

Chapter 9: Partial Response Maximum Likelihood (PRML)

9.1 Introduction to PRML Detection System

Recently, researchers from IBM laboratories reported the results of an experiment demonstrating that an areal density of 1 gigabit per square inch could be achieved for the storage and reliable retrieval of digital data on a hard disk system. This many-fold increase in density was achieved using a number of advanced techniques. One of these techniques was a different approach to combating intersymbol interference, sometimes referred to as PRML, using partial-response (PR) signaling with maximum-likelihood (ML) sequence detection.

Instead of keeping the transitions far apart using (d,k) codes, PRML allows the transitions to be close together, and the read signal, with its resulting intersymbol interference, is equalized to a frequency response known as a class4 partial-response channel. The equalized signal is then detected by a maximum likelihood sequence estimator, i.e., a Viterbi detector. In this chapter, we will give a brief summary of the PR and ML components of this system as they apply to magnetic recording.

When recording density is low, each transition written on the magnetic medium results in a relatively isolated peak of voltage and peak detection method is used to recover written information. However, when $PW50$ becomes comparable with the channel bit period, the peak detection channel can not provide reliable data detection. In chapter 8 we have discussed the peak detection systems. Superposition of pulses (linear ISI) shifts peaks of read-back signal and increased probability of errors in zero-crossing detector (refer to chapter 8). At the same time, signal amplitude is lowered and errors in the threshold detector part also increase. Whatever tricks are made with peak detection systems, they barely work at $PW50/T$ ratios above one (1).

This means that a different detection principle is needed if the density of recording is to be increased. This new detection method should not be based on voltage peaks, rather it should take into account the fact that signals from adjacent transitions interfere. In other words, the method of detection should be aware of linear ISI in the signal.

PRML is the most popular detection scheme in modern disk drives. PRML is an acronym for Partial Response Maximum Likelihood. This method was originally proposed in the early 70's by a group of IBM researchers. PRML consists of two relatively independent parts: Partial Response (PR) and Maximum Likelihood (ML) detector. In the mean time we can temporarily think of Maximum Likelihood as a magic digital "black box" which improves the error rate of the system compared to usual threshold detectors.

9.2 Partial Response (PR) System

Several partial response channels have been described by various researchers, but one in particular has shown to be suitable for use in high density magnetic recording is the PR Class IV (PR4) channel that is characterised by a D-transfer function of $(1-D^2)$. Actually, you can treat it as delay function similar to z-transform.

$$y(n) = x(n) - x(n-2)$$

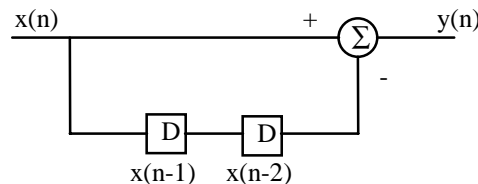


Figure 0-1: Block diagram of a PR4 channel

PRML is based on two assumptions:

- The shape of read-back signal from an isolated transition is exactly known and determined.
- The superposition of signals from adjacent transitions is linear.

A block diagram of a typical PRML system is shown in Figure 0-2 Block Diagram of typical PRML system. It is relatively simple and we will now explain how it works.

Analog signal, coming from the magnetic head should have a certain and constant level of amplification. This is done in a variable gain amplifier (VGA). To keep a signal level, VGA gets a control signal from a Clock and Gain recovery system.

