1 Introduction

The goal of an active sensing system, such as radar or sonar, is to determine useful properties of the targets or of the propagation medium by transmitting certain waveforms toward an area of interest and analyzing the received signals. For example, a land-based surveillance radar sends electromagnetic waves in the direction of the sky, where objects such as airplanes can reflect a (usually very tiny) fraction of the transmitted signal back to the radar. By measuring the round-trip time delay, the distance (called the range) between the radar and the target can be estimated since the speed of propagation for radio waves is known (3×10^8 m/s). Additional target properties can be obtained by performing further processing at the receiver side; e.g., the speed of a target can be estimated by measuring the Doppler frequency shift of the received signal.

In 1904, a German engineer named Christian Hülsmeyer carried out the first radar experiment using his "telemobiloscope" to detect ships in dense fog by means of radio waves. As to sonar, Reginald Fessenden, a Canadian engineer, demonstrated this in 1914 using a sound echo device, though not successfully, for iceberg detection off the east coast of Canada. It was amongst several other experiments and patents said to be prompted by the 1912 Titanic disaster.

Radar and sonar underwent considerable development during the two world wars and later on spread into diverse fields including weather monitoring, flight control and underwater sensing. There are two factors that are critical to the system performance, namely the *receive filter* and the *transmit waveform*. The receive filter is used to extract from the received signals the information of interest, e.g., target locations in radar or sonar applications [Skolnik 2008] or channel conditions in communications [Proakis 2001]. The transmit waveform, not surprisingly, interplays with the receive filter. A good design of the transmit waveform lends itself to accurate parameter estimation and a reduced computational burden at the receiver.

Arguably the most commonly used receive filter is the matched filter, which maximizes the signal-to-noise ratio (SNR) in the presence of stochastic additive white noise [Turin 1960]. Examples of other well-known receive filters include the mismatched filter, the use of which is also called the instrumental-variable (IV) method [Ackroyd & Ghani 1973][Zoraster 1980][Stoica, Li & Xue 2008], the Capon estimator [Capon 1969], the amplitude and phase estimation (APES) algorithm [Li & Stoica 1996][Stoica et al. 1998][Stoica et al. 1999] and more advanced data-adaptive techniques such as the iterative adaptive approach (IAA) [Yardibi et al. 2010].

We concentrate our attention on transmit waveform design in this book. In particular, we are interested in synthesizing waveforms that have good correlation properties. In radar range compression, low auto-correlation sidelobes improve the detection performance of weak targets [Stimson 1998][Levanon & Mozeson 2004]; in code-division multiple access (CDMA) systems, low auto-correlation sidelobes are desired for synchronization purposes and low cross-correlations reduce interferences from other users [Suehiro 1994][Tse & Viswanath 2005]; and the situation is similar in many other active sensing applications such as ultrasonic imaging [Diaz et al. 1999]. An emitted probing waveform with low auto-correlation sidelobes maximizes the signal-to-noise ratio, when complemented by a matched filter at the receiver side, while significantly weakening the signals from adjacent range bins.

In addition to correlation properties, the problem of transmit beampattern synthesis is also considered in this book. A classical phased array steers narrow beams toward different angles by adjusting only the waveform phases across the antenna elements. In a modern MIMO (multi-input multi-output) system, however, waveforms can be chosen freely and this waveform diversity allows for more flexibility in beampattern synthesis. One example is hyperthermia treatment for breast cancer [Guo & Li 2008], where waveform diversity enables an ultrasonic focal point to be matched to the entire tumor region without impacting surrounding healthy tissues.

