Digital Alias-free Signal Processing

Ivars Bilinskis
Institute of Electronics and Computer Science, Riga Latvia

John Wiley & Sons, Ltd

Digital Alias-free Signal Processing

Digital Alias-free Signal Processing

Ivars Bilinskis
Institute of Electronics and Computer Science, Riga Latvia

John Wiley & Sons, Ltd

Telephone (+44) 1243 779777

Email (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on www.wiley.com

Other Wiley Editorial Offices

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 42 McDougall Street, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 6045 Freemont Blvd, Mississauga, ONT, Canada L5R 4J3

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Anniversary Logo Design: Richard J. Pacifico

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN 978-0-470-02738-7 (HB)

Typeset in 10.5/13pt Times Roman by TechBooks, New Delhi, India.
Printed and bound in Great Britain by TJ International Ltd, Padstow, Cornwall.
This book is printed on acid-free paper responsibly manufactured from sustainable forestry in which at least two trees are planted for each one used for paper production.

Contents

Part 2 Processing

Preface

The rationale of replacing signal analog processing by digital, worth doing whenever it is feasible, is well known. The benefits that can be obtained are usually significant. Digital signal processing (DSP) technology is widely used in the application range characterized by relatively low frequencies and other acceptable application conditions. However, 'going digital' becomes more difficult and even problematic as application demands grow. Some nontraditional methods, techniques and algorithms are suggested in this book for widening the digital domain over the application area where analog signal processing techniques still prevail.

Much attention is paid to resolving the problems caused by overlapping of periodically sampled signal frequencies. This effect, known as aliasing, restricts application of the conventional digital signal processing methods and techniques to the frequency range where the achievable sampling rate could at least twice exceed the higher frequency present in the spectrum of the signal to be digitized and processed digitally. Attempts to eliminate the harmful impact of aliasing have led to the development of digital technology for signal processing, specifically the technology called digital alias-free signal processing, or DASP. This strengthens the competitiveness of digital techniques considerably. Successful use of special digitizing techniques for the elimination of aliasing has been an object lesson showing the significance of digitizing in the whole process of signal digital processing.

Focusing on digitizing and matching of signal processing to the specifics of signal sampling and quantizing operations has led to other significant improvements in digital signal processing methods and techniques. In fact, many signal processing problems can be resolved in this way. These considerations and the experience gained in this area are described in the book.

While this is the first full-scale book discussing DASP, there is a prehistory to it. The previous book *Randomized Signal Processing* by I. Bilinskis and

A. Mikelsons was published by Prentice Hall in 1992. With the much appreciated permission of the publishers, some material from that book has been used and is included here. Specifically, this concerns Chapters 1 to 6 and Chapters 12 to 16, where the basics of the randomized and pseudo-randomized sampling and quantizing operations, as well as parameter estimation of signals digitized in this way, are discussed. The used material represents the classical fundamentals of randomized signal processing. More advanced methods, techniques and algorithms have been developed on this basis over the last fifteen years since the referenced book was published.

An attempt has been made to reflect the earlier prehistory in Chapter 1. This chapter is an introduction, so the content of the book, the discussed problems and the innovative digital alias-free signal processing technology are outlined. It is shown that elimination of the harmful aliasing effect is crucial for achieving progress in this field. The reasons why it is essential to pay attention to signal digitizing and to match digitizing to the specifics of digital signal processing are explained.

In Chapter 2 it is shown that to achieve flexibility of the digitization techniques, essential for matching them to the given signal processing task conditions, signal sampling and quantization operations should somehow be made variable. It is therefore suggested that randomization should be used as a tool for making signal conversion operations, including sampling and quantizing, more flexible and more adaptable to specific signal processing needs. The application specifics and potential of this instrument are discussed.

Randomization of sampling is considered in Chapter 3 as a means for achieving fully digital signal processing in a much wider frequency range. To randomize signal sampling effectively, it must be known how variations of periodic sampling conditions, including variations of periodic sampling phase, affect the characteristics of the obtained sampled signals. Some issues of sampling, essential both for sampling and processing of sampled signals, are considered in this chapter.

The basics of randomized quantizing are discussed in Chapter 4. While the signal instantaneous values are always quantized by comparing them with threshold levels, they are kept in fixed positions in deterministic quantizing and are randomly varied in randomized quantizing. Although realization of the second approach is usually more complicated than the first, it is often preferable to perform quantizing of signals in this more complicated way as the properties of the quantized signals are different and various desirable benefits could be achieved. The essentials, advantages and drawbacks of this quantization approach are discussed in this chapter.

Some of the effects caused by the deliberate injection of randomness into the quantization procedure are positive while others are harmful, increasing statistical errors. Therefore the randomization level should be controlled to suppress the additional errors. A useful approach to obtaining the desirable effects is considered in Chapter 5. It is based on substitution of randomization by deterministic pseudo-randomization. Although such quantizing is usually described in probabilistic and statistical terms, it is evidently a fully deterministic process.

There are various techniques for randomization of sampling. Discussions in Chapter 6 are focused on the direct approach to such randomization. According to this approach, signal sample values are taken at time instants that are fully and directly determined by the used randomized sampling point processes. A number of sampling point processes are discussed. It is emphasized that the statistics of the signal sample taking timing process, in the case of direct sampling randomization, are signal independent.

In some cases randomizing of sampling occurs unintentionally as a side effect. Indirect randomization of sampling, discussed in Chapter 7, takes place when signal sample values are obtained at the time instants when the signals cross some thresholds or specially generated reference waveforms. This approach to signal digitization might be preferential, for example, in the cases of data acquisition from multiple signal sources on a large scale. The signal sample values are then obtained at random time instants and the randomization of the sampling process is not planned or controlled. The application rationale, advantages and limitations of these sampling techniques are considered.

The derivatives of periodic sampling are considered in Chapter 8. They are essential for composing hybrid periodic/nonuniform sampling models. Especially useful are the discussed periodic sampling point sequences with random skips. Whenever this type of sampling procedure is used, shifting of the sampling phase plays an important role. Essential relationships between the phase shifting of sampling and the sampled signal reconstruction conditions are revealed. It is suggested how to consider the estimation of Fourier coefficients as a process rather than as a calculation of a parameter.

The aliasing processes observed when the sampling operation is performed nonuniformly are studied in detail in Chapter 9 to show that randomizing of sampling only suppresses aliasing and some aliasing-related effects remain. These effects are diffused and are not as well defined as aliasing in cases of periodic sampling. The term ‘fuzzy’ fits well to the considered kind of aliasing as there are numerous multiple-frequency contributions to it. It is shown that both the sampling and signal processing stages have to be arranged in the correct way to achieve successful elimination of the aliases.

Hybrids of periodic and pseudo-randomized nonuniform sampling are discussed in Chapter 10. An attempt is made to synthesize sampling methods that have the integrated positive features of the uniform and nonuniform sampling methods and at the same time would not be impaired by their drawbacks. Available options are analysed. The concept of hybrid sampling suggests how high-performance pseudo-randomized nonuniform sampling should be realized on the basis of a few phase-shifted stabilized periodic processes. This approach is well suited also for building reduced jitter sampling drivers.

Modifications of data acquisition systems, considered Chapter 11, are basically related to the following two applications: (a) data acquisition providing alias-free processing of wideband signals and (b) data acquisition from a large quantity of signal sources. Structures of the considered data acquisition systems, the used digitizing techniques and the system performance are discussed. Deviations from the classical data acquisition schemes are substantial. The remote sampling approach is suggested and described.

Data acquisition does not necessarily mean gathering of raw data. Signal conversions performed at the stage of signal pre-processing, specifically their average parameter estimation procedures carried out close to the signal sources, are described in Chapter 12. The emphasis is on the optimization of signal quantizing and digital processing of the quantized signal sample values. An approach to optimization of the basic signal average parameter estimation is discussed.

It is shown in Chapter 13 that signal digitizing strongly impacts on the conditions for the correlation analysis and that applying the pseudo-randomized digitising techniques, if skilfully done, leads to a number of desirable effects. Specifics of processing pseudo-randomly quantized signals, especially the issues of quantized signal multiplication, are studied and put in focus. As numerous multiplications of the quantized signal sample values have to be executed to perform signal correlation analysis, it pays to rationalize them. Development of special hardware for the correlation analysis is discussed, including application of pseudo-randomized sampling for signal correlation analysis at small delay time increments.

The problem of matching signal processing to sampling is discussed in Chapter 14. Signal orthogonal and unorthogonal transforms from time to frequency domains are considered in this context. There are many typical signal processing applications where the signals have to be transformed on the basis of unorthogonal transforms. Performing of the unorthogonal transforms is considered.

As soon as the sampling procedure is randomized, the DFT of the respective signals become strongly sampling-dependent. It is shown in Chapter 15 that direct calculations of the DFT do not complete the process of signal decomposition into their components. They simply lead to acquiring intermediate signal processing

results containing valuable information. Therefore the outcome of the DFT under these conditions should not always be automatically regarded as spectrograms of the respective signals.

Nonuniformly sampled signal processing cannot be based on application of the popular fast DSP algorithms. Other approaches to the algorithm complexity reduction have to be found and exploited. Some useful techniques for that are suggested and discussed in Chapter 16. They are based on the exploitation of rectangular function sets. These functions assume only the values –1, 0 and +1 and, consequently, the application of them leads to the much simpler calculations of the Fourier coefficients as numerous multiplication operations are replaced by simple logic operations.

Anti-aliasing signal processing technology is particularly applicable in the area of spatial signal processing. Discussions in Chapter 17 are focused on the issue of the complexity reduction of the large-aperture antenna arrays. The potential of pseudo-randomization of the large-aperture antenna arrays is considered for reducing the number of sensors in the array. It is suggested that the positive experience obtained from signal processing be adjusted in the time domain to the specifics of spatial signal processing, as nonuniform sampling in the time domain is equivalent to the irregular spacing of sensors in the array.

It is shown in Chapter 18 that introduction of irregularities into a sampling process leads to the nonorthogonality of the nonuniformly sampled discrete basis functions. Apparently the pattern of the nonuniform sampling point sequence defines this nonorthogonality and the errors related to it. At purposeful intentional pseudo-randomization of sampling this pattern is *a priori* given. Therefore it should be possible to use this information to suppress the errors caused by the nonuniformities. An approach to this problem is considered. This type of adapted signal processing could be applied for processing both the temporal and spatial signals.

An application area well suited to the specifics of digital alias-free signal processing is considered in Chapter 19. This is the area where signal processing is carried out to identify objects and to evaluate their parameters by analysing the signals reflecting the reaction of these objects to some excitation signals. The parameters and characteristics of the excitation signals are usually known. Access to this valuable information makes it easier to keep algorithms relatively simple and to achieve high performance. The issues of test signal synthesis, frequency response evaluation for the objects, measuring short time intervals and demodulation of bioimpedance signals are discussed.

Considerable expertise is required to be successful at applications of the described digital alias-free signal processing technology. Substantial investments

in terms of effort, time and money are needed to gain knowledge and experience sufficient for achieving really significant positive results. A more rational approach to widening the application area for DASP is suggested in Chapter 20. It is based on the idea of developing and exploiting specific embedded systems. This approach makes it possible to encapsulate everything related to the DASP technology within these embedded systems. Then linking them to other systems could be realized on the basis of the standard techniques. Widening the digital domain in the direction of higher frequencies, simplification of data compression, complexity reduction of sensor arrays and achieving fault tolerance are mentioned among other gains immediately achievable in this way.

It might be said that this book summarizes to some extent the R&D work done in the Digital Signal Processing Laboratory of the Institute of Electronics and Computer Sciences in Riga, Latvia, over the last fifteen years. These were years after the independence of Latvia had been regained and it was a difficult time for the Institute. It took a great deal of effort just to survive, to keep the work going on, and the Institute would not have survived if the European Commission had not given us the opportunity to participate in the Copernicus programme. We consider ourselves lucky that our proposal for the project, 'Complexity-Reduced Large Aperture Arrays', against strong competition, gained the European Commission support. It was absolutely crucial for us at that time. This help and support was and still is much appreciated, and these events in 1992 were just the beginning of the international cooperation guided and supported by the European Commission. Looking back in time it is evident that the whole process of supporting the scientific community of the Eastern European countries has been effectively organized and executed. With this help we not only managed to survive but other projects followed and gradually the conditions improved considerably. Since that time there have been a lot of projects and the research results discussed in this book have mostly been obtained in the framework of work on various joint European R&D projects. The last two of them, specifically the projects EuroDASP and DASPTOOL, are especially closely related to this book. Moreover, without the encouragement and the gentle but firm pressure applied by the project officer Mr Javid Khan this book would not have happened. Thank you, Javid! Thank you for the vision, advice and help!

In conclusion, the point should be stressed that DASP supplements rather than contradicts the classical DSP theory. There is no doubt that some of the limitations of the conventional regular deterministic DSP methods can be avoided by appropriate application of the DASP techniques discussed in the book. The purpose of this book is to provide this knowledge in order to help readers to take

reasonably well-founded decisions about when and how it is appropriate to use this approach.

The book is addressed to professionals working either in the academic world or in industry who are interested in information technologies suitable for closing the gap between the real world and computers. Hopefully professors/lecturers in universities and graduate students interested in digital handling of various signals under demanding conditions will also find this book useful.

Frequently Used Symbols and Abbreviations

a_n, b_n	Fourier coefficients
A_iC_m, A_iS_m, A_mC_i, A_mS_i	Cross-interference coefficients
ADC	Analog-to-digital converter
B_iC_m, B_iS_m, B_mC_i, B_mS_i	Cross-interference coefficients
$C_x(t)$	Autocovariance function
Cov$[x, y]$	Covariance between x and y
d	Distance between sensors
DAC	Digital-to-analog converter
DASP	Digital alias-free signal processing
DFT	Discrete Fourier transform
DOA	Direction of arrival
DSP	Digital signal processing
$E[\ ]$	Expected value of [ ]
f	Frequency (Hz)
f_s	Sampling frequency (Hz)
FFT	Fast Fourier transform
$G_{xx}(f)$	Autospectral density function
IDFT	Inverse discrete Fourier transform
IFFT	Inverse fast Fourier transform
m_x	Mean value of $x(t)$
n_k, n_x, n_y	Number of threshold levels
N	Number of samples; number of sensors in an array
$p(t)$	Probability density function
$p(x, y)$	Two-dimensional probability density function

P	Mean power
PRNG	Pseudo-random number generator
q	Quantization step
RF	Radio frequencies
$R_s(i\Delta f, t)$, $R_c(i\Delta f, t)$	Periodic rectangular functions
SFDR	Spurious-free dynamic range
SNR	Signal-to-noise ratio
t	Time variable
t_k	Time instants
t_β	Half of confidence interval corresponding to confidence probability β
TDC	Time-to-digital converter
Var [ ]	Variance of [ ]
$x(t)$, $y(t)$	Signals
$X(\omega)$, $Y(\omega)$	Fourier transforms of $x(t)$ and $y(t)$ respectively
α_i, β_i	Coefficients characterizing signal decomposition in a rectangular function basis
δ	Smallest interval on a time grid
$\delta(\,)$	Delta function
ΔT	Fraction of a period
Δt	Time interval
$\Delta\varphi$	Phase angle shift
$\{\varepsilon_k\}$	Quantization noise
θ	Observation time of $x(t)$
Θ_i	Arrival angle of ith signal
λ	Wavelength
μ	Mean value of time intervals τ_k
ξ_k	Pseudo-random numbers
$\xi(t)$	Noise
ρ	Correlation coefficient
σ	Standard deviation of time intervals τ_k
σ^2	Variance of time intervals τ_k
σ_ε^2	Variance of quantization error
τ_k	Random time intervals
$\|\varphi(t)\|$	Norm of $\|\ \|$
ω	Angular frequency
ω_s	Angular sampling frequency
Ω_i	Wavenumbers
[$\hat{}$]	Estimate of [ ]

1

Introduction: Signal Digitizing and Digital Processing

The approach used to discuss digital processing of signals in this book is special. As the title of the book suggests, the central issue concerns the performance of signal digitizing and processing in a way that provides for the elimination of negative effects due to aliasing. The term 'digital alias-free signal processing' is introduced and actually covers a wide subject area. As this term also raises some questions it needs to be explained.

1.1 Subject Matter

Signals originating as continuous-time variables, usually considered as analog signals, might be and often are processed directly using analog electronics. However, prior to processing, these signals could also be converted into their digital counterparts. Using analog-to-digital conversions for digitizing the signals prior to processing has many well-known advantages and is usually preferable. When the signals are digitized, the obtained digital signals are processed according to the concepts of digital signal processing (DSP) technology. Such an approach is becoming increasingly more popular as digital computer applications spread out into new fields and there is growing dependence on them. It is helpful that basically the same principles and techniques are used in both areas of signal processing and computing. However, signals can still be treated by analog signal processing techniques whenever application of digital techniques is either impossible or technically and economically unreasonable.

Digital Alias-free Signal Processing I. Bilinskis

Therefore the analog and digital approaches can complement each other as well as compete. There are still relatively wide areas where signals are processed in an analog way simply because the available digital techniques are not applicable under given conditions or are not good enough. Application of digital techniques is limited. The dominant and most important limitation is the highest value of signal sample that is achievable under given specific conditions. It is well defined and violation of this limitation leads to distortion of the signal processing results due to frequency overlapping or the so-called aliasing effect. Attempts to eliminate the harmful impact of aliasing have led to the development of advanced digital technologies for signal processing, specifically to the development of an innovative technology called 'digital alias-free signal processing', or DASP. This strengthens the competitiveness of digital techniques considerably. The successful use of special digitizing techniques for the elimination of aliasing has been important in showing the significance of digitizing in the whole process of signal digital processing. Many other benefits could be obtained similarly by focusing on digitizing and matching it to the needs of signal processing, as suggested by DASP. This book provides answers to questions as to what can be achieved in this way and how the signal digitizing process needs to be to altered to gain these benefits.

While the application range of the suggested approach is rather wide and various benefits could be gained in this way, specific aspects need to be taken into account. This is the subject covered by the whole book. In this chapter, the first comments are made to clarify this issue, the basic one being the attitude towards digitization of signals.

The most frequent applications of the traditional DSP technology belong to the entertainment sphere, forming the basis for audio, digital radio, TV and various other multimedia systems. Quite popular is also the use of these techniques for building fixed-line and mobile phone sets. Less visible are DSP industrial applications, especially because they are often presented as embedded systems. However, the role of digital signal processing techniques in modern telecommunication, instrumentation, industrial control, biomedical, radar, navigation and many other data acquisition and processing areas is significant.

An attempt is made in this book to focus the discussions basically to industrial applications of the considered digital technology. The industrial tasks for signal processing are typically challenging and cover a wide range of signal processing conditions. Signals have to be processed in time, frequency, modulation and spatial domains. The frequency range to be covered is very wide, extending from ultra-low frequencies up to several GHz. Processing is often multidimensional and in real-time. The signal digitizing and digital processing problems considered in this book are related to these industrial applications through described methods, algorithms, hardware, and software tools.

The latest trend in the use of digital techniques for signal processing is towards the development of ambient intelligence systems, including sensor systems and networks. They require and put the emphasis on the realization of massive data acquisition functions for supplying information from a large cluster of distributed signal sources and this function becomes increasingly important. Recently it has been found that indirectly randomized nonuniform signal processing techniques are also well suited for application in this field. The first results obtained are discussed in Chapter 7. Amazingly, the deliberate indirect randomization of sampling in this particular case can be used for purposes not related to the prevention of aliasing. The sampling randomization approach, in this case, helps to separate and distance the sampling operation from the main part of the analog-to-digital conversion structure, in order to make it an extremely simple operation that can be easily executed remotely close to the signal sources.

Thus the subjects discussed in the book cover a wide area. For already explained reasons, much attention is focused on the basic operations of signal analog-to-digital conversions. Various approaches to randomization and pseudo-randomization of signal sampling and quantizing processes are discussed in detail, including issues of adapting signal digitization to conditions of specific processing. The subjects of indirect sampling randomization and hybrid periodic/nonuniform sampling are discussed. Digital processing of signals is regarded, in general, as being based on processes that decompose signals into their component parts. Signal parameter estimation, correlation and the spectral analysis of signals, represented digitally in the alias-free way, are described. The application of various signal transforms, including nonorthogonal transforms of nonuniformly sampled signals, is studied. Much attention is given to spectral analysis and signal waveform reconstruction based on it.

It can be seen from this outline of the subject area that a large part of the topics covered is not directly related to the issue of alias-free signal processing emphasized in the title of the book. Nevertheless, all of these subjects belong to the technology of digital alias-free signal processing. Indeed, to achieve the elimination of signal distortions caused by aliasing, signals need to be digitized in a special way. Once that is done, digital representation of the original signals becomes specific. This fact has to be taken into account when processing these signals in various ways. It often turns out that the algorithms developed for processing nonuniformly sampled signals are also well suited for improved processing of periodically sampled signals under some specific difficult conditions. Furthermore, the techniques used for randomization of the sampling operations are similar to those used for randomization of the quantization operations. In the case of quantizing time intervals, they are identical. This explains why many of the topics described in the book, which at first glance may seem not to be related

to the problems of avoiding aliasing, actually belong to the subject area covered by the digital alias-free signal processing technology and therefore are discussed here.

The comments given in this introductory chapter explain the approach used in this book to studies of signal processing carried out in a digital manner. It is assumed that the readers of this book are familiar with DSP basics.

DASP, of course, is a recently developed part of DSP. Integration of DASP into the general theory of DSP still has to be done. As DSP, a widely used mature technology, is well described in many excellent textbooks, there is no need here to discuss the basics of these traditional digital techniques once again. However, the subject area of this book, while significantly differing from classical DSP, is also closely related to it. For that reason, the traditional methods and techniques for processing signals digitally are discussed, but are considered in the light of their relationship to specific nontraditional DASP techniques.

1.2 Digitizing Dictates Processing Preconditions

At first glance, the alias-free digital techniques are invariant with regard to their applications. Referring to the customary classification of DSP application areas, the techniques are applicable for nonparametric, model-based and statistical signal processing. The original analog signals are always converted into their digital counterparts and the obtained digital signals are then processed as required. On the other hand, the conditions for various types of applications might differ to a large extent and, consequently, the digital techniques used for applications in various areas usually need to be specific. To organize the process of signal processing in the best possible way, conversion of the original analog signals into their digital counterparts should be carried out while taking their exact characteristics into account for usually there are several ways to represent a signal digitally. Some of the digital representations may turn out to be better for some applications than others. It is certainly beneficial to learn how to digitize a signal under a given set of conditions so that the best results are obtained. Therefore signal digitizing should be optimized whenever possible. A large part of the book is dedicated to describing various techniques for signal sampling and quantizing, including issues of optimizing digitization by matching signal digitizing techniques to the conditions of digital processing dictated by specific applications. It should be kept in mind that the present digitizing approach determines the conditions for subsequent processing of the obtained digital signals. However, this fact becomes meaningful and can be exploited beneficially only if the digitization processes can be flexibly adjusted to the needs of subsequent processing of the digital signals. As explained in Chapter 2, randomization is used as a tool for achieving this.

1.2.1 Connecting Computers to the Real-life World

As the real-life world is basically analog, so are most of the signals reflecting observed processes. Computers, on the other hand, are digital. Therefore there is a gap between the real world and computers. Signal processing techniques have the responsibility for filling this gap. To accomplish that, the original signals have to be converted into the digital form first. Only after that can these digital signals be transferred to computers either directly or after performing some preprocessing of the obtained raw digital signals.

This outlook on the basic function of digital technology for signal processing is used in the following chapters to generalize the approach to studies of the considered topics. The achieved progress of converting various types of analog signals into their digital counterparts and of processing them under difficult conditions is weighed against the requirements of the general task of connecting computers (or other digital computing and data transmission devices) to the real-life world.

The involved preprocessing functions, while secondary, are also crucial. Their contribution to linking the signal sources to computers is often invaluable. For this reason, much attention is paid to a careful consideration of special software/hardware subsystems or devices used to perform the needed preprocessing. They are usually capable of doing the job in a cost-effective way, helping the computers to carry out the required signal processing and providing the information sought.

How preprocessing is organized depends on the conditions and the specific work being done. In the present case, signals are digitized in a specific nontraditional way. Consequently, the techniques used for the raw signal preprocessing described in the following chapters are unusual.

In cases where a computer is used for decision making in a control system, the developed code, representing the reaction of the computer to the information carried by the input signals, is transformed again into a digital signal and digital techniques are used for executing the generated commands. In such cases the computer is connected to the analog world both by its input and its output. The feedback calculated and presented in the digital form at the output has then usually to be converted back to the analog form whenever digital signals are not acceptable by the objects under control. It is assumed that traditional techniques can be used in this case so this topic is not considered here.

1.2.2 Widening of the Digital Domain

One of the basic objectives of these research and development activities has been widening of the digital domain over the area where the analog

signal processing techniques are still used almost exclusively. To reach this goal, effort should apparently be focused on replacing the mixed analog–digital or analog techniques by digital. The basic part of this digital domain is formed by low-frequency digital applications. Therefore the direction in which the digital domain should be further expanded is towards higher frequency applications. That is the reason why special attention is paid in this book to digital techniques related to processing radio frequency and microwave signals.

The task of processing signals digitally at higher and higher frequencies has been attractive. However, it is not very easy to develop improvement there. The most serious obstacle preventing progress in this direction so far has been the aliasing effect that inevitably accompanys the applications of classical digital techniques. This dictates the necessity to filter off all frequencies above half of the sampling frequency or, in other words, to use a sampling frequency at least twice higher than the upper frequency in the spectrum of the signal to be processed. The aliasing-induced corruption of the signal is the penalty for not meeting this requirement. In the case of the traditional approach to signal sampling, when the sample values are taken periodically there is no way to increase the upper frequency in the signal spectra and still avoid aliasing except to go to higher and higher sampling rates. Thus the uppermost rate of taking signal sample values that is achievable at reasonable cost using the currently available microelectronic device manufacturing technologies determines the highest frequencies that could be handled digitally. As these manufacturing technologies are being continuously improved, the upper frequency limit of the digital domain is also going up. However, the digital domain enlarges in this way relatively slowly.

The alternative is to use a principally differing approach to the problem of avoiding aliasing. As shown later, avoiding frequency overlapping has to be based on a sufficiently high frequency periodic sampling or on the nonuniform digital representation of the signals. Needless to say, such an irregular digital representation of signals drastically differs from the traditionally used regular one. Furthermore, these differences do not only concern the signal sampling. When and if the original analog signals are converted into sequences of their sample values placed on the time axis nonuniformly, the following digital processing process has to be organized in such a way that it takes into account the specifics of the used digitization process. The exceptionally important role of signal digitizing needs to be recognized. Paying sufficient attention to this problem and using the right approach to represent signals in a digital form are crucial.

1.2.3 Digital Signal Representation

To convert an original continuous-time or analog signal into a corresponding digital signal, two essential operations have to be performed. Firstly, a sequence of instantaneous values of the signal, measured at discrete time instants, has to be formed. Secondly, the obtained signal readings, usually considered as signal sample values, are to be rounded off in order to express them in a numeric form.

The operation of sample value taking is referred to as 'sampling'. The rounding-off operation is quantizing. The sequence of time instants at which the samples are obtained represents a stream of uniform events, which can be depicted graphically as a sampling point process. The results of the quantization operation are used to obtain the code of signal sample values.

The digital signal obtained as a result of all the mentioned operations (sampling, quantization and encoding) is the digital substitute for the original analog signal. It is clear that in an ideal case they should be equivalent. However, in reality the digital signals always differ to some extent from the originals. In fact there is always some impact of the analog-to-digital conversion procedures on the features of the obtained sampled and quantized signals. The patterns of the sampling point processes and the specific quantization techniques used have a considerable impact on the properties of the digital signals. Under certain conditions it is possible to reconstruct the original signals from the digital ones with a high degree of precision. Nevertheless, the fact remains that the features of the analog signals and their digital counterparts differ.

In general, the requirements needed for signal analog-to-digital conversions vary over a wide range. What is good for low-frequency applications is not necessarily also good for applications at higher frequencies. The requirements for analog-to-digital conversions at low frequencies more often than not do not pose a problem. No special efforts usually have to be undertaken to meet them. That changes as signal frequencies increase. More and more attention then has to be paid to enable sampling and quantizing operations to take place. The requirements needed for technical realizations of them are obvious. These are the requirements concerning the precision of signal sample taking timing, the sampling instant jittering and the quantization threshold settling time. However, much more significant are the requirements for digital representation of the original signals.

The discovery of the essential relationships characterizing signal sampling, quantization and digital processing processes, made a long time ago, has been a dramatic achievement. As that part of history has been well documented, including references to contributions made by the involved researchers, there is no need to

discuss the matter in detail here. Only the core of these relationships will be mentioned, the essential and most famous sampling theorem. It has a very high practical value as it defines, in a simple and clear way, the basic condition for full recovery of the information carried by the original analog signals from their sample values taken at discrete time instants.

The end result of the sampling theorem is very well known and is often used in engineering practice for setting up proper working conditions for correct functioning of analog/digital electronic devices. The theorem also serves in making a choice between analog and digital approaches to a specific problem. Accordingly, the digital techniques can be used when the frequency at which sample values are taken from a signal is at least twice higher than the upper frequency of the signal to be processed. As the highest achievable sampling rate depends on the technical perfection of the electronic devices available for implementation of signal analog-to-digital conversions, the sampling theorem evidently determines the boundary limiting the field where signals could be processed in a digital manner. How wide that field is at any given moment apparently depends on the achievable quality of the microelectronic elements being produced at that time.

These considerations and widely accepted conclusions are of course true. However, a significant fact is more often than not overlooked. It is the fact that the conclusions of the sampling theorem in engineering practice are often considered in a simplified form, stating simply that the sampling rate has to be at least twice higher than the highest frequency present in the spectrum of the band-limited signal. The conditions for signal processing in reality might differ from this simple case substantially. The point is that this basic version of the theorem has been derived and actually holds fully only in cases where band-limited signals are sampled equidistantly. In other words, the simplified interpretation of the theorem is valid for the classical DSP approach developed a long time ago. As use of the periodic sampling approach is still overwhelming, it is often not realized that there could be any other type of digital version of the respective analog signals. Nevertheless, as shown later, that is the case.

The considerations mentioned above lead to the conclusion that the specifics of the used analog-to-digital conversions become increasingly important with widening of the frequency band within which the analog signals have to be processed digitally. It is clear that attempts to enlarge the digital domain in the direction of higher frequencies can be successful only when much more attention is paid to signal digitization processes than is customarily done for digitizing relatively low frequency DSP applications. This is true both for sampling and quantization principles and their technical implementations. For example, if the

discrepancies between the expected and actual sampling instants (sampling instant jitter) usually could be and are neglected in audio frequency applications, then these discrepancies have to be kept in the range of a few picoseconds or even smaller in cases where the signal upper frequency reaches hundreds of MHz. To achieve that, the involved digitizers should have special designs.

The necessity to process signals digitally at higher and higher frequencies is strongly motivated. Much money and effort has been invested in attempts to meet growing demands in this area. It is not an easy task. When looking for the obstacles that slow down expansion of the DSP application field in the direction of higher frequencies, it can easily be seen that the bottleneck is analog-to-digital conversions. Indeed, processing 16 or 32 bit words at clock frequencies measured in hundreds of MHz is easier than providing analog signal conversions into digital signals within a sufficiently wide dynamic range at very high sampling rates.

The basic parameters of analog-to-digital converter (ADCs) that characterize their precision, dynamic range and frequency range of the input signals are the quantization bit rate and the sampling rate. The dynamic range and the bandwidth, usually limited to half of the sampling frequency, are clearly traded off, and this parameter combination might serve as an indicator of achievable performance levels. Increasing the sampling rates typically leads to considerable narrowing of the system dynamic ranges to figures unacceptable for many applications. Consequently, there is often a deficiency in the DSP subsystem dynamic range for high-frequency and low-noise applications and insufficiency in sampling rates (or bandwidth) for high-precision applications.

To achieve progress in the development of high-performance digital technologies for signal processing at significantly increased frequencies, two different approaches could be used in parallel. In addition to the ongoing process of improving the DSP hardware/software tools built on the basis of well-established DSP principles, innovative techniques for signal digitizing and fully digital processing, based on nontraditional signal digitizing concepts, could be developed and used for a very wide variety of applications.

This book is devoted to the exploration of this second approach. As shown in Chapters 2, 3 and in numerous other places in the book, an effective approach to digital processing of wideband signals is based on resolving the problem of frequency overlapping, observed as the aliasing effect. It is shown that elimination of aliasing opens up the possibility of handling signals digitally at frequencies exceeding half of the sampling frequency and that this approach is applicable widely. It is based on the application of various nonuniform sampling techniques. Much effort has been spent in the book on studies and descriptions of this quite

fruitful anti-aliasing approach. Using it leads to solutions of various essential engineering problems.

1.2.4 Complexity Reduction of Systems

Digitizing techniques suitable for signal digitizing at very high frequencies are obviously much needed for the development of fully digital signal processing techniques in the frequency range up to several GHz. Such techniques are essential for many important applications including telecommunications. Attempts to use the traditional techniques, even when possible, lead to complicated and costly system designs. In the following chapters nontraditional digital techniques are investigated that might be successfully used both for expanding the digital domain and for reducing the complexity of various systems. These two types of benefits go hand in hand and are usually obtained whenever the discussed techniques, based on nonuniform sampling, pseudo-randomized quantizing and matched processing algorithms, are used correctly. Note that the second type of benefit, the simplification of designs, can often also be obtained in cases where the signals to be processed contain components that do not exceed a certain limit of relatively low frequencies and there are no principal obstacles preventing application of the classic DSP techniques.

The fact that much attention in the book is paid to elimination of aliasing might have a misleading side effect. It might lead to the wrong impression that the suggested and discussed special digital techniques are suited and recommended exclusively for applications related to processing radio frequency and microwave signals. That is definitely not so. A large part of the discussed techniques could be used with good results for developing less complicated devices and systems for signal processing than more traditional systems. Other benefits apart from elimination of aliasing could be targeted and gained.

For example, substantial complexity reduction of designs and data compression could be achieved by using the special quantization techniques and their application is actually invariant to signal frequencies. Unlike deterministic quantization, randomized and pseudo-randomized quantization provide unbiased results. Therefore, such quantizing could be used to reduce quantization errors, leading to the reduction of the quantization bit rate. It is also appropriate for applications where it is desirable to use rough quantization. This quantization mode is not only of practical interest from the viewpoint of reducing the bit streams representing quantized signals. What is more important is the fact that ADCs, when they contain only a few comparators, can be built as extremely broadband devices. They are well suited for building low-power multichannel systems and their

performance might be enhanced by applying oversampling and low-pass filtering in combination with randomization of threshold levels.

It is also possible to gain by exploiting the most remarkable properties of pseudo-randomized quantization. It can be performed in such a way that the corresponding quantization noise has very advantageous properties not provided by deterministic quantization. As shown in Chapter 5, this noise is distributed uniformly, is decorrelated from the input signal, has no spurious frequencies and has a constant power spectral density over the whole frequency range independently from the input signal and the number of threshold levels used. Consequently, application of such pseudo-randomized quantization often turns out to be optimal in the sense that the required performance is achieved by processing the minimum number of bits.

Many methods and algorithms initially developed for dealing with nonuniformly sampled signals have proved to be quite useful for resolving essential problems typical for processing signals belonging to the lower frequency range. Use of nonorthogonal transforms might be mentioned as an example illustrating this. Their first application was reconstruction of nonuniformly sampled signal waveforms. Then it was discovered that they are also good for processing signals at extremely low frequencies. For example, they can be used to remove the negative effect caused by cutting off part of a signal period when, according to the classical definition, the processing should be carried out over a number of integer signal periods (or the periods of their separate components). These transforms are also useful for decomposing signals into several basic parts simultaneously or for extraction of signal components under conditions where the spectra of the components partially overlap.

To add to the given examples, the systems mentioned above for massive data acquisition will be discussed again. They are remarkable as a showcase demonstrating that sometimes it is beneficial to use nonuniform sampling techniques for other purposes rather than for the elimination of aliasing. While such massive data acquisition systems might be used widely, the applications related to data acquisition from multiple sources of biomedical signals are especially well suited. A specific nonuniform remote sampling procedure, based on waveform crossings, is then performed in order to digitize signals as close to each of the distributed signal sources as possible, and to do it in a very simple low-power way as described in Chapter 11. The output of such a sampler, at each sampling event, is a single pulse generated so that its position in time reflects the corresponding instantaneous value of the particular input signal. These pulses are transmitted over wire or radio links to the master parts, where the whole input signal waveform is reconstructed. Application of these specific nonuniform techniques for remote

sampling leads to substitution of the standard multiplexing of analog signals by gathering multiple output signals taken off remote sampling units distributed over a wide area. This approach of using this specific kind of nonuniform sampling results in a number of benefits. The number of signal sources from which data could be acquired in this way, in comparison with the case where analog signal multiplexing is used, is increased dramatically, up to several hundreds of such sources at least.

All these examples show that it is crucial to digitize signals in a way best suited to the particular case of signal processing. How to do this is considered in the following chapters.

Recognition of the extremely important role that digitizing plays in the whole process of digital processing of analog signals is the reason why much more attention has been paid than usual to the issues of digitizing analog signals in this book. Flexible digitizing adaptable to the specific conditions of the given signal processing task is considered here as the key factor in the fruitful application of digital techniques and for obtaining in this way various significant benefits.

1.3 Approach to the Development of Signal Processing Systems

When a task for signal processing turns up, the basic concern is usually to find the best algorithm prescribing how the required processing has to be done. More often than not little attention is paid to the input signal format, assuming that it is given as a digital signal or can be easily converted to it if the signal originally is analog. That is the most often applied traditional DSP approach and in many cases there is evidently nothing wrong with it as it often leads to good results. However, this is true only conditionally.

This traditional approach is based on the assumption that there are no problems in converting the analog input signal into its digital counterpart. While that is more or less the case under conditions typical for processing relatively low frequency signals, in more demanding signal digitizing cases the situation might be quite different. Indeed, when an extremely wide dynamic range has to be achieved, when the signal to be processed is wideband and contains components at high frequencies, and in many other cases, the analog-to-digital conversions of the input signal may prove to be the crucial stage in the whole signal processing process. The point is that signal digitization is a vital component of digital processing and this fact is fully recognized in this book. This is the basic difference between the offered and discussed techniques for alias-free digital processing of signals and the widely used techniques based on periodic sampling and fixed-threshold

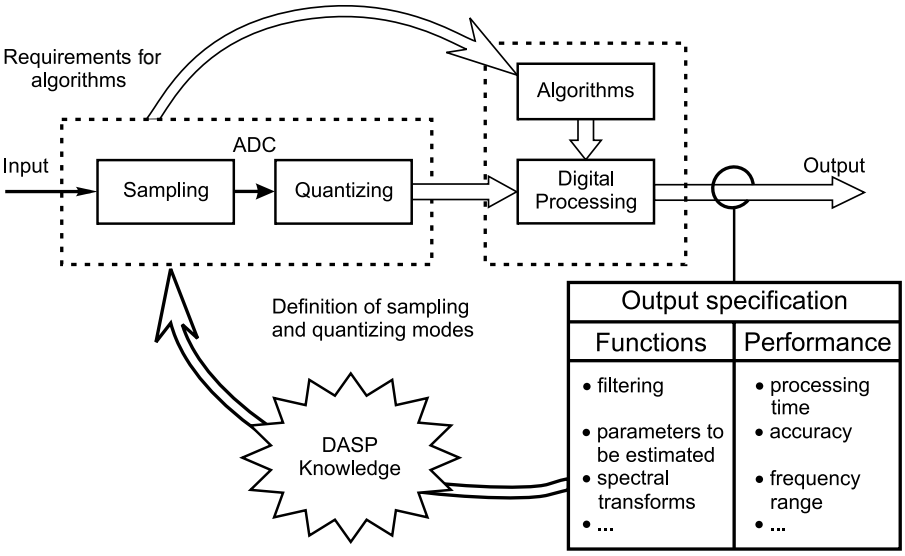


Figure 1.1 Closed-loop analysis and definition of digital signal processing system designs

quantizing. It leads to a specific approach to signal digitizing and processing. This approach will be considered more closely.

It makes sense to approach the analysis and definition of the designs of the digital signal processing system in the way suggested in Figure 1.1. The starting point for the development of a design concept of a signal processing system is the specification of its output. Such a specification has to be given in terms of functions to be fulfilled and the performance quality to be provided for. Then the organization of signal digitization procedures, based on selection of the most suitable sampling and quantization modes, is carefully considered next, targeting satisfaction of the specific processing requirements. The decisions taken at this stage are based on knowledge of the area of various possible sampling and quantization techniques, their capabilities, advantages and limitations. After it has become clear how the signal digitizing should be carried out and how the digital signal will be defined, the development cycle can be concluded by choosing or defining the algorithm for solving the given specific task with the specific digital signal features taken into account.

The choice of good algorithms for processing the digitized signals is of course important for developing any digital signal processing system. However, these should not be set up before it is clear how the input signal will look digitally. For

instance, algorithms good for processing periodically sampled signals in most cases are not directly applicable for processing nonuniformly sampled signals.

The idea of this closed-loop design development approach is simple and universal. However, it would not make much sense to apply it for optimization of traditional DSP systems. The limitations in matching the common analog-to-digital conversion process to the specifics of the given task for signal processing would prevent good results from being obtained. The problem is that the commonly used periodic sampling and fixed-threshold quantization operations, constituting the basis of a typical ADC, are rigid. Little could be done in an attempt to adapt them to the specific conditions of a given signal processing case. It is only possible to vary the time intervals between signal sample taking instants at sampling and to vary the precision of the sample value rounding-off at quantization.

There are many additional ways that operations of signal digitizing can be diversified if they are deliberately randomized. Good and not so good effects might result from this. The involved processes and relationships dictating them are quite complicated. This book is aimed to discover many of the basic ones.

1.4 Alias-free Sampling Option

The effect of aliasing, as well as the more popular means of avoiding it, of course, is well known. Traditionally the negative consequences caused by aliasing are usually accepted as unavoidable. Actually, this is not so. It is possible to avoid overlapping of signal spectral components and to distinguish them without increasing the mean sampling rate. Consider how that could be achieved.

1.4.1 Anti-aliasing Irregularity of Sampling

Assume that a digital data set, representing a signal sample value sequence, is given. This is shown graphically in Figure 1.2. Look at these signal samples and try to imagine how the signal looks from where they have been taken. It is hard to do that. The digital sample values have to be processed to reconstruct the original signal they belong to. In this particular case, the indicated sine function 1 (solid line) is found to fit the data. Therefore it should be the signal from which the sample values have been taken. However, if the reconstruction process is continued, it becomes clear that there are other sinusoids at differing frequencies, which also can be drawn exactly through the same sample value points as the first. All these sinusoids (dotted curves) are aliases and overlapping of their sample values is aliasing. The aliasing effect leads to an uncertainty. Indeed, the given sine waves of different frequencies fit equally well all of the indicated frequencies.

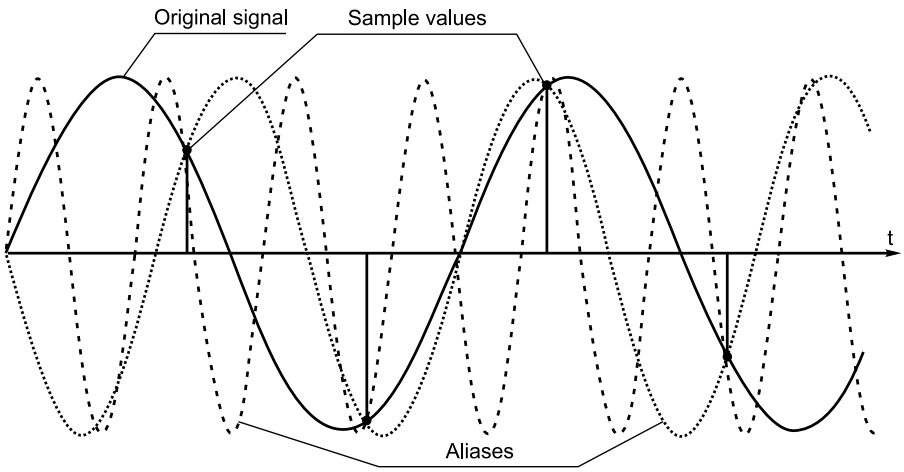


Figure 1.2 Overlapping of a periodically sampled signal component and its possible aliases

To ensure that a digital signal can provide the correct original analog signal, the bandwidth of the signal should not exceed half of the sampling frequency. If all spectral components outside the limited frequency band are filtered off the original signal or more signal sample values are taken within the same time interval, there would be no uncertainty. Either one of these possible actions impose limitations on the bandwidth of the signal, which could be sampled at the given sampling frequency without corruptions due to aliasing. Apparently, if some other way could be found to avoid aliasing, a special application of oriented digital processing of signals would be possible in a much broader frequency range. That would open up a broad area of new beneficial digital signal processing applications.

However, it is not clear whether sampling that is not corrupted by aliasing is feasible at all. In an attempt to find an alternative approach to realization of this operation, look at the diagrams of Figure 1.2 again. Notice that the time intervals between the taken signal sample values are of equal length. Now try to vary these intervals. The sample values of all indicated sinusoidal curves become different, even at small changes in distances between them. That clearly is interesting as it means that taking signal sample values irregularly disturbs the aliasing phenomenon.

To see, in a more detailed way, the consequences of this fact look at Figure 1.3. The signal shown (solid line) is the same one as given in Figure 1.2. The lower frequency sine function is again sampled and the corresponding data set is

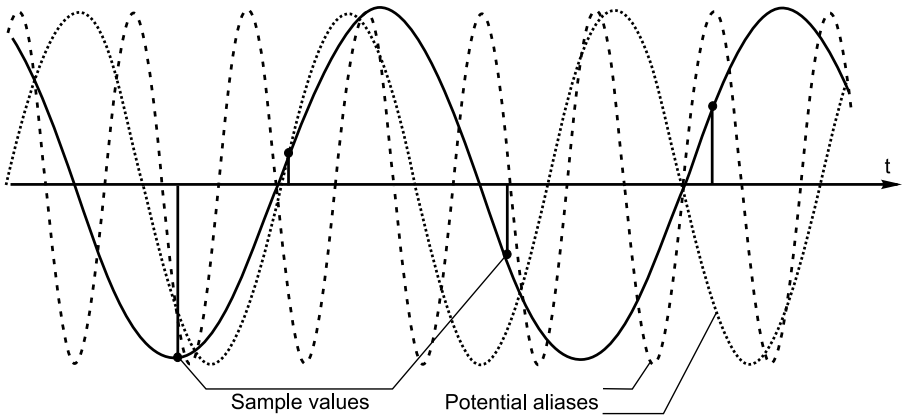


Figure 1.3 Only the signal waveform (solid line curve) exactly fits the nonuniformly spaced sample values

obtained. However, the distances between the sampling instants along the time axis now differ. They are irregular. Amazingly, this proves to be very useful. Indeed, as can easily be seen, now only one sine function can be drawn exactly through the points indicating the signal sample values. The sinusoidal curves at other frequencies simply do not fit them.

The results of this simple experiment suggest that the digital signals formed by using the nonuniform sampling operation should have features strongly differing from typical features of the digital signals obtained in the cases when signals are sampled periodically. Actually, this presumption is true. The content of the following chapters confirms this fact.

Studies of this kind of sampling show that nonuniform sampling of sinusoids at different frequencies provides differing data sets. Therefore irregularly or nonuniformly sampled signals have no completely overlapping aliases like those observed at periodic sampling. Consequently, it can be expected that application of nonuniform sampling should open up the possibility of distinguishing all spectral components of the signal, even if their frequencies substantially exceed the mean sampling rate. Studies, including experimental studies, confirm that this theoretical expectation is true. Real systems have been developed, built and exploited that are capable of processing signals fully digitally in a frequency range many times exceeding the mean sampling rate. Some of these are described in Chapter 11.

Taking signal sample values irregularly eliminates the basic conditions for aliasing to take place. That is very desirable for such sampling. Therefore it seems that

nonuniform sampling should be preferable. On the other hand, it is also evident that the end result of nonuniform sampling, the sequence of nonuniformly taken signal sample values, differs significantly from the result of periodic sampling.

Even such superficial consideration of the illustration of nonuniform sampling effects leads to the conclusion that this special approach to sampling has advantages and disadvantages. Apparently no conclusions whatsoever could be made on the grounds of only the diagrams given in Figure 1.3. The issue of nonuniform sampling is clearly too complicated for that.

1.4.2 Sparse Nonuniform Sampling

Intuitively it might be hard to accept the idea that sometimes it could be possible to take signal sample values at a rate below twice of the upper frequency of that signal and yet be able to recover the essential information carried by it. At first glance it seems that the signal sample values need to be taken often enough to keep track of the signal changes in time. To do that, the time intervals between successive sampling instants should be sufficiently short to ensure that the signal increment during the sampling interval does not exceed a certain limit. This kind of reasoning leads to the conclusion that the sampling rate has to be at least twice as high as the highest frequency present in the signal.

However, it is also possible to look at this problem in a different way. Whenever a continuous-in-time signal is to be digitized and the best sampling technique has to be found for this, the spectrum of the signal, the acceptable mean sampling rate and subsequent processing of the digitized signal need to be considered. The sampling operation should be carried out in such a way that the sequence of signal samples obtained is as closely related to the original signal as possible. However, there are other considerations that usually have to be taken into account as well. To explain what specifically is required, consider the simple example illustrated by Figure 1.4.

Figure 1.4 displays the panel of a Virtual Instrument of the DASP Lab System described in Section 2.4. This hardware/software system contains a special digitizer and a PC. The digitizer converts the analog input signals into the digital form and then this digital signal is analysed by one of the software instruments. The system operates using the digital alias-free signal processing algorithms and the mean sampling rate is equal to 80 MS/s (megasamples per second). The particular instrument shown is a Vector spectrum analyser. Normally the analog input signal to be analysed would be converted by a digitizer into a stream of nonuniformly taken sample values and then this digital signal would be analysed. However, the DASP Lab System could also be used in the rapid prototyping

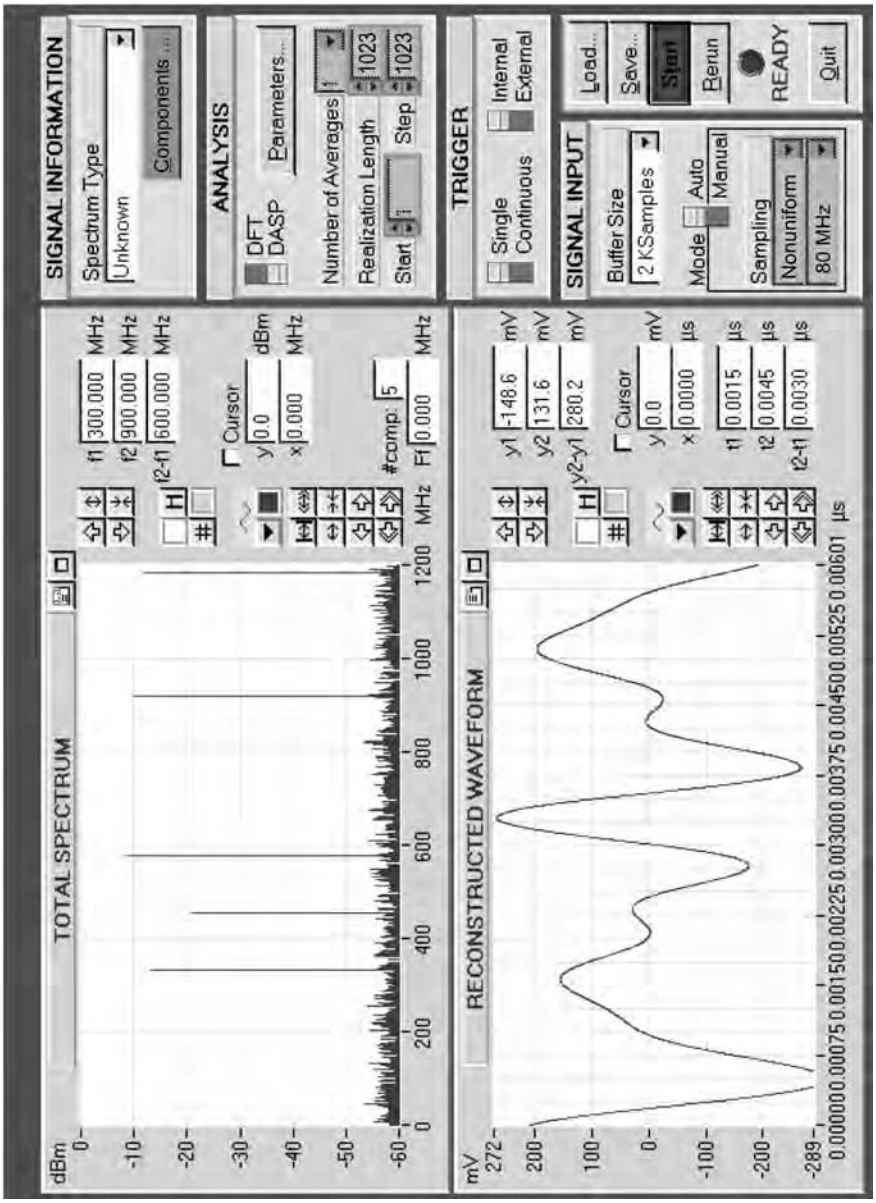


Figure 1.4 Digital signal analysis and waveform reconstruction in the frequency range up to 1.2 GHz

mode. In this case, to illustrate application of the nonuniform sampling processes, the input signal is synthesized and the Vector spectrum analyser of the DASP Lab System is used to obtain the spectrogram and the reconstructed waveform shown. The frequencies of the signal components are indicated on the given spectrogram.

Note that the frequencies of the signal components might have been chosen and placed on the frequency axis arbitrarily. As the highest frequency in this particular case is 1.185 GHz, the required sampling rate would be at least 2.370 GHz if the signal is sampled periodically. However, in the case of this example, the signal is sampled nonuniformly so the task of spectrum analysis and reconstruction of the signal waveform is resolved by using sparse sampling at a mean rate of 80 MS/s.

It is possible to reconstruct the signal waveform from such a sparse sequence of sample values because the signal is ergodic and quasi-stationary. The parameters do not vary during the time period it is being observed. Under these conditions, a reduced number of independent sample values are needed to reconstruct it by estimating all three parameters (amplitude, frequency and phase angle) of all signal components. In this case, the time intervals between the sampling instants might be large and the mean sampling rate used in this particular case is 80 MS/s. This means that it is approximately 30 times lower than it would have to be in the case of periodic sampling. Therefore about 30 times less signal sample values have been taken during the time interval the signal has been observed. The spectrogram of this particular example contains components in a wide frequency range and is shown in the upper window of the instrument panel given in Figure 1.4, while the reconstructed signal waveform is displayed in the lower window.

The example demonstrates that once aliasing is somehow eliminated, the rate of sampling required for reconstruction of the original signal does not depend on the highest frequency component of it. For instance, even a much lower sampling frequency than 80 MS/s could be used for analysis and reconstruction of the signal considered above because it is stationary and parameters of it do not change during the observation time.

This does not contradict the sampling theorem. If the sampling process is periodic, as this theorem assumes it to be, then the sampling frequency has to be high enough. Otherwise there will be aliasing and it will be impossible to estimate the signal parameters. The situation changes completely if aliasing is taken out by introducing nonuniform sampling and estimation of signal parameters in an appropriate way. Then signal sample values could be taken at a much slower rate.

Use of periodic sampling under the same conditions would result in taking many more sample values at a sampling rate about 30 times higher. Apparently the excessive sample values would not add information in this particular case. They would serve only to resolve the uncertainty caused by overlapping signal spectral components. The additional sample values would also help to reduce the impact of noise present in the signal.

This leads to the conclusion that using nonuniform sampling based anti-aliasing techniques is quite beneficial under the given conditions. Nonuniform sampling makes it possible to compress data significantly, so much simpler electronic circuitry is needed to complete the task. Just imagine how much more complicated the hardware would need to be in order to execute the periodic sampling operation at the required frequency of 2.370 GHz and to perform vector spectrum analysis of the extremely wideband digital signal.

Therefore, as the example suggests, avoiding aliasing in some other way not based on the use of high-frequency sampling should lead to a reduction in the requirements for the mean sampling rate and to other related benefits. In other words, the introduction of such sparse sampling should result in lifting the frequency limit to some higher level and in widening the digital domain in the direction of higher frequencies. In addition, the application of sparse sampling might also be good from other points of view. For instance, if a particular task for signal processing could be resolved by processing fewer signal sample values, data compression would take place, and that is always beneficial.

Application of sparse sampling (or undersampling) makes sense if the conditions are right. However, it has to be kept in mind that this sparse sampling is also necessarily nonuniform as only this kind of sampling would lead to suppression of aliasing. Consequently, processing digital signals obtained as a result of this kind of sampling has to be carried out in a special way that is suitable for handling nonuniformly sampled signals, which is a significantly more difficult task than processing periodically sampled signals.

Of course, there are also application limitations for nonuniform sparse sampling, but they differ from the limitations characterizing application conditions for periodic sampling. In the case of nonuniform sampling, the limitations on the lowest sampling rate are imposed by signal parameter variation dynamics rather than by the upper frequency of their spectra. This changes the attitude to establishing the required parameters for used sampling drivers. The signal nonstationarity issue becomes the primary consideration and analysis of the expected signal behaviour has to be carried out to determine the requirements for the designs of the sampling driver including the required mean sampling rate.

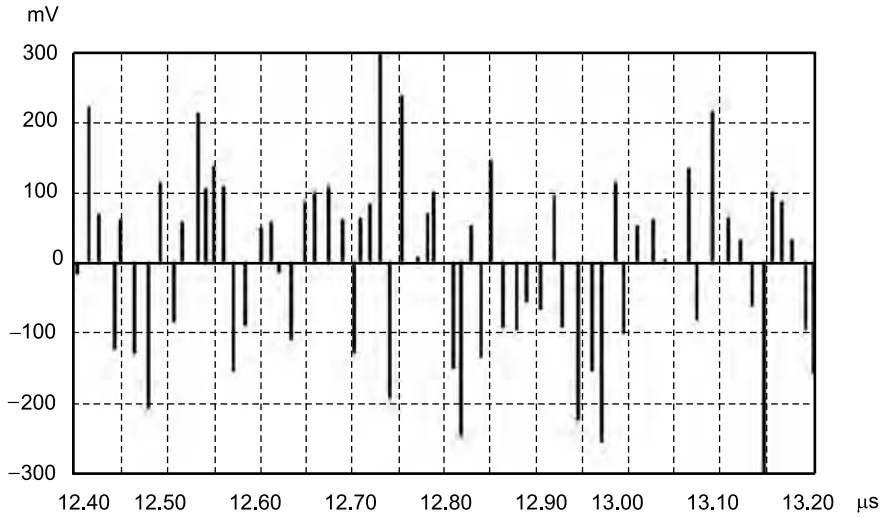


Figure 1.5 A typical block of signal sample values taken from a signal nonuniformly

1.4.3 Nonuniform Sampling Events

The crucial issue of timing nonequidistant sampling events will be mentioned in order to show some typical problems for nonuniform sampling and possible approaches to their resolution. A part of a signal sample value sequence belonging to the signal shown above is given in Figure 1.5. It can be seen that the intervals between the sampling instants vary. This represents a problem as the exact positions of all signal samples on the time axis have to be fixed, in addition to providing related sample values. In general, the digital signals formed in the case of nonuniform sampling should contain two times more bits than the digital signals obtained in the case of periodic sampling. The necessity to measure the time instants of sampling and to spend two times more bits for a digital description of a signal is difficult in itself, but this condition is especially worrisome in the light of the additional computation complexity caused by it.

However, more detailed consideration of this situation reveals that there is a much better approach to this problem; in fact it is possible to avoid doubling data volumes that have to be processed when sampling is performed nonuniformly. A special approach to realization of nonuniform sampling has to be used for that.

In principle, there are two options: measuring each sampling instant digitally or performing the sampling operation at predetermined time instants. The first

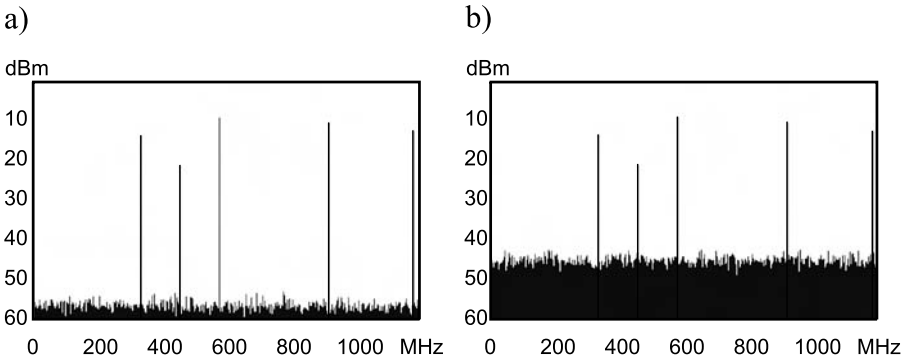


Figure 1.6 Impact of sampling jitter on spectral analysis of signals: (a) estimated amplitude spectrum of a signal in the case when the mean square value of the jitter is equal to 5 ps; (b) the same spectrogram of the same signal, obtained in the case when the mean square value of the jitter is 20 ps

option, measuring digitally the instants exactly when each signal sample has been taken, is a quite demanding engineering task, especially if the required time resolution is taken into account. Indeed, the period, for example, of the 1 GHz signal is 1 nanosecond. To sample such a signal, the smallest time digit obviously has to be equal to a few picoseconds.

The second option is much better. To realize it, the required sampling point process has to be generated and the instants when the signal sample values have to be taken need to be memorized. This information is then used both for driving the sampler and for digital processing of the sampled signal. The data indicating the exact instant of each sampling event are kept in the memory. Therefore only one digital number per sample value taken appears at the output of the digitizer performing the sampling operation in this case. As it is much easier to realize this second approach in timing nonuniform sampling events, typically it is now almost always used.

According to this approach, the signal sample values are to be taken at exactly predetermined time instants. However, in reality there is always some discrepancy between the dictated and the actual sampling instants. In other words, sampling instants jitter. Apparently this jittering has to be kept within certain margins. The impact of sampling instant jittering depends on the specific signal processing taking place. This is illustrated by the spectrograms given in Figure 1.6. It can be seen that in this particular case stronger jitter leads to significantly increased noise floor of the spectrograms.

The impact of sampling jittering is potentially even more damaging than indicated above. Not only does the average noise level increase as a result of such jittering but it might also lead to peaks at some spurious frequencies.

Imperfections of wideband signal sampling timing are measured in picoseconds. This indicates the scale of engineering problems that have to be resolved in order to achieve the high performance of digitizers based on nonuniform sampling of signals. They are quite serious and the required quality of the involved electronic designs is rather high. It is almost impossible to predict the behaviour of this kind of device theoretically. That is especially true in cases where these devices have to operate in a wide temperature range. Therefore, to obtain the data characterizing the expected performance of the digitizers, their performance has to be studied experimentally. That has been done. The obtained results and the engineering experience accumulated in this area confirm the feasibility of using such nonuniform samples in a rather wide frequency range. The upper frequencies of signals digitally processed in this way, at the time of writing this book, might reach the level of a few GHz.

1.5 Remarks in Conclusion

To conclude this first introductory chapter, some remarks will be made to summarize the basic message.

Although the quality of a modern ADC is high, signal digitization nevertheless often represents the weakest link in the chain of successive signal conversions from data acquisition to the performance of the required digital transformations. In fact, it is possible to process the digital signals in a much wider frequency range than in the cases where the original signals to be processed are analog. Therefore, digitizing actually represents the potential DSP bottleneck. This is true not only because the digitization processes determine the ultimately achievable speed of signal processing. The signal sampling and quantizing operations also impact on the quality of signal processing and this impact is much stronger than is usually realized. These facts lead to the conclusion that digitizing signals deserves much more attention than it usually receives. Although there is steady progress leading to the production of better and better microelectronic devices for signal digitizing, progress is relatively slow and costly as it is based mainly on improvements of the involved semiconductor manufacturing technologies.

Another approach to the problem of widening the domain where signals are fully processed digitally is described in this book. It is based on application of the specific DASP technology and exploits its typical advantages including elimination of aliasing. While in many DSP application cases the simplest approach

to digitizing provides results that are good enough, in cases where the signals to be handled digitally have extreme parameters, either in the frequency, time or spatial domains, the necessity to ensure sufficient flexibility of digitizing becomes crucial. Providing this flexibility is a cornerstone of the DASP methodology. The DASP hardware and software tools form the basis for advanced flexibly adaptable digital representation of signals and matched algorithms for their processing. This technology targets the execution of digitizing in the best possible way in a rather broad frequency range. Specific nontraditional techniques have to be and often are used to achieve the required functional and parametric capabilities. Both the similarities and differences between the generic DSP and special DASP techniques are discussed in the following chapters, with emphasis on the potential benefits often obtained by correct application of the suggested techniques.

However, widening of the digital domain does not exhaust all the potential benefits obtained by skilful use of DASP. In general, it is recommended that application of these techniques be considered under the conditions when classical DSP is inapplicable and/or when it is essential to make signal processing more cost efficient, meaning costs in terms of money, hardware volume, weight and processing time. Application of DASP, when realized in the correct way, should lead to replacement of the complicated microwave circuitry by substantially simpler medium frequency microelectronics and to substitution of analog signal processing blocks by digital signals. That translates into simplification of designs and, in turn, to manufacturing cost reductions, as such devices typically are considerably simpler and could possibly be built on the basis of much cheaper microelectronic chips. Various secondary benefits might be achieved as well, such as data compression or reduction of the bit flow to be processed, decorrelation of signals and their processing errors, elimination of various systematic errors, widening of the dynamic range achieved by taking out spurious frequencies, etc.

The signal dynamics (nonstationarity, the speed of signal parameter variations) rather than the upper frequency present in a spectrum of a signal serves as a criterion limiting application of this technology in specific cases. Correct application of DASP usually requires that a certain number of signal sample values, needed for resolution of the given specific signal processing task, should be taken in a given period of time. When signal parameters vary in time slowly, the mean sampling rate might be relatively low regardless of what the highest frequency present in the signal spectrum is. On the other hand, if the signal parameters vary rapidly, the applicability of nonuniform sampling in that particular case has to be checked.

One of the typical DASP drawbacks is the need for special algorithms. Unfortunately, the wealth of algorithms and computer programs developed for DSP

more often than not cannot be directly used for DASP. Usually special and more complicated algorithms have to be developed and used. In addition, not all DASP applications are as good as the similar DSP applications in the lower frequency range. These are the principal limitations and there are limitations that will probably be eliminated sometime in the future when research will find more effective methods and algorithms.

It might seem that there are some contradictions between DASP and DSP. For instance, processing of a signal sampled at a mean rate lower than the upper frequency of its spectrum, considered to be normal in the case of DASP, is simply excluded as impossible in the case of DSP. In fact there are no contradictions. The theory of DASP supplements rather than contradicts the theory and practice of generic DSP. The cornerstone of DASP is the attitude to signal digitization. Both basic operations of digitization and sampling, as well as quantizing, are considered to be most important. Adapting them to the conditions of signal processing helps in obtaining better results. However, little can be done to vary classic equidistant sampling and fixed threshold quantizing techniques. As such rigid digitizing more often than not cannot be adapted to the specifics of DSP, more flexible methods for executing the basic digitizing operations are needed. Randomization of these operations proves to be an effective instrument for making these operations flexible. The algorithms used for processing the signal digitized in this special way naturally have to be matched to the digitization specifics.

Thus randomization is crucial for DASP. An introduction to this specific way of digitizing is given in the following chapter. It is shown there that this idea is not in fact original. There have been many attempts to use this approach in order to achieve specific benefits before. The early experiences obtained in this field are described. However, there is a problem. Unfortunately randomization of digitizing leads both to good and bad consequences. The problem is how to benefit from such deliberate randomization and at the same time avoid the degradation of signal processing quality caused by it. This is not very easy to achieve. A really in-depth understanding of the involved processes is compulsory in order to resolve this problem. In fact, the full story of this book is about just that.

Bibliography

Artyukh, Yu., Bilinskis, I., Greitans, M. and Vedin, V. (1997) Signal digitizing and recording in the DASP-Lab System. In Proceedings of the 1997 International Workshop on *Sampling Theory and Application*, Aveiro, Portugal, June 1997, pp. 357–60.

Artyukh, Yu., Bilinskis, I. and Vedin, V. (1999) Hardware core of the family of digital RF signal PC-based analyzers. Proceeding of the 1999 International Workshop on Sampling Theory and Application, August 11–14, Loen, Norway, pp. 177–79.

Balakrishnan, A.V. (1962) On the problem of time-jitter in sampling. *IRE Trans. Inf. Theory*, **IT-8**(4), 226–36.

Beutler, F.J. (1966) Error-free recovery of signals from irregularly spaced samples. *SIAM Rev*., **8**(3), 328–35.

Beutler, F.J. (1970) Alias-free randomly timed sampling of stochastic processes. *IEEE Trans. Inf. Theory*, **IT-16**(2), 147–52.

Beutler, F.J. (1974) Recovery of randomly sampled signals by simple interpolators. *Inf. Control*, **26**(4), 312–40.

Beutler, F.J. and Leneman, O.A. (1966) Random sampling of random processes: stationary point processes. *Inf. Control*, **9**(4), 325–46.

Beutler, F.J. and Leneman, O.A. (1966) The theory of stationary point processes. *Acta Math*., **116**, 159–97.

Beutler, F.J. and Leneman, O.A. (1968) The spectral analysis of impulse processes. *Inf. Control*, **12**(3), 236–58.

Bilinskis, I. (1976) Stochastic signal quantization error spectrum (in Russian). *Autom. Control Comput. Sci*., **3**, 55–60.

Bilinskis, I. (1977) Quasi-stochastic coding of continuous signals. *Pr. Przem. Inst. Elektron*., **64**, 53–9.

Bilinskis, I. (1978) Random sampling of continuous signals. *Proc. Acad. Sci. LSSR*, **6**.

Bilinskis, I. and Cain, G. (1996) Digital alias-free signal processing in the GHz frequency range. In Digest of the IEE Colloquium on *Advanced Signal Processing for Microwave Applications*, 29 November 1996.

Bilinskis, I. and Mikelsons, A. (1975) Quantizing of signals by parallel stochastic weighting (in Russian). *Autom. Control Comput. Sci*., **4**, 34–8.

Bilinskis, I. and Mikelsons, A. (1978) Random sampling of continuous-time signals (in Russian). *Proc. Acad. Sci. LSSR*, **6**, 96–101.

Bilinskis, I. and Mikelsons, A. (1983) *Digital Random Processing of Continuous Signals* (in Russian). Riga: Zinatne.

Bilinskis, I. and Mikelsons, A. (1990) Application of randomized or irregular sampling as an anti-aliasing technique. In *Signal Processing*, V: *Theories and Application*. Amsterdam: Elsevier Science Publishers, pp. 505–8.

Bilinskis, I. and Mikelsons, A. (1992) *Randomized Signal Processing*. Prentice-Hall International (UK) Ltd.

Bilinskis, I., Trejs, P.P. and Nemirovski, R.F. (1972) Verfarhen zur digitalen Messung von Zeitintervallen und Einrichtung zu dessen Realisierung. Patentscrift No. 93960, DDR.

Bilinskis, I., Trejs, P.P. and Nemirovski, R.F. (1974) Zpusob digitalniho mereni casovych intervalu a zarizeni k provadeni tohoto zpusobu. Aut. osved. 163383, CSSR.

Brannon, B. (1996) Wide-dynamic-range A/D converters pave the way for wideband digital-radio receivers. In EDN, 7 November 1996, pp. 187–205.

Brown, W.M. (1963) Sampling with random jitter. *SIAM J. Appl. Math*., **11**, 460–73.

Corcoran, J.J., Poulton, K. and Knudsen, K.L. (1988) A one-gigasample-per-second analog-to-ditial converter. *Hewlett-Packard J*., **June**, 59–66.

Hejn, K. (1978) Application of Monte-Carlo methods in measurement. In Symposuim IMEKO Tech. Comm. on *Measurement Theory*, Leningrad.

Jayant, N. and Rabiner, L. (1972) The application of dither to the quantization of speech signals. *Bell Syst. Tech. J*., **51**(6), 1293–304.

Leneman, O.A.Z. (1966) Random sampling of random processes: impulse sampling. *Inf. Control*, **9**(4), 347–63.

Leneman, O.A.Z. (1966) Random sampling of random processes: optimum linear interpolation. *Franklin Inst*., **281**(4), 302–14.

Leneman, O.A.Z. (1966) On error bounds for jittered sampling. *IEEE Trans. Autom. Control*, **AC-11**(1), 150.

Leneman, O.A.Z. and Lewis, J.B. (1966) Random sampling of random processes: mean-square comparison of various interpolators. *IEEE Trans. Autom. Control*, **AC-11**(3), 396–403.

Marvasti, F. (2001) *Nonuniform Sampling, Theory and Practice*. New York: Kluwer Academic/Plenum Publishers.

Masry, E. (1971) Random sampling and reconstruction of spectra. *Inf. Control*, **19**(4), 275–88.

Masry, E. (1978) Alias-free sampling: an alternative conceptualization and its applications. *IEEE Trans. Inf. Theory*, **IT-24**(3), 317–24.

Mednieks, I. and Mikelsons, A. (1990) Estimation of true components of wide-band quasi-periodic signals. In *Signal Processing*, V: *Theories and Application*. Amsterdam: Elsevier Science Publishers, pp. 233–6.

Mikelsons, A. (1981) Estimating broadband signal parameters at a relatively low mean sampling rate. *Autom. Control-Comput. Sci*, **1**, 90–4.

Millard, J.K. (1988) A one-gigasample-per-second digitizing oscilloscope. *Hewlett-Packard J.*, **June**, 58–9.

Rabiner, L. and Johnson, J. (1972) Perceptual evaluation of the effects of dither on low bit rate PCM systems. *Bell Syst. Tech. J.*, **51**(7), 1487–94.

Roberts, L.G. (1962) Picture coding using pseudo-random noise. *IRE Trans. Inf. Theory*, **IT-8**, 145–54.

Shapiro, H.S. and Silverman, R.A. (1960) Alias-free sampling of random noise. *SIAM J. Appl. Math.*, **8**(2), 245–8.

Part 1

Digitizing

2

Randomization as a Tool

To achieve flexibility of the digitization techniques essential for *matching* them to the given signal processing task conditions, these operations should be made variable. The problem was to find out how to achieve this. Techniques existed that were applicable for variations of signal processing procedures. It simply took some time to realize that introducing an element of randomness into signal conversion procedures, an approach initially developed for other purposes, also leads to variations in the basic digitizing functions. Once that was understood, it became clear that randomization might be used as a tool for making signal conversion operations, including sampling and quantizing, more flexible and more adaptable to specific signal processing needs. The application specifics and potential of this instrument are discussed in this chapter.

However, the applicability and efficiency of this tool is arguable. Indeed, introduction of randomness into signal processing operations might also have negative consequences, as it usually leads to increased statistical errors. The end result might therefore not be acceptable. In fact this often happens whenever randomization is carried out in a way that is not sufficiently skilful. The problem of obtaining the targeted positive randomization effect without degradation of the whole signal conversion process is crucial and is not easy to resolve. Work on the development of new alias-free signal processing techniques should always be carried out with considerations of this kind taken into account. Various approaches used to obtain possible solutions of this problem are discussed in the following chapters. This is the central issue in this book.

Gradually the approach to randomization of the sampling and/or quantizing operations and for taking care of the potential negative consequences of randomization has been improved to the extent that little remains from the techniques

Digital Alias-free Signal Processing I. Bilinskis

initially used for these purposes. Although randomization techniques are still being developed, more often than not the processes involved are fully deterministic as they are actually pseudo-random rather than truly random. If that is kept in mind then the terminology used in this chapter will help to follow the gradual developments that have taken place in this area over many years and will not lead to confusion about the discussed issues.

2.1 Randomized Versus Statistical Signal Processing

Many statistical and randomized signal processing methods and techniques exist. Only the latter are considered here. Statistical processing techniques are simply used routinely whenever needed. To avoid misunderstandings, some of the definitions will be agreed for the randomized signal processing techniques studied and discussed in this book.

In general, DSP covers all kinds of signal processing algorithms, including those of statistical signal processing. This kind of signal processing is typically used when the input signals are given as random functions and the results of their processing are often random as well (for instance, the estimates of signal parameters). However, this does not mean that such signal processing is randomized. Consider signal processing techniques to be randomized only if some of the successive signal conversion or processing procedures are performed stochastically.

It should also be realized that the conditions under which randomized signal processing might take place differ. This operational mode might be either enforced or chosen willingly. The randomness present at some stage of signal processing might be either an unavoidable reality caused by uncontrollable signal acquisition conditions or it may be introduced deliberately in order to obtain some benefits.

Indeed, sometimes signals are sampled nonuniformly simply because they can be observed only at some unpredictable random time instants. Then it is virtually impossible to control the conditions used to obtain the signal sample values. This kind of real-life nonuniform signal sampling case actually occurs quite often. Much has been done in this area. However, in this book, only the second kind of randomized signal processing, based on a deliberate introduction of randomness at some stage in signal conversions and processing, is considered, with a few exceptions (see, for instance, Subsection 20.4.4).

The difference between these two approaches is distinctive. As it is impossible to control the process of sampling in the first case, the signal sample values are obtained at unpredictable time instants and it is especially difficult to process this specific digital signal. It is much easier to organize the whole process of signal sampling and processing in the second case. Then the sampling operation might be adapted to the specifics of the signal, which helps to obtain much better results

of the subsequent signal processing. There is also much in common for both of the mentioned two specific nonuniform sampling application areas. The same or at least similar mathematical descriptions, methods for modelling, algorithms for randomized signal processing and tools for statistical analysis are often applicable in both cases.

At a first glance deliberate randomization of signal processing is perceived as the addition of a noise to the signal but that interpretation seems to be absurd. Actually the essence of randomization of sampling and quantization is different. Randomization, in fact, means substitution of some fixed or stationary signal conversion procedures by their variable or nonstationary versions. In the case of sampling randomization, equidistant signal sample taking is transformed into a nonuniform procedure of sample value taking in a nonequidistant way. The distances between sampling instants are then varied in accordance with certain rules. Likewise, randomized quantization means that the signal sample values are compared with variable rather than fixed threshold levels, as in the traditional case. Thus the randomization approach actually results in making the corresponding signal conversions more dynamic. This leads to positive and negative effects. Therefore successful randomization of signal processing requires skilful usage of these techniques, making it possible to overcome the bad effects and to benefit from the good effects. Learning the lessons of the randomization prehistory might help to provide some suggestions on how to achieve that. A brief summary of the most significant achievements marking progress in this field in the past follows.

2.2 Accumulation of Empirical Experience

The origin of randomization can be identified and traced back to the 1960s. It seems that many people at that time independently came to the idea that an element of randomness present in signal processing might sometimes be more useful than damaging. Attempts were made again and again to obtain specific benefits by adding noise, irregularizing some signal conversion procedures and using statistical estimation schemes instead of deterministic calculations. Experiments confirmed that under certain conditions various positive effects could be obtained in this way.

Comments on some of the most remarkable cases follow.

2.2.1 Using Monte Carlo Methods for Signal Processing

At a time when vacuum tubes and transistors were the basic electronic devices used to perform signal processing, it was of considerable interest to find and apply methods that would simplify the involved computations. It was noticed that the

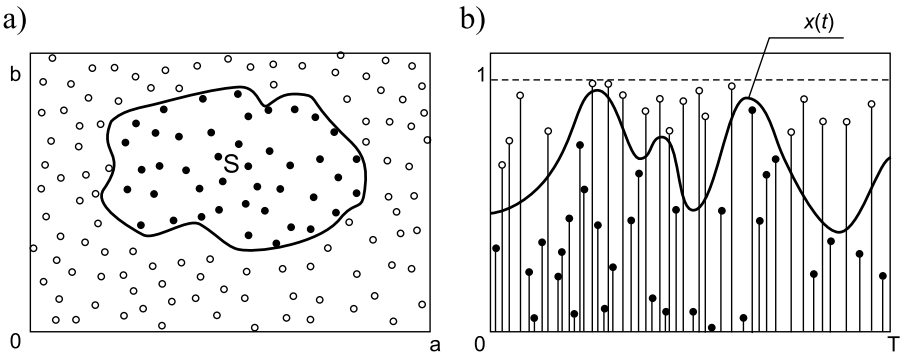


Figure 2.1 Application of the Monte Carlo method for signal processing

statistical techniques, known as Monte Carlo method, used for simplifying digital computing could also be successfully applied for the development of electronic systems for signal analysis.

Usually Monte Carlo methods had been used to estimate integrals by means of statistical experiments. The basic idea of this approach is simple. Suppose that the area S of an irregular figure, shown in Figure 2.1(a), is to be measured. By definition,

$$S = \oint_{(L)} x\,\mathrm{d}y - y\,\mathrm{d}x,$$

where L represents the contour of the irregular figure. It is evident that the solution of this task, simple at first glance, may require considerable effort because the approximation of L could be complicated.

To simplify the solution of this task, the irregular figure under consideration could be placed within a rectangle with side lengths a and b. To determine the area S by applying the statistical trial technique, N dots are placed randomly on the area ab. If the density of the dots is constant over the entire area ab, the number of dots n falling within the limits of the figure in question must be proportional to the value of its area S. Then the estimate of the area S can be calculated in a simple way, defined as

$$\hat{S} = \frac{n}{N}ab. \tag{2.1}$$

Thus the basic idea behind the Monte Carlo methods is that of substituting, where possible, complicated deterministic relationships by much simpler probabilistic ones in order to simplify the computations and programming involved. In

some fields of science and engineering, for instance in nuclear physics, these methods have been used very successfully, and in many applications they have proved to possess considerable advantages. This stimulated further developments of these methods and, as a result, new modifications emerged.

It is not surprising that Monte Carlo methods were also considered for possible applications in the field of signal parameter estimation. However, it soon became clear that only the most simple and basic of the Monte Carlo methods were suitable for developing specialized measuring devices. Only computers could realize more subtle versions of the methods, offering more advantages, especially higher precision.

Some interesting and at the same time useful hardware implementations of Monte Carlo methods were developed. They were mostly aimed at estimating integrals by means of statistical experiments. The original descriptions of the developed systems were presented in a specific way, which made it difficult to relate them to the structures of common signal processing systems. It therefore took some time before it was realized that all of the known electronic systems of this kind could be described in terms of signal analog-to-digital conversions and processing of the obtained digital signals. Using this approach revealed the interesting fact that these electronic implementations of the Monte Carlo method could be reduced to a simple scheme containing ADC and some circuitry for handling the digital signal taken off that ADC. One-bit randomized quantizing and periodic sampling techniques were typically used. Usually the mean values or sometimes the mean values of functionally converted signals were calculated.

It is evident that this approach of estimating integrals can also be applied to determine an area bounded by a coordinate axis and by some function of time $x(t)$ considered within a limited interval, as shown in Figure 2.1(b). If this function represents a signal, the mean value can be estimated by counting the random dots falling within the area limited by the signal and using the following simple formula:

$$\hat{m}_x = \frac{n}{N}. \tag{2.2}$$

Implementation of this method is easy. A scheme of an electronic device for realizing such statistical experiments is shown in Figure 2.2. To carry out the Monte Carlo statistical trials in this case, the signal instant values are compared with generated random levels. In other words, a randomized one-bit quantizer is used to round off the signal instantaneous values. The result of this operation is given as a binary number, which is assumed to be equal to binary 1 when the signal value is above the random level and to binary 0 when this value is below

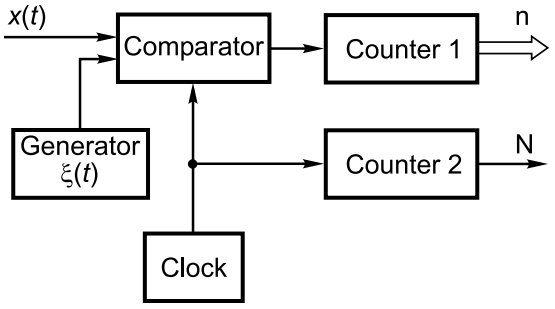


Figure 2.2 Structure of an electronic device for realizing statistical experiments

that level. The mean value of the signal is estimated simply by counting the binary 1 obtained during the whole measurement cycle of N clock periods.

At the time when this approach to statistical signal parameter estimation was developed and proposed, it was of much more interest to electronic engineers than it is now. The properties of this technique were considered advantageous and promising. The hardware was simple – a merit in itself – but also provided a digital estimation of wideband signal parameters. Only one simple block, the comparator, had to be designed as a high-frequency device. However, the main drawback of these methods – a relatively high level of statistical error – was fully present in these devices, acting as a serious negative factor and often outweighing the available benefits.

2.2.2 Polarity Coincidence Methods

Polarity coincidence methods for analysing random processes are also worth mentioning. They were developed many years ago, but cannot be recommended for practical applications today. However, they are certainly of interest as an example showing how randomization can improve the quality of signal parameter estimation. The specific signal characteristics that might be estimated using these methods are correlation functions.

The mathematical description of the basic polarity coincidence method exploits the so-called sign (polarity) function which, in the first version of the method, is given as follows:

$$\operatorname{sgn}[\dot{x}(t)] = \begin{cases} 1 & \text{for } x(t) > 0, \\ 0 & \text{for } x(t) = 0, \\ -1 & \text{for } x(t) < 0, \end{cases} \tag{2.3}$$

where $\dot{x}(t)$ is a centred random process. The polarity coincidence correlation function $R(t, t)$ is defined as an expectation of the product of polarity coincidence functions for two values of the argument

$$R_x(t_1, t_2) = E\{\text{sgn}[\dot{x}(t_1)]\,\text{sgn}\,[\dot{x}(t)]\}. \tag{2.4}$$

It has been proved that in the case of a stationary ergodic normally distributed centred signal the polarity coincidence correlation function $R_x(t_1, t_2)$ can be expressed through the probabilities of the coincidence of signal polarities at the time moments t_1 and t_2:

$$R_x(t_1, t_2) = 4p^{(++)}(t_1, t_2) - 1 = 4p^{(--)}(t_1, t_2) - 1, \tag{2.5}$$

where

$$p^{(++)}(t_1, t_2) \qquad \text{and} \qquad p^{(--)}(t_1, t_2)$$

are the probabilities of the coincidence of positive and negative signs respectively. Although the polarity coincidence correlation function characterizes the statistical relationship between remote signal instantaneous values, this function is not the same as the commonly used correlation function. Nevertheless a relation exists between them.

It has been established that the following relation exists between the polarity coincidence correlation function $R_x(\tau)$ and the normalized correlation function $\rho_x(\tau)$:

$$\begin{aligned} R_x(\tau) &= \frac{2}{\pi}\,\text{arcsin}[\rho_x(\tau)], \\ \rho_x(\tau) &= \sin\left[\frac{\pi}{2}R_x(\tau)\right], \end{aligned} \tag{2.6}$$

where $\tau = t_2 - t_1$ is the time delay. It is evident from Equations (2.5) and (2.6) that the electronic device realizing the polarity coincidence method can be relatively simple, and this simplicity is its major merit. However, the method has two essential disadvantages: firstly, the experimentally obtained estimate of the polarity coincidence correlation function is characterized by a large random error; secondly, the polarity coincidence method has a limited area of application. Indeed, expression (2.5) is valid only on condition that the distribution of the signal $x(t)$ is normal.

The above limitation can be removed if the polarity coincidence method is made more complicated, i.e. if, when determining the sign of the signal, its instantaneous values are compared not with the zero level but with some auxiliary random

function of time, uniformly distributed within the range of the signal. This is the idea behind the second kind of polarity coincidence method.

Suppose that the cross-correlation function $k_{xy}(\tau)$ is to be obtained using this method. Then two random auxiliary processes $\xi_1(t)$ and $\xi_2(t)$ are used and the following sign functions are determined:

$$\begin{aligned}
\operatorname{sgn}[z_1(t)] &= \begin{cases} 1 & \text{if } \dot{x}(t) > \xi_1(t), \\ 0 & \text{if } x(t) = \xi_1(t), \\ -1 & \text{if } \dot{x}(t) < \xi_1(t), \end{cases} \\
\operatorname{sgn}[z_2(t)] &= \begin{cases} 1 & \text{if } \dot{y}(t) > \xi_2(t), \\ 0 & \text{if } y(t) = \xi_2(t), \\ -1 & \text{if } \dot{y}(t) < \xi_2(t), \end{cases}
\end{aligned} \tag{2.7}$$

where

$$\begin{aligned}
z_1(t) &= \dot{x}(t) - \xi_1(t), \\
z_2(t) &= \dot{y}(t) - \xi_2(t).
\end{aligned}$$

It has been proved that

$$E[\dot{x}(t)\,\dot{y}(t)] = AE\{\operatorname{sgn}[z_1(t)]\operatorname{sgn}[z_2(t)]\} = AR_{xy\xi}(\tau), \tag{2.8}$$

where A is a constant that can be calculated on the basis of the $\xi_1(t)$ and $\xi_2(t)$ distribution boundaries, and $R_{xy\xi}(\tau)$ is the correlation function, determined by applying the polarity coincidence method.