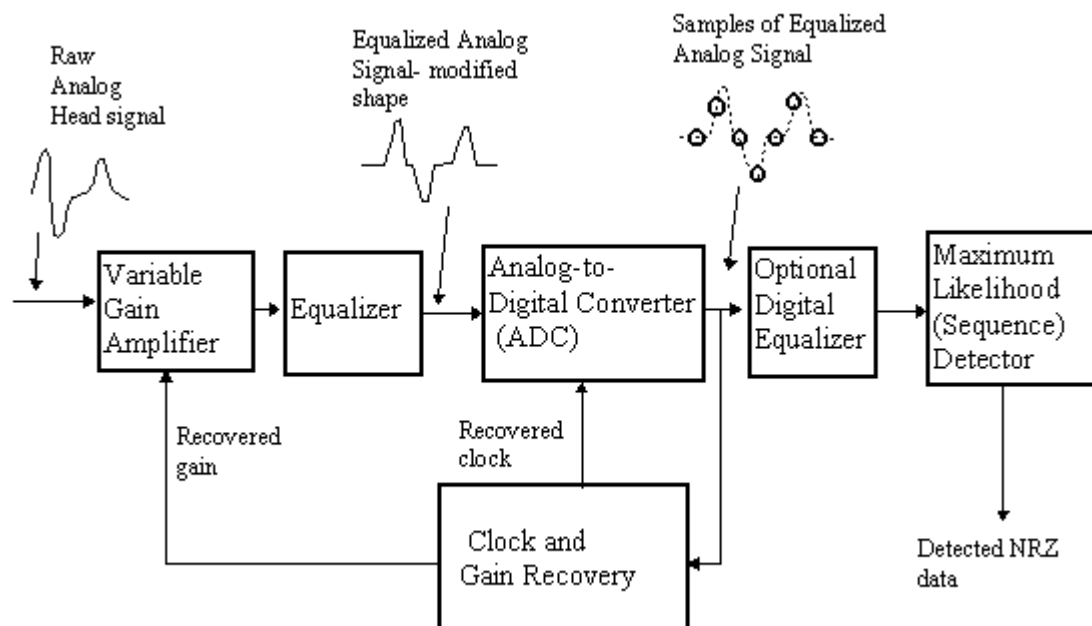


Figure 0-2 Block Diagram of typical PRML system

The shape of the read-back signal, coming from the head usually has to be modified. In the mean time we can think about this shape modification as adjustment of the pulse width to make it exactly proportional to the distance between transitions. Usually it means that the pulse, resulting from isolated transition, should have relatively flat tails. This shape modification is done by an equalizer. An equalizer is a linear programmable filter having specific frequency response. Analog signal on the equalizer output has a slightly different shape than the unmodified signal coming directly from the magnetic head.

The signal from the equalizer output is sampled by an Analog-to-Digital Converter (ADC). The sampling is initiated by clock signal exactly one time per channel bit period. Frequency and phase of the clock signal is adjusted by a clock recovery system. Signal on the ADC output is a stream of digital samples.

Digital samples are sometimes filtered by an additional digital filter. This second filtering operation can improve the quality of analog equalization.

The samples on the ADC output are used to actually detect the presence of transitions in the read-back signal. If signal quality is good, a simple threshold detector can be used to distinguish between zero signal and transition by comparing sample values to a threshold. However, much better detection quality can be provided by a Maximum Likelihood detector.

Note that no assumption that the signal should contain isolated and relatively narrow peaks was used. We now distinguish between zero signal and transition by looking at the ADC samples of the signal, and these samples are not necessarily taken at the signal peak level.

The most unclear signal transformation in Figure 0-2 is equalization. What does it mean that the pulse of voltage should have a pre-determined shape? To answer this question let us consider Class IV partial response, or PR4 system.

The isolated pulse shape in a PR4 system is shown in Figure 0-3.

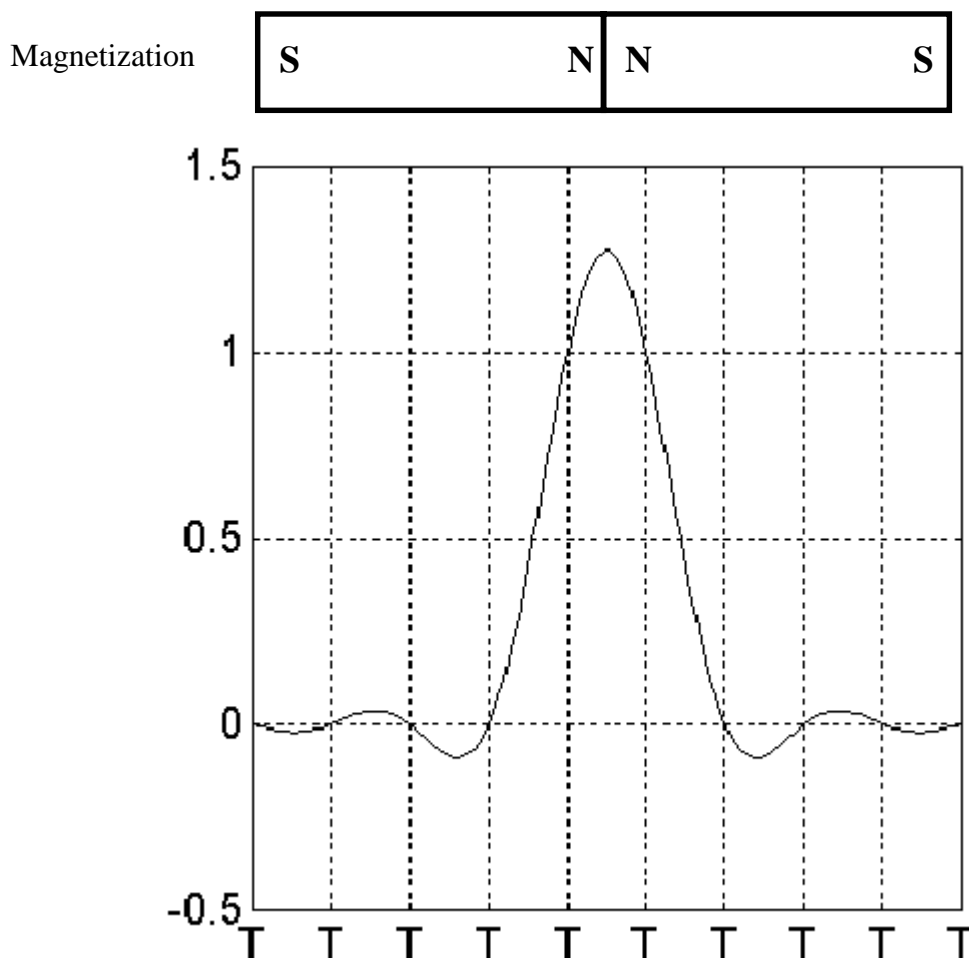


Figure 0-3 Shape of Isolated Pulse in PR4 system

The transition is written at time instant $t=0$, where T is the channel bit period. The shape is somewhat strange, it is oscillating and the pulse values at integer number of bit periods before the transition are exactly zeroes. However, at $t=0$ and at $t=T$, i.e. one bit period later, the 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. That also mean that there is a delay of $T/2$ for the write current in writing the bit on the magnetic media.