1.1 Signal model

Let $s(t)$ denote the transmitted signal, with t indicating time. Suppose that $s(t)$ consists of N symbols

$$s(t) = \sum_{n=1}^{N} x(n) p_n(t) \tag{1.1}$$

where $p_n(t)$ is the shaping pulse and $\{x(n)\}_{n=1}^{N}$ are the N symbols. The shaping pulse $p_n(t)$ (with duration t_p) can be an ideal rectangular shaping pulse

$$p_n(t) = \frac{1}{\sqrt{t_p}} \operatorname{rect}\left(\frac{t-(n-1)t_p}{t_p}\right), \quad n = 1, \ldots, N, \tag{1.2}$$

where

$$\operatorname{rect}(t) = \begin{cases} 1, & 0 \le t \le 1, \\ 0, & \text{elsewhere}, \end{cases} \tag{1.3}$$

or is equal to other pulse such as the raised-cosine pulse [Proakis 2001].

Note that the actual transmitted waveform is composed of the in-phase and quadrature components of $s(t)e^{j2\pi f_c t}$ where f_c is the carrier frequency. It is assumed that the signal demodulation has already been performed at the receiver side and thus the carrier term $e^{j2\pi f_c t}$ can be safely ignored in the analysis.

In practice, hardware components such as analog-to-digital converters and power amplifiers have a maximum signal amplitude clip. In order to maximize the transmitted power available in the system, it is desirable that the transmit sequences are unimodular or have low peak-to-average power ratios (PARs). In our design, we impose the following unit-modulus constraint whenever feasible:

$$x(n) = e^{j\phi(n)}, \quad n = 1, \ldots, N \tag{1.4}$$

where $\{\phi(n)\}$ are phases. Note that (1.1) combined with (1.4) provides a phase-coded signal representation. There are many other types of signal that are widely used or have been discussed in the literature; these include the well-known chirp waveform (see Section 1.3), discrete frequency-coded waveforms [Costas 1984][Deng 2004] and waveforms constructed from a particular set of functions such as the prolate spheroidal functions [Moore & Cada 2004] or the Hermite wave functions [Gladkova & Chebanov 2004]. In this book we have chosen to focus specifically on the phase-coded signal model, which serves as a practical and effective framework for designing waveforms with various desirable properties.

The waveform $s(t)$ is transmitted in the direction of a scene of interest and is reflected by various targets at different range locations. The reflected signals, which are time-shifted and weighted versions of $s(t)$, arrive linearly combined at the receiver side:

$$y(t) = \sum_k \alpha_k s(t - \tau_k) + e(t), \tag{1.5}$$

where τ_k is the round-trip time delay for the kth target, α_k is the coefficient related to the target reflection, for example the radar cross section (RCS), and $e(t)$ is the noise.

Suppose that we aim to estimate the coefficient $\alpha_{k'}$ by applying the filter $w(t)$ at the receiver. The estimated coefficient is given by

$$\hat{\alpha}_{k'} = \int_{-\infty}^{\infty} w^*(t) y(t) dt. \tag{1.6}$$

More precisely, according to a conventional convolution definition (1.6) is the receiver output at time instant 0 when $y(t)$ is the input and $w(-t)$ is the filter. However, we can simply refer to $w(t)$ as the filter in the receiver processing indicated by (1.6) without introducing any ambiguity in later discussions.

To determine an appropriate $w(t)$, we decompose $y(t)$ into three parts:

$$y(t) = \underbrace{\alpha_{k'} s(t - \tau_{k'})}_{\text{signal}} + \underbrace{\sum_{k \neq k'} \alpha_k s(t - \tau_k)}_{\text{clutter}} + \underbrace{e(t).}_{\text{noise}} \tag{1.7}$$

