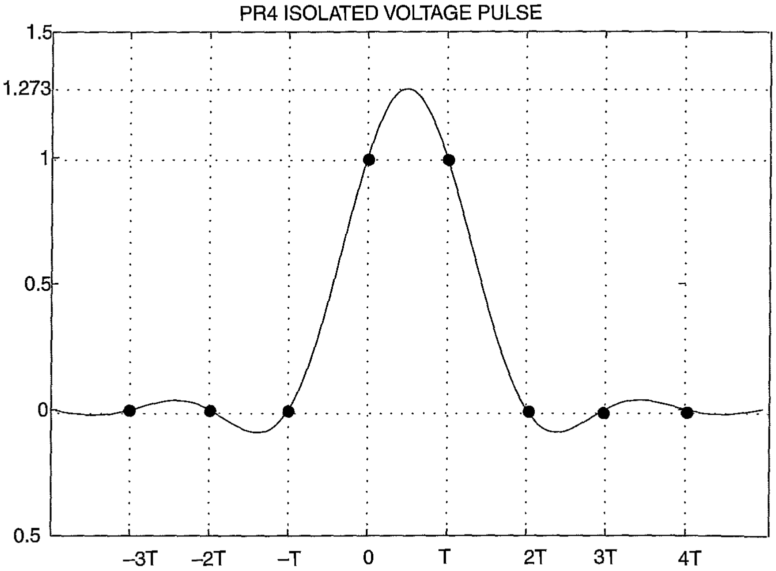


FIGURE 12.2. Shape of isolated pulse in PR4 system.

response shown in Fig. 12.3. It is generated by the linear superposition of voltages from two isolated pulses with opposite polarity:

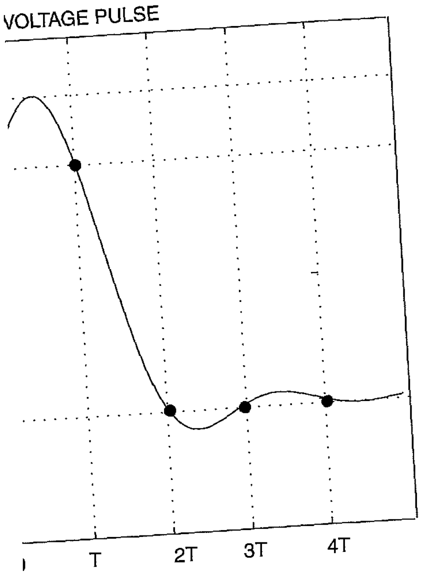
$$\begin{array}{cl} & 0\,0\,0\,1 \quad 1 \quad 0\,0\,0 \quad \text{—from the first transition} \\ + & \underline{0\,0\,0\,0 \;\; -1 \;\; -1\,0\,0} \quad \text{—from the second transition} \\ = & 0\,0\,0\,1 \quad 0 \;\; -1\,0\,0 \end{array}$$

The samples of a dibit are {. . . ,0,0,1,0,−1,0,0, . . .}.

What happens if we have three consecutive transitions (tribit)? It is easy to check that the answer is {. . . ,0,0,1,0,0,1,0,0, . . .}. Another useful pattern is the low-frequency square recording pattern consisting of transitions two channel periods apart. Obviously, the sequence of resulting samples is {. . . ,+1,+1,−1,−1,+1,+1,−1,−1,+1,+1, . . .}.

From the sequence of samples from the ADC output we can easily reconstruct any data patterns that are written on the medium. If the data are in NRZ form (i.e., 1 and 0 represents positive and negative medium magnetization, respectively, then the current value $a(k)$ of the data pattern is the sum of the current sample $s(k)$ and the bit two channel periods earlier:

$$a(k) = s(k) + a(k-2). \tag{12.1}$$



in PR4 system.

generated by the linear superposition
es with opposite polarity:

—from the first transition
—from the second transition

,1,0,−1,0,0, . . .}.
e consecutive transitions (tribit)? It is
. . . ,0,0,1,0,0,1,0,0, . . .}. Another useful
recording pattern consisting of transi-
Obviously, the sequence of resulting
-1,−1,−1,+1,+1, . . .}.
s from the ADC output we can easily
are written on the medium. If the data
resents positive and negative medium
he current value $a(k)$ of the data pattern
$s(k)$ and the bit two channel periods

$(k) + a(k-2)$. (12.1)

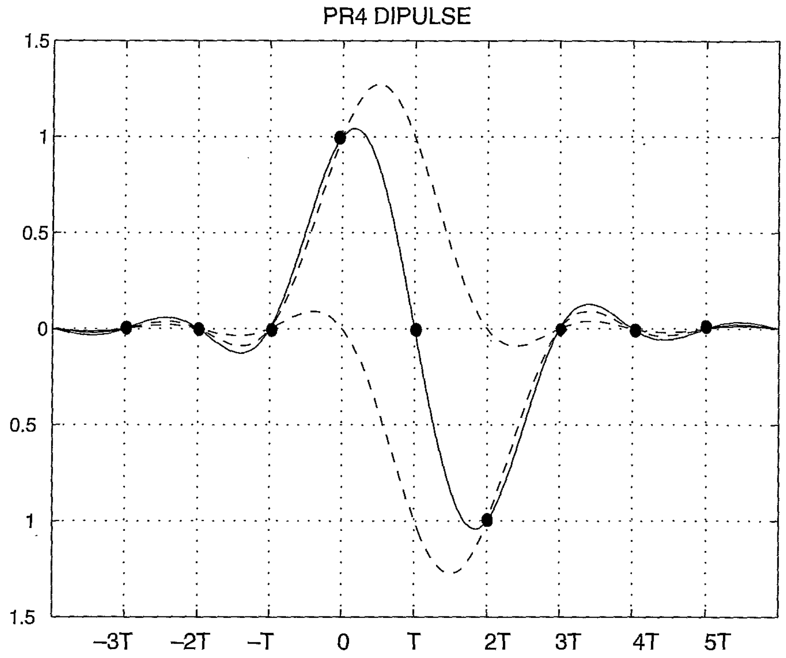


FIGURE 12.3. A dipulse response.

Applying Equation (12.1) to an isolated transition:

Samples in PRML system:	. . . 0 0 0 0 1 1 0 0 0 0 0 . . .
Recovered NRZ data:	. . . 0 0 0 0 1 1 1 1 1 1 1 . . .

For a tribit:

Samples in PRML system:	. . . 0 0 0 0 1 0 0 1 0 0 0 . . .
Recovered NRZ data:	. . . 0 0 0 0 1 0 1 1 1 1 1 . . .

The last example is especially interesting: the second pulse in the tribit has zero samples, almost completely suppressed by the first and the third transitions due to linear superposition. However, we can still easily recover the data based on the samples. Therefore, once the pulses are reduced to a "standard" shape, the data pattern is easily recovered because the superposition of signals from adjacent transitions is known. In the last example, we know that sample "1" of a transition is canceled by "−1" of the next transition. Since a positive voltage pulse is always followed by a negative pulse, and vice versa, arbitrary linear superposition of the samples of isolated pulses leads to only three possible values: {−1, 0, +1}. Therefore, ADC output in the PR4 channel consists of only three distinct sample levels: {−1, 0, +1}. In reality, however, the sample values will deviate from the nominal values because of noises, NLTS, etc.

A practical way to characterize the quality of the PR4 channel is to analyze the statistics of samples at the ADC output. If all parts of the channel are working properly, the ADC samples should take only nominal values {−1,0,+1}. As a result, a histogram of sample levels will consist of three distinct peaks, as shown in Fig. 12.4 (left). However, the presence of noise, NLTS, and nonideal equalization leads to the distortions of the sample values, and the three peaks of sample values may overlap one another, as shown in Fig. 12.4 (right).

Another way to characterize the quality of a PRML channel is to analyze the so-called "eye diagrams" or "eye patterns" generated by an oscilloscope. To obtain an eye pattern, a random data pattern is written on the disk. If the clock signal of the channel is input to the oscilloscope synchronization (trigger) and the analog equalizer output is taken as the oscilloscope signal, the superposition of random equalized waveforms will be observed on the oscilloscope screen. Since all these waveforms are synchronized to the channel clock, an interesting "focusing" pattern is observed: all the waveforms at clock points pass through the three points corresponding to {−1,0,+1}, as shown in Fig. 12.5.

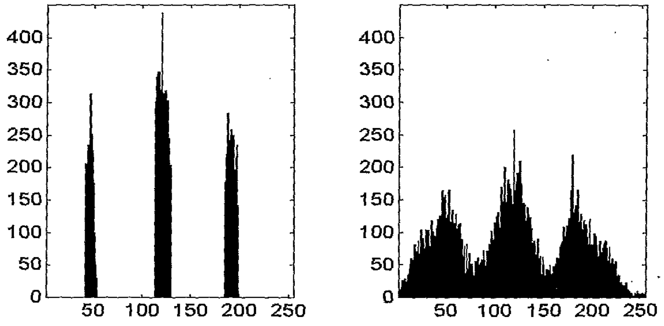
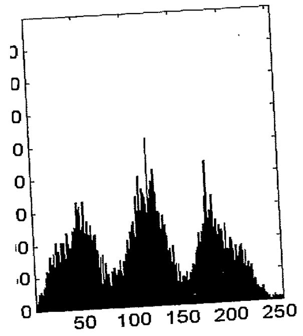


FIGURE 12.4. Sample value distributions for good (left) and poor (right) PR4 signal quality, respectively.

sitive voltage pulse is always
, arbitrary linear superposition
ıly three possible values: {−1,
channel consists of only three
y, however, the sample values
use of noises, NLTS, etc.
ıality of the PR4 channel is to
DC output. If all parts of the
ıples should take only nominal
of sample levels will consist of
: (left). However, the presence
leads to the distortions of the
mple values may overlap one

lity of a PRML channel is to
eye patterns" generated by an
andom data pattern is written
nel is input to the oscilloscope
qualizer output is taken as the
random equalized waveforms
en. Since all these waveforms
. interesting "focusing" pattern
points pass through the three
own in Fig. 12.5.



ɔr good (left) and poor (right) PR4

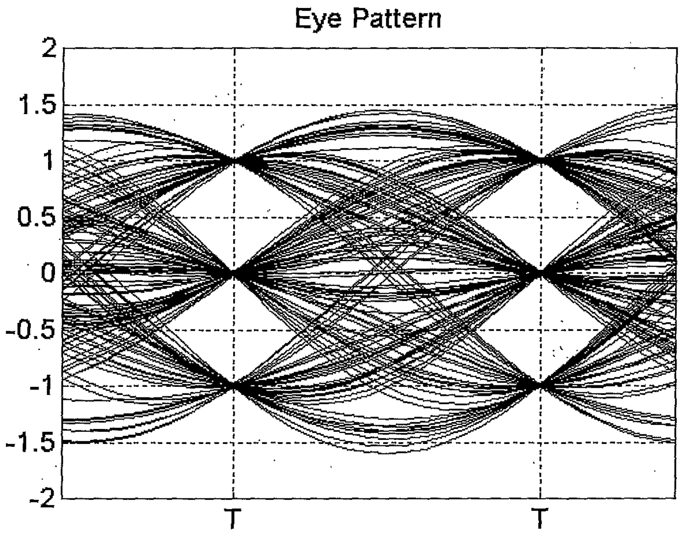


FIGURE 12.5. An ideal eye pattern for PR4 system.

If the sample distributions do not overlap, or equivalently, the "eyes" in the eye diagram are open, signal detection can be done using a simple comparator (threshold detector). In reality, however, the samples do overlap as shown in Fig. 12.4. In this case, the ML detector is needed to achieve reliable data recovery. The gain of the ML detector over a simple comparator leads to much lower error rates, typically by about three or more orders of magnitude. While it is possible that the ML detector will still be able to decode the pattern in the case of strongly overlapping samples, the probability of errors will be much lower if the histogram is nonoverlapping. We will examine the principle of the ML detector next.

12.1.2 Maximum likelihood detector

Samples at the output of ADC ideally have a small number of levels, such as {−1,0,+1} for the PR4 system. A threshold detector compares the current sample value to an amplitude threshold and immediately decides what the ideal sample value should be. For example, if sample >0.5 then ideal sample = 1, if sample <−0.5 then ideal sample = −1, if −0.5 < sample

< 0.5 then ideal sample = 0. For a stream of noisy samples such as {0.8 0.3 −0.7 −0.2 0.6 0.9 1.1 0.2}, the threshold detector output is {1 0 −1 0 1 1 1 0}. However, sequence "111" can not exist in the PR4 channel. Sequence "11" corresponds to an isolated voltage pulse, so the next transition should be of opposite polarity, meaning that the possible sequences are "1100", "1001 "11−1−1", etc., but not "1110". Apparently an error has occurred.

Unlike the threshold detector, the ML detector does not make immediate decisions on whether the incoming ADC sample is +1, 0, or −1. Instead, it analyzes a sequence of samples and then chooses the most probable sequence. Therefore, the ML detector is also called the sequence detector, or Viterbi detector after its original inventor. ML detector recognizes that sequence "111" is forbidden, and tries to find the most probable ideal sequence which matches the samples. In the above example, there are several allowable sequences: {1 0 −1 0 1 1 0 0}, {1 0 −1 0 0 1 1 0} or {1 0 −1 0 0 0 1 1}. These sequences can be compared with the received sequence of samples in order to choose the most probable sequence:

0.8	0.3	−0.7	−0.2	0.6	0.9	1.1	0.2	Samples
1	0	−1	0	1	1	0	0	Sequence #1
1	0	−1	0	0	1	1	0	Sequence #2
1	0	−1	0	0	0	1	1	Sequence #3

Sequence #1 takes 0.6 and 0.9 as "1", and 1.1 as "0"; sequence #2 takes 0.6 as "0", and 0.9 and 1.1 as "1"; sequence #3 takes 0.6 and 0.9 as"0", and 1.1 and 0.2 as "1". It appears that sequence #2 is the most probable. We can verify our intuition by calculating the *mean-squared distance* (MSD) between the samples $s(k)$ and the assumed sequence $b(k)$:

$$MSD = \sum_{k=1}^{N} [s(k) - b(k)]^2 .$$

Sequence #1: $MSD = 1.68$

Sequence #2: $MSD = 0.68$

Sequence #3: $MSD = 2.08$

Sequence #2 is the closest to the sequence of detected samples and at the same time is allowed by the PR4 channel. In other words, sequence #2 is the *most likely* among all the candidate sequences, so the ML detector takes it as the sequence that was transmitted through the channel.

The principle of ML detection can be summarized as follows:

1. Decisions are made based on a sequence of samples, instead of one current sample.

of noisy samples such as {0.8
l detector output is {1 0 −1 0
ıot exist in the PR4 channel.
ıltage pulse, so the next transi-
ıg that the possible sequences
1110". Apparently an error has

etector does not make immedi-
ADC sample is +1, 0, or −1.
es and then chooses the most
ector is also called the sequence
al inventor. ML detector recog-
l tries to find the most probable
es. In the above example, there
0 1 1 0 0}, {1 0 −1 0 0 1 1 0} or
be compared with the received
he most probable sequence:

l 0.2	Samples
0	Sequence #1
0	Sequence #2
1	Sequence #3

ıd 1.1 as "0"; sequence #2 takes
ence #3 takes 0.6 and 0.9 as"0",
equence #2 is the most probable.
g the *mean-squared distance* (MSD)
ıed sequence $b(k)$:

$- b(k)]^2$

$SD = 1.68$

$SD = 0.68$

$SD = 2.08$

ıce of detected samples and at the
ıel. In other words, sequence #2 is
te sequences, so the ML detector
ımitted through the channel.
be summarized as follows:
a sequence of samples, instead of

2. For each sequence of samples, a list of allowable sequences is generated.
3. Each of the allowable sequences is compared with the received sequence of samples and respective *MSD* (or another appropriate distance function) is calculated. The sequence with the minimum distance, i.e., the *maximum likelihood*, is selected to be the result of the detection.
4. Decisions of the ML detector are always made with some delay.

12.2 PARTIAL RESPONSE

Having introduced the basics of the PRML channels, now let us look at the PR channel (excluding the ML detector) in more detail.

12.2.1 PR4 channel bandwidth and frequency response

The presence or absence of a transition in a PRML channel is decided based upon the numerical values of the samples. The samples are taken only once per channel period, and the minimum separation between adjacent transitions equals to the channel period. One cannot help asking whether the sampling frequency in the PRML channel is high enough to recover the recorded transition.

The *sampling theorem* (Nyquist theorem)[3] states that any band-limited analog signal with a cutoff frequency of $f_{\max}$ can be uniquely recovered from its discrete samples taken with a sampling interval of $T \leq 1/2f_{\max}$. In other words, if a band-limited analog signal is sampled with a sampling rate $\geq 2f_{\max}$, the information contained in the signal is not lost. Based on the Nyquist theorem, sampling once per channel period T in the PRML channel is appropriate only if the spectrum of the analog readback signal is concentrated below the frequency $f_{\max} = 1/2T$.

The spectrum of the head readback signal is discussed in Chapters 3 and 6. The frequency spectrum of a linear channel is usually defined as the Fourier transform of its *impulse response*, i.e., the dipulse response. A good approximation of the experimental spectrum of the channel is obtained if a random pattern transmitted through the channel is displayed with a spectrum analyzer. Approximately an equal number of positive and negative transitions are written on the magnetic medium, so there should be no spectrum content at zero frequency, i.e., the DC content is zero. However, the readback signal contains high frequency components.

The highest frequency in the signal spectrum corresponds to the fastest changing slope of the signal. A fast-changing narrow pulse will have a broader spectrum than a slowly changing wide pulse. Since the random pattern contains all spectral components with frequencies of $1/2nT$, where $n = 1, 2, 3, \ldots,$ the experimental spectrum should resemble that of $kV_{sp}(k)$ or $fV_{sp}(f)$ [Equation (3.23)], where $V_{sp}(f)$ is the Fourier transform of the single (isolated) pulse.

If we fix the channel bit period and write a random pattern with increasing pulse widths, the spectral energy distribution will shift to a lower and lower frequency range. Figure 12.6 shows the spectral energy distribution for several different values of PW_{50}/T (channel density). For $PW_{50}/T = 0.5$, the highest spectral energy occurs at about $1/2T$, but significant spectral components extend up to the *clock frequency* ($1/T$). This means that for a system with low ISI, channel bandwidth should be close to the clock frequency. For $PW_{50}/T = 2$, however, the signal spectrum is effectively concentrated below the half of clock frequency given by $1/2T$. The tail of this spectrum is still outside the "half-bandwidth" ($1/2T$) range, but the power of these high frequency components is relatively small.[8]

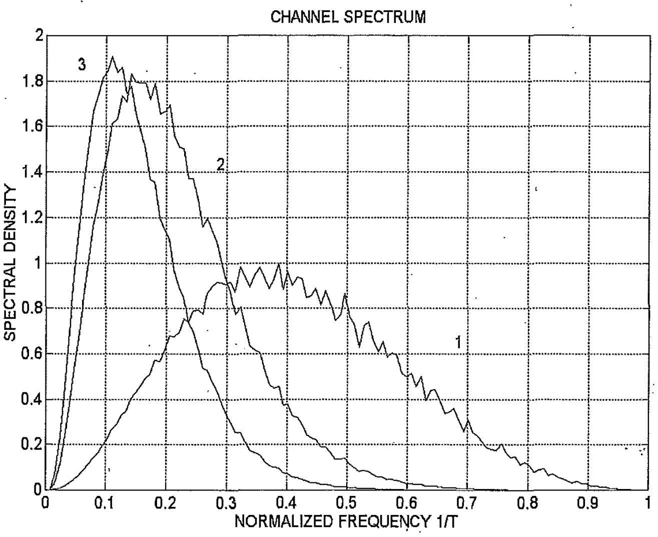
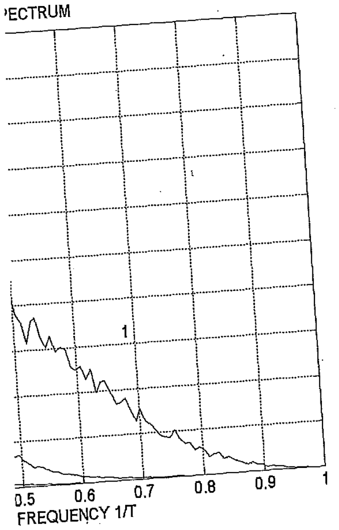


FIGURE 12.6. Magnetic recording channel spectra: (1) $PW_{50}/T = 0.5$; (2) $PW_{50}/T = 2$; $PW_{50}/T = 3$.

:trum corresponds to the fastest nging narrow pulse will have a ˌg wide pulse. Since the random with frequencies of $1/2nT$, where n should resemble that of $kV_{sp}(k)$ ·) is the Fourier transform of the

ıd write a random pattern with ɑergy distribution will shift to a re 12.6 shows the spectral energy of PW_{50}/T (channel density). For y occurs at about $1/2T$, but significa ne *clock frequency* ($1/T$). This means bandwidth should be close to the ever, the signal spectrum is effec lock frequency given by $1/2T$. The ıe "half-bandwidth" ($1/2T$) range, ɣ components is relatively small.[8]



nel spectra: (1) $PW_{50}/T = 0.5$; (2) PW_{50}/T

Therefore, the spectrum can be readily equalized into a spectrum with a cutoff frequency of $1/2T$, and sampling at a rate of $1/T$ becomes appropriate. However, note that the PRML channel may not work properly for systems with low channel densities.

If the noise power is uniform within the bandwidth, the total noise power within $f \leq 1/2T$ will be exactly half of that within $f \leq 1/T$. In this case, the fact that the PRML channel bandwidth is limited to $1/2T$ gives the PRML channel a gain of ~3 dB in SNR over the peak detection channel.

Nyquist's *interpolation formula* states that an analog signal $g(t)$, with a cutoff frequency $\leq 1/2T$ and samples of $g(nT)$, where $n = 0, 1, 2, 3, \ldots,$ can be expressed as[3]:

$$g(t) = \sum_{n=0}^{\infty} g(nT)\frac{\sin[(\pi/T)(t-nT)]}{(\pi/T)(t-nT)}. \tag{12.2}$$

In other words, each sample of "1" corresponds to a sinc(x) function (with an appropriate phase difference):

$$h(t) = \frac{\sin[(\pi/T)t]}{(\pi/T)t} = \mathrm{sinc}\left(\frac{\pi t}{T}\right),$$

which is the inverse Fourier transform of the low-pass filter. The analog signal is equal to the convolution of discrete samples and $h(t)$. The Fourier transform of Equation (12.2) is

$$G(f) = \sum_{n=0}^{\infty} g(nT)e^{-i(2\pi f)nT},\ 0<f<1/2T. \tag{12.3}$$

Since the PR4 channel has a bandwidth of $1/2T$, the equivalent representation of PR4 channel can be given as in Fig. 12.7, where an isolated

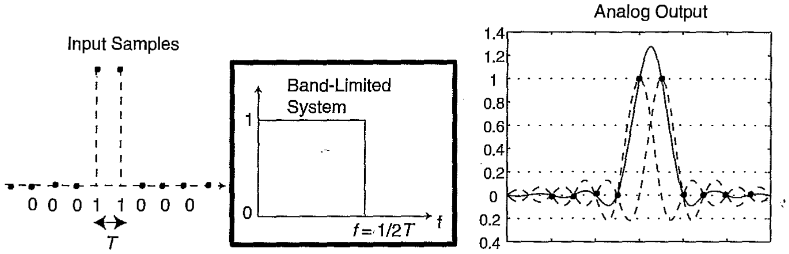


FIGURE 12.7. Nyquist interpolation formula: recovering analog waveform from discrete samples, as illustrated for the PR4 channel.

pulse represented by samples "...0110..." is passed through an ideal low-pass filter with a bandwidth of 1/2*T*. Therefore, the analog isolated pulse in the PR channel is given by Equation (12.2):

$$s(t) = \frac{\sin\frac{\pi t}{T}}{\frac{\pi t}{T}} + \frac{\sin\frac{\pi(t - T)}{T}}{\frac{\pi(t - T)}{T}}.$$

Now let us calculate the frequency response of the PR4 channel, which is the Fourier transform of the channel impulse response. The isolated pulse is called the "step response" of the PR4 channel, corresponding to a "step" (or transition) of magnetization. The samples due to the step response are {...0,0,0,1,1,0,0,0...}. The impulse response of the channel is the derivative of its step response, which generates samples {...0,0,1,0,−1,0,0...}, the same as the samples of a dipulse response. According to Equation (12.3), the frequency response of the PR4 channel is:

$$H(f) = 1 - \exp(-i4\pi fT), |H(f)| = 2\sin(2\pi fT), \qquad 0 \leq f \leq \frac{1}{2T},$$

which is shown in Fig. 12.8. Note that the spectrum peaks at the midband frequency 1/4*T*. The PR4 spectrum is very similar to the experimental spectrum of a magnetic recording channel when the channel density is ~2, as shown in Fig. 12.6. The similarity is an important clue that magnetic recording channel could be turned into a PR4 channel with a minimum amount of equalization.