Assume that an isolated transition is written on the medium and the pulse of voltage shown in Figure 0-3 Shape of Isolated Pulse in PR4 system comes to the PRML system. The PR4 system requires that the samples of this pulse taken by ADC should correspond to the bit periods. Therefore, samples of the isolated PR4 pulse on the output of ADC will be:

00...011000...

Of course value "1" is used for convenience and in reality it corresponds to some ADC level.

The fact that the isolated transition has two non-zero samples: one at the transition location and one at the next transition location is very important: we assume that if the next transition is written, **the pulses will interfere**.

What happens if we write another transition after the first one? Obviously, we will have superposition between both pulses, usually called a dipulse response as shown in Figure 0-4.

Magnetization

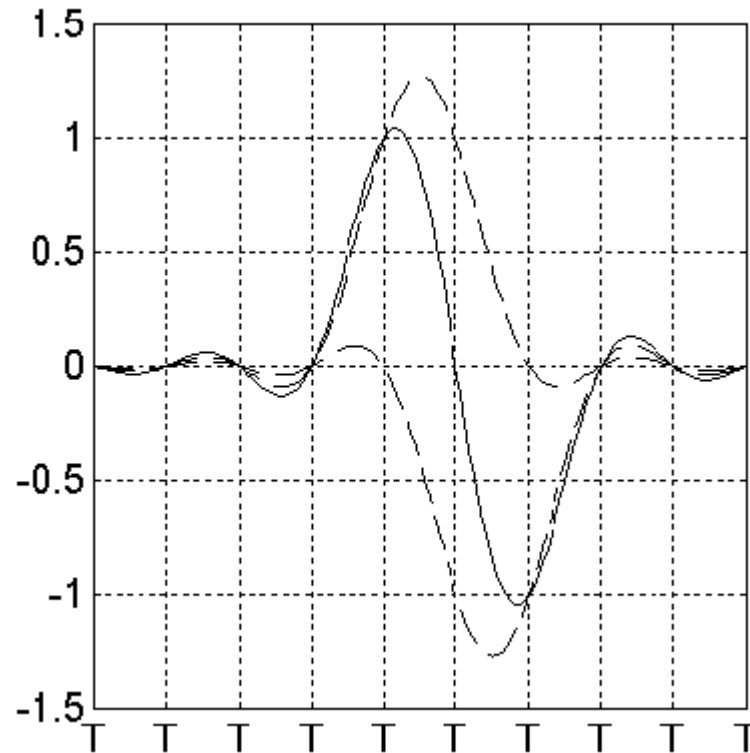


Figure 0-4: A dipulse response

Note that at sample points we have $\{ \dots, 0, 0, 1, 0, -1, 0, 0, \dots \}$.

$$\begin{array}{r}
 0 \ 0 \ 0 \ 1 \ 1 \ 0 \ 0 \ 0 \text{ - from the first transition} \\
 + \quad 0 \ 0 \ 0 \ 0 \ -1 \ -1 \ 0 \ 0 \text{ - from the second transition} \\
 \hline
 = \quad 0 \ 0 \ 0 \ 1 \ 0 \ -1 \ 0 \ 0
 \end{array}$$

The PR Class IV (PR4) channel that is characterised by a D-transfer function of $(1-D^2)$ when we change the polarity of the magnetisation at next sampling interval T . Actually, you can treat it as delay function similar to z-transform, where $D = z^{-1}$.

$$Y(D) = X(D)(1 - D^2) \quad (\text{Equation 1})$$

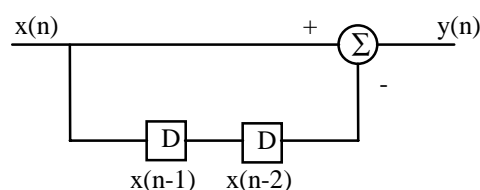


Figure 0-5: Block diagram of a PR4 channel

What happens if we have three consecutive transitions (tribit)? It is easy to check that the answer is $\{\dots, 0, 0, 1, 0, 0, 1, 0, 0, \dots\}$. Another useful pattern is "low frequency" where transitions are separated by one clock. Obviously, the sequence of resulting samples is $\{\dots, +1, +1, -1, -1, +1, +1, -1, -1, +1, +1, \dots\}$.

$$\begin{array}{rcl}
 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & - \text{ from the first transition} \\
 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & - \text{ from the second transition} \\
 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & - \text{ from the third transition} \\
 + & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 - \text{ from the fourth transition} \\
 \hline
 = & 0 & 0 & 0 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 0 & 0
 \end{array}$$

Now it is clear, that having the sequence of samples from the ADC output, we can easily reconstruct any pattern which was written on the medium. The current value of the pattern $P(k)$ in NRZ form (i.e. 1 and 0 means different medium magnetisation), is reconstructed from the current sample $y(k)$ simply as:

$$x(k) = y(k) + x(k-2) \quad (\text{Equation 2})$$

Let us consider how $x(k) = y(k) + x(k-2)$ (Equation 2 applies to an isolated transition).