If there is no clutter and $e(t)$ is the zero-mean white noise then the matched filter $w(t) = s(t - \tau_{k'})$ will give the largest signal-to-noise ratio (SNR). A direct proof goes as follows:

$$\mathrm{SNR} \triangleq \frac{\left|\int_{-\infty}^{\infty} w^*(t)\alpha_{k'} s(t-\tau_{k'})dt\right|^2}{\mathrm{E}\left\{\left|\int_{-\infty}^{\infty} w^*(t)e(t)dt\right|^2\right\}} \tag{1.8}$$

$$= \frac{|\alpha_{k'}|^2 \left|\int_{-\infty}^{\infty} w^*(t)s(t-\tau_{k'})dt\right|^2}{\sigma_e^2 \int_{-\infty}^{\infty} |w(t)|^2 dt} \tag{1.9}$$

$$\leq \frac{|\alpha_{k'}|^2 \int_{-\infty}^{\infty} |s(t-\tau_{k'})|^2 dt}{\sigma_e^2} = \frac{|\alpha_{k'}|^2 \sigma_s^2}{\sigma_e^2} \tag{1.10}$$

where E denotes the expectation operator, and σ_e^2 and σ_s^2 are the noise power and signal power, respectively. Note that (1.9) is due to the white noise assumption $\mathrm{E}\{e(t_1)e^*(t_2)\} = \sigma_e^2\delta_{t_1-t_2}$ and that (1.10) results from the Cauchy–Schwartz inequality $\left|\int w^*(t)s(t-\tau_{k'})dt\right|^2 \leq \int |w(t)|^2 dt \int |s(t-\tau_{k'})|^2 dt$. The maximum value of the SNR in (1.10) is achieved if and only if the filter $w(t)$ is a scaled version of $s(t-\tau_{k'})$, which concludes the proof.

For the purpose of normalization, the matched filter $w(t)$ is chosen as $s(t-\tau_{k'})/\int |s(t)|^2$ and the corresponding estimate of $\alpha_{k'}$ in (1.6) is given by

$$\hat{\alpha}_{k'} = \frac{\int_{-\infty}^{\infty} s^*(t-\tau_{k'})y(t)dt}{\int_{-\infty}^{\infty} |s(t)|^2 dt}. \tag{1.11}$$

Besides boosting the signal component and suppressing the noise, the matched filter can also eliminate the clutter component (as easily seen from (1.7) and (1.11)) if

$$r(\tau) = \int_{-\infty}^{\infty} s(t)s^*(t-\tau)dt, \quad -\infty < \tau < \infty, \tag{1.12}$$

is zero for all $\tau \neq 0$. The function $r(\tau)$ as defined in (1.12) is called the *auto-correlation* of $s(t)$.

1.2 Design metrics

The previous section has outlined the benefit of small auto-correlation sidelobes $r(\tau)$ (for $\tau \neq 0$). For most practical cases, we need to focus only on a delay τ that is an integer multiple of the symbol length t_p. One reason is that in modern systems digital filtering is usually performed at the receiver side, that is, the integral in (1.11) is implemented as a summation of sampled signals. In addition, if a rectangular shaping pulse (see (1.2)) is used, the values of $r(\tau)$ can be obtained exactly by the linear interpolation of two neighboring auto-correlation samples [Levanon & Mozeson 2004, Chapter 6]:

$$r(\tau) = \frac{\tau - t_1}{t_p} r(t_2) + \frac{t_2 - \tau}{t_p} r(t_1), \tag{1.13}$$

where $t_1 = \lfloor \tau/t_p \rfloor$ and $t_2 = t_1 + t_p$. Such auto-correlations at integer multiple delays $\{kt_p\}_{k=-N+1}^{N-1}$ can be calculated for $k \geq 0$ as

$$\begin{aligned} r(kt_p) &= \int_{-\infty}^{\infty} s(t)s^*(t - kt_p)dt \\ &= \int_{kt_p}^{Nt_p} \sum_{n=k+1}^{N} x(n)p_n(t)x^*(n-k)p_n^*(t)dt \\ &= (N-k) \sum_{n=k+1}^{N} x(n)x^*(n-k) \int_0^{t_p} |p_n(t)|^2 dt \\ &= (N-k) \sum_{n=k+1}^{N} x(n)x^*(n-k). \end{aligned} \tag{1.14}$$

Correlations at negative delays can be obtained from $r(kt_p) = r^*(-kt_p)$. When shaping pulses other than rectangular are used, it can still be expected that $r(\tau)$ will be well controlled as long as $r(k)$ is sufficiently small.