It is interesting to note that the possibility of removing these drawbacks was provided by the randomization approach. It was proved that the above limitation could be removed if the polarity coincidence method is made more complicated. To achieve this, the sign of the signal is determined by comparing its instantaneous values with some auxiliary random function of time (uniformly distributed within the peak-to-peak range of the signal) rather than with the zero level. This randomization idea is behind the second kind of polarity coincidence method, illustrated in Figure 2.3(b).

While at first glance the two polarity coincidence methods for correlation analysis, the simpler basic and the more complex randomized one, again seem to be specific, actually that is not the case. In essence they can normally be described in terms of sampling, quantizing and digital processing. Using that approach, it can be shown that a fixed threshold quantizer and a randomized quantizer are used in the first and second cases respectively. Therefore the benefits are due to randomization of quantization. It will be shown in Chapter 4 that these benefits

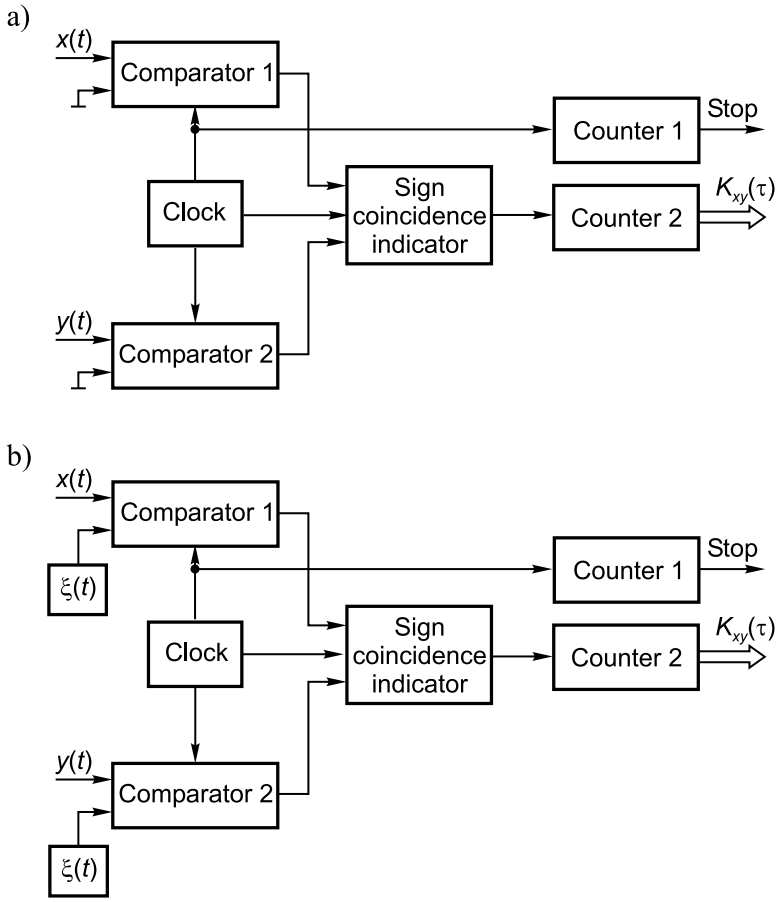


Figure 2.3 Electronic devices realizing the polarity coincidence method: (a) without and (b) with the use of an auxiliary random process

are expected as soon as deterministic fixed-threshold quantization is substituted by a randomized operation of this kind.

2.2.3 Stochastic–Ergodic Method

A new method for statistical signal analysis was announced at the beginning of the 1970s. A series of articles appeared devoted to theoretical investigations of the method, as well as descriptions of developed and factory produced instruments. This method was introduced as the stochastic–ergodic method. Here only its

fundamentals will be considered and an attempt will be made to reveal some of its peculiarities.

The stochastic–ergodic conversion of a signal is actually an analog-to-digital conversion. It is called stochastic because it uses probabilistic encoding and ergodic because the ergodicity of the signal is assumed. This is a useful description of the method, except for the fact that the estimation of a number of signal parameters also takes place in addition to the analog-to-digital conversion.

These devices are designed first of all to provide estimates of the moments of signal distribution, such as the mean and root mean square (RMS) values of the signals, as well as the autocorrelation and cross-correlation functions. Despite the diversity of measuring tasks performed, a small number of units form the basis of the different devices. The major one (according to the terminology mentioned above) is the so-called ergodic converter. In fact, the ergodic converter fulfills the functions of a randomized quantizer.

The publications dedicated to the subject of stochastic–ergodic conversion clearly show that a considerable amount of work has been successfully done in this area. The development and industrial production of instruments of this particular type deserve especial mention. For example, the signal analyser that performed measurements of signal integral parameters, such as the RMS, mean and amplitude values in the frequency range from 15 Hz to 1 MHz with errors not exceeding 1 % of the upper scale limit, was quite good for that time. The terminology used in descriptions of the techniques involved is less appropriate.

2.2.4 Stochastic Computing

Stochastic computing was proposed at about the time when second-generation digital computers were soon to be replaced by the next generation of computers built on ICs (integrated circuits). It had already been realized that the conventional digital computers, although possessing many excellent properties, were often not good enough to handle the problems associated with the simulation or control of large complex systems. Such computer applications required real-time processing of many input variables, which represented a serious difficulty for the single-CPU (central processing unit) computers since they could only perform computations on inputs sequentially.

Analog computers were better suited for such applications, since they permitted the simultaneous processing of a relatively large number of inputs. However, they also had some considerable disadvantages, intensified at that time by the relatively low element package density of ICs available then. Thus the analog computers could not be considered as economically acceptable substitutes. That stimulated further investigations and led to the development of randomized

or stochastic computing. It was proposed as a new and promising technology, capable of satisfying many of the multiprocessor system demands on the acceptance and parallel processing of a large number of varying inputs at a small cost per separate input channel.

Design specifics of a stochastic computer are determined first of all by the method used for probabilistic encoding of signals. Although conventional digital integrated circuits are used as computing elements, they in fact fulfil analog-type functions, and the stochastic computer itself is programmed as an analog computer. Stochastic computers belong to the family of analog computers, although they differ from the latter in their use of stochastic bit streams rather than direct current.

A signal or a physical variable is represented in stochastic computers by a probability $P(1)$ that a logical level in a clocked pulse sequence at the considered instant will be the binary 1. The probability $P(1)$ is an analog variable whose value ranges continuously from 0 to 1. The physical quantities represented by such stochastic bit streams therefore have to be scaled.

Binary l's in this kind of stream are represented by pulses formed at some clocking instants. The absence of pulses at other clocking instants means that 0's are located there. At each clocking instant only one of two possible events may take place: a pulse will or will not appear. Each of these pulses is generated with no statistical influence from the previously formed pulses so that the stochastic bit stream as a whole may be considered as a quasi-stationary ergodic process.

Assume that a stochastic bit stream contains m clocking time intervals. Then the ratio $P(1) = n/m$ can be used as an estimate of the probability $P(1)$, where n is the number of l's in that particular stream. If a pulse sequence of this kind is observed for m clocking intervals, the value of the probability $P(1)$ it produces can be measured with the rounding-off errors in the range of $\pm 1/2m$. To reduce the random errors, which might considerably exceed the rounding-off errors, a significantly increased length of the stochastic bit stream segment is averaged. Theoretically, only an infinite stochastic bit stream allows the exact value of an input variable to be determined.

Stochastic computers actually operate like analog computers, and so there is no fixed word length. The input signals are converted into stochastic bit streams and the stochastic quantities represented by them change relatively slowly in time following changes in the corresponding input variables. The precision of stochastic computing at a given clock frequency can be increased only by slowing down the computer and by limiting the spectra of the input signals. Hence the computing accuracy and bandwidth need to be traded off. Stochastic computers cannot be expected to provide very high computing efficiency. At reasonable clock

frequencies (about 10 MHz) an error rate of 1 % and a bandwidth of approximately 100 Hz can be achieved.

However, the moderate computing accuracy and speed are quite acceptable for many applications and the advantages of stochastic computing may well outnumber its drawbacks. When this new concept of computation was first proposed, it was strongly recommended for applications in the field of real-time automatic control and the simulation of large complex systems. It was predicted that with advances in large-scale integrated circuit manufacturing, stochastic computing would be widely and successfully employed to solve many important problems. However, these predictions did not come true. Developments in microprocessor technology allowed the design of more efficient computing systems, including multiprocessor systems. Stochastic computers could not compete and interest in stochastic computing gradually declined. However, the fact still remains that computation of stochastically represented variables can be performed in an extremely simple and economic way.

The layout and interconnections of computing elements in stochastic computers depend on the particular computations to be carried out. Several stochastic computer hardware mapping schemes are possible. The most commonly used are based on a single-line unipolar representation of the variables, which are always positive (or always negative). This scheme leads to the simplest hardware configurations of computing elements, although these elements are still quite simple even when other mapping schemes are implemented.

Multiplication is one of the most basic and frequently used computing operations. The complexity of the hardware performing this operation is an important issue for any type of computer. In the case of stochastic computers, the variables, represented by stochastic quantities, may be multiplied in an extremely simple way. For example, only a single logic AND gate is required for multiplying two unipolar variables represented by corresponding stochastic bit streams.

Many specific stochastic elements have been developed over the years. However, only relatively few types of elements are actually needed to build a stochastic computer. Such a computer is basically comprised of a large number of simple and cheap multipliers and summers.

The original signals to be processed by this kind of computer first have to be converted into stochastic bit streams. This has to be done whether the input signals are analog or digital. Therefore at least two types of converters are needed: analog–stochastic and digital–stochastic converters.

It is easy to see that a randomized one-bit quantizer can be used for conversion of an analog unipolar signal to a stochastic bit stream. If the signal at the input of such a converter varies sufficiently slowly, the obtained bit stream would represent

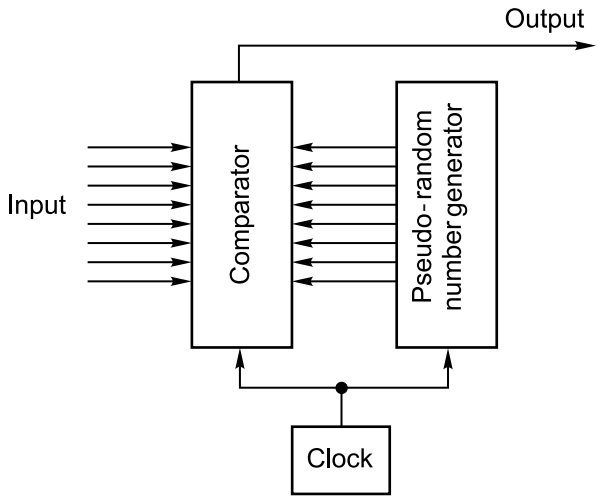


Figure 2.4 Block diagram of a digital stochastic converter

the input in a format required by the stochastic computing rules. Therefore it is not necessary to develop and use a special analog–stochastic converter as the common randomized quantizers can be used.

A block diagram of the less common digital–stochastic converter is shown in Figure 2.4. This shows that a stochastic bit stream is generated in the process of comparing the binary input numbers with the pseudo-ramdom numbers provided by the digital noise source. Binary 1's are formed at the clocking instants when the binary input numbers are larger than the pseudo-random numbers. The pseudo-random numbers used in this kind of computation are uniformly distributed in the range from 0 to 1, so that at each clock pulse instant any number within this range is generated with an equal probability. Hence the probability that an input binary number will be larger than a pseudo-random number is proportional to the input. Thus the binary 1's appear in the stochastic bit stream at the converter output with the probability equal to the properly scaled digital input signal.

This description of digital–stochastic conversion may give the wrong impression that this procedure has been developed for, and may be applied exclusively to, stochastic computing. It is true that digital–stochastic converters were independently developed to represent digital and analog signals by the stochastic bit streams needed to perform this kind of signal processing. However, the principles of their operation can also be explained in terms of sampling and quantization.

This means that the results of the following chapters may well be used for a detailed analysis of stochastic computing and for evaluating expected accuracy, bandwidth and other characteristics.

To demonstrate the possibility of such a generalization the operation of the digital–stochastic converter can be described in another way. It is obvious that the digital input signal is sampled periodically by clocking the comparator and considering the input signal only at discrete time moments. Digital samples are then rounded off to one of the two levels allowed, to 0 or 1. Note that the rounding-off operation is nothing other than quantization. It may be a little confusing that the variable being quantized in this case is digital rather than analog. However, formally there is not the slightest difference. Digital quantities may be, and often are, rounded off and a deterministic or randomized quantization procedure could be applied. In this case the threshold level changes randomly from one sampling instant to the next, and so quantization itself may be called randomized or stochastic. The digital samples of the input signal are compared with the threshold level (with the pseudo-random numbers), and when the value of the samples exceeds the threshold level they are rounded off either to level 1 or to level 0.

Thus analog–stochastic and digital–stochastic converters are actually devices performing the sampling and quantization operations. Sampling is periodic and quantization is randomized. Therefore the output signals of such converters are simply digital one-bit signals rather than some special stochastic bit streams. Amazingly, stochastic computing is not stochastic at all. It is actually deterministic rather than randomized. The only operation being randomized is quantization. As it is extremely rough and produces just one-bit digital signals, processing them is indeed simple. Multiplication, for example, can be realized by means of a logic gate. Therefore the generalized randomization model, given later in Figure 2.6, also covers the case of stochastic computing. This once again confirms the assumption that knowledge of basic randomized signal conversion procedures is widely used to analyse special cases.

The most valuable feature of stochastic computers – their simplicity – is not very attractive once the fact is taken into account that this simplicity is due to the one-bit representation of quantized data and that no more threshold levels can be used for such quantization in principle. However, it would be unwise to write off stochastic computers completely. There is still a need for economically attractive computing techniques that permit efficient parallel processing of many input variables simultaneously. With advances in microelectronics, it is now possible to develop specialized VLSI (very large scale integrated) circuits containing a very large number of stochastic computing elements operating at much higher clock

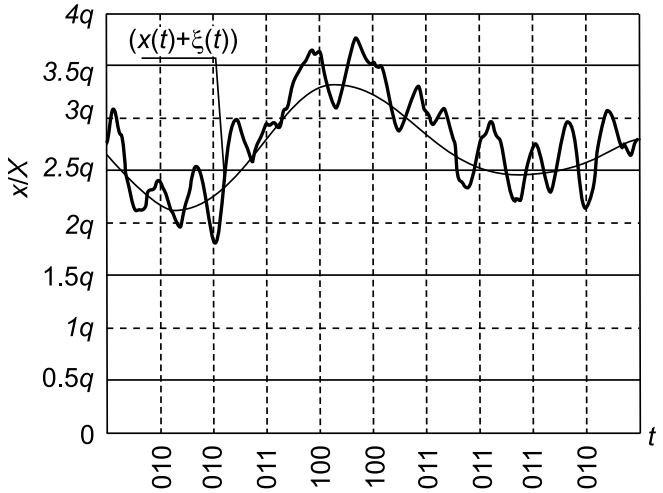


Figure 2.5 Illustration of dithering

frequencies. Stochastic computers built on the basis of a microelectronic element would offer much better precision and bandwidth. It may happen that for some special kinds of applications, e.g. for fault-tolerant computing, for solving large systems of differential equations or for sensor network applications, stochastic computing techniques might become attractive again.

2.2.5 Dithering

Dithering is one of the most popular randomization techniques. This straightforward method of randomizing signal processing is based on the addition of random noise, sometimes called dither, to analog signals to be quantized and processed. It is claimed that under certain conditions this approach provides better accuracy. It is used in a relatively large number of applications, including picture and speech coding and signal analysis. This approach is often used to take out spurious frequencies from spectra of deterministically quantized signals.

It may seem strange that improvement in the metrological characteristics of a system can be achieved by adding noise to its input. However, this is the case. Assume that a slowly varying signal is presented to an ADC and that the quantization involved is uniform and rough, as illustrated in Figure 2.5(a). Note that during some time intervals the signal does not go outside the limits of a single quantization step, i.e. it does not cross any of the threshold levels. As the exact values of the signal are fixed only when it crosses the threshold levels, and at

all other times it is only known that the signal is somewhere within the limits of these two threshold levels, some information is lost.

A method for recovering this information, at least partly, has been developed. This involves adding some auxiliary process to the signal, either periodic or random. The resulting mixture crosses the threshold levels much more often, as shown in Figure 2.5(b), which makes it possible to increase the quantization accuracy by applying short-time averaging or interpolation procedures.

It can be seen that the addition of random noise to the signal results in a marked change in the pattern of the quantizer output signal. Instead of more or less long series of one and the same digital sample value, larger and smaller sample values are obtained. The pattern of the sequence is random, because it is formed according to probabilistic rules. A larger number appears at any given sampling instant with a probability proportional to the closeness of the signal to the nearest upper threshold.

When this approach is used, the output signals of the ADCs are formed in such a way that each digital number at the output is calculated by taking into account some quantity of the quantized samples. Averaging of these samples is usually carried out to achieve that. In this way, refined results of quantization are obtained, which are represented by a sequence of digital numbers that may have fractional parts. These fractional parts can assume any value between 0 and q. However, this digital sequence can be regarded as the output of a more precise quantizer, so the least significant bit of each digital number obtained is equal to a step-size of quantization that is several times smaller. In other words, this approach allows the number of bits per sample to be increased or the signal-to-noise ratio (SNR) to be improved, since the power of quantization noise is directly related to the step-size of the quantizer.

The positive effect just described is obtained as a result of applying two procedures: dithering and averaging. The latter can also be used in cases of conventional quantization without any random noise being added to the signal. It is therefore of some interest to evaluate the effect due to dithering itself. The results of the analysis show that averaging by itself does little to improve the accuracy of quantization, as it might lead to considerable signal-dependent errors. Consequently, it is not of much use to apply this procedure to increase the number of bits per sample value without doing something to prevent these bias errors. Amazingly, this can be accomplished by adding noise to the signal.

Of course, averaging itself does not lead to these signal-dependent errors, which are caused by the mode of quantization used. In addition, the results of fixed-threshold deterministic quantization are more or less distorted by bias errors.

Although they are negligibly small in cases of multibit quantization, these errors might affect the results of rough quantization significantly. As they are systematic or signal-dependent, averaging does not help in reducing them.

Remarkably, adding a random process to the input signal leads to elimination of this kind of error, which takes place for any distribution of the signal even if quantization is extremely rough, i.e. even if only a single quantization threshold level is used. However, dithering also has negative effects. Variance of the random quantisation error increases by twice its minimum value (characterizing deterministic quantization) even if the distribution of the added noise is kept within the optimal limits $[-q/2, q/2]$, where q is the quantization step.

The considered stochastic interpolation technique can be improved. Pseudo-random noise is generated and added to a signal prior to its quantization and the same noise is subtracted from the quantizer output. Such dithering not only eliminates the systematic error or bias but also breaks up the undesirable signal-dependent patterns within quantization error sequences while the variance of the random error remains equal to the value characterizing conventional deterministic quantization.

Much more could be said about dithering, or stochastic interpolation. On the other hand, it is claimed that all known uses of randomness in the area of signal processing that have positive effects can be represented as randomization of sampling and/or quantization. If this is true then stochastic interpolation may be represented in another form. Indeed, that is the case.

As indicated above, stochastic interpolation includes the operations of adding noise to the signal and averaging the quantized samples of the mixture. There can be no doubt that averaging belongs to the domain of processing. The first of these operations is nothing other than randomized quantization. In fact, it is easy to see that adding a noise sample to a signal or subtracting it from a threshold level will produce an equal quantizer reaction.

The two kinds of dithering discussed above are based on two different modes of randomized quantization. With the first, simpler, approach, some type of truly random process is used, which is either added to the signal or subtracted from the threshold levels. Better accuracy results from implementation of the second kind of randomized quantization. In this case pseudo-randomized quantization is used as a pseudo-random rather than a random process. Such pseudo-randomized quantization, studied later in more detail, is equivalent to the dithering model, which is described as the process of adding analog pseudo-random noise to a signal at the input and subtracting the same noise in digital format from the quantized signal.

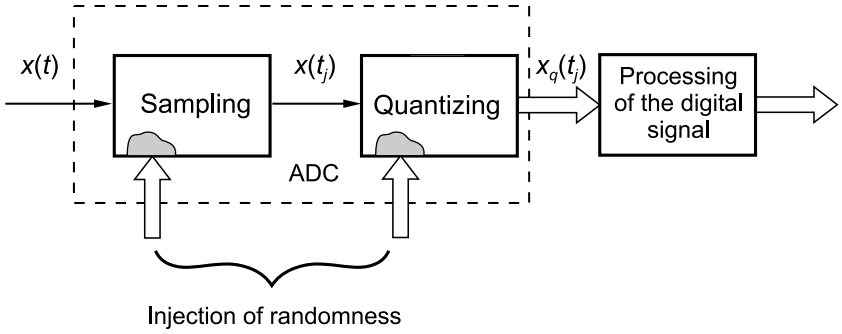


Figure 2.6 Generalized scheme, illustrating randomization of signal processing

Depending on which of these two quantization options is used, averaged quantized signal values are characterized by smaller or larger random errors. The second scheme is much more precise.

2.2.6 Generalized Scheme of Randomized Digitizing

It seems that people working in various remote areas of engineering have discovered the above methods independently. Originally they were described in quite different terms. Reading about these techniques and their implementations leads to the impression that there are many various randomized signal processing methodologies all differing in principle. However, even a simple analysis shows that this impression is wrong. It is true that the techniques of introducing randomness vary, but there are only one or two signal digitization procedures – sampling and quantizing – either or both of which have been used to introduce the element of randomness. As to the processing itself, in all known and considered cases it is performed quite deterministically. Processing is sometimes performed on the basis of classical algorithms, while in other cases specially developed algorithms are used. This generalized approach to deliberate randomization is illustrated in Figure 2.6.

Although it has been stated that randomness can be injected at both the sampling and quantizing stages, the methods described in this section have all been based on deliberately randomized or pseudo-randomized quantizing, while sampling has been assumed to be periodic. There are, of course, successful applications of randomized or nonuniform sampling, considered in some of the following chapters. They have not been discussed here because it was not necessary. When deliberate randomization of signal processing is based on randomization of

sampling techniques, the experience is usually described in terms that clearly fit the model illustrated in Figure 2.6.

This seemingly simple result of the known case analysis actually proved to be quite useful. It has led to the possibility of studying and describing the randomized signal processing methods in a generalized way. This helped to form the DASP approach, based on the concept that attention needs to be focused on the sampling and quantizing procedures and that randomized analog-to-digital conversions should be considered as a vital part of the whole digital signal processing process. Development of the DASP concept and techniques has taken many years and DASP is based on many knowledge components. The briefly described early work in the general area of deliberately randomized signal processing has played a significant role in forming that approach.

2.3 Discovery of Alias-free Signal Processing

It is typical for the early prehistory of randomized signal processing that most of the R&D (research and development) activities in that area were isolated. The applied and academic research was therefore carried out independently at various places in parallel. While engineering efforts during the 1960s and 1970s were devoted, in general, to the creation of new complexity-reduced stochastic signal processing systems, and most of them were based on the idea of deliberate randomization of quantizing, the academic research activities were mainly focused on problems of randomized sampling. The latter, typically, were focused on widening the frequency range where signals could be dealt with digitally.

2.3.1 Early Academic Research in Randomized Temporal Sampling

Shapiro and Silverman were among the first researchers to draw attention in 1960 to the fact that it might be possible to perform alias-free sampling of random noise. Their results were remarkable as they showed the reality of expanding the frequency range where signal processing is not corrupted by aliasing. A very useful randomized sampling approach called additive random sampling was suggested. This random sampling scheme is based on the assumption that successive sampling intervals are statistically independent and identically distributed. Soon other people working in this area picked up this sampling technique; it is still one of the most popular as it proved to be well suited for various applications. This method of nonuniform sampling is discussed in detail in Chapter 6.

It was also discovered that sampling instant jitter, usually considered to be harmful, could sometimes be beneficial. This kind of randomized sampling model is known as periodical sampling with jitter; it is also studied in Chapter 6. As more work was done in the field of randomized sampling, it was realized that many sampling point (sampling instant) processes might be used to perform nonuniform sampling. A relatively large variety of random point processes can, in principle, be applied to random sampling, and a number of them have been investigated to find out whether they eliminate aliasing. It was confirmed that theoretically it is possible to reconstruct signal spectra in a frequency range not limited by the mean sampling rate. The theory of such point processes was gradually developed and similar mathematical problems have also been extensively studied in the area known as renewal theory.

Academic research has done much to advance understanding of the randomized signal processing approach. The results confirmed that it is possible to avoid aliasing if signals are sampled in the correct nonuniform way. For instance, it has been proved that under certain conditions it is possible to estimate digitally signal parameters, including spectrum, in a frequency range exceeding half of the sampling rate many times. That has been a significant achievement. A relatively large number of publications on this subject appeared in the 1960s.

However, these particular results were not usually directly applicable for practical purposes as this research had basically been carried out by mathematicians. For example, the approach most often used to study randomized sampling was based on the Poisson sampling scheme as it is convenient for mathematical analysis. According to this approach, the adjacent sampling instants might appear very close together. Therefore it was impossible to use this sampling model as there are strong limitations imposed on the highest sampling rate or the smallest time interval between sampling instants for all types of ADC.

Thus the early euphoria did not last long. Soon it became clear that while it is indeed possible, for instance, to reconstruct randomly sampled signal spectra in an extremely wide frequency range, extending far beyond the Nyquist limit, the traditionally used DSP algorithms, more often than not, were not applicable and special algorithms had to be developed. Actually the first attempts to use nonuniform sampling techniques, made by people not sufficiently well experienced in this area, usually led to negative rather than positive results. The reason for such an outcome, of course, was lack of insight into the intricacies of nonuniformly sampled signals.

However, the fact remains that the academic research results were significant. They demonstrated both the theoretical possibility as well as the complexity of performing alias-free signal processing at frequencies exceeding the limit of half

the sampling rate. It gradually became clear that the practically applicable nonuniform sampling techniques are complicated and correct application of them has to be learned. Hard work, done step-by-step over the last thirty years by a relatively small number of researcher groups, led to the development of nonuniform sampling theory covering a relatively wide field of applications, including image coding and processing.

Achievements in the area of special deliberately randomized nonuniform sampling drew attention to this field, stimulating increased activities. The first International Workshop on *Sampling Theory and Applications SampTA 95*, dedicated first of all to discussions on nonuniform sampling problems, was organized in Jurmala, Latvia, in 1995. It was well attended by mathematicians and engineers working in that field in many countries of the world. The decision was taken to organize such workshops on a regular basis. Since that time SampTA Workshops have taken place in Portugal (1997), Norway (1999), USA (2001), Austria (2003) and Turkey (2005). The proceedings of all these events contain interesting and useful material showing what has been achieved and how the work has developed in different parts of the world.

2.3.2 Early Research in Randomized Spatial Signal Processing

Although historically the problems of randomizing temporal signal processing have received much more attention than the analogous problems of spatial filtering of signals, the potential usefulness of deliberate randomization for this kind of signal processing has not been overlooked. A number of publications provide evidence of this. However, at first glance, they have little in common with randomized processing of signals. The fact that a number of radar characteristics could be enhanced by randomization of their antenna design had already been discovered in the early 1960s. Specifically, it had been shown that nonuniform element spacing in arrays could lead to side-lobe reduction. It was found that when antenna elements are spaced nonuniformly rather than regularly undesirable spatial aliasing effects are reduced and that other improvements in system functioning can be achieved.

The subject of deliberate randomization of array signal processing is considered in Chapter 17. As shown there, various methods developed in the area of nonuniform temporal sampling digital processing of signals encoded in this way could also be successfully used for processing of nonuniform array signals.

It should also be noted that randomized spatial filtering of signals is one more example confirming that all known cases of deliberate randomization of signal processing are covered by the generalized scheme given in Figure 2.6. The

behaviour of the arrays of nonuniformly spaced elements can be predicted by applying the model of random sampling and the research results obtained in this area.

2.3.3 Engineering Experience

The randomized measuring and computing methods mentioned and described in Section 2.2 should not be regarded as the ultimate achievement in the field of randomized signal processing. The large amount of effort that has been put into theoretical studies of the randomization problems has resulted in the development of much more sophisticated methods, which are discussed in the following chapters. To some extent the hard work in this direction has been encouraged by successful engineering efforts and by the fact that even relatively simple randomization schemes have often proved to be useful and sometimes even considerably beneficial. This has been demonstrated repeatedly by results of successful engineering attempts to use this approach for development of high-performance electronic instruments.

Fortunately, it has been possible to realize the deliberate randomization approach in various ways in a wide complexity range. The simplest randomization techniques were implemented on an industrial scale first and further industrial success stories were based on applications of those techniques.

Hewlett-Packard stands out as a company that has achieved most spectacular results in this field. In a number of high-quality instruments developed and produced by this company, the problems of expanding the input frequency range and improving resolution and accuracy have been solved in a very simple way by using the deliberate randomization approach. This company has been one of a few that has had enough knowledge of, and confidence in, this approach to apply it over and over again with remarkable results.

They confirmed that randomization of one or more signal digitizing operations is a convenient tool for achieving some valuable practical effects. The advantages of random sampling are convincingly demonstrated by the design and performance of HP 3406, a sampling voltmeter. By applying this mode of sampling, it has become possible to broaden the signal bandwidth up to 1.2 GHz and at the same time to simplify the instrument design. Developed in the 1960s, this instrument has been successfully marketed for many years.

Implementation of randomized periodic sampling has led to performance improvements of the high-speed digitizing oscilloscope HP 54100 A/D. Resolution of 10 ps and 1 GHz bandwidth in the repetitive mode have been achieved while the mean sampling rate was only 40 MS/s. The sampling principle employed not only permitted these improvements but also led to some additional attractive

functional enhancements, like the ability to record short-time pulses repeating randomly or periodically with a very low repetition frequency. There are also cases of successful application of stochastic quantizing for highly accurate digital measurements of time intervals. Instruments of this kind are again characterized by subnanosecond resolution and very small measurement errors.

Randomized quantization has also been implemented with good results in a 1 GS/s (gigasamples per second) ADC designed for the high-speed digitizing oscilloscope model (the oscilloscope HP 54111 D). This kind of quantizing plus averaging is applied to improve measurement accuracy. In this way, the initial 6-bit ADC code is augmented with two additional bits so that the signal is represented by 8-bit sample values. Randomized quantizing (called dithering) was found to be very useful for improving dynamic performance of high-quality ADCs. Specifically, the spurious-free dynamic range (SFDR) of the 12-bit converter AD9042, designed and produced by Analog Devices, was improved at least for 25 dB using this approach.

The above examples of randomized signal processing are mentioned not only to show that this approach is already being used in industrial applications – a significant fact in itself – but also as a benchmark with which to compare the methods and applications described further. It was typical of early efforts in this area that the randomization of signal parameter measurement tasks was seen as a means of substituting, where possible, complicated deterministic relationships by much simpler statistical ones in order to be able to build simple electronic devices. At that time, when electronic instruments and systems were designed around vacuum tubes, and later transistors, this consideration was indeed important. It was popularly believed that an increase in statistical errors in these cases was inevitable and that this loss of precision was a small price to pay for the much simpler circuits obtained as a result of randomization.

Now the situation has changed. It is no longer so important to simplify electronic devices, decidedly not at the expense of substantially increased error. Use of randomized signal processing is worth considering only if it leads to achieving more, not less, precise signal processing. Current engineering activities, targeting development of alias-free innovative DASP systems and exploiting the advantages of nonuniform sampling and other randomized techniques for signal processing, are based on this concept.

2.4 Randomization Leading to DASP

A significant but previously not recognized fact was revealed by Bilinskis and Mikelsons about 15 years ago. It was shown then that only the signal digitizing operations rather than the subsequent processing procedures are always

randomized. The techniques for processing digital signals obtained as a result of randomized sampling and/or quantizing operations have, contrary to claims, always been fully deterministic. This generalization disclosed that in fact in all of the known and considered randomized signal processing cases randomizing is always based on using randomized modifications of analog-to-digital conversions. However, truly randomized operations of digitizing are used much less often than pseudo-randomized ones. In the latter cases these procedures are fully deterministic, the term 'randomized' sampling and/or quantizing being a generalized one covering a broad class of nonuniform sampling and irregularized quantizing operations.

From the discussions above, it should be clear that the achievable quality of digital handling of signals to a large extent depends on the perfection of the involved operations of digitizing and on the selection of correct and suitable sampling and quantizing options. Recognition of this fact focused attention on the necessity to improve the traditionally used DSP techniques by developing advanced techniques for signal digitization. That in turn led to the randomization of them, to the randomization of digitizing and to the development of algorithms matched to the specifics of processing signals encoded in this special way. These modifications were so productive that they gradually led to the development of a signal processing technology significantly differing from the classic DSP. Therefore it makes sense to consider it as a special case. This innovative technology, the topic of this book, is called 'digital alias-free signal processing'.

This is the first book dedicated to DASP technology and there are only a few previously published books related to this subject. However, the knowledge already accumulated in this field, including theoretical and engineering experience, is substantial, sufficient for the development of various competitive systems with properties unparalleled by the classical DSP systems.

Fortunately, the microelectronic elements, manufactured for the needs of the traditional DSP, can also be used for engineering implementation of these pseudo-randomized DASP methods. Therefore it is possible to gain from that knowledge immediately. Although the development of special DASP oriented chips, of course, would provide even better results, application of this signal processing technology makes it possible to use the currently produced ADC in a frequency range limited by the bandwidth of their analog inputs rather than by the maximal sampling rate. This approach typically widens that frequency range approximately 4–8 times. This means that application of DASP in many cases provides much better exploitation of ADC resources. Of course, there are also application limitations and drawbacks of DASP that have to be taken into account.

2.4.1 DASP Mission

The mission of the DASP technology, in a nutshell, is to provide effective methods, hardware and software tools for flexible, adaptable digitization and matched processing of signals digitized in this nontraditional way for widening the digital domain of signal processing over the areas where analog techniques still dominate.

Besides anti-aliasing measures, there are many other system performance improvement problems that the right sort of randomized signal digitizing techniques could help to resolve. Massive data acquisition from distributed signal sources, mentioned above and discussed in Chapter 7, is a good example illustrating this point. Specific sine wave crossing techniques are used for signal sampling in order to achieve a number of essential advantages, making it technically and economically feasible to gather data in a cost-effective way from a large number of signal sources. Although these randomized sampling techniques are nonuniform, they are not used for avoiding aliasing. In this case aliasing is not the central issue as the signals involved more often than not are of low frequency. Nevertheless, it makes sense to use the specific nonuniform digitization techniques in this case as well. This makes it possible to achieve other benefits, such as complexity reduction of the system gathering data from a very large quantity of sources and data compression. Similarly, there are some techniques typical for processing nonuniformly sampled signals that are useful for resolving specific problems of processing signals in the low-frequency range under specific conditions. Application of nonorthogonal transforms allowing signals with partly overlapping spectra to be separated or errors to be eliminated, which appear when the signal observation time is not equal to an integer number of signal periods, might be mentioned as additional examples of such specific problems.

Thus, as shown in the following chapters, the application range of DASP is rather wide and these applications are not exclusively in the field of radio frequency (RF) and microwave signal processing, as already explained. Many of the specific randomized signal digitizing and special digital processing methods and tools, initially developed for digitizing wideband signals, later prove to be quite useful for processing signals under certain conditions, even when their parameters also allow the use of traditional techniques.

2.4.2 Demonstrator of DASP Advantages and Limitations

Various systems have been developed and made that are capable of processing signals fully digitally in a frequency range many times exceeding the mean

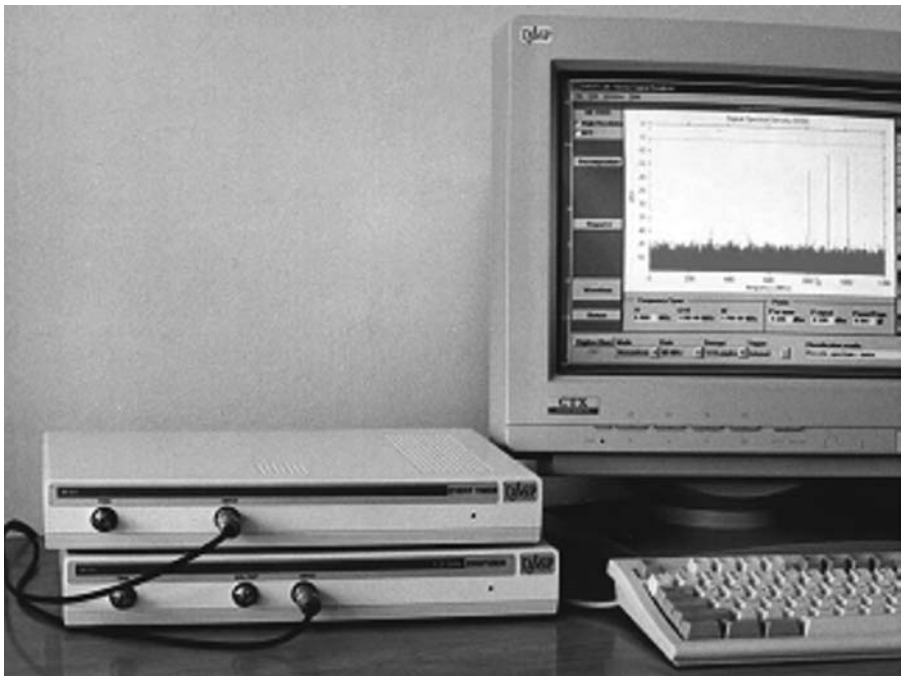


Figure 2.7 DASP-Lab System design

sampling rate. The best-known one is a computer-based system of Virtual Instruments called the DASP-Lab System. As a versatile analyser of wideband RF signals, it serves as a demonstrator showing the advantages and limitations of the DASP technology, including such an advantage as the capability to perform a fully digital radio frequency signal analysis in the time and frequency domains. This system operates in the whole frequency range up to 1.2 GHz while the mean rate of nonuniform sampling is only 80 MS/s. The system won the 1997 European IT Prize and is shown in Figure 2.7. There are two optional digitizers, one for the signal analysis in the time and frequency domains and the other for the signal analysis in the modulation domain.

The 1.2 GHz digitizer represents the hardware core of the DASP-Lab System. This device is a standalone unit connected to the host computer of the DASP-Lab System, typically a PC. It performs an input signal analog-to-digital conversion in a special way matched to the specific application conditions. The sampling operation can be chosen to be either periodic or nonuniform. The latter sampling mode is performed at predetermined time instants.

While the digitizer performs signal digitizing and data buffering functions, the computer supports them and executes data exchange via a standard interface. To obtain good signal alias-free processing results, special algorithms that are well matched to the specifics of nonuniform sampling must be and are used. Computer simulated graphical user interfaces are used to set up the digitizer and the system's operating conditions for interactive measurements, for digitized and processed signal visualization, for building up complex test signals and for performing other virtual instrumentation and virtual protyping functions.

The DASP-Lab System has three basic operational modes. It can be used for virtual instrumentation, for virtual prototyping and for signal input to the MATLAB environment. Virtual instruments are also the basis for virtual prototyping of DSP systems utilizing nonuniform sampling. In the third operational mode signals are sampled periodically and their digital replicas can be processed by the standard MATLAB DSP tools. In the first and second cases signal spectra can span the whole bandwidth of 1.2 GHz, but the MATLAB DSP tools can only be used for handling periodically sampled signals.

The virtual instruments of the system cover the following functions:

- *Digital oscilloscope*. This displays waveforms of periodic or repeated (repeatedly triggered) signals. The effective sampling rate is around 3.2 GS/s and it performs selectable processing routines and external and internal synchronization.
- *Vector spectral analyser*. This estimates and provides signal spectrum density functions, amplitude and phase spectra of signal parts having discrete spectra in the whole bandwidth of 1.2 GHz. It performs signal identification, determines signal true component spectra at arbitrary frequencies and performs high-resolution or discrete Fourier transform (DFT) spectrum analysis, signal waveform and envelope reconstruction on the basis of direct and inverse spectral transforms.
- *Power analyser*. This performs SNR measurements within four independently variable frequency intervals and estimates the signal average or peak power at arbitrary frequencies.
- *Modulation domain analyser*. This fulfils its functions on the basis of continuous time interval measurements. It performs up to 25 million such measurements per second with 20 ps single-shot least significant difference (LSD) (5 ps for up to 4 million measurements/s), clock and data jitter analysis including jitter spectrum analysis and frequency and phase modulation analysis. It is useful for resolution of many problems including jitter source identification.

Virtual prototyping helps in the development of various nonuniform sampling-based DSP systems. In this operational mode, the DASP-Lab System can be used for design verification, alias-free RF digital design prototyping, complex test-signal forming and system prototype testing with virtual instruments. In addition, virtual prototyping functions of the DASP-Lab System are helpful in acquiring DASP-based system development skills.

2.5 Some of the Typically Targeted Benefits

To illustrate the usefulness of the DASP technology, extension of the application range of ADC achievable in many cases will be considered on the basis of this technology.

As long as the traditional DSP approach is used, the highest sampling rates of the manufactured ADC define their application boundary in the frequency domain. It is possible to realize processing of a particular signal digitally only if there is an ADC available that could be used at a high enough sampling frequency. Thus the achieved highest sampling frequency of a specific type of ADC might be considered as a benchmark characterizing its application range. Much effort and money has been spent on advancing the technology for manufacturing higher-quality microelectronic devices, including faster, more precise and wider bandwidth analog-to-digital converters. As a result, better and better ADC chips, applicable for a wider and wider frequency range, are produced and offered to the market on time. Consequently, the application area of DSP techniques is continuously widening. However, progress in this field made in this way is evolutionary, relatively slow and very costly. Using DASP leads to the required results in a considerably less expensive way in a wide range of applications. In this case widening the ADC application range is achieved by using nontraditional signal digital processing methods and techniques rather than by improving the semiconductor manufacturing technologies.

The upper frequency of a signal that could be processed digitally by an ADC in a traditional way based on DSP depends on two relevant parameters: the highest sampling rate and the bandwidth of the analog input. That bandwidth then has to be equal to at least half of the highest sampling rate. However, in many cases it is much wider. Typically it is often 4 to 8 times wider than the corresponding half of the sampling rate. This valuable resource of ADC is often not exploited in classical DSP applications.

The introduction of DASP technology into system designs changes the situation drastically. Then appropriate nonuniform sampling and matched digital signal processing techniques can be used to eliminate aliasing and, consequently, the

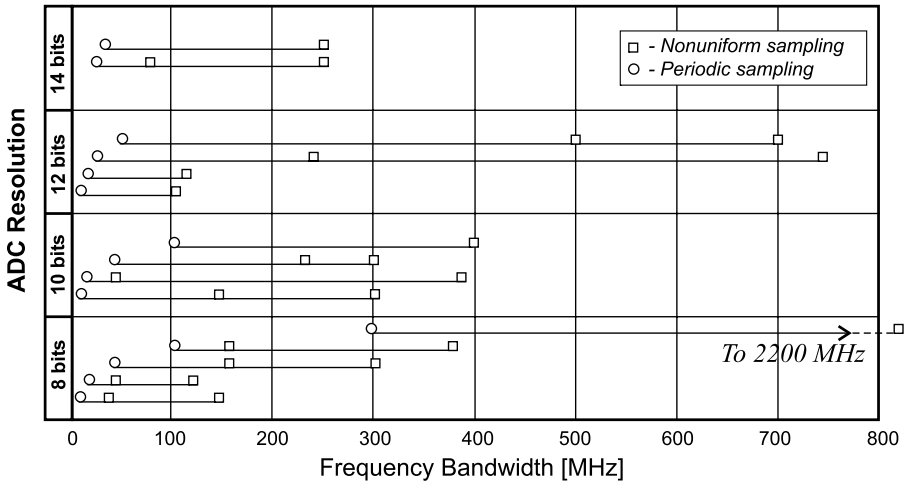


Figure 2.8 Using DASP technology leads to significant widening of the operational frequency range for the existing ADC

needed sampling rate does not directly depend on the upper frequency of the signal. Only the other mentioned ADC parameter, namely the bandwidth of the input, restricts this upper frequency. However, as in most cases this bandwidth is much wider for a given ADC than half of the permitted sampling rate, using the DASP technology typically increases the frequency limit to a level that is several times higher.

Figure 2.8 illustrates this effect. This diagram, plotted on the basis of typical ADC data taken from manufacturers' catalogues, shows how much the application range of the existing ADC could be widened in the frequency domain by paying attention to the signal digitization processes and by changing the signal processing techniques. It can be seen that using the DASP approach extends the upper frequency limits of the application ranges of the considered ADC to frequencies that are several times higher.

Note that two kinds of benefits could be obtained as a result of this effect. In the lower frequency range, covered by many types of ADC, application of DASP typically leads to cost savings as relatively low-priced converters can often be used in this frequency range instead of more expensive devices. On the other hand, only a few types of ADC could be used in the customary way in the indicated region of the higher frequencies. In this frequency range, application of DASP results in significant enlargement of the frequency range where signals can be processed fully digitally. However, the most important benefit is the possibility

of using existing ADC chips for signal conversions in a much larger frequency range that is beyond the scope for alternative DSP techniques. For example, as shown below, application of DASP makes it possible to use a 12-bit ADC with a highest sampling rate of 125 MS/s in a wide frequency range up to 700 MHz.

It could be argued that replacement of the traditional DSP techniques by those based on the DASP methods and algorithms is not always possible. That is true as there are limitations. Although they are not directly tied to the sampling rate used and are different, the limitations certainly do exist. Nevertheless, there is a really wide application field where DASP can be applied beneficially.

Bibliography

Balakrishnan, A.V. (1962) On the problem of time-jitter in sampling. *IRE Trans. Inf. Theory*, **IT-8**(4), 226–36.

Beutler, F.J. (1970) Alias-free randomly timed sampling of stochastic processes. *IEEE Trans. Inf. Theory*, **IT-16**(2), 147–52.

Beutler, F.J. and Leneman, O.A. (1966) Random sampling of random processes: stationary point processes. *Inf. Control*, **9**(4), 325–746.

Beutler, F.J. and Leneman, O.A. (1966) The theory of stationary point processes. *Acta Math.*, **116**, 159–97.

Bilinskis, I. and Mikelsons, A. (1992) *Randomized Signal Processing*. Prentice-Hall International (UK) Ltd.

Brannon, B. Overcoming converter nonlinearities with dither. *Analog Devices*, **AN-410**, Application Note.

Brown, W.M. (1963) Sampling with random jitter. *SIAM J. Appl. Math.*, **11**, 460–73.

Buslenko, N.P., *et al.* (1962) *StatisticaI Test Methods (Monte Carlo Method*) (in Russian). Moscow: Fizmatgiz.

Chu, D.C. (1974) Time interval averaging: theory problems and solutions. *Hewlett-Packard J.*, **June**, 12–15.

Corcoran, J.J., Poulton, K. and Knudsen, K.L. (1988) A one-gigasample-per-second analog-to-digital converter. *Hewlett-Packard J.*, **June**, 59–66.

Demas, G.D., Barkana, A. and Cook, G. (1973) Experimental verification on the improvement of resolution when applying perturbation theory to a quantizer. *IEEE Trans. Industr. Electron. Control Instr.*, **IECI-20**(4), 236–9.

Gaines, B.R. (1967) Stochastic computer thrives on noise. *Electronics*, **14**, 72–9.

Halliwell, J. (1970) Stochastic computer solves more problems faster. *Electron. Eng.*, **42**(512), 63–5.

Harrington, R.F. (1961) Sidelobe reduction by nonuniform element spacing. *IRE Trans. Antennas Propag.*, **AP-9**(2), 187–92.

Hewlett-Packard Catalog (1985) Measurement, computation systems, p. 171.

Hirsh, J.J. (1972) Contribution à l'étude des conversions numerique-analogique et analogique-numerique au moyen d'une représentation stochastique de l'Information. Thèse.

Jayant, N. and Rabiner, L. (1972) The application of dither to the quantization of speech signals. *Bell Syst. Tech. J.*, **51**(6), 1293–304.

Karyakin, A.I. (1974) Sign estimation of the spectral density of random processes. In *Information and Measuring Devices in Radioelectronics*. Zinatne: Riga, pp. 137–9.

Karyakin, A.I. (1976) Measuring the spectral density of random signals using the relay method with additional signals. In (in Russian) 9th All-Union Symposiu, Vol. 5, Leningrad, pp. 62–6.

Kraus, G. (1972) On the precursors of stochastic-ergodic measuring techniques. *NORMA Tech. Inf.*, **1**, 10.

Leneman, O.A.Z. (1966) On error bounds for jittered sampling. *IEEE Trans. Autom. Control*, **AC-11**(1), 150.

Lo, Y.T. (1964) Mathematical theory of antenna arrays with randomly spaced elements. *IEEE Trans. Antennas Propag.*, **AP-12**, 257–68.

Lo, Y.T. and Simcoe, R.J. (1967) An experiment on antenna arrays with randomly spaced elements. *IEEE Trans. Antennas Propag.*, **AP-15**, 231–5.

Marvasti, F. (2001) *Nonuniform Sampling, Theory and Practice*. New York: Kluwer Academic/Plenum Publishers.

Masry, E. (1971) Random sampling and reconstruction of spectra. *Inf. Control*, **19**(4), 275–88.

Masry, E. (1978) Alias-free sampling: an alternative conceptualization and its applications. *IEEE Trans. Inf. Theory*, **IT-24**(3), 317–24.

Masry, E. (1978) Spectral estimation of continuous-time processes: performance comparison between periodic and Poisson sampling schemes. *IEEE Trans. Autom. Control*, **AC-23**(4), 679–85.

Metropolis, N. and Ulam, S.S. (1949) The Monte-Carlo method. *Amer. Statist. Assoc.*, **44**(247), 335–41.

Millard, J.K. (1988) A one-gigasample-per-second digitizing oscilloscope. *Hewlett-Packard J.*, **June**, 58–9.

Montiljo, B.A. (1988) Digital filtering in a high-speed digitizing oscilloscope. *Hewlett-Packard J.*, **June**, 70–5.

Pichler, H. and Wehrmann, W. (1972) Die neue Art des Messens: stochastisch-ergodische Messenelektronik. *Radio-Elektronik Schau*, **1**, 19–23.

Poppelbaum, W.J. (1968) What next in computer technology? *Advances in Computers*, **9**, 1–21.

Porter, A. (1984) Sampling sees skinny signals. *Electronics*, **2**, 41–4.

Ribeiro, S.T. (1967) Random-pulse machines. *IEEE Trans. Electron. Comput.*, **EC-16**(3), 261–7.

Roberts, L.G. (1962) Picture coding using pseudo-random noise. *IRE Trans. Inf. Theory*, **IT-8**, 145–54.

Shapiro, H.S. and Silverman, R.A. (1960) Alias-free sampling of random noise. *SIAM J. Appl. Math.*, **8**(2), 245–8.

Sobol, I.M. (1973) *Numerical Monte-Carlo Methods* (in Russian). Moscow: Nauka.

Tumfart, S. (1975) New instruments use probabilistic principles. *Electronics*, **15**, 87–91.

Urbat, R. (1973) Stochastisch-ergodischee Mestechnik-SEM. *Elektronik-Industrie*, **4**(6), 133–4.

Urbat, R. (1973) Einfacher und besser Messen mit SEM. *Funkschau*, **25**, 965–6.

Veltmann, B.P. and Kwakernaak, H. (1961) Theorie und Technik der Polaritis-Korrelation fer die dynamische Analyse nierderfreguenter Signale und Systeme. *Regelungstechnik*, **9**, 11–18.

Veselova, G.P. (1975) On amplitude quantization with overlapping of interpolating signals (in Russian). *Avtomatika i Telemekhanika*, **5**, 52–9.

Wehrmann, W. (1971) A new line of stochastic-ergodic measuring instruments. *NORMA Tech. Inf.*, **2**, 3–12.

Wehrmann, W. (1972) The U-functionmeter – a stochastic multimeter for signal processing and data handling. *NORMA Tech. Inf.*, **1**, 11–18.

Wehrmann, W. (1972) Der stochastish-ergodische Messanalysator probability-meter. *NORMA Tech. Inf.*, **1**, 18–24.

Wehrmann, W. (1973) Stochastisch-ergodische Mesverfahren. *Elektrotechn. Masch.*, **89**(7), 287–93.

Wehrmann, W. (1973) Stochastisch-ergodische Mesmethoden. *Elektronik*, **9**, 307–10.

Wehrmann, W. (1974) Uber den Wedegang der SEM-Technik. *Elektronik*, **5**, 180–1.

Yakovlev, V.V. and Fyodorov, R.F. (1974) *Stochastic Computers* (in Russian). Leningrad: Mashinostroyenie.

Yegorov, N.I. (1968) Expanding the area of direct application of the polar correlation technique (in Russian). *Autom. Control Comput. Sci.*, **1**, 54–9.

Yermakov, S.M. (1975) *Monte Carlo Method and Related Issues* (in Russian). Moscow: Nauka.

3

Periodic Versus Randomized Sampling

Signal sampling, in general, is a well-investigated and described process. However, there are also some essential issues of sampling that are not usually given much attention. The dependence of aliasing on conditions of signal sample value taking and on the specifics of signal processing might be mentioned as examples. That is understandable. Indeed, in the case of traditional periodic sampling, the aliasing effect is not acceptable at all. Sampling then has to be performed in such a way that the sampled signals are not distorted by aliasing. Under these conditions, studying the impact of various sampling and processing conditions on the aliasing effect does not make sense. The situation is completely different when randomization of sampling is considered as a means of making the application of fully digital signal processing possible in a much wider frequency range. The processes accompanying aliasing need to be understood really well as the impact of various sampling conditions on frequency overlapping plays a very important role. To randomize signal sampling properly, it is essential to know how variations of the periodic sampling conditions, including variations of the periodic sampling phase, affect the characteristics of the obtained sampled signals.

Some issues of sampling, essential both for sampling and processing of sampled signals, are considered in this chapter. Variable phase periodic sampling processes are discussed, in addition, in Chapter 8.

3.1 Periodic Sampling as a Particular Sampling Case

Information carried by an analog signal $x(t)$ can also be represented in a digital form as a sequence of its instantaneous values $x(t_k)$ measured at time instants

Digital Alias-free Signal Processing I. Bilinskis

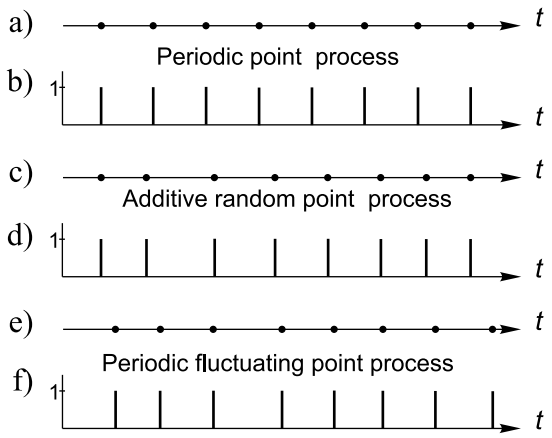


Figure 3.1 Examples of various sampling point processes and their graphical depiction

t_k, $k = 0, 1, 2, \ldots$. These signal readings obtained at discrete instants are usually considered as signal sample values and the process of taking them is referred to as sampling. The instants at which the samples are obtained form a stream of uniform events, which can be depicted graphically as a sampling point process. A few such point processes are shown in Figure 3.1. As demonstrated later, the properties of the sampled signals depend to a considerable extent on the patterns of the point processes generated and used for sampling.

When sampling is mentioned, it is usually assumed that the sampling process considered is deterministic and periodic. The model of equidistant sampling, according to which signal samples are separated by equal-length time intervals T, has been extensively studied, is now used almost exclusively and is actually considered as unique. This is readily comprehensible because such a sampling approach appears to be the most natural and obvious. It also has a number of attractive advantages. However, periodic sampling, in reality, is just one among many other possible sampling models.

It was established a relatively long time ago that application of periodic sampling alone is not sufficient. The periodic sampling model is not applicable when fluctuations in sampling instants cannot be ignored or when signal samples can be obtained only at irregular or even random time intervals. In addition, studies have revealed that randomness in sampling is not always harmful. It was discovered that random irregularities in the sampling process sometimes might even be beneficial. If properly introduced and exploited, these irregularities provide various useful effects. Basically they help to preserve the structure of the original signals, making it possible, for instance, to estimate frequencies of signal components even when these frequencies exceed half of the sampling rate.

Moreover, the sampling processes in reality are always more or less randomized. Although the randomness present at the signal sample taking process might often be negligible, it still exists. A generalized sampling model describing the sampling process analytically in a way that takes into account this sampling feature will now be considered.

3.1.1 Generalized Sampling Model

Mathematical descriptions of sampling and sampled signals $x(t_k)$ are often based on the set of Dirac distributions:

$$u(t) = \sum_{k=-\infty}^{\infty} \delta(t - kT). \tag{3.1}$$

Function (3.1) can only be applied to periodic sampling. The sampling instants, in this case, are separated on the time axis by sampling intervals T. To make it also applicable for the analysis of irregular sampling, this function is modified as follows:

$$u(t) = \sum_{n=-\infty}^{\infty} \delta(t - t_k), \tag{3.2}$$

where $\delta(t - t_k)$ is the delta function. Then a sampled signal can be given as

$$x(t_k) = x(t)u(t) \tag{3.3}$$

A sampling process may also be considered as a sequence of events taking place at some time instants t_k. Graphically this process can be depicted as a stream of points or, in other words, as a point process. There are various sampling point processes with significantly differing features. Three typical sampling point processes that are most often observed are considered. The first periodic sampling is defined as

$$t_k = kT, \quad k = 0, 1, 2, \ldots. \tag{3.4}$$

although this sampling model is the one most often used, it is evidently just a theoretical abstraction. As it is extremely simple and often very close to reality, application of this completely deterministic sampling model is fully justified. However, the sampling could never be performed in a way strictly corresponding to (3.4). The sampling events actually take place at time instants that are more or less distanced from their expected locations on the time axis. In other words, the sampling instants always fluctuate or jitter. To take this phenomenon into account, another sampling model is used. It is obtained by rewriting (3.4) as follows:

$$t_k = kT + \tau_k, \quad k = 0, 1, 2, \ldots, \tag{3.5}$$

where T is the duration of a constant time interval and τ_k is a realization of a random variable τ. Relationship (3.5) is applicable more widely than just as a description of jittering periodic sampling. It also defines a significant case of randomized sampling, including deliberate randomization of the kind discussed in detail in the following chapters.

There is yet another essential sampling model, given as

$$t_k = t_{k-1} + \tau_k, \quad k = 0, 1, 2, \ldots. \tag{3.6}$$

Equation (3.6) describes a process known as the additive random point process. Shapiro and Silverman proposed it as a convenient tool to describe randomized sampling. This point process has a number of properties valuable for practical applications, discussed in detail in Chapter 6. When the random point process (3.6) is applied for sampling, the variances of the point locations on the time axis add up, so that after some time the probability of seeing a sampling instant when looking at a signal through a narrow window will always be the same. In other words, in this case the average density of sampling points along the time axis is constant. This effect cannot be generally achieved if sampling is performed on the basis of the random point process (3.5), which is really periodic with the points fluctuating near their expected locations. As this example shows, randomization of sampling should be implemented properly so that some probabilistic requirements are met; otherwise the desired positive effects will not be achieved.

Expression (3.6) also demonstrates a notable point showing that the widely used periodic sampling is actually a particular case of randomized sampling. Note that at $\sigma/\mu = 0$ (μ and τ are the mean value and the standard deviation of the random variable τ respectively) sampling is periodic and the time intervals between the points are constant and equal to T (in this case $\mu = T$). By varying this ratio, it is possible to cover the complete range from deterministic (periodic) to extremely randomized sampling. The distribution of τ is shown in Chapter 6 to be of little importance, so this sampling scheme is applicable to a wide variety of sampling interval distributions.

The randomness introduced into the sampling process in both mentioned cases can be controlled by one parameter, the ratio σ/μ. This parameter has to be set according to the conditions of the signal processing to be performed. Seemingly logical rules for choosing the value of this parameter were defined a long time ago and have been used for many years. It was considered that, on the one hand, the randomness deliberately introduced at sampling has to be significant enough to achieve the effect targeted (e.g. the suppression of aliases), but, on the other hand, that randomness has to be minimized in order to lessen the statistical error. Therefore the value of the parameter was typically set as small as possible. The

value of this parameter might be varied in a wide range depending on specific application conditions. While it is often kept within the range of 0.01–0.1, it might often be much larger. For the random point processes shown in Figures 3.1(d) and (e) the value of σ/μ is equal to 0.2. Note that even when such a relatively large degree of randomness is introduced into the sampling point processes, its presence is not obvious.

These basic recommendations seem to be straightforward and easy to comprehend. However, recently it has been discovered that much better results may be achieved if a completely different approach to generation of the most effective randomized sampling point processes is used. Therefore these recommendations, given in many early publications, need to be taken critically. The generation of various effective sampling point processes is discussed in Chapter 10.

Substitution of expression (3.6) into Equation (3.3) gives

$$x(t_k) = x(t) \sum_{k=0}^{\infty} \delta[t - (t_{k-1} + \tau_k)]. \tag{3.7}$$

This expression describes the generalized sampling model. It covers both the randomized and periodic sampling processes and is controlled by only one parameter. Even a small variation of the ratio σ/μ would lead to substantial changes in characteristics of this sampling model. It should be emphasized that the discussed generalized model of sampling is signal independent. Deterministic as well as random signals might be sampled in this way.

Equidistant sampling, although very popular, is really only a particular mode of sampling. The achievable quality of a sampling operation directly depends on the sampling point process used. Much better results have been achieved lately by developing and using more elaborate sampling techniques. However, to develop good randomized sampling schemes, it is worth carrying out additional studies of the periodic sampling techniques, paying special attention to the crucial aliasing effect and related issues.

3.2 Spectra of Sampled Signals

An attempt will now be made to find what can be expected from randomization of the sampling operation by establishing how the spectral characteristics of the sampled signals depend on the specifics of the sampling point processes used. The spectrum of a sampled signal

$$S_s(f_n) = \lim_{N \Rightarrow \infty} \frac{2}{N} \sum_{k=1}^{N} x(t_k) \exp(-\mathrm{j}2\pi f_n t_k), \tag{3.8}$$

where N is the number of samples processed. By introducing function (3.2) and noting that

$$\int_{-\infty}^{\infty} x(t)\delta(t - t_k)\,\mathrm{d}t = x(t_k),$$

it is found from Equation (3.8) that

$$S_s(f_n) = \lim_{N\Rightarrow\infty} \frac{2}{N} \int_0^{\Theta} x(t)u(t)\exp(-\mathrm{j}2\pi f_n t)\,\mathrm{d}t, \tag{3.9}$$

where $\Theta = E[t_N]$ and $f_n = n/\Theta$.

3.2.1 Spectra of Periodically Sampled Signals

When sampling is performed periodically and the sampling frequency $f_\mathrm{s} = 1/T$, the function $u(t)$ is also periodic and its Fourier series expansion

$$\begin{aligned} u(t) &= \frac{1}{T} + \frac{2}{T}\sum_{r=1}^{\infty} \cos\frac{2\pi r}{R}t \\ &= \frac{1}{T} + \frac{1}{T}\sum_{r=1}^{\infty}\left[\exp\left(\frac{-\mathrm{j}2\pi r t}{T}\right) + \exp\left(\frac{\mathrm{j}2\pi r t}{T}\right)\right]. \end{aligned} \tag{3.10}$$

Substitution of Equation (3.10) into Equation (3.9) yields

$$\begin{aligned} S_n(f_n) &= \lim_{\Theta\Rightarrow\infty} \frac{2}{\Theta}\int_0^{\Theta} x(t)\exp(-\mathrm{j}2\pi f_n t) \\ &\quad\times\left\{1 + \sum_{r=1}^{\infty}\left[\exp\left(-\mathrm{j}2\pi\frac{rt}{T}\right) + \exp\left(\mathrm{j}2\pi\frac{rt}{T}\right)\right]\right\}\mathrm{d}t \\ &= \lim\frac{2}{\Theta}\int_0^{\Theta} x(t)\exp(-\mathrm{j}2\pi f_n t)\,\mathrm{d}t + \sum_{r=1}^{\infty}\lim_{\Theta\Rightarrow\infty}\frac{2}{\Theta}\int_0^{\Theta} x(t) \\ &\quad\times\left\{\exp\left[-\mathrm{j}2\pi\left(f_n + \frac{r}{T}\right)t + \exp\left[-\mathrm{j}2\pi\left(f_n - \frac{r}{T}\right)\right]t\right\}\mathrm{d}t\right. \\ &= S_x(f_n) + \sum_{r=1}^{\infty}\left[S_x\left(\frac{r}{T} + f_n\right) + S_x\left(f_n - \frac{r}{T}\right)\right] \\ &= S_x(f_n) + \sum_{r=1}^{\infty}\left[S_x\left(\frac{r}{T} + f_n\right) + S_x^*\left(\frac{r}{T} - f_n\right)\right], \end{aligned} \tag{3.11}$$

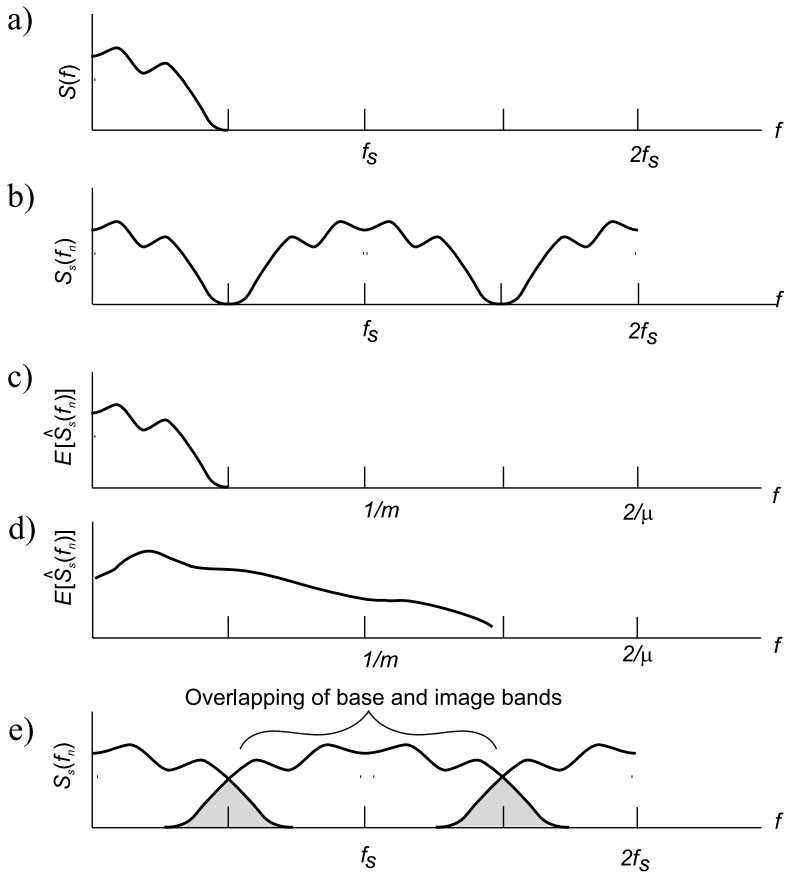


Figure 3.2 Spectra of a signal before and after periodic and random sampling: (a) spectrum of the analog signal; (b) spectrum of the periodically sampled signal; (c), (d) expectations of the spectrum estimates when the signal is sampled randomly; (e) overlapping base and image bands

where $S_x(f)$ is the spectrum of the respective original signal and $S_x^*(f)$ is the complex of the conjugate of $S_x(f)$.

Figures 3.2(a) and (b) illustrate the result. It is well known, but is not always interpreted correctly. It can be seen that the spectra of the periodically sampled signals are also periodic in the frequency domain and the sampling process is responsible for that. However, it would be wrong to conclude that this periodicity of sampled signal spectra is unavoidable. It is not the sampling itself but its periodicity in the time domain that produces this negative effect.

3.2.2 Spectra of Randomly Sampled Signals

Now assume that sampling is performed randomly according to definition (3.6). Then the estimated spectrum

$$S_x(f_n) = \lim_{N\Rightarrow\infty} \frac{2}{N} \int_0^N x(t)u(t) \exp(\mathrm{j}2\pi f_n t)\, \mathrm{d}t$$

$$= \lim_{N\to\infty} \frac{2}{N} \sum_{k=1}^{\hat{N}} N x(t_k) \exp(-\mathrm{j}2\pi f_n t_k), \tag{3.12}$$

where $\hat{N}$ is the random number of signal samples taken during the time interval $\Theta = N\mu$. Function $u(t)$ in this case is defined by Equation (3.2).

Note that the probability density function $\varphi_k(t)$ of the function $\delta(t - t_k)$ is also the same for time intervals in the range $[0, t_k]$. It is proved in Chapter 6 that additive random sampling, if properly performed, is characterized by the equation

$$\sum_{k=1}^{\infty} \varphi_k(t) = \frac{1}{\mu} = \text{constant}, \tag{3.13}$$

where $1/\mu$ denotes the mean sampling rate. If this condition holds, then it follows from Equations (3.12) and (3.13) that the expectation of the estimated spectra

$$E\left[\lim_{\Theta\Rightarrow\infty} \hat{S}_s(f_n)\right] = \lim_{\Theta\Rightarrow\infty} \frac{2}{\Theta} \int_0^{\Theta} \sum_{k=1}^{\infty} x(t) \exp(-\mathrm{j}2\pi f_n t)\varphi_k(t)\, \mathrm{d}t$$

$$= \lim_{\Theta\Rightarrow\infty} \frac{2}{N} \int_0^{\Theta} x(t) \exp(-\mathrm{j}2\pi f_n t) \sum_{k=1}^{\infty} \varphi_k(t)\, \mathrm{d}t = S_x(f_n). \tag{3.14}$$

The properties of randomly sampled signals are therefore completely different from those of periodically sampled ones. On the basis of Equation (3.14) a very important conclusion is found:

> If randomized sampling satisfies the condition (3.13), then the expectation of the estimated spectra of randomly sampled quasi-stationary signals coincides with the spectra of the respective original signals.

The spectral characteristics of a signal after periodic and randomized sampling are shown in Figure 3.2.

It can be seen from the obtained results of this analysis that randomly sampled signals, under specific conditions, might have unique spectral characteristics.

Firstly, the spectra of the randomly sampled signals are not periodic in the frequency domain. Secondly, the spectra of the original signals digitized in this way might well be broadband, with their upper frequencies considerably exceeding the Nyquist limit. Consequently, such randomization of sampling makes it possible to perform spectral analysis of high-frequency signals without distortions of their spectra that are usually observed under the same conditions when the sampling operation is performed periodically.

However, it should be noted that these spectral characteristics have been obtained on the assumption that the signals are stationary and that they are observed and digitized infinitely. What happens when digitizing conditions are more realistic remains to be found. Nevertheless, the fact that randomization of sampling might lead to elimination of spectral distortions is significant.

3.3 Aliasing Induced Errors at Seemingly Correct Sampling

The sampling theorem states that a sequence of signal samples taken periodically is an accurate representation of the original signal provided that the upper frequency of the analog signal spectrum is equal to or less than half the sampling frequency. If this condition holds, the analog signal can be reconstructed by interpolation between the sample values on the basis of the following formula:

$$x^*(t) = \sum_{k=-\infty}^{\infty} x(kT) \frac{\sin \pi(t/T - k)}{\pi(t/T - k)}. \tag{3.15}$$

The reconstructed signal $x^*(t)$ is identical to the original only if the signal spectral characteristics and the sampling frequency satisfy the requirements of the sampling theorem. The typical penalty for not meeting them is the aliasing effect, leading to increased signal reconstruction errors. On the other hand, if the requirements defining the correct sampling conditions have been met, then no aliasing and errors should be induced by this effect. It is tempting to interpret the sampling theorem as a definition of conditions that, if satisfied, would guarantee that no matter how the sampled signal is processed there would be no errors due to aliasing.

The aliasing effect manifests itself in a number of seemingly different ways and the systematic errors arising from it depend not only on the signal spectra and the mean sampling rate but also on the specific task for processing the digitized signal. For this reason, the requirements that have to be satisfied to avoid aliasing are often more severe than those defined by the sampling theorem. However, this theorem

does not say anything about the subsequent processing of the sampled signals. While it states that under the specified conditions it is possible to reconstruct the original signals from their sample sequences, this does not mean that there are guarantees against aliasing errors in cases where these sample sequences are used not for signal reconstruction but for some other kind of signal processing. An attempt will be made to clarify this seemingly conflicting statement.

Estimates of a variety of averaged signal parameters, such as the mean value, absolute mean value (rectified signal mean value), mean square value (mean power), variance, higher moments of the probability density function of x and other parameters can be defined as the expected values $E[F(x)]$ of the appropriately functionally converted signal $F(x)$. This expectation is given as

$$E[F(x)] = \int_{-\infty}^{\infty} F(x)p(x)\,\mathrm{d}x = \lim \frac{1}{2\Theta} \int_{-\Theta}^{\Theta} F(x(t))\,\mathrm{d}t. \tag{3.16}$$

Assume that the signal $x(t)$ is equidistantly sampled and try to determine how the signal spectrum should be restricted under these conditions to ensure that there is no aliasing. The Fourier series expansion of the functionally converted signal for a time interval $[0, \Theta]$ is as follows:

$$F(x) = a_0 + \sum_{n=1}^{\infty} (a_n \cos 2\pi f_n t + b_n \sin 2\pi f_n t), \tag{3.17}$$

where a_0, a_n and b_n are the Fourier coefficients, characterizing $F(x)$ for $f = n/\Theta$. Substitution of series (3.17) into Equation (3.16) yields

$$E[F(x)] = a_0 + \lim_{\Theta \Rightarrow \infty} \frac{1}{2\Theta} \sum_{n=1}^{\infty} \int_{-\Theta}^{\Theta} (a_n \cos 2\pi f_n t + b_n \sin 2\pi f_n t)\,\mathrm{d}t. \tag{3.18}$$

As the averaged value of any sinusoid is equal to zero, $E[F(x)] = a_0$. The estimate of $E[F(x)]$ can be written as

$$\hat{E}[F(x)] = E[F(x)] + \lim_{N \Rightarrow \infty} \frac{1}{2N} \sum_{n=1}^{\infty} \sum_{k=1}^{N} \left(a_n \cos \frac{2\pi f_n k}{f_s} + b_n \sin \frac{2\pi f_n k}{f_s} \right). \tag{3.19}$$

If the set of indexes n for which the ratio f_n/f_s is an integer is denoted by V, then it follows from Equation (3.19) that

$$\hat{E}[F(x)] = E[F(x)] + \sum_{n \in V} a_n. \tag{3.20}$$

The estimate $\hat{E}[F(x)]$ is therefore unbiased if the functionally converted signal $F(x)$ contains no components at frequencies for which the ratio f_n/f_s is an integer or, in other words, for which the set V is empty.

This method, when applied to determine the conditions for avoiding aliasing, provides interesting results. For instance, to ensure that there is no aliasing when the mean value of a signal is estimated, it suffices that the signal contains no components at the frequencies $f_s, 2f_s, \ldots, rf_s, \ldots$. Hence the restrictions on the signal spectrum are more loose for this kind of signal processing than those defined by the sampling theorem. Estimation of the mean value is alias free if the upper frequency of the signal spectrum does not exceed the sampling frequency f_s.

However, this case is an exception. Analysis of other cases shows that the limitations on the signal bandwidth guaranteeing the absence of aliasing errors are usually stricter than those that formally follow from the sampling theorem. To demonstrate this, consider some examples.

Example 3.1

The mean value of $x^3\,(kT)$ is to be estimated. The sampled signal

$$x(nT) = c_1 \sin(2\pi f_1 kT + \varphi_1) + c_2 \sin(2\pi f_2 kT + \varphi_2).$$

Obviously, $E[x^3(kT)] = 0$. Some specific cases will be considered.

(a) Suppose $f_1 = 0.4f_s$ and $f_2 = 0.2f_s$. It is evident the requirements of the sampling theorem are met. However, calculations made on the basis of the commonly used averaging procedure yield

$$\lim_{N\Rightarrow\infty} \frac{1}{N} \sum_{k=1}^{N} x^3(kT) = -\frac{3}{4} c_1^2 c_2 \sin(2\varphi_1 + \varphi_2) \neq 0.$$

The estimate so obtained is biased as a result of aliasing.

(b) Now consider the case where $f_1 = 0.4f_s$ and $f_2 = 1.1f_s$. In this case the conditions of the sampling theorem are not satisfied. Nevertheless, the averaging performed according to the above formula provides an unbiased estimate of $E[x^3(kT)]$, which means that in this case there is no aliasing.

Comments. It can be shown that $x^3(t)$ for the given signal contains components at frequences $3f_1, 3f_2, (2f_1 + f_2), (f_1 + 2f_2), f_1, f_2$. In the case of Example 2.1(a) aliasing occurs because $2f_1 + f_2 = f_s$. There is no aliasing in the case of Example 2.1 (b) because none of the components of the six indicated frequencies of $x^3(t)$ is equal to rf_s.

Example 3.2

The upper frequency f_n of a signal spectrum is equal to $0.25 f_s$. The demands of the sampling theorem are obviously met. The estimate $E[|x\{nT\}|]$ is calculated on the basis of the commonly used algorithm given below. The question is whether or not the estimate will be corrupted by errors due to aliasing. To find the answer, it is sufficient to consider the estimation of the mean absolute value of only one sinusoidal signal component at frequency f_n. It can be shown that

$$E\,[|x(t) = c\sin 2\pi f_n t|] = \frac{2c}{\pi}.$$

On the other hand, when this signal is sampled and digitally processed as required, the expected estimate

$$\hat{E}\,[|c\sin 2\pi f_n kT|] = \lim_{N\Rightarrow\infty}\frac{1}{N}\sum_{k=1}^{N}\left|c\sin\frac{\pi k}{2}\right| = \frac{c}{2}.$$

Comparison of the true and the estimated values reveals that the latter is biased and that aliasing therefore occurs. To understand why this is so, the Fourier series expansion of $|c\sin 2\pi f_n t|$ should be considered. It can be shown that it contains components at frequencies $2f_n, 4f_n, 6f_n, \ldots$. Under the given conditions, $4f_n = 2f_s$ and $8f_n = 2f_s, \ldots$. The components at these frequencies cause the bias errors.

The given examples show how misleading a superficial interpretation of the sampling theorem can be. They are given in order to draw attention to the fact that conditions for alias-free signal processing, in the cases where signals are to be functionally converted, differ from those stated by the sampling theorem. It is not sufficient to know the upper frequency of the signal spectrum to determine how high the sampling frequency should be to be sure that there will be no aliasing induced errors. The subsequent specific processing of the digitized signal should also be taken into account.

Note that functional conversions, performed either before or after the sampling operation, lead to the same additional restrictions on the spectra of the corresponding signals. If the original signal $x(t)$ is functionally converted before digitizing, as shown in Figure 3.3(a), then, naturally, the required sampling frequency is determined by the spectrum of the converted signal $F[x(t)]$. When the equivalent conversion is performed in the course of processing the digital signal $x(kT)$, as shown in Figure 3.3(b), then the signal is sampled at the same rate as in the first case.

This rule is applied to the cases illustrated by Examples 3.1 and 3.2. It can easily be established that the permitted bandwidth of the signal $x(t)$ is less than $1/3 f_s$ in the case of the functional conversion $F[x(t)] = x^3(t)$. In the case when

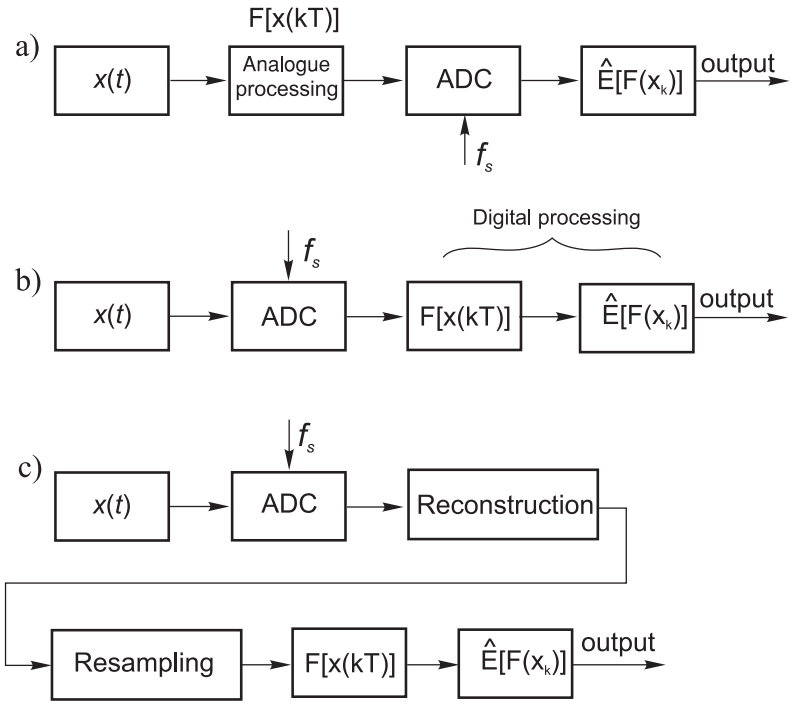


Figure 3.3 Signal parameter estimation options: (a) functional conversion is performed before analog-to-digital conversions; (b) functional conversion is performed after analog-to-digital conversions; (c) parameter estimation scheme not requiring an increase in the signal sampling rate

$F[x(t)] = |x(t)|$, the acceptable signal bandwidth is even narrower, less than $1/4 f_s$.

There is another possible way to organize sampling and processing, illustrated in Figure 3.3(c). This approach allows sampling to be performed without taking into account additional restrictions imposed by processing algorithms. Accordingly, sampling is carried out correctly as required by the sampling theorem, after which the original signal is reconstructed and resampled. This makes it possible to obtain as many sample values as required to ensure that no aliases interfere with processing the digitized signal. However, this method is much more complicated and therefore it is not suggested for practical applications.

At least two conclusions may be reached from this brief discussion and Figure 3.3. Firstly, sampling conditions should be matched to the specifics of the signal processing algorithms used to change the criterion for alias-free signal sampling as stated by the sampling theorem. Secondly, the application area of

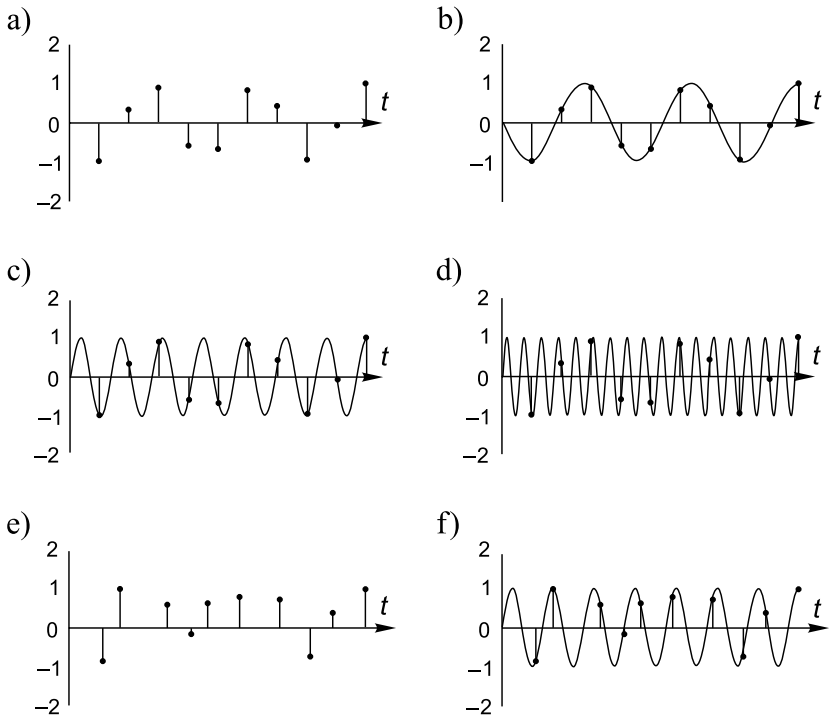


Figure 3.4 Difference between: (a) regular and (e) irregular sample value sets. Many sinusoids, for instance at the indicated frequencies (b), (c) and (d), can be drawn through the set in (a), and only one through the set given in (e)

the conventional anti-aliasing technique, in the case of periodic sampling based on low-pass prefiltering, is considerably narrower than formally follows from the theorem.

3.4 Overlapping of Sampled Signal Components

Overlapping of a periodically sampled signal components occurs when the original signal contains two or more components at the frequencies

$$f_0,\ f_s \pm f_0,\ 2f_s \pm f_0,\ \ldots,\ nf_s \pm f_0. \tag{3.21}$$

Therefore many sinusoids can be drawn through an irregular sample value set, as shown in Figure 3.4. However, it is not clear what really happens when two or more signal components overlap. It is not obvious exactly how they add up.

It is clear that adding several analog sinusoids at one and the same frequency forms one sinusoid but that does not directly apply to the case of aliasing. The situation is then different as in that case several sinusoids at different frequencies are added and the obtained sum is sampled. The relationships underpinning this process will be examined.

Consider a wideband complex signal $x(t)$ containing one component at a frequency $f_1 \in [0, 0.5f_s]$ and many other sinusoids at the frequencies indicated in (3.21):

$$x(t) = c_1 \sin(\omega_1 t + \varphi_1) + \sum_{m=1}^{\infty} \{c_{1m} \sin[2\pi(mf_s + f_1)t + \varphi_{1m}] + c_{2m} \sin[2\pi(mf_s - f_1)t + \varphi_{2m}]\}, \tag{3.22}$$

where $\omega_1 = 2\pi f_1$. Assume that this signal is periodically sampled at time instants $t_k = kT = k/f_s$. To derive a description of the sampled signal, it is useful to take into account the following equality:

$$2\pi \frac{(mf_s \pm f_1)k}{f_s} = 2\pi mk \pm \frac{2\pi f_1 k}{f_s}.$$

Then the sampled signal can be given as

$$x(t_k) = c_1 \sin(\omega_1 t_k + \varphi_1) + \sum_{m=1}^{\infty} [c_{1m} \sin(\omega_1 t_k + \varphi_{1m}) + c_{2m} \sin(\omega_1 t_k + \pi - \varphi_{2m})] \tag{3.23}$$

Another expression might also be used to describe this sampled signal. To derive it Equation (3.23) can be rewritten by substituting the following Fourier coefficients:

$$\begin{aligned} a_1 &= c_1 \sin\varphi_1, & b_1 &= c_1 \cos\varphi_1, \\ a_{1m} &= c_{1m} \sin\varphi_{m1}, & b_{1m} &= c_{1m} \cos\varphi_{m1}, \\ a_2 &= c_{2m} \sin\varphi_{2m}, & b_2 &= c_{2m} \cos\varphi_{2m}. \end{aligned}$$

Then

$$\begin{aligned} x(t_k) &= a_1 \cos\omega_1 t_k + b_1 \sin\omega_1 t_k \\ &\quad + \sum_{m=1}^{\infty} (a_{1m} \cos\omega_1 t_k + b_{1m} \sin\omega_1 t_k + a_{2m} \cos\omega_1 t_k - b_{2m} \sin\omega_1 t_k) \\ &= \left[a_1 + \sum_{m=1}^{\infty} (a_{1m} + a_{2m})\right] \cos\omega_1 t_k + \left[b_1 + \sum_{m=1}^{\infty} (b_{1m} - b_{2m})\right] \sin\omega_1 t_k \end{aligned} \tag{3.24}$$

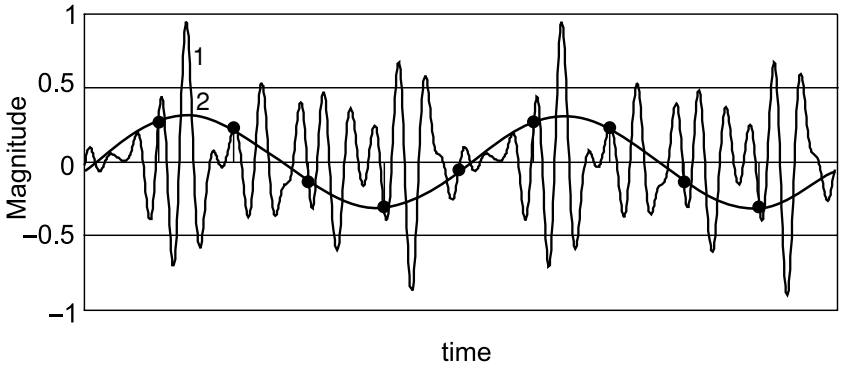


Figure 3.5 Waveform (1) of a complex analog signal and a sinusoid (2) that can be drawn through the same set of signal sample points

It follows from Equation (3.24) that

$$x(t_k) = a\cos\omega_1 t_k + b\sin\omega_1 t_k \tag{3.25}$$

where

$$\begin{aligned} a &= a_1 + \sum_{m=1}^{\infty}(a_{1m} + a_{2m)} \\ b &= b_1 + \sum_{m=1}^{\infty}(b_{1m} - b_{2m}) \end{aligned} \tag{3.26}$$

These expressions show that if a signal contains a number of components with frequencies belonging to the row (3.21), their sum can be described as one digitized sinusoid. The frequency of this sinusoid may be any from (3.21).