12.2.2 Partial-response polynomial and classification

The PR4 channel is only a special class of PR schemes. A convenient way of describing different partial response schemes is to use PR polynomials, which are somewhat different from those of cyclic codes in Chapter 8. The PR polynomials describe the correspondence between the NRZ input data pattern and the ideal samples. The NRZ input data pattern is generally given by a sequence of bits $\{a_k\}$. The PR polynomial *P* is an operator which determines the current channel sample s_k based on the current and previous NRZ bits: $s_k = P[a_k, a_{k-1}, \ldots, a_{k-n}]$. As we have already seen from Equation (12.1), the PR4 channel is described by the following operation:

$$s_k = a_k - a_{k-2}.$$

10. . ." is passed through an ideal
/2T. Therefore, the analog isolated
quation (12.2):

$$\frac{\sin\frac{\pi(t-T)}{T}}{\frac{\pi(t-T)}{T}}.$$

' response of the PR4 channel, which
nel impulse response. The isolated
: the PR4 channel, corresponding to
ıtion. The samples due to the step
he impulse response of the channel
ponse, which generates samples
:he samples of a dipulse response.
quency response of the PR4 channel

$$| = 2\sin(2\pi fT), \qquad 0 \le f \le \frac{1}{2T},$$

ıt the spectrum peaks at the midband
is very similar to the experimental
:hannel when the channel density is
:ity is an important clue that magnetic
into a PR4 channel with a minimum

nial and classification

ass of PR schemes. A convenient way
nse schemes is to use PR polynomials,
n those of cyclic codes in Chapter 8.
ırrespondence between the NRZ input
The NRZ input data pattern is gener-
}. The PR polynomial P is an operator
nel sample s_k based on the current and
. . ,a_{k-n}]. As we have already seen from
, described by the following operation:

$a_k - a_{k-2}$.

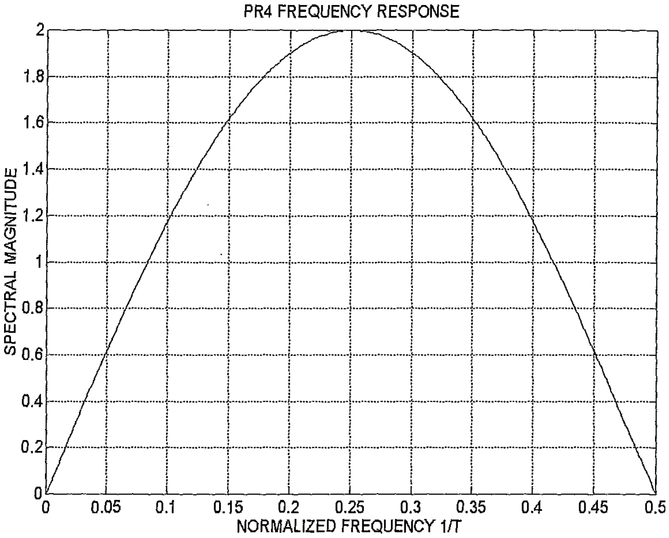


FIGURE 12.8. Frequency response of PR4 channel.

The simple operation can be presented by a PR polynomial as shown below.[4,8]

First let us define a fundamental polynomial operator D:

$$Da_k = a_{k-1},$$

which is called the "delay operator" because it delays the input bit by one bit period. The definition of delay operator allows simple arithmetical operations such as addition, subtraction, multiplication, power, etc. For example, applying the delay operator twice is defined as the second power of D:

$$D^2a_k = D(Da_k) = Da_{k-1} = a_{k-2}.$$

When a magnetic head reads a disk, it responds only to changes of magnetization, i.e., it differentiates the NRZ bits. The differentiating function of a head is described with an operator:

$$(1 - D)a_k = a_k - a_{k-1}.$$

Since NRZ bits take values 0 and 1, operator $(1 - D)$ will result in $+1$ or -1 samples, depending on the direction of the magnetization change. Operator $(1 - D)$ is the simplest polynomial, corresponding to generating positive or negative samples at the transition locations. Taking absolute values after the $1 - D$ operation generates NRZI bits. Obviously, an isolated pulse expressed in NRZ data (or magnetization) {. . . 000111. . . } is transformed into NRZI data {. . . 000100. . . } after the differentiation.

In a PR4 system each voltage pulse has two samples. In other words, if a transition of magnetization occurs, it results in a sample equal to "1" at the transition location and another sample at the next sample period. This can be described with a "spreading" operator $(1 + D)$. If $a_k = 1$, $(1 + D)a_k$ will result in sequence "11". Therefore, a PR4 system can be described as a polynomial $(1 - D)(1 + D) = 1 - D^2$. In other words, the current PR4 sample is

$$s_k = (1 - D^2)a_k = a_k - a_{k-2},$$

which is the same as Equation (12.1).

Partial-response channels, as listed in the following table, were originally proposed in digital communication to combat ISI.[2] Because of the similarity of the PR4 spectrum and the frequency response of magnetic recording channels, the PR4 channel (Class IV) became the darling of the magnetic data storage industry in the 1990s.

Name	Polynomial	Impulse response samples
PR1	$1 + D$	. . . 0 1 1 0 . . .
PR2	$(1 + D)^2 = 1 + 2D + D^2$	. . . 0 1 2 1 0 . . .
PR3	$(1 + D)(2 - D) = 2 + D - D^2$	. . . 0 2 1 −1 0 . . .
PR4	$(1 + D)(1 - D) = 1 - D^2$	. . . 0 1 0 −1 0 . . .
PR5	$-(1 + D)^2(1 - D)^2 = -1 + 2D^2 - D^4$	. . . 0 −1 0 2 0 −1 0 . . .

If we ignore the differentiating operator $(1 - D)$ and look only at the coefficients of the polynomial $(1 + D)$, we get the samples of an isolated pulse: {1,1} in the PR4 channel. The operator $(1 + D)$ determines how the transition sample is spread over the neighboring bit periods. The PR4 channel has been extended to a general family of channels as defined by the following polynomials[8]:

$$(1 - D)(1 + D)^n,\ n = 1, 2, 3, \text{etc.} \tag{12.4}$$

erator $(1 - D)$ will result in +1 or
on of the magnetization change.
mial, corresponding to generating
nsition locations. Taking absolute
nerates NRZI bits. Obviously, an
(or magnetization) {. . . 000111. . . }
100. . . } after the differentiation.
has two samples. In other words,
it results in a sample equal to "1"
sample at the next sample period.
ng" operator $(1 + D)$. If $a_k = 1$, (1
Therefore, a PR4 system can be
$D) = 1 - D^2$. In other words, the

$$= a_k - a_{k-2},$$

in the following table, were origi-
on to combat ISI.[2] Because of the
frequency response of magnetic
lass IV) became the darling of the
1990s.

	Impulse response samples
	. . . 0 1 1 0 . . .
	. . . 0 1 2 1 0 . . .
	. . . 0 2 1 −1 0 . . .
	. . . 0 1 0 −1 0 . . .
D^4	. . . 0 −1 0 2 0 −1 0 . . .

ator $(1 - D)$ and look only at the
we get the samples of an isolated
rator $(1 + D)$ determines how the
eighboring bit periods. The PR4
family of channels as defined by

$$= 1, 2, 3, \text{etc.} \tag{12.4}$$

The PR4 system corresponds to $n = 1$. If we set $n = 2$, the samples of the isolated pulse will be given by the operator $(1 + D)^2 = 1 + 2D + D^2$, which corresponds to {1,2,1}. This type of channel is called the "extended partial response 4" or EPR4 channel. If $n = 3$, $(1 + D)^3 = 1 + 3D + 3D^2 + D^3$, and the isolated pulse has samples {1,3,3,1}. This type of PR channel is called the E^{2}PR4 channel. The following table summarizes the popular PR channels used in magnetic recording:

Name	Polynomial	Isolated pulse samples	Impulse response samples
PR4	$(1 - D)(1 + D)$	. . . 0 1 1 0 . . .	. . . 0 1 0 −1 0 . . .
EPR4	$(1 - D)(1 + D)^2$	. . . 0 1 2 1 0 . . .	. . . 0 1 1 −1 −1 0 . . .
E^{2}PR4	$(1 - D)(1 + D)^3$	. . . 0 1 3 3 1 0 . . .	. . . 0 1 2 0 −2 −1 0 . . .

The isolated pulses in the extended PR4 channels are wider than that in PR4 channel: the EPR4 pulse extends over 3 bit periods and E^{2}PR4 pulse over 4 bit periods. This means that a transition in an E^{2}PR4 system will "interfere" with the next three transitions when they are detected. In other words, the channel density PW_{50}/T is much higher in E^{2}PR4 system than in PR4.

The sample values such as "2" or "3" above are somewhat arbitrary, we will normalize the maximum sample value to "1" hereafter in the text. Using this notation, an EPR4 isolated pulse has sample values {. . . 0,1/2,1,1/2,0, . . . }, as shown in Fig. 12.9. The EPR4 analog pulse shape is obtained with Nyquist's interpolation formula:

$$s(t) = \frac{1}{2}\frac{\sin\frac{\pi t}{T}}{\frac{\pi t}{T}} + \frac{\sin\frac{\pi(t-T)}{T}}{\frac{\pi(t-T)}{T}} + \frac{1}{2}\frac{\sin\frac{\pi(t-2T)}{T}}{\frac{\pi(t-2T)}{T}}. \tag{12.5}$$

The superposition of pulses in the EPR4 system results in five sampling levels: {−1,−1/2,0,1/2,1}. The current channel sample value can be found from the NRZ bit pattern using the EPR4 polynomial:

$$\begin{aligned} s_k &= (1 - D)(1 + D)^2 a_k \\ &= (1 + D - D^2 - D^3) a_k \\ &= a_k + a_{k-1} - a_{k-2} - a_{k-3}, \end{aligned} \tag{12.6}$$

which is the linear combination of the current data bit and the 3 previous data bits.

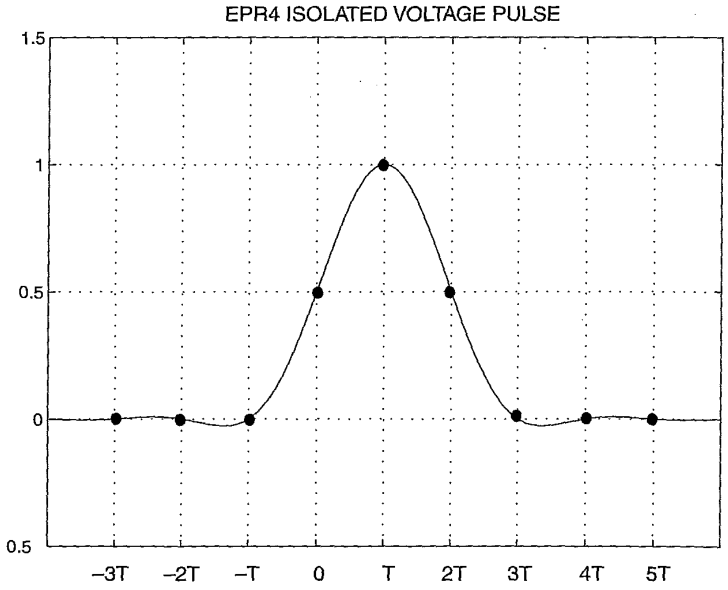


FIGURE 12.9. Isolated pulse for EPR4 system.

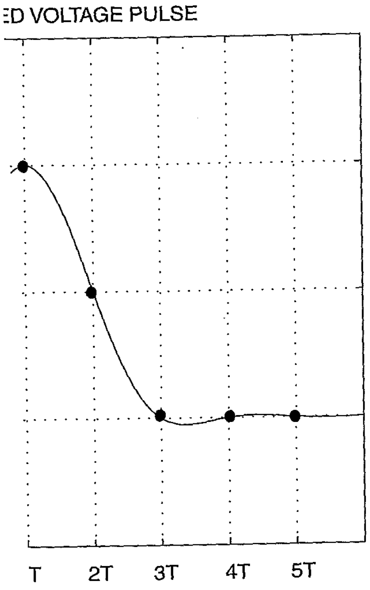
The EPR4 system is often used with 8/9(0,4) or 2/3(1,7) modulation code, which results in different impulse responses. The dipulse response (impulse response) of EPR4 system with the $d = 0$ constraint can be derived as follows:

$$\begin{array}{ccccccc}
 & 0 & \frac{1}{2} & 1 & \frac{1}{2} & 0 & \\
+ & & 0 & -\frac{1}{2} & -1 & -\frac{1}{2} & 0 \\ \hline
= & 0 & \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & 0
\end{array}$$

With the $d = 1$ constraint, the EPR4 dipulse response becomes:

$$\begin{array}{cccccccc}
 & 0 & \frac{1}{2} & 1 & \frac{1}{2} & 0 & & \\
+ & & 0 & 0 & -\frac{1}{2} & -1 & -\frac{1}{2} & 0 \\ \hline
= & 0 & \frac{1}{2} & 1 & 0 & -1 & -\frac{1}{2} & 0
\end{array}$$

The E^2PR4 isolated pulse samples are: {...,0,1/3,1,1,1/3,0,...}. The shape of the analog isolated pulse is shown in Fig. 12.10. There are seven possible sample levels in an E^2PR4 system: {−1,−2/3,−1/3,0,1/3,2/3,1} be-



tem.

ith 8/9(0,4) or 2/3(1,7) modulation e responses. The dipulse response vith the $d = 0$ constraint can be

$$\begin{array}{cccc} & & \frac{1}{2} & 0 \\ & -1 & -\frac{1}{2} & 0 \\ \hline & -\frac{1}{2} & -\frac{1}{2} & 0 \end{array}$$

ɔulse response becomes:

$$\begin{array}{ccccc} \frac{1}{2} & 0 & & \\ \frac{1}{2} & -1 & -\frac{1}{2} & 0 \\ \hline) & -1 & -\frac{1}{2} & 0 \end{array}$$

are: {...,0,1/3,1,1,1/3,0,...}. The wn in Fig. 12.10. There are seven m: {−1,−2/3,−1/3,0,1/3,2/3,1} because of linear superposition. An E^{2}PR4 system is usually used with the (1,7) encoding. The dipulse response samples for the E^{2}PR4 system with the (1,7) code is given by:

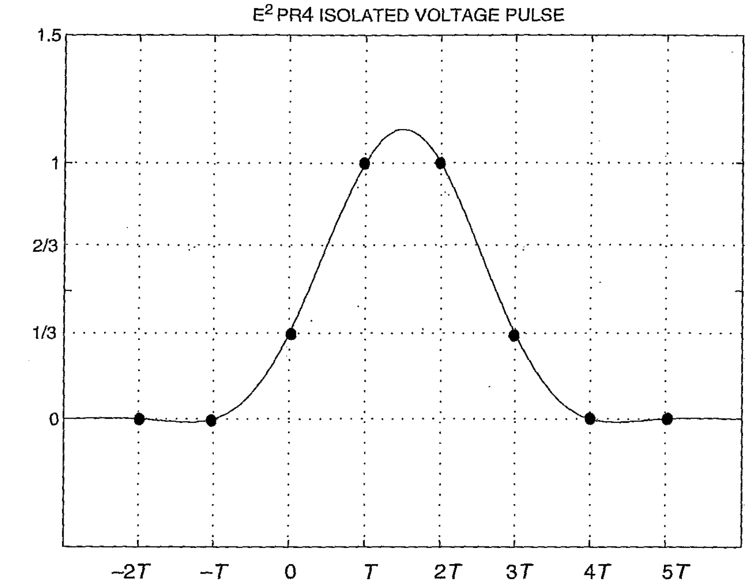


FIGURE 12.10. Isolated pulse shape for E^{2}PR4 system.

$$\begin{array}{ccccccccc} & 0 & 1/3 & 1 & 1 & 1/3 & 0 & 0 & \\ + & & 0 & 0 & -1/3 & -1 & -1 & -1/3 & 0 \\ \hline = & 0 & 1/3 & 1 & 2/3 & -2/3 & -1 & -1/3 & 0 \end{array}$$

The eye diagrams and sample value distributions of EPR4 and E^{2}PR4 systems are more complicated than those in PR4 channel. For example, Figure 12.11 presents the eye diagram for the EPR4 system. There are five distinct "focusing" points and the EPR4 histogram will consist of five separate peaks. Similarly, the E^{2}PR4 system will have seven distinct "focusing" points, as shown in Fig. 12.12, and the E^{2}PR4 histogram will consist of seven separate peaks.

Frequency responses of EPR4 and E^{2}PR4 systems are calculated similar to that of PR4 as the spectra of the dipulse responses, as shown in Fig. 12.13. Changing from PR4 to E^{2}PR4 with increasing order of channel

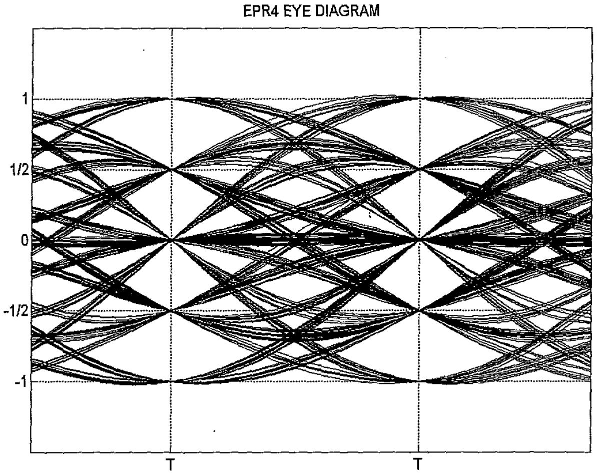


FIGURE 12.11. EPR4 system eye diagram.

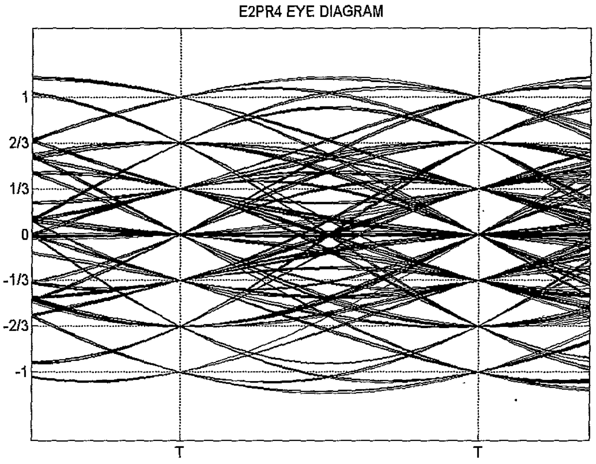
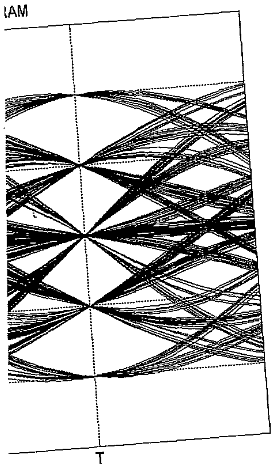
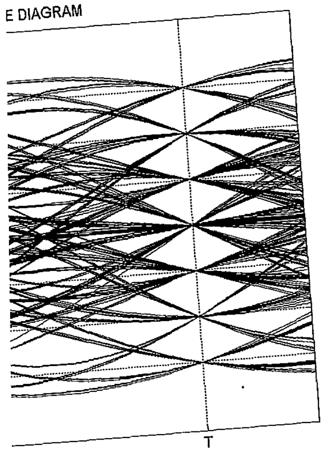


FIGURE 12.12. Eye diagram of E^2PR4 system.



n.



.4 system.

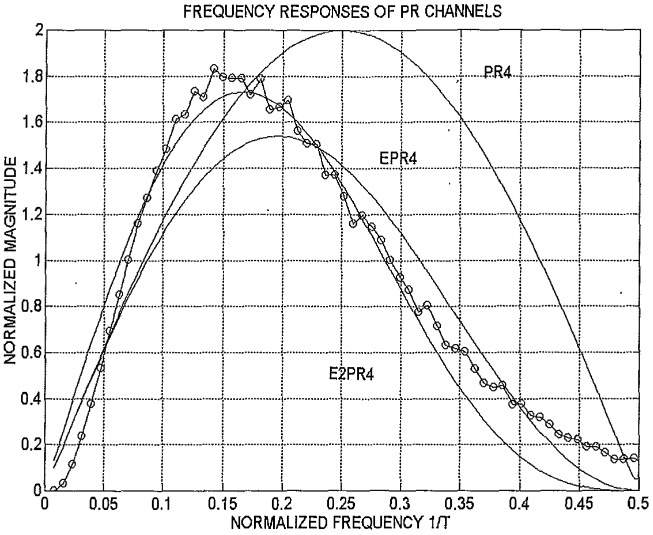


FIGURE 12.13. Frequency responses of PR4, EPR4, and E^{2}PR4 systems and experimental channel frequency response for $PW_{50}/T = 2$ (circles).

polynomials, the peaks of the frequency responses shift toward the lower frequencies. This is not surprising because the isolated voltage pulses for EPR4 and E^{2}PR4 become wider with respect to a fixed channel bit period. Note that E^{2}PR4 provides the closest fit to the channel frequency response for $PW_{50}/T = 2$. This means that little equalization is required to make the recording channel an E^{2}PR4 system.

While the PR4 channel is most widely used now, there are certain advantages to using higher-order PR systems such as EPR4 and E^{2}PR4. As shown in Fig. 12.13, the channel frequency response for $PW_{50}/T = 2$ is much closer to that of EPR4 or E^{2}PR4 than that of PR4. As a result, equalizing the channel frequency response to PR4 will require more amplification of the high-frequency components above 0.25/T. The "boost" of high-frequency components will inevitably amplify the noise in the system and degrade the channel error rate. If the noise is additive Gaussian noise, EPR4 equalization gains 2–3 dB in SNR over PR4. Therefore, EPR4 and E^{2}PR4 systems allow higher channel densities, consistent with the fact that these systems have more nonzero samples per isolated pulse.

Another advantage of the EPR4 and E^{2}PR4 systems is that they are more practical to use with the 2/3(1,7) code than PR4 if we do not want to sacrifice user bit density. While a PR4 pulse extends over two channel periods, an E^{2}PR4 pulse extends over four channel periods. Calculations show that the E^{2}PR4 system with the (1,7) code allows almost the same user density as the PR4 system with the 8/9(0,4/4) code (see the next section). Using the (1,7) code doubles the minimum distance between adjacent transitions, and thus decreases nonlinear distortions such as NLTS and partial erasure.

The main disadvantage of EPR4 and E^{2}PR4 channels is the increased complexity of the ML detector required. As we will discuss in Section 12.4, the PR4 channel allows an extremely simple realization of the ML decoder compared with the extended PRML. EPR4 and E^{2}PR4 systems also require more sophisticated schemes for the clock and gain recovery (Section 12.3).