It follows from the above discussions that the correlations of interest are given by

$$r(k) = \sum_{n=k+1}^{N} x(n)x^*(n-k) = r^*(-k), \quad k = 0, \ldots, N-1. \tag{1.15}$$

The above set $\{r(k)\}$ is called the auto-correlation of the discrete sequence $\{x(n)\}$. Note that the notation r is slightly abused in (1.12) and (1.15) in order to denote both continuous-time and discrete-time auto-correlations, yet a distinction can be made easily by examining the two different time variables.

For the set $\{r(k)\}_{k=-N+1}^{N-1}$ defined above, $r(0)$ is called the in-phase correlation and is always equal to the signal energy. All the other auto-correlations, i.e., $\{r(k),\ k = -N+1, \ldots, -1, 1, \ldots, N-1\}$, are collectively called the auto-correlation sidelobes. One of our main interests in Part I is the design of phase-coded sequences $\{x(n)\}$ whose auto-correlation sidelobes are as low as possible. Part I also discusses ambiguity function synthesis, which can be considered as a two-dimensional extension to correlation design.

More precisely speaking, the $\{r(k)\}$ defined in (1.15) constitute the *aperiodic* auto-correlation. The *periodic* auto-correlation of the sequence $\{x(n)\}$ is defined as

$$\tilde{r}(k) = \sum_{n=1}^{N} x(n)x^*((n-k) \bmod N) = \tilde{r}^*(-k) = \tilde{r}^*(N-k), \qquad k = 0, \ldots, N-1, \tag{1.16}$$

where "mod" is the modulo operator:

$$p \bmod N = \begin{cases} p - \lfloor p/N \rfloor N, & p \text{ is not an integer multiple of } N, \\ N, & \text{otherwise.} \end{cases} \tag{1.17}$$

The relationship between the aperiodic correlation (1.15) and the periodic correlation (1.16) can be easily obtained as follows:

$$\begin{aligned}\tilde{r}(k) &= \sum_{n=1}^{k} x(n)x^*(n-k+N) + \sum_{n=k+1}^{N} x(n)x^*(n-k) \\ &= \sum_{m=(N-k)+1}^{N} x(m-(N-k))x^*(m) + \sum_{n=k+1}^{N} x(n)x^*(n-k) \\ &= r^*(N-k) + r(k).\end{aligned} \tag{1.18}$$

Periodic correlations appear in many applications such as the synchronization of a CDMA (code division multiple access) system. Part II deals with the minimization of periodic correlation sidelobes.

As will be shown in Part III, waveform correlations can serve as a bridge connecting the underlying waveform to the desired beampattern of, e.g., an antenna array. In particular, the waveform diversity in a MIMO system leads to a flexible control of waveform correlations, which further leads to an agile transmit beampattern synthesis.

In the next section, we will review several well-known waveforms that have good auto-correlation properties, especially those that are phase-coded (1.4). For the sake of brevity, the word "periodic" is used explicitly whenever a periodic correlation (1.16) is referred to; otherwise, an aperiodic correlation, as defined in (1.15), is meant.

1.3 Review of existing waveforms

We start with the well-known chirp waveform. A chirp waveform is a linear frequency-modulated (LFM) pulse, whose frequency is swept linearly over a bandwidth B in a time duration T. Chirp signals have been widely used in radar applications since World War II, as they possess relatively low correlation sidelobes and are mostly tolerant to Doppler frequency shifts [Levanon & Mozeson 2004]. In addition, the power of a chirp signal is dispersed evenly throughout the frequency spectrum, which allows for high spectral efficiency.