An illustration of the summing of a number of sinusoids at frequencies from row (3.21) and the associated digital version of the sum are given Figure 3.5. The analog signal (waveform 1) is formed in accordance with Equation (3.22). In addition to the sinusoid at frequency $f_1 = 0.2f_s$, other sinusoids are included. Waveform 1 of a complex analog signal contains components at frequencies exceeding $0.5f_s$ and is characterized by the parameters given in Table 3.1 and shown in Figure 3.5. This signal is sampled and the sinusoid (waveform 2) with parameters defined by Equations (3.26) is then drawn through the sample points.

Compare waveforms 1 and 2. They are quite different. Yet there is no doubt that in the digital domain they are equivalent, because both of them are represented by one and the same sample sequence. Hence one of the digital representations is

Table 3.1

m	C_{1m}	C_{2m}	φ_{1m}	φ_{2m}
1	0.10	0.15	−1	3
2	0.20	0.15	5	−2
3	0.05	0.10	−3	2
4	0.05	0.10	1	4
5	0.10	0.05	−4	6

not adequate. Obviously, it is waveform 1 that is misrepresented. To understand why this is so, note that according to Equations (3.26) it is the Fourier coefficients of sampled signal components and their related aliases rather than the sampled waveforms that are added. This is how waveform 2 in Figure 3.5 was obtained.

In the domain of discrete signals, summing sinus waves even at different frequencies may result in one sinusoid. This will be so in cases where the frequencies of the particular sinusoids belong to the row (3.21). However, it should be noted that the sinusoid obtained in such a way is really imaginary. It shares its sample values with the associated aliases, but its parameters differ from those of the corresponding summary analog signal. To confirm this, the mean powers of the signal defined by Equation (3.22) and of its digital sinusoid can be compared. The mean power of the signal is

$$P_x = \tfrac{1}{2}\left[a_1^2 + \sum_{m=1}^{\infty}\left(a_{1m}^2 + a_{2m}^2\right) + b_{1m}^2 + \sum_{m=1}^{\infty}\left(b_{1m}^2 + b_{2m}^2\right)\right]$$

while the mean power of the sinusoid given by Equation (3.26) is

$$P_{\mathrm{s}} = \tfrac{1}{2}(a^2 + b^2).$$

It is apparent that $P_x \neq P_s$.

As the examples given above show, the manner in which aliasing affects the estimation of signal parameters is one of the essential points that have to be kept in mind to avoid corruption of signal parameter digital estimation results.

3.5 Various Approaches to Randomization of Sampling

Once the fact is accepted that other types of sampling processes might and often should be used for signal digital representation as alternatives to periodic sampling, it soon becomes clear that there are many variations of the sampling processes that have to be considered in order to choose a version most suitable

for a given specific application. Two basic sampling options, both appropriate for various applications, are considered further: periodic and nonuniform.

Periodic sampling is preferable whenever the signal bandwidth can be restricted or the sampling operation performed at the rate required by the sampling theorem. There are several reasons for this preference. Firstly, periodic sampling is the simplest method of performing this procedure. Secondly, periodic sequences of signal samples are well suited to digital processing. In addition, there are many highly efficient algorithms available for processing periodically sampled signals, including so-called fast algorithms.

Nonuniform sampling might take place either when signal sample values are obtainable and could be taken only at some random unpredictable time instants or when these sample values are taken nonuniformly in order to obtain some specifically targeted effects, like avoiding aliasing. The former cases are beyond the scope of this book. However, many of the developed techniques for handling nonuniform data are also applicable in this area. The latter cases are of most interest for many further applications.

One of the basic subjects examined in this book is nonuniform sampling, realized either as deliberately randomized or deliberately pseudo-randomized operations. They are vital for achieving the capability of processing signals in a very wide frequency range. Therefore this kind of sampling is most interesting in the light of problems considered in the following chapters. While that is true both for deliberately randomized and deliberately pseudo-randomized sampling processes, the techniques for their execution and the most suitable applications for them differ.

Randomized sampling is sometimes appropriate for applications when it is undesirable to distort signals by filtering off signal components at frequencies exceeding half of the sampling rate. However, the signal sample values at randomized sampling are taken at unknown random time instants. Therefore application of this kind of randomized sampling is limited to the relatively rare cases where information about exact sampling instants is not relevant. Estimation of some signal parameters, including signal power, and measurements of time intervals with a subnanosecond time resolution might be mentioned as some application examples for such a sampling approach.

Sampling might be randomized either directly or indirectly. To realize direct sampling randomization, signal sample values are taken at random time instants linked to pulses in a specially generated random sampling pulse process.

For indirect sampling randomization, a periodic reference waveform rather than a sequence of pulses formed at random time instants is used to perform the sampling process. The sampling process itself is then defined as the signal and the

reference waveform crossings. The basic design of the involved electronic samplers is simple and helps to locate these devices very close to the signal sources. The advantage of this approach is that the link between the devices performing the sampling operation and the master part of the whole structure is digital. This kind of indirectly randomized sampling is suggested in Chapter 7 as a technique of remote sampling used for data collection from multiple distributed signal sources. As shown there, it represents a competitive alternative to multichannel data acquisition based on multiplexing of analog signals. Digital output pulses of the comparators, which fix in time the crossings of the signal and the reference function, rather than analog signals are multiplexed in this case. Reconstruction of the signal sample values is based on the *a priori* reference waveform parameters and the fixed crossing time instants. This particular sort of sampling process is outside the mainstream sampling techniques. However, it has been found to be quite useful for temporal and spatial data acquisition from a large quantity of signal sources.

Pseudo-randomized sampling is the basic anti-aliasing technique. The indications for its use are the same as for the randomized sampling except that the sampling instants in this case are predetermined with high resolution and precision. While this type of sampling might be performed in various ways, the most often used method is the so-called additive pseudo-randomized sampling scheme. It is described in detail in Chapters 6 and 9.

Both uniform periodic and nonuniform sampling have their typical applications, advantages and limitations. Indeed, while periodic sampling has many remarkable positive properties, application of it is limited by its basic disadvantage, aliasing. To avoid aliasing, nonuniform sampling has to be used. However, nonuniform sampling also has some disadvantages. Application of it often leads to more or less noticeable statistical signal processing errors. Therefore it makes sense to combine the periodic and nonuniform sampling processes in an attempt to develop sampling techniques applicable for alias-free digitizing of signals in a wider frequency range with increased accuracy. These considerations led to the development of a quite useful hybrid sampling (HS) approach based on mixing the periodic and nonuniform sampling processes. It is discussed in Chapter 10.

HS actually covers various kinds of sampling models. The motivation for using the HS approach is the fact that potentially these techniques are very beneficial both for theoretic abstractions and practical applications. Application of HS models often helps in simplifying the analysis of various signal processing cases and HS abstractions are also useful in the development of new algorithms.

A particular HS model is hybrid double sampling (HDS). The idea underpinning HDS is the use of at least two ADCs connected in parallel. Such an approach by

Table 3.2

Especially addressed applications	Advantages and limitations
1. Directly randomized nonuniform sampling	
Signal parameter estimation not requiring sample value timing	Simple and robust technical implementation of sampling, simple designs of processing devices
Subnanosecond resolution time interval and related parameter digital measurements	Limited application range.
2. Indirectly randomized nonuniform sampling, based on reference function crossings	
Remote sampling for massive data acquisition systems	Reduced complexity low-power remote sampler designs Digital link, either wire or radio, inserted between the sampling and master parts of multi-channel ADC Data compression Limited to data acquisition from relatively low frequency sources.
3. Pseudo-randomized nonuniform sampling	
Universal digital alias-free signal processing	Wide range of alias-free versatile digital processing applications Increased complexity processing The existing fast algorithms not applicable
4. Hybrid periodic/nonuniform sampling	
Versatile signal analysis and waveform reconstruction. Widening of dynamic range Digital signal preconditioning for reduced complexity processing	Reduction of nonuniform sampling drawbacks Better suited for processing adapted to sampling irregularities Increased-complexity processing in comparison with the case of periodic sampling The existing fast algorithms not applicable
5. Hybrid double periodic/nonuniform sampling	
Signal processing requiring closely placed sample value taking, high-resolution correlation analysis, processing of signals with continuous spectra	High resolution at correlation and spectrum analysis Two ADCs connected in parallel required for sampling

itself is not new. Signal sampling by two ADCs connected in parallel is a technique often applied in cases of periodic sampling of signals. It is obvious that this is done to increase the sampling rate twice. However, in the case of HDS, the same approach is used to obtain a quite different effect. Specifically, HDS is usually performed to provide the right conditions for processing wideband signals with both discrete and continuous spectra in a broad frequency range. In such cases

the distances between the adjacent sampling instants need to be sufficiently short. When sampling is performed in accordance with the HDS model, the signal sample values, formed by the outputs of two ADCs connected in parallel, could be placed much closer than half of the sampling interval characterizing the use of a single ADC.

Thus five different nonuniform sampling approaches, in addition to periodic sampling, are studied in the following chapters. It is assumed that the advantages and drawbacks of periodic sampling are well known. All of the nonuniform sampling approaches, their typical applications, advantages and limitations are summarized in Table 3.2.

Apparently the classical way of treating signals digitally is not exclusive. There are also other different ways of doing that, as shown in this book and in many other publications. These alternatives are examined in this book. Although the nonuniform sampling techniques differ, all of them are based on taking signal sample values irregularly in time. That is necessary in order to avoid frequency overlapping and to avoid aliasing. Theoretically, it is often convenient to assume that the random (pseudo-random) intervals between the sampling instants are distributed according to one or another distribution, including the exponential distribution. However, there are practical considerations that impose some restrictions. Specifically, the interval between two successive sampling instants should never be shorter than allowed by the operational speed of the used ADC. By definition, the mean sampling rate of a typical alias-free nonuniform sampling process covering a certain frequency range is lower than the sampling frequency of the periodic sampling process that would be used for sampling the same signals. In other words, a typical nonuniform sampling process is sparse with fewer sample values taken in a given time interval.

Bibliography

Bellanger, M. (1988) *Digital Processing of Signals: Theory and Practice*. New York: John Wiley & Sons, Inc.

Beutler, F.J. (1970) Alias-free randomly timed sampling of stochastic processes. *IEEE Trans. Inf. Theory*, **IT-16**(2), 147–52.

Beutler, F.J. (1974) Recovery of randomly sampled signals by simple interpolators. *Inf. Control*, **26**(4), 312–40.

Beutler, F.J. and Leneman, O.A. (1966) Random sampling of random processes: stationary point processes. *Inf. Control*, **9**(4), 325–46.

Beutler, F.J. and Leneman, O.A. (1966) The theory of stationary point processes. *Acta Math.*, **116**, 159–97.

Beutler, F.J. and Leneman, O.A. (1968) The spectral analysis of impulse processes. *Inf. Control*, **12**(3), 236–58.

Bilinskis, I. and Mikelsons, A. (1983) *Digital Random Processing of Continuous Signals* (in Russian). Riga: Zinatne.

Bilinskis, I. and Mikelsons, A. (1990) Application of randomized or irregular sampling as an anti-aliasing technique. In *Signal Processing*, V: *Theories and Application*. Amsterdam: Elsevier Science Publishers, pp. 505–8.

Bilinskis, I., Mikelsons, A. and Vystavkin, A.N. (1986) Processing of randomly-sampled signals. In *Signal Processing*, III: *Theories and Applications*. Amsterdam: Elsevier Science Publishers, pp. 109–12.

Bland, D. and Tarczynski, A. (1997) The effect of sampling jitter in a digitized signal. In the IEEE International Symposuim on *Circuits and Systems (ISC AS'97)*, Vol. 4, Hong Kong, 9–12 June 1997, pp. 2685–8.

Bland, D. and Tarczynski, A. (1997) Optimum nonuniform sampling sequence for alias frequency suppression. In the IEEE International Symposuim on *Circuits and Systems (ISC AS'97)*, Vol. 4, Hong Kong, 9–12 June 1997, pp. 2693–6.

Brown, W.M. (1963) Sampling with random jitter. *SIAM J. Appl. Math.*, **11**, 460–73.

Cox, D.R. (1967) *Renewal Theory*. London: Chapman and Hall.

Dowski, E.R., Whitemore, C.A. and Avery, S.K. (1988) Estimation of randomly sampled sinusoids in additive noise. *IEEE Trans. Acoust., Speech, Signal Process*, **36**(12), 1906–8.

Gorelov, G.V. (1982) *Irregular Sampling*. Moscow: Radio i Svyaz.

Jimenes, J. and Agiu, J.C.(1987) Approximate reconstruction of randomly sampled signals. *Signal Processing*, **12**(2), 153–68.

Leneman, O.A.Z. (1966) Random sampling of random processes: impulse sampling. *Inf. Control*, **9**(4), 347–63.

Leneman, O.A.Z. (1966) Random sampling of random processes: optimum linear interpolation. *Franklin Inst.*, **281**(4), 302–14.

Leneman, O.A.Z. (1966) On error bounds for jittered sampling. *IEEE Trans. Autom. Control*, **AC-11**(1), 150.

Marvasti, F. (2001) *Nonuniform Sampling, Theory and Practice*. New York: Kluwer Academic/Plenum Publishers.

Masry, E. (1971) Random sampling and reconstruction of spectra. *Inf. Control*, **19**(4), 275–88.

Masry, E. (1978) Alias-free sampling: an alternative conceptualization and its applications. *IEEE Trans. Inf. Theory*, **IT-24**(3), 317–24.

Masry, E. (1978) Spectral estimation of continuous-time processes: performance comparison between periodic and Poisson sampling schemes. *IEEE Trans. Autom. Control*, **AC-23**(4), 679–85.

Masry, E. (1978) Poisson sampling and spectral estimation of continuous-time processes. *IEEE Trans. Inf. Theory*, **IT-24**(2), 173–83.

Mednieks, I. and Mikelsons, A. (1990) Estimation of true components of wide-band quasi periodic signals. In *Signal Processing*, V: *Theories and Application*. Amsterdam: Elsevier Science Publishers, pp. 233–6.

Mikelsons, A. (1981) Estimating broadband signal parameters at a relatively low mean sampling rate. *Autom. Control-Comput. Sci*, **l**, 90–4.

Mikelsons, A. (1994) Alias-free spectral estimation of signals with components of arbitrary frequencies. In *Adaptive Methods and Emergent Techniques for Signal Processing and Communications*, Ljubljana, Slovenia, April 1994, pp. 105–8.

Qu, D. and Tarczynski, A. (2006) Weighted PNS sequences for digital alias-free processing signals. In Proceedings of the 10th International Conference on *Systems*, Vouliagmeni, Athens, Greece, 10–12 July 2006, pp. 1–6.

Qu, D. and Tarczynski, A. (2006) Analysis and design of WPNS sequences for digital alias-free signal processing. *WSEAS Trans. on Signal Processing*, **2**(7), July, 933–40.

Shapiro, H.S. and Silverman, R.A. (1960) Alias-free sampling of random noise. *SIAM J. Appl. Math.*, **8**(2), 245–8.

Tarczynski, A. and Qu, D. (2005) Optimal periodic sampling sequences for nearly-alias-free digital signal processing. In IEEE International Symposium on *Circuit and Systems (ISCAS'2005)*, Kobe, Japan, 23–26 May 2005, pp. 1425–8.

Tarczynski, A. and Tzvetkov, K. (2005) Evaluation of several random sampling schemes for DASP applications. In Proceedings of *SAMPTA'2005*, Samsun, Turkey, 10–15 July 2005.

Tarczynski, A. and Qu, D. (2005) Optimal random sampling for spectrum estimation in DASP applications. Special Section on Intelligent Control and Signal Processing in *International Journal of Applied Mathematics and Computer Science*, **15**(4), December, 463–9.

4

Randomized Quantization

Quantizing is usually defined as a rounding-off operation and this definition fits both the deterministic and the randomized versions of this operation. The quantization process is always carried out according to the same generalized scheme, given in Figure 4.1. To quantize a signal, instantaneous values are first roughly measured and then the measurement results are rounded off. The differences lie in how in these procedures are implemented, leading to various properties of the quantized signals. While the signal instantaneous values are always measured by comparing them with some reference threshold levels, the latter are kept in fixed positions for deterministic quantizing and are randomly varied for randomized quantizing. This means that the rounding-off function in both cases is carried out in two different ways. It is deterministic in the first case and probabilistic in the second case. Realization of the second approach seems to be and usually is more complicated than the first. However, under certain conditions it pays to perform quantizing of signals in this more complicated way as the properties of the quantized signals are quite different, which leads to various potential desirable benefits. On the other hand, the errors of randomized quantizing are typically distributed in an interval that is twice larger than those of comparable deterministic quantization. Essentials, advantages and drawbacks of this quantization approach are discussed in this chapter.

4.1 Randomized Versus Deterministic Quantization

Assume that the signals to be quantized are within the signal amplitude range $[-A_m, A_m]$. To perform B-bit quantizing, this range is subdivided into $2(2^{B-1} - 1)$ elementary intervals q, which are known as quantization steps. In the course of

Digital Alias-free Signal Processing I. Bilinskis

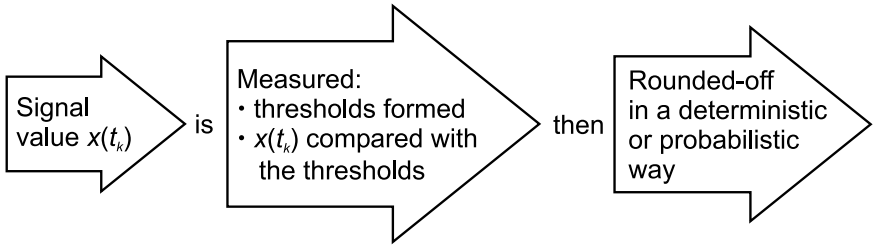


Figure 4.1 Essence of the quantizing operation

quantizing, the instantaneous values of signals have to be measured and rounded off, so the quantized signal is represented in the following form:

$$x_q = nq, \quad n = 0, \pm 1, \pm 2, \ldots, \pm(2^{B-1} - 1), \tag{4.1}$$

where

$$q = \frac{2A_m}{2^B - 2} = \frac{A_m}{2^{B-1} - 1}.$$

4.1.1 Basics

Everything said so far applies to both deterministic and the most popular versions of randomized quantizing. However, if the quantization operation is considered in some detail, then the deterministic and randomized quantizing of course differ. To continue this discussion, look at Figure 4.2, where the time diagrams illustrating deterministic quantizing of a unipolar signal are presented. The levels q, $2q$, $3q, \ldots$, to which input signal instantaneous values are rounded off, are shown in Figure 4.2(a) by dashed lines. They are in fixed positions. The distance between them is equal to the quantization step q.

To express the instantaneous values of the input signal in terms of q, they are first compared with another set of reference levels. In the case illustrated by Figure 4.2(a) these levels are constant. The distance between two of them is also equal to q, but the whole set of reference levels is located so that the first of them is elevated above the x axis by half the quantization step q. The measurement procedure is accomplished by comparing the signal with these reference levels and counting the number n of them that are below the signal's instantaneous value. When the signal being quantized is analog, as it is in the case shown, the quantized signal changes its value stepwise whenever the input signal crosses one of the reference thresholds. As can be seen from the diagrams, rough deterministic

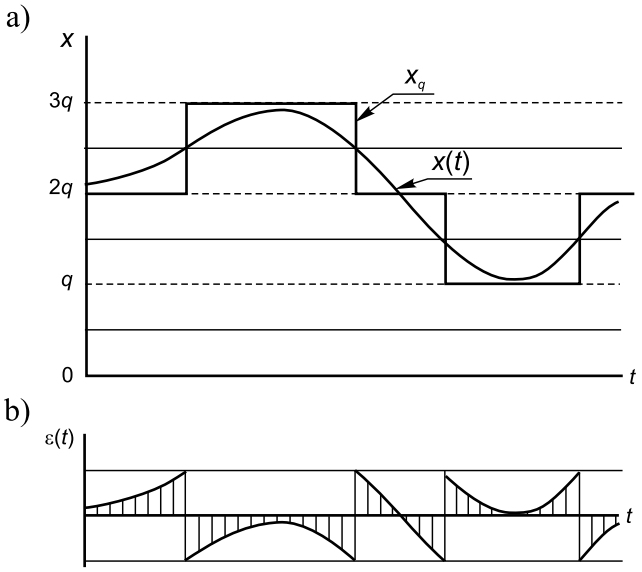


Figure 4.2 Time diagrams illustrating uniform deterministic quantizing: (a) original and quantized signals; (b) quantization noise

quantization is a typical nonlinear operation insensitive to small changes in the input signal.

Randomized quantizing follows the same procedure, but in this case the reference levels are not fixed: they change continuously or step by random step from one quantizing instant to the next, as can be seen in Figure 4.3. Randomly (or pseudo-randomly) quantized signals are henceforth denoted by $\hat{x}$. The notation $\hat{x}$ is used because it is simple. However, it should be noted that $\hat{x}$ does not denote the estimate of the corresponding signal value, unlike other similar notations used in this book. Then

$$\hat{x} = nq, \quad n = 0, \pm 1, \pm 2, \ldots, \pm(2^{B-1} - 1) \tag{4.2}$$

Formally, the rounded-off values of a randomly quantized signal are determined in the same way as for deterministic quantizing, i.e. the randomly changing threshold levels below the respective signal sample value are counted and the quantized signal $\hat{x}_k$ is defined as being equal to $n_k q$. However, the essence of this rounding-off operation in the case of randomized quantizing is quite different. As explained further, for randomized quantizing this operation is probabilistic. This means that one and the same instantaneous value of a signal with probabilities depending on

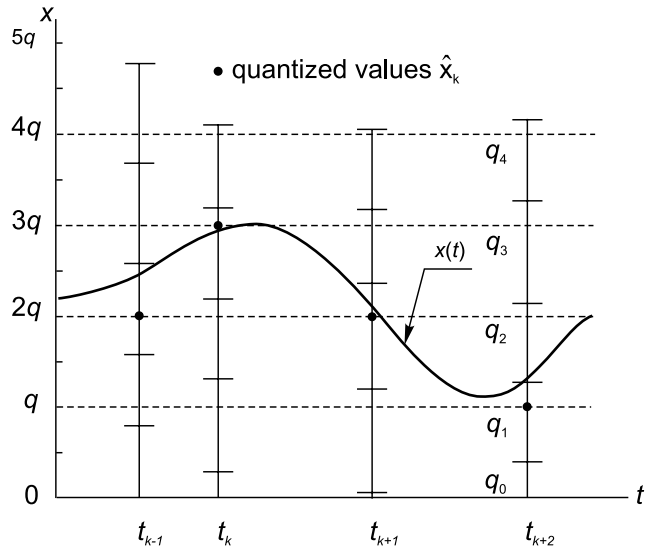


Figure 4.3 Illustration of randomized quantizing

that signal value might be rounded off to various digital levels. As a result, the properties of a randomly quantized signal differ significantly from the properties of a deterministically quantized signal.

Note that there are two seemingly different approaches to randomization of quantizing. While in the quite popular cases of so-called dithering, mentioned in Chapter 1, noise is added to the signal at the input of a deterministic quantizer, the quantization operation could also be randomized as explained above by just forming time variable sets of reference levels and using them for rounding off the input signal instantaneous values. Electronic implementations of these randomized quantization schemes are indeed different. However, the essence of them is equivalent. It does not matter exactly how the randomness is introduced at quantizing. The results are exactly the same.

4.1.2 Input–Output Characteristics

The essentials of various quantization models can also be illustrated by specific representation of the quantized signals, including their input–output characteristics, shown in Figure 4.4. While deterministic quantization has one fixed input–output relationship (Figure 4.4(a)), randomized quantization, in general, is characterized by different input–output relationships depending on the

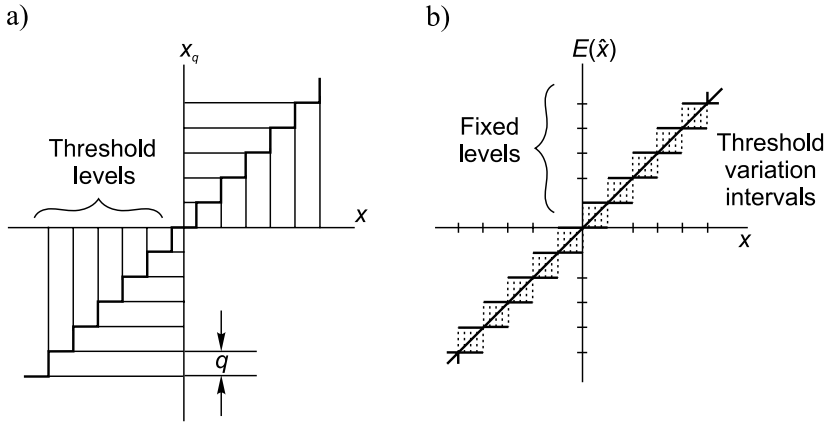


Figure 4.4 Input–output characteristics: (a) for deterministic quantizing and (b) input-expected output characteristic for randomized quantizing

specific quantizing model used. Moreover, the outcome of such quantizing, even for a specific quantizing model, is determined by instantaneous input–output characteristics. On the other hand, these instantaneous relationships are, of course, impractical. Instead, averaged input–output characteristics may be used. They are more convenient and appropriate for characterizing randomized quantizing. They may also be considered as input–expected output characteristics. The characteristics of the quantization case illustrated in Figure 4.5 are given in Figure 4.4(b). They are specific in the sense that they do not provide an exact answer to the question of what quantized signal value will be indicated in a particular case of a given input signal value. The answer to this question is obviously probabilistic. The input–expected output characteristic indicates the probabilities that specific quantized signal values will be assigned to a given input signal value. Figure 4.5 explains how these probabilities could be evaluated.

The expected value $E[\hat{x}]$ of the quantized signal after correctly executed randomized quantizing is equal to the input signal value. In the case illustrated in Figure 4.5, $x = E[\hat{x}] = 4.37$. However, at any given time instant the quantized signal can assume one of two possible values. In the case of this example, these values are 4 or 5. Figure 4.5 shows how to determine the probabilities $p(\hat{x} = 4q)$ and $p(\hat{x} = 5q)$.

Even these brief descriptions of deterministic and randomized quantizing clearly show that the deterministic and randomized approaches used to execute the quantization operation differ significantly. The properties of the quantized

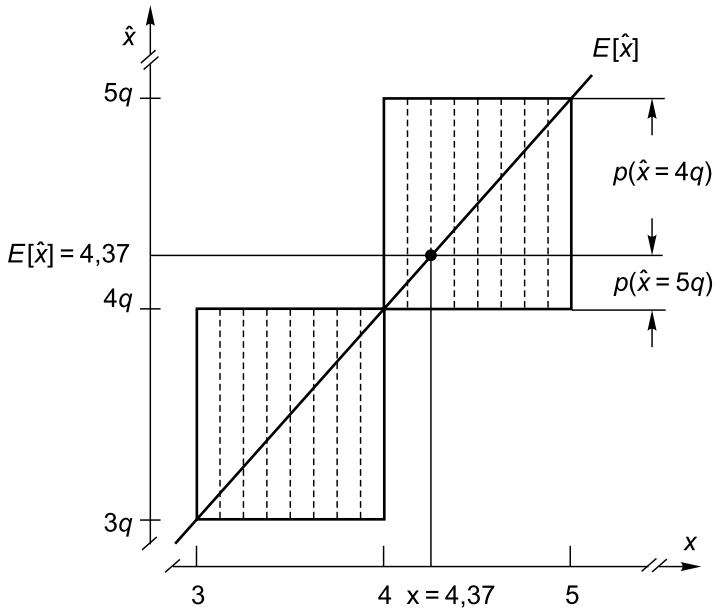


Figure 4.5 Illustration of quantizing performed according to Model 2 in detail.

signals obtained as a result of these two types of quantizing operation also differ greatly.

4.1.3 Rationale of Randomizing

The positive effect due to randomizing the quantization operation is actually obtained as a result of two procedures: randomizing the rounding off of signal instantaneous values and processing the randomly quantized signal. The simplest processing is simply averaging. Naturally, the deterministically quantized signals might be processed in this way as well. It is therefore of some interest to evaluate and compare the effect obtained when averaging is applied to both deterministically and randomly quantized signals.

In addition, it can be said that randomized quantizing is a linear operation if it can be accepted that the output signals are represented in statistical terms, i.e. as estimates of the expected quantized signal values. This obviously requires averaging of the particular quantized signal samples, which may or may not be appropriate from the viewpoint of subsequent signal processing.

Unlike deterministic quantizing, randomized quantizing is sensitive to small increments of signals, which cause proportional increases in output $\hat{E}[\hat{x}]$. This kind of quantizing can therefore be used for applications where it is desirable to use rough quantizing. It is of practical interest not only from the viewpoint of reducing the bit streams representing quantized signals. What is more important is the fact that quantizers, when they contain only a few comparators, can be built as extremely broadband devices. Such quantizers, or specialized ADCs, could be very valuable for ultra-high frequency applications. Moreover, this type of relatively simple specialized ADC can be successfully used even for relatively complicated signal processing such as correlation and spectral analysis.

Rough deterministic and rough randomized quantizing schemes differ noticeably. A single-threshold level deterministic quantizer does not give much information about the input signal. While it is possible to measure the time intervals between signal crossings of the threshold level, none of the signal basic parameters can be estimated. The randomized single-threshold level quantizing process is much more informative. By processing the randomly quantized signals presented in the form of stochastic bit streams it is possible to estimate the mean value of the original signal and to measure its peak or amplitude values.

4.2 Deliberate Introduction of Randomness

Randomness is deliberately introduced into the quantization process to perform the rounding-off operation in a probabilistic way. This means that a signal instantaneous value $x_k$within a range $(0, q)$ is rounded off to the values 0 or q depending on results of comparing it to a randomly generated value ξ_k uniformly distributed within the same range. If $x_k < \xi_k$, the value of x_k is rounded off to 0 and if $x_k > \xi_k$ then the result of rounding-off is equal to q. Therefore the outcome of the probabilistic signal rounding-off depends not only on the signal value itself but also on the random variable ξ. Even a small signal value with some probability may be rounded off to q and when that happens the particular quantization error is large. Such randomization of quantizing can be beneficial only under certain conditions. It does not make sense to randomize the quantization of separate uncorrelated signal sample values as it would then lead to worse, not better, results. On the other hand, the properties of the signal quantized in this way, as shown below, are attractive. In general, this approach might be advantageous for applications where sampled signals are quantized continuously for some duration of time and the quantized signals are later properly processed. This can be implemented in various ways.

4.2.1 Various Models

If the idea is accepted that it is possible to use time-variant threshold levels for performing signal quantization, various models for randomized quantization might be suggested. They differ first of all in the underlying statistical relationships, according to which sets of random threshold levels are formed. In some other cases the definitions of the quantized signals differ as well. Three models of randomized quantization have been selected from a relatively large quantity considered over a long period of time. They are discussed below and recommended for practical applications. In fact, they are versions of one and the same basic Model 1, illustrated in Figure 4.5.

Model 1

The intervals between the reference threshold levels (Figure 4.3), according to this model, are random and each set of levels used at time instants $t_{k-1}, t_k, t_{k+1}, t_{k+2}, \ldots$ differs from every other set. The threshold level ordinates q_{ki} should therefore be denoted by double indices indicating the time instant (first index) and the respective threshold level ordinate (second index). The threshold level sets are formed in such a way that

$$q_{ki} = q_{k(i-1)} + \tau_{ki}, \tag{4.3}$$

where τ_{ki} is a realization of the random variable.

Note that Equation (4.3) actually describes the same relationship as the one on which the additive random sampling scheme is based. As with random sampling, the intervals between the randomly positioned threshold levels $\{q_{ki}, q_{k(i-1)}\}$ should also be identically distributed and mutually independent. This is required to ensure that the quantized signal values $\hat{x}_k$ are unbiased, i.e. to ensure that $E[\hat{x}_k] = x_k$. In the case of Model 1,

$$\hat{x}_k = n_k \bar{q}, \tag{4.4}$$

where n_k is the number of threshold levels below the corresponding signal value x_k and $\bar{q}$ is the mean value of the intervals between the threshold levels.

It is apparent from Figure 4.3 that implementation of this quantization scheme is not simple. Since it makes sense to randomize the quantization of mostly high-frequency signals, the threshold level sets needed for that have to be generated during short time intervals $\{t_{k+1} - t_k\}$. This is not easy because the required statistics of those levels should satisfy certain requirements. For instance, the reference value in this case is the mean value of the intervals between the threshold levels $\bar{q}$. Therefore the stability of this parameter has to be secure. This threshold

forming process will not be dealt with in detail because quantization in accordance with Model 1 is not really recommended for analog-to-digital conversions of time-variant signals. This approach is much better suited for quantization of short time intervals, phase angles and other related physical quantities. For these applications, the quantization thresholds are in the form of short pulses generated at proper time instants and the involved techniques are those used for randomized sampling. They are considered in Chapter 6.

The randomized quantization of time intervals can therefore be analysed using virtually the same relationships as those derived in Chapter 6 for randomized sampling. Of course, this is also true for amplitude quantization performed in accordance with Model 1. For instance, it can written that the expected value of the quantized signal is

$$E[\hat{x}_k] = \bar{q}E[n_k]. \tag{4.5}$$

The expected number of threshold levels within the interval $[-X_0, x_k]$ can be determined by applying the function derived in Chapter 3. In this case,

$$\lim_{x_0 \Rightarrow \infty} E[n_k] = \lim_{x_0 \Rightarrow \infty} [P(X_0 + x_k) - P(X_0)] = \frac{x_k}{\bar{q}}. \tag{4.6}$$

Substituting Equation (4.6) into expression (4.5) gives

$$E[\hat{x}_k] = x_k. \tag{4.7}$$

Therefore the expected value of a quantized signal value is equal to the corresponding signal value. In this sense the quantization is a linear operation. Such quantization is unbiased, which holds even for very crude quantization.

To ensure the correct performance of quantizers built according to the requirements of Model 1, only one parameter of the random threshold level sets, namely the mean value $\bar{q}$, should be kept constant at a given level. Other statistical parameters may slowly drift so long as they remain within some relatively broad limits. Instantaneous input-output characteristic is given in Figure 4.6.

Model 2

This model is more versatile and practical. In fact, this is a version of the generic Model 1, characterized by $\sigma/\bar{q} \Rightarrow 0$, where σ is the standard deviation of the intervals $\{q_{ki} - q_{k(i-1)}$. Under these conditions, the intervals between thresholds are constant and equal to the quantization step q. Randomized quantization performed in accordance with this model is also illustrated by Figure 4.3. This time diagram is given for quantization of a unipolar signal. The equidistant threshold level sets change their positions randomly at time instants $t_{k-1}, t_k, t_{k+l}, \ldots$.

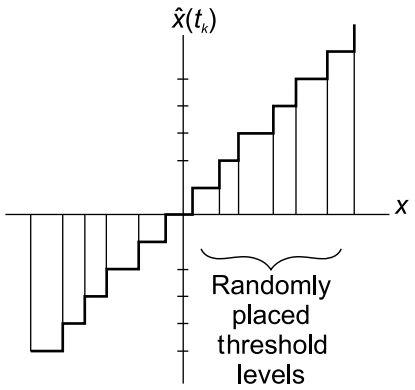


Figure 4.6 Instantaneous input-output characteristic of Model 1.

To ensure that the quantized signal value

$$\hat{x}_k = n_k q, \tag{4.8}$$

the interval q_{0k} between the zero level and the first threshold above it should be distributed uniformly within the range $[0, q]$. Note that the random interval q_{0k} at each quantization instant t_k determines the positions not only of the first but also of all the other evenly spaced threshold levels of the respective set. This means that only one random variable has to be generated and stabilized, which simplifies the implementation of this quantization method considerably. The quantization results in this case are, of course, unbiased.

Model 3

This model is also a version of Model 1, differing from it only by the definition of the reference value. Instead of relying on a relatively inconvenient basic measure $\bar{q}$, this quantization scheme relies on a reference that can be stabilized more accurately: a constant voltage level X representing the upper boundary of the signal range. The quantized signal in this case is given as

$$\hat{x}_k = \frac{n_k}{m_k} X, \tag{4.9}$$

where m_k is the number of threshold levels falling at the time instant t_k within the interval $[0, X]$. It follows from this equation that a quantization scheme such as this does not require the stability of the involved random processes. This is certainly a desirable feature, making this technique well suited for practical applications including short time interval measurements.

Note that the quantized signal values calculated according to Equation (4.5) are in fact biased. To confirm this, the quantized signal $\hat{x}$ can be considered as a function of the variables n and m. The Taylor expansion of this function around the point $(E[n], E[m-n])$ is

$$\hat{x} = \frac{E[n]}{E[m]}X + \frac{E[m-n]}{(E[m])^2}X\,(n - E[n]) - \frac{E[n]}{(E[m])^2}X\,(m - n - E[m-n]) + R, \tag{4.10}$$

where R is the remaining term.

$$E[\hat{x}] = \frac{E[n]}{E[m]}X + E[R] = x + E[R]. \tag{4.11}$$

This means that the bias of the quantised signal $\hat{x}$ is equal to $E[R]$.

Calculations show that in many cases this bias can be ignored because it is relatively small. For instance, when $\bar{q} = 0.5X$ and $\bar{q} = 0.2X$ the error due to this bias is no more than 3.5 % and 0.27 % of the maximal statistical error value respectively (99 % confidence level). Moreover, this bias error rather nontypically decreases when the quantized signal values are averaged.

It is shown in Chapter 19 that this randomized quantization model is especially well suited for measuring short time intervals with nanosecond or even picosecond resolution. At these applications, the quantizing stream of pulses generated at random time instants in accordance with the given relationships is used as an instrument for comparing the time interval being measured with a reference interval. The point is that the characteristics of the used random pulse sequence may drift around their nominal values without degradation of the measurement precision as long as the reference interval is kept stable.

4.3 Quantization Errors

Whatever quantization principle is applied, the following can always be written:

$$\begin{aligned} x(t) &= x_q(t) + \varepsilon(t) \\ \text{or} \qquad x(t) &= \hat{x}(t) + \varepsilon(t), \end{aligned} \tag{4.12}$$

where $\varepsilon(t)$ is quantization noise. An illustration of $\varepsilon(t)$ produced by the deterministic quantization of a signal is shown in Figure 4.2(b). When sampled signals are

quantized, expression (4.12) is slightly modified. Then

$$x_k = x_{qk} + \varepsilon_k \quad \text{or} \quad x_k = \hat{x}_k + \varepsilon_k \tag{4.13}$$

and the quantization noise is represented by the sequence of respective quantisation errors ε_k.

Quantization techniques are usually characterized not only by the time diagrams and input–output relationships described but also by the properties of the corresponding quantization errors or noise. In the case of deterministic quantizing, the errors $\{\varepsilon_k\}$ are usually assumed to be uniformly distributed in the range $[-0.5q, 0.5q]$. It is also assumed that

$$E[x_q] = x, \qquad E[\varepsilon(t)] = 0. \tag{4.14}$$

It follows from these assumptions that there are no bias errors and that statistical quantization errors have the following variance and standard deviation:

$$\text{Var}[\varepsilon(t)] = \frac{q^2}{12}, \qquad \sigma_\varepsilon = \frac{q}{2\sqrt{3}} \tag{4.15}$$

or, in terms of Equation (4.1),

$$\sigma_\varepsilon = \frac{A_m}{(2^B 2)\sqrt{3}}$$

These simple formulae are appropriate for quantizer evaluations in rough calculations and for other simplified calculations. For many multibit quantizer applications a more complicated analysis is often not required.

However, the assumption that the quantization errors are distributed uniformly does not hold in cases where rough quantization is applied or when signal amplitudes are comparable with the quantization step q. The quantization processes then need to be characterized in more detail.

4.3.1 Probability Density Function of Errors

Rough randomized quantization may often substitute multibit deterministic implementation of this operation. In many cases such randomized quantization proves to be more efficient because it allows the required level of precision to be achieved by processing a smaller volume of bits. However, at the same time the application success of this technique depends on the attention paid to details. Consider the vital error characteristics of such a rough randomized quantization

performed by using relatively few threshold levels. An essential characteristic is the probability density function of errors defining the distribution of the quantization errors of the quantization models.

Model 1

The probability density function for an input signal $x(t) \in [0, X]$ is denoted by $\varphi(x)$. Assume that at the quantization instant the signal is equal to x. Then the probability that there will be n threshold levels inside the interval $[0, x]$ is

$$\Pr[n_x = n] = P_n(x) - P_{n+1}(x), \tag{4.16}$$

where $P(n)$ is the probability distribution function of the nth threshold level. For any value of the signal, $x = n\bar{q} + \varepsilon$ can be written. Hence the probability density function of the quantization error is defined as

$$\Psi_1(\varepsilon) = \sum_{n=0}^{\infty} [P_n(n\bar{q} + \varepsilon) - P_{n+1}(n\bar{q} + \varepsilon)]\varphi(n\bar{q} + \varepsilon). \tag{4.17}$$

Note that the indices at $\Psi_1(\varepsilon)$ and other parameters following indicate the number of the corresponding quantization model considered. The expected value of the quantization error may be given as

$$\begin{aligned} E_1[\varepsilon] &= \int_{-\infty}^{\infty} \varepsilon \sum_{n=0}^{\infty} [P_n(n\bar{q} + \varepsilon) - P_{n+1}(n\bar{q} + \varepsilon)]\varphi(n\bar{q} + \varepsilon)\,\mathrm{d}\varepsilon \\ &= \int_{-\infty}^{\infty} \varepsilon \sum_{n=0}^{\infty} (x - n\bar{q})\,[P_n(x) - P_{n+1}(x)]\varphi(x)\,\mathrm{d}x. \end{aligned} \tag{4.18}$$

To simplify the equation it is worthwhile to introduce and use the following function:

$$\sum_{n=0}^{\infty} n[P_n(x) - P_{n+1}(x)] = H(x). \tag{4.19}$$

When $-X \Rightarrow \infty$, $H(x) = x/\bar{q}$. In addition,

$$\sum_{n=0}^{\infty} [P_n(x) - P_{n+1}(x)] = 1. \tag{4.20}$$

If these considerations are taken into account,

$$E_1[\varepsilon] = \int_0^{\infty} \left(x - \bar{q}\frac{x}{q}\right) \varphi(x)\,\mathrm{d}x = 0. \tag{4.21}$$

The variation of the quantization error

$$\begin{aligned}\mathrm{Var}_1[\varepsilon] = E_1[\varepsilon^2] &= \int_{-\infty}^{\infty} \sum_{n=0}^{\infty} (n\bar{q} - x)^2 \left[P_n(x) - P_{n+1}(x)\right]\varphi(x)\,\mathrm{d}x \\ &= \int_{-\infty}^{\infty} \bar{q}^2 \sum_{n=0}^{\infty} (n - E[n_x]\,q)^2 \left[P_n(x) - P_{n+1}(x)\right]\varphi(x)\,\mathrm{d}x.\end{aligned} \tag{4.22}$$

As it can be written that

$$\sum_{n=0}^{\infty} (n - E[n_x])^2 \left[P_n(x) - P_{n+1}(x)\right] = \mathrm{Var}[n_x], \tag{4.23}$$

Equation (4.22) may be given in the following form:

$$\mathrm{Var}_1[\varepsilon] = q^{-2} \int_0^{\infty} \mathrm{Var}[n_x]\varphi(x)\,\mathrm{d}x \tag{4.24}$$

On the basis of function (4.19), the variation $\mathrm{Var}[n_x]$ is defined as

$$\mathrm{Var}_1[n_x] = \frac{2}{\bar{q}} \int_0^x \left[H(u) - \frac{u}{q} + \frac{1}{2}\right]\mathrm{d}u. \tag{4.25}$$

Model 2
In this case $\varepsilon \in [-q, q]$ and the following equalities hold:

$$\text{For } \varepsilon < 0 \begin{cases} P_n(nq + \varepsilon) = 1 - \dfrac{|\varepsilon|}{q} \\ P_{n+1}(nq + \varepsilon) = 0 \end{cases}; \qquad \text{for } \varepsilon > 0 \begin{cases} P_n(nq + \varepsilon) = 1 \\ P_{n+1}(nq + \varepsilon) = \dfrac{|\varepsilon|}{q}. \end{cases} \tag{4.26}$$

The probability density function $\Psi_2(\varepsilon)$ of the quantization error can be obtained on the basis of these relationships. By analogy with Equation (4.17), it can be written that

$$\Psi_2(\varepsilon) = \sum_{n=0}^{\infty} \left(1 - \frac{|\varepsilon|}{q}\right) \varphi(nq + \varepsilon). \tag{4.27}$$

Equation (4.21) can be used to obtain the expected quantization error $E_2[\varepsilon]$. Following the previous reasoning, it is easy to show that

$$E_2[\varepsilon] = 0. \tag{4.28}$$

The general expression for $\mathrm{Var}_2[\varepsilon]$ is given by Equation (4.24). Substituting Equation (4.22) into Equation (4.24) yields

$$\mathrm{Var}_2[\varepsilon] = \sum_{n=0}^{\infty} \int_{nq}^{(n+1)q} (x - nq)\,(nq + q - x)\varphi\,(x)\,\mathrm{d}x$$

$$= \int_0^q n\,(q - u) \sum_{n=0}^{\infty} \varphi\,(nq + u)\,\mathrm{d}u. \tag{4.29}$$

A particular average probability density function of a quantized signal is now introduced:

$$\varphi_0\,(u) = \sum_{n=0}^{\infty} \varphi(nq + u). \tag{4.30}$$

This function is obtained by stacking all the subintervals of $\varphi(x)$ for $x \in [nq, (n+1)\,q]$ on the interval $[0, q]$ and subsequently summing them. When relatively many threshold levels are used for quantizing, $\varphi_0\,(u) \cong 1/q$ and it follows from Equation (4.29) that

$$\mathrm{Var}_2[\varepsilon] \cong \tfrac{1}{6}q^2 \tag{4.31}$$

The quantization error is characterized by this variance in cases where the probability density function for $\varepsilon \in [-q, q]$ is triangular. Indeed, substituting $\varphi_0(u)$ into Equation (4.27) gives

$$\Psi_2(\varepsilon) \cong \frac{q - |\varepsilon|}{q^2}. \tag{4.32}$$

It follows from Equations (4.17) and (4.27) that the probability density functions $\Psi_1(\varepsilon)$ and $\Psi_2(\varepsilon)$ in principle depend on the input signal x. When q decreases, i.e. when the number of threshold levels used for quantization increases, this dependence weakens and both of these functions tend to become triangular.

The probability density function $\Psi_2(\varepsilon)$ is illustrated in Chapter 5, where it is compared with analogous functions characterizing other quantization methods.

4.3.2 Variance of Randomly Quantized Signals

Two of the three quantization models considered here guarantee that the quantized signal values $\hat{x}$ are unbiased and the quantization bias errors in Model 3 are usually negligible. Hence the dominating quantization errors are random and are characterized by the variance of $\hat{x}$. Of course, these errors depend on the particular quantization conditions described by Models 1, 2 and 3.

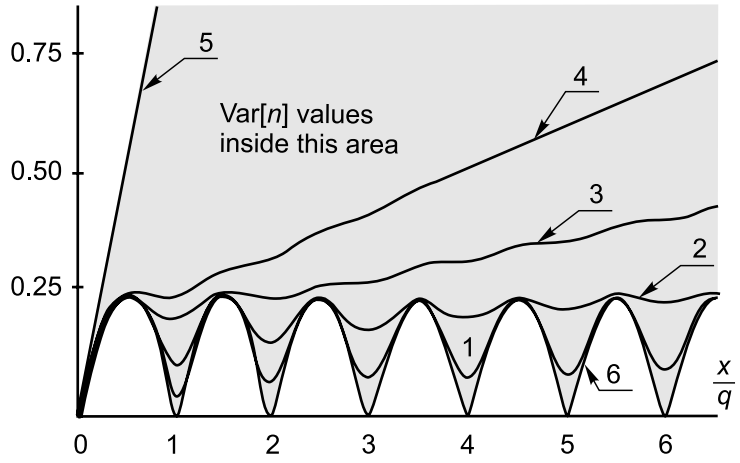


Figure 4.7 Plots of variances Var[n] versus normalized signals characterizing various quantization schemes

Model 1

When randomized quantization is performed as defined by Model 1, $\hat{x} = n\bar{q}x$ and

$$\mathrm{Var}_1[\hat{x}] = \bar{q}\ \mathrm{Var}_1[n]. \tag{4.33}$$

$\mathrm{Var}_1[n]$ is defined by Equation (4.25), which shows that this variation depends on the probability density function of the signal x, as well as on the standard deviation of the random intervals $\{q_{ij} - q_{i(k-1)}\}$. The statistical quantization errors are apparently less significant for smaller values of $\sigma/\bar{q}$. The dependence of $\mathrm{Var}_1[n]$ on $x/\bar{q}$ is illustrated in Figure 4.7. These diagrams have been calculated on the basis of Equation (4.25). Curves 1 to 4 correspond to the randomized quantization carried out under the condition that the probability density functions of the intervals between the threshold levels are normal. They are given for $\sigma/\bar{q} =$ 0.3, 0.2, 0.1 and 0.05 respectively. Line 5 corresponds to extremely randomized quantization when the ordinates of the threshold levels at each time instant t_k correspond to the Poisson point streams. Thus this line is really a boundary on the left-hand side. All other possible relationships characterizing randomized quantization performed according to Model 1 are to the right or below it. At this extreme quantization mode

$$\mathrm{Var}_1[n] = E[n] = \frac{x}{q}. \tag{4.34}$$

Model 2
Quantization, as carried out by this model, is characterized by curve 6. In this case, the variance of $\hat{x}$ is defined as

$$\mathrm{Var}_2[\hat{x}] = q^2\ \mathrm{Var}[n]. \tag{4.35}$$

For $x \in [nq, (n+1)q]$ the function $H(x) = n$. In this case, it follows from Equation (4.25) that

$$\mathrm{Var}_2[n] = \left(\frac{x}{q} - n\right)\left[1 - \left(\frac{x}{q} - n\right)\right] = -\left[\left(\frac{x}{q} - n\right) - \frac{1}{2}\right]^2 + \frac{1}{4}. \tag{4.36}$$

Therefore, under these conditions, the variances of the quantized signals are represented by parabolas assuming zero values at points where $x = nq$.

Model 3
Variance $\mathrm{Var}_3[\hat{x}]$, characterizing randomized quantization executed in accordance with Model 3, obviously depends on $\mathrm{Var}_3[m]$, $\mathrm{Var}_3[n]$, $\mathrm{Var}_3[m-n]$. All of these can be calculated from Equation (4.25). It follows from Equation (4.10) that

$$\mathrm{Var}_3[\hat{x}] = \frac{X^2}{E^4[m]}\{(E[m-n])^2\,\mathrm{Var}_3[n] + (E[n])^2\,\mathrm{Var}[m-n] - 2E[n]\,E[m-n]\,K_{n,(m-n)}\}.$$

As the covariance

$$K_{n,(m-n)} = \tfrac{1}{2}\,(\mathrm{Var}_3[m] - \mathrm{Var}_3[n] - \mathrm{Var}_3[m-n]),$$

then

$$\mathrm{Var}_3[\hat{x}] = \frac{x\,(X-x)\,\bar{q}^2}{X}\left(\frac{\mathrm{Var}_3[n]}{x} + \frac{\mathrm{Var}_3[m-n]}{X-x} - \frac{\mathrm{Var}_3[m]}{X}\right). \tag{4.37}$$

4.4 Quantization Noise

Time sequences of quantization errors are known as quantization noise. When the quantization operation is performed on a sampled signal, quantization noise is a discrete-time random process. The properties of this noise determine the usefulness of the corresponding quantization scheme. For a further comparison of the deterministic and randomized quantization approaches this section will consider the respective quantization noises. It is assumed that randomized quantization is

carried out according to Model 1. Realizations of the quantization noise characterizing deterministic and randomized quantization are given in Chapter 5. It is clear that the various types of quantization noise will differ considerably.

The noise of deterministic quantization varies within the range $[-0.5q, +0.5q]$ and during some time intervals the envelopes of this noise repeat respective segments of the signal $x(t)$. The polarity of the errors is determined only by the input signal values. The range of randomized quantization noise is twice as large, i.e. $[-q, +q]$. The envelopes of this noise also repeat the corresponding parts of the signal, although particular errors can assume either positive or negative values at random, i.e. the value of the error is either equal to $\varepsilon_k^{(+)}$ or to $\varepsilon_k^{(-)}$. At each quantization instant $\varepsilon_k^{(+)} + \varepsilon_k^{(-)} = q$. The appearance of the larger error is less probable. It can be written that

$$\frac{\Pr\left[\varepsilon_k = \varepsilon_k^{(+)}\right]}{\Pr\left[\varepsilon_k = \varepsilon_k^{(-)}\right]} = \frac{\varepsilon_k^{(-)}}{\varepsilon_k^{(+)}}, \tag{4.38}$$

where $\Pr[\varepsilon_k = \varepsilon_k^{(+)}]$ and $\Pr[\varepsilon_k = \varepsilon_k^{(-)}]$ are the probabilities of the error ε_k assuming positive or negative values respectively. Of course,

$$\Pr\left[\varepsilon_k = \varepsilon_k^{(+)}\right] + \Pr\left[\varepsilon_k = \varepsilon_k^{(-)}\right] = 1.$$

This uncertainty in error polarity, typical of randomized quantization, is beneficial because it decorrelates the input signal and its corresponding quantization noise.

4.4.1 Covariance between the Signal and Quantization Noise

Since the envelopes of quantization noise follow the corresponding parts of the input signal waveform, it seems that quantization noise is statistically dependent on the signal. To assess the degree of dependence between these two processes, consider a randomized quantization performed in accordance with the quantization Model 2. Statistical relationships of this kind are more pronounced in coarse quantization. Therefore, without loss of generality, one threshold randomized quantization scheme can be considered with $x(t) \in [0, q]$.

If at the quantization instant t_k the input signal is x_k, the quantized signal $\hat{x}_k$ can be either 0 or q. Hence the quantization error $\varepsilon_k = \hat{x}_k - x_k$ is either equal to $-x_k$ or to $(q - x_k)$. The probabilities of the respective events are

$$\begin{aligned} \Pr[\varepsilon_k = -x_k] &= \frac{q - x_k}{q}, \\ \Pr[\varepsilon_k = (q - x_k)] &= \frac{x_k}{q}. \end{aligned} \tag{4.39}$$

It can be written that the covariance between ε_k and x_k is

$$\begin{aligned}\operatorname{cov}(\varepsilon_k, x_k) &= E[\varepsilon_k x_k] \\ &= \int_0^q \left[-x_k x_k \frac{q - x_k}{q} + (q - x_k) x_k \frac{x_k}{q}\right] p(x_k)\, \mathrm{d}x_k = 0. \end{aligned} \tag{4.40}$$

This result is unexpected. Indeed, as can be seen from Equations (4.26), there is a strong dependence of ε_k on x_k. It is therefore hard to accept that the corresponding covariance is equal to zero. Nevertheless, the covariance between ε_k and x_k is equal to zero.

4.4.2 Spectrum

The extremely coarse quantization scheme is also convenient for determining the spectral characteristics of randomized quantization. Assuming that quantizing is performed according to Model 2 and taking into account Equation (4.39), the autocovariance function of quantization noise for all $m \neq 0\,(m = 0, 1, 2, \ldots)$ is given as

$$\begin{aligned} C_{\varepsilon\varepsilon}(mT) = \int_0^q \int_0^q \Bigg[& x_k x_{k+m} \frac{(q - x_k)(q - x_{k+m})}{q^2} \\ & - x_k (q - x_{k+m}) \frac{(q - x_k) x_{k+m}}{q^2} - (q - x_k)\, x_{k+m} \frac{x_k (q - x_{k+m})}{q^2} \\ & + (q - x_k)(q - x_{k+m})\, \frac{x_k x_{k+m}}{q^2} \Bigg] p(x_k, x_{k+m})\, \mathrm{d}x_k\, \mathrm{d}x_{k+m} = 0, \end{aligned} \tag{4.41}$$

where $p(x_k, x_{k+m})$ is the two-dimensional probability density function. It follows from this equation that

$$C_{\varepsilon\varepsilon}(mT) = \begin{cases} \sigma_\varepsilon^2 & \text{for } m = 0, \\ 0 & \text{for } m \neq 0, \end{cases} \tag{4.42}$$

where σ_ε^2 is the variance of the quantization noise. This result also holds for multithreshold randomized quantization. Of course, the appropriate variance characterizing the particular quantization noise should be substituted into Equation (4.42).

It follows from Equation (4.42) that the spectral density function of quantization noise $\{\varepsilon(k\Delta t)\}$ is given as

$$G_{\varepsilon\varepsilon}(f) = 2TC_{\varepsilon\varepsilon}(0) + 4T\sum_{m=1}^{\infty} C_{xx}(mT)\cos(2\pi f mT) = 2\sigma_{\varepsilon}^2 T \tag{4.43}$$

Note that the covariance between the signal and quantization noise, as well as the autocovariance function of this noise, is obtained under the assumption that the signal is sampled periodically. Under this condition and taking into account the sampling theorem,

$$G_{\varepsilon\varepsilon}(f) = \begin{cases} 2\sigma^2/f & \text{for } f \le f/2, \\ 0 & \text{for } f > f_s/2. \end{cases} \tag{4.44}$$

As quantization Models 1 and 2 show, quantizing according to Model 1 is randomized to a greater degree than quantizing as defined by Model 2. The randomization level for Model 1 is characterized by the ratio $\sigma/\bar{q}$. When the value of this ratio decreases, the correlation between quantizing errors increases. However, even at $\sigma/\bar{q} = 0$, the quantization noise is white; hence it is also white for $\sigma/\bar{q} > 0$ and Equation (4.44) is also true for quantizing performed according to Model 1.

Bibliography

Bennet, W.R. (1948) Spectra of quantized signals. *Bell Syst. Tech. J.*, **27**(7), 446–72.

Bilinskis, I. (1976) Stochastic signal quantization error spectrum (in Russian). *Autom. Control Comput. Sci.*, **3**, 55–60.

Bilinskis, I. and Mikelsons, A. (1992) *Randomized Signal Processing*. Prentice-Hall International (UK) Ltd.

Brannon, B. (1996) Wide-dynamic-range A/D converters pave the way for wideband digital-radio receivers. In EDN, 7 November 1996, pp. 187–205.

Brannon, B. Overcoming converter nonlinearities with dither. *Analog Devices*, **AN-410**, Application Note.

Brodsky, P.I., Verjushsky, V.V., Klyushnikov, S.N. and Lutchenko, A.E. (1970) Statistical quantization technique using auxiliary random signals (in Russian). *Voprosy Radioelektroniki Ser. obshchetekhnic*, **3**, 46–53.

Corcoran, J.J., Poulton, K. and Knudsen, K.L. (1988) A one-gigasample-per-second analog-to-ditial converter. *Hewlett-Packard J.*, **June**, 59–66.

Demas, G.D., Barkana, A. and Cook, G. (1973) Experimental verification on the improvement of resolution when applying perturbation theory to a quantizer. *IEEE Trans. Industr. Electron. Control Instr.*, **IECI-20**(4), 236–9.

Gedance, A.R. (1972) Estimation of the mean of a quantized signal. *Proc. IEEE*, **60**(8), 1007–8.

Jayant, N. and Rabiner, L. (1972) The application of dither to the quantization of speech signals. *Bell Syst. Tech. J.*, **51**(6), 1293–304.

Rabiner, L. and Johnson, J. (1972) Perceptual evaluation of the effects of dither on low bit rate PCM systems. *Bell Syst. Tech. J.*, **51**(7), 1487–94.

Veselova, G.P. (1975) On amplitude quantization with overlapping of interpolating signals (in Russian). *Avtomatika i Telemekhanika*, **5**, 52–9.

5

Pseudo-randomized Quantizing

Some of the effects caused by the deliberate injection of randomness into signal digitizing procedures are beneficial and some are harmful, increasing statistical errors. Fortunately, the targeted benefits are often obtained at low levels of randomization. Therefore the randomization level should be controlled to suppress additional errors and excessive randomization must, of course, be avoided in all cases. Yet another useful approach to obtaining the desirable effects is substitution of randomization by making deterministic signal digitization processes simply irregular. Such an approach is less harmful and usually leads to obtaining the same positive effects that are expected from the randomization. Pseudo-randomization of quantizing, considered in this chapter, actually represents such an irregularization. Pseudo-randomized quantizing is evidently a fully deterministic process. A description of it in probabilistic and statistical terms is a convenient way to explain that. It is further shown that skilful application of pseudo-randomization techniques leads to quite good results.

5.1 Pseudo-randomization Approach

It may be a little confusing to understand where the line separating randomized and pseudo-randomized quantization should be drawn. To randomize or pseudo-randomize quantizing, an element of randomness either has to be added at the input of the used ADC or it has to be used for randomizing the positions of the reference threshold levels used for quantizing, as shown in Figures 5.1(a) and (b)

Digital Alias-free Signal Processing I. Bilinskis

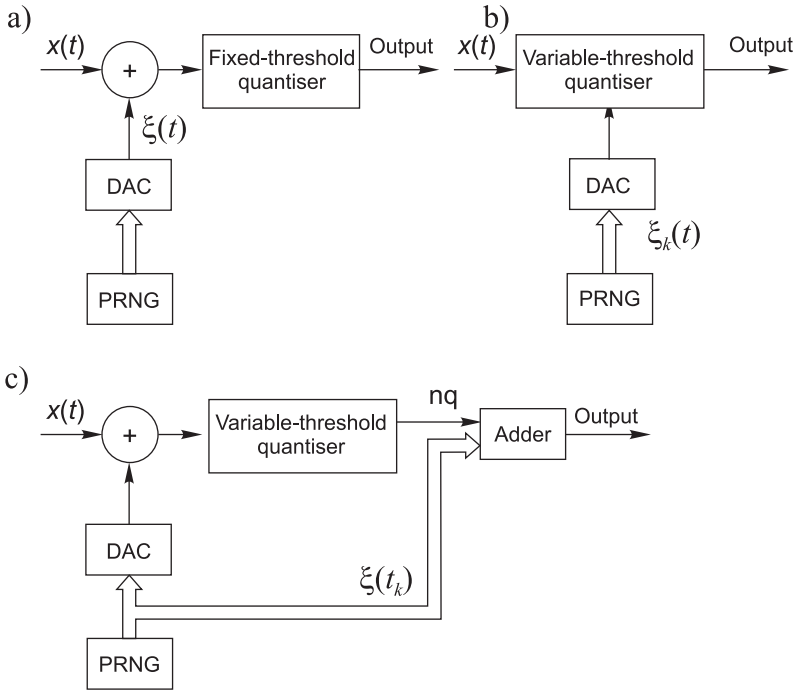


Figure 5.1 Quantizers illustrated by schemes (a) and (b) are randomized and the pseudo-randomized quantizer is illustrated by scheme (c)

respectively. Pseudo-random number generators (PRNGs) with digital-to-analog converters (DACs) added are typically used for generation of the analog auxiliary pseudo-random processes needed for that. At first glance all three quantizers shown in Figure 5.1 look equivalent. However, there are significant differences between the first two schemes and the third, shown in Figure 5.1(c). While both of the first two schemes depict a randomized quantizer, the third diagram is a scheme of a pseudo-randomized quantizer.

In all of the illustrated cases the quantizing operation is based on comparing signal instantaneous values with variable thresholds generated by a PRNG together with a DAC. When the quantizing operation is performed in this way, the positions of the threshold levels at all $\{t_k\}$ are in fact known. Whether this information is used or not is another question, but it is there, given at each t_k even in the digital form. Evidently, this information is not used at randomized quantization when the quantized signal is simply defined as $n_k q$.

The quantizer shown in Figure 5.1(c) differs distinctly from the other schemes illustrated. There are two components in the output. The quantized signal values provided are defined as a function depending both on n_k and ξ_k (or q_{0k}). Evidently the auxiliary pseudo-random process in this case is used in quite a different way. It is added to the input signal (this is equivalent to shifting the set of reference threshold levels up or down) and then the same process is used to form the output signal values. As shown below, such a pseudo-randomized quantizer has outstanding properties that cannot be obtained either from deterministic or randomized quantizers. It has the advantages of randomized quantizers without their basic disadvantages.

The question, of course, is exactly how the ξ_k values should be used in the definition of the pseudo-randomized quantization output. To find the answer to this question, an attempt will be made to discover the conditions under which the pseudo-ramdomized quantization operation might be considered to be optimal.

5.2 Optimal Quantizing

Quantizing will be considered optimal if this operation provides the best conditions for estimating the mean value m_x of a signal $x(t) \in [0, X]$. Without loss of generality, an analysis of optimal quantization can be carried out under the assumption that $x(t) = x =$ constant and that only a single threshold level is used for quantization.

5.2.1 Single-threshold Quantizing

Consider pseudo-randomized quantizing with the information provided by $\{\xi_k\}$ taken into account. The quantized signal then can be defined as

$$\hat{x}_k = \begin{cases} f_1(\xi_k) & \text{for } \xi_k \le x_k, \\ f_2(\xi_k) & \text{for } \xi_k > x_k, \end{cases} \tag{5.1}$$

and the estimate of m_x as

$$\hat{m}_x = \frac{1}{N} \sum_{k=1}^{N} \hat{x}_k.$$

Such quantizing should be performed in a way ensuring that the estimate $\hat{m}_x$ is unbiased, i.e. that the following equality holds:

$$E[\hat{m}_x] = E[\hat{x}_k] = x. \tag{5.2}$$

Then

$$\begin{aligned} E[\hat{x}_k] &= \int_0^x f_1(\xi_k)\frac{\mathrm{d}\xi_k}{X} + \int_x^x f_2(\xi_k)\frac{\mathrm{d}\xi_k}{X} \\ &= \frac{1}{X}[F_1(x) - F_1(0) + F_2(X) - F_2(x)] \\ &= x \end{aligned} \tag{5.3}$$

where

$$F_1(x) = \int f_1(x)\,\mathrm{d}x, \qquad F_2(x) = \int f_2(x)\,\mathrm{d}x.$$

It follows from Equation (5.3) that the functions $F_1(x)$ and $F_2(x)$ should meet the conditions

$$\begin{aligned} F_1(x) - F_2(x) &= xX, \\ -F_1(0) + F_2(X) &= 0. \end{aligned} \tag{5.4}$$

Any functions $f_1(\xi)$ and $f_2(\xi)$ that satisfy Equations (5.4) can be applied in order to define the quantized signal according to Equation (5.1), and all of them will provide unbiased estimates of x. Assume that $f_1(\xi)$ and $f_2(\xi)$ are linear functions

$$\begin{aligned} f_1(\xi) &= a_1\xi + a_0, \\ f_2(\xi) &= b_1\xi + b_0. \end{aligned} \tag{5.5}$$

In this case,

$$\begin{aligned} F_1(\xi) &= \tfrac{1}{2}a_1\xi^2 + a_0\xi + C_1, \\ F_2(\xi) &= \tfrac{1}{2}b_1\xi^2 + b_0\xi + C_2. \end{aligned} \tag{5.6}$$

As can be seen from equation (5.3), $E[\hat{x}_k]$ does not depend on the constants C_1 and C_2 and, therefore, they can be considered as equal to zero. Substituting Equation (5.6) into Equation (5.4) gives

$$\begin{aligned} &\tfrac{1}{2}a_1x^2 + a_0x - \tfrac{1}{2}b_1x^2 - b_0x = xX, \\ &\tfrac{1}{2}b_1X^2 + b_0X = 0. \end{aligned} \tag{5.7}$$

The first of these equations is satisfied for any x if the following conditions are met:

$$\begin{aligned} a_1 &= b_1, \\ a_0 - b_0 &= X. \end{aligned} \tag{5.8}$$

Taking these expressions into account, it follows from Equation (5.7) that

$$a_0 = \left(1 - \tfrac{1}{2}a_1\right) X,$$
$$b_0 = -\tfrac{1}{2}a_1 X.$$

The quantized signal can be defined as follows:

$$\hat{x}_k = \begin{cases} f_1(\xi_k) = a_1\xi_k + \left(1 - \tfrac{1}{2}a_1\right) X & \text{for } \xi_k \le x_k \\ f_2(\xi_k) = a_1\xi_k - \tfrac{1}{2}a_1 X & \text{for } \xi_k > x_k \end{cases} \tag{5.9}$$

Definition (5.9) holds for a_1 varying within large margins. For instance, when $a_1 = 0$, this definition describes randomized quantization. Although the estimate $\hat{m}_x$ with a_1 varying remains unbiased, other properties of the quantized signal, including the random error, of course depend on a_1. It is therefore essential to optimize such quantization or, in other words, to find the value of a_1 that provides the minimum of some criterion K.

As the variance $\mathrm{Var}[\hat{x}]$ in general depends on x, optimization can be carried out with respect to

$$K = \int_0^X \mathrm{Var}[\hat{x}]\,\mathrm{d}x. \tag{5.10}$$

It can be written that

$$\begin{aligned} \mathrm{Var}[\hat{x}] &= \int_0^x f_1^2(\xi)\frac{\mathrm{d}\xi}{X} + \int_x^X f_2^2(\xi)\frac{\mathrm{d}\xi}{X} - x^2 \\ &= \tfrac{1}{12}a_1^2X^2 + (1 - a_1)(X - x)x. \end{aligned} \tag{5.11}$$

Substituting Equation (5.11) into Equation (5.10) gives

$$K = \tfrac{1}{12}X^3\left(a_1^2 - 2a_1 + 2\right)$$

and it is clear that the minimum $K_{\min} = \frac{1}{12}X^3$ is achieved at $a_1 = 1$. Thus pseudo-randomized quantizing provides the minimum of the variance $\mathrm{Var}[\hat{x}]$ averaged over x if the quantized signal is defined in the following way:

$$\hat{x}_k = \begin{cases} f_1(\xi_k) = \xi_k + \tfrac{1}{2}X & \text{for } \xi_k \le x_k, \\ f_2(\xi_k) = \xi_k - \tfrac{1}{2}X & \text{for } \xi_k > x_k. \end{cases} \tag{5.12}$$

It can be seen from Equation (5.10) that such quantization is characterized by

$$\mathrm{Var}[\hat{x}] = \tfrac{1}{12}X^2. \tag{5.13}$$

What is important is that this variance does not depend on x. This remarkable property makes the quantization model considered here exceptional. All other

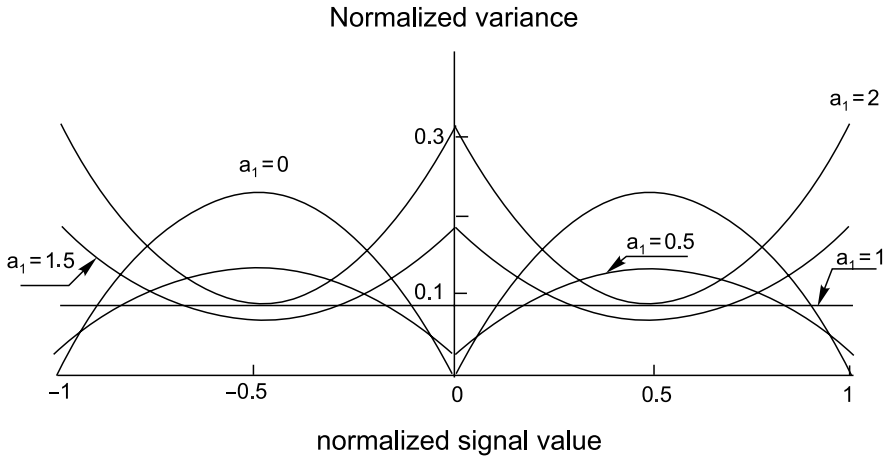


Figure 5.2 Variances of a signal quantized pseudo-randomly in various ways

quantization models defined by Equation (5.9) are characterized by $\mathrm{Var}[\hat{x}]$ depending on x, as shown in Figure 5.2. The curves given there are for the normalized signal values varying within the range of $(-1, 1)$ and five quantization models corresponding to various values of $0 \leq a_1 \leq 2$.

As optimization of the pseudo-randomized quantization is carried out under the assumption that the functions $f_1(\xi_k)$ and $f_2(\xi_k)$ are linear, it is not clear whether the results are also optimal when these functions are nonlinear. This question has been investigated and it was found that the quantization performed in accordance with equation (5.12) is also optimal when the functions $f_1(\xi_k)$ and $f_2(\xi_k)$ are nonlinear.

5.2.2 Multithreshold Quantizing

Since the quantized signals obtained in the course of pseudo-randomized quantizing are unbiased, no matter how coarse the quantization procedure, all essential relationships characterizing such quantizing are the same both for a single and a multithreshold quantization. In the latter case, it follows from Equation (5.12) that

$$\hat{x}_k = q_{0k} - \tfrac{1}{2}q + n_k q, \quad q_{0k} \in [0, q]. \tag{5.14}$$

It is sometimes more convenient to use another version of Equation (5.14):

$$\hat{x}_k = \left(\xi_k - \tfrac{1}{2}\right) q + n_k q, \qquad \xi_k \in [0, 1]. \tag{5.15}$$

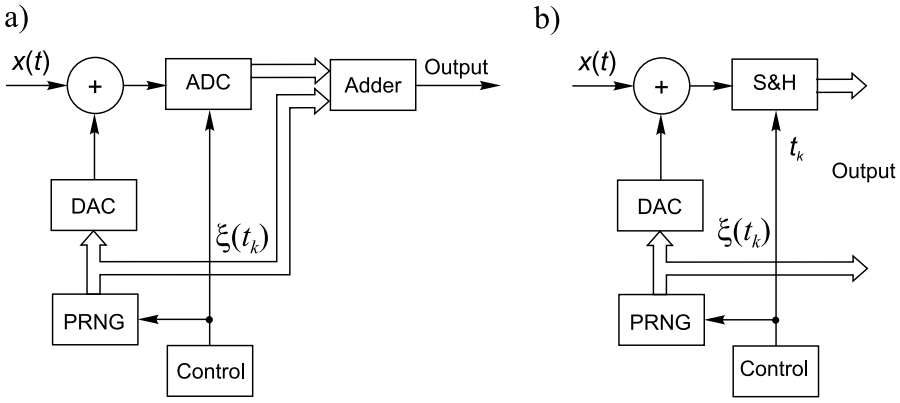


Figure 5.3 Schemes for realizing pseudo-randomized quantizing

As mentioned above, such quantizing is unbiased. Indeed,

$$E[\hat{x}_k] = qE[n_k] + E[q_{0k}] - \tfrac{1}{2}q$$
$$= q\frac{x_k}{q} + \tfrac{1}{2}q - \tfrac{1}{2}q = x_k.$$

Although in this case the reference threshold sets are also formed as described by the quantization Model 2, there is a considerable difference between the quantization performed according to this model and the pseudo-randomized quantization. The quantized signal defined by Equations (5.14) or (5.15) also assumes values, which belong to some discrete levels, but in this case intervals between them are small and equal to the smallest digit of q_{0k}.

5.2.3 Implementation Approaches

To implement the pseudo-randomized quantization method, it is not necessary to design a completely new type of analog-to-digital converter. Conventional deterministic ADCs can be used to carry out this kind of quantizing, as shown in Figure 5.3. Note that the ADC used should be a high-speed device with a sufficiently small aperture time. To perform pseudo-randomized quantizing according to the scheme given in Figure 5.3(a), the auxiliary random process generated by the pseudo-random number generator (PRNG) is converted by a DAC and the analog random process obtained is added to the input signal. The ADC converts this additive mixture and the digital random variables $\{\xi_k\}$ are subtracted from the respective values of the ADC's output signal. Thus one and the same

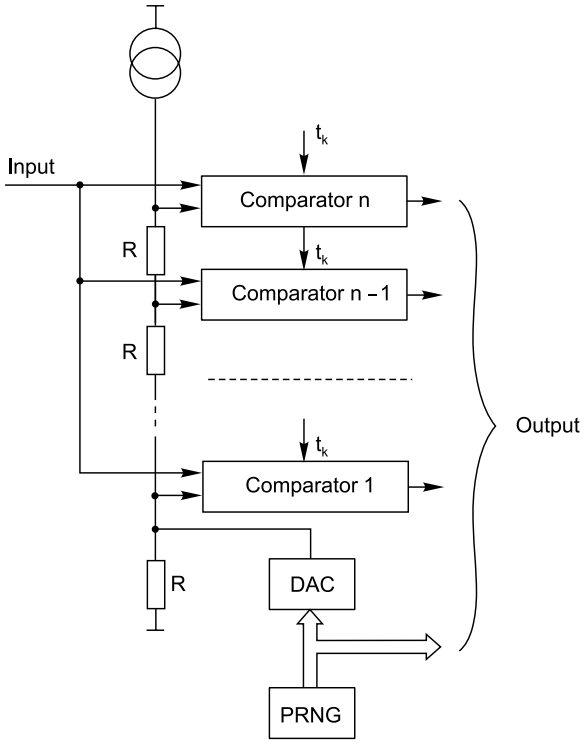


Figure 5.4 A scheme for pseudo-randomized quantising based on application of a flash converter

random process is added to the input signal and later subtracted from the quantized signal.

The operational principle of this scheme is the same as for the optimized quantization mode considered above. When the ADCs used are relatively slow, the scheme in Figure 5.3(a) can be modified as shown in Figure 5.3(b) by adding a sample-and-hold (S&H) circuit.

Another implementation approach is based on the principle of changing the reference threshold levels during the quantization process. A scheme realizing this principle is shown in Figure 5.4. The most vital elements in it are the voltage comparators strobed by short pulses at the sampling time instants $\{t_k\}$. The input signal is fed to all inputs of these comparators. The reference inputs are connected to a chain of resistors. Current from a constant current source flowing through this chain of resistors determines the reference voltages or the threshold levels. They depend not only on this current but also on the random voltage across the lowest in the chain resistor R_1, generated by the PRNG in connection with the DAC.

The threshold levels at all the comparator reference inputs follow random voltage changes, thus providing the correct operational mode for pseudo-randomized quantizing.

Note that the schemes in Figures 5.3(b) and 5.4 have two rather than one output. This is essential. Output signals of such quantizers can often be processed in two differing ways. They might be passed partly through one processing channel and partly through another. Separate algorithms are used for processing the information carried by the sequence $\{n_k q\}$ and the information given by the sequence $\{\xi_k\}$. It is possible to speed up processing and to simplify the hardware in this way, lessening the negative impact of the main disadvantage of this kind of quantizing: the increased quantity of bits needed for encoding the pseudo-randomly quantized signals.

5.3 Input–Output Relationships

Both instantaneous and expected input–output relationships for pseudo-randomized quantizing are linear. The point is that the input–expected output relationship is defined in a remarkable way, specifically $E\,[\hat{x}] = x$. Instantaneous values of $\hat{x}$ for any value of x all are within a certain area. This means that for all values of x, the respective $\hat{x}$ values are distanced from the corresponding values of $E[\hat{x}]$ by an interval not exceeding $\pm 0.5q$. For instance, when $x = x_0$, the quantized value $\hat{x}_0$ can with equal probability assume any value within the interval $[x_0 \pm 0.5q]$.

5.4 Quantization Errors

Assume that the quantized signal

$$\hat{x} = nq + q_0 - \tfrac{1}{2}q. \tag{5.16}$$

Then the conditional probability density function

$$\Psi(\varepsilon/\hat{x} = x_0) = \frac{\mathrm{d}q_0 \varphi(\hat{x}_0 + \varepsilon)}{q\ \Pr[\hat{x} = x_0]}, \tag{5.17}$$

where $\Pr[\hat{x} = x_0]$ is the probability that the quantized signal $\hat{x}$ is equal to x_0. It is obvious that

$$q\ \Pr[\hat{x} = x_0] = \frac{\mathrm{d}q_0}{q} \int_{(n-1)q+q_0}^{nq+q_0} \varphi(x)\,\mathrm{d}x.$$

To obtain the unconditional probability density function of ε, function (5.17) should be averaged over all possible digital values of $n = 0, 1, 2, \ldots$ and over all

values of q_0 that are considered to be continuous. Then

$$\Psi(\varepsilon) = \Pr[\hat{x} = x_0]\Psi(\varepsilon/\hat{x} = x_0)$$

$$= \frac{1}{q}\int_0^q \sum_n \varphi(\hat{x}_0 + \varepsilon)\,\mathrm{d}q_0. \tag{5.18}$$

The sum can be expressed by means of δ functions. This can be written as

$$\sum_n \varphi\left(nq + q_0 - \tfrac{1}{2}q + \varepsilon\right) = \int_{-\infty}^{\infty} \varphi(x) \sum \delta\left(x - nq - q_0 + \tfrac{1}{2}q - \varepsilon\right)\mathrm{d}x. \tag{5.19}$$

The Fourier transform of the sum of δ functions can be shown to be

$$\sum_n \delta\left(x - nq - q_0 + \tfrac{1}{2}q - \varepsilon\right) = \frac{1}{q}\sum_n \exp\left[\mathrm{i}\frac{2\pi}{q}n\left(x + \tfrac{1}{2}q - q_0 - \varepsilon\right)\right]. \tag{5.20}$$

Substituting Equation (5.20) into Equation (5.19) and substituting the result into Equation (5.18) gives

$$\Phi(\varepsilon) = \frac{1}{q^2}\int_0^q \mathrm{d}q_0 \int_{-\infty}^{\infty} \varphi(x) \sum_n \exp\left(\mathrm{i}\frac{2\pi}{q}n\left(x - q_0 + \tfrac{1}{2}q - \varepsilon\right)\right)\mathrm{d}x. \tag{5.21}$$

Taking into account the fact that

$$\int_{\infty}^{\infty} \varphi(x) \sum_n \exp\left(\mathrm{i}\frac{2\pi}{q}nx\right)\mathrm{d}x = \tilde{\varphi}\left(\frac{2\pi}{q}n\right),$$

where $\tilde{\varphi}\left((2\pi/q)\,n\right)$ is the value of the signal characteristic function at the argument equal to $(2\pi/q)\,n$, and that

$$\exp\left(\mathrm{i}\frac{2\pi}{q}n\frac{q}{2}\right) = (-1)^n,$$

Equation (5.21) can be rewritten as

$$\Phi(\varepsilon) = -\frac{1}{q^2}\sum_n \tilde{\varphi}\left(\frac{2\pi}{q}n\right)(-1)^n \exp\left(-\mathrm{i}\frac{2\pi}{q}n\varepsilon\right)\int_0^q \exp\left(\mathrm{i}\frac{2\pi}{q}nq_0\right)\mathrm{d}q_0. \tag{5.22}$$

Note that

$$\int_0^q \exp\left(-1\frac{2\pi}{q}nq_0\right)\mathrm{d}q_0 = \begin{cases} q & \text{for } n = 0, \\ 0 & \text{for } n \neq 0, \end{cases}$$

and $\tilde{\varphi}(0) = 1$. Finally,

$$\Phi(\varepsilon) = \frac{1}{q}, \qquad \varepsilon \in [-0.5q, +0.5q]. \tag{5.23}$$

This result leads to the following very important conclusion:

No matter how coarse the pseudo-randomized quantizing is, quantization errors do not depend on the probability density function of the signal being quantized and they are always distributed uniformly in the range $[-0.5q, +0.5q]$.

Consequently,

$$E[\varepsilon] = 0, \qquad \mathrm{Var}[\varepsilon] = \tfrac{1}{12}q^2. \tag{5.24}$$

Independence of the quantization errors from the input signal simplifies the task of deriving the variance of quantized signals considerably. In this case, it is obvious that

$$\mathrm{Var}[\hat{x}] = \mathrm{Var}[x(t)] + \mathrm{Var}[\varepsilon] = \mathrm{Var}[x(t)] + \tfrac{1}{12}q^2. \tag{5.25}$$

5.5 Quantization Noise

A realization of pseudo-randomized quantization noise is shown in Figure 5.5(d). To obtain an impression of the main properties of this noise, compare it with the deterministic and randomized quantization noises, also given in Figures 5.5(b) and (c).

Comparison of these time diagrams leads to the following conclusions:

1. Pseudo-randomized quantization noise is distributed in an interval $[-0.5q, +0.5q]$. Therefore, the quantization noise for the pseudo-randomized quantizing does not exceed the boundaries characterizing deterministic quantizing and is distributed in an interval twice smaller than in the case of the noise observed at randomized quantizing of the same signal.
2. An error of pseudo-randomized quantizing at any quantization instant t_k assumes with equal probability any value from the interval $[-0.5q, +0.5q]$.
3. There is no (visible) dependence of the pseudo-randomized quantization noise on the signal being quantized.

The latter conclusion, of course, cannot be based on visual observation. This feature of the pseudo-randomized quantization procedure has to be proved.

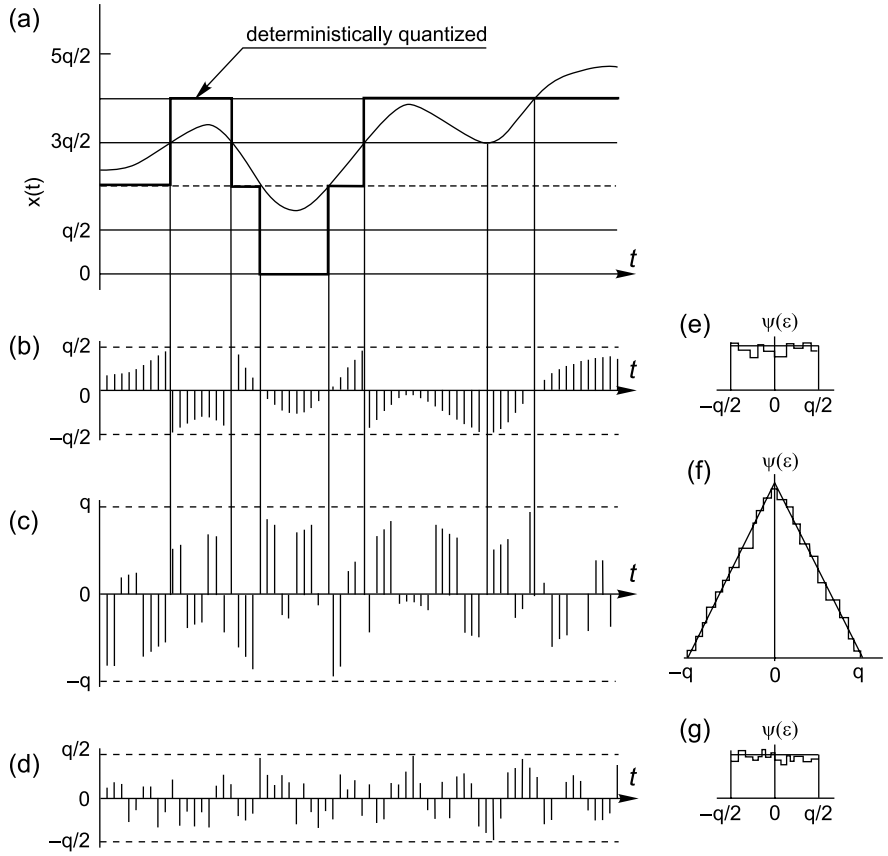


Figure 5.5 Quantizing of a signal in three different ways: (a) input signal with quantization noise and its probability density function; (b), (e) for deterministic quantizing; (c), (f) for randomized; and (d), (g) for pseudo-randomized quantizing

Thus it turns out that pseudo-randomized quantizing is unique. No other known quantization technique has such advantageous properties. At the same time, as might be expected, this quantization approach also has its drawbacks. The most serious disadvantage is the increased number of bits required for representing the quantized signals. At first glance it seems that as a result of this disadvantage the hardware for processing signals quantized in this way should be considerably more complicated than usual. Fortunately, this is not so. As shown in the following chapters, if signal processing is organized on the basis of special algorithms, well matched to the specifics of pseudo-randomized quantizing, the main disadvantage of this kind of quantizing becomes less significant.