12.2.3 Channel and user densities of PR4, EPR4, and E^{2}PR4 channels

The standard definition of channel density is the ratio of the PW_{50} of *unequalized* isolated pulse to the channel bit period T. However, PRML equalization modifies the shape of the isolated pulse. For example, if a significant high-frequency boost is present in the PRML channel, the equalized isolated voltage pulse will have a smaller PW_{50} than the unequalized one.

In an ideal case, the pulse width before equalization should match that after equalization, so the equalizer does not introduce excessive noise boost and the channel density is optimal for data recovery. In this case, the channel density before equalization should be close to that after equalization. Therefore, we can take the ideal equalized isolated pulses of PR channels to calculate the channel densities as given by PW_{50}/T, assuming that PW_{50} is measured at 50% of the unipolar pulse amplitude. Using the isolated pulses shown in Figs. 12.2, 12.9, and 12.10, we obtain that

$$D_{ch} \equiv \frac{PW_{50}}{T} = \begin{cases} 1.65 & \text{for PR4,} \\ 2 & \text{for EPR4,} \\ 2.31 & \text{for E}^2\text{PR4.} \end{cases}$$

For example, if PW_{50} of the isolated pulse from the magnetic head is 16.5 ns, the optimal channel bit period for a PR4 system will be closed to

and E^2PR4 systems is that they are
7) code than PR4 if we do not want
PR4 pulse extends over two channel
r four channel periods. Calculations
ie (1,7) code allows almost the same
ith the 8/9(0,4/4) code (see the next
les the minimum distance between
eases nonlinear distortions such as

t and E^2PR4 channels is the increased
uired. As we will discuss in Section
tremely simple realization of the ML
led PRML. EPR4 and E^2PR4 systems
emes for the clock and gain recovery

ities of PR4, EPR4, and E²PR4

el density is the ratio of the PW_{50} of
channel bit period T. However, PRML
of the isolated pulse. For example, if a
is present in the PRML channel, the
will have a smaller PW_{50} than the un-

vidth before equalization should match
alizer does not introduce excessive noise
optimal for data recovery. In this case,
zation should be close to that after equal-
he ideal equalized isolated pulses of PR
l densities as given by PW_{50}/T, assuming
the unipolar pulse amplitude. Using the
2.2, 12.9, and 12.10, we obtain that

$$l = \begin{cases} 1.65 & \text{for PR4,} \\ 2 & \text{for EPR4,} \\ 2.31 & \text{for } E^2PR4. \end{cases}$$

ated pulse from the magnetic head is 16.5
riod for a PR4 system will be closed to

16.5/1.65 = 10 ns. In practice, the channel density for PR4 tends to be ~2, which is ~20% higher than the optimal value of 1.65. This is possible by making a proper tradeoff between the pulse slimming with a PRML equalizer and the noise boost. The achievable bit error rate in practical PR4 channels with $D_{ch} = 2$ is usually better than 10^{-9}.

PRML channels are typically characterized in terms of user density: $D_u = D_{ch}R$, where R is the code rate. The user density represents how many user bits of information can be stored in a unit of the medium (as measured by PW_{50}). Let us compare PRML systems with the peak detection channel. Most peak detection systems use 2/3(1,7) encoding and the channel bit period is one-half of the flux change period B. According to Equation (3.17), the resolution of the channel is ~70% when $PW_{50}/B = 1$ or $PW_{50}/T_{ch} = 2$, which is acceptable. Therefore, the peak detection channel density is 2. The corresponding user density is $2 \times 2/3 = 4/3 = 1.33$. For a PR4 system with 8/9(0,4/4) encoding and a channel density of 1.65, the user density is 1.47. An E^2PR4 system with 2/3(1,7) encoding has a natural channel density of 2.31, which gives a user density of 1.54. Note that this is almost the same as that of the PR4 system with the 8/9 code. However, since the $d = 1$ constraint increases the actual distance between transitions on the medium, nonlinear distortions are reduced. The following table provides a comparison of user densities for different detection methods:

Method	User Density 2/3(1,7) code	User Density 8/9(0,4/4) code
Peak detection (70% resolution)	1.33	0.89
Peak detection (80% resolution)	1.00	0.67
PR4	1.10	1.47
EPR4	1.33	1.77
E^2PR4	1.54	2.05

The comparison of user densities has limited values in terms of judging the relative merit of these channels. More reliable comparison should take into consideration their sensitivity to noise, nonlinear distortions, equalization, etc. Typically about 30% gain in user density can be achieved using PRML instead of peak detection.

12.2.4 Principles of equalization

The goal of equalization is to modify the frequency response of a magnetic recording channel so as to match it with the frequency response of a desired PRML scheme. Consider channel spectra shown in Fig. 12.14. The

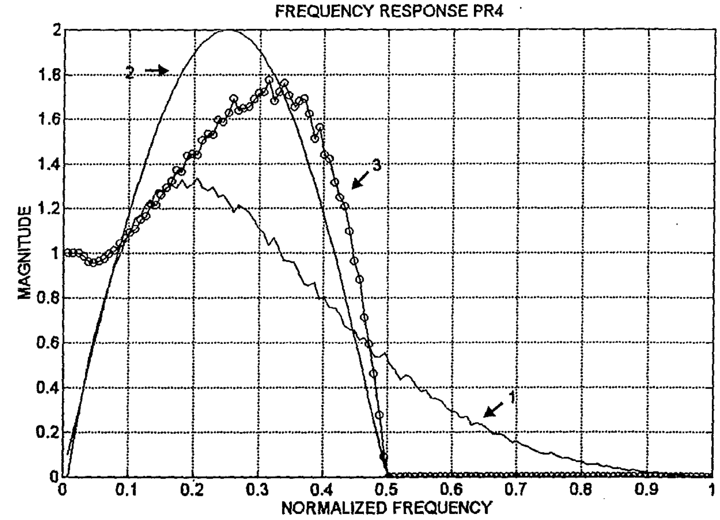


FIGURE 12.14. Channel spectrum (curve 1), and the frequency responses of PR4 channel (curve 2) and equalizer (curve 3).

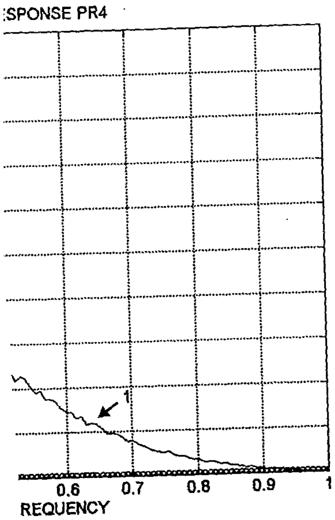
unequalized channel spectrum (curve 1) must be transformed with an equalizer so as to match the desired PR4 frequency response (curve 2). This transformation is performed with a linear filter with the frequency response given by curve 3.

The function of an equalizer can also be interpreted as that it transforms (reshapes) the sources pulse $s(t)$ into the target pulse $p(t)$.[5,7] Let $S(\omega)$ be the spectrum (Fourier transform) of $s(t)$, $P(\omega)$ the spectrum of $p(t)$, where $\omega = 2\pi f$ is the angular frequency, then the transfer function of the reshaping filter (equalizer) is

$$H(\omega) = \frac{P(\omega)}{S(\omega)}. \tag{12.7}$$

The inverse Fourier transform of $H(\omega)$ is the impulse response $h(t)$ of the filter. In time domain, the target pulse obtained by the equalizer is the convolution of the source pulse with the impulse response $h(t)$:

$$p(t) = \int_{-\infty}^{\infty} h(\tau)s(t - \tau)d\tau. \tag{12.8}$$



and the frequency responses of PR4

) must be transformed with an
4 frequency response (curve 2).
linear filter with the frequency

› be interpreted as that it trans-
nto the target pulse $p(t)$.[5,7] Let
›f $s(t)$, $P(\omega)$ the spectrum of $p(t)$,
hen the transfer function of the

(12.7)

ıe impulse response $h(t)$ of the
tained by the equalizer is the
mpulse response $h(t)$:

$\cdot\ \tau)d\tau.$ (12.8)

There are many different approaches of designing a PRML equalizer. A simple way to realize the desired frequency response is to use a *continuous time filter* (CTF) with programmable cutoff frequency and boost. By changing these two parameters, a family of frequency response functions is obtained. A more flexible solution is based on digital *finite impulse response* (FIR) *filters* or programmable analog *transversal filters*.[7]

The impulse response of a digital FIR filter has a *finite* number of nonzero samples as the name indicates. It is based on the sample version of Equation (12.8):

$$p(k) = \sum_{m=0}^{N} h(m)s(k-m), \tag{12.9}$$

where $p(k)$, $h(k)$, and $s(k)$ are the samples of $p(t)$, $h(t)$, and $s(t)$, respectively. The digital FIR filter is implemented using delay registers (denoted by D or Z^{-1}), multipliers (denoted by triangles), and adders (denoted by Σ), as shown in Figure 12.15. Each nonzero multiplier represents a *tap* (or a branch) of the filter. The more taps in the filter, the more complex it is. The sampling theorem states that if the samples are taken with a sampling frequency that is at least twice higher than the bandwidth of the channel, then the equivalent continuous signals $p(t)$, $h(t)$, and $s(t)$ can be obtained. Therefore, the PR equalization can be achieved with a digital FIR filter with sufficient sampling frequency and number of taps. Some mixed-design implementations of FIR filters use similar architecture, but analog signal levels obtained from sample-and-hold circuits instead of digital samples.

A programmable analog transversal filter realizes Equation (12.8) based on the continuous time signals $p(t)$ and $s(t)$:

$$p(t) = \sum_{m=0}^{N} h(m\tau)s(t-m\tau), \tag{12.10}$$

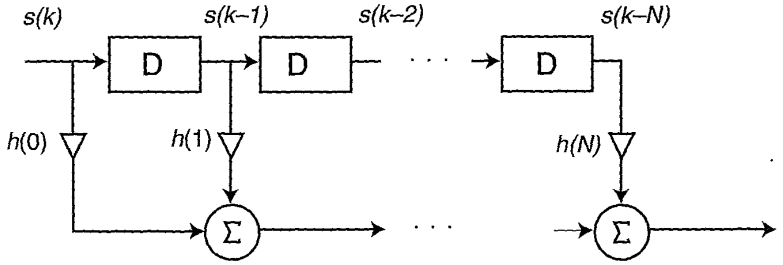


FIGURE 12.15. Structure of digital FIR filter with $N + 1$ taps.

where τ is the delay time in each delay lines, which replace the delay registers in Fig. 12.15. The architecture of the transversal filter is very similar to that of FIR filter. The source pulse $s(t)$ is taken as the input, then the output of each delay line is multiplied by some coefficient $h(m\tau)$ using an analog multiplier, and finally the outputs of all the multipliers are added together to give the target pulse $p(t)$.

To design either a digital FIR or an analog transversal equalizer properly, it is important to select the delay time and the total number of taps. In a digital FIR filter, the samples are processed at the channel clock rate. In contrast, in an analog transversal filter, the delay lines are fixed and the delay time must be $\leq 1/2f_{\max}$, where $f_{\max}$ is the channel bandwidth. For example, for a PR4 signal at 100 Mflux/s, the signal bandwidth is 50 MHz, so any delay time less than 10 ns would be appropriate.

Determining the total number of taps is somewhat more complicated than choosing the delay time. The number of taps determines the total length of the impulse response of the equalizer and, therefore, the detail in the frequency response of the filter. The acceptable rule of thumb is that the total length of the delay lines should not be less than the duration of the pulse to be equalized. In most PRML channels the function of equalization is distributed between a continuous time filter at the channel front-end and a digital FIR filter. The FIR filters typically have 6–10 programmable taps.

The frequency response of the equalizer is obtained from Equation (12.3):

$$H(\omega) = \sum_{k=0}^{N} h(k)\exp(-i\omega k\tau). \tag{12.11}$$

For example, let the delay time $\tau = 5$ ns, $N = 11$ (12 taps), and the set of coefficients $h(k) = \{1\ 1\ -1\ -1\ 1\ 1\ -1\ -1\ 1\ 1\ -1\ -1\}$, then the amplitude (absolute value) of the frequency response of the equalizer is shown in Fig. 12.16. By choosing this set of coefficients we have realized a kind of "band-pass" filter at frequencies between 40 and 60 MHz.

To provide reshaping, the FIR filter coefficients must be determined from the desired frequency response. Analytically, the "reshaping" problem can be solved if the digitized waveforms of the source isolated voltage pulse and the target isolated pulse are available. Let the input and output pulses be defined by their values in the sequence of points $t_1, t_2, \ldots, t_M$, and $s_{i,j} = s(t_i - \tau_j), p_i = p(t_i)$. The quality of equalization is measured with the *mean squared error* (MSE) or distance between the equalized and target signals:

$$E(\vec{h}) = \sum_i \left(\sum_j s_{i,j} h_j - p_i \right)^2, \tag{12.12}$$

where h_j are the equalizer coefficients.

ines, which replace the delay the transversal filter is very ılse $s(t)$ is taken as the input, ılied by some coefficient $h(m\tau)$ outputs of all the multipliers : $p(t)$.

og transversal equalizer prop- and the total number of taps. ssed at the channel clock rate. the delay lines are fixed and $_{\text{ıax}}$ is the channel bandwidth. /s, the signal bandwidth is 50 ould be appropriate.

; somewhat more complicated r of taps determines the total zer and, therefore, the detail in ptable rule of thumb is that the e less than the duration of the els the function of equalization er at the channel front-end and ıave 6–10 programmable taps. ıer is obtained from Equation

$$-i\omega k\tau). \qquad (12.11)$$

^{r} = 11 (12 taps), and the set of 1 −1 −1}, then the amplitude : of the equalizer is shown in its we have realized a kind of 40 and 60 MHz.

efficients must be determined ytically, the "reshaping" prob- s of the source isolated voltage able. Let the input and output ence of points $t_1,t_2,\ldots,t_M$, and tion is measured with the *mean* equalized and target signals:

$$- p_i\Big)^2 , \qquad (12.12)$$

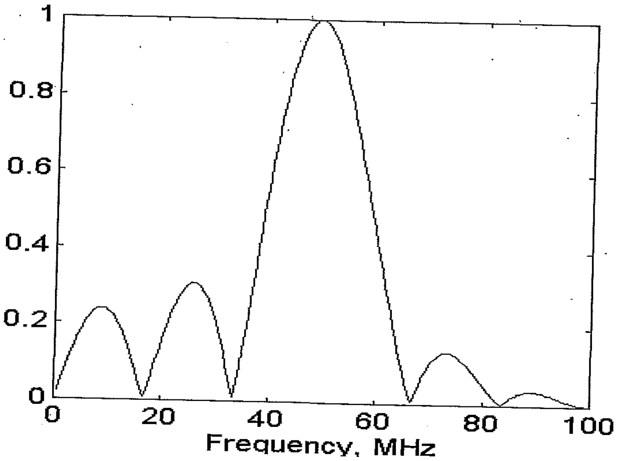


FIGURE 12.16. Example of the frequency response of an equalizer.

The best equalization is achieved when the distance is at its minimum. This occurs when

$$\frac{\partial E}{\partial h_l} = \sum_i \left[2s_{i,l} \left(\sum_j s_{i,j} h_j - p_i \right) \right] = 0,\ l = 1, \ldots, M, \qquad (12.13)$$

This results in a system of linear equations:

$$\sum_j \left[\left(\sum_i s_{i,j} s_{i,l} \right) h_j \right] = \sum_i p_i s_{i,l}, \qquad (12.14)$$

or in matrix forms:

$$\mathbf{S}^{\mathrm{T}}\mathbf{S}\mathbf{H} = \mathbf{S}^{\mathrm{T}}\mathbf{P}, \qquad (12.15)$$

where $\mathbf{S} = (s_{i,j})$, $\mathbf{H} = (h_i)$, $\mathbf{P} = (p_i)$. The solution to Equation (12.15) is given by the formula:

$$\mathbf{H} = (\mathbf{S}^{\mathrm{T}}\mathbf{S})^{-1}\mathbf{S}^{\mathrm{T}}\mathbf{P}. \qquad (12.16)$$

Therefore, if the input and output signal shapes are available, Equation (12.16) gives the set of equalizer coefficients that minimizes the MSE between the equalized and target pulses.

For example, consider the following set of coefficients:

$$h = \{0.22,-0.84,0.12,0.46,-0.04,0.67,-0.36,0.26,0.62,0.22,-1.0,0.22\}$$

These coefficients were calculated so as to reshape the response of an inductive head with $PW_{50} \approx 40$ ns to the isolated pulse of the PR4 channel at 60 Mb/s. The frequency response of the resulting equalizer is shown in Figure 12.17. If the equalizer is followed by a 30-MHz-bandwidth low-pass filter, the isolated pulse shown in Fig. 12.18 (left) will be transformed into the nearly ideal PR4 isolated pulse shown in Fig. 12.18 (right).

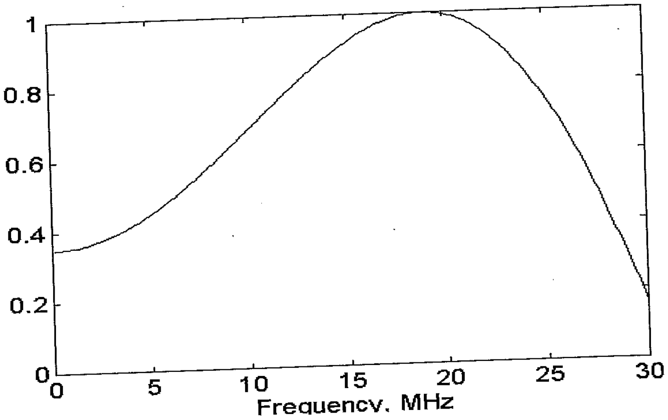


FIGURE 12.17. Frequency response of a PR4 equalizer.

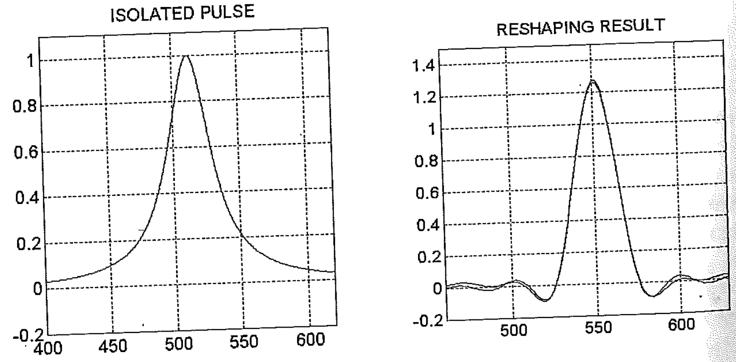
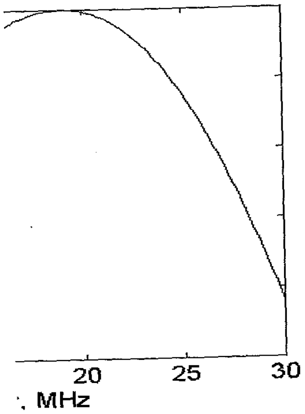
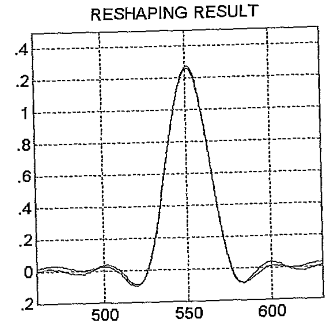


FIGURE 12.18. Isolated pulse before (left) and after reshaping (right). The reshaped pulse is shown together with the ideal PR4 pulse.

as to reshape the response of an
: isolated pulse of the PR4 channel
the resulting equalizer is shown
ved by a 30-MHz-bandwidth low-
'ig. 12.18 (left) will be transformed
: shown in Fig. 12.18 (right).



4 equalizer.



ınd after reshaping (right). The re-
al PR4 pulse.

In practice, the analytical calculation of the equalizer coefficients is time-consuming and suboptimal. Analytically calculated tap coefficients $\{h_i\}$ do not necessarily provide the best error rate of the PRML channel. The disagreement between the theory and the practice is due to the following:

1. The noise in the magnetic recording channel is correlated (e.g., medium noise), and an equalizer modifies the correlation properties of the noise. The "ideal" equalizer may create undesirable noise patterns and degrade the error rate of the ML detector. In this situation, certain "misequalization" improves the error rate performance of the PRML channel. The "optimal" equalizer allows a certain trade-off between the desirable target pulse shape and the noise boost at the equalizer output.
2. Magnetic recording channel is nonlinear. A certain residual level of nonlinear distortions is typically present in the readback signal, even after write precompensation. Therefore, the equalizer optimal for low-density readback signals (e.g., isolated pulses) becomes suboptimal for high-density signals, which often contains more nonlinear distortions. A certain trade-off between ideal equalization of isolated transitions and high-frequency patterns (dibit, tribit, and series of transitions) appears to improve the error rate of the PRML channels.

In practice, the analytically calculated equalizer coefficients are used only as the initial approximation, they are adjusted by writing a typical data pattern on the disk (e.g., random pattern) and using an adaptive procedure to minimize some "quality" measure.[8] The ultimate measure of channel quality is the bit error rate at the output of the ML detector, but a faster optimization algorithm can be based on minimizing the MSE between the received samples and their ideal values.

For example, an optimization strategy is based on the gradient optimization algorithm. If the set of coefficients at the kth step of the optimization procedure is

$$\mathbf{h}^k = \{h_1, h_2, \ldots, h_N\}, \tag{12.17}$$

the algorithm tweaks each tap coefficient h_i to minimize the MSE value $E(k)$. It changes h_i, checks the new value of MSE, and determines an update vector:

$$\nabla \mathbf{h}^k = \left[\frac{\partial E(k)}{\partial h_1}, \frac{\partial E(k)}{\partial h_2}, \ldots, \frac{\partial E(k)}{\partial h_N}\right], \tag{12.18}$$

for which MSE is decreasing. The updated set of tap coefficients is calculated as

$$\mathbf{h}^{k+1} = \mathbf{h}^k + \alpha\nabla\mathbf{h}^k. \tag{12.19}$$

The procedure is repeated until the MSE value is minimized, as schematically illustrated in Fig. 12.19.

12.3 CLOCK AND GAIN RECOVERY

The correct operation of any PRML system depends on taking samples of the readback signal in the exact "focus" positions of the eye diagram. As shown in Fig. 12.5, shifting the clock slightly from the correct position is enough to distort sample values substantially. Therefore, the correct clock signal must be recovered from the data signal.

The clock signal in a data detection channel is generated by a voltage-controlled oscillator. A clock recovery circuit, based on a phase-locked loop (PLL), adjusts the phase of the oscillator depending on the value of the phase error, as shown in Fig. 12.20. Input signal (i.e., readback data signal) is supplied to the phase detector, which calculates the phase error function by comparing the input signal with the output of the voltage-controlled oscillator. If the calculated phase error equals zero, the clock position is correct, and the oscillator frequency (and phase) stays exactly in the correct position. When the phase of the input signal and the phase of the oscillator diverge due to some reason, such as the instability of disk rotation, the phase error signal deviates from zero and the frequency of the oscillator is shifted to cancel out the phase error.

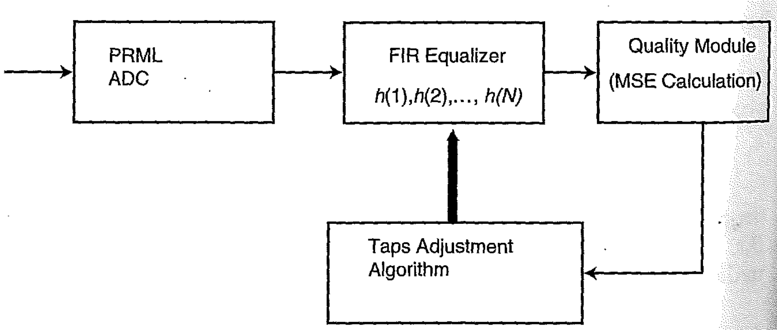


FIGURE 12.19. Block diagram of adaptive adjustment of equalizer taps.