A chirp signal $s(t)$ can be written as

$$s(t) = \frac{1}{\sqrt{T}} \exp\left(j\pi \frac{B}{T} t^2\right), \quad 0 \le t \le T, \tag{1.19}$$

where B/T is the chirp rate. Figure 1.1(a) shows the real part of $s(t)$ with parameters $T = 100$ s and $B = 1$ Hz. Figure 1.1(b) shows its auto-correlation function $r(\tau)$ (normalized by $r(0)$ and using a $20\log_{10}$ scale), where the peak sidelobe is -13.4 dB.

Cambridge University Press
978-1-107-01969-0 - Waveform Design for Active Sensing Systems: A Computational Approach
Hao He, Jian Li and Petre Stoica
Excerpt
More information

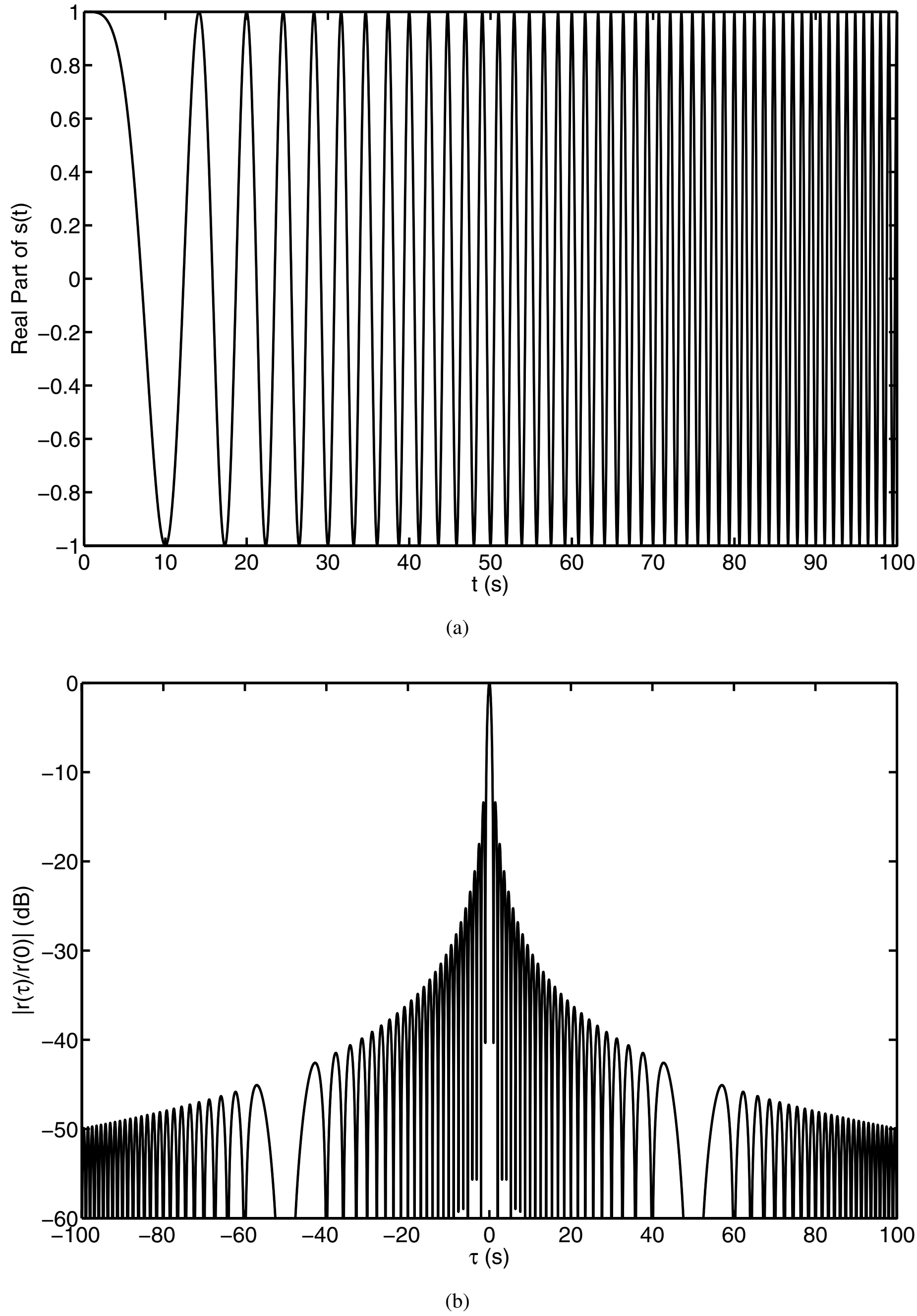


Figure 1.1 (a) The real part of $s(t)$ in (1.19) with $T = 100$ s and $B = 1$ Hz. (b) The auto-correlation function of $s(t)$.

Many phase codes can be derived from the chirp signal. We sample $s(t)$ at time intervals $t_s = n/B$ for $n = 1, \ldots, N$ ($N = BT$) and obtain the following sequence:

$$\begin{aligned} x(n) &\triangleq s(nt_s) = \exp\left[j\pi \frac{B}{T}\left(\frac{n}{B}\right)^2\right] \\ &= \exp\left[j\pi \frac{n^2}{BT}\right] = \exp\left[j\pi \frac{n^2}{N}\right], \quad n = 1, \ldots, N. \end{aligned} \tag{1.20}$$

The sequence $\{x(n)\}$ shown above has perfect *periodic* auto-correlations if N is even, meaning that all periodic auto-correlation sidelobes are zero: $\tilde{r}(k) = 0$ for $k \neq 0$.

A sequence with perfect periodic correlations for any odd N can be constructed by changing the sequence phases in (1.20) to

$$x(n) = \exp\left[j\pi \frac{n(n-1)}{N}\right], \quad n = 1, \ldots, N, \tag{1.21}$$