NRZ magnetization of medium is:	0 0 0 0 1 1 1 1 1 1 1...
Samples of PRML system:	0 0 0 0 1 1 0 0 0 0 0 ...
Recovered NRZ data will be:	0 0 0 0 1 1 1 1 1 1 1 ...

Now let us consider a tribit - three consecutive transitions.

NRZ magnetization of medium is:	0 0 0 0 1 0 1 1 1 1 1...
Samples of PRML system:	0 0 0 0 1 0 0 1 0 0 0 ...
Recovered NRZ data will be:	0 0 0 0 1 0 1 1 1 1 1 ...

This last example is especially interesting: the second pulse in the tribit has zero samples, it is almost completely suppressed by the first and the third transitions due to linear superposition! However we can easily recover the data based on the samples.

This is the main principle of PRML: we are no longer afraid of linear ISI. Once the pulses can be reduced to some "standard" shape, the data pattern is easily recovered because superposition of signals from adjacent transitions is known. In our example, we know that sample "1" is suppressed by "-1" if the next transition is written etc.

It is easy to check that all possible linear superpositions of the samples result in only three possible values: $\{-1, 0, +1\}$. A positive pulse of voltage is always followed by a negative pulse and vice versa. Therefore, a sample with a value of 2 can not be generated by any linear superposition and ADC output will consist of only three distinct levels: $\{-1, 0, +1\}$.

For a PR4 system we may look at the output of ADC on Figure 0-2 and analyze a sequence of samples. If all parts of the PRML system are working properly (i.e. the equalization, gain and timing recovery are correct and the signal is noise-free), the ADC samples should take only nominal values (such as $\{-1,0,+1\}$ for PR4 system). Noise, wrong equalization, jitter etc. distort these sample values.

A simple way to characterize the sample quality is to build their histogram (also known as sample values distribution). A histogram is a function which corresponds to the number of samples having a particular value. If a sufficiently large number of samples is taken, we will be able to see the histograms shown in Figure 0-6. The left picture is obtained for a PR4 system with good quality. Three distinct peaks corresponding to levels $\{-1,0,+1\}$ are clearly seen. The right picture has poor quality: distributions of samples overlap, i.e. in some cases zero samples look like $+1$ or -1 samples and vice versa.

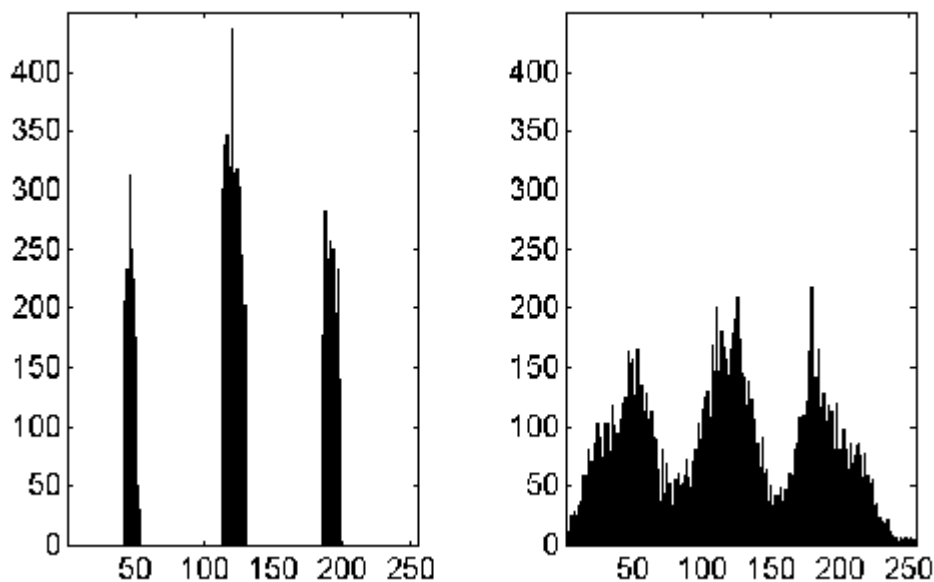


Figure 0-6: Sample Value Distributions for good (left) and poor (right) PRML signal quality.

If histograms of samples overlap similar to those shown, the ML detector will greatly improve the situation. The typical gain of an ML detector over a simple comparator is about three or more orders of magnitude in error rate. While it is possible that the Maximum Likelihood detector will still be able to decode the pattern in a case of strongly overlapping samples, the probability of errors is increased and, obviously, non-overlapping histograms are always better.

9.3 Maximum Likelihood Detector

To explain why ML detection is used, we must first look at the possible misclassification of a simple threshold detector.