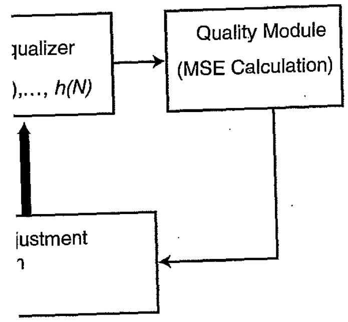
d set of tap coefficients is calcu-

$\nabla \mathbf{h}^k$. (12.19)

value is minimized, as schemati-

RY

em depends on taking samples
" positions of the eye diagram.
lightly from the correct position
tantially. Therefore, the correct
data signal.
annel is generated by a voltage-
rcuit, based on a phase-locked
ator depending on the value of
nput signal (i.e., readback data
which calculates the phase error
with the output of the voltage-
ase error equals zero, the clock
uency (and phase) stays exactly
the input signal and the phase
ason, such as the instability of
tes from zero and the frequency
ne phase error.



ljustment of equalizer taps.

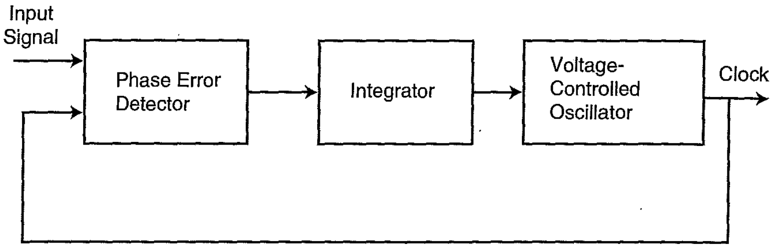


FIGURE 12.20. Typical structure of phase-locked loop for clock recovery.

The clock recovery system solves two main problems. First, prior to reading the pattern, it is necessary to align the phase of the channel clock to the correct position of the sampling. This is referred to as the *initial phase acquisition*. The second problem is *data tracking*, i.e., following the relatively slow instabilities of the disk rotational speed. To avoid fast noisy phase shifts and to provide the stability of clock recovery, the phase error signal is passed through an integrator, which is a low-pass filter. Note that the meaning of data tracking here is different from that of tracking servo to be discussed in Chapter 15.

For the initial phase acquisition, each data sector starts with a special synchronization pattern.[5,7] For the PR4 system, the synchronization pattern in NRZ form is given by {110011001100. . . }, i.e., it consists of transitions evenly separated by two channel bit periods. The ideal PRML samples of this pattern are $\{+1,+1,-1,-1,+1,+1,-1,-1\ldots\}$. The readback signal from the synchronization field resembles a sine wave, which is sampled close to its peaks, as shown in Fig. 12.21 (top). If the initial sampling phase is incorrect, as shown in Fig. 12.21 (bottom), the direction and the amount of phase shift required to reach the correct sampling point can be determined based on the first several samples. In fact, the sequence of two consecutive samples $\{s(1),s(2)\}$ determines the tabulated phase update. For example, if $\{s(1) = +1, s(2) = +1\}$, the phase update equals 0. If $\{s(1) = 0, s(2) = 1.273$ (peak of the signal)$\}$, the phase update equals $T/2$.

The correct clock phase is obtained typically within the first several bytes of the synchronization pattern, which is followed by a special address marker. When the channel recognizes the address mark, it switches itself into data tracking mode, i.e., it recognizes the end of the synchronization pattern, as shown in Fig. 12.22.

The data tracking mode requires slow updates of the channel clock to compensate the slow instabilities of disk rotation. Assume that $s_0(k)$ is

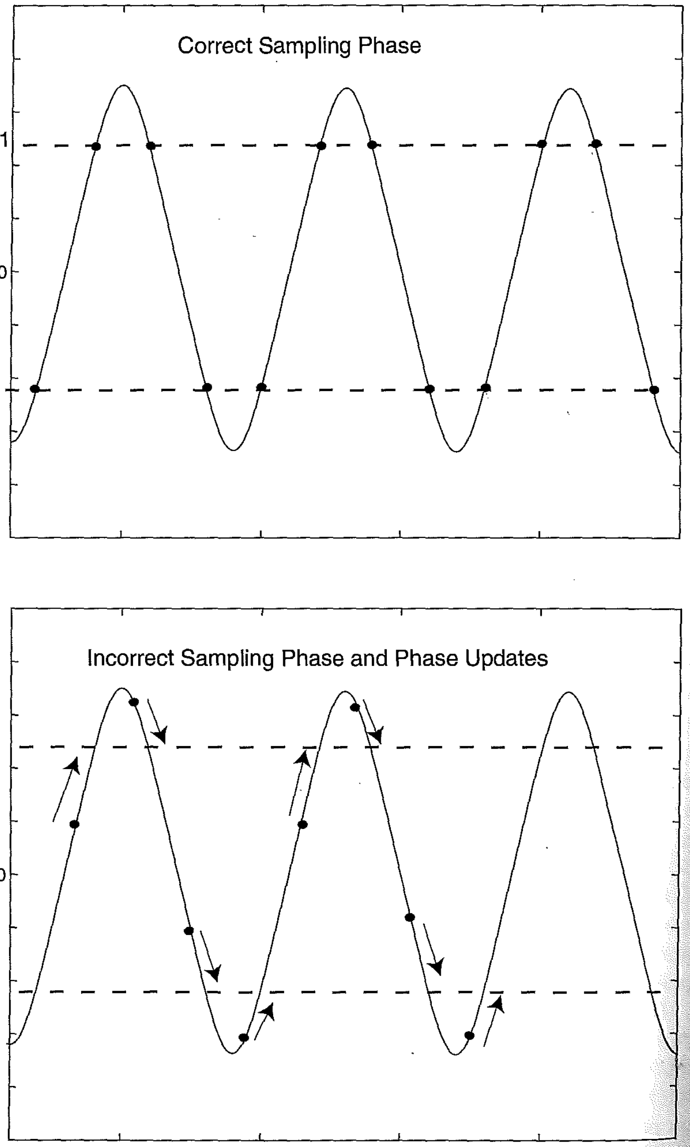
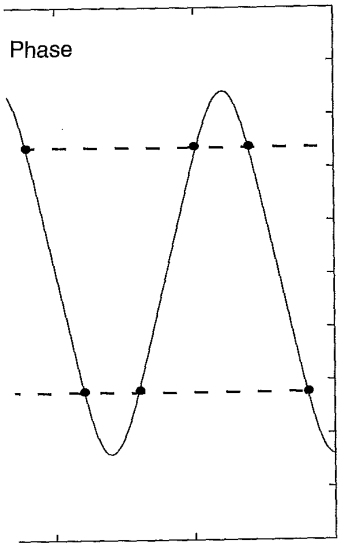
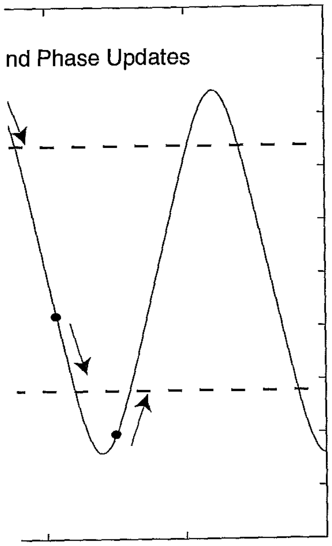


FIGURE 12.21. Synchronization pattern for PR4 channel.





ʼR4 channel.

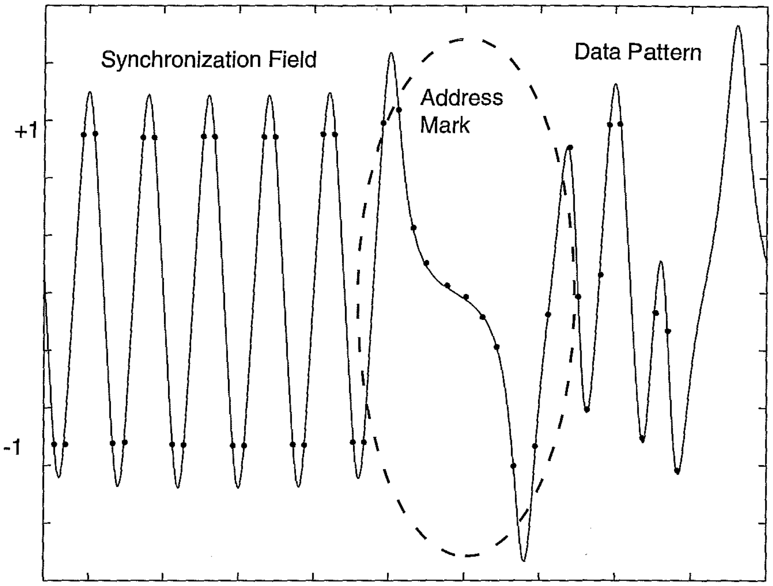


FIGURE 12.22. Synchronization pattern, address mark, and data pattern signal.

the ideal value of the signal at the kth sampling point, and instead, some value $s(k,\tau)$ is obtained with a phase shift of τ. The resulting error is $e(k,\tau) = s(k,\tau) - s_0(k)$, which depends on the value of phase shift. To minimize the error, we can use a simple gradient minimization algorithm. This can be achieved by finding the derivative of the squared error:

$$\frac{de^2(k,\tau)}{d\tau} = \frac{d(s(k,\tau) - s_0(k))^2}{d\tau} = 2e(k,\tau)\frac{ds(k,\tau)}{d\tau}. \qquad (12.20)$$

If the derivative given by Equation (12.20) is known, the next phase can be calculated from the previous phase as:

$$\tau_{k+1} = \tau_k - \alpha e(k,\tau)\frac{ds(k,\tau)}{d\tau}, \qquad (12.21)$$

where α is a positive convergence parameter. Equation (12.21) is reiterated and it will converge to the point of the minimum phase error.

Consider the signal and sampling points shown in Fig. 12.23, where four signal points {s1, s2, s3, s4} are sampled earlier than the correct clock positions, corresponding to ideal values {+1, 0, −1, 0}. The directions of

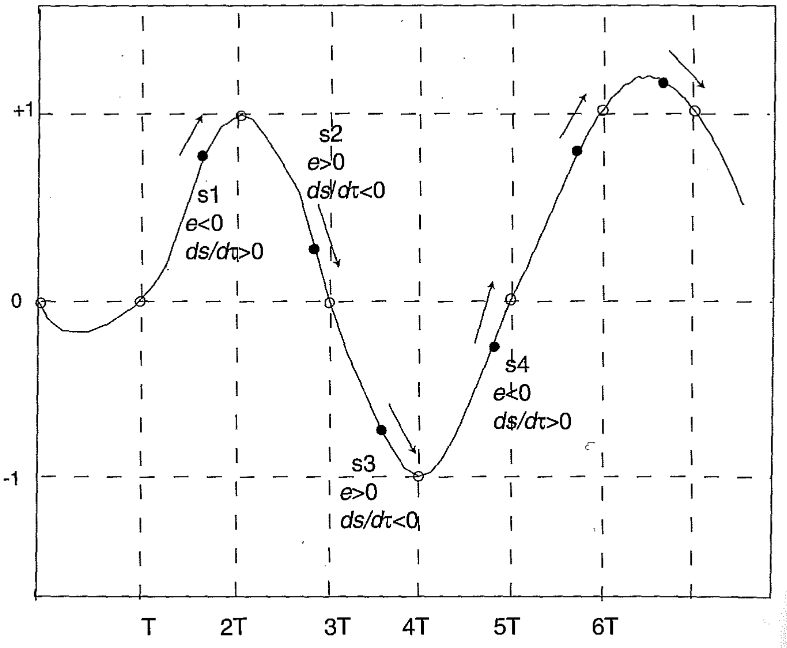
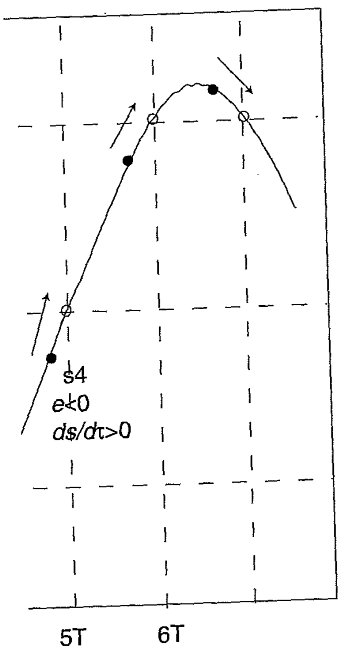


FIGURE 12.23. Calculation of phase updates for PR4 signal.

the phase update signals will be determined from Equation (12.21). At the sample point s1, the value of the signal is less than the ideal one, so $e(k,\tau) < 0$. However, the derivative $ds/d\tau$ at this point is positive, so the phase update signal is positive and the sample should be shifted to the right according to Equation (12.21). Similarly, for sample point s2, $e(k,\tau) > 0$, but $ds/d\tau < 0$, so the phase update signal is also positive. The same can be said about the other two sample points. Therefore, the algorithm given by Equation (12.21) correctly predicts the directions of the phase updates.

To modify the current phase based on Equation (12.21), we need to know the current error value $e(k,\tau)$ and the derivative of the signal $s(k,\tau)$. The error value can be easily calculated using the threshold-detected data as the ideal value—even if we make some errors, the overall phase error should be correct due to averaging of the individual signals by the integrator. In contrast, it is more complicated to calculate the derivative of the signal. It is usually estimated using a sequence of samples. For example, if the sequence of samples for a PR4 system is $\{+1,0,-1\}$, then the derivative



r PR4 signal.

from Equation (12.21). At the
ıs than the ideal one, so $e(k,\tau)$
oint is positive, so the phase
uld be shifted to the right ac-
mple point s2, $e(k,\tau) > 0$, but
ositive. The same can be said
the algorithm given by Equa-
f the phase updates.
Equation (12.21), we need to
lerivative of the signal $s(k,\tau)$.
the threshold-detected data
rrors, the overall phase error
dividual signals by the inte-
alculate the derivative of the
e of samples. For example, if
+1,0,−1}, then the derivative

in the central point is certainly negative because it is on the negative slope of a dipulse. If the sequence of samples is $\{-1,-1,+1\}$, then the derivative at the second sample point should be positive because it is on a positive slope, as shown in Fig. 12.21. A table of the expected derivatives can be constructed and used for clock recovery.

Since all the decisions about the data, samples, and clock phase updates are based on the assumption that the system amplification is correct, i.e., the "1" level is an *a priori* known value, a correct *gain control* must be provide in a PRML channel. The gain control signal is calculated from the squared error similar to that given by Equation (12.21). If a sample $s(k)$ is received instead of the ideal value $s_0(k)$, the error due to the system gain deviation is given by $e(k,\gamma) = \gamma s(k) - s_0(k)$, where γ is the system gain. The derivative of the squared error is

$$\frac{de^2(k,\gamma)}{d\gamma} = \frac{d(\gamma s(k) - s_0(k))^2}{d\gamma} = 2e(k,\gamma)s(k). \qquad (12.22)$$

Therefore, the system gain should be updated according to:

$$\gamma_{k+1} = \gamma_k - \beta e(k,\gamma)s(k), \qquad (12.23)$$

where β is another positive convergence parameter. The algorithm increases the system gain if the absolute value of a sample is smaller than that of the corresponding ideal value, and reduces the system gain if the opposite is true.

12.4. MAXIMUM LIKELIHOOD DETECTION

12.4.1 PRML state diagram and trellis

The maximum likelihood detection can be best understood based on the concept of *state diagram,* which describes all the possible states of the magnetic recording system and the transitions between these states.[2,3] The state diagram consists of two distinct parts: *states and transitions.* A state represents a unique physical situation at a given time, which may be specified by the current medium magnetization, and possibly its history, i.e., the medium magnetization one or several bit periods earlier. A convenient representation of the medium magnetization is provided by the NRZ data. A transition is the "event" relating the current state and the next state. There are only two types of transitions in magnetic recording: either the medium magnetization is changed between the current and the next bit periods or it is not. Note that the transition here is more broadly defined than a "magnetic transition" first introduced in Chapter 2.

State diagram is usually represented as a graph consisting of nodes (states) and oriented links between these nodes. The links represent transitions from one node to another. For example, a simple state diagram describing a memoryless magnetic recording system, such as PR1 and peak detection channel, is shown in Fig. 12.24. The meaning of the state diagram is very simple. The system may have two states, determined by current magnetization—either "0" or "1". If the current magnetization is "1" and the next magnetization is "1", the system stays in state "1" (right-hand loop). Similarly, if the current magnetization is "0" and the next magnetization is "0", the system stays in state "0" (left-hand loop). If the current magnetization is "0" and the next magnetization is "1", the system changes its state from "0" to "1" (upper arrow). If the current magnetization is "1" and the next magnetization is "0", the system goes to state "0" (lower arrow).

The NRZ state diagram for a PR4 channel is shown in Fig. 12.25a. The difference between the peak detection and the PR4 is that in the latter the state diagram has memory or is dependent on history. A state in the peak detection channel is completely described by the current magnetization, but a state in the PR4 channel is completely described by the current magnetization as well as the previous magnetization. This is because the isolated pulse in PR4 extends over two bit-periods and the presence of the previous transition affects the current channel sample value. The state diagram shown in Fig. 12.25a has four states: medium magnetization may be either "0" or "1" at the current or previous bit period. For example, if the current state is "10", it means that the current magnetization is "0" and the previous magnetization was "1". If the next magnetization is "1", the next state becomes "01".

The important feature of this state diagram is that we have *only two possible paths leading to each state.* This is because the medium magnetization can either stay in the same direction or change its sign. For example, we can arrive at state "10" only if the previous medium magnetization is "1" and current magnetization is "0". Since there are only two possible states

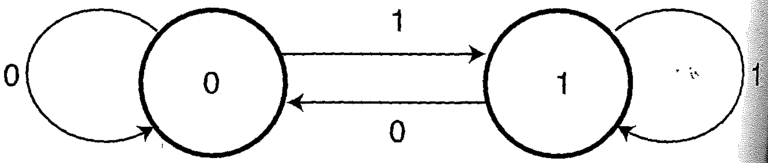
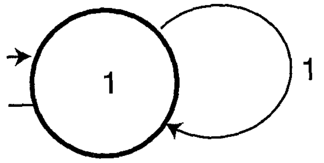


FIGURE 12.24. State diagram of a memoryless channel.

is a graph consisting of nodes
des. The links represent transi-
nple, a simple state diagram
ing system, such as PR1 and
2.24. The meaning of the state
ave two states, determined by
If the current magnetization is
ystem stays in state "1" (right-
ıetization is "0" and the next
ate "0" (left-hand loop). If the
agnetization is "1", the system
ow). If the current magnetiza-
"0", the system goes to state

ınnel is shown in Fig. 12.25a.
ınd the PR4 is that in the latter
dent on history. A state in the
bed by the current magnetiza-
letely described by the current
netization. This is because the
t-periods and the presence of
ıannel sample value. The state
s: medium magnetization may
ious bit period. For example,
e current magnetization is "0"
the next magnetization is "1",

ram is that we have *only two*
ıse the medium magnetization
nge its sign. For example, we
medium magnetization is "1"
e are only two possible states



channel.

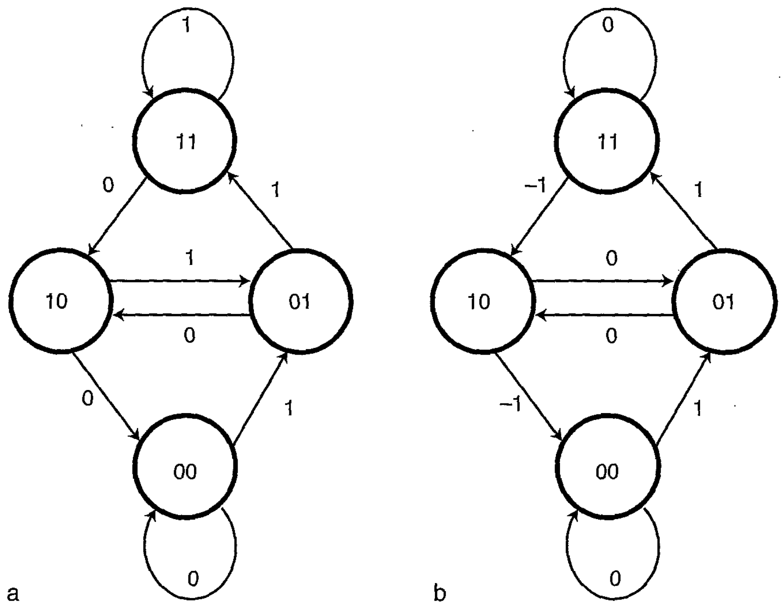


FIGURE 12.25. (a) NRZ state diagram of PR4 system. (b) PR4 state diagram with sample values.

with current magnetization being "1"—they are "11" or "01", only two paths leads to state "10".

An equivalent representation of the PR4 state diagram is shown in Fig. 12.25b, where PR4 *sample values* instead of current NRZ data are indicated along the transition paths. Each change of magnetization can generate only one sample value in the PR4 channel if we agree on the signs of magnetization. For example, change from "0" to "1" corresponds to a positive pulse, and change from "1" to "0" corresponds to a negative pulse. One has to be very careful in order not to confuse the *states* and the *samples*. First, if the magnetization between the previous state and the next state does not change for two consecutive bit periods, no magnetic transitions are generated and the samples should be equal to zero as shown by the lower and upper loops in Fig. 12.25b. If state "00" goes to state "01", the magnetization is changed from"0" to "1" and corresponds to a positive transition. This transition has samples {1,1}, so the current sample equals 1. If no magnetic transition is written next, the magnetization will change from the state "01" to state "11". This corresponds to an isolated transition and we expect to read the second sample value {1} from the isolated pulse.

To find the sample values in a state diagram of any PRML channel, we can use the corresponding PR polynomial $H(D)$. For example, the PR4 polynomial is $1 - D^2$, so the channel sample is determined by the difference of 2 NRZ bits: $s_k = a_k - a_{k-2}$. If the PR4 state changes from "11" to "10", $a_k = 0$, $a_{k-2} = 1$, thus $s_k = a_k - a_{k-2} = -1$. A *universal state diagram* can be constructed as shown in Fig. 12.26, where the number of NRZ bits to uniquely define a state is n, which is the order of polynomial $H(D)$. Therefore, PR4 and E^{2}PR4 channels require 2 and 4 NRZ bits, respectively. The state diagram with sample values is more useful because it visualizes how the ideal sample values in a PRML channel change with the magnetization pattern.

The equivalent representation of state diagram is the *trellis*, which is obtained by tracing the time sequence of state changes. Take the PR4 channel as an example, we first write down the four states at the current and at the next time instant. Then we trace the possible links between these states, as shown in Fig. 12.27. Typically we indicate the corresponding ideal sample value at each link. The concept of trellis is very important.