which is the Golomb sequence [Zhang & Golomb 1993]. The Chu sequence [Chu 1972], interestingly, is a combination of the above two sequences:

$$x(n) = \begin{cases} \exp\left[jQ\pi \frac{n^2}{N}\right], & N \text{ even}, \\ \exp\left[jQ\pi \frac{n(n-1)}{N}\right], & N \text{ odd}, \end{cases} \tag{1.22}$$

where Q is any integer that is prime to N. As expected, the Chu sequence has perfect periodic correlations for any (positive) integer N.

Besides the Golomb and the Chu sequences, there are many other phase-coded sequences whose phases are quadratic functions of n, such as the well-known Frank sequence and the P4 sequence. The Frank sequence is defined for $N = L^2$ as

$$x((m-1)L + p) = \exp\left[j2\pi \frac{(m-1)(p-1)}{L}\right], \quad m, p = 1, \ldots, L. \tag{1.23}$$

The P4 sequence is defined for any length N as

$$x(n) = \exp\left[j\frac{2\pi}{N}(n-1)\left(\frac{n-1-N}{2}\right)\right], \quad n = 1, \ldots, N. \tag{1.24}$$

Both the Frank and P4 sequences have perfect periodic correlations.

Figure 1.2 shows the auto-correlation $r(k)$ of the P4 sequence of length $N = 100$. Note that from the sequence $\{x(n)\}$ we can construct a continuous-time waveform $s(t)$ using (1.1). We will choose rectangular pulse shaping and set $t_p = 1$ s so that the signal duration is 100 s and the signal bandwidth is roughly 1 Hz (i.e., $1/t_p$); these are the same parameters as those used in Figure 1.1. The real and imaginary parts of the so-constructed $s(t)$ are shown separately in Figure 1.3(a). The auto-correlation $r(\tau)$ of this $s(t)$ is shown in Figure 1.3(b). The peak sidelobe is −26.3 dB, which is much lower than that of the chirp waveform in Figure 1.1(b).