5.5.1 Covariance between Signal and Quantization Noise

The quantization noise $\{\varepsilon_k\}$ is divided into two processes, $\{\varepsilon_k^{(+)}\}$ and $\{\varepsilon_k^{(-)}\}$. The first process contains only zeroes and positive errors and the second process, similarly, contains zeroes and negative errors. Formally, it can be written that

$$\varepsilon_k^{(+)} = \begin{cases} \varepsilon_k & \text{for } \varepsilon_k > 0, \\ 0 & \text{for } \varepsilon_k \leq 0, \end{cases} \tag{5.26}$$

$$\varepsilon_k^{(-)} = \begin{cases} \varepsilon_k & \text{for } \varepsilon_k < 0, \\ 0 & \text{for } \varepsilon_k \geq 0, \end{cases} \tag{5.27}$$

and

$$\{\varepsilon_k\} = \{\varepsilon_k^{(+)}\} + \{\varepsilon_k^{(-)}\}. \tag{5.28}$$

As the pseudo-randomized quantizing errors are distributed uniformly within the interval $[-0.5q, +0.5q]$ independently of the input signal, the expected value of the process $\{\varepsilon_k^{(+)}\}$ is

$$\mu_{\varepsilon+} = E[\varepsilon_k^{(+)}] = \int_0^{0.5q} \varepsilon_k \frac{\mathrm{d}\varepsilon_k}{q} = \frac{q}{8}. \tag{5.29}$$

The covariance between the input signal and the first component of the quantization noise $\{\varepsilon_k^{(+)}\}$ is as follows:

$$\begin{aligned} \mathrm{Cov}(\varepsilon_k^{(+)}, x_k) = &\int_0^{0.5q} \int_0^{x_k} \left(\xi_k + \frac{q}{2} - x_k\right) x_k \frac{\mathrm{d}\xi_k}{q} \\ &+ \int_{x_k+q/2}^{q} \left(\xi_k - \frac{q}{2} - x_k\right) x_k \frac{\mathrm{d}\xi_k}{q} \varphi(x_k)\, \mathrm{d}x_k \\ &+ \int_{q/2}^{q} \int_{x_k-q/2}^{q} \left(\xi_k + \frac{q}{2} - x_k\right) x_k \frac{\mathrm{d}\xi_k}{q} \varphi(x_k)\, \mathrm{d}x_k - \mu_{\varepsilon+} m_x. \end{aligned} \tag{5.30}$$

After due integration, and taking into account result (5.29),

$$\mathrm{Cov}(\varepsilon_k^{(+)}, x_k) = 0. \tag{5.31}$$

Similarly, it can be shown that

$$\mathrm{Cov}(\varepsilon_k^{(-)}, x_k) = 0. \tag{5.32}$$

Since the input signal is uncorrelated with either noise component, it is obviously also uncorrelated with the quantization noise itself.

5.5.2 Spectrum of the Pseudo-randomized Quantization Noise

In order to determine the spectrum of pseudo-randomized quantization noise, first its autocovariance function $C_{\varepsilon\varepsilon}^{(p)}(mT)$ will be obtained. As the mean value of this noise $m_\varepsilon = 0$, then

$$\begin{aligned} C_{\varepsilon\varepsilon}^{(p)}(mT) &= E[\varepsilon_k \varepsilon_{k+m}] \\ &= E\left[\left(\left(\xi_k - \tfrac{1}{2}\right) q + n_k q - x_k\right)\left(\left(\xi_{k+m} - \tfrac{1}{2}\right) q + n_{k+m} q - x_{k+m}\right)\right]. \end{aligned} \tag{5.33}$$

The random variable $(\xi_k - \frac{1}{2})q$ is uniformly distributed within the interval $[-0.5q, +0.5q]$. Assume that the variables $(\xi_k - \frac{1}{2})q$ and $(\xi_{k+m} - \frac{1}{2})q$ at $m \neq 0$ are mutually independent. Note that $(n_k q - x_k)$ represents errors of randomized quantization.

If these moments are taken into account, it follows from Equation (5.33) that

$$\begin{aligned} C_{\varepsilon\varepsilon}^{(p)}(mT) &= C_{\xi\xi}(mT) + C_{\varepsilon\varepsilon}(mT) \\ &\quad + E\left[\left(\xi_k - \tfrac{1}{2}\right)q(n_{k+m}q - x_{k+m})\right] + E\left[\left(n_k q - x_k\right)\left(\xi_{k+m} - \tfrac{1}{2}\right)q)\right], \end{aligned} \tag{5.34}$$

where the autocovariance function of the auxiliary random process $C_{\varepsilon\varepsilon}^{(p)}(mT) = 0$ for $m \neq 0$, and the autocovariance function of randomized quantization noise $C_{\varepsilon\varepsilon}^{(p)}(mD_t) = 0$ for $m \neq 0$, as shown in Chapter 4. Then

$$\begin{aligned} E\left[(n_k q - x_k)\left(\xi_{k+m} - \tfrac{1}{2}\right) q\right] &= \int_0^q \int_0^q \varphi(x_k, x_{k+m})\, \mathrm{d}x_k\, \mathrm{d}x_{k+m} \\ &\quad \times \int_{-0.5q}^{0.5q} \Big[-x_k\left(\xi_{k+m} - \tfrac{1}{2}\right)(q - x_k) \\ &\quad + (q - x_k)\left(\xi_{k+m} - \tfrac{1}{2}\right)x_k\Big] \frac{\mathrm{d}\left[\left(\xi_{k+m} - \tfrac{1}{2}\right) q\right]}{q} = 0. \end{aligned} \tag{5.35}$$

Likewise,

$$E\left[(n_{k+m}q - x_{k+m})\left(\xi_k - \tfrac{1}{2}\right)q\right] = 0. \tag{5.36}$$

Substituting Equations (5.35) and (5.36) into Equation (5.33) yields

$$C_{\varepsilon\varepsilon}^{(p)}(mT) = \begin{cases} q^2/12 & \text{for m} = 0, \\ 0 & \text{for m} \neq 0. \end{cases} \tag{5.37}$$

Thus the autocovariance function of the pseudo-randomized quantization noise does not depend on the input signal and is invariable. Consequently, the spectral density function of this noise is

$$G_\varepsilon(f) = \frac{q^2}{6f_s} = \frac{q^2T}{6}, \tag{5.38}$$

i.e. it is constant over the frequency range $[0, 0.5f_s]$.

This result leads to the following significant conclusion:

> The spectral density function of the pseudo-randomized quantization noise is invariable and depends only on the quantization step-size q and the sampling frequency f_s. This holds no matter how many reference threshold levels are used.
>
> In other words:
>
> No matter how rough pseudo-randomized quantizing is, the respective quantization noise is white and its power and other parameters do not depend on the input signal if only the auxiliary pseudo-random process used can be considered to be white.

No other quantization scheme has a similar property. It is true that randomized quantizing can also be realized in such a way that the quantization noise is white. However, the mean power of this noise depends to some extent on the variance of the input signal. The spectrum of deterministic quantization in principle largely depends on the input signal and this dependence can be weakened only by decreasing the size of the quantization step.

Note that formula (5.38), describing the spectrum of the quantization noise being discussed, is obtained under the assumption that the input signal is sampled periodically. When the signal is sampled randomly, obtaining an analytical description of the respective spectral density function is much more complicated. However, computer simulations show that when expression (5.38) is used for engineering evaluations of the noise spectra, the mean sampling rate f_s can be substituted for the sampling frequency f_s in Equation (5.38).

5.5.3 Noise Reduction by Oversampling

Assume that the spectral density function of a signal $x(t)$ being quantized is confined to the frequency range $[0, f_0]$, $f_0 < f_s/2$. The power P_ε of the corresponding rounding-off noise spectrum is distributed evenly over the whole frequency range $[0, f_s/2]$, no matter how rough the quantization is. The spectral density

function of this noise can be subdivided into two parts. The first part covers the signal bandwidth and the second part is outside this range. Accordingly, the mean power P_ε is also divided into two respective parts P_1 and P_2. The mean power of that part of the noise spectrum that belongs to the signal spectrum range is denoted by P_1 and the remaining part by P_2.

Referring to Equation (5.38),

$$P_1 = \frac{q^2 f_0}{6 f_s} \quad \text{and} \quad P_2 = q^2 \left(\frac{1}{12} - \frac{f_0}{6 f_s} \right). \tag{5.39}$$

If the quantized signal is passed through a low-pass filter with the cut-off frequency f_0, the second part of the rounding-off noise spectrum power P_2 is filtered out and the resulting quantization accuracy, characterized by the signal-to-noise ratio (SNR) $10 \log (P_x / P_1)$,increases. The SNR can be improved in this way by so-called signal oversampling. If the sampling frequency is increased four times and appropriate filtering applied, the quantization accuracy is improved by one effective bit. This approach is popular and it is used also in the cases of rough deterministic quantization. However then it might be less effective as the power of the deterministic quantization noise at a small number of reference levels used is concentrated more in the low frequency region while it is not so at pseudo-randomized quantizing.

5.6 Some Properties of Quantized Signals

The properties of quantized signals naturally differ from the properties of the original signals. These differences depend on the quantization mode applied. This phenomenon has been extensively studied for deterministic quantization applications. This problem has been considered for the case when quantizing is carried out pseudo-randomly.

The moments $a_{s,r}$ and the central moments $\mu_{s,r}$ of a system of random variables (x,ε) are defined as follows:

$$\begin{aligned} a_{s,r} &= E[x^s \varepsilon^r] \\ &= \int_\Omega \int x^s \varepsilon^r \varphi(x, \varepsilon)\, \mathrm{d}x\, \mathrm{d}\varepsilon, \\ \mu_{s,r} &= E[(x - m_x)^s (\varepsilon - m_\varepsilon)^r] \\ &= \int_\Omega \int (x - m_x)^s (\varepsilon - m_\varepsilon)^r \varphi(x, \varepsilon)\, \mathrm{d}x\, \mathrm{d}\varepsilon, \end{aligned}$$

where m_x and m_ε are the mean values of x and ε respectively, $\varphi(x, \varepsilon)$ is the probability density function of the system of the random variables (x, ε) and Ω is the space of x, ε on the plane (x, ε). As the probability density function of the quantization error ε is constant in the interval $[-0.5q, 0.5q]$ and does not depend on $\varphi(x)$, then $\varphi(x, \varepsilon) = \varphi(x)\Psi(\varepsilon)$ and

$$\begin{aligned} a_{s,r} &= \int_{-\infty}^{\infty} \int_{-0.5q}^{0.5q} x^s \varepsilon^r \varphi(x)\Psi(\varepsilon)\,\mathrm{d}x\,\mathrm{d}\varepsilon \\ &= \int_{-\infty}^{\infty} x^s \varphi(x)\,\mathrm{d}x \int_{-0.5q}^{0.5q} \varepsilon^r \frac{1}{q}\mathrm{d}\varepsilon \\ &= \begin{cases} a_s \dfrac{q^r}{(r+1)2^r} & \text{for even } r, \\ 0 & \text{for odd } r, \end{cases} \end{aligned} \tag{5.40}$$

where a_s is the sth moment $\varphi(x)$. Taking into account the fact that $m_\varepsilon = 0$, then similarly

$$\mu_{s,r} = \begin{cases} \mu_s \dfrac{q^r}{(r+1)2^r} & \text{for even } r, \\ 0 & \text{for odd } r, \end{cases} \tag{5.41}$$

where μ_s is the 5th central moment of $\varphi(x)$. Now $E[\hat{x}^s]$ can be defined. According to the definition $\varepsilon = x - \hat{x}$ and

$$\begin{aligned} E[\hat{x}^s] &= E[(x-\varepsilon)^s] \\ &= E\left[\sum_{j=0}^{s} (-1)^j C_s^j x^{s-j} \varepsilon^j\right] \\ &= \sum_{j=0}^{s} (-1)^j C_s^j E[x^{s-j}\varepsilon^j] \\ &= \sum_{j=0}^{s} (-1)^j C_s^j a_{s-j,j}. \end{aligned} \tag{5.42}$$

Substituting Equation (5.40) into Equation (5.42) and taking into account the fact that at odd j the moment $a_{s-j,j}$ is equal to zero and that the sum in

Equation (5.42) should be determined only for even j, gives

$$\begin{aligned} E[\hat{x}^s] &= \sum_{j=0}^{[s/2]} C_s^{2j} a_{s-2j,2j} \\ &= \sum_{j=0}^{[s/2]} C_s^{2j} a_{s-2j} \frac{q^{2j}}{(2j+1)2^{2j}} \\ &= a_s + \sum_{j=1}^{[s/2]} C_s^{2j} a_{s-2j} \frac{q^{2j}}{(2j+1)2^{2j}}, \end{aligned}$$

where $[s/2]$ is the integer part of $s/2$. Hence

$$a_s = E[\hat{x}^s] - \sum_{j=1}^{[s/2]} C_s^{2j} a_{s-2j} \frac{q^{2j}}{(2j+1)2^{2j}} \tag{5.43}$$

On the basis of Equation (5.43), the following corrections, essential for many engineering calculations, are obtained:

$$\left.\begin{aligned} a_1 &= E[\hat{x}], \\ a_2 &= E[\hat{x}^2] - \tfrac{1}{12}q^2, \\ a_3 &= E[\hat{x}^3] - \tfrac{1}{4}q^2 E[\hat{x}], \\ a_4 &= E[\hat{x}^4] - \tfrac{1}{2}q^2 a_2 - \tfrac{1}{80}q^4 \\ &= E[\hat{x}^4] - \tfrac{1}{2}q^2 E[\hat{x}^2] + \tfrac{7}{240}q^4. \end{aligned}\right\} \tag{5.44}$$

These corrections are the same as those obtained for deterministic quantization, but their meaning is quite different. In the case of deterministically quantized signals, they reduce the bias of the estimates $\hat{a}_s$, but their application does not guarantee its elimination. How efficient they are depends on $\varphi(x)$ and on the value of q. When Equations (5.44) are applied for correction of the estimates $\hat{a}_s$ obtained by processing pseudo-randomly quantized signals, these corrections eliminate bias errors completely, no matter what is the probability density function of the signal or the size of the quantization step q.

To obtain equations describing the estimates $\hat{a}_s$ of the corresponding moments a_s, the estimates $(1/N)\sum_{k=1}^{N} \hat{x}$ should be substituted for $E[\hat{x}_s]$. It can be shown that the estimates $\hat{a}_s$ defined in this way are unbiased. To obtain equations defining

Table 5.1

Examples of achievable effects	Quantizing	
	Randomized	Pseudo-randomized
Reduction of the quantization bit rate, widening application range of rough quantization	X	X
No systematic bias errors at estimation of signal parameters, including spectral	X	X
Wide dynamic range, no spurious frequencies	X	X
Decorrelation of signals and quantisation noise	X	X
Uniform quantization noise distribution invariant to quantization bit rate	0	X
Uniform power spectrum of the quantization noise independently of the signal and the quantization step size	0	X

the estimates of the central moments μ_s note that on grounds of Equation (5.42),

$$E[(x - m_x)^s] = \sum_{j=0}^{s} (-1)^j C_s^j m_x^j a_{s-j} \tag{5.45}$$

and the equation obtained should then be substituted into Equation (5.44). The variance of the estimates $\hat{x}^s$ is given by

$$\mathrm{Var}[\hat{x}^s] = \sum_{j=0}^{s} C_{2s}^{2j} a_{2s-2j} \frac{q^{2j}}{(2j+1)2^{2j}} - \left[\sum_{j=0}^{[s/2]} C_s^{2j} a_{s-2j} \frac{q^{2j}}{(2j+1)2^{2j}} \right]^2 \tag{5.46}$$

To evaluate the errors of estimating a_s and μ_s, the variance $\mathrm{Var}[(1/N)\sum_{k=1}^{N} \hat{x}_s]$ should be determined first by applying Equation (5.46).

5.7 Benefits

The randomization and pseudo-randomization techniques, considered in Chapters 4 and 5, represent an effective tool for making quantizing flexibly adaptable to specific requirements of an application. Skilful application of randomized and pseudo-randomized quantizing might lead to significant benefits. The most characteristic ones are indicated in Table 5.1.

Although deliberate randomizing and/or pseudo-randomizing of quantizing apparently complicate to some extent the execution of this operation, the approach is also fruitful. The point is that it makes quantizing much better adjusted to the

needs and conditions of specific applications and that in turn helps to improve the organization of digital signal processing tasks at hand.

Bibliography

Bennet, W.R. (1948) Spectra of quantized signals. *Bell Syst. Tech. J.*, **27**(7), 446–72.

Bilinskis, I. (1976) Stochastic signal quantization error spectrum (in Russian). *Autom. Control Comput. Sci.*, **3**, 55–60.

Bilinskis, I. (1977) Quasi-stochastic coding of continuous signals. *Pr. Przem. Inst. Elektron.*, **64**, 53–9.

Bilinskis, I. and Mikelsons, A. (1975) Quantizing of signals by parallel stochastic weighting (in Russian). *Autom. Control Comput. Sci.*, **4**, 34–8.

Bilinskis, I. and Mikelsons, A. (1992) *Randomized Signal Processing*. Prentice Hall International (UK) Ltd.

Bilinskis, I., Mikelsons, A. and Nemirovski, R.F. (1974) Application of random point processes for quantizing of time intervals (in Russian). *Autom. Control Comput. Sci.*, **6**, 79–85.

Brodsky, P.I., Verjushsky, V.V., Klyushnikov, S.N. and Lutchenko, A.E. (1970) Statistical quantization technique using auxiliary random signals (in Russian). *Voprosy Radioelektroniki Ser. obshchetekhnic*, **3**, 46–53.

Demas, G.D., Barkana, A. and Cook, G. (1973) Experimental verification on the improvement of resolution when applying perturbation theory to a quantizer. *IEEE Trans Industr. Electron. Control Instr.*, **IECI-20**(4), 236–9.

Jayant, N. and Rabiner, L. (1972) The application of dither to the quantization of speech signals. *Bell Syst. Tech. J.*, **51**(6), 1293–304.

Roberts, L.G. (1962) Picture coding using pseudo-random noise. *IRE Trans. Inf. Theory*, **IT-8**, 145–54.

Veselova, G.P. (1975) On amplitude quantization with overlapping of interpolating signals (in Russian). *Avtomatika i Telemekhanika*, **5**, 52–9.

6

Direct Randomization of Sampling

In an ideal case, the features of a digital signal obtained as a result of analog-to-digital conversions would copy the features of the original analogue signal. The reality is different. The sampling and quantization operations of this kind of conversion impact on the characteristics of the obtained digital signals substantially. The characteristics of the analogue signal at the input of an ADC and of its digital output are representative rather than identical. The methods used for digitization and their implementation determine how large and significant are the differences between them. Once a signal is digitized, the features of the obtained digital signal, good as well as bad, are fixed and nothing can be done to change them. Therefore it makes sense to invest a great deal of time and effort to ensure that the quality of the signal sampling outcome is high enough.

Deliberate randomization of the signal sample value taking process is one of the most promising sampling improvement options. Although the randomization approach significantly complicates the sampling operation and processing of the digital signals obtained in this way, the resulting benefits justify these complications. For instance, application of the randomized sampling technique, as shown in Chapter 1, often leads to much better exploitation of the resources of the existing ADCs.

There are various techniques used for randomization of sampling. The discussions in this chapter focus on the direct approach to such randomization. According to this approach, signal sample values are taken at time instants that are fully and directly determined by the used randomized sampling point processes. Therefore the sampling point process completely defines the signal sampling

Digital Alias-free Signal Processing I. Bilinskis

randomization scheme. It should be emphasized that the statistics of the signal sample taking timing process, in the case of direct sampling randomization, is signal independent. As shown in the following chapter, this is not always the case.

6.1 Periodic Sampling with Jitter

It is customary to depict the sampling instants $\{t_k\}$ as points on the time axis. Sequences of such points can be considered as point processes. They play a very important role because the properties of sampled signals depend first of all on the characteristics of these point processes.

A relatively large variety of random point processes can, in principle, be applied for random sampling, and a number of them have been investigated to find out whether they eliminate aliasing. The theory of such point processes has gradually been developed and interested readers should refer to the bibliography given at the end of this chapter. Similar problems have also been extensively studied in the area known as the renewal theory and many useful research results have been discovered in that area as well.

Only a few of the known and studied random point processes have been selected for study here. The explanation for this lies in the fact that now, after much effort has been expended on research in this area, it can be seen that only a few random point processes are suitable for randomized sampling applications. One of the best is the additive random point process, often used for the purpose of deliberate sampling randomization. This point process has such remarkable properties and is so flexible that it is almost ideal for random sampling applications.

However, there are several other point processes that should also be considered because they are connected to relatively frequently observed and important effects such as fluctuation of sampling instants or random deletion of samples. Fluctuation of sampling instants is a fairly common occurrence. It can even be said that it is always present, although more often than not it is insignificant enough to be ignored. How seriously fluctuation affects the precision of signal processing depends on the kind of processing being performed and, of course, the magnitude of the fluctuations.

The randomized sampling model used for analysis of this problem is known as periodic sampling with jitter. The sampling instants $\{t_k\}$ in it are given by

$$t_k = kT + \tau_k, \quad T > 0, k = 0, 1, 2, \ldots, \tag{6.1}$$

where $\{\tau_k\}$ is a family of independent identically distributed random variables with zero mean. This sampling scheme is illustrated in Figure 6.l. Probability density

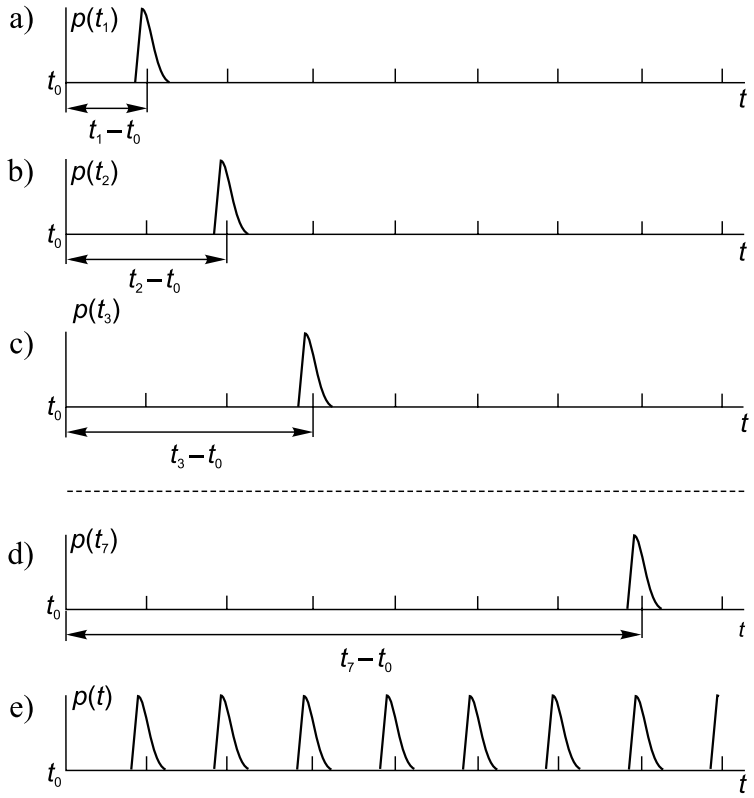


Figure 6.1 Probability density functions characterizing periodic sampling with jitter: (a), (b), (c), (d) probability density functions of time intervals $t_1 - t_0$, $t_2 - t_0$, $t_3 - t_0$ and $t_7 - t_0$ respectively; (e) resulting sampling point density function

functions of time intervals $(t_k - t_0)$ for $k = 1, 2, 3, \ldots$ are shown in Figures 6.1 (a), (b), (c) and (d). As $t_0 = 0$ in this case, the respective density functions are denoted by $p_k(t)$. The sampling point stream is characterized by the following sampling point density function:

$$p(t) = \sum_{k=1}^{\infty} p_k(t). \tag{6.2}$$

It can be seen from Figure 6.1(e) that this particular function has multiple maxima and minima. Note that as t increases the peaks shown do not decrease.

To understand the meaning of the function $p(t)$, imagine that a narrow time window Δt is moved along the t axis. Under the condition that $\Delta t \Rightarrow 0$, the

function $p(t)$ at any arbitrary time instant is equal to the probability that one of the sampling points t_s will fall within this window Δt. Then

$$p(t) = \Pr[t_s \in \Delta t] = \lim_{\Delta t \Rightarrow 0} \frac{\Delta t}{\mu} \tag{6.3}$$

Therefore if a signal $x(t)$ is sampled at the instants $\{t_k\}$, which are determined by the statistical relationship illustrated by Figure 6.1(e), some parts of the signal will be sampled with a higher probability than others. This is obviously undesirable as it will lead to signal processing errors.

There is an exception. If the time intervals $(t_k - t_0)$ are distributed uniformly in the intervals $(kT \pm 0.5T)$ respectively, then the resulting sampling point density function $p(t)$ is constant (for $t > 0.5T$). When this kind of sampling scheme is applied, all instantaneous signal values are sampled with an equal probability. It therefore seems that this method of generating random sampling points is acceptable for signal sampling implementations. However, this method in fact has a number of substantial disadvantages that prevent its wide application. The following drawbacks can be pointed out:

- The random variables $\{\tau_k\}$ should be distributed strictly uniformly within the given intervals.
- Time intervals between any two successive sampling instants t_k and t_{k+1} may be very short. Therefore technical implementations of this sampling scheme should be wideband, even at relatively low mean sampling rates.
- The randomness introduced at sampling cannot be small. It is in fact considerable.
- Statistical errors resulting from the relatively powerful randomness introduced at sampling are significant.

Up to now, periodic sampling with jitter has been considered only from the viewpoint of deliberate randomization of sampling. However, as has already been mentioned, it has another important aspect. Fluctuations in sampling instants may turn out to be highly undesirable because they may introduce significant bias and random errors. The model briefly described above is worth considering in order to find an answer to the question of how harmful these fluctuations really are under given conditions. The errors caused by such fluctuations depend considerably on the specific digitized signal processing algorithms applied. Some, like algorithms for the estimation of a number of averaged signal parameters (such as the mean power, the higher moments of signal distributions, etc.), are relatively insensitive, while others are more sensitive.

6.2 Additive Random Sampling

In the case of additive random sampling, signal samples are taken at instants

$$t_k = t_{k-1} + \tau_k, \quad k = 0, 1, 2, \ldots, \tag{6.4}$$

where τ_k is a realization of a random variable. This random sampling scheme, suggested by Shapiro and Silverman, was originally based on the assumption that successive sampling intervals $\{\tau_k, \tau_{k+1}\}$ were statistically independent and identically distributed. They were characterized first of all by their mean value μ and a standard deviation σ. Obviously, the mean sampling rate is equal to $1/\mu$.

Now consider the time intervals $[0, t_k] = \tau_1 + \tau_2 + \cdots + \tau_k$. These random variables are characterized by their respective probability density functions $\{p_k(t)\}$. Then

$$\begin{aligned} &p_1(t) = p_\tau(t), \\ &p_2(t) = p_1(t)^* p_\tau(t), \\ &\ldots\ldots\ldots\ldots\ldots \\ &p_k(t) = p_{k-1}(t)^* p_\tau(t), \end{aligned} \tag{6.5}$$

where the asterisk * denotes the composition operation. On the grounds of the central limit theorem in statistics, a very important conclusion is made:

> As the random variable $[0, t_k]$ represents the net result of a linear sum of k statistically independent constituent variables $\tau_1, \tau_2, \ldots, \tau_k$, then whatever probability density functions these constituent variables may have, the probability density of $\tau_1 + \tau_2 + \cdots + \tau_k = [0, t_k]$ will approach the normal form as k approaches infinity.

Consequently, whenever the additive random sampling scheme is applied, the density functions $p_\tau(t)$ may vary from case to case within wide boundaries without worsening sampling conditions, because the sampling point density function $p(t)$ with t increasing will always tend to the constant level $1/\mu$. Therefore, in this case

$$p(t)_{|t \geq T_a} = \frac{1}{\mu}. \tag{6.6}$$

These transformations of the density functions $p_k(t)$ in cases where $p_r(t)$ is (a) uniform and (b) close to exponential are shown in Figure 6.2. Note that the scales on these figures change from one function to another. The probability density function of a sum of r time intervals is denoted by $p_r(t)$ and the expected values

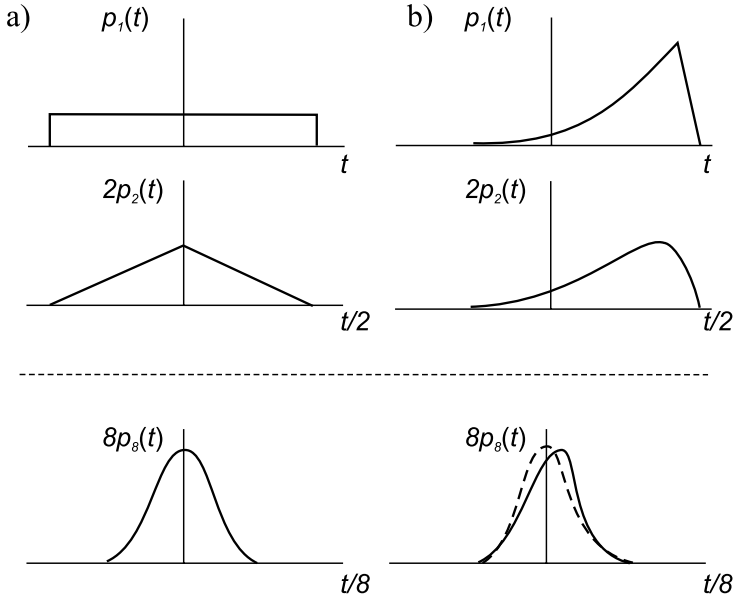


Figure 6.2 Evolutions of probability density function $p_r(t)$ when $p_1(t)$ is (a) uniform and (b) modified exponential

by μ_r. In both cases the corresponding $p_r(t)$ changes when the number r of sampling intervals summed increases from 1 to 8.

The probability density functions given Figure 6.3 illustrate the additive random sampling scheme. It can be seen that the function $p(t)$ in this case tends to become flat when t exceeds some value T_a. The value of T_a depends on $p_\tau(t)$ and, obviously, also on the acceptable deviation of $p(t)$ from the constant level $1/\mu$. However, this is the case only for synchronized sampling when $t_0 = 0$. When sampling is stationary, i.e. when t_0 is a properly distributed random variable, $p(t) = 1/\mu$ for the whole time interval over which a signal is sampled. This means that all of the instantaneous signal values are sampled with equal probability.

6.3 Sampling Function

The statistical description of random sampling point processes is not limited to the probability density functions $p_k(t)$ and the sampling point density functions $p(t)$. Other means of characterization are often more convenient, especially those based on probability distribution functions of the time intervals $\tau_1 + \tau_2 + \cdots + \tau_k$,

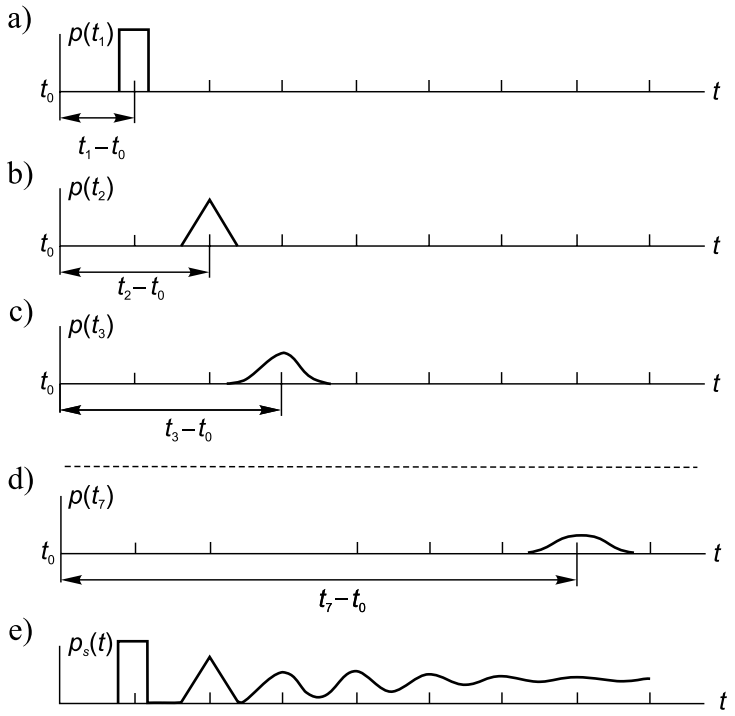


Figure 6.3 Probability density functions characterizing additive random sampling: (a), (b), (c), (d) probability density functions of time intervals $t_1 - t_0$, $t_2 - t_0$, $t_3 - t_0$ and $t_7 - t_0$ respectively; (e) resulting sampling point density function

henceforth denoted $P_k(t)$. These functions have their usual definitions. If the probability density functions of the respective random time intervals are $p_k(t)$, then

$$P_k(t) = \int_0^\infty p_k(t)\,\mathrm{d}t. \tag{6.7}$$

These functions $\{P_k(t)\}$ form a basis of yet another very useful characteristic of randomized sampling. This can be called the sampling function and can be denoted $P_s(t)$. This function was first introduced in the field of renewal theory, where it was known as the renewal function, and can be defined as follows:

$$P_s(t) = \sum_{k=1}^{\infty} P_k(t). \tag{6.8}$$

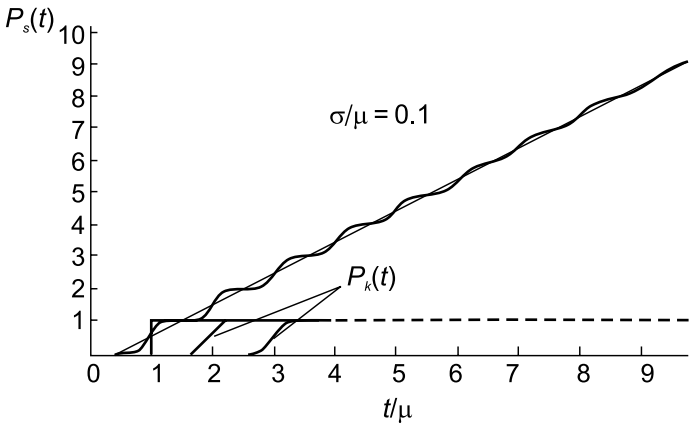


Figure 6.4 Sampling function

Its usefulness is apparent, as it allows the number $n_{t,\Delta t}$ of sampling points falling within an interval whose duration is equal to Δt to be estimated in a very simple way. Such an estimate, in a general case, is given by

$$E[n_{t,\Delta t}] = P_s(t + \Delta t) - P_s(t). \tag{6.9}$$

When the sampling process can be considered to be stationary, this kind of estimation is even simpler. Then

$$E[n_{t,\Delta t}] = \frac{\Delta t}{\mu}. \tag{6.10}$$

Although relationships (6.9) and (6.10) can be proved analytically, they are easier to understand by looking at the graphics of Figure 6.4. The probability distribution functions $P_k(t)$ correspond to the case where the respective density functions $p_k(t)$ are uniform. After some time they become normal or, more precisely, almost normal. As, according to the additive random sampling model, the random variables $\{\tau_k\}$ should be positive and normal probability density functions are defined in the range $[-\infty, \infty]$, only truncated normal probability density functions $p_\tau(t)$ are applicable in the present case. They can be defined as follows:

$$p(t) = \frac{d}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(t-\mu)^2}{2\sigma^2}\right], \tag{6.11}$$

where

$$d = \frac{1}{P[(b-\mu)/\sigma] - P[(a-\mu)/\sigma]}.$$

The quantities a and b are the left and right boundaries of the truncated density function respectively and $P[x]$ is the standardized normal probability distribution function given by

$$P[x] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(\frac{-u^2}{2}\right) \mathrm{d}u.$$

The sampling function $P_s(t)$ is obtained by summing the functions $\{P_k(t)\}$. Under conditions where $a = 0, b = \infty$ and $\sigma/\mu < \frac{1}{3}$, it can be assumed that $d \cong 1$. It then follows from Equations (6.8) and (6.11) that

$$P_s(t) = \sum_{k=1}^{\infty} \frac{1}{\sqrt{2\pi k\sigma}} \int_0^t \exp\left[-\frac{(u - k\mu)^2}{2k\sigma^2}\right] \mathrm{d}u. \tag{6.12}$$

Note that this function describes the case of synchronized sampling where the beginning of the signal sampling process coincides with one of the points from the corresponding random point process. With t increasing, this function tends to a linear form. It has been proved that when $t \Rightarrow \infty$, the sampling function $P_s(t)$ can be defined by

$$P_s(t) = \frac{t}{\mu} + \frac{\sigma^2 - \mu^2}{2\mu^2} + 0(1) \tag{6.13}$$

where $0(1)$ is a function which at $t \Rightarrow \infty$ approaches zero. This simplified version of the sampling function can be used only when the errors due to the oscillations of $P_s(t)$ are small enough, i.e. when t exceeds some value, say when $t \geq T_a$.

Specifically, this approximation can be applied to determine the parameter T_a mentioned above in connection with the problem of providing stationarity of sampling. As shown in the next section, synchronized sampling can be considered stationary if the sampling of a signal is delayed for at least $t = T_a$ relative to the instant when the corresponding random point process was started. Calculations made on the basis of the given approximation show that for $\sigma/\mu \leq 0.3$ and the error ε less than 0.0013,

$$T_a \geq \frac{0.4\mu^3}{\sigma^2} \tag{6.14}$$

The error ε in this case is defined by

$$\varepsilon = \max \frac{|p(t) - 1/\mu|}{1/\mu}, \quad t \geq T_a.$$

This parameter T_a is also applicable in other cases, for instance when conditions guaranteeing independent sampling of signal realizations have to be determined.

Obviously, if a number of signal realizations is sampled and processed, sampling of a signal realization can only be considered to be independent from the sampling of a previous realization if the distance between the two is at least equal to T_a.

6.4 Elimination of Bias Errors

It is an accepted practice to divide the errors, which occur in the course of signal processing, into two groups: random or statistical and systematic or bias errors. Ways and means of diminishing the first are discussed repeatedly throughout this book. The second kind of error is considered here. Typically they are more damaging than the errors of the first kind.

Many factors might lead to errors in signal digital processing and it is impossible to predict them all. However, as the relationships taking place at signal representation are considered in digital form, it is of interest to find what conditions for signal sampling need to be satisfied that would make it possible to avoid errors at the following processing of the sampled signals. An attempt is made to determine the conditions under which the estimation of randomly sampled signal parameters is unbiased, no matter which kind of stationary signals are being sampled and analysed. It is assumed that if a specific sampling procedure is proved to be good for some representative signal processing tasks, then it would also be better in many other cases. An estimation of functionally converted signal parameters is chosen for these representative signal processing tasks.

Consider signal parameters that can be interpreted as the mean values of functionally converted signals:

1. The mean power P_x of a signal can be represented by the following equation:

$$P_x = \frac{1}{\Theta} \int_0^{\Theta} [x(t)]^2 \, \mathrm{d}t. \tag{6.15}$$

In this case the functional conversion is carried out by squaring the signal $x(t)$.

2. The third moment of the signal $x(t)$ is defined by

$$a_{x3} = \frac{1}{\Theta} \int_0^{\Theta} [x(t)]^3 \, \mathrm{d}t. \tag{6.16}$$

The corresponding functional conversion is $F_3 = [x(t)]^3$.

3. The Fourier transform at the frequency ω is

$$S(\omega) = \frac{2}{\Theta} \int_0^{\Theta} k(t) \exp(-\mathrm{j}\omega t) \, \mathrm{d}t. \tag{6.17}$$

In this case the functional conversion is

$$F_{s(\omega)} = w(t)\exp(-\mathrm{j}\omega t).$$

Let us denote by A those signal parameters that can be interpreted on the basis of the functional conversion, F_A. Then

$$A = \frac{1}{\Theta}\int_0^{\Theta} F_A[x(t)]\,\mathrm{d}t, \tag{6.18}$$

where Θ is the duration of time during which the signal is observed. Assume that the signal is sampled at random instants $\{t_k\}$. Then the estimation $\hat{A}$ of the parameter A is given as

$$\hat{A} = \frac{1}{N}\sum_{k=1}^{M} F_A[x(t_k])], \tag{6.19}$$

where M is the number of signal samples within the time interval $[0, \Theta]$. The expectation can be written as

$$\begin{aligned} E[\hat{A}] &= \frac{1}{N}\int_0^{\Theta}\sum_{k=1}^{\infty} F_A[x(t)]p_k(t)\,\mathrm{d}(t) \\ &= \frac{1}{N}\int_0^{\Theta} F_A[x(t)]\sum_{k=1}^{\infty} p_k(t)\,\mathrm{d}t. \end{aligned} \tag{6.20}$$

Under the conditions

$$\sum_{k=1}^{\infty} p_k(t) = p(t) = \frac{1}{\mu} = \text{constant} \qquad \text{and} \qquad N\mu = \Theta, \tag{6.21}$$

it follows from Equation (6.10) that

$$\begin{aligned} E[\hat{A}] &= \frac{1}{N}\int_0^{\Theta} F_A[x(t)]p(t)\,\mathrm{d}t \\ &= \frac{1}{\Theta}\int_0^{\Theta} F_A[x(t)]\,\mathrm{d}t = A. \end{aligned}$$

Thus the following conclusion is made:

To ensure unbiased estimation of the signal parameters, which can be represented by the mean value of the correspondingly functionally converted signals, random sampling should be performed in such a way that the conditions (6.21) are met.

The conditions for unbiased estimation actually cover a much broader area of signal analysis than estimation of the mean value of the functionally converted signals. Note that satisfaction of conditions (6.21) provides unbiased estimation of signal spectra and unbiased reconstruction of signal waveforms when the reconstruction is carried out on the basis of the signal spectral component estimations. These conditions are quite logical. They state that the sampling procedure should be performed in such a way that all parts of the signals should be sampled with equal probability. This is obviously relevant for most signal processing cases. Note that periodic sampling with jitter, in general, does not meet the conditions (6.21).

Bibliography

Bilinskis, I. and Mikelsons, A. (1992) *Randomized Signal Processing*. Prentice-Hall International (UK) Ltd.

7

Threshold-crossing Sampling

In cases where sampling is being randomized, this is usually done purposefully to take signal sample values at specifically defined time instants in order to achieve some specific desirable effects. When that happens, the sampling process is randomized and is also nonuniform. In fact, randomization of sampling and nonuniform sampling are almost synonymous. However, deliberate randomization of sampling makes sense only if such a complication of this operation is justified by the end results. Apparently there has to be a good reason for randomizing the sampling operation as the equidistant sample value taking procedure has many outstanding positive qualities and is preferable whenever application of it is appropriate. The motivation considered so far for randomizing of sampling has been the widening of the frequency range for fully digital signal processing. That surely is a good enough argument in favour of deliberate direct randomizing of sampling as it leads to the irregularities of the sample taking process vital for elimination of aliasing. However, avoidance of aliasing is not the only possible reason why sometimes it makes sense to use this approach. There might also be some other motivational factors, like complexity reduction of systems for data acquisition, transmission and processing. Specific types of sampling randomizing, indirect randomizing and targeting goals other than suppression of aliases are discussed in this chapter.

The operational environment of the data acquisition systems often dictates the requirements and defines the conditions for sampling. Sometimes there are restrictions imposed on the accessibility to the signal sources. Then the signals might be observed and their sample values taken only at random unpredictable time instants. The obtained sequences of the signal sample values under these

Digital Alias-free Signal Processing I. Bilinskis

conditions are unavoidably nonuniform and have to be treated as such. In some other cases randomizing of sampling occurs unintentionally as a side effect. This kind of indirect randomization of sampling takes place when the sampling process is arranged so that the signal sample values are obtained from signals at the time instants when they cross some thresholds, set up either by constant levels or specially generated reference waveforms. As explained in this chapter, this approach to signal digitization might be preferred, for example, in cases of data acquisition from multiple signal sources on a large scale. The signal sample values are then obtained at random time instants so that they are nonuniformly spaced. However, in this case the randomization of the sampling process is not planned or controlled. It occurs as a consequence of the used principle for the signal sample value taking. The application rationale, advantages and limitations of these sampling techniques are considered.

7.1 Sampling at Input and Reference Signal Crossings

In addition to the amplitude sampling, so far considered as the basic signal sample value taking operation for representing signals in a digital form, yet another approach to this operation, so-called threshold-crossing sampling, might be used. This kind of sampling is a tool that may be effectively applied in a number of areas, including demodulation of variously modulated signals and massive data acquisition from sensor arrays, clusters and networks. There have also been attempts to exploit the threshold-crossing sampling for many other applications related to antenna designs, image reconstruction, etc. While the use of amplitude sampling is universal and is emphasized in this book, sampling at the input signal $x(t)$ and reference signal $r(t)$ crossings, although more specific and less often applied, deserves serious attention as well.

At this type of sampling, signals are compared with various types of reference functions. Achieving the equality $x(t_k) = r(t_k)$ of a signal $x(t)$ and a reference function $r(t)$ is considered as a crossing of the threshold. The events of a signal crossing this threshold are detected and the time instants $\{t_k\}$ of these events carry the information about the signal sample values. The term 'threshold-crossing sampling' covers many things. Specifically, it encompasses sampling based on zero crossings, constant level crossings, multiple level crossings and input $x(t)$ and time-variant reference signal $r(t)$ crossings. Recovery of the signal sample values from the crossing time instants, straightforward at the zero and constant level crossings, requires some calculations at the time-variant reference function crossings.

7.1.1 Level-crossing Sampling

The simplest and most popular threshold-sampling variety is based on zero crossings. Obviously, crossings of zero or any other single level provide information limited with regard to the scale factor, as the signal changes below and above the threshold levels are not reflected in any way at all. The signal sample values are then fixed and only the time instants $\{t_k\}$ when the signal assumes these constant values are detected. Relatively many useful zero-crossing applications have been developed. In particular, this technique serves well in cases where the parameters to be estimated do not depend on the amplitude, e.g. at applications related to signal phase angle and frequency measurements. The direct current component usually has to be extracted from the signal.

Evidently multilevel crossings are more informative. For instance, it is possible to use such an approach when performing signal asynchronous quantizing. Analog-to-digital converters built on this basis are specific as the signal sample values in this case are obtained only at the time instants when the signal crosses one of the quantization levels. Therefore the signal sampling operation is randomized, nonuniform and signal dependent. On the one hand, special techniques are needed for processing digital signals formed in this way, which represents a drawback. On the other hand, this approach has lately received a lot of attention as it has a significant power saving potential, essential for many applications.

The main attractiveness of the level-crossing sampling approach is the extreme simplicity of the electronic circuits needed for realization of it. Only one comparator is usually exploited for detecting the signal and threshold-crossing events. The crossings repeatedly happening at time instants $\{t_k\}$ are converted into a nonuniformly spaced one-bit stream. However, the quality of this kind of sampling depends on the performance of the comparator under given specific conditions. If there is virtually no noise present, reaching the equality $x(t_k) = r(t_k)$ is fixed precisely as occurring at the correct time instant t_k. Unfortunately, when the signal values approach the threshold, the comparator becomes sensitive to noise, so even relatively weak noise affects the outcome of this type of sampling. If there is an additive noise, the signal and noise mixture crosses the threshold sooner or later than the signal itself. Under the impact of the noise, there might even be some repeated crossings observed as 'ringing' of the comparator. As a result the time instant when the signal crosses the threshold is indicated with some error. How large this is depends on various factors, including the threshold-crossing angle. Of course, this error decreases if the signal rise is steeper. If the threshold is fixed at a constant level, the crossing angle totally depends on the signal

characteristics and nothing can be done to widen this crossing angle. There are various techniques that can be used to avoid the ringing effect.

7.1.2 Time-variant Threshold Crossings

The sampling operation and techniques for processing signals sampled in the threshold-crossing manner are quite specific, but this subject lies beyond the scope of this book. However, there are some ramifications of this sampling approach that are much closer to the core subject of here. Specifically, these are the sampling techniques based on time-variant threshold crossings.

Threshold-crossing sampling becomes more dynamic if this process is arranged as signal crossings with a time-variable reference function. Then the threshold changes in time and typically covers the whole dynamic range of the signal being sampled. Although the sampling instants then still depend on the signal, the reference function could be used for controlling, to some extent, the sampling process and that helps to resolve the basic inherent problems of the threshold-crossing sampling approach.

The subject of time-variant threshold crossings is not a simple one. Many specific methods and techniques have been proposed in this area and a great deal of time and effort has been invested in trying to resolve many theoretical and engineering problems arising at application attempts. The experience obtained in this direction is valuable; it is relatively well documented and some bibliography related to this subject is given at the end of the chapter. Although this experience is taken into account, it is not described here as it makes sense to use a different approach to analysis and description of this kind of sampling. Indeed, no matter how specific variable threshold-crossing sampling processes are, they still belong to the class of nonuniform sampling processes discussed in this book in detail. Consequently, it is not necessary to cover the full scope of the subject. The knowledge obtained in other areas of nonuniform sampling makes it easier to discover the features and characteristics of the sampling based on the time-variant threshold crossings, which is demonstrated in the following section.

7.2 Representing Signals Using Timing Information

For function-crossing sampling, detection of the time instants $\{t_k\}$ at which a signal $x(t)$ intersects a reference waveform is the basic operation. Therefore it is crucial to realize how this type of timing information could be used to represent sampled signals. The best way of approaching this problem is to focus on finding exactly how the precise signal sample values $x(t_k)$ could be recovered after these time

Table 7.1

Input	Sampling: Comparing	Sampling: Transmitting	Digital output
Signal → Sine wave →	Analogue values of crossing time instants	Analogue values of sampling instants	Recovery of digital reference sample values ⇕ Recovery of digital sampling instants

instants $\{t_k\}$ have been fixed. In the case of a well-defined and stabilized reference function $r(t)$, this task of the signal sample value recovery is actually not very difficult. The signal sample values $x(t_k)$ are evidently equal to the corresponding discrete value of the reference function $r(t_k)$. Therefore, the digital version $x(t_k)$ of the original signal $x(t)$ is obtained as a sequence of the reference signal sample values taken at the crossing instants $\{t_k\}$. It looks like a typical nonuniform signal represented by a sequence of signal sample values randomly located on the time axis at instants $\{t_k\}$. That confirms the point that the sampling process based on the time-variant threshold crossings could be reduced to the typical nonuniform sampling processes considered in other chapters.

Various approaches to realization of this kind of sampling could be used. A typical set of operations leading to recovery of the digital values of signal samples and their taking instants is given in Table 7.1. According to this scheme, the input signal is compared with the reference function and the timing information obtained is then transmitted over wire, radio or optical channels using an analog signal. Then the digital values of the reference signal $r(t_k)$ and the sampling instants $\{t_k\}$ are recovered, not necessarily in that order.

The fixed signal and sine-wave crossing instants $\{t_k\}$ and the corresponding sine-wave values $r(t_k)$ recovered from them fully describe the randomly digitized signal properly tied to the time axis. Evidently the most responsible functions are both the sine-wave crossing time instant $\{t_k\}$ fixing and the reference signal value $r(t_k)$ recovery. The quality of signal sampling based on sine-wave crossings to a large extent depends on the accuracy achieved at the performance of these functions. Some aspects of their execution are discussed in Section 7.5.

At first glance the outcome of function-crossing sampling, the sequence of signal sample values, is similar to the digital signal obtained as a result of applying the additive or periodic sampling processes with jitter. However, this similarity is

superficial. Actually the sampled signal, formed as a result of function-crossing sampling with subsequent recovery of the signal sample values, has features differing from the features of the digital signals obtained in the cases of direct deliberate pseudo-randomization of sampling. The point is that the sampling process based on the reference function crossings is basically random and, consequently, the sampling intervals are random continuous-value variables. This means that the sampling irregularities are not *a priori* known, which clearly represents a disadvantage, especially in the cases where adapting signal processing to the sampling irregularities is indicated.

As randomizing of the reference function crossings happens unintentionally, there is no guarantee that the nonuniform output signal, representing the outcome of such a function-crossing sampling, will provide effective suppression of aliases unless special arrangements for controlling the sampling irregularities are made. Nevertheless, under certain conditions, it is feasible. A particular approach to this problem of achieving an alias-suppression capability is shown in the following section.

7.3 Sine-Wave Crossings

Choosing an appropriate type of reference function is vital for effective implementation of the function-crossing sampling. Although various reference functions might be exploited, sinusoidal functions are often preferable as they are narrowband and can be easily generated, stabilized and used for reconstruction of the input signal. For that reason, the focus of the following discussions is on signal and sine-wave crossings.

7.3.1 Recovery of Signal Sample Values

The signal sampling process, based on sine-wave crossings, can be arranged in various ways. Direct application of this sampling concept is illustrated in Figure 7.1. According to this, a comparator is used for comparing the signal with the sine-wave reference function. Binary signal 1 is formed at the output of the comparator marking the parts of the signal waveform where it exceeds the reference sine wave. The rising and falling edges of the comparator output signal indicate the intersections of both waveforms occurring at the time instants t_k, $k = 0, 1, 2, \ldots$, when the equality $x(t_k) - r(t_k) = 0$ takes place. Recovery of the signal sample values $x(t_k)$ is performed by reading the values $r(t_k)$ of the reference waveform corresponding to the sampling instants t_k.

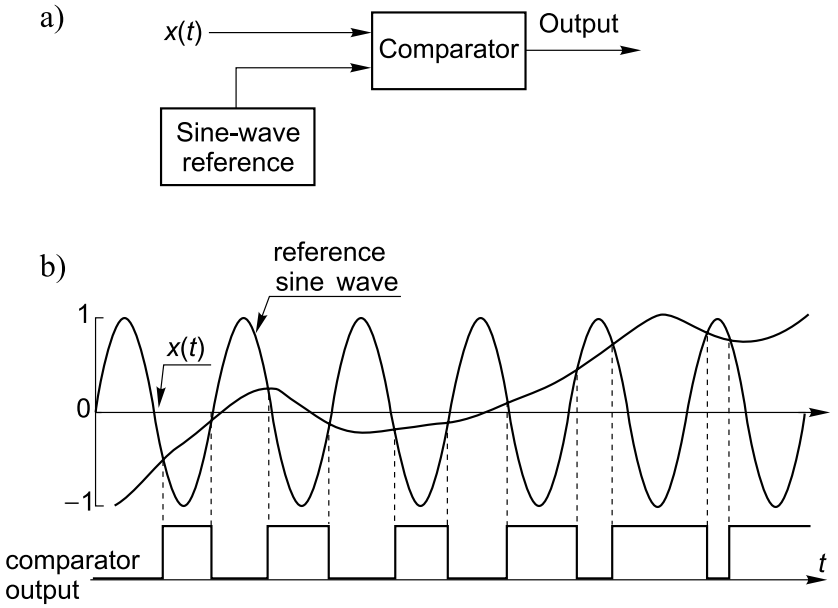


Figure 7.1 Signal sampling based on sine-wave crossings: (a) diagram illustrating realization; (b) time diagram of the involved signal interaction

In general, two coordinates need to be given for each signal sample value taken randomly. In the case of direct sampling randomization, the sampling instants are pre-planned and to fully determine each signal sample only its value has to be measured. At the threshold-crossing sampling, the input and reference signal crossings occur randomly in time. Therefore it might seem that this type of indirect sampling randomization suffers from the handicap of not knowing when the signal sample values are taken. That is true, but on the other hand the sample value of the reference function at the crossing instant actually indicates both coordinates of the corresponding signal sample. Indeed, the time instant t_k of taking each signal sample value $x(t_k)$ (and equal reference sine-wave value) functionally depends on the value of this sample. It simply has to be recovered from the corresponding instantaneous value of the reference function $r(t_k) = A_\mathrm{r} \sin 2\pi f_\mathrm{r} t_k = x(t_k)$, where A_r is the amplitude of the reference sinusoid. In the case where this type of sampling is realized on the basis of the scheme illustrated in Figure 7.1, there is an uncertainty that has to be resolved. This is related to the fact that there are two phase angle values (except the phase angles equal to $\pi/4$ and $3\pi/4$) corresponding to each particular reading of the reference sine-wave value. The scheme shown in

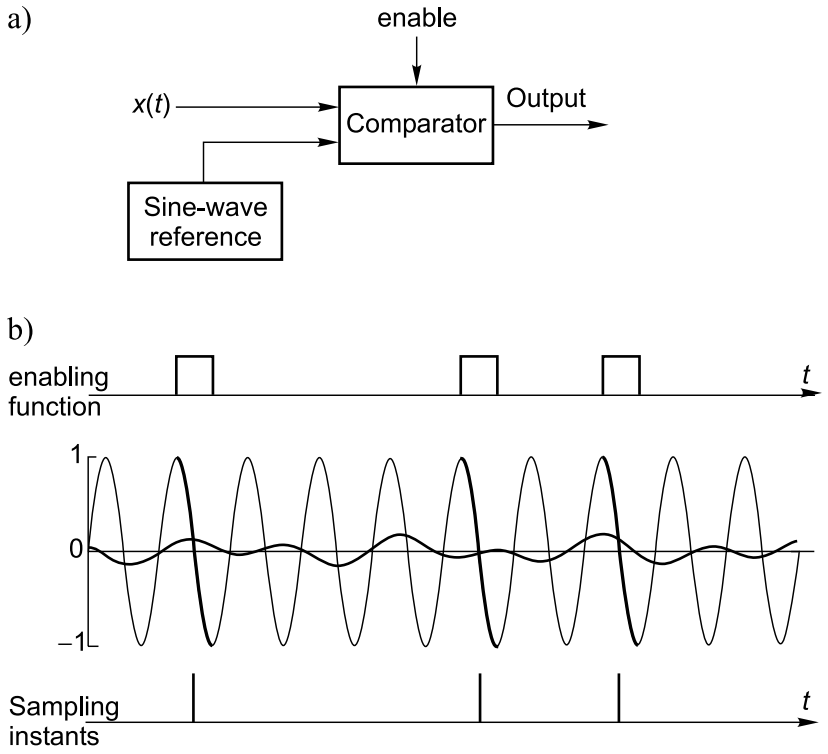


Figure 7.2 Illustration of sampling based on sine-wave crossings pseudo-randomly activated for half-periods of the reference function: (a) diagram illustrating realization; (b) time diagram of the sampling processes

Figure 7.2 takes care of this problem and is more flexible in some other aspects as well.

When the sampling operation is carried out in a particular way according to the time diagrams given in Figure 7.2, it is obtained from the equality $r(t_k) = A_\mathrm{r} \sin 2\pi\, \mathrm{f_r} t_k = x(t_k)$, where the sampling time instants depend on the reference sample value $r(t_k)$ as follows:

$$t_k = \left(k + \frac{1}{4}\right) T_\mathrm{r} + \frac{T_\mathrm{r}}{2\pi} \arcsin \frac{r\,(t_k)}{A_\mathrm{r}}, \tag{7.1}$$

where T_r is the period of the reference function and $k = 0, 1, 2, \ldots$ is the number of reference function periods that also indicate the number of sampling events taking place within this period of the sine wave.

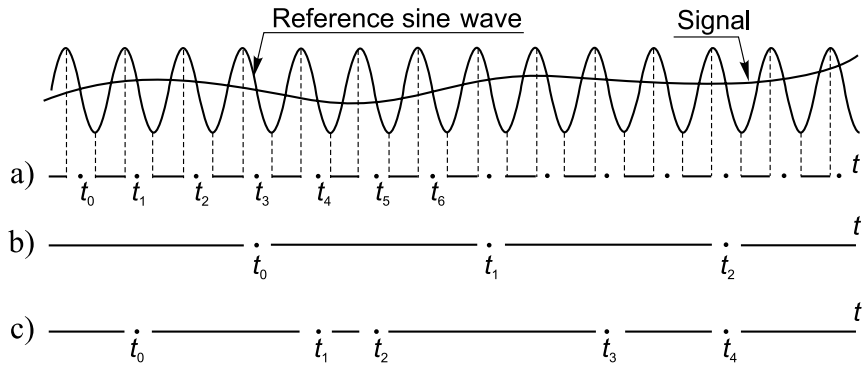


Figure 7.3 Typical applications of sampling based on sine-wave crossings in the case where the signal is band-limited and the frequency of reference sinusoid exceeds the upper frequency of the signal spectrum

7.3.2 Various Realizations

A very useful instrument that can be used to adjust the sampling scheme to various exploitation conditions is the enabling function indicated in the scheme given in Figure 7.2. This function effectively controls the scheme that provides the realization of specific sampling scenarios. It enables the scheme for signal sampling only during certain time slots. The width of these slots is typically equal to the half-period of the used reference sine wave. This type of control has a number of essential functions. Firstly, it leads to resolving the uncertainty. When crossings of the signal and the sine wave take place within the half-period time slots, each crossing provides a single reference function value. Secondly, in many cases the power supply for the comparator can be activated only during the time intervals when it is enabled for comparison of the input and reference signals. It is obvious that it is a power-saving measure crucial for many applications. Thirdly, use of this enabling function makes it possible to arrange data acquisition from a number of signal sources, as described in Section 7.4. Fourthly, this control function can be used to match the sampling mode to the specific requirements of a signal processing system.

In many cases this kind of sampling is used under conditions where the signal is band limited and the frequency of the reference sinusoid exceeds the upper frequency W of the signal spectrum by at least two times. Some of these cases are illustrated in Figure 7.3. The most intensive sampling takes place when there is only a single signal source connected to a single comparator. This comparator is enabled periodically with only half-period intervals between successive open

time slots, as shown in Figure 7.3(a). If there is more than one input and several comparators are used to perform the sampling operation, a particular comparator is kept closed for a number of reference signal periods. Figure 7.3(b) illustrates such a situation.

Note that in both of these cases the time slots within which the signal and sine-wave crossings occur randomly are repeated periodically. Therefore the sampling point process belongs to the category of periodic sampling processes with jitter and the sampling instants are defined as

$$t_k = \left(\frac{1}{4} + \frac{1}{2}\,\mathrm{km}\right) T_\mathrm{r} + \tau_k, \tag{7.2}$$

where m denotes the number of reference sine-wave half-periods between the time slots open for sampling and τ_k is the random time interval between the edge of the time slot and the crossing instant.

Figure 7.3(c) illustrates a particular realization of sampling based of sine-wave crossings providing for alias-free signal digitizing. The comparator in this case is enabled pseudo-randomly. This means that between successive slots within which sampling events are located randomly the comparator is kept inactive for pseudo-random numbers of the reference sine-wave periods. The resulting stream of sampling events belongs to the category of additive sampling processes, with the random interval between the sampling instants distributed specifically. The sampling instants are then given as

$$t_k = t_{k-1} + (n_k T_\mathrm{r} + \tau_k) \tag{7.3}$$

where $(n_k T_\mathrm{r} + \tau_k)$ denotes the random variable representing the interval between the sampling instants t_{k-1} and t_k.

While the first component $n_k T_\mathrm{r}$ of this interval $(n_k T_\mathrm{r} + \tau_k)$ is equal to a pseudo-random number n_k of the reference sine-wave periods T_r, the second component τ_k is a random variable. It is distributed within a half-period slot of the reference function and the exact distribution itself is not predictable as it depends on the signal.

Performing this type of sampling in a pseudo-randomized way suppresses the aliasing effect and makes sense in multichannel data acquisition cases. Therefore the mean sampling rate in a particular channel depends not only on the frequency of the reference signal but also on the number of input channels and the need for anti-aliasing measures might arise.

Sine-wave crossing sampling is also applicable for digitizing of signals at frequencies exceeding the frequency of the reference signal. How the sampling process develops under these distinctly different conditions is shown in Figure 7.4.

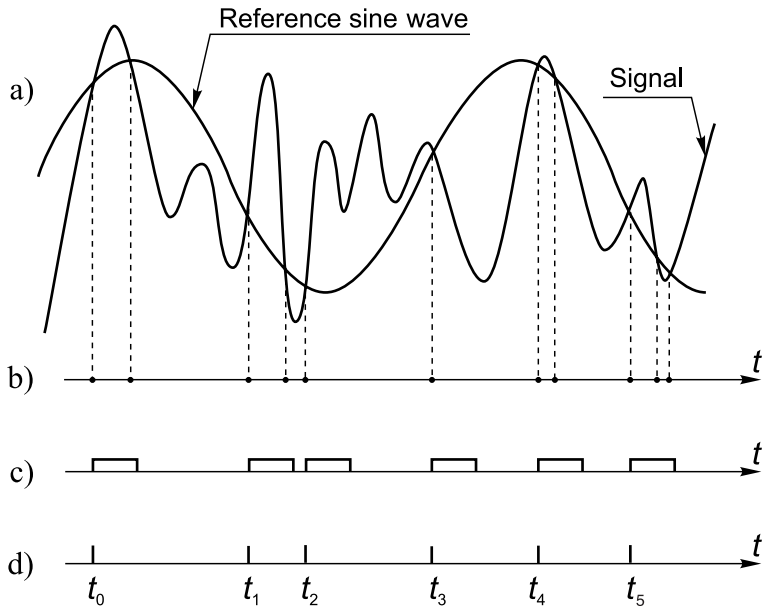


Figure 7.4 Sampling based on the signal and sine-wave crossings in the case where the frequency of the reference signal is lower than the upper frequency of the bandpass signal

The scheme used for performing the sampling operation is similar to that given in Figure 7.2(a), except that a disabling rather than an enabling function is used for blocking the comparator, which is normally kept in the active operational state. The comparator forms a signal whenever a crossing of the input and reference signals occurs during the time intervals when it is not especially disabled.

However, this sampling mode should be applied with caution. There is a problem, which is related to the fact that frequencies equal to the reference signal frequency will not be encoded correctly. As the upper frequency of the input signal in this case exceeds the frequency of the reference signal, the frequencies of both signals might overlap and in that case the obtained digital signal would not reflect the input signal correctly. This type of sampling can only be used if the frequency of the reference signal is eliminated from the input bandpass signal. This sampling approach is quite useful as it expands the applicability of the sampling method based on sine-wave crossings, which therefore also widens the application range. The expansion of applicability of sine-wave crossings is especially interesting in the light of fast DFT algorithms, discussed in Section 16.4, which are used exclusively to process nonuniformly presented digital signals obtained as a result of sampling based on this approach.

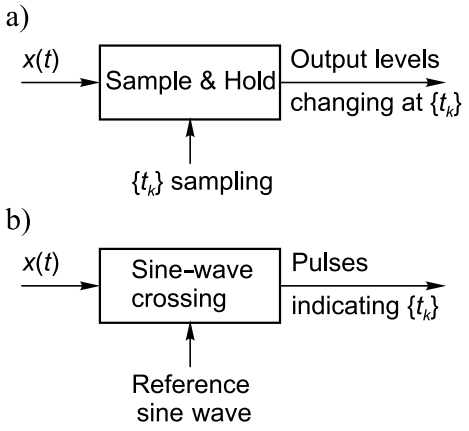


Figure 7.5 Output signal representations at various approaches to sampling

7.4 Remote Sampling Based on Sine-Wave Crossings

The signal digitizing and processing priorities vary with conditions that are typical for a specific task. Data acquisition from multiple signal sources, by definition, is based on data gathering and transporting over shorter or longer distances. Therefore their performance depends directly on the quality of the techniques used to fulfil these functions. The sampling techniques, based on timing of the signal and reference waveform crossing events, are actually considered in this book first of all because they are especially well suited for meeting the requirements of such data acquisition systems. For that reason, the suggested sampling approach is considered here in the context of this special remote sampling application.

The differences in various typical sampling realizations are emphasized in Figure 7.5. The diagram given in Figure 7.5(a) corresponds to the classical so-called 'Shannon sampling', which is usually based on application of a S&H circuit. Analog voltage (current) levels, changing at time instants t_k when the latest sample value replaces the previous one, represent the output of such a sampler in this case. As this kind of analog signal is sensitive to noise, they are usually immediately converted into their digital counterparts by a quantizer located in the near vicinity. More often than not, both schemes of the sampler and quantizer are put on a single crystal that is installed in an ADC chip.

The uniqueness of the sampling based on the signal and sine-wave crossings, illustrated in Figure 7.5(b), is a consequence of the involved operational principle and the distinctly different sort of signal representing the output of such a sampler.

This output signal is given as a nonequidistantly spaced single bit stream. In the operational context of a particular signal processing system, it could be considered as a signal carrying the timing information vital for recovering the input signal in a digital form. Physically, it is a sequence of pulses appearing precisely at the input signal and sine-wave crossing time instants.

If this signal is compared to the analog voltage levels representing the sampler output in the first case, then an important conclusion can be made:

> Sampling based on sine-wave crossings is better suited for execution of this operation from a distance than the conventional amplitude sampling.

This conclusion is actually based on the differences in the sampled signal representations in both cases. Signals sampled on the basis of sine-wave crossings carrying the timing information are relatively insensitive to the impact of surrounding noise. Consequently, they could be transmitted with small distortions over considerably longer distances than the signals reflecting the results of the conventional amplitude sampling. That is obviously crucial for realization of sampling as a remote operation.

Architectures of signal processing systems apparently depend on the type of samplers used. Embedding samplers based on sine-wave crossings into signal processing systems make it possible to restructure these systems on the basis of remote sampling and introduction of the concept for distributed analog-to-digital conversions. If this approach is used, a single distributed ADC could replace a group of ADCs digitizing signals from a multitude of signal sources. To achieve this, the classical general structure of an ADC is converted into the structure of a distributed ADC. Therefore a sufficient number of samplers should be used with communication links inserted between the samplers and a single common quantizer. The issue of massive data acquisition based on this type of distributed ADCs is discussed in Chapter 11.

Note that information in this case is again carried by nonuniform flows of symbols. However, this nonuniformity is usually not and often even could not be exploited for realization of the aliasing suppressing function. Although suppression of aliasing is also achievable in this case, special techniques need to be used. The application potential of sampling based on sine-wave crossings depends more on its capability to provide other desirable benefits exclusively obtainable at signal sampling. However, there are also problems. The point is that, for a number of reasons, the specific sampling based on sine-wave crossings considered here is applicable in a limited frequency range and there are also other factors narrowing the application range of these techniques. Attention is drawn to the most significant of these limitations.

7.5 Advantages and Disadvantages

There are various types of restrictions imposed on potential applications of the sampling techniques based on sine-wave crossings. A start will be made by pointing out some parametric limitations.

First of all, there should be a trade-off between the sampling resolution measured in an achievable number of bits and the angle of signal/sine-wave crossings. On the one hand, it is better if the frequency of the reference sine-wave crossing the signal is relatively high. The steepness of the reference waveform crossing the input signal should be sufficient so that the crossing time instant is fixed more precisely and the impact of noise is less damaging. On the other hand, the error in crossing time definition increases with this steepness and that translates into more pronounced errors in detection of the corresponding reference signal sample values. Therefore the reference signal frequency has to be chosen with these considerations in mind.

The most reliable devices, the quality of which to a large extent determines the achievable bandwidth and sampling precision, are microelectronic comparators. They form the output signals, fixing the input and the reference signal crossing instants, with certain delays. In an ideal case, this kind of delay should be constant. In reality, it more or less depends on the threshold crossing overdrive. The impact of these overdrive-induced delay variations has to be kept within given margins.

For these reasons, the frequency of the input signals sampled on the basis of sine-wave crossings should be limited so that it does not exceed a certain level. Basically this sampling technique is well suited for handling low-frequency signals. The achievable upper frequency given in absolute figures depends, of course, on the quality of the microelectronic products produced and offered in any given period of time. The performance level of the microelectronic products continuously goes up and, consequently, the frequency range for the usable reference signals widens. At the time when this book is being written, acceptable sampling precision has been obtained in the reference frequency range stretching up to 10–30 MHz.

Yet another limitation of a more important nature is related to the fact that the threshold-crossing sampling can not be regarded as a universal technique for signal digitizing. Actually, this type of sampling is quite specific and its successful usage requires appropriate processing of the digitized signals. The basic difficulty, of course, stems from the unusual nonuniformity of the sampling event spacing, as this leads to the necessity of using special algorithms that take care of all related problems and signal processing in this particular area is not yet sufficiently mature.

Nevertheless sinusoid-crossing sampling has some very attractive features unparalleled by other sampling techniques. Therefore it has a high application potential in some areas. If skilfully used, exploitation of these techniques could provide remarkable benefits. Moreover, the difficulties caused by the nonuniform spacing of the sampled signal values, under certain conditions, could be overcome in an amazingly simple and effective way. An example of encoding and reconstructing an electrocardiogram (ECG) signal is demonstrated below.

As sampling based on sine-wave crossings in general are to be used for digitizing signals at relatively low frequencies, these techniques compete with the traditional periodic sampling positioned rather strongly in this application range. Therefore there has to be a good reason to take the decision to use the sinusoid-crossing sampling instead of this type of classic mature periodic sampling. The specific threshold-crossing techniques, of course, are not competing with periodic amplitude sampling in the whole digital signal processing area. Their application area is limited. However, there is a relatively broad area where application of specific sampling techniques based on sine-wave crossings could play a significant positive role. Specifically, they represent an attractive option for building systems that bridge the gap between computers and the real world. Data acquisition systems, built on the basis of this sampling concept, have features that make them competitive in this field.

Sinusoid-crossing sampling is well suited for executing the sampling operation from a distance and the designs of the microelectronic devices realizing such remote sampling are extremely simple, with all of the consequences that relate to this fact. If this type of remote action sampler is incorporated into a structure of a distributed ADC and is used for building systems for data acquisition on a large scale, a number of significant advantages could be achieved. Some of them, specifically, are the following: appropriate conditions for sampling signals very close to the signal source; extremely simple designs of the input devices; low power consumption of the data acquisition front devices; one-bit representation of data positioned in time; transmission of data that are relatively insensitive to noise over relatively long distances; fast and robust logic driven sequential activation of the multiple input channels substituting input signal switching (multiplexing) sensitive to errors; a very large number of inputs measured at least in hundreds.

More is said about obtaining these massive data acquisition features in Chapter 11, where this subject is discussed in some detail. These massive data acquisition techniques look interesting and are worth considering as a possible basis for development of various sensor systems and networks. Biomedical applications may be mentioned as an area especially well suited for their use in picking up signals from a large collection of different types of sensors.

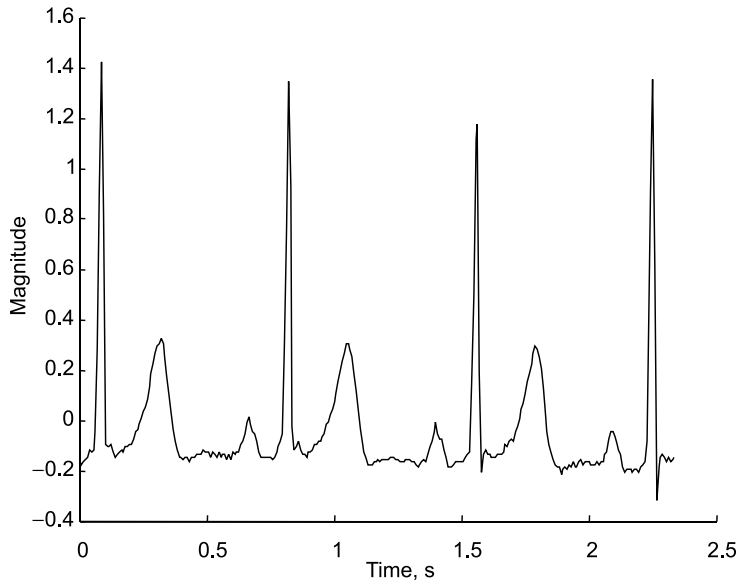


Figure 7.6 ECG waveform reconstructed from a digital signal obtained as a result of sampling performed on the basis of sine-wave crossings

On the other hand, while the listed advantages look attractive, the fact remains that the data acquisition process is indirectly randomized, which may in some cases discourage applications of it. To show that this potential disadvantage is not always damaging and that using unfamiliar specific algorithms for handling nonuniformly represented data is not always necessary, consider an example related to encoding and reconstruction of a typical signal for biomedicine, an ECG signal, illustrated in Figure 7.6.

A picked-up and conditioned cardio-signal was sampled according to the scheme of sine-wave crossings given in Figure 7.3(b). The frequency of the reference signal was chosen to be 500 kHz and the mean sampling rate was set close to 1 kS (kilosample)/s. This means that only one half-period in 1000 was used as a time slot within which the crossings of the input and the reference signals were taken into account. The values of the reference signal corresponding to the crossing instants occurring during these time slots were taken as the sampled input signal values. Under the specified sampling conditions, the randomness of the sampling intervals plays an insignificant role. Therefore it was assumed that the sine-wave crossings occur in the middle of sampling time slots that are 1 μs long and that the sampling process is consequently periodic. The waveform

reconstructed under this assumption is very close to the waveform recovered taking into account exact sampling conditions. It is impossible to see the difference between the cardiograms reconstructed in the exact and in the approximate way ignoring the nonuniformity of the sampling intervals. This means that the assumption about the periodicity of the sampling process is justified under the given conditions, which simplifies reconstruction of the sampled signals significantly. Signals are reconstructed directly from the values of the reference signal detected at the input and reference signal crossings.

Of course, there are no problems when these signals are sampled periodically by performing the sampling operation in the traditional amplitude-sampling manner as the signal spectrum is band limited and the upper frequency is relatively low. However, application of the sampling based on sine-wave crossings might prove to be preferential for designing, on this basis, multichannel massive data acquisition systems. Such sampling performed under the given parameters of the sampling process would provide data acquisition for up to 500 inputs as only one in 500 periods of the reference signal is used for sampling a particular input signal in this case. Therefore up to 499 periods of the reference signal are free and could be activated for sampling additional signals picked up from other signal sources.

The achievable quantization resolution largely depends on the duration of the time slots within which the devices compare the values of the input and the reference signals. In the case of this particular example, 1 μs long time slots should provide at least 12-bit quantization resolution for such devices as, for example, time-to-digital converters TDC-GPX have resolutions up to 27 ps.

Bibliography

Daria, V.R. and Saloma, C. (2000) High-accuracy Fourier transform interferometry, without oversampling, with a 1-bit analog-to-digital converter. *Applied Optics*, **39**(1), 1 January.

Kumaresan, R. and Wang, Y. (2001) On representing signals using only timing information. *J. Acoust. Soc. Am.*, **110**(5), Pt. 1, November.

Lim, M. and Saloma, C. (1998) Direct signal recovery from threshold crossings. *Physical Review E*, **58**(5), November.

Logan, B.F. (1977) Information in the zero crossings of band-pass signals. *Bell Syst. Tech. J.*, **56**, 487–510.

Marvasti, F. *A Unified Approach to Zero-Crossings and Nonuniform Sampling of Single and Multidimensional Signals and Systems*. Princeton, NJ: NONUNIFORM Publication.

Nazario, M.A. and Saloma, C. (1998) Signal recovery in sinusoid-crossing sampling by use of the minimum-negativity constraint. *Applied Optics*, **37**(14), 10 May.

Zeevi, Y.Y., Gravriely, A. and Shamai-Shitz, S. (1987) Image representation by zero and sine-wave crossings. *J. Opt. Soc. Am. A*, **4**, 2045–60.

8

Derivatives of Periodic Sampling

It is not easy to benefit from application of nonuniform sampling and at the same time not to suffer from the negative effects that usually accompany processing of the signals sampled in this way. To make these anti-aliasing techniques practically applicable, the typical negative effects corrupting this kind of signal processing have somehow to be suppressed. A search for a good approach to this problem has led to the recent development of hybrid sampling methods, discussed in Chapter 10. They are based on the idea of mixing the elements of periodic and randomized sampling in order to gain from exploiting the advantages of both. The derivatives of periodic sampling discussed in this chapter are considered as essential building blocks for composing such hybrid periodic/nonuniform sampling models. Especially useful are periodic sampling point sequences with random skips. As shown in the following chapters, whenever this type of sampling procedure is used, the consequences of the sampling phase shifting have to be understood in order to take them into account. To reveal the essential relationships between the phase shifting of sampling and the sampled signal reconstruction conditions, it is desirable to visualize the involved signal transformation processes. To do this it is suggested that estimation of Fourier coefficients should be considered as a process rather than a calculation of a parameter and that this approach should be used to visualize the dynamics of the involved processes. This proves to be a convenient and productive technique for studies and will be repeatedly used henceforth.

Digital Alias-free Signal Processing I. Bilinskis

Generation of the pseudo-randomized sampling pulse sequences used for analog-to-digital conversions suitable for alias-free signal processing is yet another essential application of the model for periodic sampling with random skips. The basic considerations that have to be taken into account when designing generators operating on the basis of this model are discussed as well as practical experience obtained in this area.

8.1 Phase-shifted Periodic Sampling

Introducing phase variations into periodic sampling, if performed correctly, leads to suppression of aliasing. To gain from such an approach, it is essential to realize exactly how the conditions for aliasing change when the phase of the sampling process is varied.

8.1.1 Dependence of Aliasing on the Sampling Phase

The popular signal model will be used according to which a signal $x(t)$ is composed of a multitude of sinusoidal components at arbitrary frequencies in a limited frequency band. Suppose that such a signal is sampled periodically. The sampling frequency is denoted as f_s. Then the sampling point process is periodic with the period $T = 1/f_s$ and the signal sample values are taken at sampling instants $t_k = kT + \Delta T$, $k = 0, 1, 2, \ldots$, where $\Delta T < T$ is the phase angle of the sampling point process.

It was discovered a long time ago that the following equality holds for a signal component at frequency f_x and corresponding aliasing frequencies:

$$A \sin(2\pi f_x t_k + \varphi_x) = A \sin[2\pi(mf_s \mp f_x)t_k + \varphi_m], \quad m = 0, 1, 2, 3, \ldots, \tag{8.1}$$

where φ_m is the phase of the corresponding aliasing frequency. Equality (8.1) is based on the relationships

$$2\pi(mf_s \mp f_x)t_k = 2\pi(mf_s \mp f_x)(kT + \Delta T) = 2\pi mk + \frac{2\pi m\ \Delta T}{T} \mp 2\pi f_x t_k$$

and

$$\sin(2\pi mk) = 0, \quad \cos(2\pi mk) = 1.$$

Then

$$\begin{aligned}\sin[2\pi(mf_{\mathrm{s}} \mp f_x)t_k] &= \sin\left(2\pi mk + \frac{2\pi m\ \Delta T}{T} \mp 2\pi f_x t_k\right)\\ &= \sin(2\pi mk)\cos\left(\frac{2\pi m\ \Delta T}{T} \mp 2\pi f_x t_k\right)\\ &\quad + \cos(2\pi mk)\sin\left(\frac{2\pi m\ \Delta T}{T} \mp 2\pi f_x t_k\right)\\ &= \sin\left(\frac{2\pi m\ \Delta T}{T} \mp 2\pi f_x t_k\right),\end{aligned}$$

which leads to Equation (8.1). This means that the signal sample values, taken from signal components at frequencies belonging to the row $mf_{\mathrm{s}} \mp f_x$, do overlap and that these frequencies are indistinguishable.

Conditions of this overlapping are reflected by the relationships defining the dependence of ϕ_x on the phase angle of the sampling point process:

$$\varphi_x = \begin{cases} -\left(\varphi_m + \dfrac{2\pi m\ \Delta T}{T} + \pi\right), & \text{for } (mf_{\mathrm{s}} - f_x),\\ \varphi_m + \dfrac{2\pi m\ \Delta T}{T}, & \text{for } (mf_{\mathrm{s}} + f_x).\end{cases} \tag{8.2}$$

In the case where the sampling phase $\Delta T = 0$ and $t_k = kT, k = 0, 1, 2, \ldots,$ the phase angle φ_x does not depend on the phase shift of sampling. Under these conditions the relationship between the phase angle of the signal down-converted to the baseband and the phase of the aliasing frequencies is as follows:

$$\varphi_x = \begin{cases} -(\varphi_m + \pi), & \text{for } (mf_{\mathrm{s}} - f_x),\\ \varphi_m, & \text{for } (mf_{\mathrm{s}} + f_x).\end{cases} \tag{8.3}$$

In all other cases, the phase angle φ_x also depends on the phase shift of the sampling pulse process. For instance, when the particular aliasing frequency converted down to the baseband $[0, 0.5f_{\mathrm{s}}]$ is $f_1 = f_{\mathrm{s}} - f_x$, the following equation holds:

$$x_{1k} = A_1 \sin(2\pi f_1 t_k) = A_1 \sin\left(2\pi f_x t_k - \frac{2\pi\,\Delta T}{T} - \pi\right). \tag{8.4}$$

Similarly,

$$x_{2k} = A_2 \sin(2\pi f_2 t_k) = A_2 \sin\left(2\pi f_x t_k + \frac{2\pi\,\Delta T}{T}\right) \quad \text{for } f_2 = f_\text{s} + f_x, \tag{8.5}$$

$$x_{3k} = A_3 \sin(2\pi f_3 t_k) = A_3 \sin\left(2\pi f_x t_k - \frac{4\pi\,\Delta T}{T} - \pi\right) \quad \text{for } f_3 = 2f_\text{s} - f_x, \tag{8.6}$$

and

$$x_{4k} = A_4 \sin(2\pi f_4 t_k) = A_4 \sin\left(2\pi f_x t_k + \frac{4\pi\,\Delta T}{T}\right) \quad \text{for } f_4 = 2f_\text{s} + f_x. \tag{8.7}$$

These relationships demonstrate a remarkable fact. They show that conditions for frequency overlapping or aliasing essentially depend not only on the sampling frequency but also on the sampling phase. The fact that the phase angle of the signal, downconverted by aliasing to the baseband, depends on the phase of the sampling point process is illustrated in Figure 8.1. The original signal in this case has just one sinusoidal component at frequency f_1 (Figure 8.1(a)). When it is sampled so that $\Delta T = 0$ (Figure 8.1(b)), the phase angle φ_{a1} of the alias reconstructed within the baseband [0, 0.5 f_s] is equal to the phase φ_x of the original signal (Figure 8.1(d)). If the same signal is sampled at the same sampling frequency but all sampling instants are shifted to a position characterised by $\Delta\varphi T = 0.75T$ (Figure 8.1(c)), the phase angle φ_{a1} of the reconstructed alias changes in accordance with the relationships (8.2), as shown in Figure 8.1(d), diagram 2.

8.1.2 Reconstruction of Sampled Signals

Consider what happens when reconstruction of a signal sampled at time instants belonging to variable-phase periodic sampling point processes is attempted on the basis of direct and inverse Fourier transforms. The Fourier coefficients are estimated at all frequencies within the frequency band of the signal including the frequencies of a particular row $mf_\text{s} \mp f_x$. The signal might have a component at the frequency f_x or not, and other components at some of the indicated frequencies mf_s might overlap this specific frequency. In general, the results of Fourier coefficient estimation at this frequency depend on the true signal component, the overlapping aliases and the sampling conditions, specifically the phase of the periodic sampling

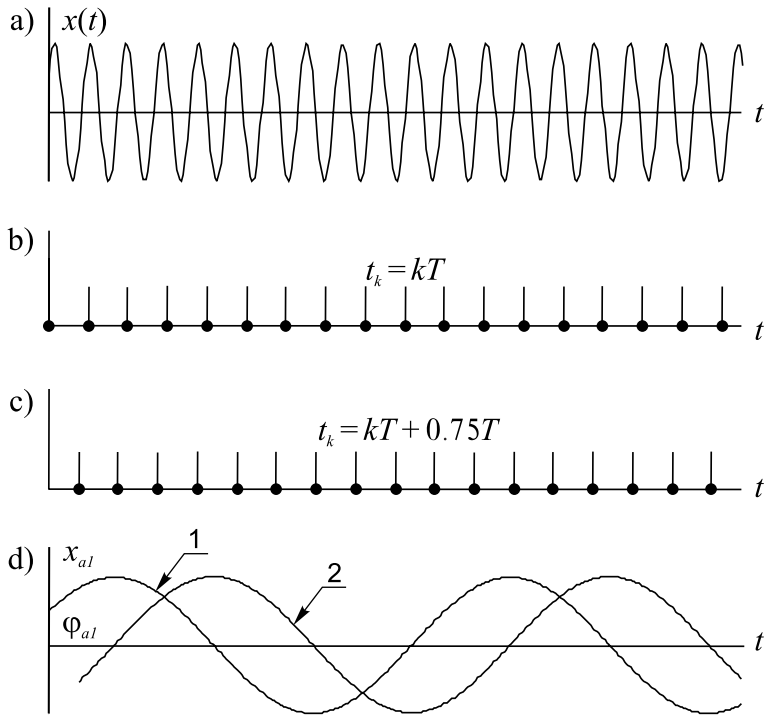


Figure 8.1 Illustration of the fact that the phase angle of the signal, downconverted by aliasing, depends on the phase of the sampling point process

point process. Attention is drawn to the fact that the phase shift of the sampling point process affects only the overlapping aliases. Variations of this phase do not affect the contribution of the true signal component to the Fourier coefficient estimate at all. There are at least two significant consequences of that:

1. The phase of the periodic sampling point process affects the waveforms reconstructed from the obtained signal sample values.
2. Using variable phase periodic sampling opens up the possibility of avoiding overlapping of the aliasing frequencies.