As we remember, samples on the output of ADC ideally have a small number of levels, such as $\{-1,0,+1\}$ for the PR4 system. Evidently, a threshold detector could be set to classify a current sample value comparing it to an amplitude threshold, e.g. if $\text{sample} > 0.5$ then $\text{sample} = 1$, if $\text{sample} < -0.5$ then $\text{sample} = -1$, if $|\text{sample}| \leq 0.5$ then $\text{sample} = 0$. Imagine that we are looking at a stream of noisy samples :

0.8 0.3 -0.7 -0.2 0.6 0.9 1.1 0.2

For this sequence of samples, the threshold detector output would be:

1 0 -1 0 1 1 1 0

If we look at these values, we notice a strange sequence of three "ones" in the row. This sequence of three "ones" can not exist! Indeed, "11" always means an isolated transition. The next transition should be of opposite polarity, therefore the only possible combinations are "011", "110", "-111", etc.

The Threshold detector knows nothing about the previous and subsequent samples and compares each sample with the threshold. An ML detector is smarter. It "knows" that "111" is a forbidden sequence of samples and tries to decide which was the most probable data pattern which caused this sequence of samples.

It is easy to propose several close allowable sequences:

{1 0 -1 0 1 1 0 0} or

{1 0 -1 0 0 1 1 0} or

{1 0 -1 0 0 0 1 1}.

Which one is the most probable? Let us look at these sequences again:

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

y(n)	0.8	0.3	-0.7	-0.2	0.6	0.9	1.1	0.2	sum of sq	
y1(n)	1	0	-1	0	1	1	0	0		
x(n)	1	0	0	0	1	1	1	1		
	0.04	0.09	0.09	0.04	0.16	0.01	1.21	0.04	1.68	1.296
y2(n)	1	0	-1	0	0	1	1	0		
x(n)	1	0	0	0	0	1	1	1		
	0.04	0.09	0.09	0.04	0.36	0.01	0.01	0.04	0.68	0.825
y3(n)	1	0	-1	0	0	0	1	1		
x(n)	1	0	0	0	0	0	1	1		
	0.04	0.09	0.09	0.04	0.36	0.81	0.01	0.64	2.08	1.442

y4(n)	1	0	-1	0	1	1	-1	-1		
x(n)	1	0	0	0	1	1	0	0		
	0.04	0.09	0.09	0.04	0.16	0.01	4.41	1.44	6.28	2.506
y5(n)	1	0	-1	0	0	1	0	0		
x(n)	1	0	0	0	0	1	0	1		
	0.04	0.09	0.09	0.04	0.36	0.01	1.21	0.04	1.88	1.371

Sequence #1 assumes that 0.6 and 0.9 values are "1" and 1.1 value is "0" which does not look very probable. Sequence #2 assumes that 0.6 is in fact "0" while 0.9 and 1.1 are "1". Sequence #3 assumes that 0.6 and 0.9 are "0" and 1.1 and 0.2 are "1". Obviously, Sequence #2 is the most reasonable assumption. We can verify it by calculating the Sum-of-Squared Error (SSE) between samples $s(k)$ and assumed sequence $b(k)$:

$$SSE = \sqrt{\sum_{k=1}^N (s(k) - b(k))^2} \quad (\text{Equation 3})$$

Sequence #1: SSE = 1.2961

Sequence #2: SSE = 0.8246

Sequence #3: SSE = 1.4422

As seen, Sequence #2 is the closest to the original samples sequence and at the same time satisfies our constraints. In other words, this sequence is the most likely among other candidate sequences.

This simple example demonstrates the principles of ML detection:

Decisions are made based on a sequence of samples, instead of one current sample. For each sequence of samples a list of allowable sequences is generated. Each of the allowable sequences is compared with the received sequence and SSE (or any other distance function) is calculated.

The sequence having the minimum distance (maximum likelihood) is selected to be the result of the detection.

Decisions of the ML detector are always done with some delay.

Of course, practical realisation of the described ML detector is difficult. We do not know how to choose the sequence length. If we start from the first sample and analyse all possible sequences, after each additional sample, the number of possible combinations which should be analysed grows exponentially. Another problem is that detection should be done in real time together with the rate of incoming samples and delays should be small enough to avoid overflow. The solution to these problems was found by A.Viterbi in 1967. Fascinating features of the Viterbi algorithm are that it is equivalent to full-blown Maximum Likelihood detection and that it can be realized in real time using moderately complex hardware.

What should be stressed here is that the Viterbi algorithm is only (although smart and ingenious) a practical realisation of the Maximum Likelihood detection principle which opened the way for effective real-time data detection. For a while we can think about the ML detector as a "black box" accepting ADC samples and providing decoded data on its output.

9.4 Ideal PR4 signal

Sampling theorem states that any analog band-limited signal having frequencies only below a highest frequency f_{max} can be uniquely recovered from its discrete samples taken with sampling interval $T/2f_{max}$. In other words, if a band-limited signal is sampled with a sampling period of at least $1/2f_{max}$ or smaller, the information content of the signal is not lost.