The auto-correlation properties of the Golomb, Chu or Frank sequences are similar to those of the P4 sequence and are omitted here.

Another widely used sequence is the maximum length sequence (m-sequence) [Proakis 2001], which is one of the most commonly known PN (pseudo-noise)

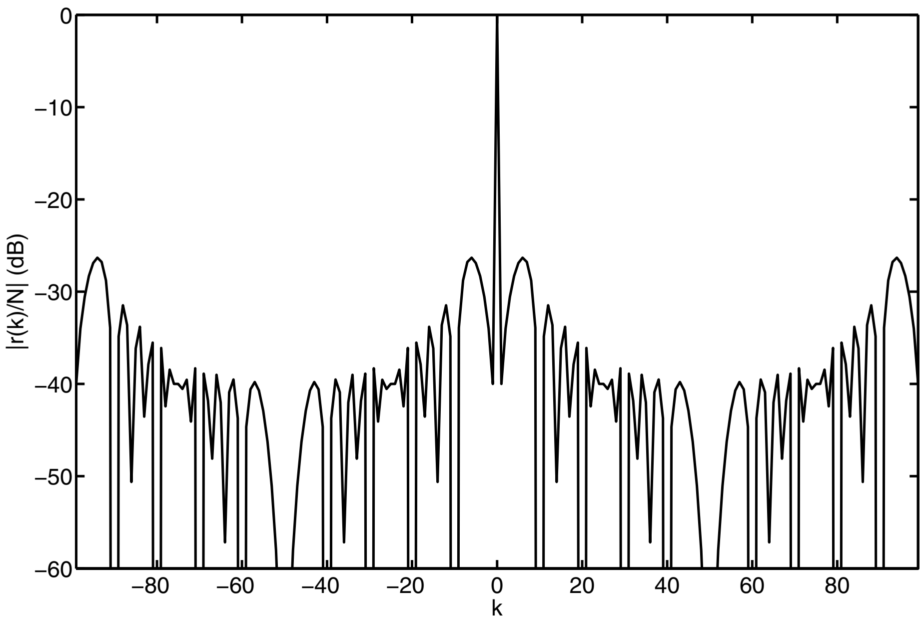
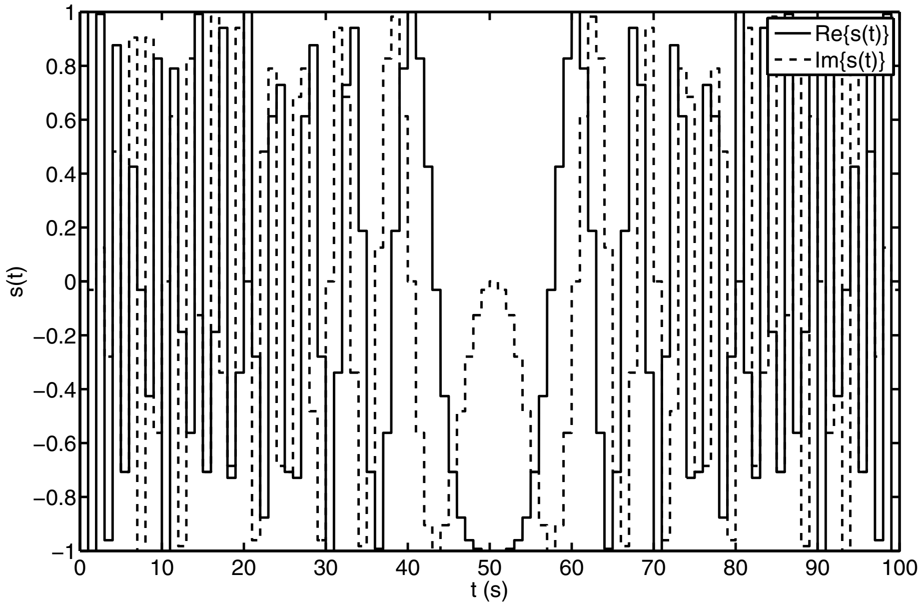


Figure 1.2 The auto-correlation function $r(k)$ of the P4 sequence, as defined in (1.24), of length $N = 100$.