How this possibility might be realized is discussed in Chapter 10. Before studying this essential issue in some detail, variation of the complementary waveforms of signal components at frequencies belonging to the row of the aliasing frequencies $mf_s \mp f_x$ will briefly be considered. The diagrams shown in Figure 8.2 illustrate this kind of waveform variation, which take place when they are reconstructed

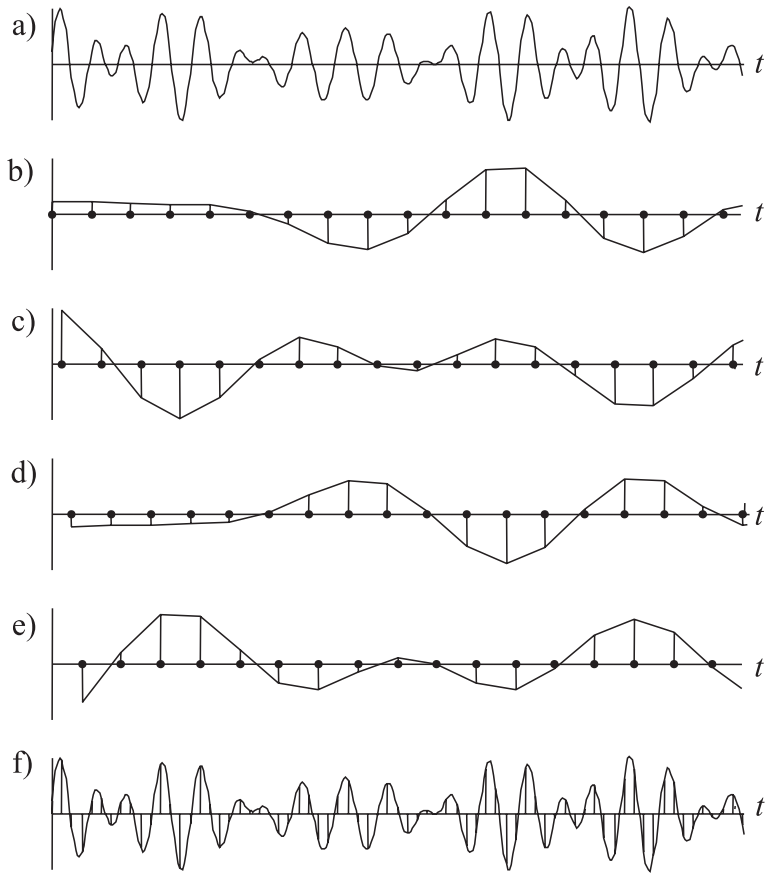


Figure 8.2 Waveform variations in the case when they are reconstructed from the sampled signal obtained as a result of variable-phase periodic sampling. The original signal (a) is periodically sampled periodically at different phase shifts: $\Delta T = 0, 0.25T, 0.5T, 0.75T$. Adding the respective reconstructed waveforms (b, c, d, e) provides the waveform (f)

from the signal sample values obtained under conditions of variable phase periodic sampling. In this particular case, the analog signal consists of three sinusoids.

It can be seen from Figure 8.2 that, each of the particular waveforms is actually meaningless. However, using all of them together leads to reconstruction of the original signal. Therefore the information carried by the signal sample values, obtained at variable-phase periodic sampling, makes it possible to avoid aliasing and to reconstruct the original signal not corrupted by frequency overlapping.

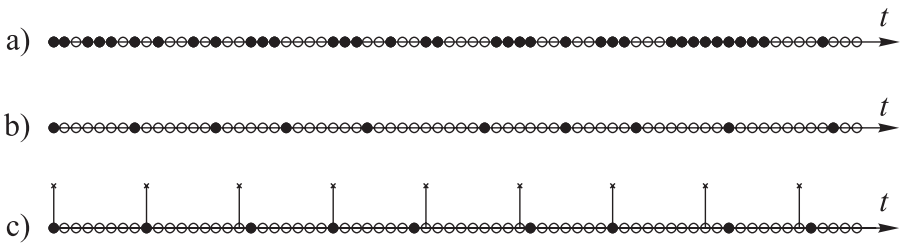


Figure 8.3 Various periodic sampling point processes with random or pseudo-random skips

The result obtained in this particular case is not surprising. Indeed, using four phase-shifted periodic sampling processes in parallel is obviously equivalent to sampling at a frequency four times higher. Actually the purpose of giving this example is just to show that variations of the sampling phase makes it possible to obtain data that might be useful for avoiding aliasing. In the case of processing a stationary signal, such phase-shifted sampling should not necessarily be performed in parallel.

8.2 Periodic Sampling with Random Skips

In many cases of basically periodic sampling, taking of signal sample values at some of the periodically repeating sampling instants is omitted. This kind of sampling might take place either in result of some faults in system functioning or it might be pre-planned and carried out in this way intentionally. We are basically considering the second case. Then the periodic sampling process with random skips is organised and carried out in order to obtain some specific effect. Before discussing particular applications of such sampling, let us consider this general sampling model in some detail.

8.2.1 General Model

The basis of the sampling approach is a periodic point process with randomly or pseudo-randomly excluded sampling points. There are a number of sampling point processes belonging to this category. The difference between them is in the pattern of the skipped sampling points, illustrated in Figure 8.3.

A periodic sampling point process with random skips is shown in Figure 8.3(a). The sampling events take place in this case with some probabilities p_k. Therefore the probability that no sample value will be taken at the sampling time instant t_k is equal to $(1 - p_k)$. As shown in the following chapters, this sampling model

is actually a generic one. It represents the basis for decompositions of various sampling point processes and, in addition, it plays an essential role for the hybrid sampling schemes discussed in Chapter 10. It may also be used, for example, to describe a sampling process when some faults occur as a result of electronic system malfunctioning or when faults are due to signal collisions with interfering bursts of noise.

The sampling point processes with pseudo-random skips given in Figures 8.3(b) and (c) define the conditions of sampling performed according to the earlier discussed models of additive and periodic sampling with jitter respectively. However, the sampling schemes in this case are digital and the smallest digit is equal to the period δ of the time grid. Otherwise these schemes are defined as explained in Chapter 6. In the case of additive random sampling illustrated in Figure 8.3(b), signal samples are taken at sampling time instants

$$t_k = t_{k-1} + \tau_k, \quad k = 0, 1, 2, \ldots, \tag{8.8}$$

where τ_k is a realization of a digital pseudo-random variable.

In a typical pseudo-randomized sampling case, this digital variable is distributed uniformly within the interval $[\mu \pm 0.5\Delta\mathrm{t}]$, where μ is the mean value of the intervals between the sampling events and the interval $\Delta t < \mu$ so that the distance between two successive sample value taking instants is never shorter than some specified limit. When sampling is performed according to the sampling point pattern displayed in Figure 8.3(c), the sampling instants $\{t_k\}$ in it are given as

$$t_k = kT + \tau_k, \quad k = 0, 1, 2, \ldots. \tag{8.9}$$

The period $T = n\delta$ and the digital variable τ_k is again distributed in an interval shorter than the period.

Of course, these few cases do not exhaust the various sampling schemes based of the model of periodic sampling with random skips. There are more. Nevertheless, all of them have some common features. A high-frequency clock provides their time basis. The frequency $1/\delta$ is usually stabilized. Specific sampling models are then implemented by choosing and setting up a particular scheme, according to which the sampling points are skipped. The pattern of the missing sampling points define most of the characteristics of the resulting specific sampling model. This approach is well suited for organizing various pseudo-randomized sampling modes. The essential advantage of this type of sampling process is the possibility of using for signal processing the information about the exact sample value taking instants and the fixed pattern of the empty slots between the sampling time instants.

8.2.2 Typical Use

Periodic sampling point processes with random skips cover a broad class of various processes often used in the area of randomized digitizing of signals. First of all, the pseudo-randomized clock pulse sequences should be mentioned as the central factor dictating the time instants exactly when the signal sample values are taken according to the time schedule. In addition, they also control other related functions of the respective digitizer. The quality of the sampling pulse sequences, generated according to the preset sampling regime, has to be high as the performance standard of the corresponding system largely depends on it. Execution of the signal processing algorithms is usually based on the assumption that the signal sample values are taken exactly at the predetermined time instants. The reference values generated in the computer are typically tied to these time instants as well. Therefore any discrepancy between the predetermined and realized signal sample value taking acts leads to serious negative consequences. The generation of these pseudo-randomized clock pulse sequences is a responsible function, which needs to be based on a well-developed knowledge basis. A typical example of a developed and successfully exploited electronic system of this kind is given in Section 11.1.

Another area where the periodic sampling point processes with random skips play an essential role is hybrid sampling systems. Although there are various options on how to build such hybrid systems on the basis of combining elements of deterministic and randomized signal processing, they typically use the considered periodic sampling point processes with random skips as essential components, discussed in Chapter 10.

The last but not the least important application area of periodic sampling point processes with random skips is signal processing adapted to sampling nonuniformities. This particular technique has a high application potential and is considered in Chapter 18.

8.3 Compensation Effect

Decomposing periodic sampling point processes with random skips into various frequency periodic phase-shifted components is a technique helping to provide an insight into aliasing processes. In turn this is essential for analysis and synthesis of digital alias-free signal processing algorithms. This technique often provides good results, especially when it is used in combination with the approach to Fourier transforms that presents them as processes rather than just calculations of

Fourier coefficients. Then an interesting effect, discussed here as a compensation effect, can be observed. It reveals the essence of alias suppression and shows how this anti-aliasing mechanism works. This effect and examples of its use are described in this section.

8.3.1 Display of Fourier Transforms

For studies of various side effects of sampling, it is often convenient to use a special approach to Fourier transforms. To realize how and why something happens, application of special functions makes it possible to observe these transforms as processes developing in time. These functions are cumulative sums that are actually the basis for Fourier transforms. They can be given as

$$\begin{aligned} \tilde{a}_{\mathrm{a}}(n) &= \frac{2}{N}\sum_{k=0}^{n-1} x_k \cos(2\pi f_{\mathrm{a}} t_k) \\ \tilde{b}_{\mathrm{a}}(n) &= \frac{2}{N}\sum_{k=0}^{n-1} x_k \sin(2\pi f_{\mathrm{a}} t_k) \end{aligned} \tag{8.10}$$

where f_{a} is the frequency of DFT filtering.

Consider a signal model

$$\begin{aligned} x_k &= a_{\mathrm{r}} \cos(2\pi f_{\mathrm{r}} t_k) + b_{\mathrm{r}} \sin(2\pi f_{\mathrm{r}} t_k) \\ &= a_{\mathrm{r}} \cos[2\pi(mf_{\mathrm{s}} \mp f_{\mathrm{a}})t_k] + b_{\mathrm{r}} \sin[2\pi(mf_{\mathrm{s}} \mp f_{\mathrm{a}})t_k], \quad k = \overline{0,\ N-1} \end{aligned}$$

where N is the number of the signal samples, $f_{\mathrm{r}} = mf_{\mathrm{s}} \mp f_{\mathrm{a}}$, $m = 0, 1, 2, \ldots$ is the signal frequency and f_{s} is the sampling rate. Substituting into Equations (8.10) yields

$$\begin{aligned} \tilde{a}_{\mathrm{a}}(n) &= a_{\mathrm{r}}\frac{2}{N}\sum_{k=0}^{n-1} \cos(2\pi f_{\mathrm{r}} t_k)\cos(2\pi f_{\mathrm{a}} t_k) + b_{\mathrm{r}}\frac{2}{N}\sum_{k=0}^{n-1} \sin(2\pi f_{\mathrm{r}} t_k)\cos(2\pi f_{\mathrm{a}} t_k) \\ \tilde{b}_{\mathrm{a}}(n) &= a_{\mathrm{r}}\frac{2}{N}\sum_{k=0}^{n-1} \cos(2\pi f_{\mathrm{r}} t_k)\sin(2\pi f_{\mathrm{a}} t_k) + b_{\mathrm{r}}\frac{2}{N}\sum_{k=0}^{n-1} \sin(2\pi f_{\mathrm{r}} t_k)\sin(2\pi f_{\mathrm{a}} t_k) \end{aligned} \tag{8.11}$$

To demonstrate the usefulness of this approach, graphical images reflecting DFT filtering of true and aliasing signal components are compared.

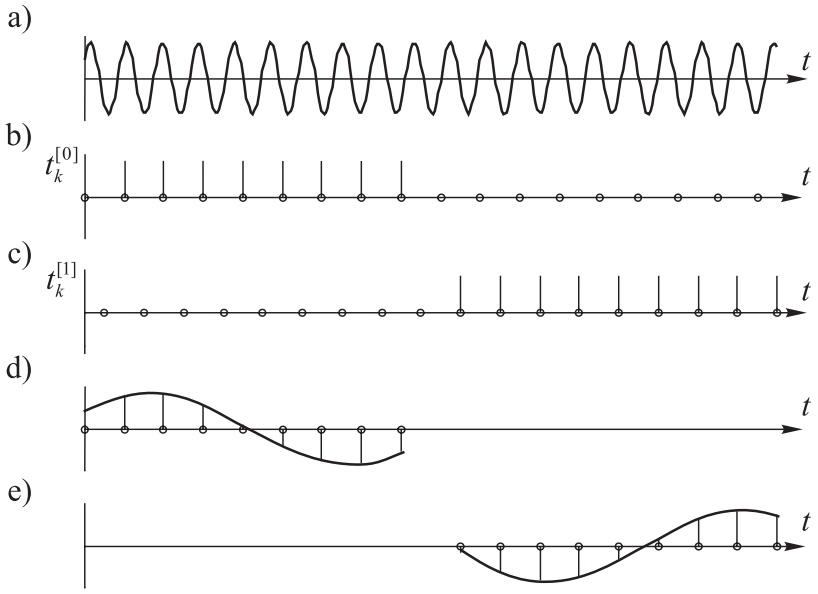


Figure 8.4 Results of sampling a signal according to a two-segment phase-shifted sampling scheme: (a) original signal; (b), (c) phase-shifted sampling point sequences; (d), (e) downconverted signal segments

8.3.2 Observing the Aliasing Processes

Suppose that the defined signal, depicted in Figure 8.4(a), is sampled periodically and after taking the first batch of $N/2$ sample values the sampling pulse stream is shifted for half of the sampling period so that the following $N/2$ sample values are taken with this phase shift introduced as shown in Figures 8.4(b) and (c). As the sampling frequency, in the illustrated general case, is lower than the signal frequency, aliasing occurs and the signal is downconverted into the baseband frequencies. The phase angles of the downconverted signal segments change following phase shifting of the sampling process. This effect could be used for reconstruction of the original signal although the sampling frequency is below the Nyquist limit. The reconstructed downconverted signal waveforms are given in Figures 8.4(d) and (e).

Now the cumulative sums introduced in the previous section can be used as a convenient tool for observing the processes related to aliasing. Consider a particular case where

$$x_k = a_\mathrm{r} \cos(2\pi f_\mathrm{r} t_k) + b_\mathrm{r} \sin(2\pi f_\mathrm{r} t_k)$$

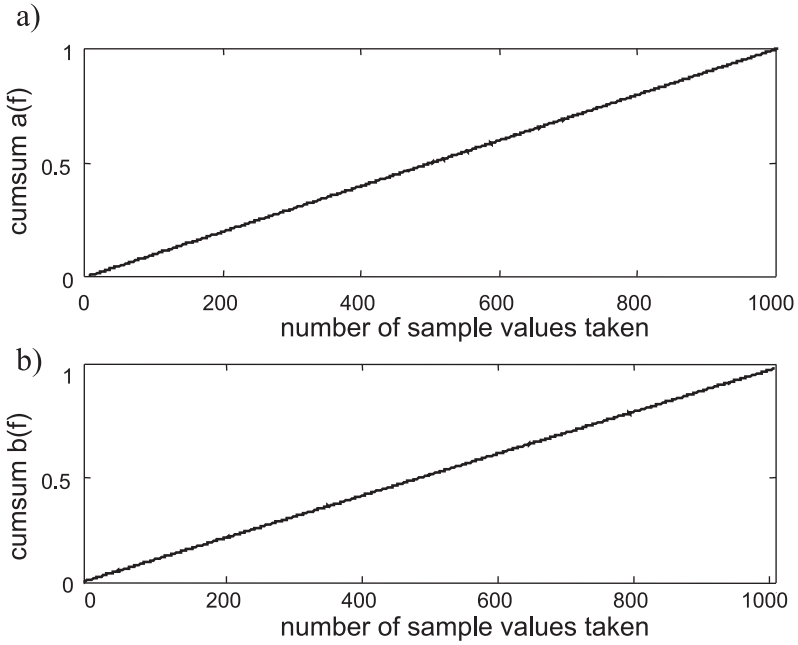


Figure 8.5 Cumulative sums (cumsum) for an existing real signal component

and $a_r = 1$ and $b_r = 1$. The signal frequency f_r is set at first below the Nyquist limit and is equal to the filtering frequency f_a. The cumulative sum in this particular case graphically looks like a linearly increasing function for both segments of the sampling process and is shown in Figure 8.5. If the filtering frequency f_a is chosen so that it is equal to an aliasing frequency, then the cumulative sum looks like the broken lines shown in Figure 8.6.

It can be seen from Figure 8.6 that the cumulative sum, obtained for a particular Fourier coefficient in the case where there is a signal component aliasing to the frequency of DFT filtering, is a positively increasing linear function for the first segment of the sampling point process and is equally a negatively increasing function for the second shifted segment of the sampling point process. In other words, the cumulative sum formed at the stage of the first segment is compensated by the cumulative sum accumulated during the time of the second segment of the sampling point process. The end result is close to zero, which it should be in this case when there is no signal at the frequency of analysis and only an aliasing signal component is present.

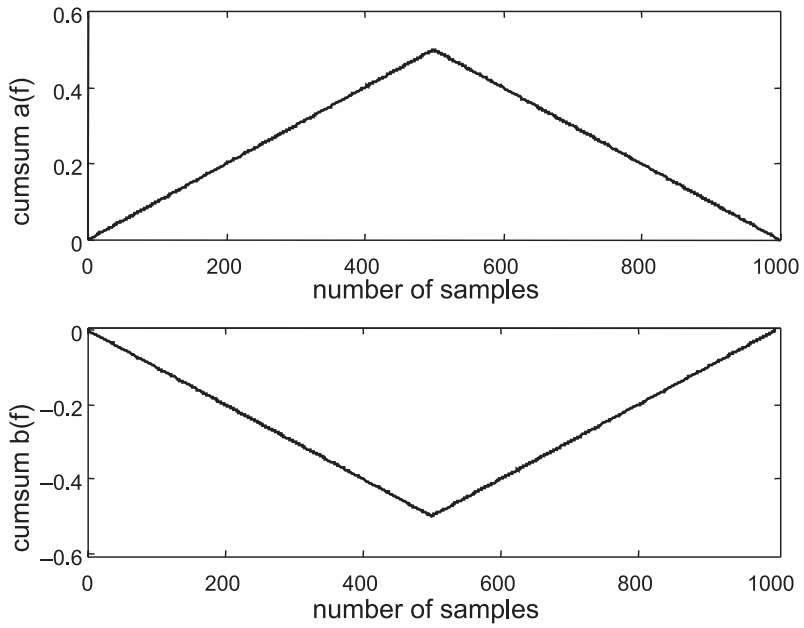


Figure 8.6 Cumulative sums in the case where DFT filtering is performed at one of the aliasing frequencies

Therefore, using the sampling point processes formed by appropriate shifting of segments of the sampling point process produces the compensation effect, taking out the signal aliases as described above. This compensation effect naturally depends on various factors. Nevertheless, it is there and could be exploited with results better than those usually obtained by application of nonuniform sampling. If the involved Fourier transforms could be carried out simply as calculations of the Fourier coefficient estimates, it would be easy not to notice this significant effect, not to observe it and not to take it into account.

The sampling pulse streams, shown in Figures 8.4(b) and (c), are respectively defined as follows:

$$t_k^{[0]} = kT_\mathrm{s}, \quad k = \overline{0, N/2-1},$$

$$t_k^{[1]} = kT_\mathrm{s} + 1.5T_\mathrm{s} + t_{N/2-1}^{[0]} = kT_\mathrm{s} + \frac{N+1}{2}T_\mathrm{s}, \quad k = \overline{0, N/2-1},$$

where $T_\mathrm{s} = 1/f_\mathrm{s}$ is the period of sampling. Formulae for the respective cumulative sums can be written as follows:

1. For signal frequencies $mf_s - f_a$ and $n \leq N/2$, it is found that

$$\tilde{a}_a\left(n \leq \frac{N}{2}\right) = \frac{a_r}{N}\left\{n + \frac{\sin(2\pi f_0 T_s n)}{\sin(2\pi f_0 T_s)}\cos[2\pi f_0 T_s(n-1)]\right\} - \frac{b_r}{N}\left\{\frac{\sin(2\pi f_0 T_s n)}{\sin(2\pi f_0 T_s)}\sin[2\pi f_0 T_s(n-1)]\right\}, \tag{8.12}$$

$$\tilde{b}_a\left(n \leq \frac{N}{2}\right) = \frac{a_r}{N}\left\{\frac{\sin(2\pi f_0 T_s n)}{\sin(2\pi f_0 T_s)}\sin[2\pi f_0 T_s(n-1)]\right\} - \frac{b_r}{N}\left\{n - \frac{\sin(2\pi f_0 T_s n)}{\sin(2\pi f_0 T_s)}\cos[2\pi f_0 T_s(n-1)]\right\}. \tag{8.13}$$

2. For signal frequencies $mf_s - f_a$ and $n > N/2$,

$$\tilde{a}\left(n > \frac{N}{2}\right) = \frac{a_r}{N}\left[\frac{N}{2} + \frac{\sin(\pi f_0 T_s N)}{\sin(2\pi f_0 T_s)}\cos\alpha_1\right] - \frac{b_r}{N}\left[\frac{\sin(\pi f_0 T_s N)}{\sin(2\pi f_0 T_s)}\sin\alpha_1\right] + \frac{(-1)^{m(N+1)}a_r}{N}\left[n - \frac{N}{2} + \frac{\sin\alpha_2}{\sin(2\pi f_0 T_s)}\cos\alpha_3\right] - \frac{(-1)^{m(N+1)}b_r}{N}\left[\frac{\sin\alpha_2}{\sin(2\pi f_0 T_s)}\sin\alpha_3\right], \tag{8.14}$$

$$\tilde{b}_a\left(n > \frac{N}{2}\right) = \frac{a_r}{N}\left(\frac{\sin(\pi f_0 T_s N)}{\sin(2\pi f_0 T_s)}\sin\alpha_1\right] - \frac{b_r}{N}\left(\frac{N}{2} + \frac{\sin(\pi f_0 T_s N)}{\sin(2\pi f_0 T_s)}\cos\alpha_1\right) + \frac{(-1)^{m(N+1)}a_r}{N}\left[\frac{\sin\alpha_2}{\sin(2\pi f_0 T_s)}\sin\alpha_3\right] - \frac{(-1)^{m(N+1)}b_r}{N}\left[n - \frac{N}{2} - \frac{\sin\alpha_2}{\sin(2\pi f_0 T_s)}\cos\alpha_3\right], \tag{8.15}$$

where

$$\alpha_1 = 2\pi f_0 T_s\left(\frac{N}{2} - 1\right), \qquad \alpha_2 = 2\pi f_0 T_s\left(n - \frac{N}{2}\right), \qquad \alpha_3 = 2\pi f_0 T_s\left(n + \frac{N}{2}\right). \tag{8.16}$$

Similar formulae can be derived for signal frequencies $mf_s + f_a$.

This technique for visualizing the processes related to aliasing is a convenient tool. Application of it for analysis of various signal processing operations often proves to be effective for displaying and presenting them in a comprehensive form. This is demonstrated repeatedly in the following chapters.

8.4 Generation of Randomized Sampling Pulse Trains

Most of the algorithms for digital signal processing are based on the assumption that the time instants of signal sample value taking are known. This condition does not represent a problem for the traditional practice of signal processing. The intervals between the sampling instants are of a constant and known duration and the needed timing data are obtained simply by counting the sampling events. The circumstances are different for processing nonuniformly sampled signals as the intervals between the sampling instants vary by more or less wide margins. Theoretically there are two options: either to measure each sampling interval or to generate the sampling pulses at precisely predetermined instants. As implementing the first option is quite demanding, the second approach to the problem is usually preferable.

Thus the randomized sampling pulse trains, on the one hand, need to be generated according to the requirements of the chosen specific sampling point process and, on the other hand, each pulse has to be generated at the exact predetermined time instant. The generated sampling pulse trains are then used for clocking ADCs and initiating in this way the signal sample value taking.

8.4.1 Basic Approach

The basic principles of generating sampling pulse trains can be easily illustrated using examples of their analog implementations.

Periodic Jittering Pulse Trains

When the distribution of time intervals between successive pulses in the pulse sequence vary within relatively narrow boundaries, pulse trains can be formed in a simple way. A sinusoidal function is generated, a random process $\xi(t)$ is added and the pulses are formed at the time instants when the combined process crosses the zero level.

Additive Random Pulse Trains

It is convenient to use multivibrators to generate this kind of pulse train. Operation of a multivibrator is based on a relatively slow exponential charging and abrupt

discharging process of a capacitor. The capacitor is discharged at time instants when the capacitor voltage $u_c(t)$, which increases during the process of charging, reaches some threshold level. In the case of this application, a random process $\xi(t)$ is added to the normally constant threshold level so that this threshold is the sum of a constant voltage $u_0(t)$ and $\xi(t)$. When the power of $\xi(t)$ is negligible, the multivibrator relaxes almost periodically. In cases where the power of $\xi(t)$ is more or less substantial, the discharging instants are more or less influenced by this random process. Multivibrators are appropriate for this application because after the discharges of the capacitor they begin the next charging cycle, disregarding the prehistory without taking the timing of past oscillations into account. This means that intervals between generated sampling pulses are statistically independent as required.

However, this is true only under the assumption that the multivibrators being used are ideal. Under real conditions these intervals may be slightly correlated, but it has been found to be of no real significance. It should be noted that multivibrators become very sensitive to noise during the short time intervals preceding capacitor discharges. This means that the sampling pulse generators of this kind have to be properly screened to prevent their synchronization with outside interference.

Pseudo-random Pulse Trains

The applicability of the analog methods for generating sampling pulse trains is limited. Much more universal are the digital methods. It is important to understand that by using them the problem of forming the sampling pulses is solved in a predetermined way so that their exact positions on the time axis are known beforehand. Such digital sampling pulse trains are actually pseudo-random.

Generation of pseudo-random sampling pulse trains, in principle, is straightforward. They are built on the basis of a periodic clock fulfilling the essential time reference function. To form these pseudo-random pulse trains, pulses are randomly taken from the clock pulse sequence in accordance with the sampling point process to be implemented. A pseudo-random number generator is used to generate the random variables needed for dictating the distances between the generated pulses.

The properties of the pulse trains generated in this way are influenced both by the periodic clock pulse sequence and by the sequence of pseudo-random numbers used. While the latter determines the sampling point density function, characterizing the pulse trains obtained, the former is responsible for aliasing, which may occur if the signals sampled by means of such pulse trains contain components exceeding half the clock frequency.

The most valuable feature of the pseudo-random pulse trains is, of course, that they are fully determined by the clock frequency and by the sequence of

pseudo-random numbers used. This means that: (a) such pulse trains can be reproduced at will whenever this is necessary; (b) the nth pulse will always appear at an exactly predetermined instant; (c) descriptions of particular pseudo-random pulse trains can be stored in computer memory and this information can then be used to calculate the time instants when some other required functions or signals should be read out, so that they and the signals are sampled simultaneously.

8.4.2 Practical Experience

While the matter of pseudo-randomized sampling pulse generation at first glance seems to be simple, in reality the situation is different. The performance of digital alias-free signal processing systems directly depends on the quality of the generators producing these sampling pulse sequences. To provide wideband operation in the frequency range up to several GHz and to achieve high precision for signal processing, the sampling pulses often have to be generated with a very short time digit and to be formed very close in time to the respective predetermined instants so that errors do not exceed a few picoseconds. The sampling pulse sequence also has the responsibility of providing a stable and precise time reference. The ever-present jitter needs to be kept within a strictly limited margin as well. Excessive jitter, for instance, leads to spurious frequencies that corrupt signal processing in the frequency domain.

The first decision that has to be taken in the initial stage of designing such a generator is related to the problem of obtaining a precise and stable time reference. One option is to depend on the clock frequency. Implementation of this approach leads to quite good results in the case of relatively low frequency designs. However, the situation is different when the smallest time digit is in the subnanosecond range. For example, when microwave signals are processed digitally at frequencies up to 1 or 2 GHz, the required clock frequency would be twice as much and the necessity to build electronic systems operating at these frequencies substantially complicates the system designs.

In such cases another option based on the use of signal propagation delays as the time reference seems to be more appropriate. This approach could be implemented on the basis of considerably lower frequency electronics and therefore the system designs would then be much simpler. The practical experience obtained confirms this assumption. High-performance sampling pulse generators have been developed and made on this basis. Exploitation of various systems containing this kind of sampling pulse generator has provided good results. However, there is also an intrinsic drawback characterizing this approach. These are problems related to the jitter control. The generation process is then based on switching various time delay elements to generate the sampling pulses with a predetermined proper

pattern of nonuniform distances between them. It is relatively difficult to ensure in this case that all generated sampling pulses are placed exactly on the time grid. Hence there are imperfections in the generation process, which are observed as jitter of the sampling pulse generation instants.

It seems that the best approach to the generation of sampling pulse trains is to combine both techniques. This method is illustrated later in Figure 11.3 where a block diagram of a generator implementing this method is shown. Such a combination of these techniques makes it possible to use moderate clock frequencies and the quantity of the used delay elements is reduced to a relatively small number. As a result, the generator shown in Figure 11.3 generates sampling pulses with high precision while the jitter does not exceed a few picoseconds. To form sampling pulses according to a predetermined nonuniform pattern, the pulse forming process is based on controllable division of the clock frequency followed by controllable introduction of time delays. The generator of pseudo-random numbers provides the code for digital control of this process. This is discussed in more detail in Chapter 11.

To continue the illustration of this example, some figures characterizing a particular generator are given. The periodic clock pulses were generated at the rate of 669.3266 MHz. The clock frequency was divided by a random integer ranging from 9 to 16. Then the output pulses were expanded in width (up to approximately 5 ns). An adjustable delay line and a high-speed multiplexer were used to implement a single-bit controllable delay block. As a result, the intervals between the generated sampling pulses at the output of the generator were pseudo-randomly varied from 13.447 to 24.652 ns with the smallest time digit equal to 747 ps. This generator is adapted for joint operation over a wide range of ADCs with the maximum sampling rate over 80 MS/s and a broad analog bandwidth. Although in the particular considered case the mean sampling rate is only 53.546 MS/s, the equivalent sampling rate is much higher, providing for digital alias-free signal processing in the bandwidth up to 669.3 MHz. It should be emphasized that the introduced tight control of the sampling pulse jitter made it possible to achieve a bandwidth free from the spurious frequencies due to sampling imperfections.

Bibliography

Artjuhs, J. and Bilinskis, I. (2006) Method and apparatus for alias suppressed digitizing of high frequency analog signals. EP 1 330 036 B1, European Patent Specification, Bulletin 2006/26, 28.06.2006.

Artyukh, Yu., Bilinskis, I. and Vedin, V. (1999) Hardware core of the family of digital RF signal PC-based analyzers. In Proceedings of the 1999 International Workshop on *Sampling Theory and Application*, Loen, Norway, 11–14 August 1999, pp. 177–9.

Artyukh, Yu., Bilinskis, I., Boole, E., Rybakov, A. and Vedin, V. (2005) Wideband RF signal digititising for high purity spectral analysis. In Proceedings of the International Workshop on *Spectral Methods and Multirate Signal Processing* (*SMMSP 2005*), Riga, Latvia, 20–22 June 2005, pp. 123–8.

Bilinskis, I. and Artjuhs, J. (2006) Method and apparatus for alias suppressed digitizing of high frequency analog signals. United States Patent US 7,046,183 B2, 16 May, 2006.

Bilinskis, I. and Mikelsons, A. (1983) *Digital Random Processing of Continuous Signals* (in Russian). Riga: Zinatne.

Bilinskis, I. and Mikelsons, A. (1990) Application of randomized or irregular sampling as an anti-aliasing technique. In *Signal Processing*, V: *Theories and Application*. Amsterdam: Elsevier Science Publishers, pp. 505–8.

9

Fuzzy Aliasing

This chapter is dedicated to detailed studies of the aliasing processes observed in cases where the sampling operation is performed nonuniformly. It is shown that randomizing of sampling can attenuate the aliasing effect. The aliases are not taken out completely at this stage. Some aliasing related effects are still present but they are diffused and not so well defined as in cases of periodic sampling. The term 'fuzzy' fits well to this kind of aliasing as there are numerous multiple-frequency contributions to this kind of aliasing process. Various aspects of fuzzy aliasing are discussed. It is emphasized that successful elimination of aliases, crucial for high-performance digital signal processing at frequencies exceeding half the sampling rate, could be achieved only by a skilful combination of randomized sampling techniques with application of specific algorithms. Both the sampling and signal processing stages have to be arranged in the correct way. The signal processing algorithms provide resolution of the frequency-overlapping problem, which is made feasible by proper randomization of sampling. Some of these special algorithms for proper alias-free signal processing are described in Part 2 of this book.

As demonstrated in the following chapter, a clear insight into the processes underpinning fuzzy aliasing is an essential precondition for development of new techniques for signal processing, including hybrid algorithms and algorithms adapted to sampling nonuniformities.

9.1 Meaning of the DFT of a Nonuniformly Sampled Signal

Suppose that a single tone signal at frequency f_x is sampled nonuniformly according to the definition of the additive sampling point process. If the DFT of this

Digital Alias-free Signal Processing I. Bilinskis

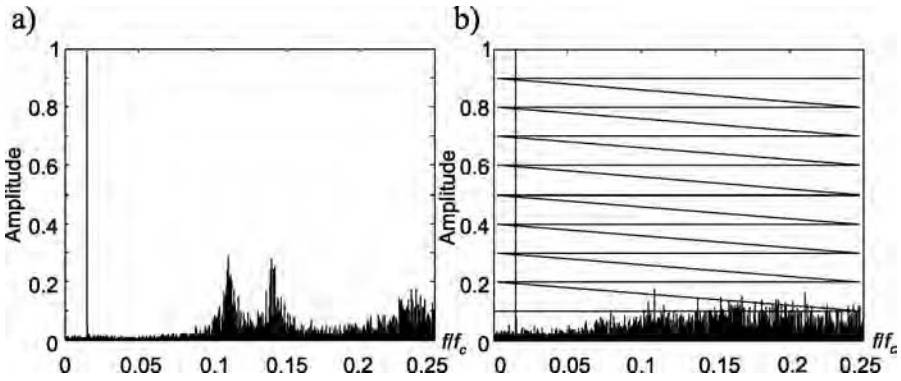


Figure 9.1 Typical result of the DFT of a nonuniformly sampled single tone signal: (a) diagram obtained in the case of relatively small ratio σ/μ; (b) diagram obtained at an increased value of the ratio σ/μ

sampled signal is performed, typical diagrams, usually considered to be spectrograms, are obtained as a result. They are given in Figure 9.1.

In the first case, the mean additive sampling rate is equal to $1/\mu = 1/8\delta$, the sampling intervals are equal to $\mu + \tau_k$ and $\mu = 8\delta$ and the random variable τ_k assumes the values $(0, \pm\delta)$ with equal probability. The second diagram has been obtained under similar conditions. The only difference is that the random variable τ_k in the second case is distributed in a wider interval. Specifically, it assumes, with equal probability, one of the values $(0, \pm\delta, \pm 2\delta, \pm 3\delta, \ldots)$. It can be seen from these diagrams that the first of them has distinct peaks at some frequencies. It is found that the peaks are located in the vicinity of the frequencies belonging to the following row: $f_x, 1/\mu \pm f_x, 2/\mu \pm f_x, \ldots$. There are no such peaks in the second diagram. This means that increasing the width of the distribution of the random variable τ_k indicated above has led to a much more pronounced suppression of aliasing.

At first glance it seems that everything is clear. To avoid aliasing, the sampling process simply has to be randomized to a sufficiently high degree. Then there would be no aliases, only some background noise like the noise present in the diagram of Figure 9.1(b). Thus it is easy to get the illusion that everything is more or less fine and that randomizing of sampling indeed provides elimination of aliases. It is also realized that this positive effect is to some extent spoiled by the noise that appears as a result of the sampling randomization. However, it is usually accepted as an unavoidable negative effect as the appearance of it is a natural consequence of the randomness introduced at the sampling. Therefore

the conclusions often made are that the dynamic range characterizing spectrum analysis of nonuniformly sampled signals is poor and that nothing can really be done to improve it.

Such conclusions would, however, be wrong. The results of the DFT obtained for a nonuniformly sampled signal under the given conditions should be considered from a different angle. The fact that the signal has been sampled nonuniformly changes the situation. The customary approach used in the periodic sampling cases is not fully applicable automatically when signals are sampled nonuniformly. While the DFT of periodically sampled signals provides spectrograms indicating how the amplitudes of signal components are distributed in the frequency domain, the estimates of Fourier coefficients, obtained as a result of a direct DFT of a band-limited nonuniformly sampled signal, reflect the features of both the signal and the involved sampling point processes. Therefore the multitude of these spectral estimates represents only a half-ready spectrogram of the respective signal. The effects observed and accepted as aliasing related ones are in fact more complicated than it might seem at first glance. In fact they contain numerous multifrequency contributions closely related to these aliasing processes. This is the essence of fuzzy aliasing, which distorts the raw spectrograms obtained as a result of direct application of the DFT. To obtain final alias-free spectrograms of signals not corrupted by side effects of the sampling process, additional processing of these raw spectrograms is carried out in some special ways.

Apparently, comprehending the impact of fuzzy aliasing on the nonuniformly sampled signals is crucial for achieving the capability of alias-free digital processing of signals in wide dynamic and frequency ranges. Various aspects of fuzzy aliasing are discussed in the following sections.

9.2 Concept of Fuzzy Aliasing

While pseudo-randomized sampling might be realized in accordance with various sampling models, this kind of sampling is by definition digital. This means that every sampling instant is locked to a regular time grid. Therefore all parameters of the sampling point process built on this basis are expressed in numbers of digital quantities with the smallest time digit equal to the period of the respective clocking pulse sequence.

9.2.1 Generic Periodic Sampling with Random Skips

As the basis of the sampling approach considered here is a periodic point process with pseudo-randomly excluded sampling pulses, it is essential to focus studies

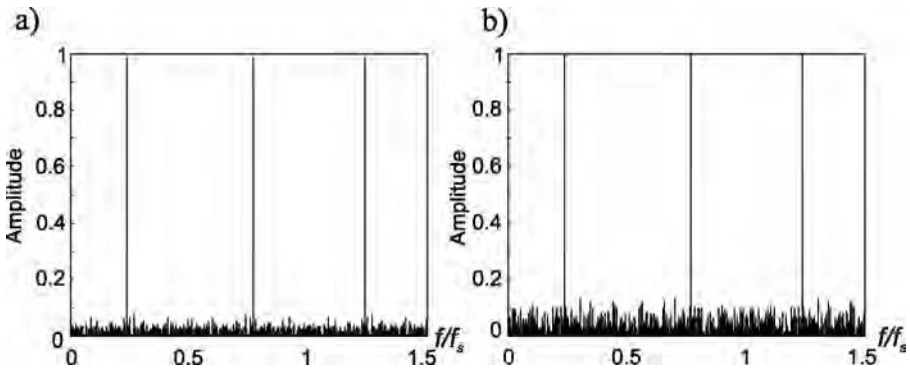


Figure 9.2 Spectrograms obtained in the case when sample values of a single-tone signal are taken periodically with some of them pseudo-randomly skipped: (a) $N = 2048$; (b) $N = 512$

on features of this basic component of the sampling process. The goal of these studies is to discover practical applicable methods for reduction of errors that impair spectral analysis and waveform reconstruction of a wide class of signals. The targeted error reduction is to be based on more extensive use of all *a priori* information. The pattern of pseudo-random skips is typically fixed, so first a method must be found showing how this information can be used more effectively.

Aliasing for the periodic sampling pulse streams with pseudo-random skips is well-defined, shown by the spectrograms given in Figure 9.2. It can be seen that the aliases are located symmetrically to the sampling frequency and their amplitudes are equal to the amplitude of the signal. However, the noise due to the cross-interference is still present.

Broadly speaking, periodic sampling with random skipping of signal sample values belongs to the class of randomized sampling processes. However, the features of this kind of randomized sampling differ greatly from those characteristic of the randomized sampling processes discussed, specifically, in Chapter 3. Indeed, the basic motivation for deliberate randomization of sampling so far has been the avoidance of aliasing. In this particular randomization case this effect is not achieved. Therefore there must be other reasons why such a sampling model is considered at all.

Actually there are some good reasons. Firstly, it is important to use fault-tolerant signal processing to discover how to reconstruct spectra and waveforms of signals when the process of their sampling has been impaired and reduced to the process of periodic sampling with random skips by some interference and noise. Secondly,

estimation of signal spectral parameters in the case of such sampling is a vital component of more complex algorithms for alias-free signal sampling and processing.

In both cases, it is essential to find a good approach to digital signal parameter estimation improvement applicable under the considered specific signal sampling conditions. To achieve this, the dominating relationships characterizing such sampling and processing of the digital signals obtained in this way need to be studied. It is clear that the targeted error reduction should be based on more effective use of all *a priori* information than has been done in this particular area so far. The real challenge is to take out, or at least suppress, the impact of the interference due to irregular sampling of other signal components. This problem is considered in Chapter 18.

9.2.2 Primary and Secondary Aliasing

Consider an additive pseudo-randomized sampling point process. It is formed on the basis of the periodic clock process so that the sampling instants correspond to the definition of this kind of sampling. Note that in general it is essentially a specific periodic process with pseudo-randomly skipped points. The frequency of the basic periodic process is $f_c = 1/\delta$ and the mean sampling rate $1/\mu$ is several times lower. The ratio of these frequencies is evidently equal to μ/δ. To simplify the analysis of fuzzy aliasing, assume that the distribution of the intervals between the sampling instants is uniform. It is characterized simply by the mean value μ, the distribution interval of the random variable τ_k or the standard deviation σ.

To observe the aliasing effects affecting the sampling processes carried out according to this periodic sampling scheme with randomized skips, the diagram given in Figure 9.3 is used. It reflects the results of the DFT obtained for a sinusoidal signal at frequency f_x sampled in accordance with the digital additive sampling scheme in the frequency range exceeding the clock frequency f_c.

At some frequencies there are full-scale aliases and at other frequencies there are smaller diffused peaks. It is easy to find that no matter what is the pattern of the missing sampling points, frequencies belonging to the row f_x, $f_c \pm f_x$, $2f_c \pm f_x$, ... fully overlap. This happens as a result of the aliasing effect usually observed. In the context of this analysis, consider this kind of aliasing as primary aliasing. At pseudo-randomised sampling the clock frequency is usually set up at a sufficiently high level to guarantee that there are no primary aliasing frequencies within the bandwidth of the input signal.

However, there is also a secondary aliasing effect. This depends on the pattern of skipped sampling points and is not so well defined and might be described

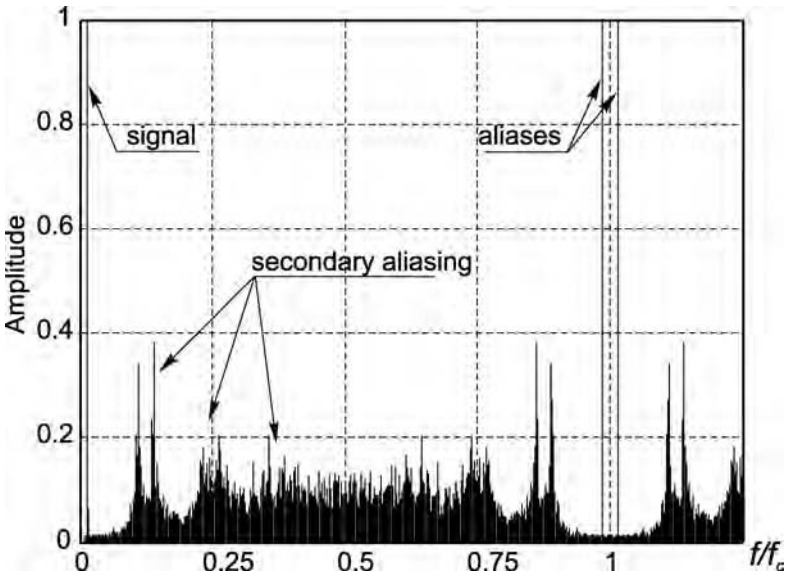


Figure 9.3 Typical results of the DFT obtained for a single-tone signal sampled according to the digital additive sampling scheme reflecting the primary and secondary aliasing effects

in probabilistic terms. In the illustrated case of additive sampling, such aliasing takes place in the close vicinity of the frequencies defined as f_x, $1/\mu \pm f_x$, $2/\mu \pm f_x$, ... Frequency overlapping at these frequencies is suppressed, but is observed not only at the indicated frequencies but also in the areas near to them. The peaks observed at this secondary type of aliasing are more or less pronounced depending on the parameters of the corresponding sampling process. In the illustrated case, these peaks depend on the interval of distribution of the random variable τ_k (or on the ratio σ/μ) and on the number N of the sample values taken. They become more suppressed with the distribution interval and N increasing and at some randomization level these peaks are completely washed out. Then the secondary aliasing is not noticed at all as there is a more powerful noise in the background. It is tempting to say that under these conditions there is no aliasing. Actually this is not true. Frequency overlapping still takes place and the observed background noise is actually the outcome of the aliasing processes occurring in virtually the whole frequency range. The nature of this secondary aliasing effect, which in fact is the basis of fuzzy aliasing, is relatively complicated. As fuzzy aliasing plays an important role in digital alias-free signal processing, it is crucial to obtain a very clear picture of it. The essence and the anatomy of fuzzy aliasing are revealed

in the following sections. The key elements of it are the decompositions of the sampling point processes, which are considered further.

9.2.3 Decomposition of Sampling Point Processes

The quality of nonuniformly sampled signal processing often depends on the degree of sampling-specific error suppression. These errors are mainly due to the imperfection of the considered periodic point process with random skips as most of the used randomized sampling schemes can be tied to this sampling model. Apparently successful error reduction should be based on more extensive use of all *a priori* information available. The pattern of pseudo-random skips is typically fixed and known. Therefore it has to be learned first of all how to use this information more effectively. It is useful to decompose the periodic sampling point process with random skips into a number of component sets at various frequencies. Each of the component sets contain at a specific frequency in turn a number of components shifted in phase for different phase angles.

Once such a decomposition task is approached, it is easy to see that the basic point process with the period δ could be decomposed into two periodic sampling point processes with pseudo-random skips having a period 2δ two times larger and the phase shift between them equal to half of this period. Therefore there are at least two equivalent representations of the basic sampling point process. Actually there are more variable parameter decompositions. Their specifics depend on the type of sampling point process decomposed.

Additive Random Sampling

Figure 9.4 illustrates the decomposition of a periodic sampling point process with random skips in the case where the pattern of the skipped points meets the requirements of the additive random sampling. The decomposition of the basic sampling point process with a period equal to δ (Figure 9.4(a)) into three phase-shifted sampling processes with a period equal to 3δ that is three times larger is shown in Figure 9.4(b). The same approach can be used to decompose the first sampling process into three components at the frequency of a basic randomly decimated periodic process equal to $\frac{1}{3\delta}$. This decomposition has three components shifted in phase for 0, $2\pi/3$ and $4\pi/3$ radians. Apparently this decomposition approach makes it possible to decompose the initial basic sampling point process at frequencies $\frac{1}{4\delta}$, $\frac{1}{5\delta}$, $\frac{1}{6\delta}$, ..., $1/n\delta$, each containing 4, 5, 6, ..., n phase-shifted components respectively. Theoretically, only the number N of the taken signal sample values limits the number of possible decompositions. Indeed, at each

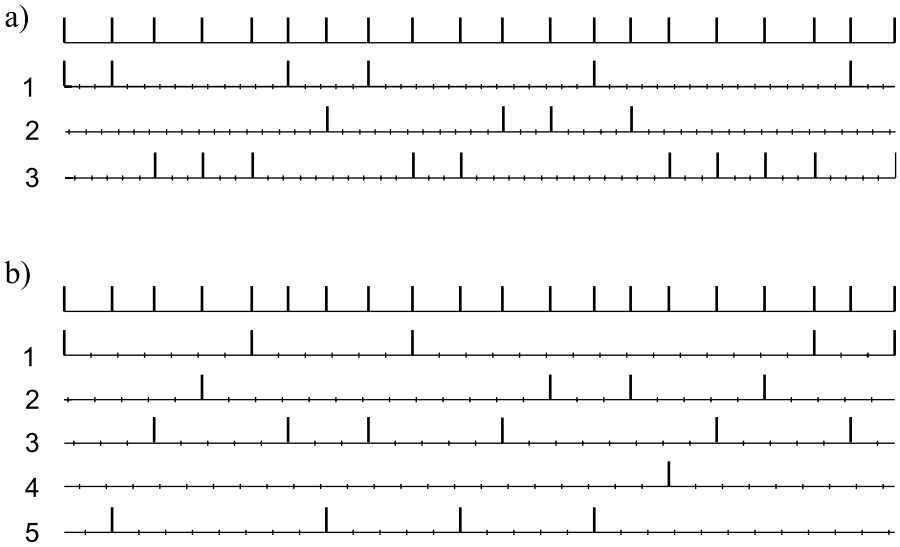


Figure 9.4 Decomposition of a periodic sampling point process with random skips into various components in the case of additive random sampling

decomposition all of the signal sample values are subdivided nonequally between the n components of the particular nth decomposition.

Periodic Sampling with Jitter

Figure 9.5 illustrates the decomposition of periodic sampling points with jitter. In this case the features of the obtained decompositions differ from the characteristic features of the decompositions obtained when the primary sampling point process belongs to the class of additive pseudo-random sampling. Specifically, the periodic sampling points with jitter are decomposed into only three out of eight components. The remaining five other components are empty and do not contain any sampling points. Therefore the conclusion is found that a spectrogram obtained by performing a DFT for a pseudo-randomly sampled signal contains, in addition to estimates of the signal components, multiple aliases due to the primary and the secondary aliasing effects.

The decomposition of the randomized sampling point sequences is an approach that is convenient for studies of various processes related to processing of signals sampled in this way. It can also be used both for analysis of various processes and synthesis of algorithms. For instance, the usefulness of this decomposition

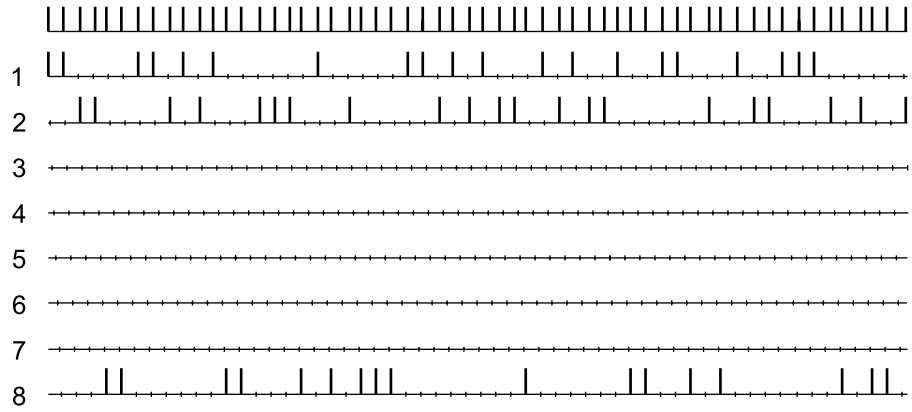


Figure 9.5 Decomposition of a periodic sampling point process with random skips into various components in the case of periodic sampling with jitter

idea is demonstrated in the following chapter, where it is used for development of hybrid sampling models.

9.3 Anatomy of Fuzzy Aliasing

The suggested concept of randomized sampling decomposition provides the instrument for tracking various contributions to the overall fuzzy aliasing phenomenon needed for revealing the anatomy of fuzzy aliasing.

9.3.1 Tracking of Particular Contributions

The row defining the frequencies for primary aliasing is based on the clock frequency f_c. The secondary aliasing takes place at frequencies belonging to a similar frequency row:

$$f_x,\ f_c/n + f_x,\ 2f_c/n + f_x,\ 3f_c/n + f_x, \ldots; \quad n = 2, 3, 4, \ldots. \tag{9.1}$$

where f_c/n are the frequencies of the considered decompositions. The sampling point process is decomposed at each frequency f_c/n into n phase-shifted sampling point streams so that the total number N of the sampling points is subdivided into n random parts. However, the power of the secondary aliases is not distributed in the frequency domain in a way that is obviously related to them. Instead, under certain conditions, the secondary aliasing is dominated by peaks close to the frequencies in the row: $f_x,\ 1/\mu \pm f_x,\ 2/\mu \pm f_x, \ldots$ Therefore not all sampling point

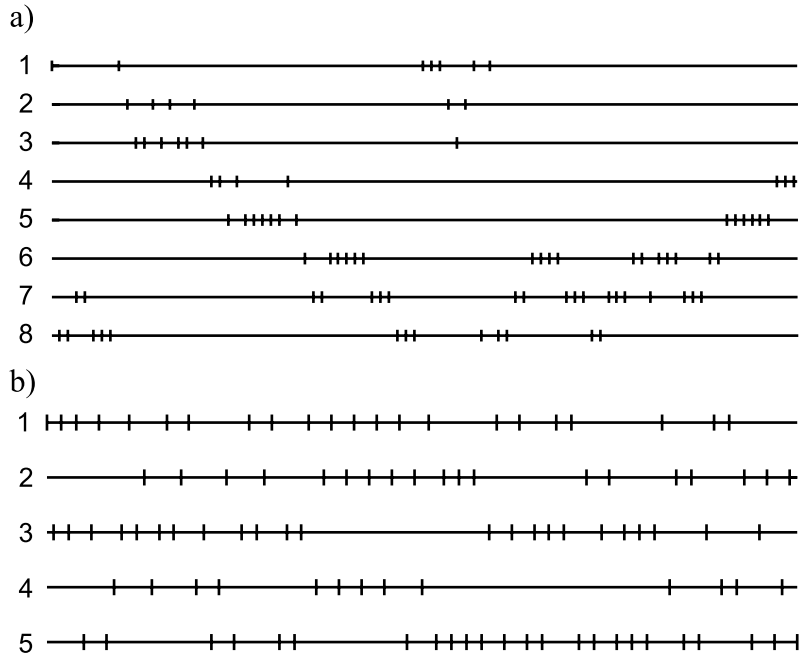


Figure 9.6 Typical differing decompositions of an additive sampling point process: (a) decomposition into eight components; (b) decomposition into five components

decompositions play an equally important role in secondary aliasing. The decomposition into $n = \mu/\delta$ components is special. The question is: why it is so special?

To answer this question, look at the features characterizing the considered decompositions a little more carefully. Two typical decompositions of an additive sampling point process for the cases where $n = \mu/\delta$ and $n \neq \mu/\delta$ are illustrated in Figures 9.6(a) and (b) respectively. The decomposed additive sampling point process is characterized by $\mu = 8\delta$, the sampling intervals are equal to $\mu + \tau_k$ and the random variable τ_k assumes the values $(0, \pm\delta)$ with equal probability. It can be seen that the sampling points drift across the decomposition levels in the first case where $n = \mu/\delta$ and jump chaotically between these levels in the second case where $n \neq \mu/\delta$. That leads to the characteristic decomposition distributions illustrated in Figure 9.6.

In general, the distributions of the sampling points between the constituents represented by various phase-shifted particular periodic processes with random skips are shown in the cases where $n \neq \mu/\delta$ are closer to uniform (histograms a, b, c and e in Figure 9.7). The differences in the quantities of the sampling points belonging to these randomly decimated periodic processes with particular phase

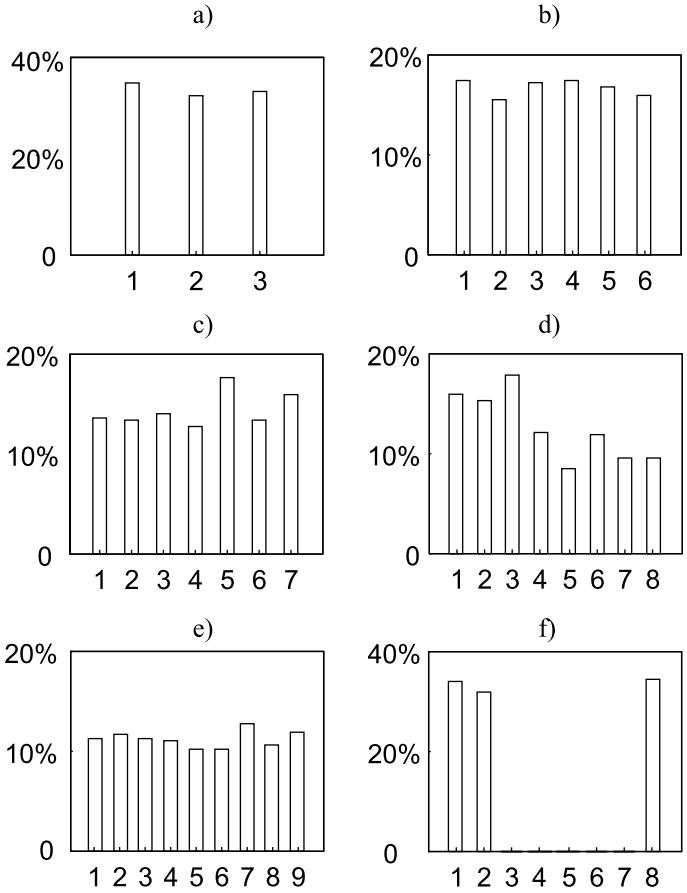


Figure 9.7 Histograms characterizing the distribution of the sampling points between the constituents represented by variously phase-shifted periodic processes with random skips

shifts are then relatively small. They are much more pronounced in the cases where $n = \mu/\delta$ (histograms d and f in Figure 9.7). Consequently, aliasing occurring at decomposition frequencies characterized by $n \neq \mu/\delta$ is better compensated than aliasing taking place at frequencies for which $n = \mu/\delta$.

9.3.2 Incomplete Compensation of Aliases

The background noise, present in the diagrams displaying the results of DFT performed for nonuniformly sampled signals, are actually residues of incompletely compensated aliasing taking place at many frequencies. Compensation of the

phase-shifted particular aliases plays a significant role in the whole multifrequency aliasing mechanism. The particular errors in the estimates of Fourier coefficients are residues of incompletely compensated aliases. This in turn happens because the components of particular decompositions contain an unequal number of signal sample values.

The significance of this compensation effect could be demonstrated by the example of processing periodically sampled signals. The point is that a periodic sampling point stream can also be decomposed in the way described. However, the decompositions would then have different characteristics. For instance, the total number of signal sample values would be subdivided equally between the decomposition components. It is easy to check that the aliases occurring in this case at a particular decomposition frequency would be fully compensated.

9.3.3 Aliasing at Multiple Frequencies

As explained in Section 9.2, every sampling point process could be represented by a number of equivalent decompositions at frequencies $f_c/n, 2f_c/n, 3f_c/n, \ldots$ and each of these decompositions contain n phase-shifted components given as periodic sampling point processes with random skips. Therefore all of them provide some aliasing related effects, which happen for the multitude of indicated frequencies. In addition, the sampling conditions at each of the decomposition frequencies are dictated by several phase-shifted components of the sampling point process. Therefore the compensation mechanism, described in Section 8.3, acts as well and the resulting estimate of the Fourier coefficient calculated at this particular frequency is affected by all of these factors. The secondary aliasing conditions change from frequency to frequency. The total number of the taken signal sample values is subdivided at each of the decomposition frequencies into n nonequal parts. Therefore the contribution to the summary secondary aliasing given by these separate decomposition components, with n growing, declines. Consequently, the total aggregated secondary aliasing effect is rather complicated. It is suggested that the term 'fuzzy aliasing' is used to refer to it.

The impact of fuzzy aliasing on processing of nonuniformly sampled signals is strong and has to be taken into account. Attention is drawn once again to the fact that the processing of nonuniformly sampled signals should be matched to the specifics of randomized sampling. An example elaborating this point follows.

9.4 Object Lesson

The necessary condition for avoiding aliasing is taking signal sample values at nonequally distanced time instants. Meeting this condition of sampling

nonuniformity is obligatory as there is no other way to avoid overlapping of frequencies at signal digitizing and digital processing. However, it does not mean that using nonuniform sampling is sufficient for achieving the desirable elimination of aliases. In concluding the discussion of studies on fuzzy aliasing, the point must be stressed that nonuniform sampling by itself does not completely take out the aliases and does not automatically lead to alias-free signal processing. Although the negative consequences of aliasing in many cases could be avoided, it is impossible to achieve that by relying exclusively on nonuniform sampling alone. This kind of nonequispaced sampling only provides the necessary preconditions for signal processing not corrupted by aliasing.

The perception of the role that randomized sampling plays in digital alias-free signal processing should therefore be adjusted to this concept. Correctly performed nonuniform sampling actually leads to two useful effects. Firstly, it suppresses the magnitude of the aliases. Secondly, it also opens up the possibility of resolving the aliasing problem at the stage when the digitized signals are processed. In fact, the secret of successful digital alias-free signal processing lies in using the correct organization of both sampling and processing operations so that they are effectively combined for elimination of the aliases. The development of algorithms for a particular signal processing should be based on the reality that aliases at randomized sampling are suppressed rather than taken out completely. Special procedures for elimination of aliases should be included in those algorithms.

Figure 9.8 illustrates this point. A wideband signal was nonuniformly sampled and the obtained data were used two times. Firstly, the diagram shown in Figure 9.8(a) was calculated as described in Chapter 15. It represents DFT results of a nonuniformly sampled wideband signal. Evidently it has a high level of background noise, which can now be recognized as the product of fuzzy aliasing. If the transformed signals have been sampled periodically, this diagram would represent the signal spectrogram. In this case, where the sampling procedure has been randomized, the DFT of the sampled signals provides data affected by the randomness of sampling. To obtain a spectrogram in this case, it has somehow to be filtered out of the raw data obtained as a result of the DFT.

To demonstrate this point, the same data were used once more to perform the spectrum analysis more correctly on the basis of a special algorithm taking out the impact of fuzzy aliasing. The resulting spectrogram is given in Figure 9.8(b). It was obtained specifically by using the SECOEX algorithm discussed in Chapter 20. The achieved improvement is obviously significant. Thus this example confirms the earlier comment on the DFT of randomly sampled signals. The results of the DFT should therefore not be considered as spectrograms of the respective signals before the errors imposed by fuzzy aliasing have been eliminated. This

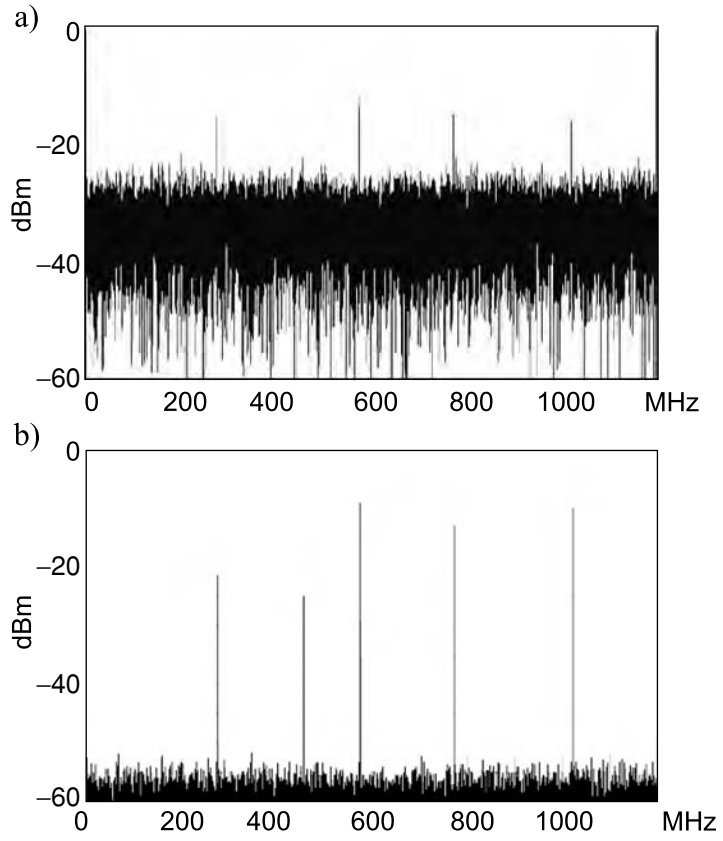


Figure 9.8 Spectrograms of a nonuniformly sampled wideband signal obtained: (a) as a result of the DFT and (b) with special anti-aliasing procedures added at the stage of processing the digitized signal

example also shows how important it is to use the correct tools to process the nonuniformly sampled signals.

Bibliography

Bilinskis, I. and Ziemelis, Z. (2006) Decomposition of random sampling point processes. *Electronics and Electrical Engineering*, **5**(69), 45–8.

10

Hybrid Sampling

Both uniform periodic and nonuniform randomized sampling have their advantages and disadvantages. They are specific and different. However, it can be said that the drawbacks of periodic sampling are countered to some extent by advantages of nonuniform randomized sampling and vice versa. Recognition of this situation has led to attempts to synthesize sampling methods that have integrated positive features of uniform and nonuniform sampling methods and at the same time would not be impaired by their drawbacks. This kind of periodic and pseudo-randomized nonuniform sampling hybrids is discussed in this chapter.

Although purposeful development of hybrid sampling is an activity that began only recently, the basic principles of this approach have been in use for a long time. For instance, pseudo-randomized sampling may be mentioned as a case where a periodic clocking process is combined with pseudo-random skipping of the sampling pulses. Therefore it is not very easy to draw a line between pseudo-randomized and hybrid sampling models. In fact they partly overlap. The hybrid sampling techniques discussed here are more elaborate and specific than the basic pseudo-randomized sampling techniques.

These sampling techniques also offer an additional opportunity to improve the certainty that the signal sample values will be taken at exact predetermined time instants. This is essential for high-performance signal processing as the sampling instants then have to be known with sufficient resolution no matter how the sampling process is organized. Although good results have been achieved in this area and it is possible to pinpoint the sampling instants with picosecond resolution, the electronic blocks ensuring that sampling is performed exactly at predetermined time instants are complicated and represent the most expensive part of alias-free digitizers. The concept of hybrid sampling reveals how

Digital Alias-free Signal Processing I. Bilinskis

pseudo-randomized nonuniform sampling might be realized on the basis of a few phase-shifted stabilized periodic processes and this approach is well suited for building sampling drivers characterized by reduced sampling instant jitter.

10.1 Hybrids of Periodic and Random Sampling

The development of hybrid sampling schemes is based on various combinations of periodic and randomized sampling elements. The features of these combinations differ. The basic issue is how to build the hybrid sampling schemes so that they provide results at minimal incurred costs. Such costs are given in terms of design hardware and software complexity. The discussions will start with the subject of hybrid sampling by considering a model based on the superposition of pseudo-randomly decimated periodic sampling point processes.

10.1.1 Basic Approach

Suppose that a sampling pulse stream, uniform or nonuniform, is used for sampling a signal and sample values of it are taken at instants $\{t_k\}$ given on a discrete time grid with a uniform step of the grid equal to δ. The step δ of the time grid determines the Nyquist frequency limit for the given stream of sampling pulses. This upper frequency limit f_{u} is apparently defined as $f_{\mathrm{u}} = \frac{1}{2\delta}$ and holds both for the uniform and nonuniform sampling pulse streams formed on the basis of this time grid.

A particular approach to signal sampling is based on the use of n periodic sampling pulse streams. Although the periods $T = n\delta$ of all these pulse streams are the same, they are each shifted in phase for $\Delta\varphi = 2m\pi/n, m = 0, 1, 2, \ldots, (n-1)$, and in time for $\Delta t = m\delta$. Thus the smallest shift in time is equal to δ. The resulting sampling pulse stream is obtained by superposition of all these partial streams.

Obviously, the formed stream represents just a periodic sampling process with frequency n times higher than the frequency of the partial streams. Using such shifted interleaving periodic sampling pulse streams for signal sampling by a number of ADCs connected in parallel is a common practice. This approach is used when the required sampling rate exceeds the capabilities of the ADC used. Now suppose that the sampling operation is performed by a single ADC under the conditions that the highest sampling rate applicable for it is several times lower than the frequency of the resulting sampling process obtained in this way. It is evident that some quantity of the sampling pulses has to be taken out. In other words, the partial sampling pulse streams have to be decimated. That has

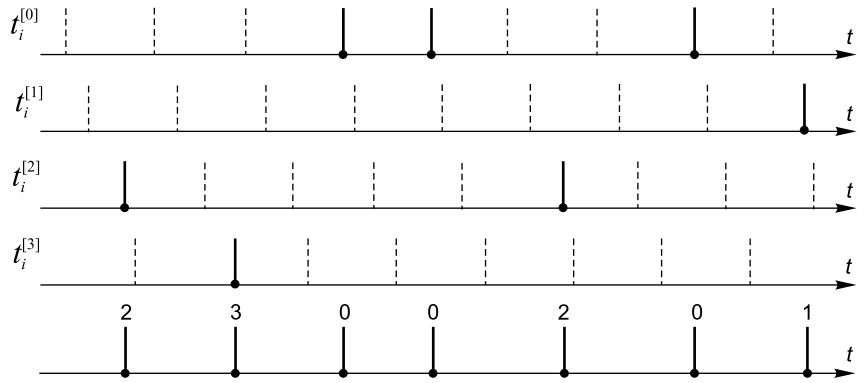


Figure 10.1 Generation of a nonuniform stream of sampling pulses from partial pseudo-randomly decimated phase-shifted periodic streams

to be done in a pseudo-random way so that the smallest distance between two adjacent sampling instants is either equal to or larger than the shortest sampling time interval permitted for the specific ADC used. Figure 10.1 illustrates the formation of this kind of nonuniform stream of sampling pulses from partial pseudo-randomly decimated shifted periodic streams.

It can be seen that the sampling process formed in this way represents a classic pseudo-randomized sampling point process. It actually corresponds to the additive sampling process and, as shown in the previous chapter, could be interpreted as a superposition of periodic sampling point processes with random skips.

However, this model of hybrid sampling essentially differs from the basic pseudo-randomized sampling model. Firstly, each sampling instant is marked with the number of the specific periodic sampling point process to which it belongs. Secondly, sample values marked by the same numbers at digital processing of the sampled signals are grouped into separate subsequences that are processed separately so that the phase shift of the particular periodic sampling pulse stream that they belong to is taken into account. Thirdly, the suggested approach to generation of sampling point processes based on superposition of a number of periodic phase-shifted sampling point processes with random skips could be implemented in a way that is much better than the earlier approach used to generate of the additive sampling point processes based on the use of a set of controllable time delay elements. More will be said about this in Section 10.4.

To use this approach successfully, processing of the signal digitized in this manner has to be properly arranged and some specific conditions have to be taken into account.

10.1.2 Arrangements for Sample Value Processing

Basically there are two different sources of errors distorting the results of processing nonuniformly sampled signals. The first type of signal parameter estimation error is linked to incompletely suppressed aliasing. The second kind of error is related to sampling irregularities. It should be realized that the hybrid sampling technique provides for elimination of errors due to aliasing and that benefit is due to the fact that hybrid sampling is based on superposition of partial sparse periodic sampling pulse streams. This sampling approach is much better for elimination of aliases than the more conventional nonuniform sampling techniques. However, it does not automatically suppress the errors caused by random skipping of the signal sample value taking. These error components need to be diminished at the stage of signal processing in some other way. For instance, the signal processing methods adapted to sampling nonuniformities, discussed in Chapter 18, are effective for that.

As shown in Chapter 8, variable-phase periodic sampling, if performed correctly, provides the conditions necessary for suppression of aliasing. The considered method for hybrid sampling exploits the dependencies described there. To explain how aliases present in the digital signal obtained as a result of hybrid sampling can be eliminated, the signal model will be used again, according to which a signal $x(t)$ is composed of a multitude of sinusoidal components at arbitrary frequencies in a limited frequency band. Suppose that such a signal is sampled periodically and that it is done according to superposition of a number of phase-shifted periodic sampling point processes. Consider the case where the frequency range is extended four times. In that case, the highest frequency of the signal might exceed two times the frequency of the periodic sampling point processes used. The sampling point processes are periodic with the period $T = 1/f_s$ and the signal sample values are taken at sampling instants $t_k^{[m]} = t_k^{[0]} + m\delta$, $m = \overline{0, 3}$, where $m\delta = mT/4$ is the phase angle of the mth sampling point process. Superposition of the particular sampling point processes apparently has to be performed so that the minimal distance between adjacent sampling instants in the resulting sampling process should not be shorter than the smallest sampling interval allowed for that ADC. Under the specified conditions, each of the signal basic frequencies f_0 within the frequency range $(0 - \frac{1}{4} f_u)$ has three potentially aliasing frequencies: f_1, f_2, f_3. The task is to recover the signal components at all indicated frequencies. At least four periodic sampling point processes are used.

To compose equations describing the sampling process under these conditions, account is taken of the fact that all aliasing frequencies represented by sample

values taken at instants belonging to a particular periodic decimated sampling point process completely overlap. Therefore it can be written that the signal component at the frequency f_0, filtered out of the signal sampled by one of the four periodic decimated sampling point processes, represents a sine function. The parameters of this sine function differ for all involved phase-shifted decimated periodic sampling point processes. The filtered-out sine functions at frequency f_0 for each of the mentioned four sampling point sequences are denoted by $x(t_k^{[0]})$, $x(t_k^{[1]})$, $x(t_k^{[2]})$ and $x(t_k^{[3]})$ respectively. As overlapping of the aliasing frequencies is fully determined only by the phase angles of the particular randomly decimated periodic sampling point processes, the dependencies characterizing aliasing in the case of variable-phase periodic sampling discussed in Chapter 8 are also fully applicable in this case no matter what pattern of the randomly skipped sampling points there is. This approach leads to the following set of equations:

$$\begin{aligned}
x\left(t_k^{[0]}\right) &= (a_0 + a_1 + a_2 + a_3)\cos\left(2\pi f_0 t_k^{[0]}\right) + (b_0 - b_1 + b_2 - b_3)\sin\left(2\pi f_0 t_k^{[0]}\right),\\
x\left(t_k^{[1]}\right) &= (a_0 + b_1 + b_2 - a_3)\cos\left(2\pi f_0 t_k^{[1]}\right) + (b_0 + a_1 - a_2 + b_3)\sin\left(2\pi f_0 t_k^{[1]}\right),\\
x\left(t_k^{[2]}\right) &= (a_0 - a_1 - a_2 + a_3)\cos\left(2\pi f_0 t_k^{[2]}\right) + (b_0 + b_1 - b_2 - b_3)\sin\left(2\pi f_0 t_k^{[2]}\right),\\
x\left(t_k^{[3]}\right) &= (a_0 - b_1 - b_2 - a_3)\cos\left(2\pi f_0 t_k^{[3]}\right) + (b_0 - a_1 + a_2 + b_3)\sin\left(2\pi f_0 t_k^{[3]}\right).
\end{aligned} \tag{10.1}$$

This system describes how the power of the signal component at frequency f_0 is subdivided between all four aliasing frequencies f_0, f_1, f_2 and f_3. Therefore it can be used to estimate all Fourier coefficients at any arbitrary frequency f_0 for all related overlapping signal components. The fact that it can be done is significant.

Note that the signals on the left side of all equations of (10.1) are sine functions, characterized by the frequency f_0 and differing amplitudes and phase angles. These parameters, as (10.1) shows, depend on the parameters of the aliasing frequencies. If the sine functions on the left-hand side of these equations are represented by their quadratic components, the following system of equations is obtained:

$$\begin{aligned}
a_0^{[0]} &= a_0 + a_1 + a_2 + a_3, \qquad & b_0^{[0]} &= b_0 - b_1 + b_2 - b_3;\\
a_0^{[1]} &= a_0 + b_1 + b_2 - a_3, \qquad & b_0^{[1]} &= b_0 + a_1 - a_2 + b_3;\\
a_0^{[2]} &= a_0 - a_1 - a_2 + a_3, \qquad & b_0^{[2]} &= b_0 + b_1 - b_2 - b_3;\\
a_0^{[3]} &= a_0 - b_1 - b_2 - a_3, \qquad & b_0^{[3]} &= b_0 - a_1 + a_2 + b_3.
\end{aligned} \tag{10.2}$$

Next, this equation system can be rewritten as follows:

$$\begin{aligned}
a_0 &= \tfrac{1}{4}\big(a_0^{[0]} + a_0^{[1]} + a_0^{[2]} + a_0^{[3]}\big), \quad b_0 = \tfrac{1}{4}\big(b_0^{[0]} + b_0^{[1]} + b_0^{[2]} + b_0^{[3]}\big);\\
a_1 &= \tfrac{1}{4}\big(a_0^{[0]} + b_0^{[1]} - a_0^{[2]} - b_0^{[3]}\big), \quad b_1 = \tfrac{1}{4}\big(-b_0^{[0]} + a_0^{[1]} + b_0^{[2]} - a_0^{[3]}\big);\\
a_2 &= \tfrac{1}{4}\big(a_0^{[0]} - b_0^{[1]} - a_0^{[2]} + b_0^{[3]}\big), \quad b_2 = \tfrac{1}{4}\big(b_0^{[0]} + a_0^{[1]} - b_0^{[2]} - a_0^{[3]}\big);\\
a_3 &= \tfrac{1}{4}\big(a_0^{[0]} - a_0^{[1]} + a_0^{[2]} - a_0^{[3]}\big), \quad b_3 = \tfrac{1}{4}\big(-b_0^{[0]} + b_0^{[1]} - b_0^{[2]} + b_0^{[3]}\big).
\end{aligned} \tag{10.3}$$

It can be seen from the equations of (10.3) that parameters of all aliasing frequencies are explicitly expressed through the amplitudes of the quadratic components of the sine functions filtered out of the original signal at the frequency f_0 in the cases where that signal has been sampled by all four of the shifted decimated periodic sampling processes. The obtained mathematical description of this relationship is very useful. It simplifies calculations significantly and reveals the essential interdependencies well.

On the other hand, nothing is gained by obtaining estimates of $a_0^{[m]}$ and $b_0^{[m]}$ for all four ($m = 0, 1, 2, 3$) mentioned sampling subroutines. That needs to be done in another way. A direct DFT-based parameter estimation is the first possibility that comes to mind. However, this approach would provide only rough estimates corrupted by errors due to random skipping of the sampling instants in all four periodic components of the total sampling point process. To obtain improved results of such a signal analysis, more elaborate signal spectral parameter estimation procedures, better suited for processing nonuniformly sampled signals, need to be used to process the signal sample values obtained by applying the hybrid sampling techniques. Various ways of how to do this are considered in Part 2. As shown there, while the DFT is still typically involved, the raw estimates obtained as a result of these transforms could be improved on the basis of either iterative or adaptive procedures exploiting available *a priori* information. The specific pattern of the sampling instants excluded from all four ($m = 0, 1, 2, 3$) mentioned periodic sampling point processes represents the most valuable information of this kind.

It follows from the equations of (10.3) that the frequency range of the signal is subdivided into two parts and different techniques are used for estimating Fourier coefficients at frequencies within these two parts. While the coefficients $a_0^{[m]}$ and $b_0^{[m]}$in the baseband range are estimated by filtering signal components out of all frequencies within this range, calculations based on Equations (10.3) are carried

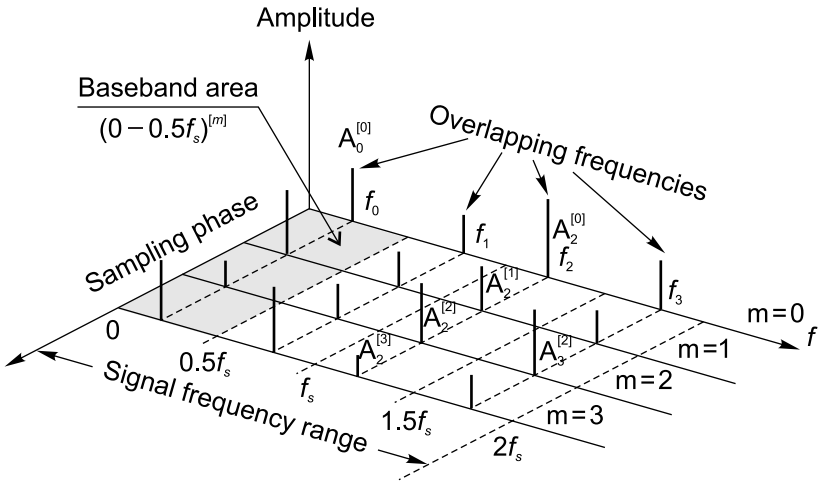


Figure 10.2 Illustration of the signal parameter estimation arrangement in the case where a signal is sampled according to the model of hybrid sampling

out to estimate these parameters at the higher frequencies outside the baseband, illustrated by Figure 10.2.

An estimation of Fourier coefficients $a_0^{[m]}$ and $b_0^{[m]}$, requiring execution of a large quantity of multiplication operations, represents the main computational burden. This concerns estimations only in the baseband frequency/sampling phase area (shaded in Figure 10.2) that are equal to $1/m$ of the whole signal frequency/sampling phase range. However, these coefficients apparently have to be estimated m times for all of the used m phase-shifted periodic sampling point processes with random skips. Calculating Fourier coefficient estimates within the other three frequency/sampling phase areas, $(0.5f_s - f_s)^{[m]}$, $(f_s - 1.5f_s)^{[m]}$ and $(1.5f_s - 2f_s)^{[m]}$, is a much simpler task, requiring only a few arithmetic operations carried out in accordance with Equations (10.3). To simplify the illustration, only the varying amplitudes of the signal components are shown in Figure 10.2. Of course, the phase angles of these components, estimated on the basis of processing signal sample values obtained at different phase shifts of the sampling processes, normally differ as well.

A block diagram of a device that performs hybrid sampling and preprocessing of the sampled signal in the described way is shown in Figure 10.3. A special sampling driver is used to form the sampling pulse sequence required to perform the hybrid sampling operation. While different approaches may be used for doing

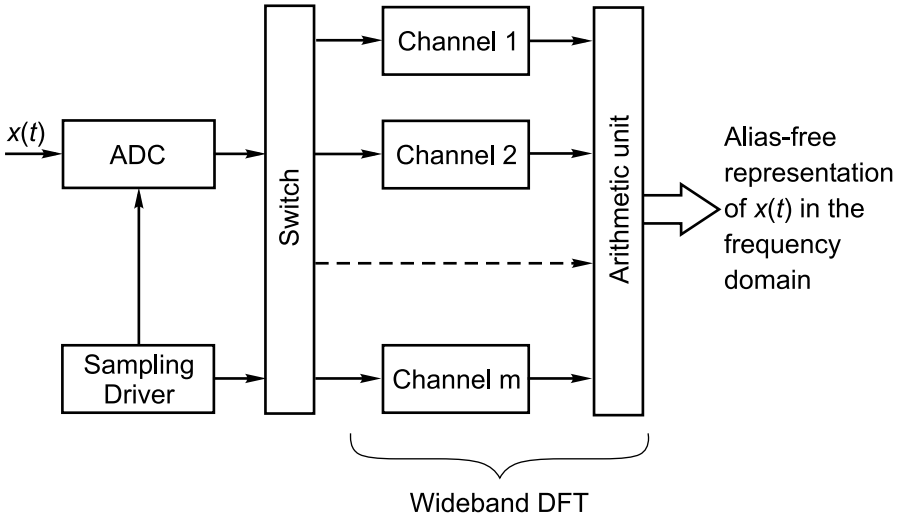


Figure 10.3 Block diagram of a device for hybrid sampling providing alias-free digital representation of the signal in the frequency domain

that, the pulses formed for sampling are taken from a number of variable-phase periodic pulse processes. The signal sample values obtained relate to specifically phased sampling and are transferred to the corresponding signal preprocessing channel for separate preprocessing. An electronic switch controlled by the sampling driver is used for that. At the stage of signal preprocessing an estimation of Fourier coefficients $a_0^{[m]}$ and $b_0^{[m]}$ for the baseband frequencies is performed in parallel. These estimates are then used to calculate the wideband Fourier coefficient estimates according to Equations (10.3), as described above. The obtained alias-free data then represent the signal in the whole frequency range, which is m times wider than the baseband. How they are used, of course, depends on the specific application. However, signal reconstruction and periodic resampling at m times higher frequency than the frequency of the periodic processes used initially would obviously represent the most flexible approach. It would make it possible to exploit the classical algorithms for signal processing.

10.2 Hybrid Double Sampling

Problems of digital alias-free signal processing could often be more effectively tackled at the stage of original analog signal conversion into the digital format rather than later at the stage of digitized signal processing. As signal characteristics

and conditions for their processing vary over a wide range, adapting signal digitizing to them represents a challenge. To face this, the used digitizing methods and means need to be sufficiently flexible. The hybrid sampling approach might be considered as a competitive option for developing specific sufficiently adaptable sampling methods and sampling driver designs. Hybrid double sampling techniques discussed in this section develop the basic hybrid sampling idea further. Specifically, organization of the sampling operation in a way ensuring that the signal sample value can be taken at sufficiently short sampling intervals is considered.

Hybrid double sampling is implementation of hybrid sampling based on the use of at least two ADCs connected in parallel. It is aimed at obtaining the right conditions for processing signals with both discrete and continuous spectra. To achieve this, the distances between adjacent sample values often have to be shortened and application of hybrid double sampling makes it possible to perform sampling with small discrete sampling interval increments. This kind of sampling is based on various combinations of periodic and nonuniform sampling elements.

Note that some of the hybrid sampling schemes can be realized either as a single ADC version or as a hybrid double sampling scheme built on the basis of two ADCs. For instance, this applies to the hybrid sampling model mixing pseudorandomly decimated phase-shifted periodic sampling point streams, considered above.

A few specific models of hybrid double sampling models are considered in this section. These models are considered to be generic and many variations of them are possible. Their characteristics differ. Some sampling models are better suited for the needs of special application conditions than others. Therefore it is possible to adapt, to some extent, the hybrid double sampling mode and parameters to specific applications. It also means that it is better if the microelectronic sampling drivers used to implement these sampling techniques are reprogrammable.

10.2.1 Providing for Short Sampling Intervals

There is a typical drawback of scarce nonuniform sampling that leads to some limitations imposed on essential applications of this kind of sampling. As the distribution of the sampling intervals for nonuniform sampling has to guarantee that the distance between two consequent sampling instants is equal or larger than the shortest sampling time interval permitted for the specific ADC to be used, it is impossible to obtain signal sample values located more closely on the time axis. This limitation of scarce nonuniform sampling essentially impacts a particular type of wideband signal processing. For instance, this applies to the spectrum

analysis of continuous spectrum signals based on an estimation of correlation functions. A solution of the short sampling interval problem is offered by hybrid double sampling.

An estimation of signal spectral density functions is often carried out on the basis of the DFT performed over the autocorrelation functions $P_x(\tau_i)$. These functions usually have to be estimated for the discrete time delay values τ_i varying in a given range including small values. This represents a problem in cases where the signals have been sampled nonuniformly. There is also an additional condition that has to be taken into account in these cases. This is related to the obvious necessity of correctly averaging the products $x(t_k)x(t_k + \tau_i)$, which are then calculated for a varying number N_i of discrete values of the time delay in accordance with the definition of the estimate of the autocorrelation function for a particular value τ_i:

$$\rho_{xx}(\tau_i) = \frac{1}{N_i} \sum_{i=1}^{N_i} x(t)x(t + \tau_i). \tag{10.4}$$

The values of N_i in general depend on the used sampling point process and vary for different delay time values. In the case of additive pseudo-random sampling, the histogram of N_i values for the varying delay times is shown in Figure 10.4 This histogram shows the size of the number of averaged products $x(t_k)x(t_k + \tau_i)$ for specific delay time values within the given delay time range. Therefore this kind of empirical N_i distribution could be used to calculate the estimates of autocorrelation functions in the cases of predetermined pseudo-random sampling.

It can be seen that the number N_i is equal to zero or is not defined for small values of τ_i. This is a problem, as that part of autocorrelation functions is especially essential for estimation of wideband signal spectra. Next, it is also desirable to have flat histograms of N_i as the autocorrelation function estimates would then have approximately the same statistical estimation error for all delay time values. The histogram of N_i values, given in Figure 10.4, is also not very good from this point of view.