Now we also can construct the origin of the isolated pulse shape shown in Figure 0-3 Shape of Isolated Pulse in PR4 system. Since the PRML system has a bandwidth of $1/2T$, the only band-limited function which is represented by samples "...0110..." is the response of the ideal low-pass filter on this sample sequence. Each "1" sample on the channel output will result in $\sin(x)/x$ type function and, therefore, the ideal isolated pulse for PR4 system is given by:

$$s(t) = \frac{\sin\left(\frac{\pi t}{T}\right)}{\left(\frac{\pi t}{T}\right)} + \frac{\sin\left(\frac{\pi(t-T)}{T}\right)}{\left(\frac{\pi(t-T)}{T}\right)} \quad (\text{Equation 4})$$

Note that if the bandwidth of the channel is larger than $1/2T$ we can find different shapes of isolated pulses satisfying our criteria (as an extreme example of almost infinite bandwidth this shape could even be a rectangular pulse given by the same samples {...0,0,1,1,0,0,...}).

Let us now calculate the frequency response of the PR4 channel. As is well known, the frequency response is the Fourier transform of the channel impulse reaction. Our isolated pulse is called "step response" of the PR4 channel, or reaction to a "step" of magnetisation. Using NRZ notation for channel input, this step is described by a sequence

000001111111... A derivative of the step reaction will be given by ...00010000. In other words, it represents localised change of magnetisation or a dipulse. Therefore, we have to samples of {...10-1....} as before.

9.5 Partial Response Polynomials. EPR4 and E²PR4 Systems

Different partial response schemes are often described using polynomials. These polynomials historically came from discrete signal processing terminology.

Let us describe an input data pattern in NRZ terms, i.e. "0" stands for one particular direction of medium magnetisation and "1" for another. The data pattern is given by the sequence of bits a_k . When a magnetic head reads the signal, it responds only to changes of magnetisation, i.e. it differentiates the signal.

The differentiating function of the head can be described if we introduce a "delay operator" D , given by $D a_k = a_{k-1}$. Obviously, the head acts on the NRZ pattern as $(1-D)a_k = a_k - a_{k-1}$. The operator given by $(1-D)$ will result in +1 or -1 samples, depending on the direction of the

magnetisation change. Operator $(1-D)$ is the simplest polynomial, corresponding to generating positive or negative samples.

In a PR4 system each pulse of voltage has 2 samples. In other words, if a transition of magnetisation occurs, it results in a sample equal to "1" at the transition location and another sample at the next sample period. This is equivalent to "spreading" or delaying and adding the current sample, given by operator $(1+D)$. Indeed, if $a_k = 1$, the operator $(1+D) a_k$ will result in sequence "1,1". Therefore, a PR4 system in polynomial terms is described as $(1-D)(1+D) = 1-D^2$, where operator D^2 is the result of a product DD and delays the current sample two bit periods.

If we discard the differentiating part of the polynomial given by $(1-D)$ and look only at $(1+D)$ we will get the samples of the isolated pulse: {1,1}. In other words, the term $(1+D)$ determines how the transition samples are "spread" over the neighbouring bit periods. Partial Response 4 is a particular case of more general family of PR polynomials, given by the general equation:

$$(1-D)(1+D)^n \quad (\text{Equation 5})$$

PR4 corresponds to $n=1$. If we set $n=2$ the samples of the isolated pulse will be given by the term $(1+D)^2 = 1+2D+D^2$ which corresponds to (1,2,1). This type of PRML is called "Extended Partial Response 4" or EPR4. When $n=3$, the polynomial is given by $(1+D)^3 = 1+3D+3D^2+D^3$ and the isolated pulse has samples {1,3,3,1}. This type of partial response is called E²PR4.

The following Table 0-1 summarizes different PRML schemes used in magnetic recording.

Table 0-1: different PRML schemes used in magnetic recording

{PRIV ATE}Name	Polynomial	Isolated Pulse Samples
PR4	$(1-D)(1+D)$	0 1 1 0...
EPR4	$(1-D)(1+D)^2$	0 1 2 1 0...
E ² PR4	$(1-D)(1+D)^3$	0 1 3 3 1 0 ...

As we see from this table, isolated pulses become wider: EPR4 pulse extends over 3 bit periods and E²PR4 pulse over 4 bit periods. This means that the transition in an E²PR4 system will be "felt" by the next 3 transitions in the pattern.

Of course, the value of samples "2" or "3" are imaginary. Further in this text we will always normalise the isolated pulse amplitude to "1". Using this notation, an EPR4 pulse has sample values {...,0,1/2,1,1/2,0,...} and is shown in Figure 0-7: Isolated pulse for EPR4 system.