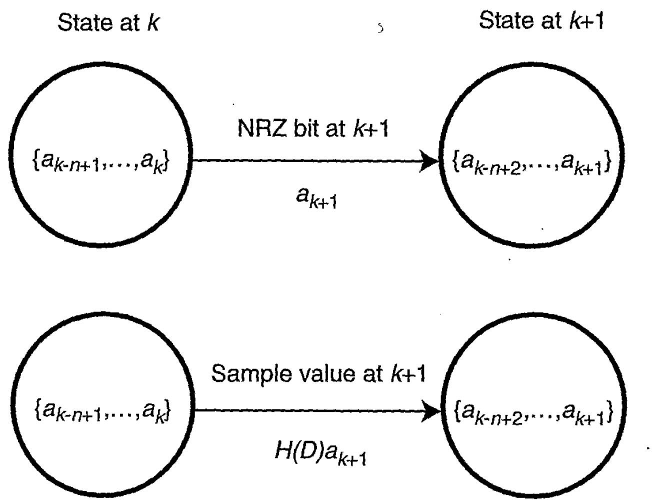
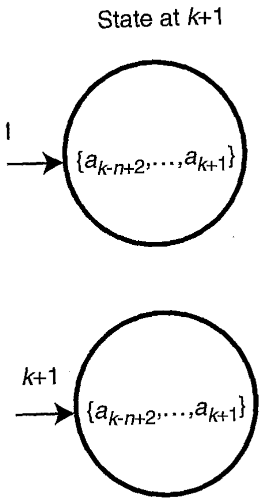

State at k
State at k+1
NRZ bit at k+1
$\{a_{k-n+1},\ldots,a_k\}$
$\{a_{k-n+2},\ldots,a_{k+1}\}$
a_{k+1}
Sample value at k+1
$\{a_{k-n+1},\ldots,a_k\}$
$\{a_{k-n+2},\ldots,a_{k+1}\}$
$H(D)a_{k+1}$

FIGURE 12.26. Universal state diagram with paths indicated by the current NRZ bits (top) or the current sample values (bottom).

:e "11". This corresponds to an
ɩ the second sample value {1}

iagram of any PRML channel,
ial $H(D)$. For example, the PR4
ɔle is determined by the differ-
'R4 state changes from "11" to
$= -1$. A *universal state diagram*
where the number of NRZ bits
:he order of polynomial $H(D)$.
2 and 4 NRZ bits, respectively.
ore useful because it visualizes
ɩnnel change with the magneti-

diagram is the *trellis*, which is
f state changes. Take the PR4
n the four states at the current
ɩce the possible links between
ly we indicate the correspond-
cept of trellis is very important.



ɩths indicated by the current NRZ
).

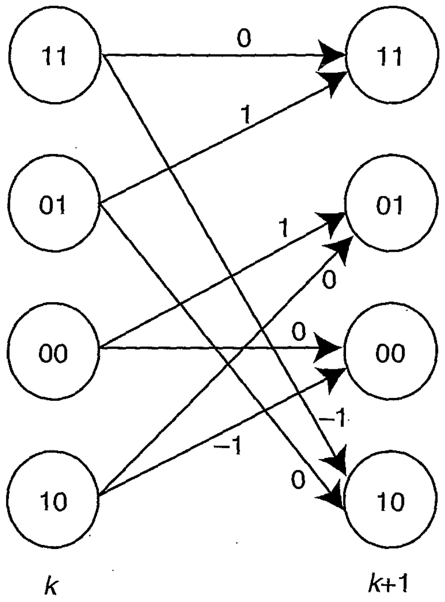


FIGURE 12.27. Trellis for PR4 system.

When the ML detector makes its decisions, it actually "extends" the trellis frame shown in Fig. 12.27 for several consecutive time instants k, $k + 1$, $k + 2$, etc., and estimates the likelihood of possible trajectories in the trellis structure.

Now consider the trellises for EPR4 and E^2PR4 channels, which are more complicated than that of PR4. In the EPR4 system we have to trace the history of three consecutive changes of magnetization which gives us $2^3 = 8$ possible states: "111", "110", "101", "100", "011", "010", "001" and "000", as shown in Figure 12.28. To find the samples corresponding to the transitions between EPR4 states, we use EPR4 polynomial (Equation 12.6). For example, when state "110" changes to state "101", the sample value is $s_k = a_k + a_{k-1} - a_{k-2} - a_{k-3} = 1 + 0 - 1 - 1 = -1$. Since we typically normalize all the samples with the maximum value, which is 2 for EPR4, so the normalized sample value is $-1/2$.

The trellis for the E^2PR4 system has $2^4 = 16$ states. However, if $d = 1$ encoding is used, the number of states in E^2PR4 will be reduced. In fact, the $d = 1$ constraint will eliminate all the states which have adjacent

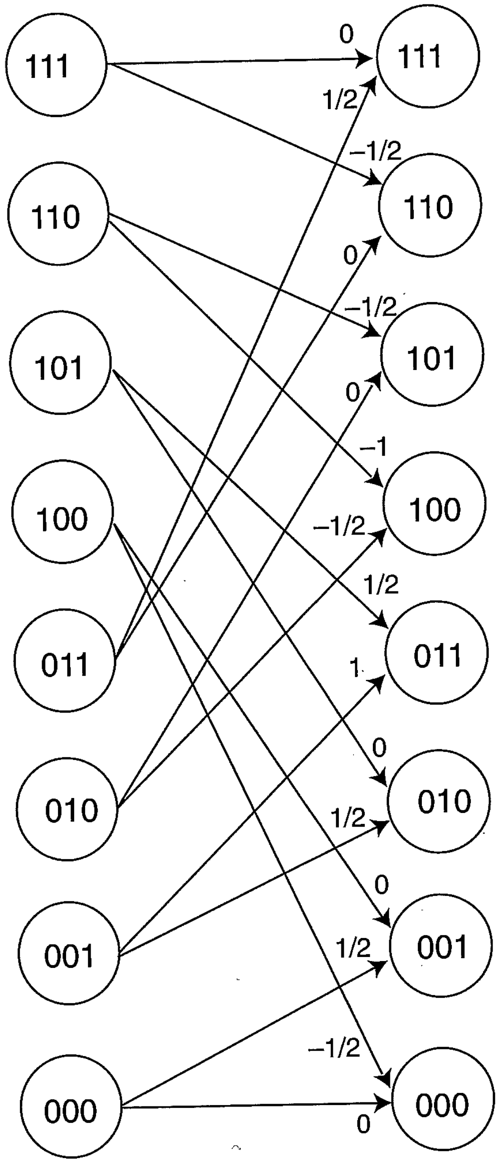
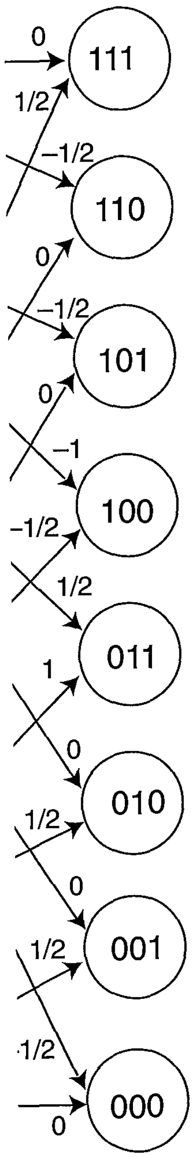


FIGURE 12.28. Trellis for EPR4 system.



magnetic transitions. In other words, the states with NRZ data "1101", "1011", "1010", "0101", "0100", and "0010" will be eliminated. The resulting trellis is shown in Fig. 12.29, which is greatly simplified from the unconstrained case.

12.4.2 Maximum likelihood or Viterbi detection algorithm

When several transitions in a PR4 state machine are traced from state to state, the PR4 trellis is extended by adding the same diagram at each subsequent time instant. Consider the trellis extension shown in Fig. 12.30. Assume that at time instant k the exact state of the system is known to be "01". When the next channel sample comes from the ADC output, the system goes to another state. If the ideal channel sample equals 1, the system will go from state "01" at time k to state "11" at time $k + 1$. If the channel sample equals 0, the system will go from state "01" at time k to state "10" at time $k + 1$. However, the incoming channel sample is distorted by noise, so both trajectories are possible between time instants k and $k + 1$. Therefore, no decisions are taken at this point, and two states ("11" and "10") are considered as possible candidates for the system at time $k + 1$.[2,3]

Each of the possible states is assigned a certain number, or a "path metric," which corresponds to the squared difference between the channel sample $s(k + 1)$ and the ideal channel sample $s_0(k + 1)$. For example, the path metrics for states "11" and "10" at time $k + 1$ are

$$M_{k+1}(\text{"11"}) = [s(k + 1) - 1]^2,$$
$$M_{k+1}(\text{"10"}) = [s(k + 1) - 0]^2.$$

The next channel sample $s(k + 2)$ causes the system to go to a different state at time $k + 2$. As seen from Fig. 12.30, there are four alternatives at this time: state "11" may change into states "11" and "10", while state "10" may change into states "00" and "01". Therefore, all four states are to be considered at time $k + 2$. Each trajectory leading to a particular state will have an accumulated path metric. For example, the metric for state "10" at time $k + 2$ is obtained by adding the squared difference between the channel sample $s(k + 2)$ and the ideal sample (-1) to the metric $M_{k+1}(\text{"11"})$:

$$M_{k+2}(\text{"10"}) = M_{k+1}(\text{"11"}) + (s(k + 2) + 1)^2.$$

When the next sample $s(k + 3)$ is received, the system is going from time instant $k + 2$ to $k + 3$. As seen from Fig. 12.30, there are eight possible

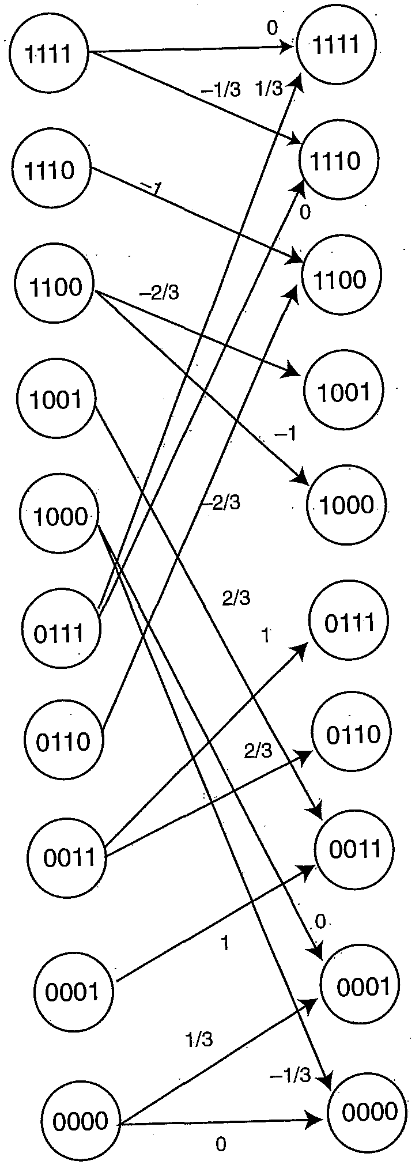
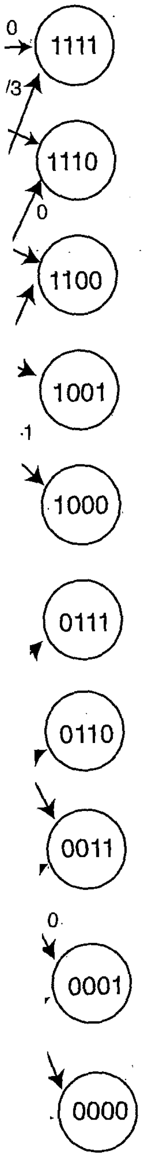


FIGURE 12.29. E^{2}PR4 trellis with $d = 1$ constraint.



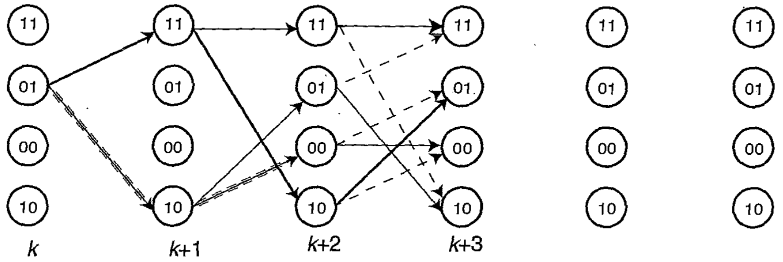


FIGURE 12.30. Candidate trajectories in PR4 trellis.

trajectories at this step, since there are four possible state at time $k + 2$. However, at time $k + 3$ two competing trajectories are converging to a state. Distance metrics (or path metrics) have been assigned the two competing trajectories, one of which should be discarded by comparing the path metrics. For example, two trajectories converging to state "11" at time $k + 3$ will have the following metrics:

$$M(1)_{k+3}(\text{"11"}) = M_{k+2}(\text{"11"}) + (s(k + 3) - 0)^2,$$
$$M(2)_{k+3}(\text{"11"}) = M_{k+2}(\text{"01"}) + (s(k + 3) - 1)^2.$$

If $M(1)_{k+3}(\text{"11"}) < M(2)_{k+3}(\text{"11"})$, then the trajectory "01"–"11", shown in Fig. 12.30 with a dashed line, will be discarded. In this case, the accumulated path metric at time $k + 3$ for state "11" will be equal to $M_{k+3}(\text{"11"}) = M_{k+2}(\text{"11"}) + (s(k + 3) - 0)^2$. Note that after four of the eight competing trajectories are discarded, only four possible paths are continued through the trellis.

The Viterbi algorithm is based on the fact that only one surviving trajectory will remain after several consecutive steps through the trellis. The process of discarding erroneous paths from the trellis is illustrated in Fig. 12.31. As the next sample comes to the PRML channel, the trellis is extended to time instant $k + 4$. Similarly to Fig. 12.30, there are eight possible trajectories at this time and four of them will be discarded by comparing their path metrics. However, note that all the dashed trajectories coming from states "11" and "10" at time $k + 3$ are discarded at time $k + 4$. Therefore, all the trajectories leading to these states up to time $k + 3$ should also be discarded. This leaves us with only two surviving trajectories up to step $k + 3$: {"01", "11", "10", "01"} and {"01", "10", "00", "00"}.

The continuation of this process to time instant $k + 5$ is illustrated in Fig. 12.32. Now all the trajectories coming from state "00" at time $k + 4$

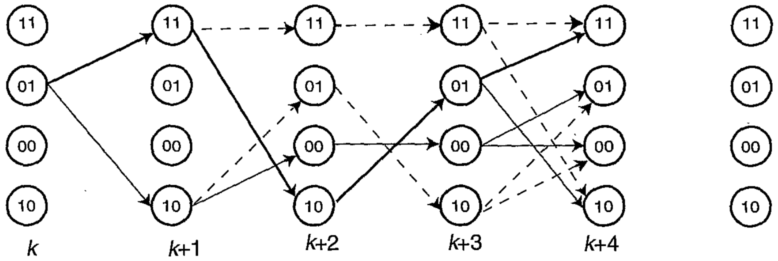


FIGURE 12.31. Extension of trellis shown in Fig. 12.30 to time instant $k + 4$.

are to be discarded. Tracing all the paths in the trellis coming to state "00", then the trajectory {"01", "10", "00", "00"} starting at time k is eliminated. This leaves only one surviving trajectory {"01", "11", "10", "01"} up to time $k + 3$, which constitutes the solutions of the Viterbi algorithm. Now we have to start from time instant $k + 3$, and repeat the process to find the next only surviving trajectory.

The Viterbi algorithm can be summarized as follows:

1. Starting from the known state, calculate the path matric for each possible trajectory in the trellis leading to the current state.
2. If two trajectories converge to the same state, select the one with a smaller path metric and discard the other. Trace the trellis and eliminate all the discarded trajectories.
3. Continue this process until only one surviving trajectory is left at some number of steps N behind the current step k. The trajectory constitutes the output of the ML detector up to step $k - N$.

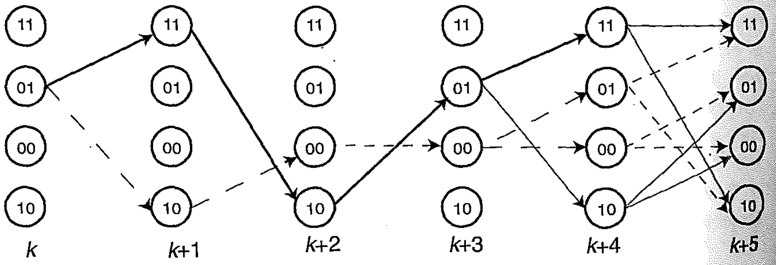
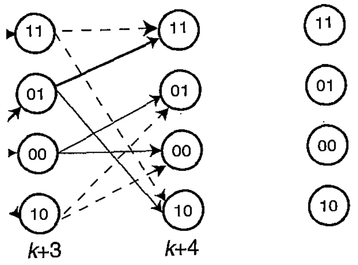


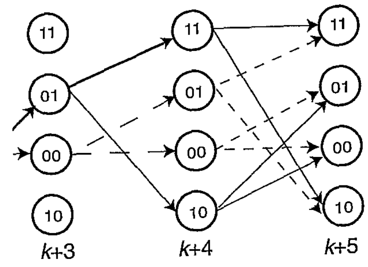
FIGURE 12.32. Extension of trellis (Fig. 12.31) to time instant $k + 5$.



ι Fig. 12.30 to time instant $k + 4$.

hs in the trellis coming to state
'00", "00"} starting at time k is
'ing trajectory {"01", "11", "10",
ıtes the solutions of the Viterbi
ime instant $k + 3$, and repeat the
trajectory.
arized as follows:

alculate the path matric for each
ading to the current state.
e same state, select the one with
'd the other. Trace the trellis and
tories.
one surviving trajectory is left at
the current step k. The trajectory
detector up to step $k - N$.



31) to time instant $k + 5$.

It is important to understand that the ML detector can provide decisions only after some delay, i.e., after one surviving path is found. Therefore, a special "path memory" is required to trace the history of the trajectories in the trellis. In general, it is impossible to predict how fast the competing paths will converge, but the probability of disagreement decreases exponentially with the number of steps.

Note that for any PRML scheme, once a trajectory is split into competing paths, zero samples will not update the ML decisions. A sequence of zero samples will effectively continue the "parallel" paths through the trellis. Therefore, a long sequence of zero samples, or equivalently a long string of zeroes in the NRZI data pattern, will increase the delay of the Viterbi algorithm. To eliminate long strings of zeroes, RLL encoding with (0,4/4) or (1,7) codes are used.

Another problem in the ML detector is the occurrence of a catastrophic or quasi-catastrophic sequence, which creates parallel paths through the trellis, i.e., the trajectories that will never converge to the same state. These sequences exist for a periodic pattern and in general have a very low probability of occurrence. To get rid of these sequences, special scramblers (randomizers) are used.

The number of steps that should be kept in the ML memory depends on the number of states n in the trellis. Theoretically, a memory of about one hundred steps may be necessary. In practice, however, a trajectory with the minimum metrics after a reasonable number of steps is considered to be the winning trajectory. Therefore, keeping about $5(n + 1)$ steps in the ML memory is generally enough for most practical applications.

12.4.3 Interleave and sliding threshold in PR4 channel

The PR4 polynomial is $H(D) = 1 - D^2$. The current sample $s(k)$ is obtained from NRZ data $\{a_k\}$ as $s(k) = (1 - D^2)a_k = a_k - a_{k-2}$. It means that the current sample depends only on the current NRZ bit and the NRZ bit two channel periods earlier, but not on the previous NRZ bit. Since the rule holds for any time instant, we can split the sequence of samples in PR4 channel into even and odd sequences and process them independently. For each of the even or odd sequences, we obtain a greatly simplified trellis as shown in Fig. 12.33, which is often called the *interleaved* PR4 trellis.

The interleaving property of the PR4 greatly simplifies the design of PR4 systems: even and odd samples are processed simultaneously in two simple ML detectors, each running at half of the original channel data rate, and are later combined into a single NRZ data stream, as shown in Fig. 12.34.

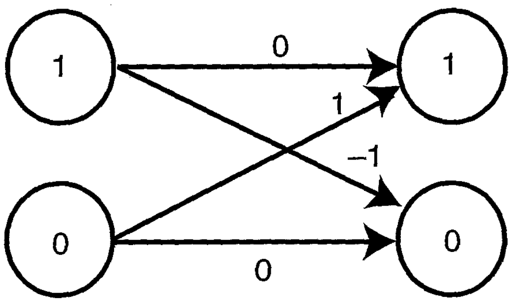


FIGURE 12.33. Interleaved PR4 trellis, equivalent to $1 - D$ channel.

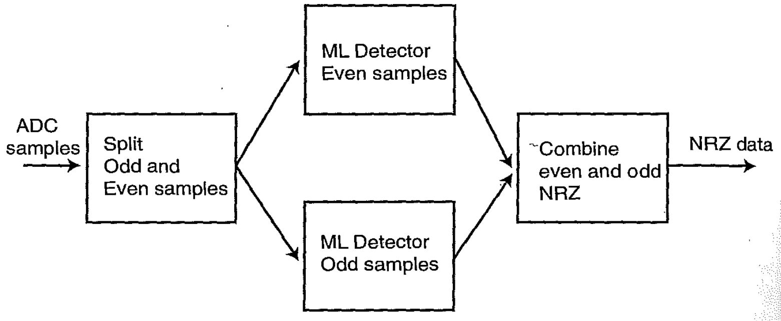
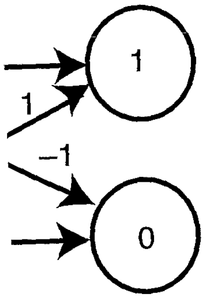


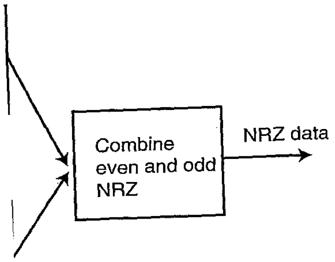
FIGURE 12.34. Interleaving ML detection for PR4 channel.

An interesting feature of PR4 interleaving is that the ML detector for the interleaved PR4 trellis shown in Fig. 12.33 can be realized with a simple threshold-detector-type scheme. To see how the ML detection can be obtained, we need to calculate the error metrics for the interleaved PR4 trellis. The metric is calculated by squaring the difference between the channel sample $s(k)$ and the ideal sample which may cause the corresponding transition from one state to another. The table of possible transitions and the corresponding metrics of the ML detection follows:

From State	To State	Ideal Sample	Squared Error
1	1	0	$[s(k) - 0]^2 = s^2(k)$
1	0	−1	$[s(k) - (-1)]^2 = s^2(k) + 2s(k) + 1$
0	0	0	$[s(k) - 0]^2 = s^2(k)$
0	1	1	$[s(k) - 1]^2 = s^2(k) - 2s(k) + 1$



lent to 1 − D channel.



ʔR4 channel.

ing is that the ML detector for
12.33 can be realized with a
see how the ML detection can
ɔr metrics for the interleaved
ıaring the difference between
e which may cause the corres-
ır. The table of possible transi-
ML detection follows:

Squared Error
$s(k) - 0]^2 = s^2(k)$
$s(k) - (-1)]^2 = s^2(k) + 2s(k) + 1$
$s(k) - 0]^2 = s^2(k)$
$s(k) - 1]^2 = s^2(k) - 2s(k) + 1$

We can now perform a simple transformation. Subtract $s^2(k)$ from the error metric of every possible path and divide the resulting error by 2, we obtain the following table:

From State	To State	Ideal Sample	[Squared Error − $s^2(k)$]/2
1	1	0	0
1	0	−1	$s(k) + 1/2$
0	0	0	0
0	1	1	$-s(k) + 1/2$

The squared error is now reduced to the sum or difference of the sample $s(k)$ and the threshold (1/2). This is very important for the hardware implementation of the Viterbi algorithm. The actual calculation of the metrics and their updates is demonstrated in Fig. 12.35. Assume that at instant k, state "1" has an accumulated metric $M1(k)$ and state "0" $M0(k)$. When the next sample $s(k+1)$ comes, the metric is updated as follows:

$$\begin{aligned} M0(k+1) &= \min\{M0(k), M1(k) + s(k+1) + 1/2\}, \\ M1(k+1) &= \min\{M1(k), M0(k) - s(k+1) + 1/2\}. \end{aligned} \tag{12.24}$$

There are only three possible extensions from the current step k to the next step $k + 1$ in the trellis, as shown in Fig. 12.36. The first extension takes place when the current state is "0". In this case, both the trajectories leading from "0" to "0" and from "0" to "1" should win over the competing

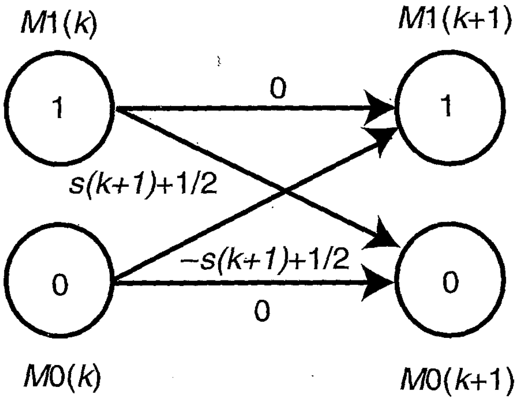


FIGURE 12.35. Metric updates for interleaved PR4 trellis.