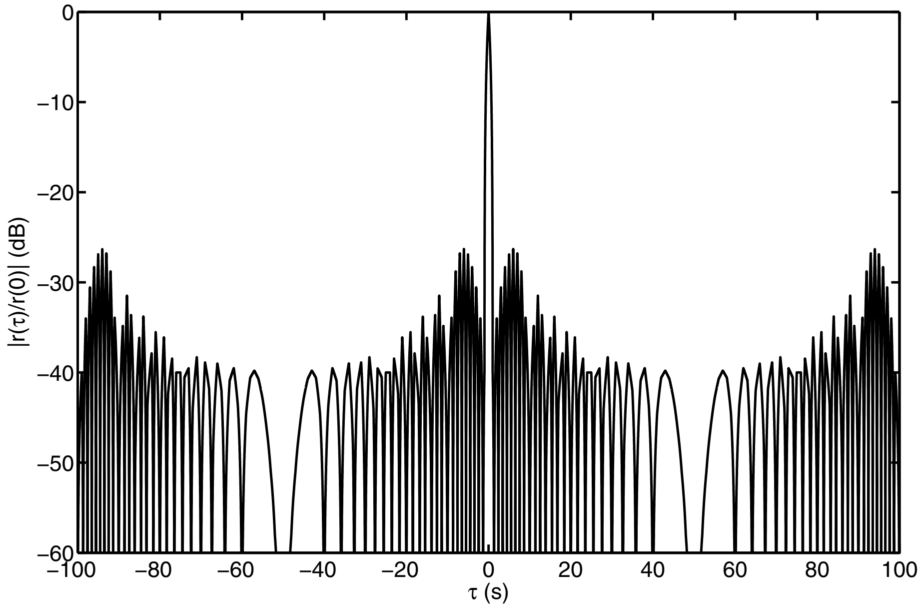
sequences. An m-sequence is a type of pseudo-random binary sequence that is generated by a maximal linear feedback shift register (LFSR). Figure 1.4 shows a length-3 LFSR where the plus operator indicates "exclusive or". Each register block can store 0 or 1, so three blocks amount to eight different states. When fed with any initial binary sequence (not all zeros), such a shift register will cycle through all eight states except for the all-zero state. For instance, starting from 001, the register in Figure 1.4 will pass repeatedly through the following seven states 001, 100, 010, 101, 110, 111, 011. By taking only the output from the third block and replacing 0 with -1, we obtain a length-7 m-sequence, $\{1, -1, -1, 1, -1, 1, 1\}$. Its aperiodic as well as periodic correlations are shown in Figure 1.5.

One of the prominent features of an m-sequence is that its *periodic* correlation sidelobes are always equal to -1, as can be observed from Figure 1.5(b). Its aperiodic correlation sidelobes, though, do not have a regular pattern and can be relatively high. Also note that the LFSR can be efficiently implemented in hardware, which greatly facilitates the use of m-sequences in practice.

Figure 1.6 compares the aperiodic auto-correlation of an m-sequence of length 127 and that of a Golomb sequence (see (1.21)) of the same length. The polyphase Golomb sequence exhibits notably lower correlation sidelobes than the binary m-sequence. Although a polyphase sequence does not necessarily have lower correlation sidelobes than a binary sequence, allowing more phase values leads to more degrees of design freedom. As will be shown in Chapter 2, by allowing arbitrary phases between 0 and 2π



(a)



(b)

Figure 1.3 (a) The real and imaginary parts of $s(t)$ in (1.1) when a P4 sequence of length 100 and rectangular shaping pulses are used. (b) The auto-correlation function $r(\tau)$ of this $s(t)$.