When wideband signals having components at very high frequencies need to be analysed, the signal sample values have to be taken with very short intervals between successive sampling instants t_k and t_{k+1} to perform an estimation of autocorrelation functions for small delay time values. That requirement could be satisfied either by using a very high frequency ADC or by using double hybrid sampling carried out using two ADCs connected in parallel. The second approach

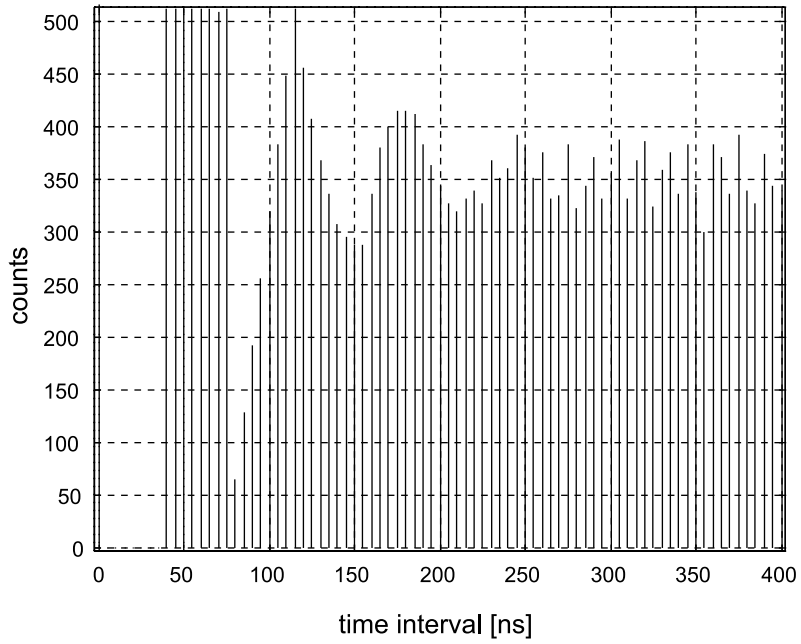


Figure 10.4 Histogram of N_i values for varying delay time values illustrating conditions for signal autocorrelation analysis in cases where the signal sample values are taken at the sampling instants according to the definition of additive pseudo-random sampling

is often preferable. Then the smallest delay step would no longer depend on the highest sampling rate of the involved ADCs in principle, very small delay step values could be obtained and that would lead to the possibility of analysing signals at very high frequencies. However, there are various hybrid double sampling schemes and some of them are better suited to applications involving the signal correlation analysis than others.

10.2.2 Double Periodic Sampling with Jitter

The model of double periodic sampling with jitter is based on the already described model of ordinary periodic sampling with jitter for which the sampling instants $\{t_k\}$ are defined as $t_k = kT + \tau_k, T > 0, k = 0, 1, 2, \ldots$, where $\{\tau_k\}$ is a multitude of independent identically distributed random variables with zero mean and T is the mean period of sampling. Usually random variables $\{\tau_k\}$

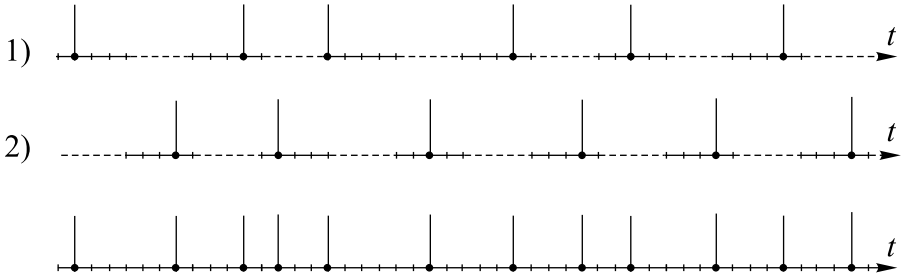


Figure 10.5 Example of double periodic sampling point stream with jitter

are distributed uniformly within a smaller or larger part of the period T. In the case of double periodical sampling with jitter, the sampling process is performed according to two sampling point processes acting in parallel. The sampling instants $\{t_k\}$ in the first and second sequences of the sampling instants $\{t_k^{[1]}\}$ and $\{t_k^{[2]}\}$ are defined as follows:

$$\begin{aligned} t_k^{[1]} &= 2kT + \tau_{2k}, \quad \tau_{2k} \in [0, (T-\delta)], \quad k = 0, 1, 2, \ldots, \\ t_k^{[2]} &= (2k+1)T + \tau_{2k+1}, \quad \tau_{2k+1} \in [0, (T-\delta)], \quad k = 0, 1, 2, \ldots. \end{aligned} \tag{10.5}$$

Hybrid double periodical sampling point processes with jitter is a superimposition of sequences $\{t_k^{[1]}\}$ and $\{t_k^{[2]}\}$:

$$t_k = kT + \tau_k, \quad \tau_k \in [0, (T-\delta)], \quad k = 0, 1, 2, \ldots. \tag{10.6}$$

This means that theoretically this double sampling model is equivalent to the case of ordinary periodic sampling with jitter where the random variable τ_k is uniformly distributed in the interval $[0, (T-\delta)]$ so that the shortest distance between two sampling instants is equal to δ, illustrated in Figure 10.5.

The most interesting characteristic of this model of double periodic sampling with jitter is the histogram showing how the distances between the sampling instants are distributed. A typical histogram of this kind is shown in Figure 10.6. It can be seen that the distribution of distances between the sampling instants, in this case while relatively flat in a larger scale, varies in the area of short distances and there are relatively few intervals between the sampling instants, which are equal to δ, 2δ, 3δ, ... This illustration has been obtained in the case where $\delta = 0.25T = 0.5$ ns.

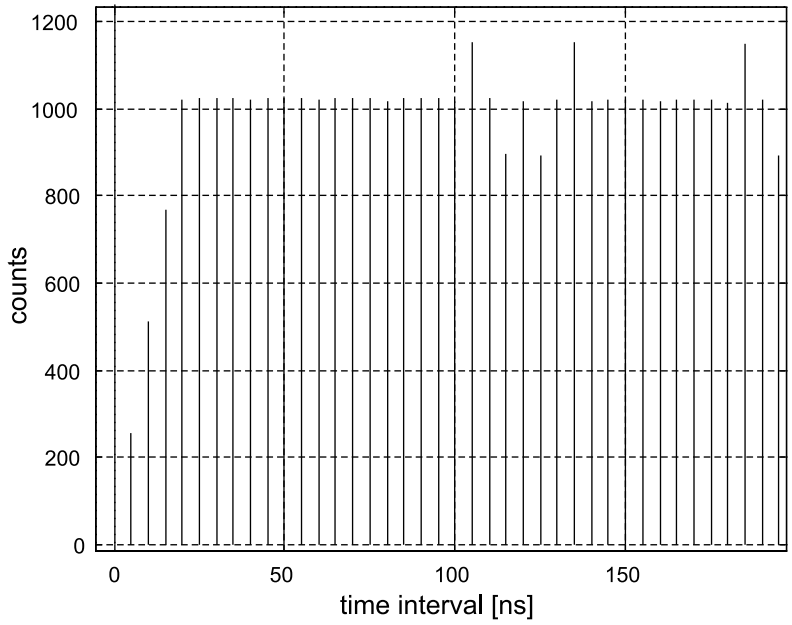


Figure 10.6 Histogram of the sampling intervals in the case of double periodic sampling with jitter

10.2.3 Double Additive Pseudo-random Sampling

The double additive pseudo-random sampling point stream $\{t_k\}$ is a superimposition of sequences $\{t_k^{[1]}\}$ and $\{t_k^{[2]}\}$. They are given as follows:

$$\begin{aligned} t_n^{[1]} &= t_{n-1}^{[1]} + \tau_n, \quad \tau_n \in [0, \mu), \quad n = 0, 1, 2, \ldots, \\ t_k^{[2]} &= t_{k-1}^{[2]} + \tau_k, \quad \tau_k \in [0, \mu), \quad k = 0, 1, 2, \ldots. \end{aligned} \tag{10.7}$$

Only one sample value is taken at the overlapping instants $t_k = t_n^{[1]} \equiv t_k^{[2]}$.

A particular realization of such a double additive pseudo-random sampling point process is shown in Figure 10.7. The histogram of the distances between the sampling instants characterizing this particular double sampling approach is shown in Figure 10.8. According to it, the distribution of small sampling intervals, while not perfect, seems to be acceptable for many cases of practical applications. This illustration again has been calculated for the case where $\delta = 0.25T = 0.5$ ns.

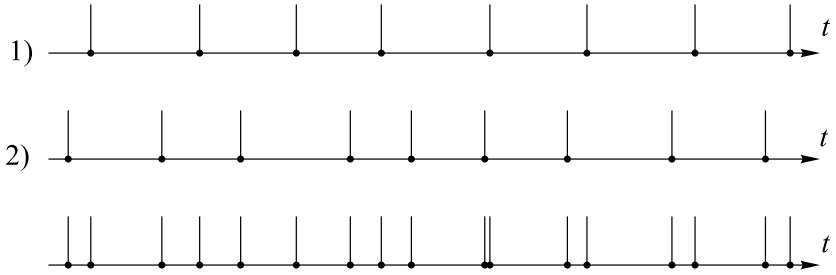


Figure 10.7 Illustration of a double additive pseudo-random sampling point process

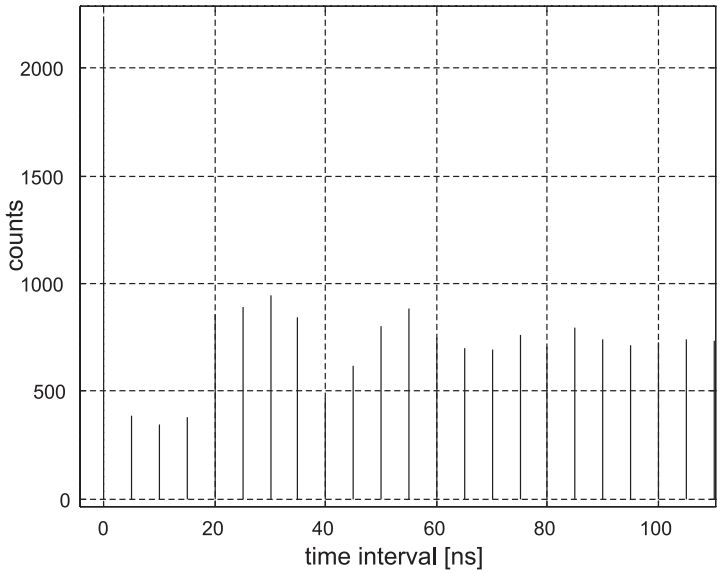


Figure 10.8 Histogram of the sampling intervals in a double additive pseudo-random sampling case

10.2.4 Periodic/Additive Pseudo-random Sampling

In the case of periodic/additive pseudo-random sampling the hybrid process is a superimposition of sequences $\{t_k^{[1]}\}$ and $\{t_k^{[2]}\}$, defined as follows:

$$\begin{aligned} t_k^{[1]} &= kT_\mathrm{s}, \quad k = 0, 1, 2, \ldots, \\ t_k^{[2]} &= t_{k-1}^{[2]} + \tau_k, \quad \tau_k \in [0, \mu), \quad k = 0, 1, 2, \ldots. \end{aligned} \tag{10.8}$$

Of course, there are also overlapping instants $t_k = t_n^{[1]} \equiv t_k^{[2]}$.

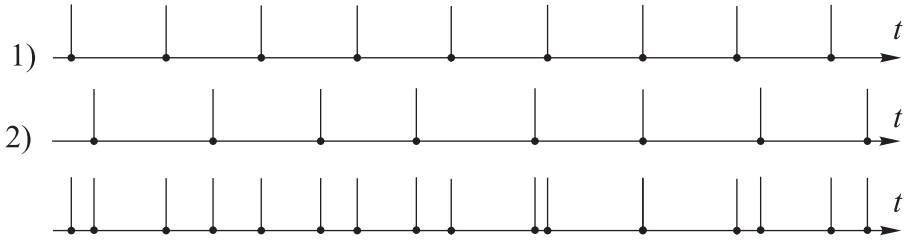


Figure 10.9 Example of the hybrid periodic/additive pseudo-random sampling point stream

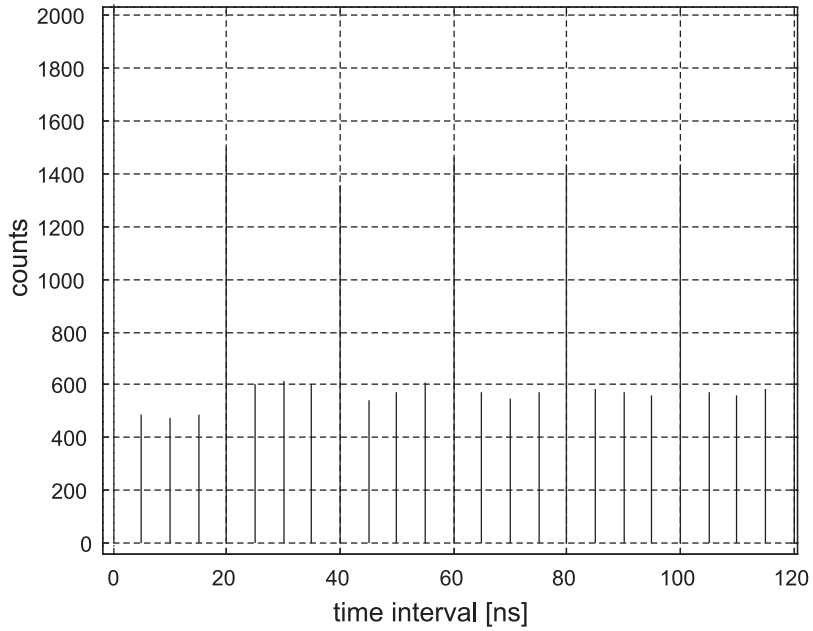


Figure 10.10 Histogram of the sampling interval distribution characterizing superimposed periodic and additive sampling point streams

An illustration of this particular type of hybrid double sampling scheme is given in Figure 10.9. The histogram showing the sampling interval distribution in this case is given in Figure 10.10. It can be seen that this distribution of the sampling intervals is closer to the desirable uniform distribution. It has been calculated using the same conditions as those for the histogram displayed in Figure 10.8.

This histogram clearly shows that this kind of sampling provides acceptable sampling conditions essential for an autocorrelation function estimation in the

range of small delay times. However, this particular composition of periodic and additive pseudo-randomized sampling processes has a much wider application range than just the correlation analysis. An interesting and useful modification of it is discussed in the following section.

10.3 Mixing Hybrid Sampling with Periodic Sampling

While the development of models for hybrid sampling so far has focused basically on achieving a uniform distribution of the sampling intervals, especially in the range of short intervals, it is not actually the exclusive motivation for arranging the sampling process as a hybrid double one. Hybrid double sampling of signals might also play an essential role for implementation of special procedures needed for an effective reduction of errors caused by aliasing. Consider a model of hybrid double sampling targeting the achievement of this benefit.

The hybrid sampling process already described in Section 10.1 may be used as a platform for a hybrid double sampling scheme of this kind. As explained above, this sampling scheme uses n periodic sampling point processes with pseudo-randomly skipped sampling points. Each of them is shifted in phase for $\Delta\varphi = 2m\pi/n, m = 0, 1, 2, \ldots, (n-1)$, or in time for $\Delta t = m\delta$. The important point is that the signal sample values obtained at the instants belonging to a specific periodic pseudo-randomly decimated sampling point process are marked and preprocessed separately. To arrange double sampling on this basis, the rules for forming the scheme for hybrid sampling as the basis are slightly changed and an additional periodic sampling process is added. The sampling points falling into the spaces between the sampling instants $\{kT\}$ are related to the periodic pseudo-randomly decimated processes shifted in phase for a phase angle that differs from the phase angle of the periodic process $\{kT\}$. That represents an additional restriction not imposed on the hybrid sampling process when it is used separately.

The sampling point stream obtained in this way is shown in Figure 10.11. Note that the frequency $1/T$ of the added periodic sampling point process $\{kT\}$ is equal to the frequency of the pseudo-randomly decimated periodic components of the used hybrid sampling process. Moreover, this periodic sampling point process is used as the basis of the whole hybrid double sampling scheme and the sampling points from all other pseudo-randomly decimated periodic components interleave with this process $\{kT\}$.

A block diagram of an electronic implementation of this hybrid double sampling mode is given in Figure 10.12. It is based on two ADCs connected in parallel. Periodic and hybrid sampling drivers are used to execute the sampling operation

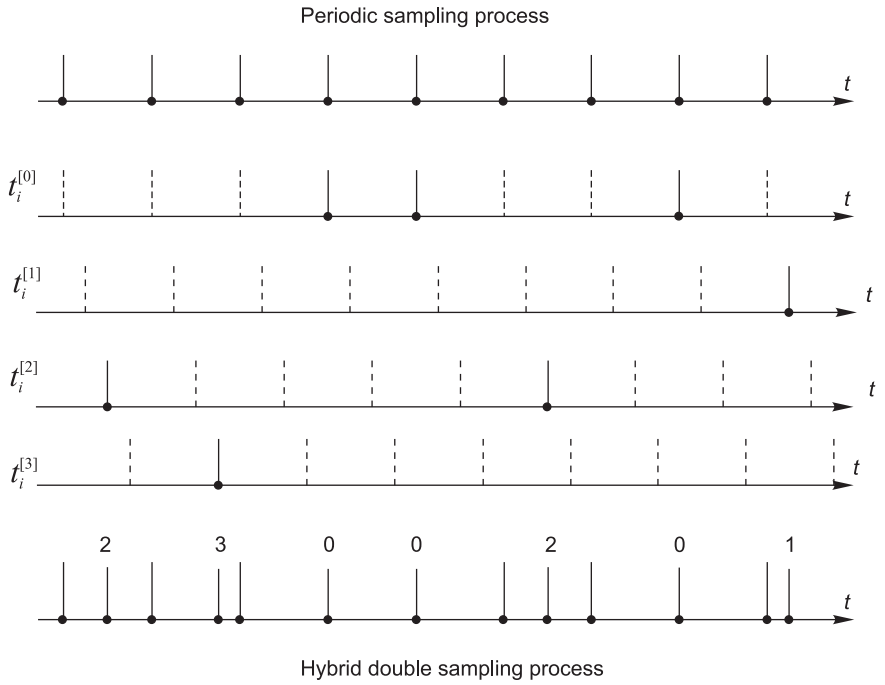


Figure 10.11 Time diagrams of the particular and resulting hybrid double sampling point processes

in accordance with the described sampling approach. The output signals of both ADCs, marked so that they are tied to specific periodic components of the hybrid double sampling process, are put together and transferred for processing.

The algorithms used to process this type of digital signal evidently have to be matched to the signal hybrid double digitizing specifics. In general, the arrangements for processing signal sample values, taken according to the rules of hybrid double sampling, should be the same as those already discussed in Section 10.1 with regard to the case of hybrid sampling. However, using two ADCs in this way leads to better conditions for processing the obtained digital signal. In this case, no sample values are missing from the sequence $x(t_k^{[0]})$ for which all the signal sample values are taken periodically at instants $\{kT\}$. Therefore, considerably more accurate estimates of $a_0^{[m]}$ and $b_0^{[m]}$ are obtained for $m = 0$ as they are not corrupted by errors due to random skipping of the sample values. The specific consequences of that depends, of course, on the particular algorithm used for signal processing. However, more often than not the fact that the digital signal

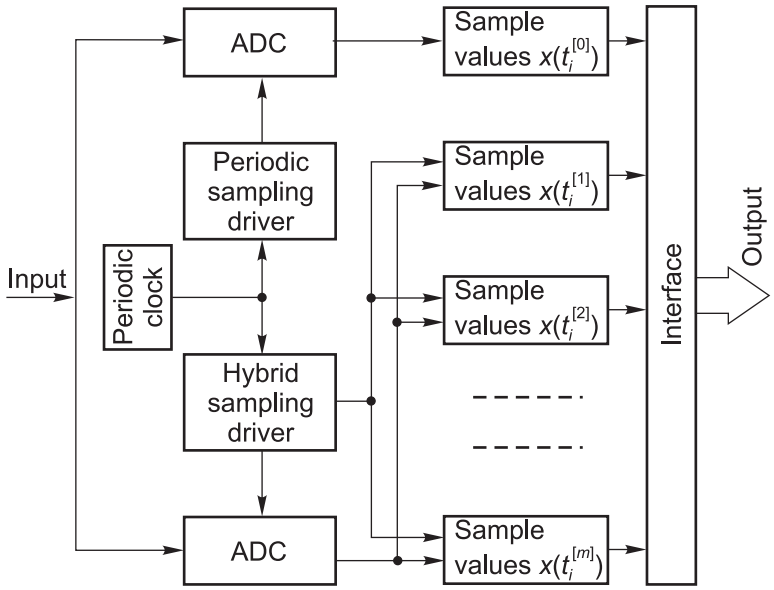


Figure 10.12 Block diagram of a device for signal digitizing according to the hybrid double sampling model

then contains periodically taken sample values is very useful. For instance, this makes it possible to use fast algorithms for a precise estimation of the frequencies of the peaks in signal spectra, which is crucial for obtaining good results in many cases of digital alias-free signal processing. This is demonstrated in Part 2 of this book by examples of iterative direct and inverse Fourier transforms and adapting signal processing to sampling nonuniformities discussed in Chapters 20 and 18 respectively.

Although sampling of signals according to the model for hybrid double sampling usually targets a more precise execution of various algorithms for digital alias-free signal processing, this kind of sampling is also well suited for applications related to signal correlation analysis. The sampling interval distribution in this case is close to the distribution displayed by the histogram given in Figure 10.10. That is understandable as for this particular application area the hybrid double sampling scheme is close (but not identical) to the model based on the superposition of periodic and additive sampling point streams.

10.4 Comments in Conclusion

The basic advantage of this hybrid sampling approach over pseudo-randomized sampling is in the improved suppression of aliasing. Aliases are actually

eliminated in this way rather than being suppressed. This advantage is largely due to the decomposition of the original additive pseudo-randomized sampling process into a number of variable-phase periodic sampling point streams with pseudo-random skips of sampling instants. It is a fruitful approach leading to several desirable consequences:

1. This approach leads to the avoidance of fuzzy aliasing. Aliasing, of course, is still taking place, but now it is well-defined and on a full scale with parameters depending on the phases of the periodic sampling instant streams but not on the pattern of randomly skipped sample values. That makes it possible to reveal the aliases and to take them out.
2. Processing of the digital signals, obtained as a result of the described hybrid double sampling, can often be realized on the basis of well-developed fast algorithms such as the fast Fourier transform (FFT). It is also advantageous that the DFT-based estimation of Fourier coefficients in this case can be carried out in parallel simultaneously in m channels in a frequency range reduced m times.
3. There are also some fringe benefits. Specifically, this kind of hybrid sampling could be implemented in a way characterized by significantly reduced jitter of the sampling instants. Suppressing sampling jitter is crucial as it causes serious problems. In general, jittering corrupts processing of wideband nonuniformly sampled signals because most of the algorithms are sensitive to fluctuations of the sampling time instants with regard to their predetermined values. Before the introduction of hybrid sampling, the development of special microelectronic sampling drivers was seen to be the only possible way of decreasing this kind of jitter. Although this approach still remains as an option, the hybrid sampling approach offers a better solution. The problem with implementation of the additive sampling techniques is that usually a relatively large number of code-controlled discrete delay elements are involved in the generation of additive pseudo-random sampling point processes. The difference between the predetermined and actual sampling instants is often due to deviations of the delay times. For that reason, the development and use of special high-performance microelectronic sampling drivers does not lead to effective elimination of that sort of jitter. The performance of hybrid sampling, on the other hand, does not require the use of many time delay elements. The pulse streams used for hybrid sampling might be formed on the basis of a few periodic phase-shifted pulse processes stable in time. Taking out some pulses from them does not lead to the introduction of jitter either. Therefore using hybrid sampling techniques is also seen as an option to the jitter suppression problem.

Bibliography

Artjuhs, J. and Bilinskis, I. (2006) Method and apparatus for alias suppressed digitizing of high frequency analog signals. EP 1 330 036 B1, European Patent Specification, Bulletin 2006/26, 28.06.2006.

Bilinskis, I. and Artjuhs, J. (2006) Method and apparatus for alias suppressed digitizing of high frequency analog signals. United States Patent US 7,046,183 B2, 16 May 2006.

Lo, K.C. and Purvis, A. (1993) On interlacing and concatenating additive random sampling sequences. In Proceedings of the International Conference on *Signal Processing*, Beijing, China, October 1993, pp. 48–51.

Lo, K.C. and Purvis, A. (1994) Hybrid additive random sampling and its realization. In Proceedings of the IEEE International Symposium on *Circuits and Systems*, London, May–June 1994, Vol. 2, pp.105–8.

Artyukh, J., Bilinskis, I., Boole, E., Rybakov, A. and Vedin, V. (2005) Wideband RF signal digititising for high purity spectral analysis. In Proceedings of the International Workshop on *Spectral Methods and Multirate Signal Processing (SMMSP 2005)*, Riga, Latvia, 20–22 June 2005, pp. 123–8.

Part 2

Processing

11

Data Acquisition

Digital data reflecting information carried by an analog signal can be represented either by the signal sample values given at the sampling instants or by the results of some preliminary signal processing procedure. Signal sample value preprocessing, carried out as data acquisition subroutines, is considered as signal parameter estimations in Chapter 12 and as complexity-reduced calculations of DFT estimates in Chapter 16.

The first category of the data acquisition systems considered here could often be realized on the basis of a single ADC with a few devices added for signal conditioning, data buffering and interfacing. However, this basic data acquisition scheme has to be diversified as soon as the conditions for data acquisition become more complicated and more demanding. Then some nontraditional approach, such as randomized sampling, might prove to be an interesting option. First of all, it concerns the methodology according to which signals are sampled and quantized. As shown in Part 1 of this book, these operations could be carried out in various ways so that they could be matched to specific functional and performance requirements. There are various signal digitizing techniques available and analysis of them is helpful in finding an appropriate approach to use in particular special signal digitizing and data acquisition problems. The structure and performance of data acquisition systems can be modified on the basis of the selected digitizing techniques with the application conditions taken into account. The resulting deviations from the classical data acquisition scheme might be substantial. The modifications of data acquisition systems considered in this chapter are basically related to the following two topics: (a) data acquisition providing alias-free processing of wideband signals and (b) data acquisition from a large quantity of signal sources.

Digital Alias-free Signal Processing I. Bilinskis

11.1 Data Acquisition from Wideband Signal Sources

The central issue for data acquisition from wideband signal sources is elimination of limitations imposed on digitization of these signals by the aliasing effect. It makes sense to use nonuniform sampling techniques to obtain the capability of digitizing signals at frequencies much higher than the limit marked by half of the sampling rate. As shown in Part 1 of the book, the task of alias-free signal sampling can theoretically be resolved in a number of ways and signals can be digitized without corruptions imposed by aliasing in bandwidths that are several times wider.

11.1.1 Practical Results Confirming the Theory

These considerations are not purely theoretical. In fact, only physically realizable and practically applicable methods, algorithms and techniques are considered in this book. Engineering experience accumulated over many years in this area confirms the feasibility of them. The example of a wideband digitizer, shown in Figure 11.1, illustrates this point.

In Section 2.5, it is shown that using the digital alias-free signal processing technology leads to significant widening of the operational frequency range, even for existing ADCs. The digitizer shown in Figure 11.1 is designed on the basis of a 12-bit ADC characterized by the maximal sampling rate and the input bandwidth equal to 125 MS/s and 700 MHz respectively. This means that this digitizer, operating on the basis of periodic sampling performed at the highest achievable for this ADC rate, is applicable for digitizing signals in the frequency range up to 62.5 MHz. Application of randomized sampling opens up the possibility of widening this frequency range by about 10 times. This digitizer performs alias-free signal digitization in a wide dynamic frequency range (0.1–669 MHz) and the achieved spurious free spectral purity is near to that achievable by the high-performance 12-bit ADC (in this case, AD9433 from Analog Devices). Using this digitizer and special DASP software makes it possible to analyse signals completely digitally within this wide frequency band.

A spectrogram and the respective reconstructed waveform obtained by processing an amplitude-modulated signal are given in Figure 11.2(a). A similar spectrogram and waveform are shown in Figure 11.2(b). They have been obtained using the same hardware and software tools in the case where a weak signal is estimated in the presence of a powerful one. Note that in the case of DSP, the minimal sampling rate required for performing analysis of these signals within the indicated frequency range would be 1400 MS/s. That sampling rate

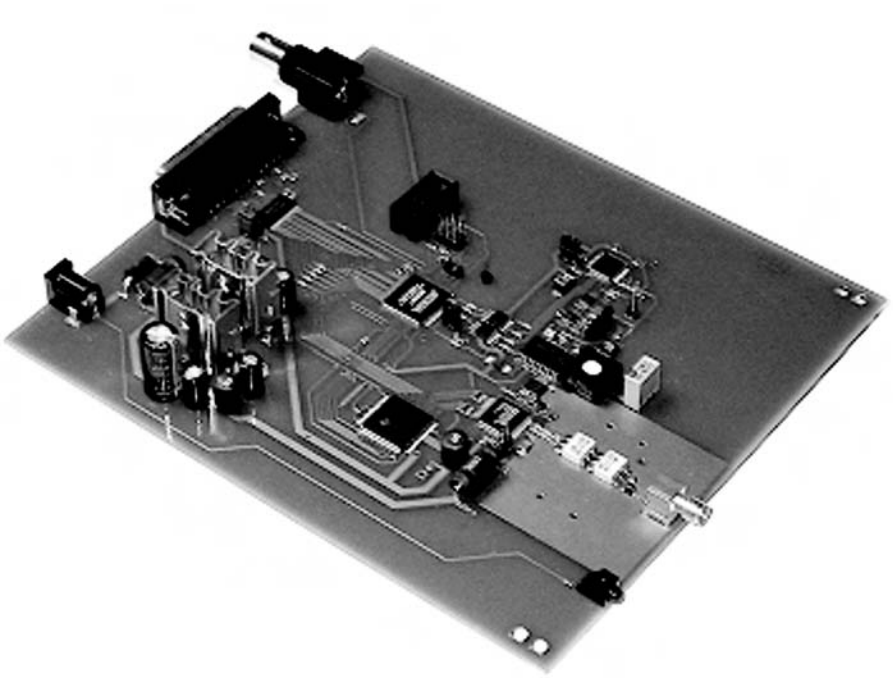


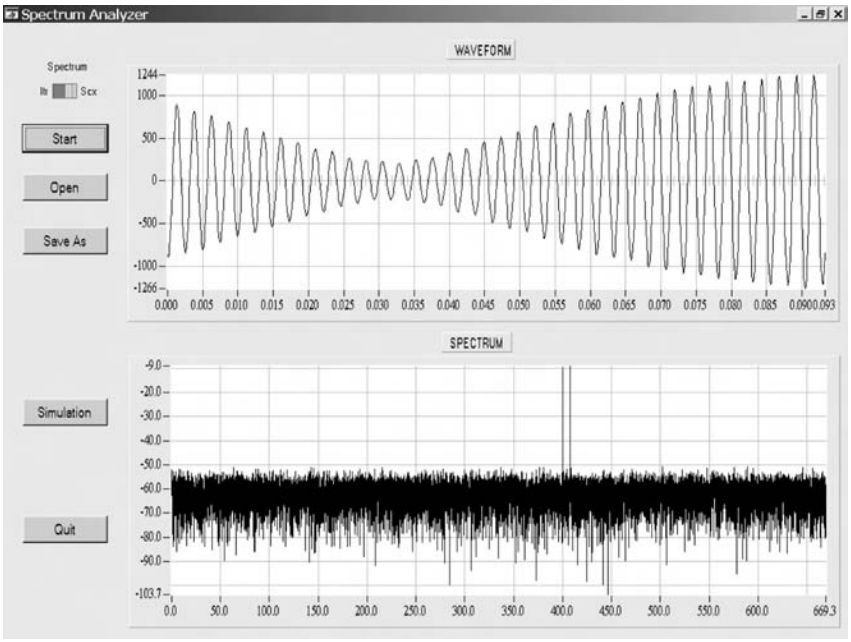
Figure 11.1 Digitizer built on the basis of a 12-bit ADC with the 125 MS/s maximum sampling rate applicable for a DASP based wideband signal analysis in the frequency range 0.1–700 MHz

would be about 25 times higher than the mean sampling rate used to obtain the given spectrograms.

11.1.2 Sampling with Reduced Uncontrolled Jitter

While the example above illustrates the point that appropriate nonuniform sampling makes it possible to widen the frequency range for input signal alias-free digitizing, the mentioned frequency of 669 MHz of the input signal spectrum does not represent the achievable upper limit. The considered digitization techniques are applicable also for handling signals digitally at frequencies in the GHz range. However, with the upper frequency going up, the perfection level of the engineering implementation of these alias-free digitization techniques has to be increased as well. Various engineering aspects are concerned, including designs of the involved PCBs (printed circuit boards), ASIC and FPGA chips.

a)



b)

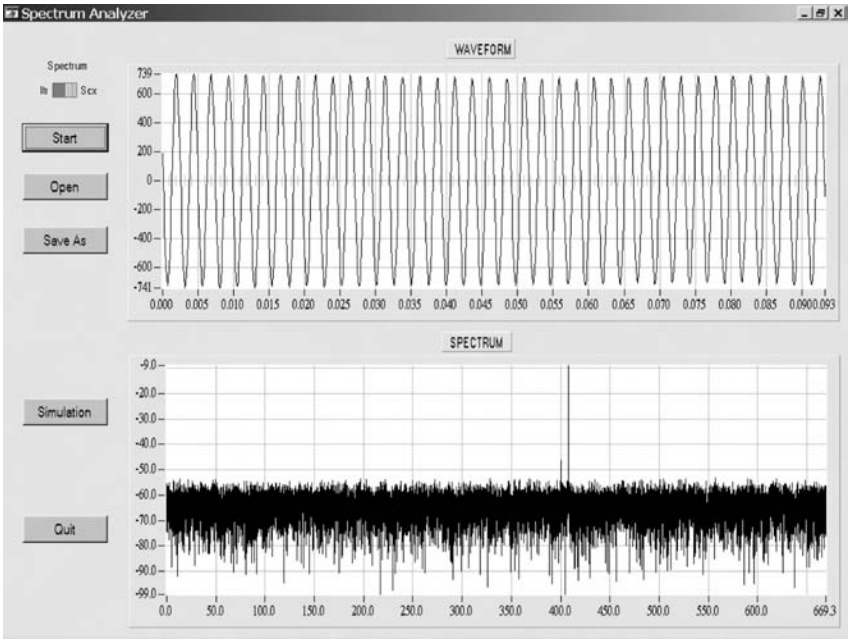


Figure 11.2 Alias-free spectrograms and reconstructed signal waveforms: (a) for an amplitude-modulated signal; (b) for a weak signal estimated in the presence of a strong signal component

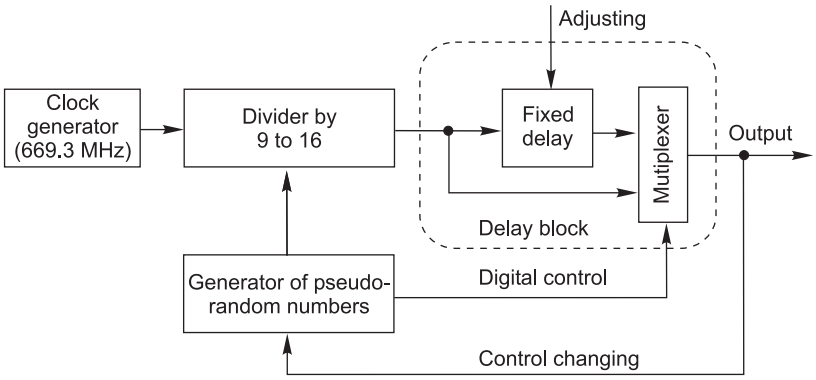


Figure 11.3 Block diagram of a sampling pulse former characterized by low-level sampling instant jitter

A particular problem that has to be dealt with is ensuring that the uncontrolled jitter present at signal sampling is kept at an acceptable sufficiently low level. In many cases this means that the standard deviation of the jitter should not exceed a few picoseconds or it should be even smaller. It has been found that to provide for at least 50 dB of spurious-free dynamic range (SFDR) in the bandwidth up to 700 MHz the discrepancy between the expected and real sampling instants should not exceed 2 ps.

To design and manufacture digitizers operating at this performance level is a challenging engineering task and one of the most responsible electronic devices in the digitizer structure is the sampling pulse former. A block diagram of it is given in Figure 11.3. According to the patented operational principle, the pulses dictating the sampling instants are formed by controlled division of the clock frequency with subsequent pseudo-random delay of the pulses obtained in this way. The generator of pseudo-random numbers supervises both of these frequency division and delay controlling functions.

It can be seen from the scheme given in Figure 11.3 that the frequency of the clock pulse repeating (669.3266 MHz) is divided in this particular case by a random integer ranging from 9 to 16. Then the output pulses are expanded in width and passed to the one-bit controllable delay block. This is designed on the basis of an adjustable delay line and a high-speed multiplexer. Operating in this way, the sampling pulse former generates a sequence of sampling pulses with the intervals between adjacent pulses pseudo-randomly varied from 13.447 to 24.652 ns, with the smallest step-size equal to 747 ps. This design ensures that the deviation of sampling time instants from the points on a strictly uniform

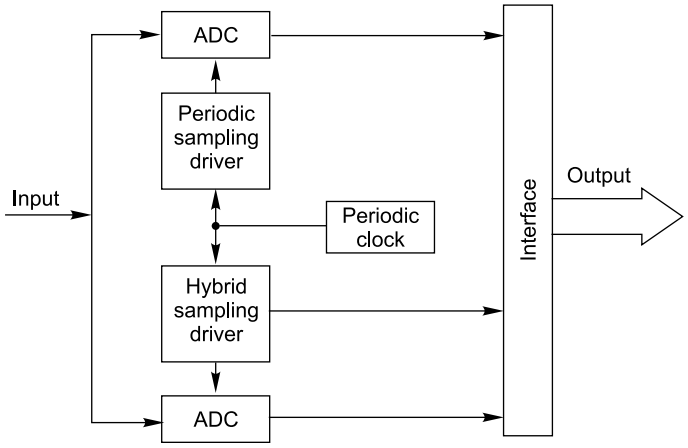


Figure 11.4 Block diagram of a device for data acquisition according to the hybrid double sampling model

timing grid do not exceed a few picoseconds. Although in this case the mean sampling rate is only 53.546 MS/s, the equivalent sampling rate is much higher, providing for alias-free signal bandwidth up to 669.3 MHz.

11.2 Application of Hybrid Double Sampling

A block diagram of an electronic implementation of a data acquisition system based on hybrid double sampling discussed in Chapter 10 is given in Figure 11.4. It contains two ADCs connected in parallel, with good reason. Periodic and hybrid sampling drivers are used to execute the sampling operation in accordance with the described sampling approach. The output signals of both ADCs, marked so that they are tied to specific periodic components of the hybrid double sampling process, are put together and transferred for processing.

The algorithms used to processing this type of digital signals evidently have to be matched to the signal hybrid double digitizing specifics. This issue is covered in some detail in Chapter 18. In general, the arrangements for processing signal sample values, taken according to the rules of hybrid double sampling, should be the same as those already discussed in Section 10.1 in regard to the case of hybrid sampling. However, using of two ADCs in the described way leads to specific conditions for processing the obtained digital signal. It is essential that no sample values are missing from the sequence $x(t_k^{[0]})$ for which they are taken periodically. In result much more accurate estimates of $a_0^{[0]}$ and $b_0^{[0]}$are obtained.

Although they are corrupted by the aliases these aliases are well defined so that the sample value sequences pseudo-randomly taken by the second ADC could be used for taking these aliases out. The sampling interval distribution in this case is close to the distribution displayed by the histogram given in Figure 10.10.

Although data acquisition performed according to the discussed model for hybrid double sampling usually targets more precise execution of various algorithms for digital alias-free signal processing, this kind of sampling is also well suited for real-time applications based on signal parallel processing as the data at the output of the second ADC are decomposed according to the scheme shown in Figure 10.12. The fact that the digital signal at the output of this data acquisition scheme contains periodically taken sample values is very useful also for other reasons. Firstly, this makes it possible to use the fast algorithms for precise estimation of the frequencies of the peaks in signal spectra. Secondly, the data stream in this case is represented by a periodic sample value sequence with ramdom skips and, consequently, there is no secondary aliasing. Thirdly, the intervals between two successive sample values could be very short and that is crucial for obtaining good results at processing wideband signals with continuous spectra including signal processing for the correlation analysis to be performed at very small delay time increments.'

11.3 Pseudo-randomized Multiplexing

A popular approach to data acquisition from multiple sources of information is based on sequentially connecting these signal sources to the input of a single ADC. Technical realization of such a data acquisition scheme is typically carried out on the basis of a multiplexer performing the required switching function in a regularly rotating manner. The multiplexer then connects the input signals to the input of the ADC for brief time intervals Δt in a fixed order. This means that the signal sample values are taken in this case from each of the input signals periodically, as shown in Figure 11.5 (left side). The problem is that period T of sampling under these conditions depends not only on the highest sampling rate of the given ADC but also on the number of inputs that have to be sequentially connected to it. That narrows the bandwidth of the input signals within which the signals can be processed without errors due to aliasing. In general, periodic multiplexing imposes additional limitations on the input signal spectra and on the number of inputs so that these parameters have to be traded off.

Application of deliberate randomization of this type of multiplexing can be considered as a possible approach to the resolution of this problem. When multiplexing is arranged on this basis, the order in which the input signals are connected

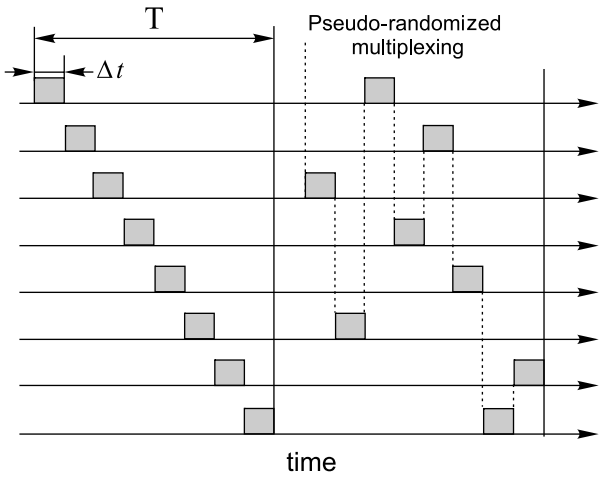


Figure 11.5 Switching patterns for periodic and pseudo-randomized multiplexing of input signals

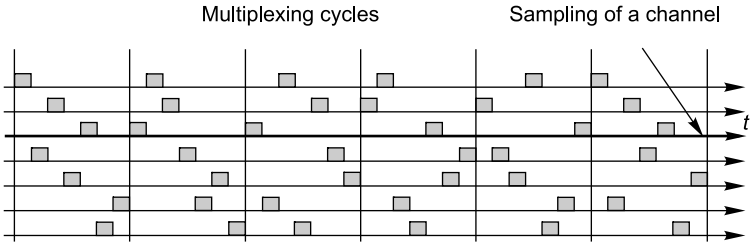


Figure 11.6 Nonuniform sampling of input signals taking place when their multiplexing is deliberately randomized

to the input of the ADC is changed from regular to pseudo-random, as shown in Figure 11.5. The signals in all input channels are then sampled pseudo-randomly in a specific pre-planned way controlled by a pseudo-random number generator. Figure 11.6 illustrates this sampling mode of the input signal multitude.

Evidently, this type of pseudo-random multiplexing results in nonuniform sampling of the information carriers. However, in this case the mean sampling rate could be below the carrier frequency. Therefore the duration of the multiplexing cycle could be prolonged. Signals in all channels could be reconstructed and tied to a common time reference.

While the errors of signal processing due to aliasing might be taken out and the limitations put on the number of input channels made less restrictive in this

way, the consequence of obtaining these benefits is increased complexity of the algorithms used to process the sampled signals. Whether it is acceptable or not needs to be decided in every given case.

The mentioned drawback of the classic multi-input ADC scheme based on the customarily used multiplexing of analog input signals, related to worsening the conditions for aliasing, is not the most damaging one. Actually this approach to digitizing a multitude of analog input signals based on collecting and passing them over shorter or longer wires with subsequent switching one by one to the input of the ADC also has other serious disadvantages. Transients accompanying the analog signal switching process, which have to be kept within sufficiently narrow margins, and the analog signal line sensitivity to surrounding noise might be mentioned in addition. Consequently, the application range for data acquisition systems based on analog signal multiplexing is limited. They cannot be used for gathering data from a large number of input signal sources. Another data acquisition concept is suggested for that in the next section.

11.4 Massive Data Acquisition

Conditions for data acquisition from multiple sources of analog signals, performed for the purpose of monitoring, analysing and/or controlling various systems and processes, vary over wide margins. Consequently, the concepts affecting how data acquisition systems are built differ as well. While relatively simple data acquisition systems quite often satisfy the requirements of specific applications, the need for very complicated wireless sensor networks arises more and more often and much attention has recently been made to develop them.

Referring back to the remote sampling principle discussed in Chapter 7, consider this approach to signal sampling in the context of data acquisition from multiple sources of analog signals. In order to draw attention to the advantages and limitations of the suggested remote sampling techniques based on sine-wave crossings, compare the capabilities of this sampling technique with the two most often met approaches to the technical realization of multichannel data acquisitions.

11.4.1 Specifics of Multichannel Data Acquisition

The approaches to data acquisition from many sources of information carried by analog signals mostly differ in organization of analog-to-digital conversions of the original analog signals. In the case of the first discussed scheme, there is a central ADC in the master part of the system. It converts the output signal of an analog multiplexer with its inputs connected by wire links to the multiple

analog signal sources. Although implementations of this first scheme are typically relatively simple, the used single central ADC is placed at a distance from the signal sources and analog signals have to be transmitted over the links between the front-end part and the master parts. As already pointed out, analog signal multiplexing imposes additional serious limitations as well.

In the second case, analog-to-digital conversions are moved as close to the sources of the original signals as possible and digital signals carry the data from multiple ADCs to the common master part of the data acquisition system. In this case each front-end device picking up an input signal contains an ADC. This leads to technically more complicated realizations than those typical for the first scheme. However, the data acquisition systems of the second kind have an essential advantage: the links between the front-end (sensor) part and the master part are digital and for this reason they are well protected against external noise, making it possible to use relatively long-distance links, including radio links.

The complexity of the data acquisition systems of this kind is apparently related to the bit rate of the used ADCs. In order to simplify the acquired data transmission from the sensor, a special type of ADC forming output signals as one-bit streams, known as $\Sigma\Delta$ modulators, is usually considered to be the best choice. They compare the input signal sample values with the reference signal and, depending on the result of this comparison, subtract or add an increment to that reference and form the corresponding binary output signal instantaneous value. Thus the reference signal is actually the accumulated sum of the previous discrete bipolar increments and it tracks the input signal waveform. To achieve good enough accuracy, tracking of the input signal has to be sufficiently close. That leads to the necessity of performing signal oversampling. Therefore the rate of bits transmitted between the front-end (sensor) part and the master part significantly exceeds the upper frequency in the spectrum of the input signals. Consequently, the whole data acquisition system, including the multichannel data reconstruction part, is not simple at all. That is true even for each one of the involved multiple sensors containing the $\Sigma\Delta$ modulators.

Thus existing engineering practice in this field shows that the first basic problem with the data acquisition systems is how to reduce their relatively high complexity. The point is that they often are too complicated. This complexity is bad for cost-effective massive data collection from a really large number of widely scattered sensors and, what is especially annoying, this makes it difficult to achieve sufficiently low power consumption.

The second problem is the relatively low number of data acquisition channels, the number of signal sources from which data might be collected by a single master part achievable by reasonably complicated and costly systems. The third

problem is limitations in operational speed related to application limitations in the frequency domain.

11.4.2 Reconfigurable Distributed Structure ADC

In general, the data acquisition techniques based on the remote sampling concept belong to the category of the second mentioned data acquisition scheme as the input signal digitization operations are moved close to the signal sources. However, there the similarity ends. The input signal sampling operations are performed in a quite different way. The sine-wave crossing remote sampling scheme, discussed in Chapter 7, and the data acquisition systems based on it do not have the drawbacks of the popular data gathering techniques. Exploitation of this signal digitizing concept, relying on the remote sampling approach, leads to unusual developments in the area of data acquisition systems. Once the idea is accepted that sampling operations might be distanced from quantizing, the structure of the data acquisition system could actually be converted into a structure of a distributed ADC having multiple remotely acting samplers placed close to the signal sources. The big advantage of this approach is that it leads to extremely simple low-power designs of the used samplers and that the information that has to be transmitted between the samplers and the master part of the whole data acquisition system is the timing information. That in turn makes such systems well suited for data acquisition from a very large number of signal sources. The gain for that is indirect randomization of the data acquisition process and the resulting necessity to process nonuniformly sampled signals digitally.

For technical realizations of the suggested data acquisition approach attention is focused on the complexity reduction of the data acquisition front-end designs. Obtaining extremely simple front-end designs is targeted, even at the cost of complicating the master part. To achieve this, the sampling procedure has to be organized in a way that permits the insertion of the communication link (e.g. a radio link) in the structure of the data acquisition system as close to the signal sources as possible. While this link usually connects each ADC with the master part of the data acquisition system, in this case the sampling operation and the whole structure of analog-to-digital conversions are organized so that the communication links are inserted inside the structure of the multi-input ADC. A block diagram of such a remote sampling ADC is given in Figure 11.7.

It can be seen from Figure 11.7 that the structure of such a distributed multichannel remote sampling ADC is subdivided into two parts. The first part is a cluster of extremely simple design input devices, each one containing a sampler. The samplers compare input signals during the time intervals when they are

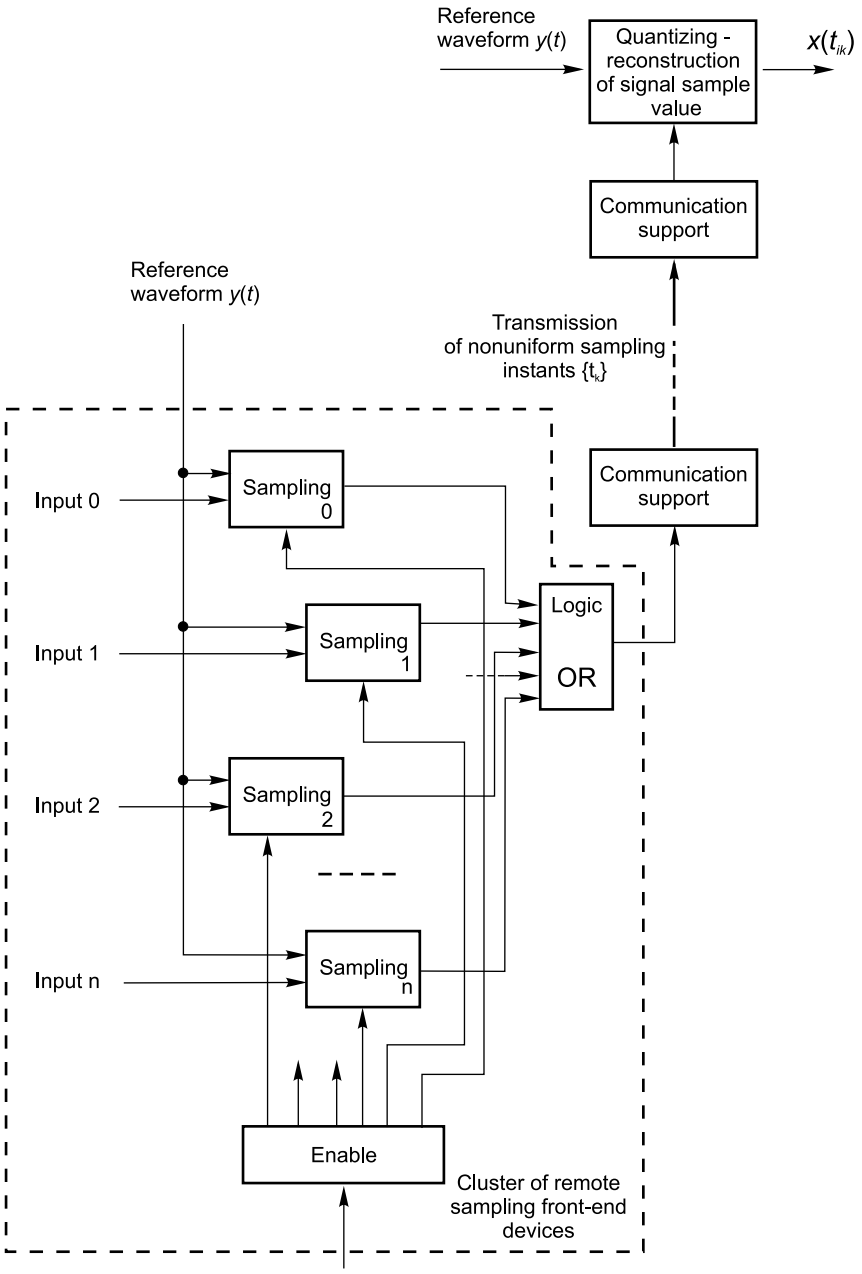


Figure 11.7 Distributed structure of a remote sampling ADC

enabled. Outputs of all samplers of a particular cluster are connected to an interface supporting transmission of the timing information, reflecting the sampling results over the communication link to the second part of the ADC. The circuits fulfilling the remaining analog-to-digital conversion functions are located in the second output part of this distributed ADC structure.

Bibliography

Artjuhs, J. and Bilinskis, I. (2006) Method and apparatus for alias suppressed digitizing of high frequency analog signals. EP 1 330 036 B1, European Patent Specification, Bulletin 2006/26, 28.06.2006.

Artyukh, Yu., Bilinskis, I. and Vedin, V. (1999) Hardware core of the family of digital RF signal PC-based analyzers. In Proceedings of the 1999 International Workshop on *Sampling Theory and Application*, Loen, Norway, 11–14 August 1999, pp. 177–9.

Artyukh, J., Bilinskis, I., Boole, E., Rybakov, A. and Vedin, V. (2005) Wideband RF signal digititising for high purity spectral analysis. In Proceedings of the International Workshop on *Spectral Methods and Multirate Signal Processing (SMMSP 2005)*, Riga, Latvia, 20–22 June 2005, pp. 123–8.

Bilinskis, I. and Mikelsons, A. (1992) *Randomized Signal Processing*. Prentice-Hall International (UK) Ltd.

Bilinskis, I. and Rybakov, A. (2006) Iterative spectrum analysis of nonuniformly undersampled wideband signals. *Electronics and Electrical Engineering*, **4**(68), 5–8.

Daria, V.R. and Saloma, C. (2000) High-accuracy Fourier transform interferometry, without oversampling, with a 1-bit analog-to-digital converter. *Applied Optics*, **39**(1), 1 January.

Kumaresan, R. and Wang, Y. (2001) On representing signals using only timing information. *J. Acoust. Soc. Am.*, **110**(5), Pt. 1, November.

Lim, M. and Saloma, C. (1998) Direct signal recovery from threshold crossings. *Physical Review E*, **58**(5), November.

Logan, B.F. (1977) Information in the zero crossings of band-pass signals. *Bell Syst. Tech. J.*, **56**, 487–510.

Marvasti, F. *A Unified Approach to Zero-Crossings and Nonuniform Sampling of Single and Multidimensional Signals and Systems*. Princeton, NJ: NONUNIFORM Publication.

Nazario, M.A. and Saloma, C. (1998) Signal recovery in sinusoid-crossing sampling by use of the minimum-negativity constraint. *Applied Optics*, **37**(14), 10 May.

Zeevi, Y.Y., Gravriely, A. and Shamai-Shitz, S. (1987) Image representation by zero and sine-wave crossings. *J. Opt. Soc. Am. A*, **4**, 2045–60.

12

Quantizing-specific Signal Parameter Estimation

Data acquisition does not necessarily mean the gathering of raw data. If the signals carrying information could be preprocessed in a simple and cost-effective way, then sometimes it is possible and makes sense to perform some kind of signal preprocessing carried out close to the signal sources. If the results of this preprocessing in separate data channels are acceptable for further processing of total information received from all channels then this approach is useful, as it apparently leads to data compression and less demanding requirements for the whole data gathering process. Of course, conditions for data acquisition, their processing and usage vary from case to case. Nevertheless, some typical and relatively often used signal preprocessing functions can be selected. In particular, an estimation of signal parameters seems to be a relatively popular class of signal conversions performed at the stage of signal preprocessing. They are partly discussed in this chapter and some aspects are considered in the next chapter.

A signal parameter estimation is discussed here. Imagine that an average parameter of a continuous-time signal, for instance its mean power, has to be estimated or measured by digital processing of quantized samples of the signal. To solve this seemingly simple task correctly, many fairly obvious questions have to be answered:

1. For how long should the signal be observed?
2. How precisely are the signal samples quantized and how many of them need to be processed to measure the mean power of the signal with the required accuracy?

Digital Alias-free Signal Processing I. Bilinskis

3. What are the restrictions on sampling intervals in this case? Should the requirements of the sampling theorem be satisfied?

In this chapter, these and other related questions are studied in an attempt to work out an appropriate approach to the problem of optimizing a signal average parameter estimation. Clearly, this kind of estimation can be considered to be most efficient when the estimation result is obtained in the shortest time possible and by processing a minimal quantity of bits. It will be shown that randomizing sampling and quantizing procedures helps to achieve optimal estimation conditions. Estimations of both random and periodic signal parameters are considered. In fact, when the input signals are sampled randomly, in most cases it does not matter whether these signals are random or periodic. However, processing the periodic signals demands a more careful approach, because if they are not processed over an integer number of periods, additional estimation errors may occur.

12.1 Theoretical Limits

Answers will now be found to questions 1 and 2 in the case when the rth moment

$$\mu_r = \lim_{\Theta \Rightarrow \infty} \frac{1}{\Theta} \int_0^{\Theta} x^r(t)\,\mathrm{d}t \tag{12.1}$$

of a stationary ergodic signal $x(t)$, observed during a time interval Θ, is to be estimated by applying digital processing methods.

12.1.1 Minimal Observation Time

Assume that the signal is sampled periodically and is quantized sufficiently accurately for the quantization errors to be negligible. The sampling intervals are equal to T. Under these conditions, the estimate $\hat{\mu}_r$ and its variance can be calculated on the basis of the following equations:

$$\hat{\mu}_r = \frac{1}{N} \sum_{k=1}^{N} x^r(kT), \tag{12.2}$$

$$\mathrm{Var}[\hat{\mu}_r] = \mathrm{Var}[x^r] \left[\frac{1}{N} + \frac{2}{N^2} \sum_{m=1}^{N-1} (N-m)\,\rho_{xr}(mT) \right], \tag{12.3}$$

where ρ_{xr} is the normalized autocorrelation function of the signal $x^r(t)$.

As the properties of this variance mostly depend on the second term in brackets, this term should be analysed first. This can be denoted as

$$L = \frac{2}{N^2} \sum_{m=1}^{N-1} (N - m)\, \rho_{xr}\,(mT). \tag{12.4}$$

It seems that it is more convenient to consider the spectral density function of the signal rather than the corresponding autocorrelation function. According to definition,

$$\rho_{xr}\,(t) = \frac{1}{\mathrm{Var}\,[x^r]} \int_0^\infty G_{xr}\,(f) \cos(2\pi f)\, \mathrm{d}f.$$

Substituting this equation into Equation (12.4) yields

$$L = \frac{1}{N^2 \mathrm{Var}\,[x^r]} \int_0^\infty \frac{\sin^2 \pi f N T}{\sin^2 \pi f T} G_{xr}(f)\, \mathrm{d}f - \frac{1}{N}, \tag{12.5}$$

where $G_{xr}(f)$ is the spectral density function of the signal $x^r(t)$. By substituting Equation (12.5) into Equation (12.3),

$$\mathrm{Var}[\hat{\mu}_r] = \frac{1}{N^2} \int_0^\infty \frac{\sin^2 \pi f N T}{\sin^2 \pi f T} G_{xr}(f)\, \mathrm{d}f. \tag{12.6}$$

On the basis of Equation (12.4), the limit of L for $N \Rightarrow \infty$, $T \Rightarrow 0$ and $NT = \Theta$ is given by

$$\begin{aligned} \lim_{\substack{N \Rightarrow \infty \\ T \Rightarrow 0}} L &= \lim_{\substack{N \Rightarrow \infty \\ T \Rightarrow 0}} \frac{2}{N^2} \sum_{m=0}^{N-1} (N - m)\, \rho_{xr}(mT) \\ &= \lim_{\substack{N \Rightarrow \infty \\ T \Rightarrow 0}} \frac{2}{N^2 T^2} \sum_{m=1}^{N-1} (NT - mT)\, \rho_{xr}\,(mT)\, T \\ &= \frac{2}{\Theta^2} \int_0^\Theta (\Theta - t)\, \frac{\Theta}{\rho_{xr}}\,(t)\, \mathrm{d}t. \end{aligned} \tag{12.7}$$

Similarly, from Equation (12.5),

$$\lim_{\substack{N \Rightarrow \infty \\ T \Rightarrow 0}} L = \frac{1}{\mathrm{Var}\,[x^r]} \int_0^\infty \frac{\sin^2 \pi f \Theta}{(\pi f \Theta)^2} G_{xr}(f)\, \mathrm{d}f. \tag{12.8}$$

It follows from Equations (12.7) and (12.8) that the value of the considered limit will always exceed zero if, for $t > 0$, $\rho_{xr}(t) \not\equiv 0\,(G_{xr}(f) \not\equiv 0)$. This value

becomes equal to zero only under the condition that the signal is observed and processed infinitely, i.e. when $\Theta = \infty$. It is not possible to suppress the variance $\mathrm{Var}\,[\hat{\mu}_r]$ and the random estimation errors due to it completely simply by enlarging the number N of the signal samples taken and processed. Indeed, if the observation time is fixed, then as N increases, $\mathrm{Var}\,[\hat{\mu}_r]$ tends to the value $\mathrm{Var}\,[x^r] \lim_{N \Rightarrow \infty, T=0} L$. On the other hand, if the sampling intervals T are increased till they exceed the correlation interval of the function $x^r(t)$, the component L is suppressed to zero and the value of the first component increases because, for a fixed Θ, enlarging T means decreasing N. In this case, the variance of the estimate $\hat{\mu}_r$ will be equal to $\mathrm{Var}[x^r]/N$.

Suppose that the estimation error of the parameter μ_r should not exceed the value εX^r with a probability β, where ε is the relative error and X is the range of the input signal $x(t)$. The minimal possible time interval Θ_0 during which the signal should be observed in order to satisfy the given requirements can be determined using Equation (12.7). As this equation presumes that for $N \Rightarrow \infty$, $\Delta t \Rightarrow 0$ and $NT = \Theta_0$, the value Θ_0 in fact represents a theoretical limit, characterizing signal processing under idealized conditions.

The value Θ_0 can be obtained from the following equation:

$$\lim_{\substack{N \Rightarrow \infty \\ \Delta t \Rightarrow 0}} L = \frac{2}{\Theta^2} \int_0^{\Theta_0} (\Theta_0 - t)\, \rho_{xr}\,(t)\, \mathrm{d}t$$

$$= \frac{\varepsilon^2 X^{2r}}{t_\beta^2 \mathrm{Var}\,[x^r]} = \alpha, \tag{12.9}$$

where t_β is half of the confidence interval corresponding to the confidence probability β. If the right-hand side of Equation (12.9) is denoted by α, then

$$\alpha = \frac{\varepsilon^2 X^{2r}}{t_\beta^2 \mathrm{Var}\,[x^r]}. \tag{12.10}$$

Equation (12.9) leads to the following conclusion. No matter how small the sampling interval T and how high the precision of quantizing, signal average parameters can be estimated with estimation errors below some given level only under the condition that signals are observed and processed at least during a time interval Θ_0. The value of Θ_0 depends on the parameter to be estimated, on the required estimation precision and on the autocorrelation function of the signal.

12.1.2 Sufficient Number of Signal Samples

It may sometimes be more important to take some essential estimation characteristic other than the estimation time to its limit. For instance, it is often desirable to minimize the number N of signal samples needed to obtain the required estimate with a specified accuracy. The equations given above can also be used to solve this task.

In general, to minimize N, the sampling interval T should be enlarged. If the value of Δt exceeds the maxima l correlation interval $\tau_{c,\max}$ of the signal $x^r(t)$, then $L = 0$ and under this condition it follows from Equation (12.3) that $\mathrm{Var}[\hat{\mu}_r] = (1/N)\mathrm{Var}[x^r]$ and $N \geq 1/\alpha$, where α is defined by Equation (12.10). Therefore, the minimal number of samples, denoted by N_0, is given by

$$N_0 = \frac{1}{\alpha} = \frac{t_\beta^2 \mathrm{Var}\,[x^r]}{\varepsilon^2 X^{2r}}. \tag{12.11}$$

Note that the corresponding estimation time (for $\Delta t = \tau$), defined as

$$\Theta_1 = N_0 \tau_{c,\max}, \tag{12.12}$$

exceeds the minimal value Θ_0 given by Equation (12.9).

The equations obtained lead to the following conclusion. Signal average parameters can be estimated with the demanded accuracy by processing $N_0 \leq N < \infty$ samples. If the signals are sampled at intervals equal to their maximal correlation intervals, the estimation can be performed by observing the signal during the time interval Θ_1 and by taking and processing the minimal number N_0 of samples. A further increase in observation time does not allow N to be reduced below the limit N_0.

The relationship of N versus $\Theta \in [\Theta_0, \Theta_1]$ can be derived from Equation (12.3), if the right-hand side of this equation is considered to be equal to $\varepsilon^2 X^{2r}/t_\beta^2$. Taking into account the fact that $T = \Theta/N$ yields

$$\frac{1}{N} + \frac{2}{N^2} \sum_{m=1}^{N-1} (N - m)\rho_{xr}\left(\frac{m\Theta}{N}\right) = \frac{\varepsilon^2 X^{2\mathrm{r}}}{t_\beta^2 \mathrm{Var}[x^r]} = \alpha. \tag{12.13}$$

12.1.3 Influence of Quantization Errors

Recall that the theoretical limits of Θ and N were obtained under the condition that the quantization errors were negligible. In reality this is not usually so because, as

shown below, better overall estimation characteristics are achieved if the signal samples are quantized in a coarse manner with relatively few thresholds.

It will now be established how the equations given above defining the theoretical limits of Θ and N change if quantization errors are taken into account, i.e. when the estimates $\hat{\mu}_r$ are obtained by processing quantized signal samples $\{\hat{x}_k\}$, rather than the samples $\{x_k\}$ themselves. Suppose that a sample x_k^r of the signal $x^r(t)$ is quantized. The quantized value of this sample $\hat{x}_k^r = x_k^r + \varepsilon_k$, where ε_k is the corresponding quantization error. Assume that the quantization noise $\{\varepsilon_k\}$ is not correlated with the signal $\{x_k^r\}$. As shown in Chapters 4 and 5, this assumption is justified in the cases of randomized and pseudo-randomized quantizing. Under the given conditions,

$$\text{Var}[\hat{\mu}_r] = \frac{1}{N}(\text{Var}[x^r] + \text{Var}[\varepsilon]) + \frac{2\text{Var}[x^r]}{N^2}\sum_{m=1}^{N-1}(N-m)\rho_{xr}(mT). \quad (12.14)$$

The following variable is introduced:

$$A = 1 + \frac{\text{Var}[\varepsilon]}{\text{Var}[x^r]}. \quad (12.15)$$

It is obvious that A depends on the quantization method applied. Now Equation (12.14) can be rewritten as follows:

$$\text{Var}[\hat{\mu}_r] = \text{Var}[x^r]\left[\frac{A}{N} + \frac{2}{N^2}\sum_{m=1}^{N-1}(N-m)\,\rho_{xr}(mT)\right]. \quad (12.16)$$

For randomized quantizing

$$A \cong 1 + \frac{q^r}{6\ \text{Var}[x^r]} \quad (12.17)$$

and for pseudo-randomized quantizing

$$A \cong 1 + \frac{q^r}{12\text{Var}[x^r]}. \quad (12.18)$$

However, it is often the signal samples x_k rather than the values x_k^r that are quantized and the estimate $\hat{\mu}_r$ is obtained by processing the quantized samples $\hat{\mu}_k$. In this case the variable A is defined by a more complex equation than Equations (12.17) and (12.18). For each specific estimation algorithm, the corresponding variance $\text{Var}[\hat{\mu}_r]$ should be found first. Then there will be no problem in defining the relative value of A.

When the quantizing errors are taken into account, the relationship N versus Θ, describing the conditions for estimating the signal average parameters μ_r with the required precision, is given by

$$\frac{A}{N} + \frac{2}{N^2} \sum_{m=1}^{N-1} (N - m)\, \rho_{xr}(T) = \alpha. \tag{12.19}$$

This equation shows how quantizing errors influence the estimation time and accuracy. For a more detailed analysis of these relationships, the specific signal autocorrelation or spectral density functions should be substituted into the above equations.

12.1.4 Estimation of Periodic Signal Parameters

The point is stressed once again that systems containing randomized or pseudo-randomized quantizers and employing random sampling techniques can be applied equally well to process both random and periodic signals. The randomization of sampling and/or quantizing does not orient these systems to some specific kind of signals. However, this does not mean that the conditions for processing periodic and random signals are always the same. These two categories of signals have different properties, which is important when the specifics of their processing are considered. These differences should be taken into account regardless of whether the corresponding signal processing systems are randomized or not.

The analytical results and conclusions obtained above are first of all applicable to processing random signals. When signals are periodic, other considerations should be taken into account. As the problems of processing periodic signals are really beyond the scope of this book, they will not be discussed in depth but the main points will simply be outlined:

1. In general, an estimation of periodic signal average parameters may be performed during shorter time intervals and by processing fewer signal samples. For instance, the mean value of a sinus wave can be measured with only two samples. This means that the processing of periodic signals may be more sensitive to quantization errors and, consequently, may require the application of more precise quantizers than those which are optimal for encoding random signals.
2. There are specific error sources that affect periodic signal processing. If no special algorithms are used then, as a rule, such signals should be processed

during an integer number of their periods. When the signal period is unknown, the signal parameter estimation time depends on the lowest frequency in its spectrum. The sampling rate depends on the upper frequency of the spectrum and on the particular parameter that is to be estimated. For instance, if the mean power of a signal is estimated, the sampling interval should be less than or equal to one-fourth of the period of the highest frequency present in the signal spectrum. This is true under conditions where the estimation is performed without preliminary reconstruction and resampling of the signal.

Thus periodic signal parameter estimation conditions are closely related to the specifics of the respective signals. In the following sections, the efficiency of randomized signal encoding for this kind of application is compared with the efficiency of deterministic encoding methods.

12.2 Optimal Estimation

Equation (12.19) can be used to optimize conditions of signal average parameter estimation with regard to different criteria. This equation shows how changing one of the variables A, N and Θ affects other variables. As the value of A for the given quantization model depends on the quantization step size q or on the number z of the threshold levels used, optimizing on the basis of this equation allows the best quantization conditions to be determined, ensuring an efficient estimation of the signal parameters required.

12.2.1 Minimizing the Number of Signal Samples

Criterion $N\Theta$

For a selected quantization method and the fixed number z of the threshold levels used, Equation (12.19) describes the relationship between N and Θ. Consider the estimation to be optimal when criterion $N\Theta$ reaches its minimum. The solution to this task depends first of all on the signal autocorrelation function. If this is given in an analytical form there are usually no serious problems in finding the minimum. In some other cases, the solution can be found only by applying less convenient methods of numerical analysis.

Consider an example of optimizing the conditions for estimating the mean value μ_x of a signal $x(t)$, which is characterized by the following triangular

autocorrelation function:

$$p_x(t) = \begin{cases} 1 - |t/t_0| & \text{for } |t| \le t_0, \\ 0 & \text{for } |t| > t_0. \end{cases} \tag{12.20}$$

The corresponding spectral density function is given by

$$G_{xx}(\omega) = \frac{4\sigma_x \sin^2 \omega t_0}{t_0 \omega^2}, \tag{12.21}$$

where σ_x^2 is the variance of the signal. Substituting Equation (12.20) into Equation (12.19) (after omitting manipulations) yields

$$N = \begin{cases} \dfrac{3\,(A-1)\,\Theta^2\omega_0^2}{t_0^2 + 3\Theta^2\alpha - 3\Theta t_0} & \text{for } \Theta \in [\Theta_0, \Theta_2]\,, \\ A/\alpha & \text{for } \Theta \ge \Theta_2, \end{cases} \tag{12.22}$$

where

$$\Theta_0 = \frac{(1 + \sqrt{1 - 4\alpha/3})t_0}{2\alpha} \tag{12.23}$$

and

$$\Theta_2 = \frac{A\tau_{c,\max}}{\alpha} = \frac{At_0}{\alpha}$$

is the estimation time, corresponding to the estimation of μ_x by processing uncorrelated signal samples.

Equation (12.22) can be used to define the optimal values of N and Θ with regard to the criterion $N\Theta$. Denote the respective optimal values of N and Θ by $N_{\text{opt}}^{(1)}$ and $\Theta_{\text{opt}}^{(1)}$. Then

$$\begin{aligned} \Theta_{\text{opt}}^{(1)} &= \frac{(1 + \sqrt{1-\alpha})t_0}{\alpha} \cong 2\Theta_0, \\ N_{\text{opt}}^{(1)} &= \frac{3(A-1)\,(1 + \sqrt{1-\alpha})^2}{\alpha(3 - 2\alpha + 3\sqrt{1-\alpha})} \cong \frac{2\,(A-1)}{\alpha}. \end{aligned} \tag{12.24}$$

It follows from Equation (12.24) that, under the conditions in question, the optimal sampling interval is given by

$$T_{\text{opt}}^{(1)} = \frac{\Theta_{\text{opt}}^{(1)}}{N_{\text{opt}}^{(1)}} \cong \frac{t_0}{A-1}. \tag{12.25}$$

However, the obtained estimation time will be optimal, in the given sense, only for $\Theta_{\text{opt}}^{(1)} \in [\Theta_0, \Theta_2]$. For $\Theta_{\text{opt}}^{(1)} \geq \Theta_2$, the observance time Θ_2 will be optimal, i.e. the optimal estimation will be performed under the condition that the signal samples are taken at the minimum time intervals for which these samples are still uncorrelated.

In the case under discussion this is $T = t_0$. Processing of uncorrelated samples is optimal for $A \leq 2$. For $A > 2$, the optimal sampling interval is equal to $T_{\text{opt}}^{(1)}$.

Example 12.1
The input signal is characterized by $X/\sigma_x = 10$, $t_0 = 10^{-6}$ s. Therefore the signal bandwidth is $f_0 = 0.5/t_0 = 500$ kHz. The mean value μ_x of the signal should be estimated with the relative random error $\varepsilon \leq 10^{-3}$. The confidence probability $\beta = 0.95(t_\beta = 2)$. The signal is quantized roughly and only two threshold levels are used.

Under the given conditions,

$$\begin{array}{lll} \Theta = 0.04 \text{ s}, & A = 3.083, & \alpha = 25 \times 10^{-6}, \\ \Theta_2 = 0.12 \text{ s}, & N_2 = 123\,320, & N_2\Theta_2 = 14\,798 \text{ s}, \\ \Theta_{\text{opt}}^{(1)} = 0.08 \text{ s}, & N_{\text{opt}}^{(1)} = 166\,640, & \dfrac{N_{\text{opt}}^{(1)}\Theta_{\text{opt}}^{(1)}}{N_2\Theta_2} = 0.9. \end{array}$$

The sampling frequency corresponding to the optimal estimation conditions is $f_{\text{s}} = 2.083 \times 10^3$ Hz.

12.2.2 Simplifying Hardware

The complexity of ADCs and the hardware used for processing digitized signals first of all depends on the number of bits of the corresponding quantizer. It is therefore of considerable practical interest to determine the minimum number of quantizer threshold levels at which it is still possible to solve the given estimation task with the required accuracy.

Criterion Nz
Minimizing this criterion allows the best conditions to be determined for estimating the parameter μ_x by processing a relatively small number of few-bit signal samples. Assume that the signal autocorrelation function is given by Equation (12.20) and that sampling is performed in such a way that the signal samples taken are uncorrelated. The variable z is a positive integer. The conditions are

found under which $Nz_1 < Nz_2$ if $z_1 < z_2$. To do this the following inequality should be satisfied:

$$\frac{Nz_1}{Nz_2} = \frac{A_1 z_1/\alpha}{A_2 z_2/\alpha} = \frac{A_1 z_1}{A_2 z_2} = \frac{z_1 + X^2/\left(12\sigma_x^2 z_1\right)}{z_2 + X^2/\left(12\sigma_x^2 z_2\right)} < 1, \tag{12.26}$$

where A_1 and A_2 correspond to $z = z_1$ and $z = z_2$. It follows from Equation (12.26) that

$$z_1 z_2 > \frac{X^2}{12\sigma_x^2}. \tag{12.27}$$

The inequality obtained can be applied to find the value of z that provides the minimum of criterion Nz. This procedure can be explained by an example.

Example 12.2
Let $X/\sigma_x = 10$. Substituting this value into the inequality (12.27) yields

$$z_1 z_2 > 8\tfrac{1}{3}. \tag{12.28}$$

Let $z_1 = 1$. Then it follows from expression (12.28) that it is better to use one threshold level than $z \geq 9$. At the same time, it is better to use from two to eight levels than one. Now let $z_1 = 2$. In this case it is better to use two threshold levels than $z \geq 5$. On the other hand, application of two threshold levels is less desirable than $z = 3$ or $z = 4$. By comparing the cases when $z = 3$ and $z = 4$, it is found that, under the given conditions, the best solution minimizing the criterion Nz is $z = 3$. The relationship of z versus the ratio σ_x/X, minimizing this criterion, is shown in Figure 12.1.

12.2.3 Minimizing Bit Flow

Criterion Nn
Minimizing this criterion allows the optimal number of quantization threshold levels to be determined, which guarantees that the estimate $\hat{\mu}_x$ will be obtained with the required accuracy by processing the minimum bits. Since values of z and n are connected by $z = 2^n - 1$, optimization can be carried out in the same way as in the previous case. If the minimum of N is reached by processing uncorrelated signal samples, as in the case of optimizing with regard to criterion Nz, the minimum of criterion Nn can be determined on the basis of the

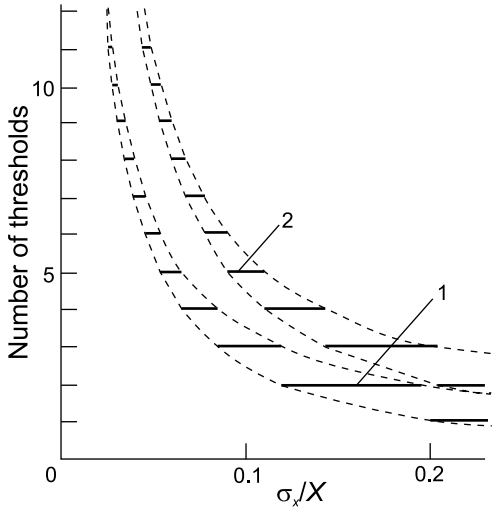


Figure 12.1 The optimal number of threshold levels versus the normalized signal root mean square

following equation:

$$\text{Criterion } Nn = \frac{n}{\alpha} + \frac{X^2 n}{12\alpha\sigma_x^2 (2^n - 1)^2}. \tag{12.29}$$

For instance, if $X/\sigma_x = 10$, it is found that the best solution is $n = 3$. This means that under the given conditions the estimation of μ_x can be performed with the required accuracy if the quantizer used for the signal encoding uses seven threshold levels. The relationship of n versus σ_x/X that minimizes criterion Nn is given in Figure 12.2.

Criterion $N\Theta z$

It is advisable to optimize the signal average parameter estimation with regard to this criterion in cases where the observation time obtained by optimizing with regard to criterion Nz is too long. The optimization procedure begins by optimizing with regard to criterion $N\Theta$ and then finding the value of z that gives the minimum value for criterion $N\Theta z$. This last part of the optimization procedure is performed in the same way as optimization of criterion Nz.

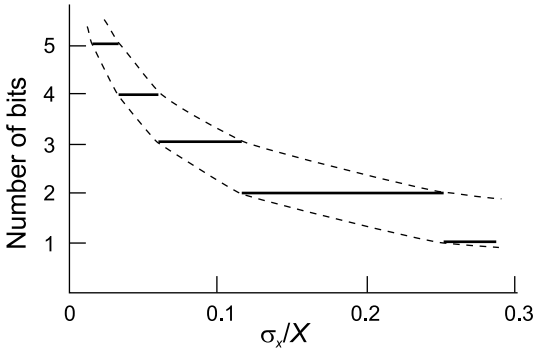


Figure 12.2 The optimal number of quantization bits as a function of the normalized signal root mean square

On the basis of the results obtained above in Equation (12.24),

$$\text{Criterion } N\Theta z = N_{\text{opt}}^{(1)}\Theta_{\text{opt}}^{1} z = \begin{cases} \dfrac{4\,(A-1)\,z t_0}{\alpha^2} = \dfrac{X^2 t_0}{3\alpha^2 \sigma_x^2 z} & \text{for } A > 2, \\ \dfrac{t_0 A^2 z}{\omega_0 \alpha^2} = \dfrac{t_0 \left[z^2 + X^2/\left(12\sigma_x^2\right)\right]^2}{\omega_0 \alpha^2 z^3} & \text{for } A \le 2. \end{cases} \tag{12.30}$$

It can be seen from Equation (12.30) that criterion $N\Theta z$ for $A > 2$ is inversely proportional to z. Of the many possible z values the optimal value is the largest providing $A > 2$. This optimal value is denoted by z_0.

After z_0 has been found, the value of criterion $N\Theta z\,(z_0)$ should be compared with the values of criterion $N\Theta z$ for $z > z_0$. If $N\Theta(z) < N\Theta(z_0)$, the minimum of criterion $N\Theta z$ should be determined by using the second equation of system (12.30) for which $A \le 2$.

Example 12.3
Assume that $X/\sigma_x = 10$. From the inequality

$$\begin{aligned} A &= \frac{1 + X^2}{12\sigma_x^2 z^2} \\ &= \frac{1 + 25}{3z^2 > 2}, \end{aligned}$$

it is found that $z_0 = 2$. Then it is found which of the z values above the level z_0 satisfies the inequality

$$\frac{\text{Criterion } N\Theta z\,(z_0)}{\text{Criterion } N\Theta(z)} < 1. \tag{12.31}$$

By applying Equation (12.30) it is easy to find that inequality (12.31) is not met, even for $z = 3$. Consequently, criterion $N\Theta z$ has a minimum of $z > 2$. This means that the minimum is achieved when the signal samples processed are uncorrelated. On the basis of Equation (12.30), it is found that the minimum of criterion $N\Theta z$, corresponding to $A < 2\,(z > 2)$, is $z = 5$. Hence

$$A = 4/3, \qquad N = 4/(3\alpha), \qquad \Theta = 4t_0/(3\alpha), \qquad T = t_0.$$

These results can be compared with respective values obtained by optimizing under the same conditions with regard to criterion Nz. In this case,

$$A = 5z/27, \qquad N = 5z/(27\alpha), \qquad \Theta = 5zt_0/(27\alpha), \qquad T = t_0.$$

The estimation optimized for criterion $N\Theta z$ can be performed during a time interval and by processing a number of signal samples, which are only 69.23 % from the respective values obtained by optimizing for criterion Nz. The sampling frequency in both cases can be the same, i.e. it should be equal to or less than ω_0, provided that the signal samples taken are uncorrelated. A reduction in observation time and the number of signal samples processed is achieved by increasing the number of quantizer threshold levels from 3 to 5. The function of z versus σ_x/X, providing the minimum for criterion $N\Theta z$, is shown in Figure 12.1 (the broken curve 2).

12.2.4 Deviations from Optimal Conditions

The optimal estimation conditions given above, providing minima for different criteria, depend on the normalized signal parameter σ_x/X. Although it is assumed that the input signal is stationary and, therefore, that it does not change its parameters during the estimation time interval, its mean power and, correspondingly, the parameter $\sigma_x\,X$ may, of course, change over time considerably within the whole range from zero to the maximum. The estimation conditions that are optimal for some value of $\sigma_x\,X$ will not really be optimal for other values of this parameter.

The theoretical lower limit of N is obtained for $z = \infty$. This limit is closely approached at small z values. It is apparent that nothing much can be gained (in the sense of minimizing N) by setting the number of quantizer threshold levels higher than four. Moreover, it seems that the best choice is $z = 2$, i.e. when one voltage comparator is used for positive half-waves of the signal and one for negative

half-waves. In this case, the quantizer and the processor are both simple and this reduction in the quantizer threshold level number can easily be compensated for by taking and processing a slightly increased number of signal samples.

Thus the analysis confirms the assumption that application of rough pseudo-randomized quantizing provides signal average parameter estimation conditions that are nearly optimal. Application of multibit quantizers only unduly complicates the processing hardware and slows down the estimation process rather than providing increasing accuracy.

12.2.5 Comments

The optimization of signal average parameter estimation considered above holds only under a number of assumptions and for the estimation of signal moments. The results obtained should therefore not be regarded as universally applicable. Of course, the estimation of other signal average parameters has also to be studied by taking specific estimation conditions into account. In relatively simple cases, this can be done analytically and the analysis above can be followed. In more complicated cases, however, the application of computer simulations of particular estimation procedures may prove to be more efficient for finding out the mode of quantizing that will provide the best estimation characteristics under the conditions given.

Although no presumptions should be made about the best mode of quantizing, analysis of specific signal average parameter estimation cases will, nevertheless, more often than not show that in this field of applications rough quantizing is preferable. This is not an entirely unexpected conclusion. It has been known for a long time that even when the quantization is deterministic it does not make sense to apply fine quantizing to digital estimation of correlation functions. It is usually considered that only six to eight threshold levels are needed to ensure nearly optimal estimation conditions. A further decrease in the number of threshold levels used is prevented by bias errors appearing when deterministic quantization becomes too coarse.

When quantizing is randomized or pseudo-randomized no bias errors occur, even with extremely rough quantizing; this approach therefore widens the application area of this kind of quantizing considerably. Randomized or pseudo-randomized rough quantizing can usually be performed by using fewer threshold levels than are required for carrying out this operation under similar conditions deterministically.

These considerations apply not only to the estimation of moments and correlation functions but also to other more important signal parameters, such as

Fourier, and similar, coefficients estimated in the course of spectral transforms, which are performed on an exponential and other orthogonal function basis. Their estimation is considered in the following chapters. As shown there, optimizing is carried out by taking into account some additional factors, for instance, whether or not the signal is noisy. Both fine and rough quantizing modes are applicable for estimating these parameters.

12.3 Specifics Related to Pseudo-randomized Quantizing

The previous sections have clearly shown that when signal average parameters have to be estimated, it is advantageous to quantize signals pseudo-randomly. However, as most of the existing algorithms for digital estimation of such parameters have been developed for processing deterministically quantized signals, the application of pseudo-randomized quantizing is possible only if the processing algorithms are appropriately modified.

Attempts to replace deterministic by pseudo-randomized quantizers without adapting the processing algorithms and hardware to the specifics of such quantizing may well be unsuccessful. Although the digital signals obtained by pseudo-randomized quantizers have considerably superior properties, the price for this improvement is an increased bit rate and corresponding hardware complications.

Recall that pseudo-randomly quantized signals are defined as

$$\hat{x}_k = \left(\xi_k - \tfrac{1}{2}\right) q + n_k q. \tag{12.32}$$

The first term of this equation indicates the positions that the threshold levels occupy at the quantization time instants t_k when the rounding-off results $n_k q$ are obtained. In the case of deterministic quantizing, the thresholds are fixed and it is not necessary to add this information to the descriptions of the quantized signals. Pseudo-randomized quantizer outputs therefore have more bits in principle, and for this reason their straightforward applications may sometimes be questionable. Overall good results from the application of such quantizers may only be expected if their output signals are processed in a way best suited to the given conditions.

However, as the conditions of signal processing can differ considerably, it is not possible to give precise recipes for every particular case; for success the general technical principles must be understood. The material in this chapter may be found to be helpful in this.

It is first recommended that the digital signal processing be organized in such a way that the sequences $(\xi_k - \frac{1}{2})q$ and $n_k q$ are processed separately. Next, it should be taken into account that the process $(\xi_k - \frac{1}{2})q$ is predetermined and can

be repeated. This approach sometimes permits some kind of preprocessing of this pseudo-random sequence where the results are stored in a memory and then used for corrections.

There may be various approaches to the problem of processing the quantized signal components $(\xi_k - \frac{1}{2})q$ and $n_k q$ separately. Of course, this processing should be matched with the corresponding algorithms, but new, unconventional ways of dealing with the auxiliary pseudo-random process can still be found. Several typical approaches will now be examined.