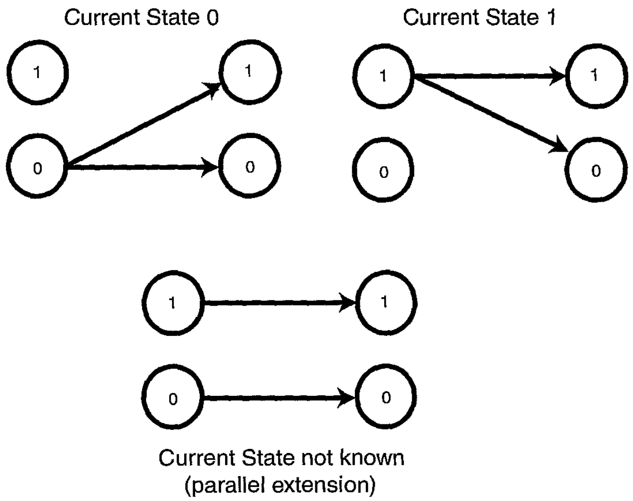


FIGURE 12.36. Three possible trellis extensions in interleaved PR4 trellis.

trajectories from state "1". Writing Equation (12.24) for this case, we obtain that

$$M0(k) - s(k+1) + 1/2 < M1(k), \qquad (12.25)$$
$$M0(k) < M1(k) + s(k+1) + 1/2.$$

The inequalities (12.25) are equivalent to the following:

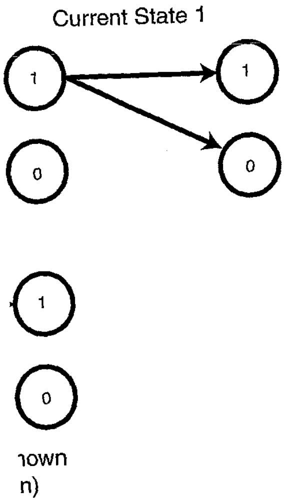
$$M0(k) - M1(k) < s(k+1) - 1/2. \qquad (12.26)$$

This means that it is enough to only look at the "difference metric":

$$D(k) = M0(k) - M1(k),$$

to determine whether state "0" was detected. If we now calculate the "updates" of the metrics $M0(k+1)$ and $M1(k+1)$ using Equation (12.24), the resulting "difference metric" at step $k+1$ equals: $D(k+1) = s(k+1) - 1/2$.

Similarly, we may consider the cases of the extension from state "1" and of parallel extension. In summary, we obtain the following set of simple rules for calculating the metrics and updates:



ons in interleaved PR4 trellis.

on (12.24) for this case, we obtain

'2 < $M1(k)$, (12.25)
+1) + 1/2.

the following:

:+1)−1/2. (12.26)

: at the "difference metric":

$M1(k)$,

ected. If we now calculate the 1(k + 1) using Equation (12.24), + 1 equals: $D(k + 1) = s(k+1)$

of the extension from state "1" ve obtain the following set of d updates:

Current State = 0: $s(k+1) > D(k) + 1/2;\ D(k + 1) = s(k+1) - 1/2.$ (12.27)

Current State = 1: $s(k+1) < D(k) - 1/2;\ D(k + 1) = s(k+1) + 1/2.$ (12.28)

Current State Unknown: $D(k) - 1/2 < s(k+1) < D(k) + 1/2;$ (12.29)
$D(k + 1) = D(k).$

For example, Equation (12.27) means that if the channel sample $s(k+1)$ exceeds the current difference metric $D(k)$ by 1/2, state "0" is detected and the new difference metric is given by $s(k+1) - 1/2$.

Equations (12.27)–(12.29) allow us to realize the PR4ML channel with an extremely simple scheme called the *sliding threshold,* as shown in Fig. 12.37. Assume that $D(1) = 0$ at the first step. If the incoming sample $s(2)$

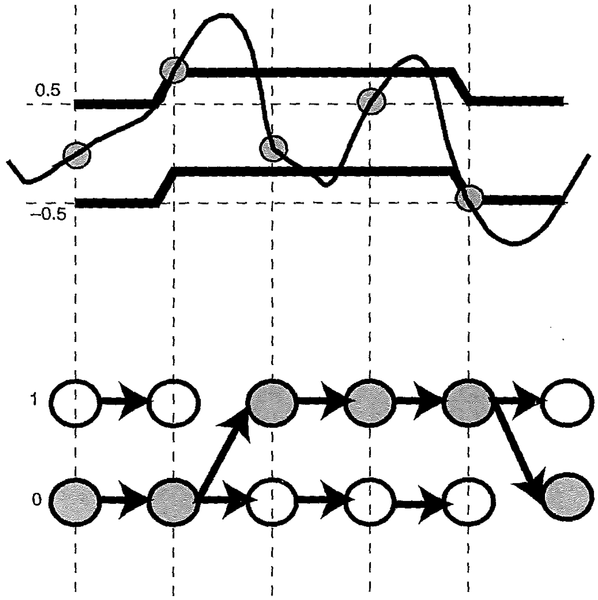


FIGURE 12.37. Realization of PR4ML detector with sliding threshold (top), and the corresponding trellis (bottom). The shaded states are the ML detector output.

$> 1/2$, e.g., $s(2) = 0.75$, state "0" is detected and the difference metric $D(2)$ is set to 0.25, i.e., the metric plays the role of an adjustable threshold. If the next sample is within the window from -0.25 to 0.75, the state cannot be determined, the difference metric is kept unchanged and the decision of the detector is postponed. However, once the incoming sample goes below the window, state "1" is detected and the parallel trajectory associated with state "0" in the trellis is discarded. In essence, the difference metric D is the center of a *sliding* detection window from $D - 0.5$ to $D + 0.5$. The window follows the samples, and the decisions of the ML detector are made after comparing the sample with the detection window.

Unfortunately, this simple realization of the ML detector can be used only for the interleaved PR4 system. Other types of PRML detectors cannot be simplified to such an extent, and usually require real-time metric calculation for each state (8 states for EPR4 and up to 16 states in E^2PR4).

12.4.4 Error events in maximum likelihood detection

An error occurs in the ML detector when the accumulated path metric for the wrong path through the trellis is smaller than the path metric of the correct path. While there are many possible trajectories that may diverge and converge in the trellis, only some of them are the dominant sources of errors.

When a sequence of channel samples $s_k = \{s_1, s_2, \ldots, s_N\}$ is coming to the ML detector, each of these samples equals the sum of the "ideal" noise-free sample s_k^0 and noise n_k: $s_k = s_k^0 + n_k$. Assume that two trajectories in the trellis diverge at $k = 1$ and converge back at $k = N$. The correct trajectory has noise-free samples s_k^0, while the erroneous trajectory has different sample values s_k^e. When the correct and the erroneous trajectories converge at time $k = N$, the accumulated path metrics for these two trajectories are:

$$M\ (\text{Correct}) = \sum_{k=1}^{N} (s_k - s_k^0)^2 = \sum_{k=1}^{N} n_k{}^2$$

$$M\ (\text{Wrong}) = \sum_{k=1}^{N} (s_k - s_k^e)^2 = \sum_{k=1}^{N} (s_k^0 + n_k - s_k^e)^2$$

$$= \sum_{k=1}^{N} (s_k^0 - s_k^e)^2 + 2 \sum_{k=1}^{N} ((s_k^0 - s_k^e)n_k + \sum_{k=1}^{N} n_k{}^2$$

ed and the difference metric D(2) ole of an adjustable threshold. If om −0.25 to 0.75, the state cannot cept unchanged and the decision once the incoming sample goes and the parallel trajectory associarded. In essence, the difference tion window from $D - 0.5$ to D es, and the decisions of the ML mple with the detection window. n of the ML detector can be used Other types of PRML detectors d usually require real-time metric R4 and up to 16 states in E^{2}PR4).

elihood detection

en the accumulated path metric smaller than the path metric of possible trajectories that may some of them are the dominant

$s_k = \{s_1, s_2, \ldots, s_N\}$ is coming to equals the sum of the "ideal" $+ n_k$. Assume that two trajectoverge back at $k = N$. The correct le the erroneous trajectory has ct and the erroneous trajectories ed path metrics for these two

$$n_k^2$$

$$(s_k^0 + n_k - s_k^e)^2$$

$$\sum_{k=1}^{N}((s_k^0 - s_k^e)n_k + \sum_{k=1}^{N} n_k^2$$

The ML detector will make an error if $M(\text{Correct}) - M(\text{Wrong}) > 0$, i.e.,

$$M\,(\text{Correct}) - M\,(\text{Wrong}) = -\sum_{k=1}^{N}(s_k^0 - s_k^e)^2 - 2\sum_{k=1}^{N}(s_k^0 - s_k^e)n_k > 0. \quad (12.30)$$

Only the second term in Equation (12.30) may become positive. Therefore, the smaller the amplitude of the first sum, the greater the probability of error of the ML detector. The first sum is called the *squared distance* between the two trajectories in the trellis.

For any type of PRML channels, certain trajectories in the trellis have a *minimum squared distance* between one another:

$$d_{\min}^2 = \sum_{k=1}^{N}(s_k^0 - s_k^e)^2. \quad (12.31)$$

For example, the PR4 system trellis shown in Fig. 12.30 has a minimum squared distance of 2. Other sequences have squared distances equal to 4, 6, etc. A wrong trajectory or its corresponding samples sequence is usually called an "error event." The squared distance between the error event and the correct sequence of samples is called the *squared distance of the error event.*

Finding *all* the possible error events is a complicated task, which is difficult to achieve in general. However, knowing the *typical* and *dominant* error events is necessary to analyze the error performance of PRML channels. The following table describes the sequences of samples corresponding to *all* the possible transitions from a state at step k to a state at step $k + 3$ in the PR4 trellis:

To: **From:**	**"11"**	**"01"**	**"00"**	**"10"**
"11"	0 0 0 −1 0 1	0 −1 0 −1 −1 1	0 −1 −1 −1 −1 0	0 0 −1 −1 0 0
"01"	1 0 0 0 0 1	1 −1 −1 0 −1 0	1 −1 −1 0 −1 0	1 0 −1 0 0 0
"00"	1 1 0 0 1 1	1 0 0 0 0 1	1 0 −1 0 0 0	1 1 −1 0 1 0
"10"	0 1 0 −1 1 1	0 0 0 −1 0 1	0 0 −1 −1 0 0	0 1 −1 −1 1 0

This table is very interesting. It demonstrates that there are only two possible three-element sequences between each two states. Note that all these sequences differ by the same sequence {−1,0,1}. These are exactly the samples of a dipulse in the PR4 system. This means that a *typical* short error event in the PR4 system looks like an extra or missing dipulse and no other short error events are possible. Therefore, the ML detector in the PR4 channel is sensitive to a specific noise burst resembling a dipulse response as shown in Fig. 12.38. In comparison, a single extra pulse or two unipolar noise bursts will not confuse the ML algorithm to the same degree. In other words, a dipulselike noise burst is more likely to cause an error event in the Viterbi detector.

The theory of Viterbi detection proves that longer error events have smaller probability to occur in the trellis. The probability of the error event typically falls off almost exponentially with the length of the error event. This means that the shortest error events create the main contribution to the total error rate of the system. The dominant error events in PRML channels are thus found to be as follows:

1. For the PR4 system with 8/9 rate code, the most probable error events are {1, 0, −1}, {1, 0, 0, 0, −1}, {1, 0, 0, 0, 0, 0, −1}, etc. Of course, both polarities of dipulses are possible, e.g., {−1, 0, 1}. These error events have a squared distance of 2, which is the

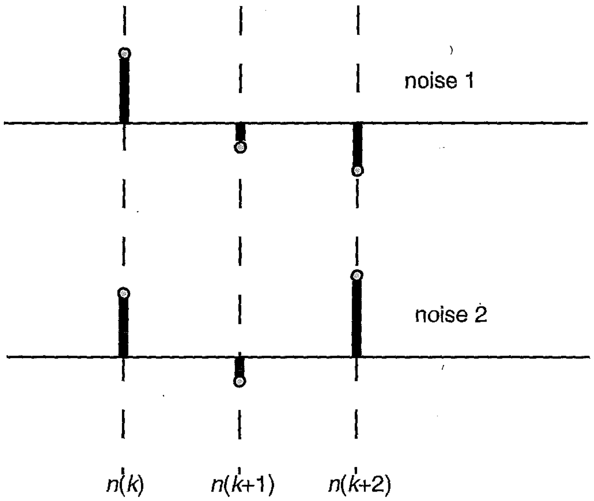
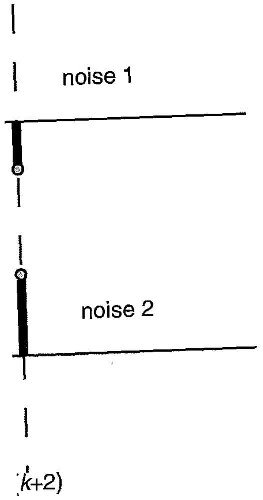


FIGURE 12.38. Illustration of dangerous (noise 1) and "harmless" (noise 2) noise samples in PR4ML detection.

ence {−1,0,1}. These are exactly
n. This means that a *typical* short
an extra or missing dipulse and
herefore, the ML detector in the
oise burst resembling a dipulse
parison, a single extra pulse or
e the ML algorithm to the same
ise burst is more likely to cause

es that longer error events have
is. The probability of the error
ally with the length of the error
events create the main contribu-
.. The dominant error events in
follows:

: code, the most probable error
-1}, {1, 0, 0, 0, 0, 0, −1}, etc. Of
s are possible, e.g., {−1, 0, 1}.
ed distance of 2, which is the



se 1) and "harmless" (noise 2) noise

minimum squared distance. Other error events have a squared distance of 4 or higher, and their contribution to the error rate of the system is negligible.

2. For the EPR4 system with 8/9 rate code, the most probable error events are {1/2, 1/2, −1/2, −1/2}, {1/2, 1/2, 0, 0, −1/2, −1/2}, {1/2, 0, −1/2, 0, −1/2, 0, 1/2}, etc. The minimum squared distance is 1. Other error events may have a squared distance of 1.5, such as {1/2, 0, −1, 0, 1/2} and {1/2, 1/2, −1/2, 0, 1/2, −1/2, −1/2}.
3. For the E^{2}PR4 system with 8/9 rate code, the most probable error event is {1/3, 1/3, −1/3, 0, 1/3, −1/3, −1/3} with a minimum squared distance of 6/9 = 0.67.
4. For the E^{2}PR4 system with (1,7) code, the only minimum distance error sequence is {1/3, 2/3, 0, −2/3, −1/3}, so the minimum squared distance is 10/9 = 1.11. The next probable error event has a squared distance of 12/9 = 1.33 and is given by {1/3, 2/3, 0, −1/3, 1/3, 0, −2/3, −1/3}

The knowledge of minimum distance error events can be used to estimate the error rate of the ML detector. In the case that there is only *additive white Gaussian noise* (AWGN) in the channel, the error rate can be readily determined analytically. The error events may be caused by the noises in the samples so that the received samples are $s(k) = s_0(k) + n(k)$ where $s_0(k)$ are the ideal samples. Using the *vector representation,* we may plot the correct and wrong sequences of samples in an N-dimensional space, where N is the length of the error event, as shown in Fig. 12.39. In any real channels, we receive samples plus noise instead of ideal samples. The Gaussian noise in the N-dimensional space has an isotropic distribution and is described by a noise "sphere" in which all directions are equivalent. Therefore, any received sequence of samples is lying somewhere inside the sphere. A correct sequence will be recognized by the ML detector if the received vector is closer to the correct sequence than to the wrong sequence. According to Equation (12.30), if the distance between the correct and wrong sequences is d, the probability of an error event is simply the probability of the following event:

$$-\sum_{k=1}^{N}(s_k^0 - s_k^e)n_k > d^2/2.$$

In the case that the sample noises are AWGN, the *probability of an error event* in the ML detector is $Q(d^2/2\sigma)$, where the Q-function is defined in Chapter 11 [see Equations (11.2)-(11.5)], and σ is the standard deviation

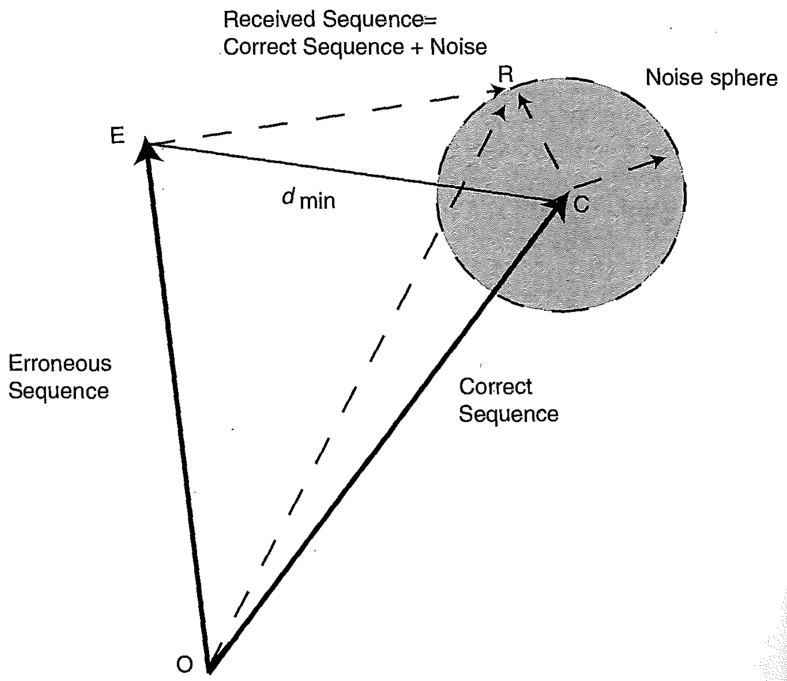
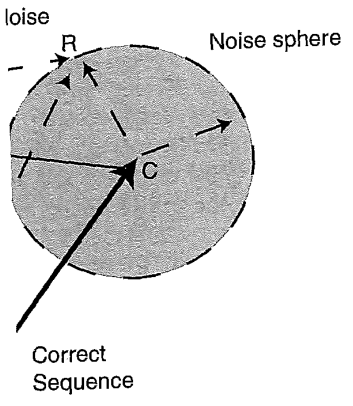


FIGURE 12.39. Vector representation of correct/wrong sequences and noise. The ML detector makes no error if |CR| < |ER|.

of the noise term $-\Sigma_{k=1}^{N}(s_k^0 - s_k^e)n_k$. Note that the Q-function is a very steep function when the argument is >10.

At reasonably low error rates, the main contribution to the error rate is from the minimum distance error events because other error events have much larger distances from the correct vector and the contributions from such error events are typically several orders of magnitude smaller. Therefore, the error rate of the ML detector is determined by $Q(d_{\min}^2/2\sigma)$, where $d_{\min}$ is the minimum distance in the channel trellis.

Now let us compare the error rate of the ML detector with that of a simple threshold detector. The distance between the PRML ideal samples is A, which is 1 for the PR4 system, 0.5 for the EPR4 system, and 1/3 for



ct/wrong sequences and noise. The

; that the Q-function is a very
).
nain contribution to the error
or events because other error
n the correct vector and the
re typically several orders of
r rate of the ML detector is
the minimum distance in the

the ML detector with that of a
tween the PRML ideal samples
r the EPR4 system, and 1/3 for

the E^2PR4 system, respectively. The threshold detector can distinguish between any two ideal samples if the sample noise is less than $A/2$. Therefore, the error rate of the threshold detector is given by $Q(A/2\sigma_s)$, where σ_s is the rms noise of each sample, which is *different* from σ! Note that, for the minimum distance error events,

$$\sigma = d_{\min}\cdot\sigma_s,$$

so the ML detector error rate is

$$P_{ML} = Q\left(\frac{d_{\min}^2/2}{\sigma}\right) = Q\left(\frac{d_{\min}/2}{\sigma_s}\right).$$

The last expression means that if one of the sample values in the received vector exceeds half of the minimum distance, an error event will occur, which is consistent with Fig. 12.39.

For the PR4 system, $d_{\min} = \sqrt{2}$, $A = 1$, so the ML detector error rate is $Q(0.7/\sigma_s)$, while that of the threshold detector is $Q(0.5/\sigma_s)$. Simply put, the PR4ML detector has an SNR gain of 3 dB over the simple threshold detector. For the EPR4 system, $d_{\min} = 1$, $A = 1/2$, so the ML detector error rate is $Q(0.5/\sigma_s)$, while that of the threshold detector is $Q(0.25/\sigma_s)$. An SNR gain of 6 dB over the simple threshold detector is realized. For the E^2PR4 system with (1,7) encoding, $d_{\min} = 10/9$, $A = 1/3$, so the ML detector error rate is $Q(0.556/\sigma_s)$, while that of the simple threshold detector is $Q(0.167/\sigma_s)$. The SNR gain in the last case is 10.5 dB.

The above results do not tell us the relative merit of the three PRML detectors, which is ultimately established by error margin analysis (see Section 12.5). Note that the value of σ_s in each case is normalized by the maximum sample amplitude just like the sample values. Consequently, for a given readback signal and noise, the σ_s value will be dependent on what PRML channel is implemented. The above results are obtained without taking noise correlation, nonlinearities, media defects, and interference signals into account. If the noise statistics is not known, the probability of errors must be estimated based on comparing the path metrics for the correct and the erroneous trajectories.

The ML detector makes an error when a wrong path through the trellis is chosen instead of a correct one. The received sequence of samples at the input of the ML detector is noisy and can be represented by $s(k)$ $= s_0(k) + n(k)$, where $s_0(k)$ are ideal PRML sample values, and $n(k)$ are the noise samples. If an error is made during the detection process, a "wrong" sequence of samples $b(k)$ is detected. Note that an arbitrary "wrong" sequence $b(k)$ can always be represented as $b(k) = s_0(k) + m(k)A$, where $m(k)$ is the number of levels between $s_0(k)$ and $b(k)$ and A is the minimum step between ideal PRML sample values. For example, $A = 1$ for the PR4ML system. Assume that an error event has a finite length of N samples, then an error occurs in the ML detector if

$$\sum_{k=1}^{N} [s(k) - s_0(k)]^2 > \sum_{k=1}^{N} [s(k) - b(k)]^2. \tag{12.32}$$

Substituting $s(k) = s_0(k) + n(k)$ and $b(k) = s_0(k) + m(k)A$ into Equation (12.32), we obtain that

$$\sum_{k=1}^{N} [n(k)]^2 > \sum_{k=1}^{N} [n(k) - m(k)A]^2, \tag{12.33}$$

Opening the brackets in Equation (12.33), we obtain the following condition for the occurrence of an error event in the ML detector:

$$\frac{1}{\sum_{k=1}^{N} [m(k)]^2} \sum_{k=1}^{N} m(k)n(k) > \frac{A}{2}. \tag{12.34}$$

This means that an error occurs if a linear combination of noise samples, given by the left side of Equation (12.34), exceeds a threshold equal to $A/2$. The coefficients $m(k)$ describe how the "correct sequence" samples are different from the "wrong" sequence samples, i.e., they are the sample values of the error events. Therefore, the ML detector works like a linear noise *filter.* Remember that a typical error event for the PR4 channel with (0,4/4) code is the error sequence $m(k) = \{-1,0,1\}$, corresponding to detecting a false dipulse instead of zero signal or missing a real dipulse. It follows from Equation (12.34) that such an error occurs when $|n(k) - n(k-2)| > 1$.