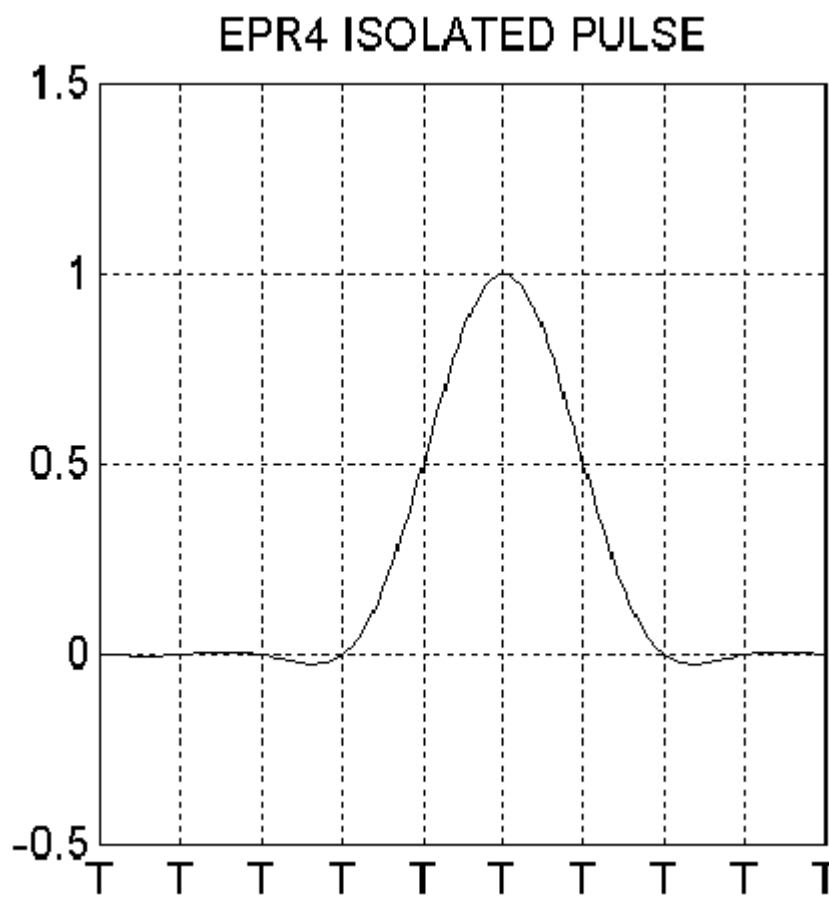


Figure 0-7: Isolated pulse for EPR4 system

Again, this analogue pulse shape is obtained assuming that the channel bandwidth is limited by $1/2T$. In this case each sample generates $\text{sinc}(x) = \sin(x)/x$ function where $x=t/T$ and the isolated pulse shape is obtained by delaying and adding the corresponding sinc functions with the weights $(1/2, 1, 1/2)$. The superposition of pulses for EPR4 results in 5 PRML sampling levels: $\{-1, -1/2, 0, 1/2, 1\}$.

The EPR4 system is used both with (8/9) encoding and with (1,7) code (using NRZI modulation). A dipulse response for EPR4 system is given by:

d=0 constraint:

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

with d=1 constraint EPR4 dipulse :

$$\begin{array}{cccccc}
 0 & \frac{1}{2} & 1 & \frac{1}{2} & 0 & \\
 + & 0 & 0 & -\frac{1}{2} & -1 & -\frac{1}{2} & 0
 \end{array}$$

$$= 0 \quad \frac{1}{2} \quad 1 \quad 0 \quad -1 \quad -\frac{1}{2} \quad 0$$

E2PR4 isolated pulse samples are: $\{\dots, 0, 1/3, 1, 1, 1/3, 0, \dots\}$. The shape of this pulse is shown in Figure 0-8: Isolated pulse shape for E2PR4 system.

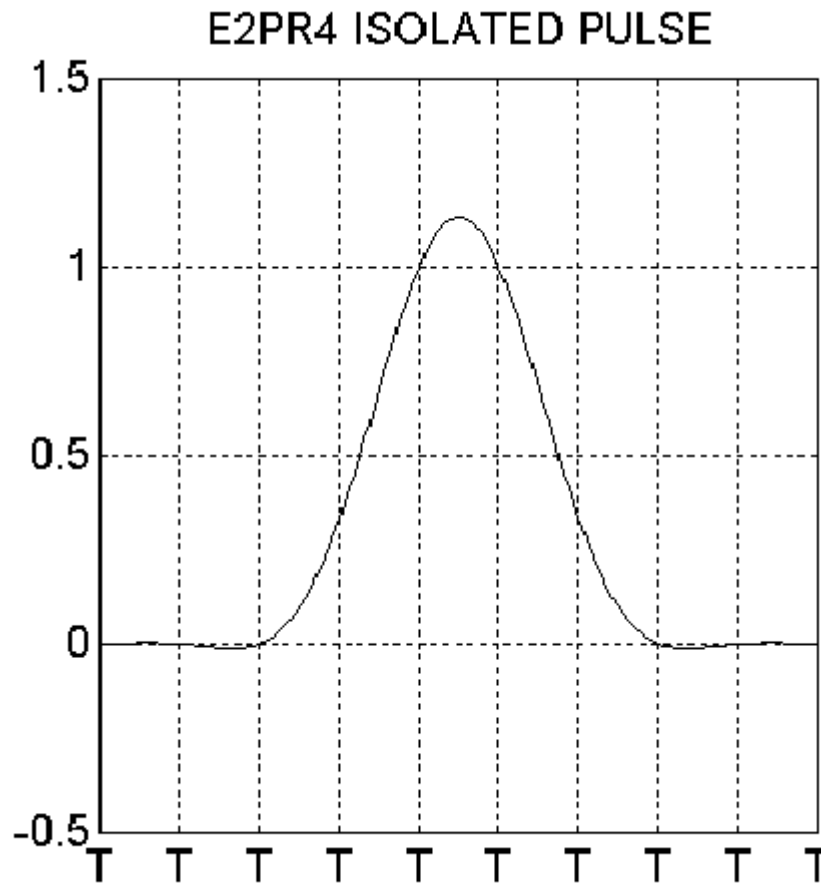


Figure 0-8: Isolated pulse shape for E2PR4 system

There are 7 possible sample levels in an E2PR4 system: $\{-1, -2/3, -1/3, 0, 1/3, 2/3, 1\}$.