12.3.1 Avoiding Processing of the Dither Process

Under some conditions it is possible to simplify processing of pseudo-randomly quantized signals in such a way that there is no need to take into account the information carried by the sequence $\{\xi_k\}$. To elucidate this an estimation of the absolute mean value $m_{|x|}$ of a stationary ergodic signal $x(t)$ is considered. This signal parameter is defined as

$$m_{|x|} = \frac{1}{\Theta} \int_0^{\Theta} |x(t)| \, \mathrm{d}t. \tag{12.33}$$

The corresponding estimate is given by

$$\hat{m}_{|x|} = \frac{1}{N} \sum_{k=1}^{N} |\hat{x}_k|.$$

If the quantized signal $\{\hat{x}_k\}$ is obtained by means of a pseudo-randomized quantizer, this estimate becomes

$$\hat{m}_{|x|} = \frac{q}{N} \sum_{k=1}^{N} \left| \xi_k - \frac{1}{2} + n_k \right| = \frac{q}{N} \sum_{k=1}^{N} \left(\xi_k - \frac{1}{2} \right) + \frac{q}{N} \sum_{k=1}^{N} n_k. \tag{12.34}$$

As the process $\{\xi_k\}$ is pseudo-random, clearly the first sum in Equation (12.34) can be calculated in advance or, under certain conditions, not calculated at all. If the auxiliary pseudo-random dithering process $\{\xi_k\}$ is generated in such a way that the instantaneous values of ξ_k do not repeat themselves within a certain cycle, and an integer number of those cycles is equal to N, this sum is equal to zero. Under this condition, the estimate $\hat{m}_{|x|}$ can be obtained in the following simplified way:

$$\hat{m}_{|x|} = \frac{q}{N} \sum_{k=1}^{N} n_k. \tag{12.35}$$

Although in this case the information carried by the process $\{\xi_k\}$ is seemingly not taken into account, this is actually not the case. Equation (12.35) holds only under the given condition, and this means that quantizing is performed while the threshold level sets are in predetermined positions. These positions are not tied to the corresponding quantizing time instants, but this is of no importance, because the definition of this estimate does not require that they should be. The probabilistic characteristics of $\hat{m}_{|x|}$ obtained from Equation (12.35) will therefore be the same as if this estimate had been calculated in a more conventional way, in accordance with Equation (12.34), without using special versions of the pseudo-random process $\{\xi_k\}$.

12.3.2 Simplified Processing of the Dither Process

There can be different approaches to the problem of simplifying the processing of pseudo-randomly quantized signals. Naturally, they depend first of all on the processing algorithms. However, success in this to a large extent also depends on the ability to look at this problem in an inventive way. Consider, for instance, the estimation of the mean power P of a stationary ergodic signal $x(t)$ which is observed during a time interval $[0, \Theta]$. According to the definition,

$$P = a_2 = \lim_{\Theta \Rightarrow \infty} \frac{1}{\Theta} \int_0^{\Theta} x^2(t)\, \mathrm{d}t.$$

When this signal is sampled and quantized, the corresponding estimate is given by

$$\hat{P} = \frac{1}{N} \sum_{k=1}^{N} \hat{x}_k^2 - \frac{1}{12} q^2 \tag{12.36}$$

where the term $\frac{1}{12}q^2$ is the correction that renders the estimate (12.36) unbiased.

The pseudo-randomly quantized signals can be represented in the following way:

$$\hat{x}_k = q(n_k + c_k), \tag{12.37}$$

where $c_k = \xi_k - \frac{1}{2}$. Substituting this expression into Equation (12.36) yields

$$\hat{P} = \frac{q^2}{N} \sum_{k=1}^{N} \left(n_k^2 + 2n_k c_k + c_k^2\right) - \frac{1}{12} q^2. \tag{12.38}$$

Note that the term $(1/N)\sum_{k=1}^{N} c_k^2$ is an estimate of $E[c_k] = \frac{1}{12}$. Hence,

$$\hat{P} = \frac{q^2}{N}\sum_{k=1}^{N}\left(n_k^2 + 2n_k c_k\right) = \frac{q^2}{N}\sum_{k=1}^{N}\left(n_k^2 + 2n_k\xi_k - n_k\right). \tag{12.39}$$

In this way, the term containing ξ_k^2 can be cancelled. This simplifies calculations, because very rough quantizing is preferable for this kind of application and the value of n_k is then not large.

The most interesting and also the most practical overall results can be obtained by applying rough quantizing with only one or two (one for negative and one for positive signal half-waves) threshold levels. In this case n_k can assume only the values 0 and 1 (the signal cannot be positive and negative simultaneously) for $n_k^2 = n_k$ and Equation (12.39) becomes

$$\hat{P} = \frac{2q^2}{N}\sum_{k=1}^{N} n_k\xi_k \tag{12.40}$$

This estimate $\hat{P}$ can then be obtained without performing multidigit multiplication operations. In this approach, only one adder is required to carry out the estimation and, consequently, the electronic implementation of this estimation scheme is extremely simple. This is important for at least two reasons. Firstly, the simplicity of the input circuits allows the design of very wide band devices. Secondly, the time intervals between quantizing instants can be small and a high speed of estimation can be achieved.

Devices for estimating the absolute mean values and the mean power of signals are described in the following sections. It remains to find out how widely they can be applied. The above conclusion, stating that the application of extremely rough pseudo-randomized quantizing provides nearly optimal conditions for estimating signal average parameters, is based on the optimization of the mean value estimating process. This conclusion does not necessarily hold for the estimation of higher-order signal distribution moments. Fortunately, the results obtained and discussed in the following sections show that this quantization technique is well suited for more complicated average parameter estimation cases as well. Although the approach to the problem of simplifying processing of the quantized signal component $(\xi_k - \frac{1}{2})q$ discussed provides good results, even better effects can sometimes be achieved. For some pseudo-randomized quantizing applications it is possible to process the sequence $\{\xi_k\}$ beforehand and to store the results in a memory for applying as corrections to the corresponding estimation results. This approach is effective for spectrum analysis.

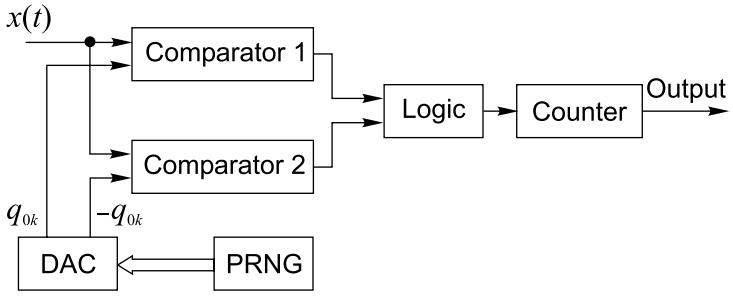


Figure 12.3 Scheme of an electronic device for measuring the absolute mean value of signals

12.4 Estimation of the Absolute Mean Value

An average parameter estimation of signals can be complicated depending on the definition of a particular parameter. It is a relatively simple matter to estimate signal moments and related parameters, so it seems logical to begin by considering the technique for estimating them.

A randomized scheme for estimating a signal absolute mean value is studied here. Its performance is first of all evaluated in terms of random estimation errors. Note that there can be instrumental as well as software implementations of this and other estimation schemes, provided that signal quantizing is randomized or pseudo-randomized.

12.4.1 Electronic Device

A block diagram of an electronic device for estimating the absolute mean values of pseudo-randomly quantized signals is shown in Figure 12.3. This device comprises a pseudo-randomized ADC and some functional blocks for processing the quantized signals. It can in fact be used to estimate two signal parameters in parallel: the absolute mean value and the mean power. The price of adding one more function, in this case, is very low. Only an additional simple logic gate and adder are needed.

In accordance with the optimization results given above, the pseudo-randomized quantizer contains only two voltage comparators, one for positive and one for negative signal values. The quantizing thresholds q_{0k}^{-} and q_{0k}^{+}, pseudo-randomly changing in time, are generated by the means of the pseudo-random number generator and a DAC. The latter has two outputs, the first for q_{0k}^{+} and the second for q_{0k}^{-}.

To ensure proper implementation of the estimation algorithm, the following requirements should be met:

1. The absolute values of the threshold levels q_{0k}^- and q_{0k}^+ at each instant should be equal.
2. Pseudo-random numbers ξ_k, generated by the PRNG, are not repeated within a cycle of N_0 numbers.
3. Estimating is carried out until $N = nN_0$ signal samples are taken and processed, where n is a positive integer.

Consider a centred signal $x(t) \in [-q, q]$, $m_x = 0$. Under the given conditions, the estimate of its absolute mean value is given by

$$\hat{m}_{|x|} = \frac{q}{N} \sum_{k=1}^{N} n_k, \tag{12.41}$$

where $n_k = n_k^+ + n_k^-$ is the sum of the comparator C1 and C2 outputs. For $x_k \geq 0$, $n_k^- \equiv 0$ and for $x_k < 0$, $n_k^+ \equiv 0$. Hence n_k can assume only the value 1 or 0. Therefore $\hat{m}_{|x|}$ is proportional to the number of logical 1's accumulated in the counter.

12.4.2 Estimation Errors

The estimate that is being considered is unbiased. Therefore estimation inaccuracy is completely determined by random errors. As $\hat{m}_{|x|}$ can be considered as the estimate of the expected value of $|x(t)|$, then

$$\mathrm{Var}[\hat{m}_{|x|}]_q = \frac{1}{N}\mathrm{Var}[|\hat{x}_k|] = \frac{1}{N}\mathrm{Var}[|x_k| - \varepsilon_k], \tag{12.42}$$

where ε_k is the random error of pseudo-randomized quantizing and the index q at $\mathrm{Var}[\hat{m}_{|x|}]$ indicates that this variance is due to quantizing. Taking into account the properties of pseudo-randomized quantizing,

$$\mathrm{Var}[\hat{m}_{|x|}]_q = \frac{1}{N}\left(\sigma_{|x|}^2 + \frac{1}{12}q^2\right). \tag{12.43}$$

This equation describes the variance of the considered estimate under the condition where the pseudo-random numbers $\{\xi_k\}$ are mutually statistically independent. In this case, this condition is not satisfied because the auxiliary pseudo-random process is generated in such a way that the numbers ξ_k cannot repeat themselves within a certain cycle. However, the impact of this is negligible.

By definition, the random estimation error caused by the quantization errors is given as

$$\varepsilon_q \leq t_\beta \sqrt{\mathrm{Var}[\hat{m}_{|x|}]_q}. \tag{12.44}$$

To ensure that the error will not exceed some given value ε, more than N quantized sample values should be averaged. It follows from Equations (12.43) and (12.44) that

$$N \geq \frac{t_\beta^2}{\varepsilon^2}\left(\sigma_{|x|}^2 + \frac{1}{12}q^2\right). \tag{12.45}$$

This inequality can be used to calculate number N of signal samples that have to be processed to ensure with probability β that the random estimation error will not exceed the predetermined value ε.

To evaluate the efficiency of the estimation method, it can be compared with others, specifically with those based on extremely rough randomized quantizing and on multibit deterministic quantizing. The variances of the estimates $\hat{m}_{|x|}$ obtained in those cases are given, respectively, by the following equations:

$$\mathrm{Var}[\hat{m}_{|x|\mathrm{r}}]_q = \frac{1}{N}\hat{m}_{|x|}(q - \hat{m}_{|x|}), \tag{12.46}$$

$$\mathrm{Var}[\hat{m}_{|x|\mathrm{id}}]_q = \frac{1}{N}\sigma_{|x|}^2, \tag{12.47}$$

where $\hat{m}_{|x|}$ is the absolute mean value and $\sigma_{|x|}^2$ is the variance of the rectified signal. The indexes *r* and *id* indicate that the corresponding estimates are obtained by applying the extremely rough randomized and the extremely fine idealized deterministic quantizers respectively.

Obtaining of the estimate $\hat{m}_{|x|\mathrm{id}}$ by processing the signal samples, taken without quantizing errors, is more complicated, because in this case multibit rather than one-bit numbers have to be averaged. To obtain the estimates $\hat{m}_{|x|\mathrm{r}}$ and $\hat{m}_{|x|\mathrm{id}}$, it is necessary to take and to process N_r, N_id signal samples respectively. It follows from Equations (12.46) and (12.47) that

$$N_\mathrm{r} \geq \frac{t_\beta^2}{\varepsilon^2}m_{|x|}(q - m_{|x|}), \tag{12.48}$$

$$N_\mathrm{id} \geq \frac{t_\beta^2}{\varepsilon^2}\sigma_{|x|}^2. \tag{12.49}$$

Although the absolute mean values of signals are not measured very frequently, this signal processing operation can fairly often be found in more complicated algorithms. The given analysis shows what happens if the signals to be estimated are quantized randomly, pseudo-randomly and deterministically. It can be seen from the given relationships that application of pseudo-randomized quantizing allows an estimation to be performed in a very simple way and the loss

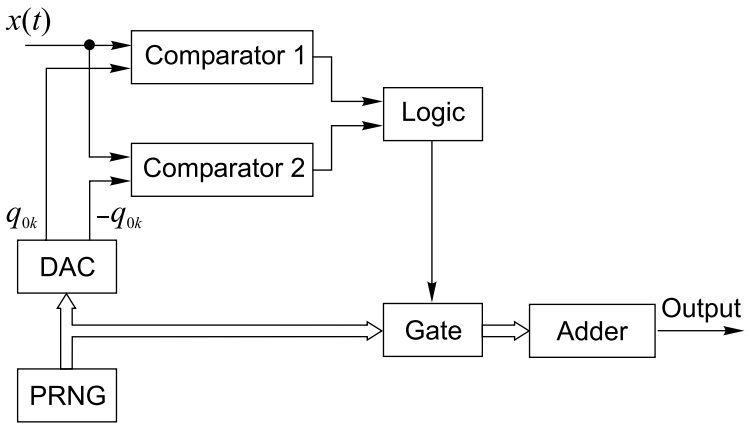


Figure 12.4 Block diagram of a device performing an estimation of the signal mean power

in estimation precision is not very significant. If it is possible to enlarge the data block processed, the estimation errors can easily be suppressed to the same level as when the signal samples are obtained virtually without quantization error.

12.5 Estimation of the Mean Power

Consider estimation of the mean power of a stationary ergodic zero mean signal $x(t) \in [-q, q]$. Suppose that this procedure is realized on the basis of the estimation method described by Equation (12.40). The electronic device applicable for such an estimation is shown in Figure 12.4. It can be seen from Equation (12.40) and the block diagram of this device that the pseudo-random numbers ξ_k generated by the PRNG are passed through a logic gate to an adder. This gate is open at the instants t_k when $n_k = 1$. The value of the estimated signal mean power is read out from the adder at the end of each estimation cycle.

Note that the method of estimating the mean power of signals considered here is very suitable for high-frequency applications. Its electronic implementation is extremely simple and can easily be developed as a very wide band device. To do this only the two voltage comparators at the input need to be designed as ultrahigh-frequency electronic elements, and the restrictions on the input signal spectra imposed by aliasing can be avoided by applying randomized sampling. There are no problems in accomplishing this.

12.5.1 Estimation Efficiency

The estimate $\hat{P}$ of the signal mean power P, as defined by Equation (12.40), is unbiased. Then

$$E[\hat{P}] = 2q^2 \left[\int_0^q \phi(x) \int_0^{x/q} \xi \, \mathrm{d}\xi \, \mathrm{d}x + \int_{-q}^0 \phi(x) \int_0^{-x/q} \xi \, \mathrm{d}\xi \, \mathrm{d}x \right]$$
$$= \int_{-q}^q x^2 \phi(x) \, \mathrm{d}x = P, \tag{12.50}$$

where $\phi(x)$ is the probability density function of the signal. The variance of this estimate can be derived from Equation (12.40). By definition,

$$\mathrm{Var}[\hat{P}]_q = \frac{4q^4}{N} \mathrm{Var}[\xi_k n_k] + \frac{8q^4}{N^2} \sum_{m=1}^{N-1} \sum_{k=1}^{N-m} \mathrm{Cov}(\xi_k n_k, \xi_{k+m} n_{k+m}), \tag{12.51}$$

where $\mathrm{Cov}\,(\xi_k n_k, \xi_{k+m}, n_{k+m})$ is the covariance of the random variables $\xi_k n_k$, ξ_{k+m} and n_{k+m}. Equation (12.51) can be given as

$$\mathrm{Var}[\hat{P}]_q = \frac{4}{3N} q[|x|^3] - \frac{1}{N} P^2 + \frac{2}{N^2} \sum_{m=1}^{N-1} (N - m) \, C_{x2}(mT), \tag{12.52}$$

where $C_{x2}(mT)$ is the autocovariance function of $x^2(t)$.

To evaluate the efficiency of this estimation method it should be compared with other methods. When the signals to be estimated are quantized randomly, the estimate of the mean power is given as

$$\hat{P}_\mathrm{r} = -\frac{q^2}{N} \sum_{k=1}^{N} n_{1k} n_{2k}, \tag{12.53}$$

where n_{1k} and n_{2k} are the output signals of the voltage comparators and the index r for P indicates that $x(t)$ is quantized randomly. The variance of this estimate is obtained in the same way as the variance described by Equation (12.52). In this case

$$\mathrm{Var}[\hat{P}_\mathrm{r}]_q = \frac{1}{N}(q^4 - P^2) + \frac{2}{N^2} \sum_{m=1}^{N-1} (N - m) \, C_{x2}(mT). \tag{12.54}$$

When the signal samples are taken without the quantization error, the estimate is defined by

$$\hat{P}_\mathrm{id} = -\frac{q^2}{N} \sum_{k=1}^{N} x_k^2 \tag{12.55}$$

and the variance of this estimate is given by

$$\mathrm{Var}[\hat{P}_{\mathrm{id}}]_q = -\frac{1}{N}(E[x^4] - P^2) + \frac{2}{N^2}\sum_{m=1}^{N-1}(N-m)\,C_{x2}(mT). \quad (12.56)$$

The random estimation error due to quantizing imperfections is defined as

$$\varepsilon_q \le t_\beta \sqrt{\mathrm{Var}[\hat{P}]} \quad (12.57)$$

The random errors characterizing the estimation of variously quantized signals can therefore be calculated by substituting the corresponding variances into inequality (12.57).

Equations (12.52), (12.54) and (12.56), defining the variances of the estimates can be used for deriving expressions for calculating the number N of the signal samples needed to obtain these estimates with errors not exceeding the preset values of ε. Then

$$N \ge \frac{t_\beta^2}{\varepsilon^2}\left(\frac{4}{3}qE[|x|^3] - P^2\right), \quad (12.58)$$

at random quantizing

$$N_{\mathrm{r}} \ge \frac{t_\beta^2}{\varepsilon^2}(q^2 - P^2) \quad (12.59)$$

and without quantizing error

$$N_{\mathrm{id}} \ge \frac{t_\beta^2}{\varepsilon^2}(E[x^4] - P^2). \quad (12.60)$$

12.6 Errors Due to Randomized Sampling

Randomizing various signal processing techniques is often controversial, because negative as well as positive effects may arise. Whether the desirable effects prevail or not depends on the particular signal processing conditions and the outcome is not always easy to predict.

The conditions of signal average parameter estimation are usually well suited to the specifics of the randomized signal processing as far as randomized quantization is concerned. Consider what happens if the same parameters are estimated by processing randomly sampled signals. The following considerations are for the case when signals are sampled according to the direct randomization scheme described in Chapter 6.

12.6.1 Absolute Mean Value Estimate

Suppose that a stationary ergodic signal $x(t)$ is sampled randomly at the instants $\{t_k\}$ and that the signal samples $\{x_k\}$ are taken without quantization error. The estimate $\hat{m}_{|x|}$ of the absolute mean value is given by

$$\hat{m}_{|x|} = \frac{1}{N}\sum_{k=1}^{N} |x_k| \tag{12.61}$$

and the variance of this estimate by

$$\mathrm{Var}[\hat{m}_{|x|}]_s = 2\int_0^{\omega,\max} \mathrm{Var}[\hat{m}\,(\omega)]\, G_{|x||x|}\,(\omega)\,\mathrm{d}\omega, \tag{12.62}$$

where $G_{|x||x|}$ is the spectral density function of the rectified signal $|x\,(t)|$. The index s of $\mathrm{Var}[\hat{m}_{|x|}]$ indicates that this variance of the estimate $\hat{m}_{|x|}$ is caused by sampling irregularities.

12.6.2 Mean Power Estimate

Consider estimate (12.56). The variance of this estimate

$$\mathrm{Var}[\hat{P}_{\mathrm{id}}]_s = 2\int_0^{2\omega,\max} \mathrm{Var}[\hat{m}\,(\omega)] G_{yy}\,(\omega)\,\mathrm{d}\omega, \tag{12.63}$$

where

$$G_{yy}(\omega)\;\mathrm{d}\omega = \frac{1}{2\pi}\int_0^{\infty} G_{xx}(v)\, G_{xx}(\omega - v)\,\mathrm{d}v \tag{12.64}$$

and $G_{xx}(\omega)$ is the spectral density function of the signal $x(t)$.

12.6.3 Overall Estimation Errors

Consider the estimate $\hat{P}$. Its random error ε_0 can be given as

$$\varepsilon_0 \le t_\beta \sqrt{\mathrm{Var}[\hat{P}_{\mathrm{id}}]_q + \mathrm{Var}[\hat{P}]_s} \tag{12.65}$$

This error naturally depends on the variances caused by both the rounding-off errors and the sampling irregularities. The relationships underlying it are in general fairly complicated. This is true especially with regard to that component of the error that is tied to random sampling. It can be seen from Equations (12.63) and (12.64) that the corresponding variance is determined by the frequency-dependent

variance Var[$\hat{m}(\omega)$] of the mean value of an elementary signal and by the spectral density function of $x(t)$.

Fortunately, in the case considered, it is not necessary to calculate ε_0 in detail. The maximal value of the relative estimation error is around $0.5/q^2$ for $N = 1$. To suppress this error, for instance 500 times, 2.5×10^5 signal samples have to be taken and processed. This number of samples is not needed to compensate for excessive rounding-off errors; it is needed to measure the mean power of the signal according to the definition of this parameter. No matter how the signal is quantized, the number N of signal samples processed should always be large enough to achieve acceptable estimation precision. Consequently, the variance Var [$\hat{m}(\omega)$] will to a considerable extent also be suppressed. Calculations show that in this signal processing area the errors due to random sampling irregularities are usually suppressed to negligible values.

Bibliography

Bilinskis, I. and Mikelsons, A. (1983) *Digital Random Processing of Continuous Signals* (in Russian). Riga: Zinatne.

Bilinskis, I. and Mikelsons, A. (1992) *Randomized Signal Processing*. Prentice-Hall International (UK) Ltd.

Bilinskis, I., Mikelsons, A. and Skageris, A. (1981) Device for randomized estimation of root mean square values of periodical signals. Author's Certificate 869019, USSR, Bulletin of Inventions 36, p. 282.

Bilinskis, I., Mikelsons, A., Skageris, A. and Schvetski, B.I. (1979) Quasi-randomized method of estimating the mean power of signals (in Russian). In *Analog-Digital Conversion of Signals*, Vol. 4. Riga: Zinatne, pp. 3–10.

Bilinskis, I., Mikelsons, A. and Viksna, J. (1974) Functional sweeping transforms based on random tests (in Russian). *Integrated Circuits in Digital Technology*, **2**, 145–54.

Bilinskis, I., Mikelsons, A. and Vystavkin, A.N. (1986) Processing of randomly-sampled signals. In *Signal Processing*, III: *Theories and Applications*. Amsterdam: Elsevier Science Publishers, pp. 109–12.

Gedance, A.R. (1972) Estimation of the mean of a quantized signal. *Proc. IEEE*, **60**(8), 1007–8.

Hejn, K. (1978) Application of Monte-Carlo methods in measurement. In Symposuim IMEKO Tech. Comm. on *Measurement Theory*, Leningrad.

13

Estimation of Correlation Functions

Continuing the discussion on the specifics of processing nonuniformly quantized signals, started in the previous chapter, the issues of quantized signal multiplication are studied and put into focus here. They are of special interest in many cases where two or more pseudo-randomly digitized signals are to be processed together. An estimation of correlation functions represents the most typical case. As a rule, numerous multiplications of the quantized signal sample values need to be executed in order to perform signal correlation analysis. As these operations are relatively time consuming, it is essential to rationalize them. That is especially important in cases of special hardware development for correlation analysis. The most suitable quantization method has to be selected first. Referring back to Chapter 5, it is clear that application of pseudo-randomized quantizing should be considered. However, it seems that multiplication of two pseudo-randomly quantized signals is more complicated than multiplication of randomly or deterministically quantized signals. On the other hand, the properties of pseudo-randomized quantizing are superior to the analogous properties of other quantizing techniques.

Application of pseudo-randomized sampling for signal correlation analysis is also discussed. It is shown that signal digitizing strongly impacts on the conditions for the correlation analysis and that applying pseudo-randomized digitizing techniques, if done skilfully, leads to certain desirable effects. Therefore it is worth examining various processes related to this kind of signal correlation analysis more closely.

Digital Alias-free Signal Processing I. Bilinskis

13.1 Multiplication of Quantized Signals

In order to highlight the problems arising when two or more pseudo-randomly quantized signals are processed together, multiplication of two quantized signals $\hat{x}$ and $\hat{y}$ are considered, where

$$\hat{x} = q(c_x + n_x), \qquad \hat{y} = q(c_y + n_y)$$

and

$$c_x = \xi_x - \tfrac{1}{2}, \qquad c_y = \xi_y - \tfrac{1}{2}.$$

13.1.1 Expected Value of Multiplied Quantized Signals

The product

$$z = \hat{x}\hat{y} = q^2(c_x c_y + c_x n_y + c_y n_x + n_x n_y) \tag{13.1}$$

is obviously a random variable. Its expected value is

$$E[z] = E[\hat{x}\hat{y}] = q^2(E[c_x c_y] + E[c_x n_y] + E[c_y n_x] + E[n_x n_y]).$$

It can easily be shown that the first three expectations are equal to zero. Hence

$$E[z] = E[\hat{x}\hat{y}] = q^2 E[n_x n_y]. \tag{13.2}$$

Usually it makes sense to apply pseudo-randomized quantizing only if it can be coarse. To simplify the following analysis, assume that the quantization considered is extremely coarse, i.e. that

$$n_x = \begin{cases} 1 & \text{for } q\xi_x \le x, \\ 0 & \text{for } q\xi_x > x, \end{cases} \qquad n_y = \begin{cases} 1 & \text{for } q\xi_y \le y, \\ 0 & \text{for } q\xi_y > y. \end{cases}$$

It then follows from Equation (13.2) that

$$\begin{aligned} E[z] &= q^2 \int_0^q \int_0^q \int_0^x \int_0^y \frac{\mathrm{d}(q\xi_x)}{q} \frac{\mathrm{d}(q\xi_y)}{q} \varphi(x, y)\, \mathrm{d}x\, \mathrm{d}y \\ &= E[xy] = a_{11}. \end{aligned} \tag{13.3}$$

To minimize the bias error, the multiplication of $\hat{x}$ and $\hat{y}$ can therefore be reduced to multiplying n_x and n_y, without taking into account other terms of Equation (13.1). However, this does not mean that those other terms are redundant. In principle, their addition reduces the statistical error. On the other hand, these additional terms also complicate calculations.

13.1.2 Variance of Multiplication Results

As it is assumed that n_x and n_y are not large (typically they are smaller than the respective values ξ_x and ξ_y), then the most inconvenient term is $\xi_x\,\xi_y$, because it requires the multiplication of multidigit numbers. To decide which of the terms can be cancelled without unreasonable increase in the number of random errors, various multiplying options have to be evaluated.

The following notation is introduced:

$$\begin{aligned} z^{(1)} &= q^2(c_x n_y + c_y n_x + n_x n_y), \\ z^{(2)} &= q^2 n_x n_y. \end{aligned} \tag{13.4}$$

As all three of the estimates z, $z^{(1)}$ and $z^{(2)}$ are unbiased, they differ only from the viewpoint of their random errors, which are defined by the respective variances $\mathrm{Var}[z]$, $\mathrm{Var}[z^{(1)}]$ and $\mathrm{Var}[z^{(2)}]$. Then

$$\begin{aligned} \mathrm{Var}[z] = q^4 \{&\mathrm{Var}[c_x c_y] + 2K(c_x c_y, c_x n_y) + 2K(c_x c_y, c_y n_x) \\ &+2K(c_x c_y, n_x n_y) + \mathrm{Var}[c_x n_y] + 2K(c_x c_y, c_y n_x) + 2K(c_x n_y, n_x n_y) \\ &+\mathrm{Var}[c_y n_x] + 2K(c_y n_x, n_x n_y) + \mathrm{Var}[n_x n_y]\}. \end{aligned} \tag{13.5}$$

To evaluate this variance, its components should be derived first. The variance of n_x is given by

$$\mathrm{Var}[n_x] = \int_0^q \frac{x}{q}\varphi(x)\,\mathrm{d}x - \left[\int_0^q \frac{x}{q}\varphi(x)\,\mathrm{d}x\right]^2 = \frac{m_x}{q}\left(1 - \frac{m_x}{q}\right). \tag{13.6}$$

Likewise,

$$\mathrm{Var}[n_y] = \frac{m_y}{q}\left(1 - \frac{m_y}{q}\right). \tag{13.7}$$

As the random numbers c_x and c_y are mutually independent and do not depend on the signals x, y,the variance

$$\mathrm{Var}[c_x c_y] = 1/144. \tag{13.8}$$

On the basis of Equations (13.6) and (13.7),

$$\begin{aligned} \mathrm{Var}[c_x n_y] &= \mathrm{Var}[c_x]\mathrm{Var}[n_y] + (E[n_y])^2\mathrm{Var}[c_x] \\ &= \frac{m_y}{12q}\left(1 - \frac{m_y}{q}\right) + \frac{m_y}{12q^2} = \frac{m_y}{12q}. \end{aligned} \tag{13.9}$$

Similarly,

$$\mathrm{Var}[c_y n_x] = \frac{m_x}{12q}. \tag{13.10}$$

The variance of $n_x n_y$ is

$$\begin{aligned}\mathrm{Var}[n_x n_y] &= \int_0^q \int_0^q \int_0^x \int_0^y \frac{\mathrm{d}(q\xi_x)}{q}\frac{\mathrm{d}(q\xi_y)}{q}\varphi(x, y)\,\mathrm{d}x\,\mathrm{d}y - (E[n_x n_y])^2 \\ &= \frac{1}{q^2}a_{11} - \frac{1}{q^4}a_{11}^2. \end{aligned}\tag{13.11}$$

Now the correlation moments involved will be derived. The correlation between $c_x c_y$ and $c_x n_y$ is given by

$$\begin{aligned}K(c_x c_y, c_x n_y) &= E[c_x^2 c_y n_y] - E[c_x c_y]E[c_x n_y] \\ &= E[c_x^2]E[c_y n_y] \\ &= \frac{1}{12}\int_0^q \varphi(y)\,\mathrm{d}y \int_0^y \left(\xi_y - \frac{1}{2}\right)\frac{\mathrm{d}(q\xi_y)}{q} \\ &= \frac{1}{24q^2}(\sigma_y^2 + m_y^2 - qm_y), \end{aligned}\tag{13.12}$$

where $\varphi(y)$ and σ_y^2 are the probability density function and the variance of signal y respectively. Likewise,

$$K(c_x c_y, c_y n_x) = \frac{1}{24q^2}(\sigma_x^2 + m_x^2 - qm_x). \tag{13.13}$$

The correlation between $c_x c_y$ and $c_x n_y$ is given by

$$\begin{aligned}K(c_x c_y, n_x n_y) &= E[c_x n_x c_y n_y] \\ &= K[c_x n_y, c_y n_x] \\ &= \int_0^q \int_0^q \int_0^q \left(\xi_x - \frac{1}{2}\right)\frac{\mathrm{d}(q\xi_x)}{q}\int_0^y \left(\xi_x - \frac{1}{2}\right)\frac{\mathrm{d}(q\xi_y)}{q}\varphi(x, y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \frac{1}{4q^4}(a_{22} - qa_{12} - qa_{21} + q^2 a_{11}), \end{aligned}\tag{13.14}$$

where the moments $a_{ij} = E[x^i y^j]$, $i, j = 1, 2, \ldots$. Next, it can be shown that

$$K(c_x n_y, n_x n_y) = E\left[c_x n_x n_y^2\right] = \frac{1}{2q^3}(a_{21} - qa_{11}) \tag{13.15}$$

and

$$K(c_y n_x, n_x n_y) = \frac{1}{2q^3}(a_{12} - qa_{11}). \tag{13.16}$$

Substituting Equations (13.8) to (13.16) into Equation (13.5) gives

$$\mathrm{Var}[z] = \frac{q^4}{144} + \frac{q^2}{12}\left(m_x^2 + m_y^2 + \sigma_x^2 + \sigma_y^2\right) - a_{11}^2 + a_{22}. \tag{13.17}$$

By cancelling the first four terms of equation (13.5) and by substituting the equations of the respective variances and the correlation moments gives

$$\mathrm{Var}\left[z^{(1)}\right] = \frac{q^3}{12}(m_x + m_y) - \frac{q^2}{2}a_{11} - a_{11}^2 + \frac{q}{2}(a_{12} + a_{21}) + \frac{a_{22}}{2}. \tag{13.18}$$

Note that the variance of $z^{(2)}$ is determined by Equation (13.11).

13.1.3 Optional Approaches

To evaluate the efficiencies of the different approaches to the problem of multiplying pseudo-randomly quantized signals, the variances characterizing the corresponding estimates z, $z^{(1)}$ and $z^{(2)}$ should be compared with the minimal variance $\mathrm{Var}[xy]$, calculated on the assumption that the signal samples are obtained without quantizing errors. By definition,

$$\begin{aligned}\mathrm{Var}[xy] &= \int_0^q \int_0^q x^2 y^2 \varphi(x, y)\,\mathrm{d}x\,\mathrm{d}y - (E[xy])^2 \\ &= a_{22} - a_{11}^2. \end{aligned} \tag{13.19}$$

Assume that the probability density function of the signal $x(t) \in [-q, q]$ is normal and that the signal $y(t) = x(t + \tau)$. This normal random process is characterized by the mean value m_x and by the correlation function $K(\tau)$. Under these conditions, the a_{ij} are given by

$$\begin{aligned} a_{12} &= a_{21} = m_x^3 + m_x\sigma_x^2 + 2m_x K(\tau), \\ a_{11} &= m_x^2 + K(\tau), \\ a_{22} &= m_x^4 + 2m_x^2\sigma_x^2 + \sigma_x^4 + 4m_x^2 K(\tau) + 2K^2(\tau). \end{aligned} \tag{13.20}$$

Substituting Equation (13.20) into Equations (13.17), (13.18) and (13.11) yields

$$\begin{aligned} \mathrm{Var}[z] &= \tfrac{1}{144}q^4 + \tfrac{1}{6}q^2 m_x^2 + \tfrac{1}{6}q^2\sigma_x^2 + 2m_x^2\sigma_x^2 + \sigma_x^4 + 2_x^2 K(\tau) + K^2(\tau), \\ \mathrm{Var}\left[z^{(1)}\right] &= \tfrac{1}{6}q^3 m_x - \tfrac{1}{2}q^2 m_x^2 + q m_x^3 - \tfrac{1}{2}m_x^4 + q m_x\sigma_x^2 + m_x^2\sigma_x^2 + \tfrac{1}{2}\sigma_x^4 \\ &\quad + 2q m_x K(\tau) - \tfrac{1}{2}q^2 K(\tau), \\ \mathrm{Var}\left[z^{(2)}\right] &= q^2 m_x^2 - m_x^4 + q^2 K(\tau) - 2m_x^2 K(\tau) - K^2(\tau). \end{aligned}$$

These variances characterize different multiplication methods. The lowest variance is

$$\mathrm{Var}[x(t)x(t+\tau)] = 2m_x^2\sigma_x^2 + \sigma_x^4 + 2m_x^2K(\tau) + K^2(\tau).$$

This represents the lower boundary of the considered variances and takes place whenever the errors due to the signal quantization can be ignored. Multiplying two randomly quantized signals, carried out in the simplest possible way, naturally leads to maximal random errors.

The estimate $z^{(1)}$ of the product $\hat{x}\hat{y}$ is the best choice, because its variance exceeds the variance of the estimate z negligibly, while the multiplying process itself is considerably simpler as the multiplication of the two multidigit numbers c_x and c_y is omitted.

Equations (13.1) and (13.4) and their respective variances lead to the conclusion that the most efficient way of multiplying pseudo-randomly quantized signals is described by the following equation:

$$z^{(1)} = \hat{x}\hat{y} = q^2(c_xn_y + c_yn_x + n_xn_y) = q^2\left[\left(\xi_x - \tfrac{1}{2}\right)n_y + \left(\xi_y - \tfrac{1}{2}\right)n_x + n_xn_y\right]. \tag{13.21}$$

This multiplication scheme is very suitable for applications in the field of correlation analysis. The correlometer described in the next section is based on this scheme.

13.2 Correlation Analysis of Pseudo-randomly Quantized Signals

Suppose that a cross-correlation function $R_{xy}(t)$ of two stationary signals $x(t)$ and $y(t)$ has to be experimentally evaluated at the instants $t = mT$, $m = 0, 1, 2, \ldots$. This can be accomplished by averaging the products of the respective quantized signal sample values$\{\hat{x}(k\Theta), \hat{y}(k\Theta + mT)\}$, $k = 0, 1, 2, \ldots, N$. Then

$$R_{xy}(mT) = \frac{1}{N}\sum_{k=0}^{N-1}\hat{x}(k\Theta)\hat{y}(k\Theta + mT), \tag{13.22}$$

where Θ is a time interval whose duration exceeds the correlation interval of $x(t)$, $y(t)$.

13.2.1 Estimation Procedure

If $\hat{x}(k\Theta)$, $\hat{y}(k\Theta + mT)$ are obtained by the means of pseudo-randomized quantizers, then, as the results given in the previous section show, estimate (13.22) is unbiased and evaluation of the acceptable level of the quantizing random errors leads to choosing the number of quantizing thresholds to be used. Using two threshold levels, one for the positive and one for the negative signal values, seems to be the best choice. Thus, each of the ADCs used in both channels actually consists of two one-bit pseudo-randomized quantizers, one covering the positive half of the input signal range $[0, q]$ and the other the negative half of the range $[-q, 0]$.

Let $q\xi_k$ be the positive threshold level at the instant $t_k = k\Theta$ and $q\eta_k$ the negative threshold level, where ξ_k and η_k are random variables uniformly distributed within the ranges $[0, 1]$ and $[0, -1]$ respectively. Then, for the x channel,

$$\hat{x}_k = \begin{cases} q\left(\xi_k - \frac{1}{2}\right) + qn_k^+ & \text{for } x_k \geq 0, \\ q\left(\eta_k + \frac{1}{2}\right) + qn_k^- & \text{for } x_k < 0, \end{cases} \tag{13.23}$$

where

$$n_k^+ = \begin{cases} 1 & \text{for } q\xi_k \leq x_k, \\ 0 & \text{for } q\xi_k > x_k, \end{cases}$$

and

$$n_k^- = \begin{cases} -1 & \text{for } q\eta_k \geq x_k, \\ 0 & \text{for } q\eta_k < x_k. \end{cases}$$

13.2.2 Essential Relationships

To simplify the implementation of this quantizing scheme, the relation $\eta_k = \xi_k - 1$ can be used. Note that $n_k^- \equiv 0$ for $x_k > 0$ and $n_k^+ \equiv 0$ for $x_k < 0$. Hence Equation (13.23) can be represented by

$$\hat{x}_k = q\left(\xi_k - \tfrac{1}{2}\right) + qn_k, \tag{13.24}$$

where $n_k = n_k^+ + n_k^-$. Now Equation (13.22) can be rewritten as

$$\hat{R}_{xx}(mT) = \frac{q^2}{N}\sum_{k=1}^{N}\Big[\left(\xi_k - \tfrac{1}{2}\right)\left(\xi_{k+m} - \tfrac{1}{2}\right) + \left(\xi_k - \tfrac{1}{2}\right)n_{k+m} + \left(\xi_{k+m} - \tfrac{1}{2}\right)n_k + n_k n_{k+m}\Big]. \tag{13.25}$$

This scheme for obtaining $\hat{R}_{xy}(mT)$ corresponds to the full product z, defined by Equation (13.1). If the correlation function is to be calculated in accordance with the simplified multiplication procedures $z^{(1)}$, $z^{(2)}$, the respective equations are

$$\hat{R}_{xx}^{(1)}(mT) = \frac{q^2}{N}\sum_{k=1}^{N}\left[\left(\xi_k - \tfrac{1}{2}\right)n_{k+m} + \left(\xi_{k+m} - \tfrac{1}{2}\right)n_k + n_k n_{k+m}\right], \quad (13.26)$$

$$\hat{R}_{xx}^{(2)}(mT) = \frac{q^2}{N}\sum_{k=1}^{N} n_k n_{k+m}. \quad (13.27)$$

For the signal $x(t) \in [-q, q]$, Equations (13.17), (13.18) and (13.11) give

$$\mathrm{Var}[\hat{R}_{xx}(mT)] = \frac{q^4}{N}\left[\frac{1}{144} + \frac{\sigma_x^2}{6q^2} + \frac{1}{q^4}a_{x22}(mT) - \frac{1}{q^4}R_{xx}^2(mT)\right], \quad (13.28)$$

$$\mathrm{Var}\left[\hat{R}_{xx}^{(1)}(mT)\right] = \frac{q^4}{N}\left[\frac{m_{|x|}}{6q} - \frac{1}{2q^2}a_{|x|11}(mT) + \frac{1}{q^3}a_{|x|12}(mT) + \frac{1}{2q^4}a_{x22}(mT) - \frac{1}{q^4}R_{xx}^2(mT)\right], \quad (13.29)$$

$$\mathrm{Var}\left[\hat{R}_{xx}^{(2)}(mT)\right] = \frac{q^4}{N}\left[\frac{1}{q^2}a_{|x|11}(mT) - \frac{1}{q^4}R_{xx}^2(mT)\right], \quad (13.30)$$

where

$$m_{|x|} = E[|x(t)|] \quad a_{|x|ij} = E\left[\left|x_k^i\right|\left|x_{k+m}^j\right|\right]. \quad (13.31)$$

These variances determining the random errors of $\hat{R}_{xx}(mT)$ when the signals $x(t)$ and $y(t)$ are roughly quantized can be compared with the minimal value of this kind of variance characterizing the idealized case when the correlation function $R_{xx}(mT)$ is calculated from the true sample values of the signals obtained without any quantization errors. Then

$$\hat{R}_{xy}^{(0)}(mT) = \frac{1}{N}\sum_{k=0}^{N} x_k x_{k+m} \quad (13.32)$$

and

$$\mathrm{Var}\left[\hat{R}_{xx}^{(0)}(mT)\right] = \frac{q^4}{N}\left[\frac{1}{q^4}a_{x22}(mT) - \frac{1}{q^4}R_{xx}^2(mT)\right], \quad (13.33)$$

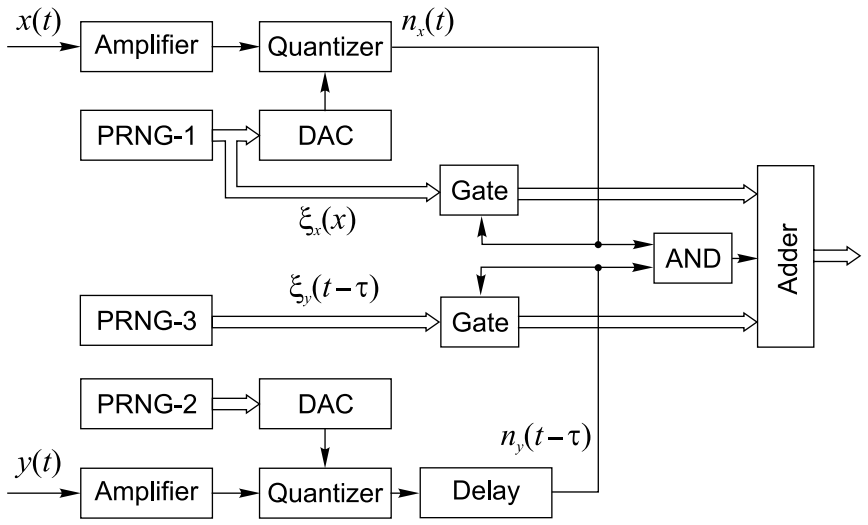


Figure 13.1 Block diagram of a correlometer processing pseudo-randomly quantized signals

where $R_{xx}(mT)$ is the autocovariance function of the signal. By comparing the random errors in $R_{xx}(mT)$ it can easily be seen that if the relative simplicity of the multiplication scheme (13.21) is taken into account, the estimate $\hat{R}_{xx}^{(1)}(mT)$ provides the best compromise between complexity and the estimation accuracy.

13.2.3 Implementation Issues

A block diagram of the electronic device for estimation of correlation functions in accordance with the considered basic scheme is given in Figure 13.1. It contains two pseudo-randomized ADCs for digitizing the signals $x(t)$ and $y(t)$, some simple logic circuits and an adder for executing the multiplication operations according to Equation (13.21).

The quantization operations in both channels are performed as defined by Equation (13.23). This means that each of the pseudo-randomized ADCs contains a pseudo-random number generator (PRNG), a DAC and a quantizer. The latter compares signals $x(t)$ and $y(t)$ with the auxiliary pseudo-random processes $\{\xi_k\}$ and $\{\eta_k\}$ at some digital time instants. The quantized signals are described by the outputs of these comparators together with the corresponding discrete values of $\{\xi_k\}$ and $\{\eta_k\}$.

To obtain the correlation function, the quantized sample values of $x(t)$ are multiplied by the quantized sample values of $y(t)$ delayed for m time intervals of

T. When quantizing is performed as described above, the values of n_y, as well as of ξ_y, should be delayed. While n_y can easily be delayed by means of two registers, many more registers are required to delay ξ_y. Fortunately, another method can be used. A special pseudo-random number generator is used for generation of the delayed sequence of ξ_y. In the scheme shown in Figure 13.1, the generator PRNG 3 is used for this purpose.

Multiplication of the quantized $x(t)$ sample values with the delayed quantized $y(t)$ sample values is performed as follows. When, for instance, $n_x = 1$ and $n_y = 0$, the gate in the y channel is open and delayed for the m time interval values T and $(\xi_y - \frac{1}{2})q$ is read from PRNG 3 and entered into the adder to be summed to the content of the mth cell of the memory, where the averaged values of the correlation function being calculated are stored. At this time moment, the AND gate is closed, so nothing else is put through to the adder.

When $n_x = 0$ and $n_y = 1$, the situation is similar, only inverse. Then the gate in the x channel is open and the output signal of PRNG 3 is passed to the adder. When $n_x = 1$ and $n_y = 1$, both of the gates and the logical element AND are open and three digital signals are passed to the adder. They are m time intervals T, $(\xi_x - \frac{1}{2})q$; the delayed value of m time intervals T, $(\xi_y - \frac{1}{2})q$; and the logical 1 from the output of the AND gate. Averaging of the multiplied values of $\hat{x}(k\Theta)$ and $\hat{y}(k\Theta + mT)$ is carried out in this way by means of the adder and the memory.

This method of obtaining the correlation functions has several advantages. Firstly, the multiplication operations required for calculating the correlation functions are replaced by the summing operations. Secondly, the randomized ADCs used in both channels can easily be designed as extremely broad band devices; to do this only two wideband strobed voltage comparators are needed for each of the ADCs. Thirdly, there are no bias errors, although extremely rough quantizing is applied. The random errors are suppressed by averaging. The application of two-threshold pseudo-randomized quantizing to correlation analysis makes it possible to use this discrete technique for direct analysis of extremely high-frequency signals without the need for any analog preprocessing. A bandwidth of several GHz could be achieved relatively easily by such correlometers.

13.3 Correlation Analysis of Pseudo-randomly Sampled Signals

Specific signal sampling affects the correlation analysis of signals even more than unusual quantizing of their sample values. In general, the motivation for

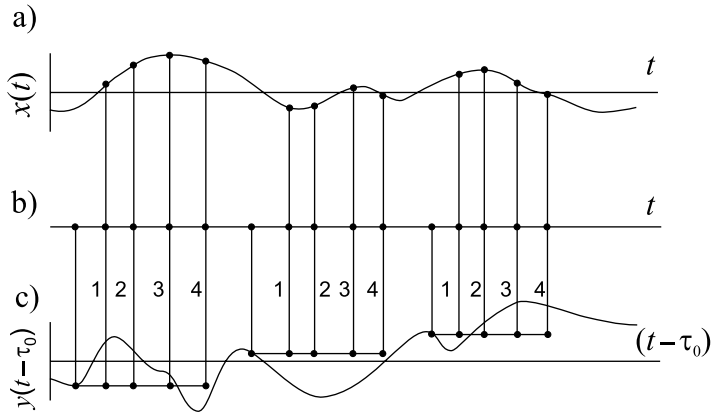


Figure 13.2 Estimation of the cross-correlation function by repetitively multiplying the indicated pairs of $x(t)$ and $y(t - \tau_0)$ sample values

using nonuniform sampling in a signal correlation analysis is widening of the frequency range within which the signals could be processed digitally, as this approach not only permits the analysis of broadband signals but is also better suited for performing correlation analysis with smaller time delay steps T.

The time diagram given in Figure 13.2 illustrates some specific points typical for correlation analysis carried out by processing pseudo-randomly sampled signals. In fact it is a simplified diagram showing what are the specifics of the correlation analysis in this case. The cross-correlation function values taken at discrete time moments are estimated by multiplying sample values of both input signals $x(t)$ and $y(t)$ as usual. Each of the signal $y(t)$ sample values are multiplied by the signal $x(t)$ sample values delayed for 1, 2, 3, . . . pseudo-random sampling intervals, as shown in Figure 13.2. In addition, for reasons explained below, the signal $y(t)$ is also delayed for a stable time interval τ_0.

This illustration shows how a cross-correlation function is estimated by taking and processing N sample values of $y(t)$. The same quantity of samples is taken from $x(t)$at each of the pseudo-randomly varied delays determined by 1, 2, 3 or 4 sampling intervals. The delays, determined by the sampling intervals, are discrete pseudo-random variables and the delay step T can be very small. This delay step is equal to the time digit δ gcharacterizing the pseudo-randomized sampling point process. A histogram of delay times, obtained by computer simulations of this kind of correlation analysis in the case where the signals are sampled according to the additive pseudo-random sampling point process, is shown in Figure 10.4. This histogram shows how many of the particular sample value products are delayed

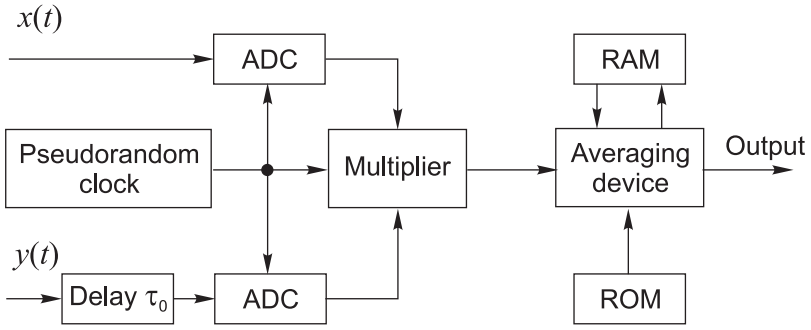


Figure 13.3 Block diagram of a correlometer based on the formation of time delays by pseudo-randomization of the input signal sampling

for any given discrete delay time interval. It also reveals that for a certain delay time interval close to zero there are no sample value products. To fill this gap, a constant time delay τ_0 is inserted into the y channel.

It can be seen from Figure 10.4 that the distribution of $x(t_k)y[t_k - (\tau + \tau_0)]$ along the time delay τ axis varies. Consequently, for proper averaging of these products, it is essential to know how many of them fall into each of the time slots. So long as the pseudo-random sampling technique is used, the functioning of such a correlometer is fully deterministic and the information needed for correct averaging can be prepared and stored in a memory (ROM (read-only memory) in Figure 13.3). Of course, in this case, when varying numbers of the products are averaged, the random estimation error is not constant for all τ values.

The diagram given in Figure 13.3 illustrates the implementation of the described approach to correlation analysis. In accordance with this scheme the signals $x(t)$ and $y(t - \tau_0)$ are pseudo-randomly sampled and quantized by both ADCs and the digitized signal values are then multiplied as shown in Figure 13.3. The obtained products are then added to the contents of the corresponding cell of the random access memory (RAM). The acquisition and processing of a predetermined number N of $y(t - \tau_0)$ samples completes a particular cycle of estimating the correlation function, and the results are stored in the RAM. To read out this correlation function, the varying coefficients stored in the read-only memory (ROM) should be normalized. These coefficients are inversely proportional to the numbers of the particular products $x(t_k)y[t_k - (\tau + \tau_0)]$ accumulated in the respective cells of the RAM.

The basic drawback of this approach to correlation analysis is related to the fact already stated in Chapter 10 that distribution of the nonuniform sampling

intervals has to guarantee that the distance between two consequent sampling instants is equal to or larger than the shortest sampling interval characterizing the used ADC. Therefore in general it is impossible to obtain signal sample values located more closely on the time axis. That is the reason why the signal y is delayed for the time interval τ_0. This limitation of scarce nonuniform sampling essentially impacts on a particular type of wideband signal processing, specifically the spectrum analysis of signals having continuous spectra based on an estimation of correlation functions.

The problem of providing for sufficiently short digital delays can be resolved on the basis of hybrid double sampling. It has already been considered that input signal digitizing according to the additive sampling scheme leads to the histogram of N_i values for the varying delay times given in Figure 10.4. This histogram, showing how large is the number N_i of averaged products $x(t_k)x(t_k + \tau_i)$ for specific delay time values within the given delay time range, changes drastically if the input signals are sampled according to the hybrid double sampling scheme. This histogram, as can be seen, for example, in Figure 10.10, does not have empty delay time slots like the ones in the histogram in Figure 10.4. Therefore, if the input signals are sampled according to the hybrid double sampling approach, then there is no need to insert an analog delay into the y input channel. Under these conditions of sampling it is possible to estimate the values of the correlation functions at discrete delay time instants within the whole considered delay time range with the area near zero delays included. In the cases of predetermined pseudo-random sampling, these empirical N_i distributions are kept in a memory as they contain information needed for correct averaging of the estimated values of the correlation functions.

13.4 Comments

The above schemes for signal correlation analysis, reflecting the specifics related to the application of pseudo-randomized signal sampling and quantizing, could be considered as a core of different systems performing signal detection, demodulation and estimation of the correlation structure of a signal or cross-correlation between signals. Essential applications of such systems are related to an estimation of signal power spectra and to cross-correlation based demodulation of transmitted signals.

No matter in which particular application area this type of system is used, it makes sense to digitize signals pseudo-randomly basically in two cases: firstly, whenever a multiplier-less structure of the system is preferable and, secondly,

when the correlation function values have to be estimated for delay time intervals much smaller than the signal sampling intervals.

Bibliography

Bilinskis, I. and Mikelsons, A. (1992) *Randomized Signal Processing*. Prentice-Hall International (UK) Ltd.

14

Signal Transforms

The idea of exploiting sampling randomization to achieve the capability of alias-free signal processing in a wide frequency range is attractive and often attracts attention. However, to achieve good results processing of nonuniformly represented signals must be carried out using special algorithms that take into account the specifics of this sampling approach. While this may seem to be obvious, it is less clear what criteria are required for evaluating the degree of matching the specifics of signal sampling and processing. This is not a trivial question. The following discussions explain to some extent why this is so. More about this is discussed in Chapter 18.

14.1 Problem of Matching Signal Processing to Sampling

These discussions will start by considering an example. Suppose that the mean power P_x of a wideband signal $x(t)$ component at frequency ω_i that exceeds the mean sampling frequency ω_s has to be estimated. The application of random sampling is clearly indicated. At first glance it seems that this task can be solved by estimating the Fourier coefficients a_i and b_i at the frequency ω_i on the basis of the often applied formulae

$$\begin{aligned} \hat{a}_i &= \frac{2}{N}\sum_{k=1}^{N} x(t_k)\cos\omega_i t_k, \\ \hat{b}_i &= \frac{2}{N}\sum_{k=1}^{N} x(t_k)\sin\omega_i t_k. \end{aligned} \tag{14.1}$$

The required estimate is then given as

$$\hat{P}_x = \frac{1}{2}\left(\hat{a}^2 + \hat{b}^2\right).$$

Digital Alias-free Signal Processing I. Bilinskis

Although these equations look like their conventional counterparts, they are in fact modified versions because the sample values of the signal and the functions cos ωt and sin ωt are taken irregularly and simultaneously at the random time instants $\{t_k\}$. This is done in an attempt to realize processing of the nonuniformly sampled signal in a proper manner. If the sampling operation is performed in accordance with the additive random point process, then, as shown in Chapter 6, the estimates are virtually unbiased and consequently they and the estimate $\hat{P}_x$ contain only random errors.

Now the question may be asked as to whether the processing algorithm in this case is matched to sampling or not. On the one hand, the answer to this question could be affirmative: to some extent this estimation method is matched to the specifics of random sampling. The absence of the bias errors might be considered as evidence showing this to be the case. On the other hand, the answer to this question could just as well be negative. The degree of match between sampling and processing could be regarded as poor, first of all because the processing algorithm considered, which was initially developed for estimating the orthogonal signal components, in this case is applied to estimating signal components that after random sampling are no longer orthogonal. However, this fact is ignored.

The mutual unorthogonality of the randomly sampled functions sin $\omega_i t_k$ and cos $\omega_i t_k$ means that the estimation of one is influenced by the other. Moreover, these two randomly sampled functions are also nonorthogonal with regard to all of the other signal components. Consequently, the so-called random estimation errors in this case are not really random at all. They are actually caused, at least partly, by the interaction of all the signal components.

This sort of cross-interference between signal components is a basic phenomenon playing a significant negative role in processing of nonuniformly sampled signals. It must not be ignored. A particular approach to digital processing of signals that helps to get rid of this disadvantage is based on unorthogonal transforms. The mathematical apparatus of unorthogonal transforms is a powerful tool for dealing with many problems arising in advanced processing of irregularly sampled signals. Although many signal processing tasks may be reduced to unorthogonal transforms, application of them is effective only under certain conditions.

When the conditions for signal processing meet the requirements of correct unorthogonal transforms, the obtained results are really good. Then the distortions of the signal processing results due to the impact of sampling irregularities are taken out completely, which is demonstrated in this chapter. In other cases when it is not applicable, adapting signal processing to the specific nonuniformities of the sampling process might be attempted, as shown in Chapter 18.

14.2 Bases of Signal Transforms

Suppose that a signal, represented by the function $x(t)$, can be given as

$$x(t) = \sum_{i=1}^{\infty} c_i \phi_i(t), \tag{14.2}$$

where $\phi_i(t)$ are functions from a system

$$\Phi = \{\phi_1(t), \phi_2(t), \ldots, \phi_i(t), \ldots\}. \tag{14.3}$$

Depending on the properties of the function system Φ, the coefficients $\{c_i\}$ of the series (14.2) are obtained in different ways. The equations establishing the relationships between $x(t)$ and the coefficients $\{c_i\}$ of the respective system Φ define the corresponding signal transform.

The usefulness of such signal transforms is obvious. First of all, they can be applied to decompose the signals into their components, whose definition will depend on *a priori* information about signal properties and signal source. Solving most signal processing tasks, such as filtering, signal enhancement and extraction from noise, spectral analysis, identification and recognition, data compression and many others, involves some kind of signal transform and it can be shown that conventional orthogonal transforms do not always provide the best results.

14.2.1 Required Properties of the Transform Bases

The function system Φ is the basis of the corresponding signal transforms. There are certain requirements that it has to satisfy:

1. Completeness. A system Φ is considered to be complete for a certain space of functions if any function $x(t)$ from this space can be represented by the series given by Equation (14.2).
2. Linear independence. All functions $\{\phi_i(t)\}$ of the system Φ should be linearly independent.

To prove this, consider the case when, for instance, $\phi_i(t)$ is linearly dependent on other functions of the corresponding system. Then

$$\phi_i(t) = \sum_{i=2}^{\infty} h_i \phi_i(t). \tag{14.4}$$

Substitution of Equation (14.4) into Equation (14.2) gives

$$x(t) = \sum_{i=2}^{\infty} (c_i + c_1 h_i)\phi_i(t),$$

which means that it is possible to express the function $x(t)$ by all other functions of the corresponding system Φ without the function $\{\phi_i(t)\}$. Hence this function, when it linearly depends on the other functions of the system Φ, is redundant.

It seems that no more than these two conditions have to be satisfied for a correct application of the signal transforms discussed here. The discussion of when and why the orthogonality of the basis functions is required follows.

14.2.2 Transforms by Means of a Finite Number of Basis Functions

According to Equation (14.2), the signal $x(t)$ is in general represented by an infinite series. In practice, the number of terms of such a series is always finite. Therefore, it is more appropriate to talk about approximating the signal. To do this, the series

$$x^*(t) = \sum_{i=1}^{m} c_i \phi_i(t) \tag{14.5}$$

has to be constructed, so that $x^*(t)$ approximates $x(t)$ sufficiently closely. The least squares approximation error can be used as a criterion for evaluating this closeness.

Now the approximation task both for analog and digital signals will be considered.

Analog Processing

The coefficients $\{c_i\}$ for the series (14.5) can be determined by solving the following minimization task:

$$\int_0^{\Theta} \left[\sum_{i=1}^{m} c_i \phi_i(t) - x(t) \right]^2 \mathrm{d}t = \min. \tag{14.6}$$

To find the minimum of the integral (14.6), all the individual derivatives of $\{c_j\}$ should be considered as being equal to zero. Then

$$2 \int_0^{\Theta} \left[\sum_{i=1}^{m} c_1 \phi_i(t) - x(t) \right] \phi_j(t)\, \mathrm{d}t = 0 \quad \text{for } j = \overline{1, m}.$$

Therefore m linear equations are obtained, which can be incorporated into the following system:

$$\left\{\sum_{i=1}^{m} c_i \int_0^{\Theta} \phi_i(t)\phi_j(t)\,\mathrm{d}t = \int_0^{\Theta} x(t)\phi_j(t)\,\mathrm{d}t\right\} \quad j = \overline{1,m}. \tag{14.7}$$

This system of equations can be represented by

$$\begin{aligned}
&\alpha_{11}c_1 + \alpha_{12}c_2 + \alpha_{13}c_3 + \cdots + \alpha_{1m}c_m = \hat{c}_1\\
&\alpha_{21}c_1 + \alpha_{22}c_2 + \alpha_{23}c_3 + \cdots + \alpha_{2m}c_m = \hat{c}_2\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
&\alpha_{m1}c_1 + \alpha_{m2}c_2 + \alpha_{m3}c_3 + \cdots + \alpha_{mm}c_m = \hat{c}_m
\end{aligned} \tag{14.8}$$

where

$$\alpha_{ji} = \int_0^{\Theta} \phi_i(t)\phi_j(t)\,\mathrm{d}t, \tag{14.9}$$

$$\hat{c}_j = \int_0^{\Theta} x(t)\phi_j(t)\,\mathrm{d}t. \tag{14.10}$$

Solving the equation system (14.8) provides the coefficients $\{c_i\}, i = \overline{1,m}$. Obviously, the number of linearly independent functions of the system $\Phi = \{\phi_1(t), \phi_2(t), \ldots, \phi_m(t)\}$ is equal to the number of linearly independent equations of system (14.8).

Digital Processing

The above minimization task in this case is given by

$$\sum_{k=1}^{N}\left[\sum_{i=1}^{m} c_i\phi_i(t_k) - x(t_k)\right]^2 \mathrm{d}t = \min, \quad j = \overline{1,m}, \tag{14.11}$$

which leads to the following equation system:

$$\left\{\sum_{i=1}^{m} c_i \sum_{k=1}^{N} \phi_i(t_k)\phi_j(t_k) = \sum_{k=1}^{N} x(t_k)\phi_j(t_k)\right\} \quad j = \overline{1,m}. \tag{14.12}$$

This system can also be represented by system (14.8), although in this case

$$\alpha_{ji} = \sum_{k=1}^{N} \phi_i(t_k)\phi_j(t_k), \tag{14.13}$$

$$\hat{c}_j = \sum_{k=1}^{N} x(t_k)\phi_j(t_k). \tag{14.14}$$

A system of digital functions $\Phi = \{\phi(t), k = \overline{1, N}\}$ cannot contain more than N linearly independent functions. To check this, assume that, contrary to this statement, the system Φ contains $(N + 1)$ linearly independent functions. Express the $(N + 1)$th function $\phi_{N+1}(t_k)$ through the other functions as

$$\begin{gathered} d_1\phi_1(t_1) + d_2\phi_2(t_1) + d_3\phi_3(t_1) + \cdots + d_N\phi_N(t_1) = \phi_{N+1}(t_1) \\ d_1\phi_1(t_2) + d_2\phi_2(t_2) + d_3\phi_3(t_2) + \cdots + d_N\phi_N(t_2) = \phi_{N+1}(t_2) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ d_1\phi_1(t_N) + d_2\phi_2(t_N) + d_3\phi_3(t_N) + \cdots + d_N\phi_N(t_N) = \phi_{N+1}(t_n) \end{gathered} \tag{14.15}$$

where $\{d_i\}$, $i = \overline{1, N}$, are unknown variables. Each of the digital functions $\phi_i(t_k)$, $k = \overline{1, N}$, can be considered as an N-dimensional vector and the matrix of the equation system (14.15) represents these vectors in such a way that the ith column defines the ith vector $\phi_i(t_k)$, $k = \overline{1, N}$. According to the assumption all N vectors are linearly independent, the determinant of system (14.15) differs from zero and this equation system has only one solution. The kth equation of this system can be written as

$$\phi_{N+1}(t_k) = \sum_{i=1}^{N} d_i\phi_i(t_k) \quad \text{for all } k = \overline{1, N},$$

which shows that the function $\phi_{N+1}(t_k)$, $k = \overline{1, N}$, represents a linear combination of the other N functions. As this contradicts the assumption, the initial statement is proved to be correct.

Thus, in the case of the discrete transforms, the number m of the functions of series (14.15) is limited by the number N of the signal sample values processed and the following unequality should be satisfied:

$$m \le N. \tag{14.16}$$

14.3 Orthogonal Transforms

The linearly independent functions of the system $\Phi = \{\phi_1(t),\ \phi_2(t), \ldots, \phi_m(t)\}$ are orthogonal if in the time interval $[0, \Theta]$ the following conditions are satisfied:

$$\int_0^{\Theta} \phi_i(t)\phi_j(t)\,\mathrm{d}t = \begin{cases} \alpha_{ii} > 0 & \text{for } i = j, \\ 0 & \text{for } i \ne j. \end{cases} \tag{14.17}$$

Denote $\sqrt{\alpha_{ii}}$ by $\|\phi_i(t)\|$. This parameter of the function $\phi_i(t)$ is considered to be its norm. If for all i

$$\sqrt{\alpha_{ii}} = \|\phi_i(t)\| = 1,$$

the corresponding system of functions is described as orthonormalized. In the case of digitized signals, conditions (14.17) become

$$\sum_{k=1}^{N} \phi_i(t_k)\phi_j(t_k)\,\mathrm{d}t = \begin{cases} \alpha_{ii} > 0 & \text{for } i = j, \\ 0 & \text{for } i \neq j. \end{cases} \tag{14.18}$$

Now consider solving the equation system (14.18) in the case when the basis functions are orthogonal. Then

$$\alpha_{ji} = \begin{cases} \alpha_{ii} = \|\phi_i(t)\|^2 > 0 & \text{for } j = 1, \\ 0 & \text{for } j \neq i, \end{cases} \tag{14.19}$$

and the equation system in question is reduced to the following system of equalities:

$$\begin{aligned} c_1 &= \frac{1}{\|\phi_1(t)\|^2}\hat{c}_1 \\ c_2 &= \frac{1}{\|\phi_2(t)\|^2}\hat{c}_2 \\ &\cdots\cdots\cdots\cdots \\ c_m &= \frac{1}{\|\phi_m(t)\|^2}\hat{c}_m. \end{aligned} \tag{14.20}$$

Thus we come to the conclusion that the orthogonal basis functions are a particular case of linearly independent basis functions and their application considerably simplifies the corresponding signal transforms. The coefficients $\{c_i\}$ calculated in the course of the orthogonal transforms can be obtained directly without solving the equation system under consideration.

14.3.1 Analog Processing

It follows from Equations (14.10) and (14.20) that the coefficients $\{c_i\}$ in the case of the orthogonal transforms are given as

$$c_i = \frac{1}{\|\phi_i(t)\|^2}\int_0^{\Theta} x(t)\phi_i(t)\,\mathrm{d}t, \tag{14.21}$$

where the norm

$$\|\phi_i(t)\| = \left[\int_0^{\Theta} \phi_i^2(t)\right]^{1/2} \mathrm{d}t, \qquad i = \overline{1, m}.$$

Note that in this case the number m of the approximation terms is unlimited, as long as the orthogonality conditions are satisfied.

14.3.2 Digital Processing

Similarly, it follows from Equations (14.14) and (14.20) that

$$c_i = \frac{1}{\|\phi_i(t)\|^2} \sum_{k=1}^{N} x(t)\phi_i(t_k), \tag{14.22}$$

where

$$\|\phi_i(t_k)\|^2 = \sum_{k=1}^{N} \phi_i^2(t_k), \quad i = \overline{1, m}.$$

The theory of discrete orthogonal transforms is well developed and their properties, with emphasis on the positive aspects, are well known. It can even be said that modern signal processing techniques are to a large extent founded on the concepts of discrete orthogonal transforms. They are so popular, and so widely discussed in numerous publications, that it is not necessary to discuss them here in detail. This category of transforms is considered first of all for the purpose of a survey to indicate the role of orthogonal transforms in the general scheme of signal transforms as a whole.

There can be no doubt that discrete orthogonal transforms have significant advantages. They can be carried out by performing relatively few multiplication and summing operations, and they are well suited to the application of so-called fast algorithms. On the other hand, discrete orthogonal transforms also have considerable drawbacks. The following are a few of these:

1. The orthogonal transforms of signals actually represent their approximations within a given time interval, and these approximations are not applicable for extrapolating or predicting the signals outside this interval.
2. When they are used, it is difficult, or sometimes impossible, to gain from *a priori* information about the input signal source or structure.
3. Orthogonal transforms do not allow signals to be decomposed into their true components if the latter are mutually unorthogonal.
4. This category of transforms is not applicable to irregularly sampled signals.

The last disadvantage calls for comment. Sometimes the fact that a signal has been sampled randomly or irregularly can be ignored and the orthogonal

transforms can be applied. In this case additional estimation errors will result from sampling irregularities. If this is acceptable, irregularly sampled signals can be transformed in this way.

14.4 Discrete Unorthogonal Transforms

A signal transform is considered unorthogonal if the system of functions $\Phi = \{\phi_i(t_k), k = \overline{1, N}\}$, applied to perform the transform, is unorthogonal. It is agreed that the system Φ does not contain any functions identically equal to zero. Under this condition, $\|\phi_i(t)\|^2 > 0$ for all $i = \overline{1, m}$.

A few typical signal processing applications will now be mentioned where the signals need to be transformed on the basis of unorthogonal transforms:

1. Signal transforms carried out by means of a system of analog functions, which are originally orthogonal and become unorthogonal in the course of digitizing, as when the sampling applied is irregular. As an example to illustrate this case, refer to Chapter 15, where the discrete Fourier transforms of randomly sampled signals are discussed.
2. Short-time periodic signal spectrum analysis, when the signals have to be transformed under the condition that the observation time interval $[0, \Theta]$ is shorter than the given period of the signal. In this case the frequencies of the true signal components are known but the system of the functions $\{\phi(t)\}$, chosen correspondingly, is unorthogonal.
3. Spectrum analysis of quasi-periodic signals, which are actually nonstationary with a slowly varying period.
4. Spectrum analysis of signals containing components at frequencies irregularly spaced along the frequency axis.
5. Decomposing signals into true components, which are mutually unorthogonal.

Even this incomplete list of unorthogonal transform applications shows that the problems approached in this way are significant. In this book, the most interest lies in the first application.

Now consider a system of functions $\Phi = \{\phi_i(t_k), k = \overline{1, N}\},\ i = \overline{1,\ m}$, which is unorthogonal. Some nonzero coefficients α_{ij}, $j \neq i$, which are not placed on the diagonal of the corresponding matrix defined by the equation system (14.12), definitely exist. For this reason, this equation system cannot be reduced to the system of equalities (14.20). Hence, the coefficients $\{c_i\}, i = \overline{1, m}$, have to be calculated on the basis of the equation system (14.12) and the condition (14.16) should be met.

The equation system (14.12) can be represented in the form of the following matrix equation:

$$\mathbf{A}.\mathbf{C} = \hat{\mathbf{C}}, \tag{14.23}$$

where the matrix

$$\mathbf{A} = \begin{Bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} & \cdots & \alpha_{1m} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} & \cdots & \alpha_{2m} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \alpha_{m1} & \alpha_{m2} & \alpha_{m3} & \cdots & \alpha_{mn} \end{Bmatrix} \tag{14.24}$$

is square and contains $(m \times m)$ elements. It follows from definition (14.13) of its elements α_{ji} that this matrix is also symmetric. The vectors of the coefficients $\{c_i\}$, $i = \overline{1, m}$, and their estimates $\{c_i\}$ are given by the respective columns

$$\mathbf{C} = \begin{bmatrix} c \\ c_2 \\ \cdots \\ c_m \end{bmatrix}, \qquad \hat{\mathbf{C}} = \begin{bmatrix} \hat{c} \\ \hat{c}_2 \\ \cdots \\ \hat{c}_m \end{bmatrix},$$

where the estimates $\{\hat{c}_i\}$ are defined by Equation (14.14). It follows from Equation (14.23) that

$$\mathbf{C} = \mathbf{A}^{-1}\hat{\mathbf{C}}, \tag{14.25}$$

where $\mathbf{A}^{-1}$ is the inverse of matrix $\mathbf{A}$.

It can be seen from Equation (14.14) that each estimate $\{\hat{c}_j\}$ is obtained from a scalar multiplication of the vectors $\{x(t_k)\}$ and $\{\phi_j(t_k)\}$, $k = \overline{1, N}$. To obtain the coefficients themselves, the vector $\mathbf{C}$ according to Equation (14.25) should be multiplied by the matrix $\mathbf{A}^{-1}$. Hence,

$$c_j = \sum_{i=1}^{m} \alpha_{ji}^{*} \hat{c}_j, \quad j = \overline{1,\ m}, \tag{14.26}$$

where α_{ji}^{*} are elements of the matrix $\mathbf{A}^{-1}$.

Thus, to carry out such unorthogonal transforms, all the elements of the matrix $\mathbf{A}$ should be determined while the estimates $\{\hat{c}_j\}$ are being calculated; then the matrix $\mathbf{A}$ should be inverted in order to obtain the coefficients. As the inversion of such matrices is time consuming, this unorthogonal transform procedure is not suitable for real-time applications. Fortunately, it is often possible to compose the matrix $\mathbf{A}$ and invert it beforehand. Then the coefficients $\{\hat{c}_j\}$ can be calculated much more quickly.

Such an approach can be applied when the unorthogonal transforms are performed in order to eliminate the errors due to sampling irregularities. If the sampling point process $\{t_k\}$ is generated pseudo-randomly, all the functions of the system Φ are fully defined and the matrix $\mathbf{A}$ and its inverse can be obtained beforehand. Although in this case there are no problems and this method of performing unorthogonal transforms can be widely applied, it can still be improved, as the next section will show.

14.5 Conversion of Unorthogonal Transforms

Suppose that a signal is sampled pseudo-randomly and that the sampling point process $\{t_k\}$ is fixed, Denote the values $\phi_j(t_k)$ by ϕ_{ik}. Then the system Φ of the given values can be written as

$$\mathbf{\Phi} = \begin{bmatrix} \phi_{11} & \phi_{12} & \phi_{13} & \phi_{1N} \\ \phi_{21} & \phi_{22} & \phi_{23} & \phi_{2N} \\ \cdots & \cdots & \cdots & \cdots \\ \phi_{m1} & \phi_{m2} & \phi_{m3} & \phi_{mN} \end{bmatrix}. \tag{14.27}$$

If the digital signal is represented as

$$\mathbf{X} = \begin{bmatrix} x(t) \\ x(t_2) \\ \cdots \\ x(t_N) \end{bmatrix},$$

it follows from Equation (14.14) that

$$\hat{\mathbf{C}} = \mathbf{\Phi X}. \tag{14.28}$$

Substitution of expression (14.28) into Equation (14.23) yields

$$\mathbf{CA} = \mathbf{\Phi X}.$$

Multiply both sides of this equation by the inverse matrix $\mathbf{A}$. Then

$$\mathbf{C} = \mathbf{A}^{-1}\mathbf{\Phi X} = \mathbf{\Psi X}, \tag{14.29}$$

where

$$\mathbf{\Psi} = \mathbf{A}^{-1}\mathbf{\Phi}. \tag{14.30}$$

It follows from Equation (14.30) that

$$\mathbf{A} = \mathbf{\Phi}\mathbf{\Phi}^{\mathrm{T}}, \tag{14.31}$$

where $\mathbf{\Phi}^{\mathrm{T}}$ is the transpose of matrix $\mathbf{\Phi}$. Hence

$$\mathbf{A}^{-1} = (\mathbf{\Phi}\mathbf{\Phi}^{\mathrm{T}})^{-1}. \tag{14.32}$$

Substitution of this expression into Equation (14.30) yields

$$\mathbf{\Psi} = (\mathbf{\Phi}\mathbf{\Phi}^{\mathrm{T}})^{-1}\mathbf{\Phi}. \tag{14.33}$$

As $\mathbf{A}$ is a square $m \times m$ matrix, the inverse matrix $\mathbf{A}^{-1}$ is also square and also contains $m \times m$ elements.

The matrix $\mathbf{\Phi}$, as can be seen from its definition (14.27), is a rectangular $m \times N$ matrix. Hence the matrix $\mathbf{\Psi}$ is also rectangular and contains $m \times N$ elements. This matrix can be given as

$$\mathbf{\Psi} = \begin{bmatrix} \psi_1(t_1) & \psi_1(t_2) & \psi_1(t_3) & \cdots & \psi_1(t_N) \\ \psi_2(t_1) & \psi_2(t_2) & \psi_2(t_3) & \cdots & \psi_2(t_N) \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \psi_m(t_1) & \psi_m(t_2) & \psi_m(t_3) & \cdots & \psi_m(t_N) \end{bmatrix}. \tag{14.34}$$

It follows from Equations (14.29) and (14.33) that

$$c_i = \sum_{k=1}^{N} x(t_k)\psi_i(t_k). \tag{14.35}$$

Equations (14.33) and (14.35) lead to the following conclusion. If the initial system $\mathbf{\Phi}$ of the unorthogonal basis functions is replaced by the equivalent system $\mathbf{\Psi}$, defined by Equation (14.33), the unorthogonal transforms can be calculated directly.

This method of performing the unorthogonal transforms is especially efficient under conditions where the functions $\{\psi(t_k)\}$ can be calculated beforehand, so that the corresponding data can be stored in a memory. When the unorthogonal transforms are used for excluding the errors caused by sampling irregularities, these functions can be determined only for given particular realizations of the sampling point processes and, for this reason, it is convenient to use pseudo-random sampling.

Note that when the unorthogonal transforms are performed according to this approach, this method can be implemented by the same electronic devices that are commonly used for the DFT. The only difference is that when they are applied for calculating the unorthogonal transforms, the discrete values of $\{\sin \omega_i t_k,\ \cos \omega_i t_k\}$, which are kept stored in a memory unit, should be replaced by the sample values of the functions $\{\psi(t_k)\}$. Even the operational characteristics, like the number of operations per second and the accuracy of the results

obtained, will be the same in both cases. The only disadvantage of performing the unorthogonal transforms in this way is that there are no fast algorithms for calculating them. At least, they have not yet been discovered.

Bibliography

Bilinskis, I. and Mikelsons, A. (1992) *Randomized Signal Processing*. Prentice-Hall International (UK) Ltd.

15

DFT of Nonuniformly Sampled Signals

In the classical case of processing periodically sampled signals on the basis of the DFT, the results of the performed DFT reflect the structure of the respective signals in the frequency domain. In other words, the DFT of periodically sampled signals leads to signal decomposition and to obtaining their spectra. Therefore it is normal to expect that when the DFT of nonuniformly sampled signals are calculated the results will also represent the spectra of these signals. However, these expectations are typically not fulfilled. As soon as the sampling procedure is randomized, the DFT of the respective signals become strongly sampling-dependent. As explained in the previous chapter, the basis functions for the DFT under the conditions of nonuniform sampling become unorthogonal. If DFT could be performed as unorthogonal transforms then the preference should be given to such an approach. However, that is not always possible. Direct calculations of DFT have to be undertaken while clearly realizing that these transforms will not complete the process of signal decomposition into their components. Estimating the Fourier coefficients then actually leads to acquiring intermediate signal processing results containing valuable information. Therefore the outcome of the DFT under these conditions should not always be automatically regarded as spectrograms of the respective signals. Some comments on this are given in Section 15.1. To obtain spectrograms showing the signal components in the frequency domain, this information has to be additionally processed taking into account the mutual unorthogonality of the basis functions leading to cross-interference between the signal components being estimated. It is not a trivial matter. A lot of effort has been spent on research in this area and gradually several

Digital Alias-free Signal Processing I. Bilinskis

approaches to processing transformed nonuniformly sampled signals have been developed. Some of the methods and techniques used for this type of signal processing and their potential are discussed in this chapter.

It is worth paying a lot of attention to calculating raw DFT of nonuniformly sampled signals, as the estimates obtained in this way play a vital role in the subsequent digital processing of the transformed signals. Such DFT support all kinds of digital signal filtering and other signal processing procedures, including signal waveform reconstruction, and the quality of the final signal processing results often depends on the performance of the involved DFT.

15.1 Problems Related to Sampling Irregularities

To achieve the capability of processing signals in a broad frequency range not limited by half of the used sampling rate, sampling must be nonuniform. However, once the signal digitizing is based on nonuniform sampling, processing of the signals then has to be organized with the specifics of this technique taken into account. Apparently it is essential to obtain a clear insight into the details of processing this type of digital signal.

15.1.1 Alternative Approaches to DFT

General considerations of this have been given in Chapter 3. If the sampling point process satisfies condition (3.13), then it follows from Equation (3.14) that

$$E\left[\lim_{N\Rightarrow\infty}\hat{a}_i\right]=a_i \qquad \text{and} \qquad E\left[\lim_{N\Rightarrow\infty}\hat{b}_i\right]=b_i.$$

This means that asymptotically there should be no bias error and the statistical estimation errors should tend to c^2/N (c is the amplitude of the signal component being estimated). Therefore, for large N, interval systematic bias and statistical error at least should be small. That would be so if the side effect caused by the nonorthogonality of the basis functions and the cross-interference effect due to it did not exist. This effect certainly cannot be ignored. If these considerations are taken into account, the following randomized estimation schemes can be suggested:

1. The signal $x(t)$ is sampled pseudo-randomly at predetermined time instants $\{t_k\}$. The values of $\cos(2\pi f_i t_k)$ and $\sin(2\pi f_i t_k)$ are calculated beforehand, stored in a memory and then later used for estimation. To reduce estimation errors to an acceptable level by averaging, a relatively large data block is processed. The errors due to the cross-interference remain, fact that has to be taken into account at further processing of the obtained raw estimates of the Fourier coefficients.

2. The signal $x(t)$ is sampled pseudo-randomly at predetermined instants $\{t_k\}$ and this timing information is exploited not only when calculating the corresponding set $\{\sin(2\pi f_i t_k), \cos(2\pi f_i t_k)\}$ but also to compensate for errors due to the sampling irregularities.

The first approach is used most often and leads to satisfactory results if the DFT are carried out to obtain raw Fourier coefficient estimates used in the process of further calculations. The second estimation scheme, although much more complicated than the first one, is in principle more advanced. The mechanism of error reduction then is not averaging, and the number N of the sample values to be processed depends only on the complexity of the signal $x(t)$ being processed, i.e. on the quantity of its components. Therefore it is applicable under more dynamic conditions for short-time estimations. Several methods for the DFT that belong to this category are discussed in this chapter in addition to the unorthogonal transform approach described in the previous chapter. However, before considering them, two possible techniques for estimating the Fourier coefficients are compared. One of the alternative techniques is based on direct calculations of the DFT and the second possible approach can be realized in accordance with the best-fitting procedure.

15.1.2 Best-fitting Procedure Versus Direct DFT

Consider an arbitrary frequency f and the following row of frequencies:

$$f,\ f_s - f,\ f_s + f,\ \ldots,\ nf_s - f,\ nf_s + f. \tag{15.1}$$

Apparently in the case where signals are sampled periodically these frequencies overlap. Suppose a monoharmonic signal with an amplitude equal to 1 and the frequency f_i belonging to row (15.1) is randomly sampled at N sampling instants $\{t_k\}$, so that the following digital signal is obtained:

$$x(t_k) = \sin(2\pi f_i t_k + \varphi_i), \quad k = \overline{1, N}.$$

As the sampling operation is randomized, the aliasing effect has to be eliminated or at least suppressed. The graphical interpretation of this effect is given in Figure 15.1. The sample values of $x(t_k)$ are shown as points. To find out how randomization of sampling helps to suppress aliasing, an attempt will be made to draw some other sinusoids with frequencies from row (15.1) through the indicated set of points. The outcome of such a best-fit operation strongly depends on the mode of applied sampling. The least mean square errors of drawing a sine wave at one of the frequencies f_j from this row through the points $\{x(t_k)\}$ will be equal to zero only in the case of periodic sampling, as shown in Figure 15.1(a). Even

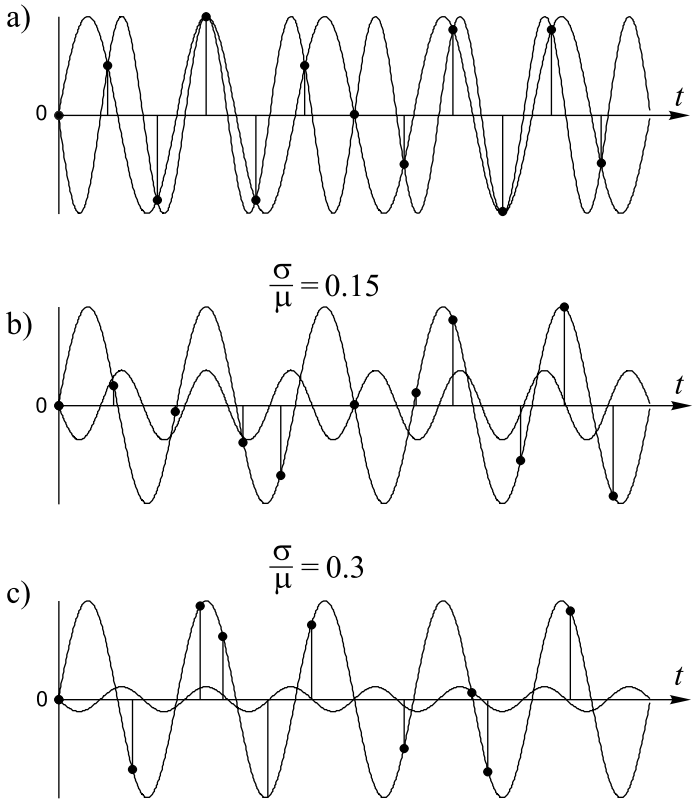


Figure 15.1 Suppression of an aliasing signal component in the case where the signal has been sampled nonuniformly

a slight randomization of sampling will magnify this error considerably. With an increasing degree of sampling randomization, the error in drawing a sine wave at any of the frequencies (15.1) through the sample points tends to become more and more significant. This is illustrated by Figures 15.1(b) and (c). They show that at larger values of σ/μ the suppression of the alias is more noticeable.