The noise correlation greatly affects the error rate of the PRML channel, and the correlated noise in multidimensional space is no longer spherical. An intuitive model of the correlated noise is shown in Fig. 12.40. The "negatively" correlated noise, corresponding to the type 1 noise in Fig. 12.38 distorts the spherical distribution into the direction of the erroneous sequence, while the "positively" correlated noise bends the noise distribution along the correct sequence of data. Obviously, the negative noise

te that an arbitrary
$b(k) = s_0(k) + m(k)A$,
ıd $b(k)$ and A is the
For example, $A = 1$
ıas a finite length of
if

^{2}. (12.32)

$n(k)A$ into Equation

(12.33)

the following condi-
detector:

(12.34)

on of noise samples,
a threshold equal to
: sequence" samples
, they are the sample
tector works like a
t for the PR4 channel
,0,1}, corresponding
1 or missing a real
n error occurs when

e of the PRML chan-
e is no longer spheri-
wn in Fig. 12.40. The
type 1 noise in Fig.
tion of the erroneous
ds the noise distribu-
, the negative noise

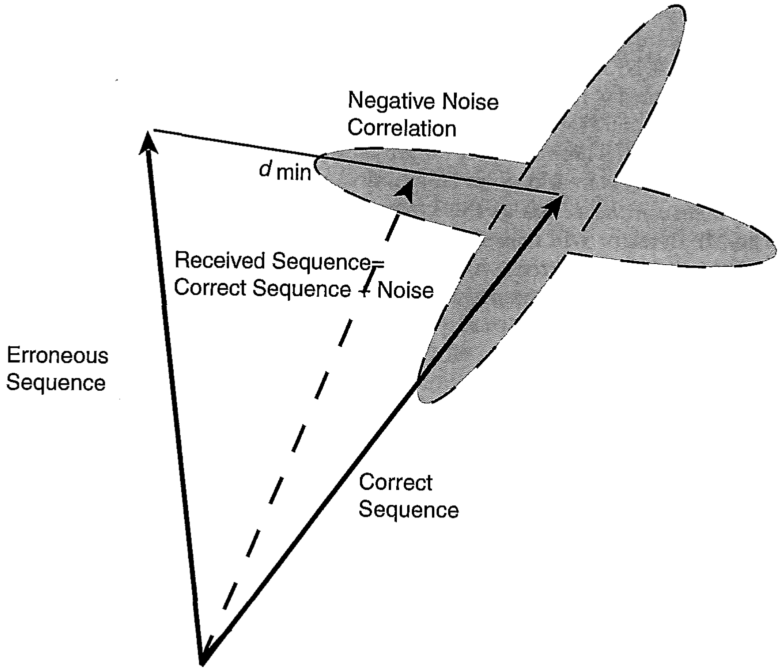

Positive Noise Correlation
Negative Noise Correlation
d_{min}
Received Sequence= Correct Sequence + Noise
Erroneous Sequence
Correct Sequence

FIGURE 12.40. Influence of correlated noise on error rate of the ML detector.

correlation is not good for achieving a lower error rate and thus should be avoided.

12.5 PRML ERROR MARGIN ANALYSIS

Like any magnetic recording systems, PRML channels should provide extremely low BER (10^{-9} or better). When the system is optimized, the channel BER is low and direct error rate measurements from the output of the PRML channel are time consuming. For example, if the error rate is on the order of 10^{-8} and at least 100 errors are to be processed to obtain a reliable statistical estimate of the channel error rate, this means that at least 10^{10} bits are to be read from the channel and analyzed. Assume that the channel data rate is 100 Mbit/s, then the error rate measurement will take at least 1.5 minutes just to acquire the data.

A standard way of analyzing the error rate performance of the channel is stressing the system and checking how many errors appear at the system output. There are several ways to stress the PRML channel.[10–12] One approach is to add noise from a noise generator to the system input and measure the errors at the system output for different noise levels. The ML detector will make a large number of errors at high noise levels and the number of errors will gradually decrease when the amplitude of the external noise becomes smaller. Another way is to move the head away from the on-track position in order to increase the noise level in the readback signal. In this "off-track" test the dependence of the number of errors on the off-track position can be obtained.

While both of the described approaches are widely used, they do not fully reflect the system in normal conditions. Externally injected noise has different correlation and spectral energy distribution from head and medium noises. Off-track stressing is often used to characterize off-track system performance (Chapter 14), but it does not reflect the normal system in the on-track position. The degradation of PRML system performance is often due to NLTS, PE, and transition noises, which are different from the off-track noise and are not fully reflected in the off-track test.

A flexible way for stressing the PRML system is provided by the *sequenced amplitude margin* (SAM) method,[11] which is somewhat similar to the time margin analysis for the peak detection channel. The SAM algorithm is based on determining how close the ML detector is to making an error when it selects surviving trajectories in the trellis.

Consider a portion of the trellis for the PR4 system as shown in Fig. 12.41, where we assume that the correct data pattern is known *a priori.* There are two cases in which we can make this assumption. One is when there is an actual Viterbi detector in the system and the error rate is low so that the outputs of the Viterbi detector can be trusted. Another case is more related to a test equipment environment. In this case, the known pattern of data is written on a disk, and proper synchronization between written and read data is provided.

The correct path through the trellis is shown in Fig. 12.41 as the thick line, i.e., a transition from state "01" to state "11" has occurred, with an ideal sample value of $s_0(k+1) = 1$. In a real system, some noisy sample value $s(k+1) = s_0(k+1) + n(k+1)$ is received. At step $k+1$ we already have

rrors are to be processed to obtain
nel error rate, this means that at
annel and analyzed. Assume that
the error rate measurement will
the data.
r rate performance of the channel
ow many errors appear at the
to stress the PRML channel.[10–12]
ise generator to the system input
output for different noise levels.
ber of errors at high noise levels
decrease when the amplitude of
other way is to move the head
er to increase the noise level in
t the dependence of the number
obtained.
es are widely used, they do not
itions. Externally injected noise
rgy distribution from head and
n used to characterize off-track
es not reflect the normal system
of PRML system performance
oises, which are different from
cted in the off-track test.
ML system is provided by the
d,[11] which is somewhat similar
k detection channel. The SAM
se the ML detector is to making
ries in the trellis.
e PRML system as shown in Fig.
data pattern is known *a priori.*
this assumption. One is when
ystem and the error rate is low
can be trusted. Another case is
ment. In this case, the known
oper synchronization between

hown in Fig. 12.41 as the thick
te "11" has occurred, with an
al system, some noisy sample
l. At step k+1 we already have

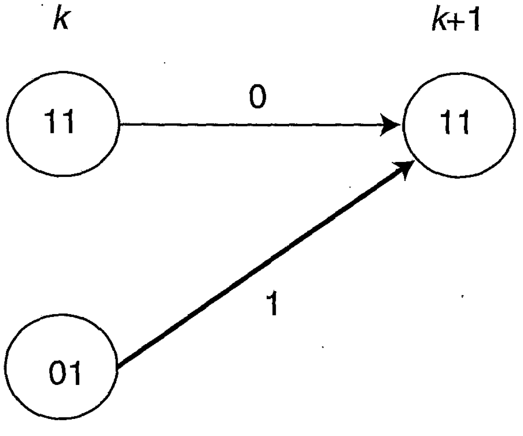


FIGURE 12.41. Part of the PR4 system trellis. The correct trajectory is shown as the thick line.

two accumulated path metrics: $M(''11'')$ and $M(''01'')$. The Viterbi detector makes a correct decision at step $k + 1$ if

$$M(''01'') + [s(k+1) - 1]^2 < M(''11'') + [s(k+1) - 0]^2,$$

or

$$B(k+1) = M(''01'') - M(''11'') + [n(k+1)]^2 - [s(k+1) - 0]^2 < 0. \qquad (12.35)$$

More generally, the left side of Equation (12.35) can be rewritten as

$$B(k+1) = M(''S1'') - M(''S2'') + [s(k+1) - m1]^2 - [s(k+1) - m2]^2, \qquad (12.36)$$

where $S1$ and $S2$ are the two previous states leading to the correct state, $m1$ and $m2$ are the corresponding ideal sample values, and $s(k+1)$ is the sample value.

The sample $s(k+1)$ equals $m1 + n(k+1)$. The difference between $m1$ and $m2$ is constant, at least for the standard realization of Viterbi detectors. For examples, $|m1 - m2| = 1$ for PR4, $|m1 - m2| = 1/2$ for EPR4, and $|m1 - m2| = 1/3$ for E^{2}PR4 system. Therefore,

$$B(k+1) = M(''S1'') - M(''S2'') - (m1 - m2)^2 - 2(m1 - m2)n(k+1), \qquad (12.37)$$

An error will occur if $B(k+1) > 0$.

If we know path metrics for the two previous states leading to the correct state, we may easily calculate the left side of Equation (12.37).

Even if $B(k+1)$ is always less than zero, it has a distribution of values as shown in Fig. 12.42. Since the distribution (histogram) of $B(k+1)$ is known, we may integrate it from a variable threshold C to $+\infty$:

$$Q(C) = \int_C^{\infty} P(B)dB, \tag{12.38}$$

where P is the probability density function (PDF) of B. If it is Gaussian with a mean value of zero, then the function is the same as the Q-function defined previously. The value of $Q(0)$ corresponds to the error rate at the nominal threshold since an error occurs if $B(k+1) > 0$. Similarly, $Q(C)$ represents the expected error rate if the decision threshold in the Viterbi algorithm is reduced to C. In this case, an error occurs if $B(k+1) > C$.

In essence, the SAM algorithm is based on obtaining a histogram of the differences of the path metrics and integrating it with a variable threshold. During the normal operation of the ML detector, the threshold is set to "0", i.e., the correct trajectory is selected when the accumulated path metric $M(\text{correct}) - M(\text{wrong}) < 0$. This nominal threshold value

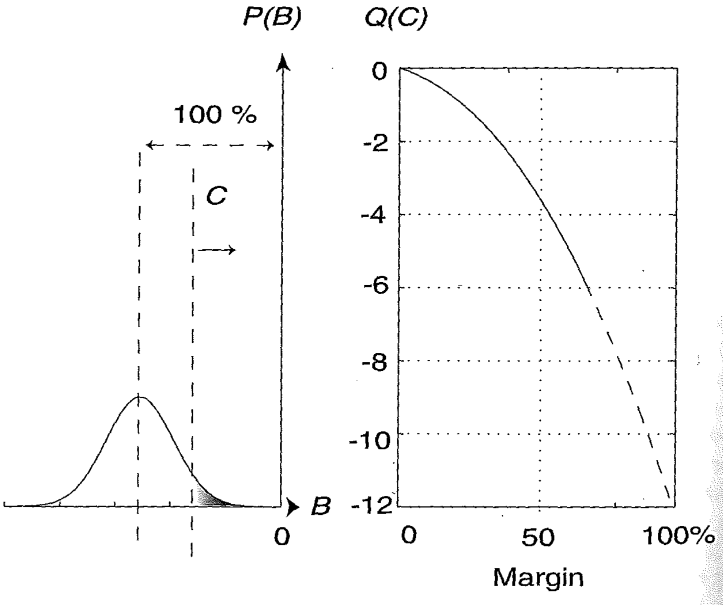
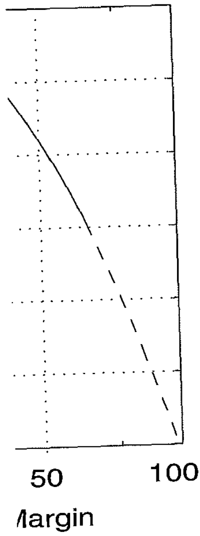


FIGURE 12.42. Illustration of the SAM plot generation. Left: distribution of values $B(k+1)$. Right: SAM plot obtained by integrating $P(B)$ with variable threshold.

ɜtribution of values as
n) of $B(k+1)$ is known,
ı $+\infty$:

(12.38)

B. If it is Gaussian with
the Q-function defined
ror rate at the nominal
ilarly, $Q(C)$ represents
he Viterbi algorithm is
$> C$.
taining a histogram of
ng it with a variable
detector, the threshold
vhen the accumulated
minal threshold value



.eft: distribution of values
vith variable threshold.

corresponds to a *sequenced amplitude margin* of 100% in Fig. 12.42. Decreasing the decision threshold to C such that the correct paths are selected based on $M(\text{correct}) - M(\text{wrong}) < C$, where C is a negative number, causes the errors at the output of the ML detector to increase accordingly. A plot of the error rate vs. the decision threshold is called the *sequenced amplitude margin* (SAM) *plot*, as shown in Fig. 12.42 (right). The very small error rates are often *extrapolated* from the experimental points at the relatively large error rates to save time. The mean value of B tends to be the minimum squared distance in the trellis, so the *sequenced amplitude margin* (%) can be defined as follows:

$$\text{Margin (\%)} = \frac{C + d_{\min}^2}{d_{\min}^2}$$

Note that the definition here is different from the *operation margin*, which is understood as the "room" left to the nominal decision threshold to maintain the error rates below the specified level. If the decision threshold is $C < 0$ at the specified error rate level, then

$$\text{Operation Margin (\%)} = 1 - \text{Margin (\%)} = \frac{-C}{d_{\min}^2}.$$

For example, if the specified error rate is 10^{-10}, which is reached when the decision threshold is reduced to minus one-tenth of the minimum squared distance, then the operation margin at 10^{-10} is 10%. Obviously, a steeply descending SAM plot is desired to have a large operation margin.

A typical histogram of path metric difference is illustrated in Fig. 12.43. The minimum distance errors correspond to the main peak of the distribution. The next distance error events may create one or more additional peaks. However, the dominant source of errors in the PRML channel is generated by the minimum distance error events, so the reasonable normalization of the sequenced amplitude margin is to assign the 0% margin to the peak of this distribution, or the mean value of B, as was done previously.

Note that an equivalent SAM plot can be obtained based on Equation (12.34). If the probability density function of the left side is $p(f)$, where

$$f = \frac{\sum_{k=1}^{N} m(k)n(k)}{\sum_{k=1}^{N} [m(k)]^2},$$

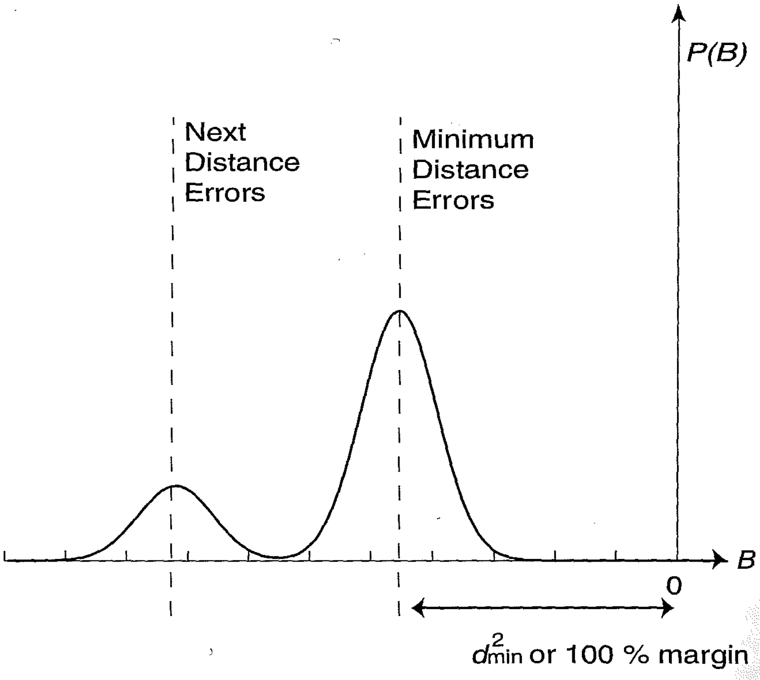


FIGURE 12.43. Histogram of the path metric differences in the ML detector.

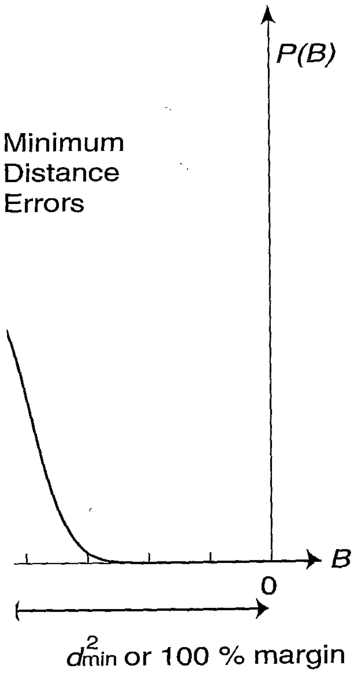
then the error rate of the PRML channel is

$$Q(A/2) = \int_{A/2}^{\infty} p(f)df.$$

In the case of the PR4ML channel, the 100% margin is assigned to $A = 1$, and the 0% margin is assigned to $A = 0$. Note that the standard deviation of f is not the same as that of B.

Some experimental SAM plots are shown in Fig. 12.44, where the differences between equalization result in horizontal shifts of the SAM plot. Misequalization boosts the noise in the system, and off-track position introduces additional noise into the system. Both reduce the SAM plot slope and the operation margin.

The SAM plot is a powerful tool for evaluating the performance of PRML channels. Changes in the slope and shifts of the SAM plot similar to those shown in Fig. 12.44 are important clues to determining



c differences in the ML detector.

is

$p(f)df$.

0% margin is assigned to $A = 1$,
Note that the standard deviation

shown in Fig. 12.44, where the
n horizontal shifts of the SAM
ne system, and off-track position
em. Both reduce the SAM plot

or evaluating the performance
e and shifts of the SAM plot
mportant clues to determining

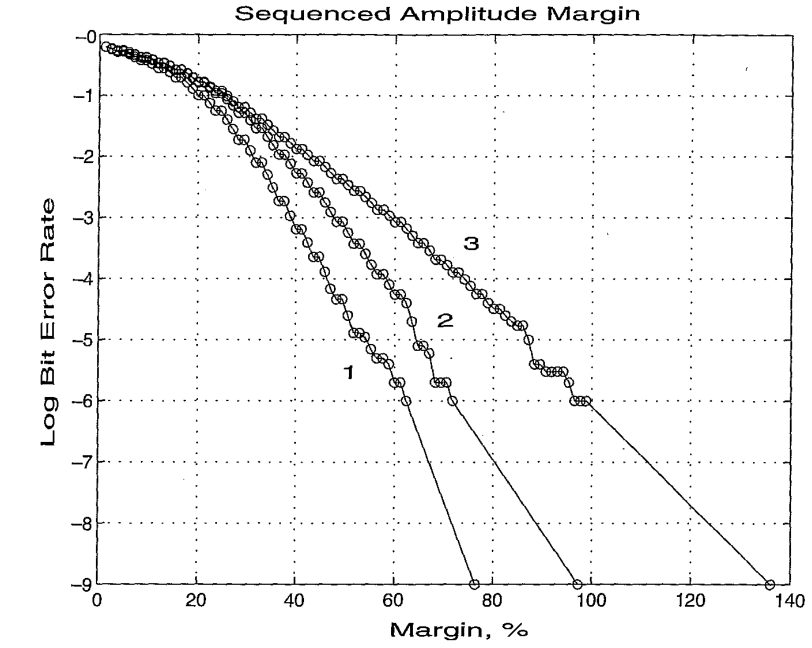


FIGURE 12.44. Examples of SAM plot: optimal equalization and on-track measurement (curve 1); misequalized signal (curve 2); off-track measurement (curve 3).

what are affecting the error rate. The SAM plot also predicts the actual BERs in the real operation conditions. For example, the SAM plots in Fig. 12.44 are calculated based on 10^6 samples and extrapolated to 10^{-9} level so that the operation margin at an error rate of 10^{-9} can be predicted.

12.6 PERFORMANCE OF PRML CHANNELS

PRML channels achieve their best performance when the readback signal is accurately equalized to the desired target shape without excessive noise amplification, and simultaneously the magnetic recording channel is maintained linear. Any deviation of the isolated pulse shape and any nonlinear distortion in the channel will greatly degrade the error rate of the PRML channel. However, these do not translate into the same impact on the ML detection process. As discussed in the previous sections, the ML detector is affected by the noise correlation properties. The noises and nonlinear distortions that do not match the probable error patterns are much less critical. For example, relatively low-frequency modulation of the system gain will not affect the performance of the ML detector.[13,14]

A simplified analysis of the error rate of the PRML system is based on considering two different noise sources. The first noise source is a random, irregular noise n, introduced by head, media, electronics, and clock jitter. The second noise source is "shape distortion" noise d. This distortion term describes regular deviations of the pulse shape from ideal due to misequalization, pulse asymmetry, and nonlinear distortions. These distortions are regular in the sense that they may be predicted. For examples, distortions caused by NLTS may be described as a deviation from an ideal dipulse shape; distortions caused by misequalization can be described as a deviation from a target isolated pulse.

Using the definition of the random and shape distortion noise terms, the total noise of each sample equals $n(k) + d(k)$. Equation (12.34) can now be rewritten as

$$\frac{\sum_{k=1}^{N} m(k)n(k)}{\sum_{k=1}^{N} [m(k)]^2} + \frac{\sum_{k=1}^{N} m(k)d(k)}{\sum_{k=1}^{N} [m(k)]^2} > \frac{A}{2}. \tag{12.39}$$

which is the condition of an error event. Note that the last term in Equation (12.39) represents the "correlation" of the shape distortion pattern with the error events in the ML detector. In other words, if the shape distortion "matches" the shape of the error event, this term will become large and the probability of error will increase.

Consider the PR4 system. The most probable error pattern for PR4 detection is given by $m(k) = \{1,0,-1\}$ or $\{-1,0,1\}$. Therefore, NLTS and PE are extremely dangerous because they match the error pattern. In fact, the reduction of dipulse amplitude by any means is "matched" with the most probable error pattern. Similarly, any misequalization that results in different signs of distortion two bit periods apart is also critical for a PR4 system.

recording channel is
pulse shape and any
grade the error rate of
e into the same impact
previous sections, the
properties. The noises
robable error patterns
-frequency modulation
of the ML detector.[13,14]
PRML system is based
first noise source is a
nedia, electronics, and
stortion" noise d. This
pulse shape from ideal
inear distortions. These
e predicted. For exam-
ed as a deviation from
isequalization can be
lse.
distortion noise terms,
. Equation (12.34) can

$$\frac{A}{2}. \qquad (12.39)$$

e last term in Equation
listortion pattern with
if the shape distortion
will become large and

error pattern for PR4
herefore, NLTS and PE
error pattern. In fact,
is "matched" with the
ualization that results
rt is also critical for a

The left side of Equation (12.39) consists of two terms, each being the linear combination of the corresponding noise component with the coefficients of the error sequence $m(k)$. Let the *probability density function* (PDF) of the first term be $p(f)$ and that of the second term be $g(f)$, and assume that the two noise terms are statistically independent, then the PDF of the left side of Equation (12.39) is given by the convolution of both distributions:

$$P(f) = \int_{-\infty}^{\infty} p(f - x)g(x)dx. \qquad (12.40)$$

Therefore, the error rate of the PRML channel is

$$G(A/2) = \int_{A/2}^{\infty} P(f)df = \int_{-\infty}^{\infty} Q(A/2 - x)g(x)dx, \qquad (12.41)$$

where

$$Q(A/2) = \int_{A/2}^{\infty} p(f)df,$$

Note that A is the distance between the ideal PRML sample values. Therefore, *the total error rate $G(A/2)$ of a PRML system is the convolution of the error rate function $Q(A/2)$ of a linear and ideally equalized system with the probability density function of the distortion term.* The latter includes the shape distortions due to nonlinearities and misequalization.[13]

Function $Q(A/2)$ represents the error rate due to the random noise term. It has a well-known bell-like (or "waterfall") shape, with its slope determined mainly by the standard deviation of the random noise and by the particular set of coefficients $m(k)$. Note that the standard deviation of the noise in the ML detector depends on the system equalization, which may alter the noise bandwidth and may boost noise at some frequencies.