An E2PR4 system is usually used with (1,7) encoding (using NRZI modulation). A dipulse for an E2PR4 system and (1,7) code is given by:

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

What are the advantages of using higher order PRML systems?

One advantage is obvious: we have more samples per clock period, therefore we can increase the recording density. In other words, if the pulse width is kept the same, the relative channel bit period can be increased.

Another, less obvious advantage, is that the shapes of the pulses shown in Figure 0-7: Isolated pulse for EPR4 system and Figure 0-8: Isolated pulse shape for E2PR4 system are more "natural" and smooth compared to the PR4 isolated pulse. EPR4 and E2PR4 pulses have less oscillations and their bottom part is more extended. The peak of spectral density lies at $1/4T$ frequency for PR4 pulse (the center of the bandwidth range) and shifts to lower frequencies for EPR4 and E2PR4 pulses. These lower frequency spectrum distributions are closer to a typical frequency content of raw non-equalised pulses. Therefore, equalisation for extended PRML can become less critical and requires a less high frequency boost which may improve signal to noise ratio.

Finally, the last advantage of high order PRML schemes is that since we have some extra density gains compared to the PR4 (remember, our pulse now extends over 3 - 4 samples), we can well afford to lose it and to return to the (1,7) code (i.e. write zero between each pair of transitions). While this hardly seems to be a rational solution, it can in fact have advantages when strong non-linear distortions are present. Nonlinear distortions strongly depend on the degree of pulse overlapping. If EPR or E2PR4 pulses are separated by two clock intervals, their overlapping parts are on the "tails" of the pulses, while superposition of PR4 pulses cover about $2/3$ of the pulse width. Therefore, resulting non-linearities for EPR4 and E2PR4 are weaker. At the same time, an E2PR4 scheme with (1,7) code can achieve the same density as PR4 with (0,4/4).

The main disadvantage of extended PRML schemes is the increased complexity of circuits and decoders. While the PR4 system works with 3 levels of samples, the E2PR4 system works with 7 levels. It requires a higher resolution of ADC, a complicated timing and gain recovery circuit and a sophisticated ML detector. In fact, the complexity of ML detectors grows exponentially with the number of levels. Realization of the fast Viterbi detector for an extended PRML is extremely complicated and sub optimal Sequence Detectors are typically used. Another problem with a higher order PRML schemes is that they have higher sensitivity to noise compared to PR4.

9.6 References

- John C.Mallinson "*The Foundations of Magnetic Recording*" - Academic Press, 1993. A good introduction into principles of magnetic recording, contains short description of PRML principles.
- B. Sklar "*Digital Communications. Fundamentals and Applications*" - Prentice Hall, 1988. A useful textbook, contains description of partial response, convolutional codes and Viterbi decoding algorithm.
- P.H.Siegel and J.K.Wolf "Modulation and Coding for Information Storage" - *IEEE Communications magazine*, December 1991, pp. 68-86. An informative review paper describing the principles of Peak Detection, PRML and coding.
- R.D.Cidecyan, F.Dolvio, R. Hermann, W.Hirt, W. Schott "A PRML System for Digital Magnetic Recording" - *IEEE Journal on Selected Areas in Communications*, - 1992, vol.10, No.1, pp.38 -56. The only available relatively detailed review of the PRML system, including principles, Viterbi algorithm, Encoding, Gain and Clock Recovery.
- Howell, D.P. McCown et al. "Error Rate Performance of Experimental Gigabit per Square Inch Recording Components" - *IEEE Transactions on Magnetics*, 1990, vol . 26, 5, pp. 2298-2302. A short description of the IBM prototype PRML channel, includes useful discussion of equalization and non-linear distortions.
- J.D. Cocker, R.L. Galbraith, G.J. Kerwin, J.W. Rae and P.Ziperovich "Implementation of PRML in a Rigid Disk Drive" - *IEEE Trans. Magnetics*, vol. 2,6, 1991, pp. 4538-4543.
- A short description of first IBM PRML drive J.Hong, R. Wood and D. Chan "An experimental 180 Mb/sec PRML Channel for Magnetic Recording" - *IEEE Trans. Magnetics*, vol. 27, 6 pp.4532-4537, 1991.
- A very useful paper for understanding PRML, equalization and timing recovery H.Thapar and A.Patel "A Class of Partial Response Systems for Increasing Storage Density in Magnetic Recording" - *IEEE Trans. Magnetics*, vol. 23, 5, pp.3666-3668 1987 - Discussion and comparison of high order PRML systems.