In the presence of *misequalization,* which excludes the boost effect of equalization on random noise, the PRML signal samples at clock locations contain a discrete set of noise values $d(k)$, resulting from the superposition of the misequalization errors at current and neighboring pulses. For example, if an ideal isolated PR4 pulse has values {0,1,1,0} and the misequalized pulse samples are {0,0.9,1,0.1}, then the set of $d(k)$ is {±0.1,±0.2}. Any linear combination $d(k)$ in Equation (12.39) will result in additional errors. In the case of $m(k) = \{1,0,-1\}$, the distribution $g(f)$ for a random data pattern can take these discrete values: {0,±0.05,±0.1,±0.15}.

The main impact of NLTS and PE is to decrease PRML sample amplitudes for adjacent transitions, which generates a set of $d(k)$ and additional discrete values in the distribution $g(f)$. Figure 12.45 demonstrates the histogram of $g(f)$ in a low-noise channel. The distribution was obtained using simulated signals with artificially introduced NLTS equal to 25% of the bit period. Note that $g(f)$ consists of a main peak at the zero location and a number of additional peaks. The most distant noticeable peak at approximately $f = \pm.33$ has an amplitude of about 10% of the main peak amplitude.

According to Equation (12.41), the convolution of the distribution of the random noise term with $g(f)$ gives the total error rate, which is approximately the superposition of many shifted distributions $Q(A/2)$ weighted with the amplitudes of the corresponding peaks. The result of the convolution is shown in Figure 12.46. Note that the resulting error rate at about 33% margin is close to 10% of the original $Q(A/2)$ value at 0% margin.

Nonlinear distortions are especially critical in PRML channels because they are matched to the probable error patterns. This is true for all PRML systems such as PR4, EPR4 and E^2PR4 because the most probable error

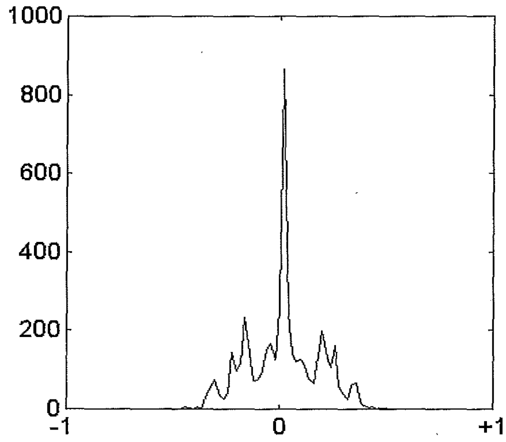
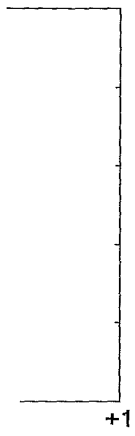


FIGURE 12.45. Distribution $g(f)$ (not normalized) in a low-noise channel with 25% NLTS.

ecrease PRML sample ampli-
es a set of $d(k)$ and additional
gure 12.45 demonstrates the
e distribution was obtained
oduced NLTS equal to 25%
ain peak at the zero location
t distant noticeable peak at
about 10% of the main peak

volution of the distribution
e total error rate, which is
hifted distributions $Q(A/2)$
onding peaks. The result of
ote that the resulting error
he original $Q(A/2)$ value at

in PRML channels because
s. This is true for all PRML
se the most probable error



a low-noise channel with

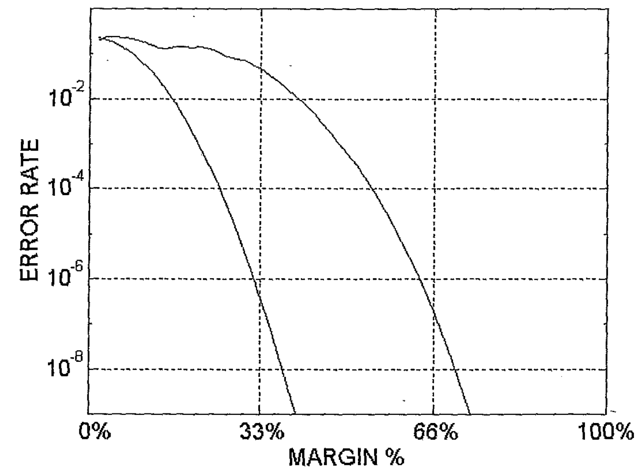


FIGURE 12.46. Convolution of $Q(f)$ (left curve) with $H(f)$ in Fig. 12.46. The result (right curve) is close to a shifted copy of $Q(f)$.

events for all of them have a dipulse shape. The minimum distance error sequences are listed in the following:

PR4:	$\{1, 0, -1\}$
EPR4:	$\{1/2, 1/2, -1/2, -1/2\}$
E^{2}PR4:	$\{1/3, 2/3, 0, -2/3, -1/3\}$

The shape of these error events indeed matches the dipulse shape. Therefore, any nonlinear distortions, which reduce the dipulse amplitude, will increase the probability of errors in the ML detector.

A moderate amount of NLTS can be *precompensated* in the write process by intentionally delaying a transition in the presence of previous transitions. The effects of write precompensation are demonstrated in Figs. 12.47 and 12.48. The sample histograms without and with precompensation are shown in Fig. 12.47, which indicates that the separation between the sample values is greatly improved by precompensation.

The SAM plots corresponding to the histograms in Fig. 12.47 are shown in Fig. 12.48. Precompensation of NLTS greatly increases the slope of the SAM plot, just like reducing the random noise in the channel. Consequently, the operation margin of the channel increases accordingly.

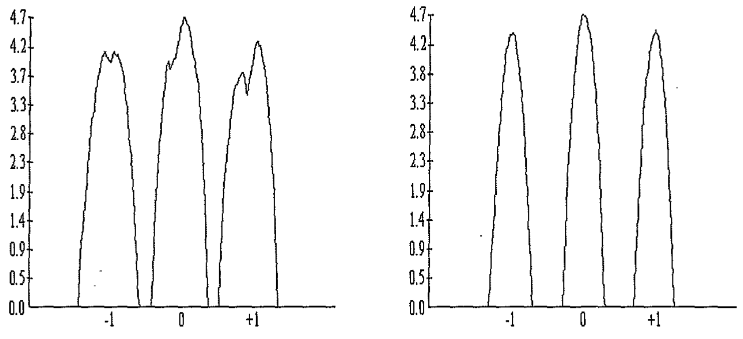


FIGURE 12.47. Sample values distributions without precompensation (left) and with precompensation (right). NLTS equals 10% of the bit period. The histograms are shown on a logarithmic scale.

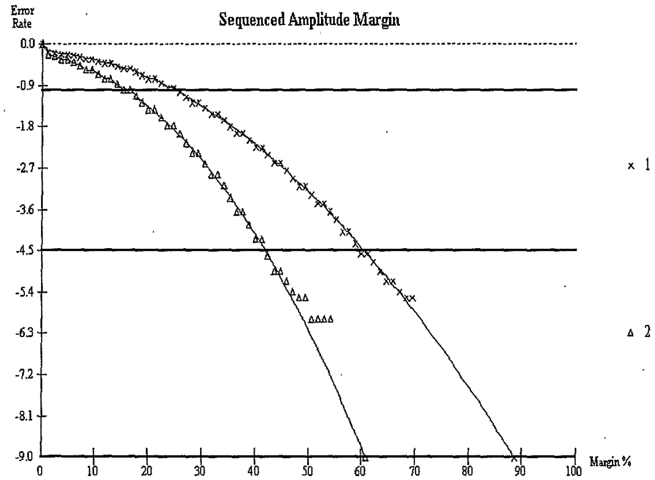
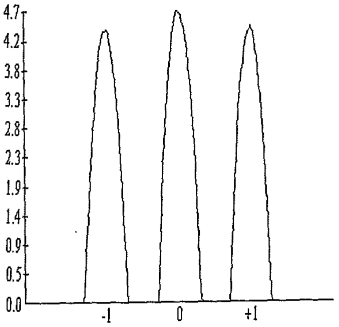
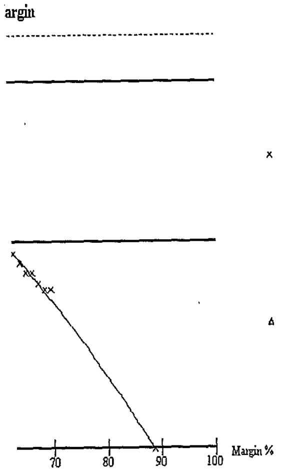


FIGURE 12.48. SAM plots for PRML channel without precompensation (right) and with precompensation of NLTS (left).



without precompensation (left) and .0% of the bit period. The histograms



:l without precompensation (right)

At high recording densities when NLTS approaches 20% of the bit period, the situation becomes more complicated. If all the transitions in a random pattern are precompensated the same amount that is optimal for a dipulse, the error rate performance is *not* optimized, as illustrated in Fig. 12.49. If a random data pattern is written on the disk and all the transitions in the pattern are precompensated using the same precompensation parameter that is optimal for a dibit pattern, then curve 1 is generated. Apparently, this strategy does not work well because it may *overcompensate* many transitions. If the first, second, and third transitions in a group of adjacent transitions are precompensated with individually optimized parameters and the fourth and further transitions are assigned a precompensation parameter equal to that of the third transition, the SAM plot is improved significantly (curve 3). In contrast, for a pattern consisting of isolated transitions and dibits, i.e., without three or more transitions adjacent to one another, precompensating the second transition in each dibit produces the best SAM plot (curve 2) among the three cases.

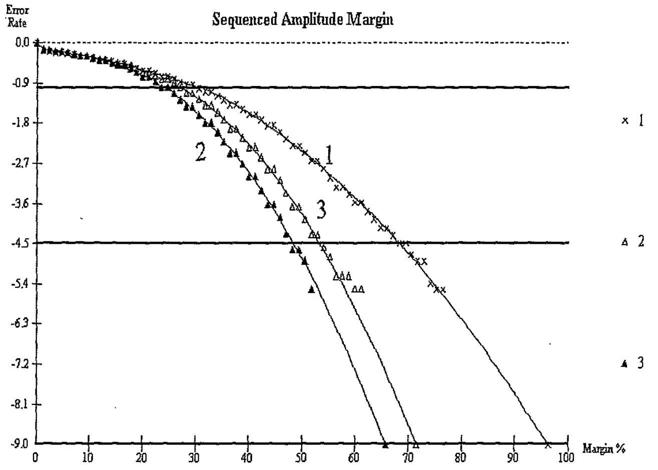


FIGURE 12.49. SAM plots with different precompensation strategies: (1) random pattern, same precompensation for all the transitions; (2) isolated dibit pattern, same precompensation; (3) random pattern, precompensation for the first, second, and third transitions individually.

The adjacent transition interactions are very important when considering NLTS and write precompensation.

As discussed in Chapter 10, the degree of interactions between transitions in a random pattern is determined by the head/medium combination. Another important nonlinear distortion is the interaction between old information and new information, which is called the hard transition shift (HTS) and will also affect the sample values like NLTS does. Figure 12.50 demonstrates an extreme case in which the sample histogram is strongly dependent on the HTS. The HTS value for this particular head and medium combination is about 12% of the bit period, while the NLTS value is close to 25% of the bit period. When the medium is effectively AC-erased and NLTS is precompensated, the channel samples are concentrated at their ideal values. However, when the medium is DC-erased and rewritten, the resulting interactions between NLTS and HTS (Chapter 10) cause strong distortions of the distributions, as shown in Fig. 12.50 (right).

The related SAM plots are shown in Fig. 12.51. In the case where NLTS is not precompensated, the operation margin is the smallest (curve 1). When the pattern is written with optimal precompensation on an AC-erased medium, the operation margin improves greatly (curve 2). If the same is done on a DC-erased medium, however, the SAM plot shifts to the right (curve 3). To compensate the influence of the HTS on the data pattern, the polarity of each transition is taken into account and the value of the HTS is subtracted from the precompensation parameter of each hard transition. The procedure shifts the SAM plot to the left and improves the operation margin (curve 4). Unfortunately, the HTS shift cannot be

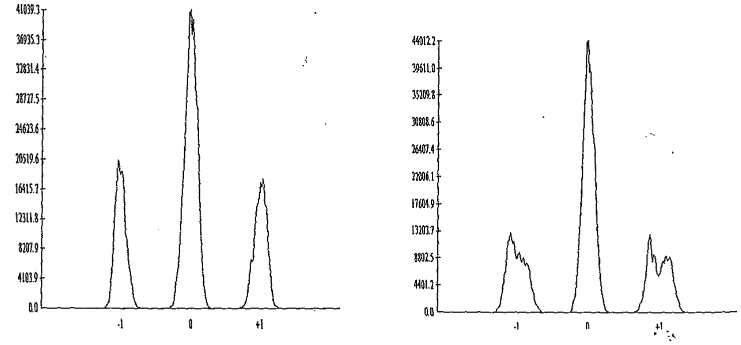
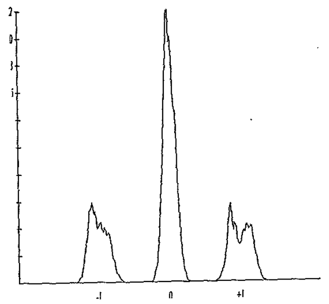


FIGURE 12.50. Sample histograms with NLTS precompensation for AC-erased medium (left) and for positive DC-erased medium (right).

very important when considering

ree of interactions between transi-
ɔy the head/medium combination.
n is the interaction between old
h is called the hard transition shift
lues like NLTS does. Figure 12.50
the sample histogram is strongly
for this particular head and me-
bit period, while the NLTS value
ː medium is effectively AC-erased
nel samples are concentrated at
ıedium is DC-erased and rewrit-
ILTS and HTS (Chapter 10) cause
s shown in Fig. 12.50 (right).
in Fig. 12.51. In the case where
ion margin is the smallest (curve
mal precompensation on an AC-
nproves greatly (curve 2). If the
however, the SAM plot shifts to
ıfluence of the HTS on the data
taken into account and the value
ɔmpensation parameter of each
AM plot to the left and improves
nately, the HTS shift cannot be



S precompensation for AC-erased
lium (right).

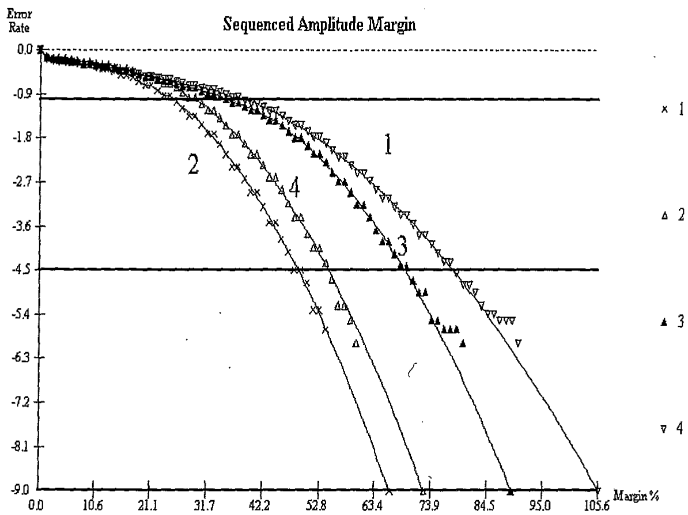


FIGURE 12.51. SAM plots for a PR4 system: (1) without NLTS precompensation; (2) with AC-erased medium and NLTS precompensation; (3) with DC-erased medium and NLTS precompensation; (4) with DC-erased medium and precompensation of both NLTS and HTS.

precompensated because the old information pattern is not known in a real detection channel.

As mentioned previously, equalization affects the error rate of the PRML channel by modifying the system SNR and introducing specific shape distortions if equalization is not ideal. System SNR considerations are very important. To match the frequency response of the channel to the PRML frequency response, a boost in certain frequencies is required of equalization. Theoretical calculations for Gaussian noise and the optimal density range predict that the SNR loss can be 2–3 dB for PR4, and 1–1.5 dB for EPR4 and E^2PR4 systems. When actual measurements are done in real recording systems, spectral distributions of medium and electronic noises are mixed and the results are usually less pessimistic for PR4 and more pessimistic for extended PRML schemes. The results may be further altered by the pulse shape, which can be asymmetrical and very different from the Lorenzian pulse. Therefore, the actual degradation of SNR caused by equalization is almost impossible to predict, and should be measured experimentally.

Shape distortions introduced by an equalizer can be destructive or beneficial. Some of the shape distortions are more destructive than the others. For example, if the isolated PR4 pulses samples are {. . . , 0, 1.1, 0.9, 0, . . . }, the dipulse samples are {. . . , 0, 1.1, −0.2, −0.9, 0, . . . }; if the isolated pulse sample are {. . . , 0, 0.9, 1, 0.1, 0, . . . }, then the dipulse samples are {. . . , 0, 0.9, 0.1, −0.9, −0.1, 0, . . . }. The first dipulse will cause less PRML errors because its first sample is pushed up to 1.1. In contrast, the second dipulse will be more subject to errors because its first sample is pushed down to 0.9 and the third sample is −0.9. Like random noises and nonlinear distortions, pulse shape distortions due to misequalization will be more prone to errors if they reduce the dipulse amplitude.

A real recording channel is somewhat nonlinear, so the optimal equalization is actually pattern dependent. Consider the SAM plots shown in Fig. 12.52. At first, an isolated pulse is captured and the equalizer coefficients are calculated so as to reshape it to the targeted isolated pulse. Next, the equalizer thus obtained is applied to a data pattern consisting of isolated transitions. Finally a SAM plot (curve 1) is generated. Obvi-

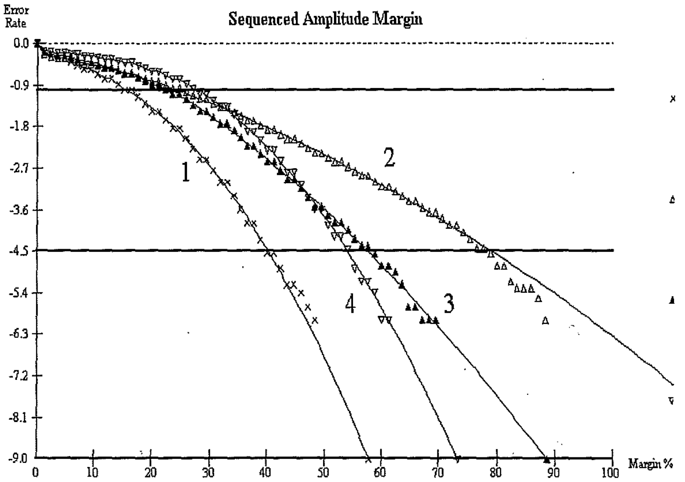
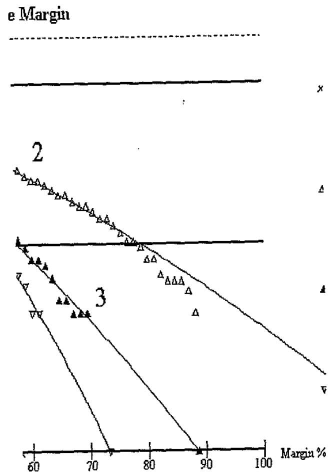


FIGURE 12.52. SAM plots for PR4ML channel: (1) optimal equalization for isolated transitions, pattern consisting of isolated transitions; (2) optimal equalization for isolated transitions, random pattern with NLTS precompensation; (3) equalization adjusted for random pattern, random pattern with NLTS precompensation; (4) equalization adjusted for random pattern, pattern consisting of isolated transitions.

ı equalizer can be destructive or
ns are more destructive than the
4 pulses samples are {. . . , 0, 1.1,
, 0, 1.1, −0.2, −0.9, 0, . . . }; if the
1, 0.1, 0, . . . }, then the dipulse
ı, . . . }. The first dipulse will cause
e is pushed up to 1.1. In contrast,
to errors because its first sample
nple is −0.9. Like random noises
istortions due to misequalization
uce the dipulse amplitude.
ıt nonlinear, so the optimal equal-
onsider the SAM plots shown in
aptured and the equalizer coeffi-
t to the targeted isolated pulse.
ılied to a data pattern consisting
lot (curve 1) is generated. Obvi-



el: (1) optimal equalization for iso-
transitions; (2) optimal equalization
LTS precompensation; (3) equaliza-
ırn with NLTS precompensation; (4)
rn consisting of isolated transitions.

ously, the equalizer works well because the operation margin is large. However, if the same equalization is applied to a random pattern with the optimal write precompensation, a much worse SAM plot (curve 2) is generated. If the equalizer is adjusted for the random pattern with NLTS precompensation so as to optimize the system error rate, an improved SAM plot (curve 3) is obtained, but this equalizer is not optimal for the pattern consisting of isolated transitions (curve 4). In short, different data patterns have different optimal equalizations, which must be carefully considered in the design and characterization of PRML channels.

References

1. H. Kobayashi and D. T. Tang, "Application of partial response channel coding to magnetic recording system," *IBM J. Res. Develop.*, July 1970, p. 368; R. W. Wood and D. A. Petersen, "Viterbi detection of class IV partial response on a magnetic recording channel," *IEEE Trans. Commun.*, **34**, 454, 1986.
2. E. R. Kretzmer, "Generalization of a technique for binary data communication," *IEEE Trans. Commun. Technol.*, **14**, 67, 1966; G. D. Forney, Jr., "The Viterbi algorithm," *Proc. IEEE*, **61**(3), 268, 1973.
3. B. Sklar, *Digital Communications: Fundamentals and Applications*. (Englewood Cliffs, NJ: Prentice Hall, 1988).
4. P. H. Siegel and J. K. Wolf, "Modulation coding for information storage," *IEEE Communications Magazine*, December 1991, (12), 68.
5. R. D. Cidecyan, F. Dolvio, R. Hermann, W. Hirt, and W. Schott, "A PRML system for digital magnetic recording," *IEEE Journal on Selected Areas in Communications*, **10**(1), 38, 1992.
6. T. Howell, *et al.*, "Error rate performance of experimental gigabit per square inch recording components," *IEEE Trans. Magn.*, **26**(5), 2298, 1990.
7. J. D. Cocker, R. L. Galbriath, G. J. Kerwin, J. W. Rae, and P. Ziperovich, "Implementation of PRML in a rigid disk drive," *IEEE Trans. Magn.*, **27**(6), 4538, 1991.
8. H. Thapar and A. Patel, "A class of partial response systems for increasing storage density in magnetic recording," *IEEE Trans. Magn.*, **23**(5), 3666, 1987.
9. J. Hong, R. Wood, and D. Chan, "An experimental 180 Mb/sec PRML channel for magnetic recording," *IEEE Trans. Magn.*, **27**(6), 4532, 1991.
10. A. Taratorin, "Method and apparatus for measuring error rate of magnetic recording devices having a partial response maximum likelihood data detection channel," U.S. Patent 5,355,261, 1996.
11. T. Perkins "A window margin like procedure for evaluating PRML channel performance," *IEEE Trans. Magn.*, **31**(2), 1109, 1995.
12. A. Kogan *et al.*, "Histograms of processed noise samples for measuring error rate of a PRML data detection channel," U.S. Patent 5,490,091, 1997.

13. A. Taratorin, "Margin evaluation of PRML channels: non-linear distortions, misequalization and off-track noise performance," *IEEE Trans, Magn.*, **31**(6), 3064, 1995.

14. P. Ziperovich, "Performance degradation of PRML channel due to nonlinear distortions," *IEEE Trans. Magn.*, **27**, 4825, 1991.