The effectiveness of the suppression of aliasing may be characterized by the ratio $\gamma_{ji} = P_j/P_i$, where P_i and P_j are the mean powers of the sinusoids at the frequencies f_i and f_j respectively. The procedure of drawing a sinusoid at frequency f_j through the sample points $\{x(t_k)\}$ with the least square error should be performed according to the following condition:

$$\sum_{k=1}^{N} [\tilde{a}_j \cos 2\pi f_j t_k + \tilde{b}_j \sin 2\pi f_j t_k - \sin(2\pi f_i t_k + \varphi_i)]^2 = \min . \quad (15.2)$$

Estimates $\tilde{a}_j$ and $\tilde{b}_j$ of the Fourier coefficients characterizing the sinusoid at frequency f_j drawn through the sample points with the least square error are obtained by solving the system of equations corresponding to Equation (15.2):

$$\begin{aligned} \tilde{a}_j &= \frac{\hat{a}_{ji}\alpha_s - \hat{b}_{ji}\alpha_{cs}}{\alpha_c\alpha_s - \alpha_{cs}^2}, \\ \tilde{b}_j &= \frac{\hat{b}_{ji}\alpha_c - \hat{a}_{ji}\alpha_{cs}}{\alpha_c\alpha_s - \alpha_{cs}^2}, \end{aligned} \tag{15.3}$$

where

$$\begin{aligned} \hat{a}_{ji} &= \frac{2}{N}\sum_{k=1}^{N} \sin(2\pi f_i t_k + \varphi_i)\cos 2\pi f_j t_k, \\ \hat{b}_{ji} &= \frac{2}{N}\sum_{k=1}^{N} \sin(2\pi f_i t_k + \varphi_i)\sin 2\pi f_j t_k \end{aligned} \tag{15.4}$$

and

$$\begin{aligned} \alpha_c &= \frac{2}{N}\sum_{k=1}^{N} \cos^2 2\pi f_j t_k, \\ \alpha_s &= \frac{2}{N}\sum_{k=1}^{N} \sin^2 2\pi f_j t_k, \\ \alpha_{cs} &= \frac{2}{N}\sum_{k=1}^{N} \sin 2\pi f_j t_k \cos 2\pi f_j t_k. \end{aligned} \tag{15.5}$$

The solution to the minimization task allows the degree of aliasing γ_{ji} to be defined as follows:

$$\gamma_{ji} = \tilde{a}_j^2 + \tilde{b}_j^2. \tag{15.6}$$

Now consider the notation in Equation (15.5). Obviously, $\alpha_c \cong 1$ and interval $\alpha_{cs} \cong 0$. This suggests that aliasing can also be evaluated by the following approximate equality:

$$\gamma_{ji} = \hat{a}_{ji}^2 + \hat{b}_{ji}^2. \tag{15.7}$$

Substitution of Equation (15.7) for Equation (15.6) is very desirable, because it considerably simplifies further analysis. The errors arising in this case were evaluated by computer simulations and they are small enough to justify this substitution. Therefore there is no need to use the more correct but computationally significantly more complicated best-fitting approach to estimate the Fourier

coefficients. Direct calculations performed on the basis of Equation (14.1) provide good enough results.

Taking into account the random nature of sampling, aliasing should actually be characterized by the expected value:

$$E[\gamma_{ji}] \cong E\left[\hat{a}_{ji}^2\right] + E\left[\hat{b}_{ji}^2\right]. \tag{15.8}$$

15.1.3 Sample Values Partly Fitting to Any Frequency

This analysis confirms the already mentioned (see Chapter 9) conclusion that elimination of aliases cannot be achieved simply by direct substitution of the periodic sampling by randomization. While intensification of sampling randomization helps to reduce the amplitude of the aliasing frequency drawn through the sampling points of the true signal, the amplitude of the aliasing sinusoid, estimated on the basis of Equation (15.2), will not be zero even when there are no aliasing frequencies present. The least squares error of drawing an aliasing sinusoid through the sampling points of the true signal component depends on the specific irregularities of the involved sampling point process and the power of other signal components. Therefore randomization of sampling provides the necessary preconditions for effective elimination of aliasing. More elaborate specific anti-aliasing methods are needed and should be used to achieve complete elimination of the aliases. The interpretation suggested in Chapter 9 of aliasing processes taking place when processing randomly sampled signals called 'fuzzy aliasing' provides the key to resolution of this problem.

So far the discussion in this section has been focused on the anti-aliasing issue. Although the above mathematical expressions were derived for evaluation of inadequate suppression of aliases, they are also applicable for an analytical description of the best-fit operation in those cases where an attempt has been made to draw a sinusoid at any arbitrary frequency f_j through a set of sample value points belonging to another sinusoid at a different frequency f_i according to Equation (15.2). It is irrelevant that both of these frequencies belong to the row (15.1). The estimates of the Fourier coefficients defining the amplitude of this sinusoid at frequency f_j depends on the parameters of other sinusoids at frequency f_i and on the specific nonuniformities of the involved sampling point process. Actually a set of signal sample values, obtained at nonuniform sampling of a signal, partly fits to any frequency. It is emphasized that the best-fitting procedure as well as direct calculations of the Fourier coefficient estimates, working well in cases where signals are sampled periodically, do not provide sufficiently accurate representation of signals in the frequency domain whenever signals have been sampled nonuniformly.

Actually the obtained DFT results then represent simply some kind of raw estimates of the Fourier coefficients. Nevertheless, they are often needed and have to be calculated at some stage of the complete signal processing process.

The given relationships are relevant also to the cross-interference effect related to the nonorthogonality of the nonuniformly sampled discrete basis functions.

15.2 Cross-interference Corrupting DFT

Deliberate randomization of sampling, while crucial for achieving the capability of alias-free processing of signals digitally in a wide frequency range, also leads to some problems. In particular, the basis functions for the DFT under conditions of nonuniform sampling become nonorthogonal, which leads to cross-interference between signal components. The impact of this cross-interference is observed as signal parameter estimation errors. They are both sampling and signal-dependent.

To find exactly how the cross-interference due to sampling nonuniformities impact on the process of the DFT, computer simulations of the DFT, performed under conditions of pseudo-randomized additive sampling, could be carried out. It is convenient to start this experiment by using at first a single tone variable frequency signal and a DFT filter tuned to frequency f_1 for estimating the signal parameters at this frequency, specifically the Fourier coefficients, amplitude and phase. If the signal is sampled periodically, the performance of this filter would be close to ideal. Figure 15.2 illustrates the results obtained in this case. At the signal frequency $f_i = f_1$ the calculated values of the Fourier coefficient estimates (for growing amplitude of the signal) are close to their true values (as shown by the dashed line 1) and these estimates are near to zero for signal frequencies deviating from f_1.

Now suppose that a single tone signal $x(t)$ at frequency f_1 is sampled nonuniformly and the Fourier coefficient a_1 is estimated under the conditions where the signal phase is equal to $\pi/2$. Then $b_1 = 0$ and the value of a_1 is estimated as being equal to the signal amplitude if there are no estimation errors (line 1). However, under the conditions of this experiment, the sampling process is noticeably irregular, which leads to these errors. As shown in Figure 15.2, this error increases proportionally to the growing true value of a_1 (line 2). In this particular case, $\hat{a}_1 = a_1 (A_1 C_1)$, where the coefficient $A_1 C_1$ is a constant indicating the strength of the impact of the sampling imperfections on the estimation of the coefficient a_1.

Actually the behaviour of the filter is even more complicated. There are also noticeable estimation errors caused by nonuniform sampling of other signal components and these errors depend on the power and frequencies of these signal

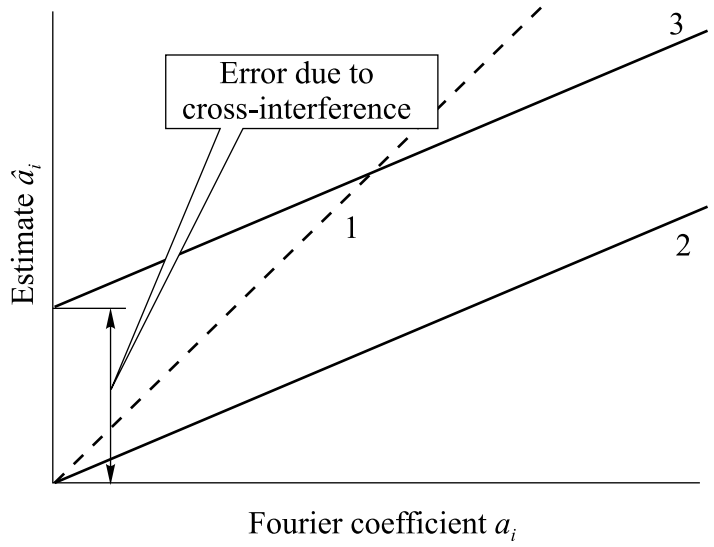


Figure 15.2 Illustration of the impact sampling irregularities on the errors when estimating the Fourier coefficients

components. In fact, the irregularities of the sampling point stream leads to cross-interference between the signal components. This phenomenon can be observed if filtering of a nonuniformly sampled signal with two or more components is studied. It can be seen that each signal component impacts on the parameter estimation of all other signal components and that this impact directly depends on the specific nonuniformities present in the involved sampling process. To observe this phenomenon, the experiment should be repeated. Exactly the same sampling point process should be used for the signal sampling and a sine function at frequency f_m should be added to the previously used cosine function with gradually increased amplitude. The results obtained in this case are displayed as line 3 in Figure 15.2. It can be seen that the presence of the irregularly sampled component at frequency f_m adds to the estimation error. However, the value of the coefficient A_1C_1 remains the same; it is not changed by the second component. Therefore it seems that in this second case $\hat{a} = a\,(A_1C_1) + b_m(A_1S_m)$, where the coefficient A_1S_m indicates the strength of the impact of the sinusoidal signal component at frequency f_m on the estimation of the coefficient a_1 when the signal is sampled in that particular way. The impact of the specific nonuniformities of the sampling point process can be characterized by constant coefficients. Thus it seems that these coefficients characterizing the cross-interference between the

signal components might be introduced as follows:

$$\begin{aligned}\hat{a}_i &= \sum_{m=0}^{M-1} [a_m(A_i C_m) + b_m(A_i S_m)], \quad i = \overline{0, M-1},\\ \hat{b}_i &= \sum_{m=0}^{M-1} [a_m(B_i C_m) + b_m(B_i S_m)], \quad i = \overline{0, M-1}.\end{aligned} \tag{15.9}$$

These coefficients are actually the weights of the errors introduced by nonuniform sampling of the signal sine (or cosine) component at frequency f_m that corrupts the estimation of a Fourier coefficient a_i (or b_i) at frequency f_i. They are derived mathematically in Chapter 18. More detailed studies of this subject reveal that the description of the experiment given above reflects the involved relationships only approximately. A system of equations has to be composed and used for a more accurate description, as shown in Chapter 18. The usefulness of the introduced interference coefficients is demonstrated there by showing how this concept could be exploited for adapting alias-free processing of signals to the sampling nonuniformities.

15.3 Exploitation of FFT

Performing the DFT requires a lot of calculations. To reduce the number of involved multiplication operations, fast algorithms are widely used, especially the FFT. However, the existing fast algorithms have been developed on the assumption that the intervals between signal sample values are constant. Therefore, in general, they cannot also be directly used for processing nonuniformly sampled signals. Nevertheless, if the nonuniform sampling is performed as a pseudo-randomized procedure, under certain conditions it is possible to exploit the FFT for reduction of the computational complexity of the DFT, and that could be done relatively often. Different approaches to this problem have to be used to handle digital signals formed on the basis of either directly or indirectly randomized sampling discussed in Chapters 6 and 7 respectively.

15.3.1 Application of FFT for Processing Nonuniformly Sampled Signals

To process the directly randomized nonuniformly sampled signals by using algorithms developed for processing periodically amplitude-sampled signals, the so-called zero padding method is typically used. According to this method, zeroes

are inserted into the nonuniform signal sample value sequence at those periodically repeating time instants where there are no signal sample values, so that the sampling process is transformed into a periodic one. However, zero padding is a crude method and its application leads to the introduction of substantial errors. Indeed, each zero replaces a signal sample value when this technique is used, which means that each substitution of this kind introduces a particular instantaneous error equal to the respective missing sample value. If there is a large quantity of this kind of error then the spectrograms obtained are naturally of a poor quality. Whether it is acceptable or not depends on the proportion between the sampling slots filled by the signal sample values and the empty slots marked as zeroes. The achievable performance is illustrated in Figure 15.3 where spectrograms obtained at various sample value/zero ratios are displayed.

It can be seen that the zero padding procedure and application of the FFT leads to a more or less significant distortion of the spectrograms depending on the specific sample value/zero ratios. In the cases of Figures 15.3(b) and (c) half and two-thirds of the sample values are randomly replaced by zeroes respectively. These spectrograms can be compared with the true signal spectrogram obtained under conditions of correct periodic sampling and are displayed in Figure 15.3(a). Obviously zero padding has led to the introduction of substantial errors.

The idea of using the FFT to obtain estimates of the Fourier coefficients is of course attractive, as application of this fast algorithm drastically reduces the amount of calculations. However, using this approach evidently leads to large errors in estimating the Fourier coefficients. The achievable accuracy of these estimates is typically not acceptable. Therefore this procedure, if and when used, should be regarded as a preliminary calculation providing only raw intermediate results. The degree of its usefulness depends on the specific algorithm used for signal analysis.

The algorithm for signal spectrum analysis and waveform reconstruction considered in Chapter 20 confirms the fact that such usage of the FFT makes sense and could lead to good results. This algorithm is focused on the realization of an obvious idea that *a priori* information should be used as often as possible. To achieve that, the initially inserted zeroes at some processing stage should be replaced by approximate estimates of the signal sample values roughly estimated as a result of the FFT. It is shown that this leads to a dramatic reduction in the mentioned errors. Even better results are obtained when an iterative procedure of spectral analysis and waveform reconstruction is used. In is case the missing uniform sample values are substituted first by zeroes, then estimated signal values are inserted in those places and after that even more accurately estimated signal

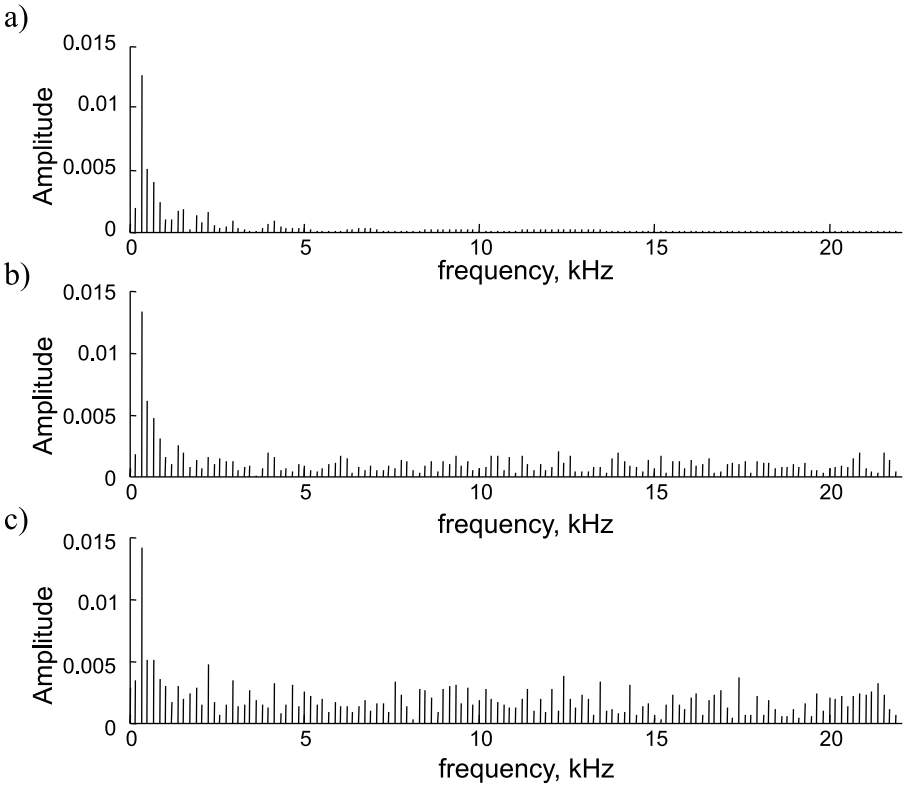


Figure 15.3 Spectrograms obtained by using the FFT in the case where a signal is sampled pseudo-randomly and zeroes are inserted to regularize the sequence of the sample values

sample values are used. That leads to a significant improvement in the estimation accuracy. Such an approach makes it possible to rationalize signal processing on the basis of the FFT.

This specific example, considered in more detail in Chapter 20, demonstrates the fact that under certain conditions it is possible to use already existing fast algorithms, first of all the FFT, for an alias-free estimation of nonuniformly sampled wideband signal parameters in the frequency domain and for subsequent waveform reconstruction carried out on the basis of the inverse FFT. The fact is noteworthy as it suggests that probably there are also some other possibilities still not discovered of using fast algorithms that are effectively applicable for processing nonuniformly sampled signals.

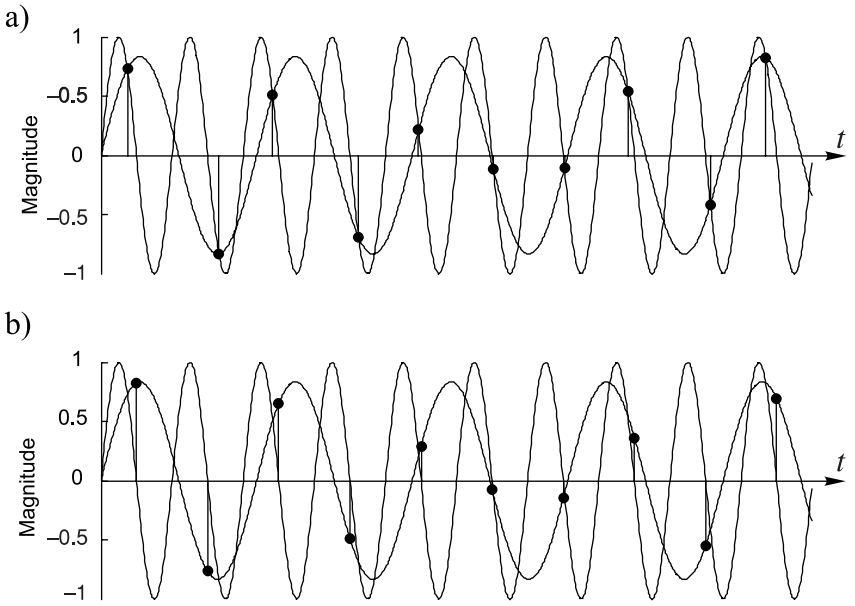


Figure 15.4 Regularization of the sampling operation based on sine-wave crossings

15.3.2 Fast Transforms of Signals Sampled at Sine-Wave Crossing Instants

To achieve the applicability of the standard DSP computer programs for processing nonuniform digital signals, the representation of these signals apparently must be somehow regularized. Some varieties of nonuniform sampling techniques could be regularized more easily than others. The sampling method based on sine-wave crossings, discussed in Chapter 7, might be considered as belonging to the first category as the specifics of this kind of sampling are typically favourable for regularization. Firstly, the nonuniformity of the sampling operation in this case is just a consequence of the specific signal sample value encoding based on timing of the signal and the reference waveform crossings. The irregularities observed with this type of sampling just occur; they are not introduced intentionally to achieve some desirable effect like suppression of the aliases. Secondly, it helps that the frequencies of the signals being sampled are usually below half of the mean sampling rate. The approach to regularization of the sampled signals in this case is simple and is illustrated in Figure 15.4.

The signal sample values $\{x_k\}$ are taken at the signal and the reference sine-wave crossing instants, as shown in Figure 15.4(a). To regularize this type

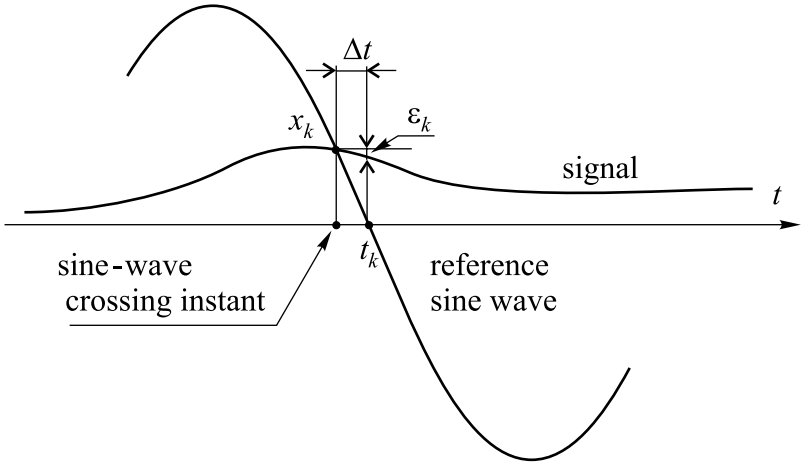


Figure 15.5 Regularization error occurring as a result of ascribing the signal instantaneous value x_k to the reference function zero-crossing instant

of sampling, all signal sample values should be assigned to the time instants t_k, $k = 0, 1, 2, \ldots$, which actually are the instants when the reference sine wave crosses the zero or some other constant level. Thus this regularization approach is simple and straightforward and leads naturally to the introduction of some noise $\varepsilon(t_k)$. The signal values at these time instants differ from those corresponding to the signal reference function crossing instants. As shown in Figure 15.5, the value of a particular error ε_k depends on the angle of the signal/sine-wave crossing and on the time interval Δt between this crossing instant and the reference function zero-crossing instant t_k. The fact that the time interval Δt is smaller for smaller values of the sampled signal helps to keep the relative errors close to a constant level for small to medium signal values.

When the sampling results are regularized as explained, most of the classic DSP algorithms, including the fast ones like the FFT, could be used for processing the digital signals obtained in this way. However, the errors introduced at the sampling result regularization impose some limitations on the applicability of this approach. To assess these limitations, it is shown in Figure 15.6 how the power of the regularization noise $\varepsilon(t_k)$ depends on the reference sine-wave frequency. Actually, the given curve is pessimistic as it has been calculated to conditions that are near to the worst. Specifically, the regularization errors ε_k were calculated for a full amplitude single-tone signal at the frequency just below the Nyquist limit. The time intervals Δt and the related regularization errors ε_k then tend to their maximal values.

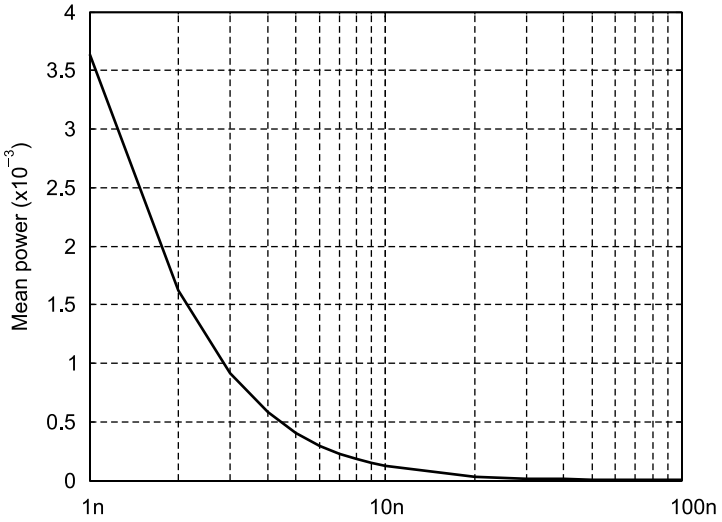


Figure 15.6 Mean power of the regularization noise versus normalized reference sine-wave frequency

These estimation errors ε_k were calculated under the sampling conditions emulating the specifics of massive data acquisition from a variable number of signal sources. Specifically, the power of the regularization errors ε_k are given in Figure 15.6 for the case where the single tone signal is characterized by frequency f_s and amplitude A. The mean sampling rate f_{ms} depends both on the reference frequency f_r and the number n of signal sources as $f_{ms} = f_r/n$. In this case the errors ε_k are calculated for various values of n while the mean sampling rate f_{ms} is kept constant and equal to 2.18753 f_x. To keep the mean sampling rate f_{ms} constant, the reference frequency f_r is appropriately changed for each $n$value. Note that n is equal to the reference frequency normalized with regard to the mean sampling rate $n = f_r/f_{ms}$. This means that the power P_ε of the regularization noise $\varepsilon(t_k)$, given for the growing numbers of signal sources, also indicates its dependence on the normalized reference frequency.

It can be seen from the diagram obtained and displayed in Figure 15.6 that the mean power of the regularization noise, even in the worst case, decreases quickly at relatively small numbers n of the signal sources encoded in this way. It seems that in many cases of data acquisition from multiple signal sources that the regularization approach should be applicable if the number is about 10 and more. The feasibility and rationale of this approach is further illustrated in Figure 15.7.

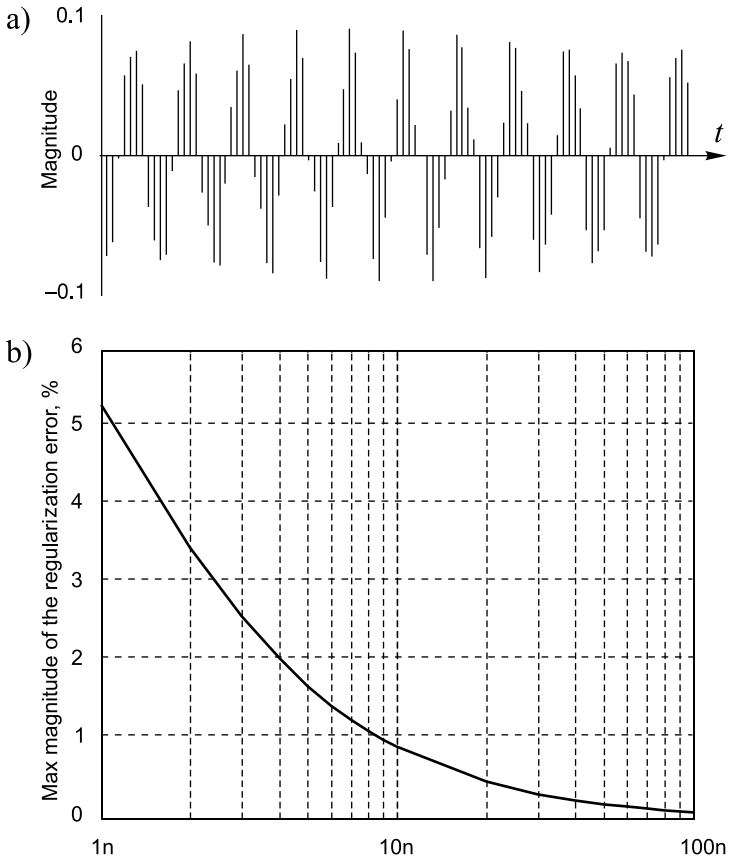


Figure 15.7 Illustration of (a) the digital regularization noise and (b) the maximal value of the regularization errors versus the normalized reference sine-wave frequency (number of signal sources) of the nearly worst case of sine-wave crossing sampling of a single-tone signal at a frequency close to the Nyquist limit

The digital regularization noise is shown in Figure 15.7(a) and the maximal value of the regularization errors versus the normalized reference sine-wave frequency (number of signal sources) is given in Figure 15.7(b).

The regularization noise $\varepsilon(t_k)$ impacts on the processing of signals sampled on the basis of the sine-wave crossing approach like any other additive noise present in the signal. Figure 15.8 illustrates the impact of this noise on the estimation of a Fourier coefficient. To realize conditions under which the impact of regularization

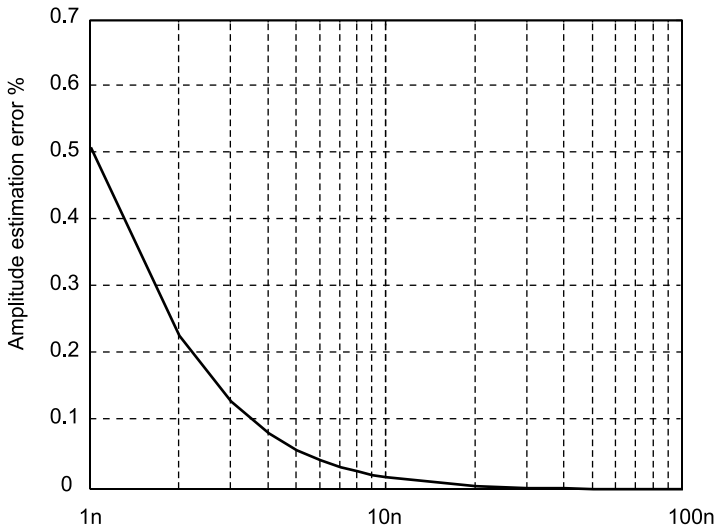


Figure 15.8 Impact of the sampling regularization on the estimation of the amplitude of a single-tone signal.

errors ε_k is most damaging, the errors in estimating the Fourier coefficients were calculated for a full amplitude single-tone signal at a frequency close to the Nyquist limit. The estimation errors were again calculated for the varying normalized reference frequency or, what is the same, for the varying number of signal sources.

The given curve corresponds to the nearly worst case. Errors corrupting the estimation of more complex signal parameters containing components at other lower frequencies are below the given curve. It can be seen from Figures 15.7 and 15.8 that even relatively small increases in the normalized reference frequency significantly reduce the noise $\varepsilon(t_k)$ caused by the regularization procedure and the negative effects due to it. As a result the impact of the sampling regularization on the signal processing results is insignificant in a large area of the normalized reference frequency. The errors due to the nonuniform sampling regularization in the case where this particular sampling technique is used for remote sampling at massive data acquisition from many signal sources become insignificant when the number n of the signal sources exceeds 5 to 10. Therefore the errors related to the considered procedure of sampling regularization under the conditions typical for data acquisition from a large number of signal sources can often be ignored. This fact leads to the following conclusion:

If the remote sampling operations carried out in the process of massive data acquisition are performed on the basis of sine-wave crossings, the digital signals obtained in this way can usually be regularized and presented as periodically sampled signals even at relatively small numbers of signal sources (about 10 or more). Consequently, most of the well-developed traditional DSP algorithms for processing periodically sampled signals are applicable for processing digital signals reconstructed from the data acquired in this way from a relatively large quantity of remote signal sources.

15.4 Revealing the Essence of the Fourier Coefficient Estimation

There is no doubt that the Fourier transform plays a significant role in various algorithms for signal processing. Therefore it is crucial to perform this transform and use the results correctly. While the use of the DFT for dealing with periodically sampled signals is a well-established routine, the involved estimation of the Fourier coefficients becomes more problematic whenever the signal to be transformed is represented by the sequence of its sample values obtained in the process of randomized sampling. Therefore the conditions under which the estimates of the Fourier coefficients are calculated should be well understood. Unfortunately they are estimates of a parameter calculated for a given set of data, which makes the use of them obscure. Often it is much better to use the approach introduced in Section 8.3, based on the representation of the DFT by cumulative sums. The point is that this approach makes it possible to observe the estimation of the Fourier coefficients as a process. This subject will now be expanded.

The DFT performed in the course of Fourier analysis actually represents estimation of a spectral parameters a_a or b_a characterizing a signal $x(t_k) = x_k$ at a given frequency f_a. According to definition,

$$\begin{aligned} \hat{a}_a &= \frac{2}{N}\sum_{k=0}^{N-1} x_k \cos(2\pi f_a t_k), \\ \hat{b}_a &= \frac{2}{N}\sum_{k=0}^{N-1} x_k \sin(2\pi f_a t_k). \end{aligned} \tag{15.10}$$

Calculations carried out according to these equations only provide the numerical values of the estimates $\hat{a}_a$ and $\hat{b}_a$. This kind of DFT might be considered as selective digital filtering. Application of such filters for signal decomposition and spectral analysis is quite popular. Their frequency response is given in

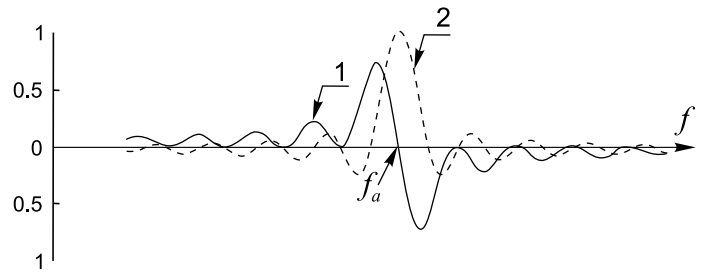


Figure 15.9 Frequency responses of the selective DFT filters for extracting the quadrature components of the signal.

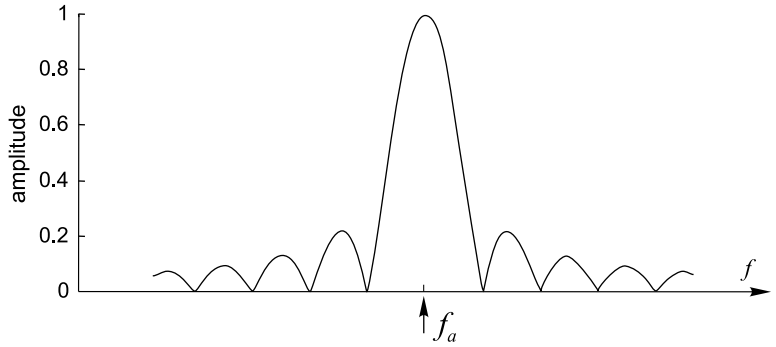


Figure 15.10 Typical amplitude–frequency response of a DFT filter

Figures 15.9 and 15.10 just as a reference for further analysis of various aspects of this type of filtering.

This kind of filtering will now be considered from a different angle. It can be seen from the expressions (15.10) that these coefficients are obtained by averaging the accumulated products of signal sample values each multiplied by the sample values of sine or cosine functions of a particular frequency taken at the same instants as the respective signal sample values. This accumulation process is stopped when a predetermined number N of sample values has been used. The Fourier coefficients are obtained as the end result of this accumulation process. No attention is paid to that process itself.

As pointed out in Chapter 8, the cumulative process is highly informative and substantially more information can often be extracted if this kind of spectral estimation is observed as a process. So far this approach has been used only for analysis of the effects from phase shifting of a periodic sampling point process. The proposed exploitation of the cumulative process has proved to be an efficient tool for discovery of well-hidden relationships. This fact has been confirmed by

demonstration of the compensation effect that takes place when properly phase-shifted periodic sampling point processes are used. Now it will be shown that analysis of this sort of cumulative sum forming is often productive and useful. Emphasis is put on the application of this approach to the resolution of various signal filtering tasks.

The cumulative sums related to the Fourier transform can be given as

$$\begin{aligned} \tilde{a}_a(n) &= \frac{2}{N}\sum_{k=0}^{n-1} x_k \cos(2\pi f_a t_k), \\ \tilde{b}_a(n) &= \frac{2}{N}\sum_{k=0}^{n-1} x_k \sin(2\pi f_a t_k), \end{aligned} \tag{15.11}$$

where f_a is the frequency to which this type of DFT filter is tuned.

The calculation and use of these cumulative sums makes it possible to observe estimation of a Fourier coefficient as a process depending both on the signal being analysed and on the sampling point process used. Observation and analysis of these processes leads to the ability to obtain additional information.

The properties of these cumulative sums will now be described, starting with a case where the signal being filtered is given as

$$x_k = a_0 \cos(2\pi f_0 t_k), \quad k = \overline{0, N-1}.$$

In this particular case the cumulative sums become

$$\begin{aligned} \tilde{a}_a(n) &= a_0 \frac{2}{N}\sum_{k=0}^{n-1} \cos(2\pi f_0 t_k)\cos(2\pi f_a t_k), \\ \tilde{b}_a(n) &= a_0 \frac{2}{N}\sum_{k=0}^{n-1} \cos(2\pi f_0 t_k)\sin(2\pi f_a t_k). \end{aligned} \tag{15.12}$$

The properties of these functions depend on the frequencies f_0 and f_a and on the sampling point stream. Naturally, they are determined more easily in the case of periodic sampling. The cumulative sums shown in Figure 15.11 have been obtained for the coefficients $a(f_0)$ for various values of the frequencies f_0 and f_a. Note that in the case where $f_0 \neq f_a$ the value of the Fourier coefficient differs if it is estimated for various values of N.

It can be seen that these cumulative sums are quite informative. By observing them, it is possible to obtain more information about the involved signals than by just calculating the values of the corresponding Fourier coefficients. Therefore it often makes sense to use them in order to reveal the essence of various processes. It is easier to achieve that if it is clear what various shapes of these functions mean.

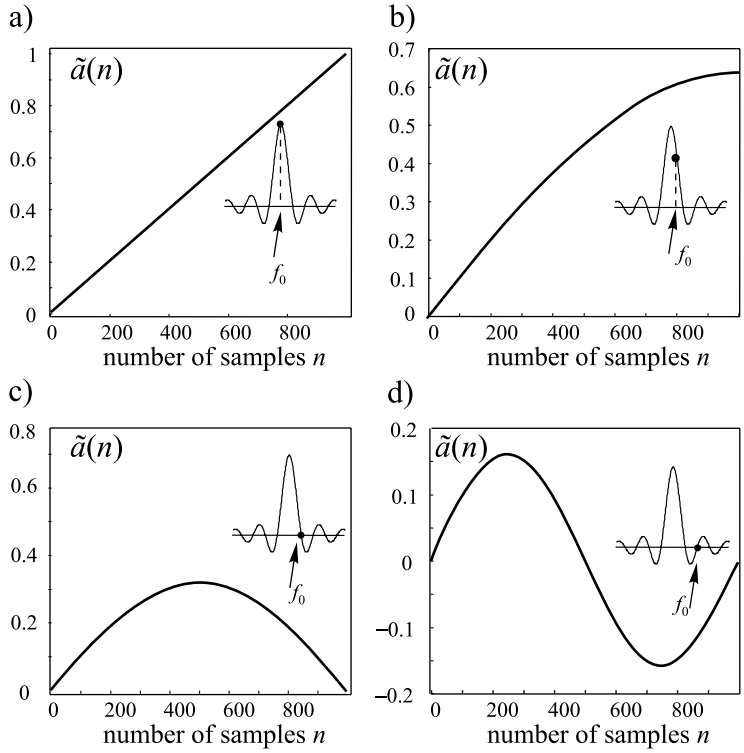


Figure 15.11 Cumulative sums for various locations of frequencies f_0 with respect to the central filter frequency f_a in the case where the difference between them is relatively small

Some typical cumulative sums obtained under specific conditions are shown in Figure 15.12.

The given examples show typical cumulative sums for the case of a single tone signal. The signal structure that might be expected in reality is of course much more complicated. Nevertheless, much can be learnt from observations of this kind of typical cumulative sum. First of all, the cumulative sum for a signal component at a frequency equal to the frequency to which the respective DFT filter is tuned graphically looks like a linearly increasing function. Therefore the presence of a linearly increasing component in a cumulative sum is evidence that the signal contains a component at the frequency of DFT filtering. This fact has been used before in Chapter 8 to demonstrate the alias compensation effect.

However, in general the analysis of these cumulative sums might prove not to be so simple. Apparently it has to cover cases where signals are composed

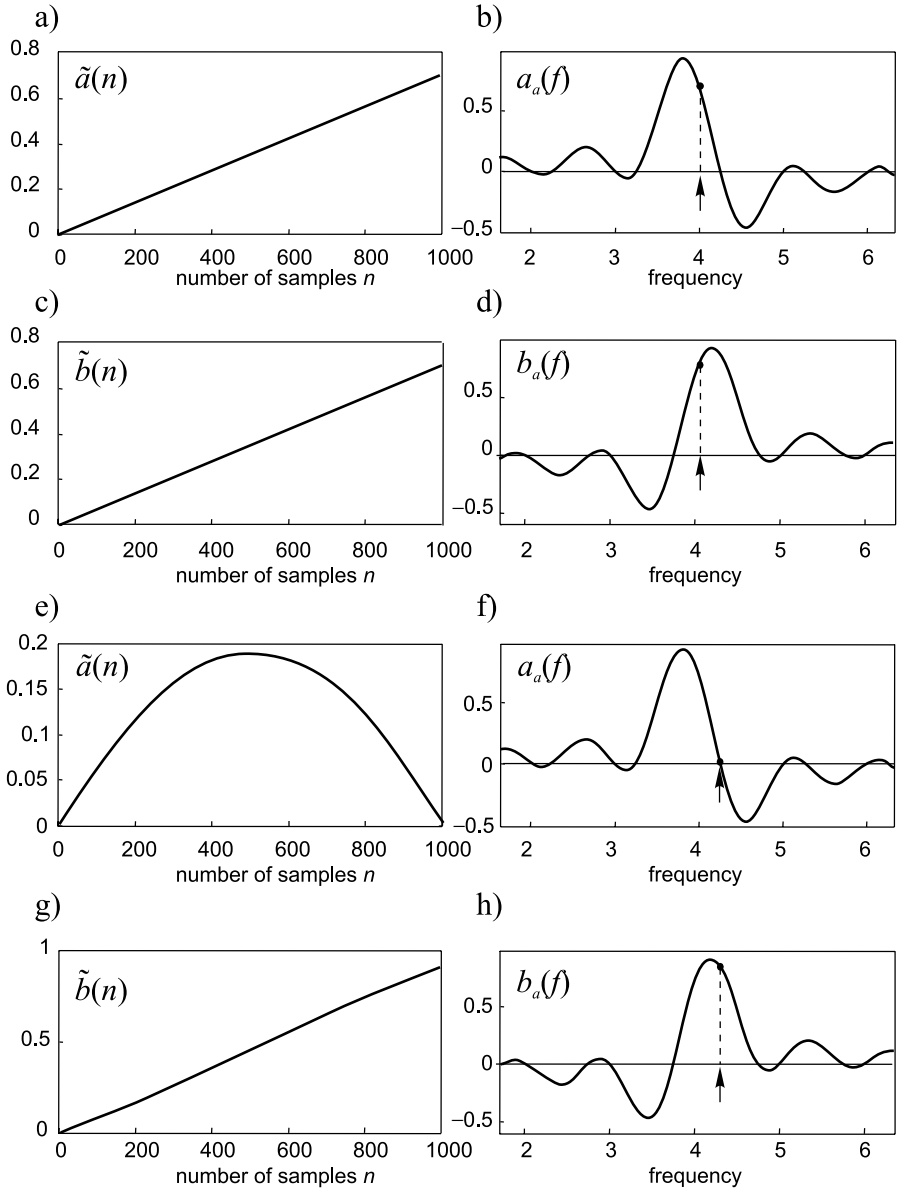


Figure 15.12 Cumulative sums for both Fourier coefficients calculated for various locations of frequencies f_0 with respect to the central filter frequency f_a in the case where the signal phase angle $\varphi = 45^0$

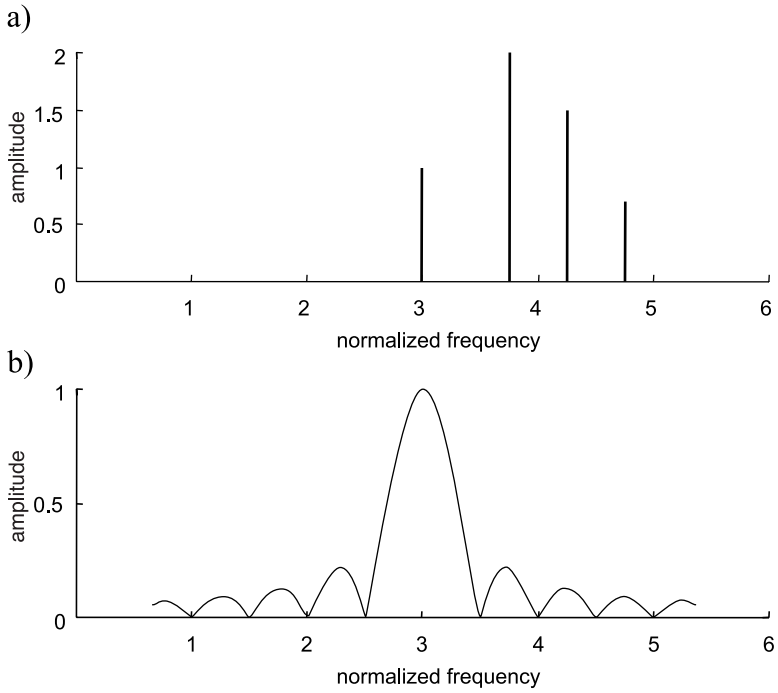


Figure 15.13 Example of (a) a multifrequency signal filtered by (b) a DFT filter tuned to the frequency of the first signal component in its spectrum

from multiple components. Figure 15.13 illustrates the forming and analysing of cumulative sums in the case of multifrequency signals containing four components at the indicated frequencies.

An ideal selective filter would filter out just one of these components. It would filter out the component at the frequency equal to the frequency at which the filter is tuned. In terms of cumulative sums, this component forms the linear increasing function 1 shown in Figure 15.14. The values indicate the amplitude of the corresponding component with the number N of signal samples taken into account. All other signal components would have to be filtered out. In reality the DFT filter only weakens their impact. These components form the particular cumulative sums 2, 3 and 4. The resulting cumulative sum 5 provides the result for the conventional DFT. The difference between the cumulative sums 1 and 5 represents the DFT errors. Two particular errors for $N = 512$ and $N = 1024$ are indicated. It is shown that there might be even bigger errors taking place at different N values. This example also shows what has to be achieved to improve

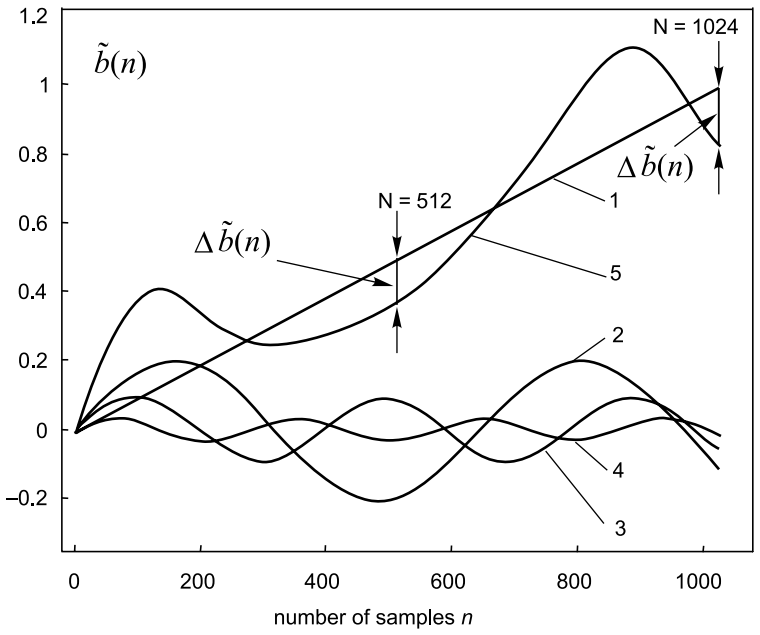


Figure 15.14 Cumulative sums formed for the multifrequency signal given in Figure 15.13

DFT filtering. In fact, only the linear component of the total cumulative sum needs to be preserved. All other oscillating components have to be taken out.

Attention should be drawn to the essential point that the traditional digital filtering methods could be used for processing or special filtering of the cumulative sums as the discrete argument (the sample-taking instants) in this case is regular. It is true that there are pseudo-random skips of the sampling instants, but each sampling event is placed on a regular grid. While filtering of the signal sample value sequences require application of special techniques as the sampling intervals are nonuniform, transformation of the task to filtering of cumulative sums leads to regularization of the filtering. This is a significant fact and can be exploited in various ways in order to increase the accuracy of signal transforms from time to frequency domains.

Note that the necessity to process signals in this way comes from the fact that the frequency response of the DFT filter is not ideal. It has side lobes. The usual approach used to reduce the harmful effects due to them is based on windowing. The described approach based on application of cumulative sums represents an alternative to application of the windowing techniques.

Bibliography

Allay, N. and Tarczynski, A. (2004) Spectral magnitude and phase estimation of digitally modulated signals from their nonuniform samples. In Proceedings of the IASTED International Conference on *Communications Systems and Networks (CSN 2004)* Marbella, Spain, 1–3 September 2004, pp. 237–42.

Bilinskis, I. and Mikelsons, A. (1992) *Randomized Signal Processing*. Prentice-Hall International (UK) Ltd.

Bland, D. M., Laakso, T. and Tarczynski, A. (1996) Analysis of algorithms for nouniform-time discrete Fourier transform. In International Symposium on *Circuits and Systems* (*ISCAS'96*), Vol. 2, Atlanta, Georgia, 12–15 May 1996, pp. 453–6.

Bland, D. M., Laakso, T. and Tarczynski, A. (1996) Application of NUT-DFT for reasampling nonuniformly sampled signals. In Proceedings of the Third IEEE International Conference on *Electronics, Circuits, and Systems* (*ICECS'96*), Vol. 1, Rhodos, 13–16 October 1996, pp. 586–9.

Bland, D.M., Tarczynski, A. and Laakso, T.I. (1996) Spectrum analysis of nonuniformly sampled signals using the NUT-DFT. In Proceedings of the IEEE Nordic Signal Processing Symposium *NORSIG'96*, Vol. 1, Helsinki, 24–27 September 1996, pp. 307–10.

Tarczynski, A. (2002) Spectrum estimation of nonuniformly sampled signals. In Proceedings of the 14th International Conference on *Digital Signal Processing* (*DSP2002*), Vol. 2, Santorini, Greece, 1–3 July 2002, pp. 795–8. 7504-1.

Tarczynski, A. and Allay, N. (2003) Digital alias-free spectrum estimation of randomly sampled signals. In the 7th World Multi-Conference on *Systemics, Cybernetics and Informatics* (*SCI'03*), Vol. IV, Orlando, Florida, 27–30 July 2003, pp. 344–8.

Tarczynski, A. and Allay, N. (2004) Spectral analysis of randomly sampled signals: suppression of aliasing and sampler jitter. *IEEE Trans. on Signal Processing*, **52** (12), December, 3324–34.

Tarczynski, A., Bland, D. M. and Laakso, T. (1996) Spectrum estimation of non-uniformly sampled signals. In Proceedings of the IEEE International Symposium on *Industrial Electronics* (*ISIE'96*), Vol. 1, Warsaw, 17–20 June 1996, pp. 196–200.

Tarczynski, A. and Valimaki, V. (1996) Modifying FIR and IIR filters for processing signals with lost samples. In Proceedings of the IEEE Nordic Signal Processing Symposium (*NORSIG'96*), Vol. 1, Helsinki, 24–27 September 1996, pp. 359–62.

16

Complexity-reduced DFT

As the regularity of signal sample value positioning in time is a crucial precondition for operation of the fast signal processing algorithms, these types of algorithms, in general, are inappropriate for processing the digital signals obtained as a result of deliberate randomization of the sampling process. There are, however, exceptions, some of which are described in Section 15.3. These exceptions actually prove the rule as the applicability of the FFT in those cases is based on the rarely usable regularization of the randomized sampling procedure. Thus the fact remains that the strategy for rationalizing the nonuniform sampled signal processing cannot be based on application of the popular fast DSP algorithms. Other approaches to the algorithm complexity reduction have to be found and exploited. Some useful techniques for that are suggested and discussed in this chapter.

The possibility of reducing the computational burden for the DFT, achieved by applying the methods for complexity-reduced spectral analysis, while attractive, is always especially interesting for increasing the speed of on-line estimations of Fourier coefficients, which often have to be carried out using various types of signal preprocessing operations. An example illustrating this kind of signal preprocessing and the benefits that can be obtained is described in Section 19.6.

16.1 Potential Gains from Application of Rectangular Function Sets

The number of multiplication operations that have to be carried out in order to obtain signal spectra in the frequency domain from their digital waveforms given

Digital Alias-free Signal Processing I. Bilinskis

in the time domain is the basic characteristic roughly indicating the complexity of the respective algorithms for such signal transforms. Although the execution of multiplication operations by means of specialized hardware nowadays costs considerably less than in the recent past, this kind of algorithm requiring a smaller quantity of multiplication operations is still in demand. Apparently it is much more difficult to develop this type of algorithm if the digital signals to be processed are nonuniform. The experience accumulated in the area of developing fast algorithms for transforming the periodic sampled signals simply cannot be used in this case. However, in addition to the application of these fast algorithms, there is also an alternative approach to the resolution of this complexity reduction problem that is much better suited for rationalizing this type of signal transform. This approach is based on the exploitation of rectangular function sets. Their attractiveness is defined by the fact that these functions assume only the values $-1, 0$ and $+1$. Consequently, application of them leads to the replacement of numerous multiplication operations by simple logic operations. While there are various types of these function systems and they are used in different ways, application of them could be beneficial both for processing periodically as well as nonuniformly sampled signals. On the other hand, the conditions for successful use depend on the specifics of the digital signals to be processed together with these rectangular functions, and this fact has to be taken into account.

In general, it is possible to gain benefits from using these types of rectangular function in two different ways: (a) performing sequential analysis with the orthogonal rectangular function sets used as the basis functions and (b) exploiting this type of function as a tool for complexity reduction of the DFT.

16.1.1 Use of Orthogonal Rectangular Functions

Sequential analysis performed on the basis of the Walsh, Haar and other similar functions can be mentioned as the most impressive developments in this area. The main factor making application of these function systems attractive is the fact that they are rectangular and assume only the values $-1, +1$ and sometimes also 0. Therefore when these rectangular waveforms are used as the basis functions for execution of the spectrum analysis, no multiplication operations need to be executed at all. That is the case, for instance, when obtaining the Walsh spectrum without performing many of the relatively complicated multiplication operations. However, this spectrum, a representation of signals in the frequency domain, although easily obtainable, is specific. It is then expressed in terms of

Walsh functions. In many cases it makes them inapplicable or at least inconvenient, because such spectra cannot be directly interpreted in terms of sinusoidal signal spectra commonly used in radioelectronics. This has stimulated many attempts to develop Walsh-to-Fourier spectral conversion methods. Although under certain conditions such conversions are feasible, they are computationally burdensome, so the rationale for this is questionable. What is gained when calculating the Walsh coefficients might well be lost during the following spectral conversions.

Formally, this type of transform might be performed both for periodic and nonuniform sampled signals. However, the topic of signal spectrum analysis on the basis of Walsh and similar orthogonal functions as a specific one is beyond the scope of this book. Attention is focused on the second of the two mentioned approaches. In this case the rectangular functions are used as a tool for performing the DFT in a computationally effective way while the signal spectra are obtained in the commonly used and easily interpretable form.

16.1.2 Reduction in the Computational Burden for DFT

The outline of the suggested approach is as follows. The Fourier transform is accomplished in two stages. In the first stage, the signal to be analysed is formally decomposed on the basis of some type of special rectangular function assuming only the values -1, 0 and $+1$, so that the decomposition can be carried out without multiplication. In the second stage a spectral conversion is performed. The set of coefficients obtained at the first stage is converted into a set of Fourier coefficients. The complexity of the calculations largely depends on the properties of the used rectangular functions.

This outline, in general, coincides with the Walsh-to-Fourier conversion scheme mentioned above, according to which the input signal at the first stage is decomposed on the basis of orthogonal Walsh functions. In both cases the spectral analysis is carried out in two steps. However, there the similarity ends. According to the suggested approach the signal spectrum analysis is carried out on the basis of the DFT while the involved sets of rectangular functions just support the involved calculations providing for the complexity reduction. The point is that they should be appropriate and well suited for carrying out the calculations at the second stage and it is not a requirement that they are orthogonal. Although the basic principles of the method are deterministic, the application of random sampling helping to avoid aliasing is essential for its digital implementation. A description of this method follows.

16.2 Complexity-reduced DFT Exploiting Rectangular Functions

It is essential that the rectangular functions used in the first stage of the suggested kind of complexity-reduced DFT do not necessarily have to be orthogonal. They should satisfy only one condition: their inner structure, displayed by their spectra, should be appropriate for the purposes of the conversions to be carried out in the second stage. Once this position is understood, it is relatively easy to find suitable functions and a number of them have been selected.

16.2.1 Essentials of the Method

The most useful two-function systems, which are considered in more detail in the following discussion, are given in Figures 16.1 and 16.2. Consider system 1 of the rectangular functions. The functions belonging to this system are periodic and nonorthogonal. They are in fact the sign functions of the sine and cosine functions at respective frequencies. At each of the considered frequencies there is a pair of functions. The functions of each pair are shifted against one another by $\pi/2$. These periodic rectangular functions are denoted by $R_{\mathrm{s}}(i\,\Delta f, t)$, $R_{\mathrm{c}}(i\,\Delta f, t)$. The indices s and c indicate that they are the sign functions of the corresponding sine and cosine waves respectively.

The formal decomposition of a signal $x(t)$ on the basis of the given rectangular functions actually involves the calculation of a set of coefficients $\{\alpha_i, \beta_i\}$ covering the whole frequency range of interest. If the signal $x(t)$ is observed during the time interval $(t_b - t_a)$ then, for example, the following equations can be written:

$$\begin{aligned}\beta_i &= \frac{1}{t_b - t_a}\int_{t_a}^{t_b} x(t) R_{\mathrm{s}}(i\,\Delta f, t)\,\mathrm{d}t \\ &= \frac{1}{t_b - t_a}\left[\int_{t_a}^{t_1} x(t)\,\mathrm{d}t - \int_{t_1}^{t_2} x(t)\,\mathrm{d}t + \int_{t_2}^{t_3} x(t)\,\mathrm{d}t - \cdots\right], \qquad (16.1)\end{aligned}$$

where $t_1, t_2, t_3, \ldots$ are the instants when the function $R_{\mathrm{s}}(i\,\Delta f, t)$ changes sign. It can be seen from this equation that the calculation of the coefficient β_i (likewise also the coefficient α_i) does not require execution of any multiplication operations.

To find out what this coefficient means, consider the structure of $R_{\mathrm{s}}(i\,\Delta f, t)$. This function can be expanded into the following Fourier series:

$$R_{\mathrm{s}}(i\,\Delta f, t) = b_{\mathrm{s}1}\sin(2\pi i\,\Delta f t) + b_{\mathrm{s}3}\sin(2\pi 3 i\,\Delta f t) + b_{\mathrm{s}5}\sin(2\pi 5 i\,\Delta f t) + \cdots. \qquad (16.2)$$

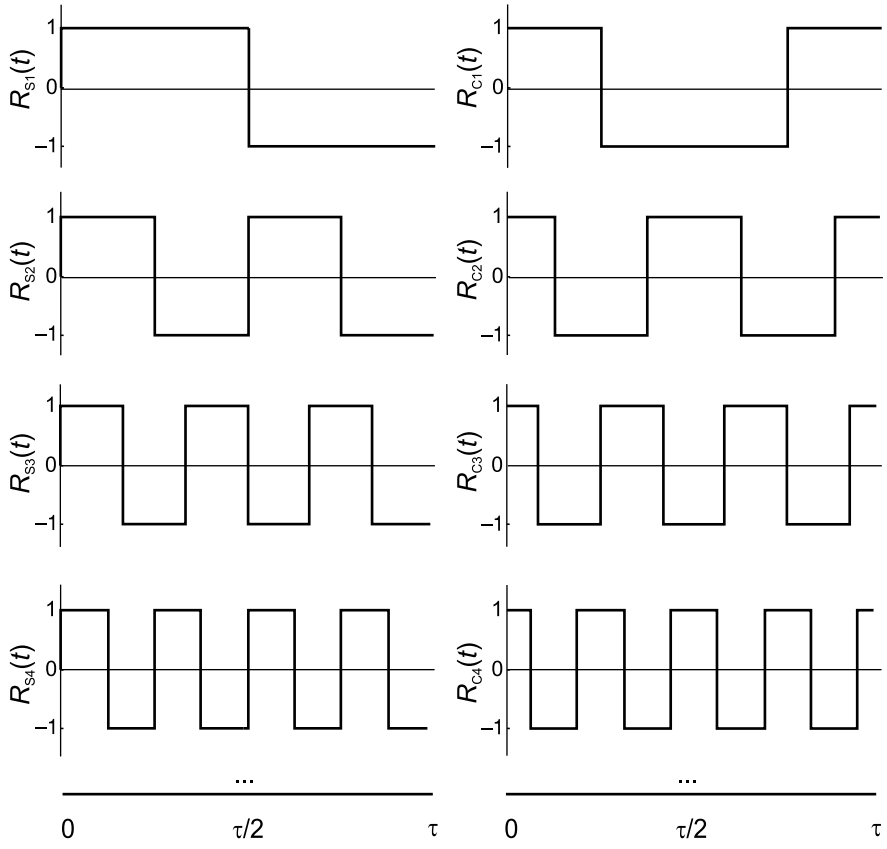


Figure 16.1 Rectangular function system 1

It follows from Equations (16.1) and (16.2) that

$$\beta_i = \frac{1}{t_b - t_a}\left[b_{s1}\int_{t_a}^{t_b} x(t)\sin(2\pi i\,\Delta f\,t)\,\mathrm{d}t\right] + b_{s3}\int_{t_a}^{t_b} x(t)\sin(2\pi 3i\,\Delta f\,t)\,\mathrm{d}t + b_{s5}\int_{t_a}^{t_b} x(t)\sin(2\pi 5i\,\Delta f\,t)\,\mathrm{d}t + \cdots. \tag{16.3}$$

As this equation shows, the value of the coefficient β_i, calculated for the frequency $i\,\Delta f$, actually depends on the signal components at the frequencies $3i\,\Delta f, 5i\,\Delta f, 7i\,\Delta f, \ldots$. Hence a separate pair of coefficients α_i and β_i is

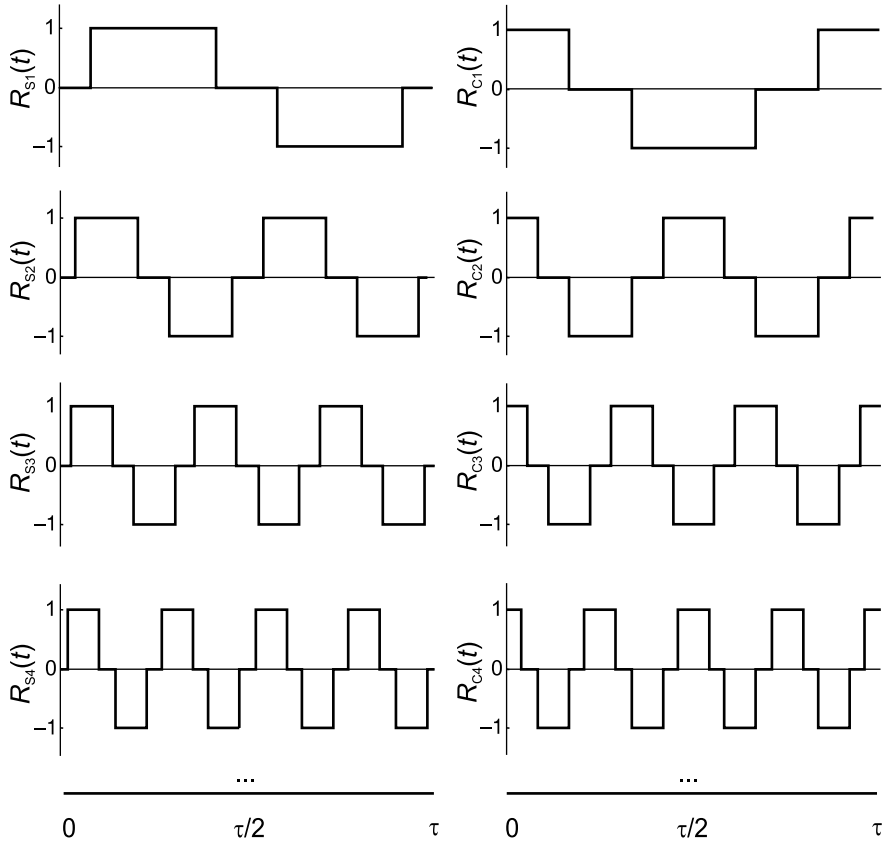


Figure 16.2 Rectangular function system 2

meaningless. Only a full set of them, calculated for the whole signal frequency range, carries information about the signal spectrum and, as shown in the next subsection, the conversion of these coefficients $\{\alpha_i, \beta_i\}$ into Fourier coefficients in this case is then a simple matter.

16.2.2 Mathematical Description

Suppose that a signal is represented by a real centred function $x(t)$ that is absolutely integrable over the time interval $[0, T]$ and that the upper frequency of the signal spectrum is f_0. Consider solving the task of calculating the Fourier series expansion of $x(t)$ on the basis of the method briefly described in the previous section.

Consider the rectangular functions of system 1 shown in Figure 16.1. They are defined by

$$\begin{aligned}
&R_{\mathrm{s}}(i\,\Delta f, t) = R_{\mathrm{s}}(i\,\Delta f, t + rT_i), \quad T_i = \frac{1}{i\,\Delta f}, \quad r = 0, 1, 2, \ldots, \\
&R_{\mathrm{c}}(i\,\Delta f, t) = R_{\mathrm{s}}\left(i\,\Delta f, t + \frac{T_i}{4}\right), \quad i = 1, 2, \ldots, \\
&R_{\mathrm{s}}(i\,\Delta f, t) = R_{\mathrm{s}}(i\,\Delta f, -t), \quad R_{\mathrm{s}}(i\,\Delta f, t) \in [1, -1], \\
&\int_0^{T_i} R_{\mathrm{s}}(i\,\Delta f, t)\,\mathrm{d}t = 0
\end{aligned} \tag{16.4}$$

The signal $x(t)$ is first decomposed on the basis of these functions. Then

$$\begin{aligned}
\alpha_i &= \frac{2}{T}\int_0^T x(t) R_{\mathrm{c}}(i\,\Delta f, t)\,\mathrm{d}t, \\
\beta_i &= \frac{2}{T}\int_0^T x(t) R_{\mathrm{s}}(i\,\Delta f, t)\,\mathrm{d}t.
\end{aligned} \tag{16.5}$$

On the other hand, the Fourier series expansions of the functions $R_{\mathrm{s}}(i\,\Delta f, t)$, $R_{\mathrm{c}}(i\,\Delta f, t)$, given by

$$\begin{aligned}
R_{\mathrm{s}}(i\,\Delta f, t) &= \sum_{r=1}^{\infty} b_{sr} \sin(2\pi r i\,\Delta f t), \\
R_{\mathrm{c}}(i\,\Delta f, t) &= \sum_{r=1}^{\infty} a_{cr} \cos(2\pi r i\,\Delta f t),
\end{aligned} \tag{16.6}$$

where

$$a_{sr} = \frac{4(-1)^{r+1}}{\pi(2r-1)}, \qquad b_{cr} = \frac{4}{\pi(2r-1)}$$

are the Fourier coefficients for the frequencies $r i\,\Delta f$.

Substituting expressions (16.6) into Equations (16.5) yields

$$\begin{aligned}
\alpha_i &= \frac{2}{T}\int_0^T x(t) \sum_{r=1}^{\infty} a_{cr} \cos(2\pi r i\,\Delta f t)\,\mathrm{d}t \\
&= \sum_{r=1}^{\infty} a_{cr} \frac{2}{T}\int_0^T x(t) \cos(2\pi r i\,\Delta f t)\,\mathrm{d}t.
\end{aligned} \tag{16.7}$$

Similarly, it can be shown that

$$\beta_i = \sum_{r=1}^{\infty} b_{sr} \frac{2}{T} \int_0^T x(t) \sin(2\pi r i \Delta f t)\, \mathrm{d}t. \tag{16.8}$$

As the Fourier coefficients of $x(t)$ at the frequency $ri\Delta f$ are defined as

$$a_{ri} = \frac{2}{T} \int_0^T x(t) \cos(2\pi r i \Delta f t)\, \mathrm{d}t,$$

$$b_{ri} = \frac{2}{T} \int_0^T x(t) \sin(2\pi r i \Delta f t)\, \mathrm{d}t,$$

it follows from Equation (16.7) that

$$\alpha_i = \sum_{r=1}^{\infty} a_{cr} a_{ri}, \qquad \beta_i = \sum_{r=1}^{\infty} b_{sr} b_{ri}. \tag{16.9}$$

In the cases when the spectra of signals are restricted, Equations (16.9) are finite. Then

$$\alpha_i = \sum_{r=1}^{[n/i]} a_{cr} a_{ri}, \qquad \beta_i = \sum_{r=1}^{[n/i]} b_{sr} b_{ri}, \quad i = \overline{1, n}, \tag{16.10}$$

where $n = f_0/\Delta f$ is the number of frequencies considered and $[n/i]$ is the integer part of n/i. Thus the relationship between the coefficients $\{\alpha_i, \beta_i\}$ and the Fourier coefficients sought are described by two similar systems of n linear independent equations containing n unknown variables.

The equation system for determining the Fourier coefficients $\{a_i\}$ can be given as the following matrix equation:

$$\begin{bmatrix} \nu_{11} & \nu_{12} & \nu_{13} & \cdots & \nu_{1n} \\ 0 & \nu_{22} & \nu_{23} & \cdots & \nu_{2n} \\ 0 & 0 & \nu_{33} & \cdots & \nu_{3n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \nu_{nn} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \cdots \\ a_n \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \cdots \\ \alpha_n \end{bmatrix}. \tag{16.11}$$

An element of this matrix is denoted by ν_{ij}, $i = \overline{1, n}$; $j = \overline{1, n}$. For $i < \mathrm{j}$, $\nu_{ij} = 0$ and

$$\nu_{ij} = \begin{cases} a_{cm} & \text{if } m = i/j = [i/j], \\ 0 & \text{if } m = i/j \neq [i/j]. \end{cases} \tag{16.12}$$

When the Fourier coefficients characterizing the respective rectangular functions are given, expression (16.12) allows all the elements ν_{ij} to be found. Specifically, it follows from this expression that $\nu_{11} = \nu_{22} = \nu_{33} = \cdots = \nu_{nn} = a_{c1}$.

If the functions $R_c(i\Delta f, t)$ are even, the equation describing the relationship between the coefficients $\{\alpha_i\}$ and the Fourier coefficients $\{a_i\}$ is given as

$$\alpha_i = \sum_{r=0}^{\infty} a_{c(2r+i)} a_{(2r+1)i}. \tag{16.13}$$

In the case where the signal spectrum is restricted and the Fourier coefficients at frequencies exceeding $n\Delta f$ can be considered to be equal to zero,

$$a_{(2r+1)i} = 0 \quad \text{for } (2r+1)i > n \tag{16.14}$$

and Equation (16.13) is then given by

$$\alpha_i = \sum_{r=0}^{[n/2i-1/2]} a_{c(2r+i)} a_{(2r+1)i}. \tag{16.15}$$

It follows that $\alpha_i = a_{c1} a_i$ for all i satisfying the inequality

$$\frac{n}{2i} - \frac{1}{2} < 1.$$

Therefore

$$\alpha_i = a_{c1} a_i \quad \text{for } i > [n/3]. \tag{16.16}$$

In the case where, for example, $n = 16$, the relationships can be described by the following system of equations:

$$\begin{aligned}
\alpha_1 &= a_{c1}a_1 + a_{c3}a_3 + a_{c5}a_5 + a_{c7}a_7 + a_{c9}a_9 + a_{c11}a_{11} + a_{c13}a_{13} + a_{c15}a_{15} \\
\alpha_2 &= a_{c1}a_2 + a_{c3}a_6 + a_{c5}a_{10} + a_{c7}a_{10} \\
\alpha_3 &= a_{c1}a_3 + a_{c3}a_9 + a_{c5}a_{15} \\
\alpha_4 &= a_{c1}a_4 + a_{c3}a_{12} \\
\alpha_5 &= a_{c1}a_5 + a_{c3}a_{15} \\
\alpha_6 &= a_{c1}a_6 \\
&\cdots\cdots\cdots \\
\alpha_{16} &= a_{c1}a_{16}.
\end{aligned} \tag{16.17}$$

From this example, for $n = 16$ all the Fourier coefficients $a_i, i > 5$, are obtained simply by normalizing the respective coefficient α_i. On the basis of

Equations (16.16) and (16.17),

$$a_i = \begin{cases} \dfrac{\alpha_i}{a_{c1}} & \text{for } i = \overline{n, [n/3]+1}, \\ \dfrac{\alpha_i - \sum_{r=1}^{[n/2i-1/2]} a_{cr} a_{(2r+1)i}}{a_{c1}} & \text{for } i = \overline{[n/3], 1}. \end{cases} \tag{16.18}$$

Similarly,

$$b_i = \begin{cases} \dfrac{\beta_i}{b_{s1}} & \text{for } i = \overline{[n/3],1}, \\ \beta_1 - \sum_{r=1}^{[n/2i-1/2]} b_{sr} b_{(2r+1)i} & \text{for } i = \overline{[n/3], 1}. \end{cases} \tag{16.19}$$

Thus the spectral conversion at the second stage of the Fourier analysis performed in accordance with the approach suggested is indeed simple, and only a small number of multiplication operations are needed to obtain the Fourier coefficients, far fewer than are needed for the FFT.

These calculations can be simplified even further by using the rectangular functions of system 2, shown in Figure 16.2. The Fourier series representing these functions not only does not contain terms with even indices but each third term with odd indices is also missing, i.e. all the coefficients $a_{c3} = a_{c9} = a_{c15} = a_{c21} = \cdots$ are equal to zero. This, of course, leads to further simplification of the calculations at the second stage. Figure 16.3 illustrates how the conditions for the Fourier coefficient calculations depend on which of the systems 1 or 2 of the rectangular functions is used. Two-stage calculations have to be carried out only in the frequency range $[0, 1/3 f_0]$ when system 1 is used. In the cases where system 2 is used, direct single-stage estimation of the Fourier coefficients could be carried out even in a broader frequency range, covering four-fifths of the whole signal bandwidth.

16.2.3 Digital Implementation

Both the hardware and software implementations of the described method are basically simple, which represents a significant advantage of it. After the input signal $x(t)$ has been digitized by an ADC, it is processed to obtain the estimates of the coefficients $\hat{\alpha}_i$ and $\hat{\beta}_i$. The following equations are used:

$$\begin{aligned} \hat{\alpha}_i &= \frac{2}{N} \sum_{k=1}^{N} x(t_k) R_c(i\Delta f, t_k), \\ \hat{\beta}_i &= \frac{2}{N} \sum_{k=1}^{N} x(t_k) R_s(i\Delta f, t_k). \end{aligned} \tag{16.20}$$

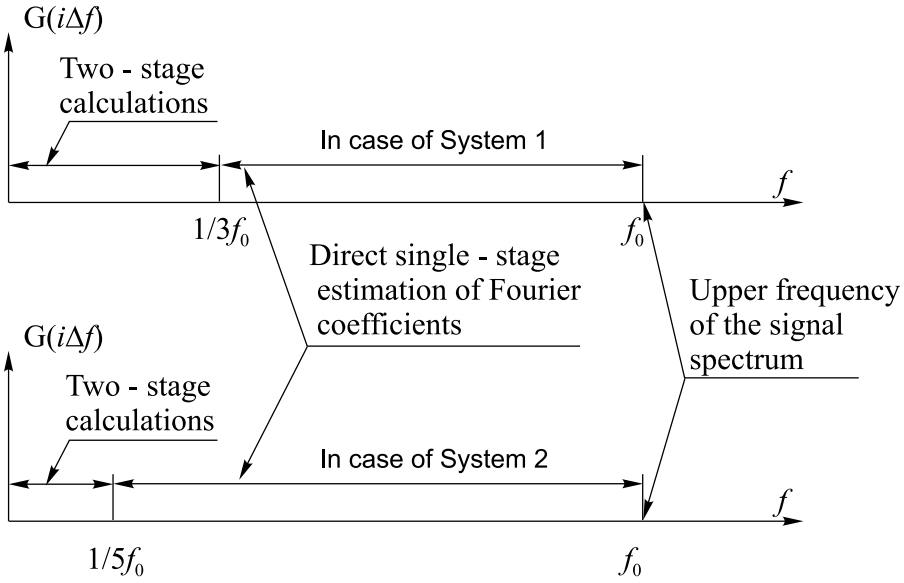


Figure 16.3 Dependence of the conditions for direct single-stage estimation of the Fourier coefficients on the systems of the rectangular functions used

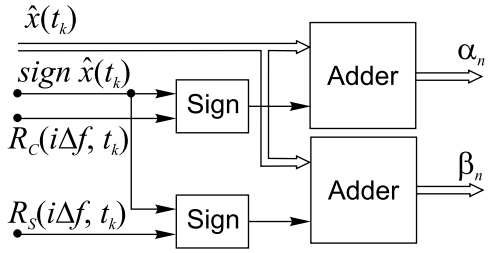


Figure 16.4 Block diagram of the basic digital filter for estimation of the coefficients α_i and β_i

A simple electronic device is used to estimate the α_i and β_i coefficients. A block diagram of it is given in Figure 16.4.

As the functions $R_s(i\Delta f, t)$ and $R_c(i\Delta f, t)$ assume only the values -1, 0 and $+1$, only two adders and a couple of logic elements are needed to perform an estimation of both coefficients. To perform full-scale DFT, a number of such devices, provisionally called filters, are connected and used in parallel. In addition to the bank of these filters, there is also an ADC and a memory unit, which is used for storing the discrete values of the functions $R_s(i\Delta f, t)$ and $R_c(i\Delta f, t)$.

The functioning of this instrument is obvious. Once the estimates $\{\hat{\alpha}_i, \hat{\beta}\}_i$ are obtained, they are converted into the respective estimates of the required Fourier coefficients.

Although implementation of the method looks extremely simple, the relationships underlying discrete Fourier transforms performed in this way are more complicated than they seem at first glance. The behaviour of such a system mainly depends on the sampling method by which the input signal and the applied rectangular functions are sampled. The problem is that the spectra of the functions $R_s(i\Delta f, t)$ and $R_c(i\Delta f, t)$ are infinite and cannot in principle be restricted by low-pass filtering. It is true that the coefficients of the respective Fourier series steadily decrease, but there are still many high-frequency components that may cause aliasing errors if these functions are sampled periodically. To suppress the aliases and the errors due to them, randomizing of the sampling process often has to be performed. Randomization of sampling is crucial if the input signals have components at frequencies exceeding the Nyquist limit. This means that the performance of such a system mainly depends on the quality of its software and on the perfection of the used methods and algorithms for processing nonuniformly represented digital signals. Randomization of sampling, while necessary and irreplaceable as a tool for the elimination of aliasing, is not a reliable method for getting sufficiently accurate signal spectra, as shown in the next section. Special signal processing procedures have to be used in addition to obtain high precision at spectral analysis carried out in a complexity-reduced way. Fortunately it is now feasible, as such techniques have recently been developed. The achievable level of performance in this area is demonstrated in Chapter 18, where adapting the considered discrete Fourier transforms to nonuniform sampling irregularities is discussed.

16.3 Computer Simulations of the Rectangular Function-based DFT

To gain an impression of the properties of the complexity-reduced DFT based on application of the rectangular basis functions, the computer simulation results given below may be considered. However, to come to the correct conclusions while doing that, it is essential to keep in mind that the method used to estimate the Fourier coefficients is aimed at obtaining valuable intermediate signal processing results rather than signal spectra, as explained in the introduction to Chapter 15. This means that the end results at this stage of the approach to the DFT are estimates of the Fourier coefficients appropriate for various methods of further processing, including high-performance spectral analysis.

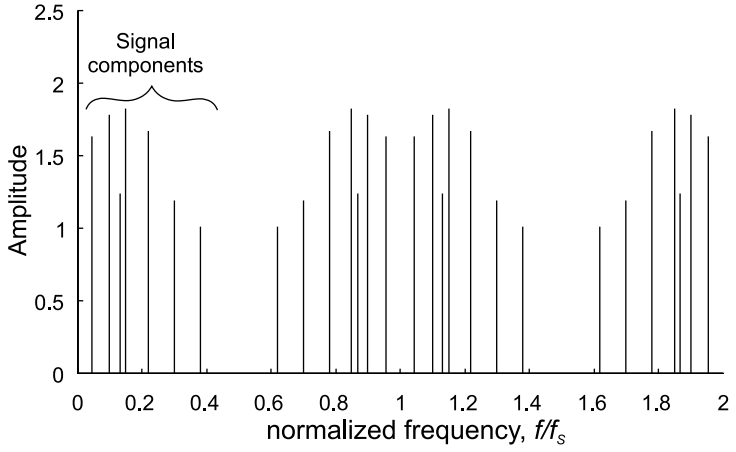


Figure 16.5 Illustration of rectangular function based DFT of a periodically undersampled signal

Figure 16.5 illustrates what happens when DFTs are carried out for a signal sampled periodically under aliasing conditions. In this particular case, the sampling rate is about two times smaller than the rate ensuring alias-free operation. Apparently full-scale aliasing takes place under the given conditions and the DFT results are distorted. To avoid this kind of frequency overlapping effect, the signal sample values have to be taken nonuniformly. Estimates of the α_i and β_i coefficients obtained when the same signal is sampled according to the additive sampling scheme are given in Figures 16.6(a) and (b) respectively. The end results of DFTs carried out under these specific conditions are shown in Figure 16.7.

The spectral estimates displayed in Figure 16.7 are much closer to the true signal spectrum than the estimates distorted by aliasing given in Figure 16.5. Randomization of sampling has clearly led to considerable suppression of the aliases and the amplitudes of the signal basic components are estimated relatively well. However, there is noticeable background noise not present in the input signal. Therefore this noise has appeared in the spectrogram due to imperfections of the algorithm used for the DFT. In this particular case the rectangular basis functions of system 1 have been used to reduce the number of multiplication operations needed for carrying out the DFT. Therefore the question arises: what is the real source of this background noise? Does it appear as a result of sampling randomization or do these errors appear because rectangular basis functions are used instead of exponential ones?

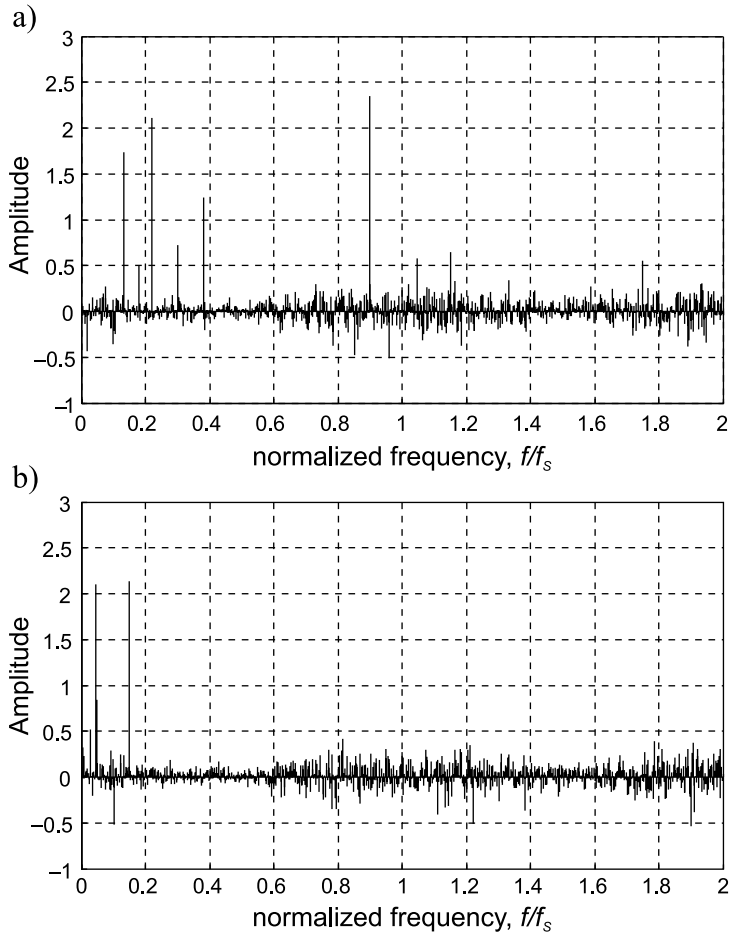


Figure 16.6 Estimates of (a) α_i coefficients and (b) β_i coefficients

To find the answer to these questions, DFTs of the given signal have been repeated by using the exponential basis functions under other equal conditions. The results are given in Figure 16.8 and are close to the estimates shown in Figure 16.7. This fact leads to the conclusion that the imperfections of the used algorithm for DFTs are related to insufficiently well-handled processing of the nonuniformly sampled signal rather than to the used rectangular basis functions.

Thus the computer simulations show that the estimates of the Fourier coefficients obtained as a result of the suggested method for the complexity-reduced DFT, in comparison with DFT using the exponential basis, are of the same level of quality. Application of both approaches leads to DFT results corrupted by

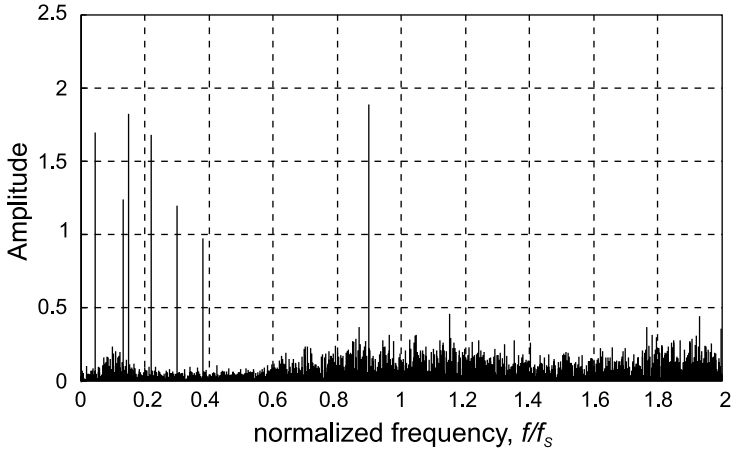


Figure 16.7 DFT of the test signal sampled according to the additive sampling scheme and performed on the rectangular function basis

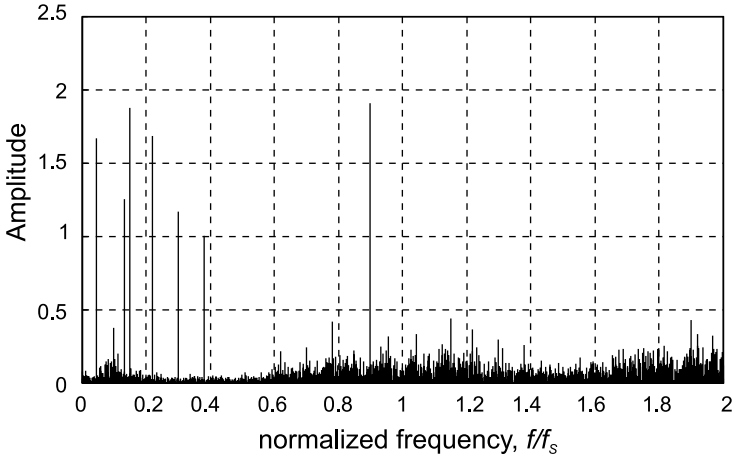


Figure 16.8 DFT of the test signal sampled according to the additive sampling scheme and performed on the exponential function basis

nonuniform sampling at a comparable level. Therefore in many cases it makes sense to simplify the DFT-related calculations in the described way. The obtained DFT results represent valuable data that are needed and could be successfully used for resolution of many signal processing tasks. With regards to the background noise, it is actually not a noise at all. The observed fluctuations can be tracked down to the cross-interference between signal components related to

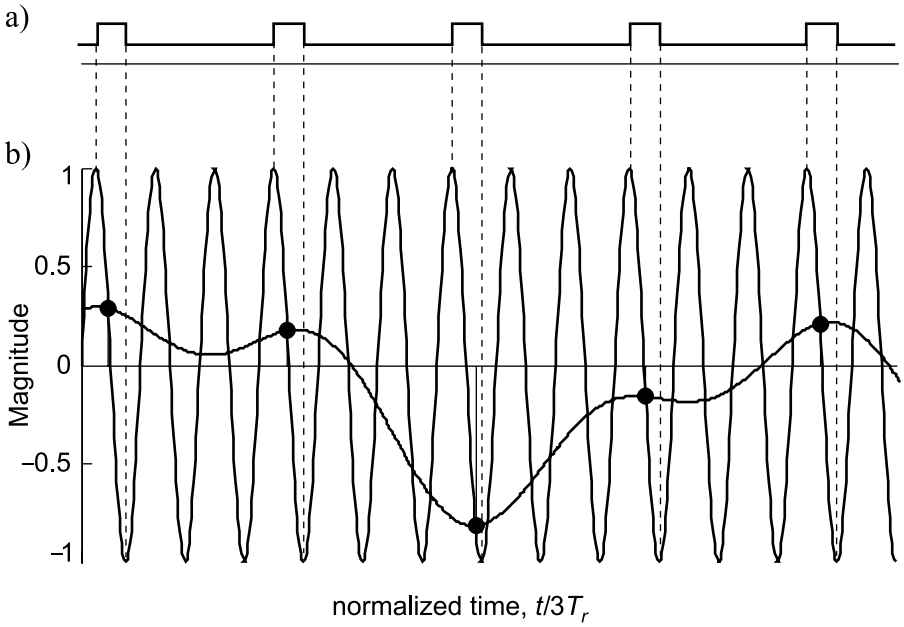


Figure 16.9 Illustration of the fact that sampling based on sine-wave crossings always results in a sequence of reference sine-wave sample values even if the signal sampled contains multiple components: (a) a function enabling the sampling operation; (b) a diagram of the sampling operation

specific sampling irregularities. In fact, these fluctuations appear as a result of the typical fuzzy aliasing effect described in Chapter 9. The problem of eliminating this fuzzy aliasing effect and the DFT errors due to it has been resolved and the techniques used to achieve this are discussed in Chapters 18, 19 and 20.

16.4 Fast DFT of Sine-Wave Crossings

Properties of digital signals obtained as a result of sampling realized as signal and reference sine-wave crossings substantially differ from the typical features of digital signals formed on the basis of the commonly used signal sample value taking at predetermined time instants. Figure 16.9 illustrates this type of sampling. A periodic pulse sequence, shown in Figure 16.9(a), is used to enable the sampling operation at a specific signal source periodically during half-periods of the reference function. The frequency of these enabling pulses depends on the number of signal sources connected to the respective system.

Attention is drawn to the amazing fact that when the sampling method based on the sine-wave crossings is used, the sampled signal represents a sequence of nonuniformly spaced sample values of the reference sinusoid no matter what spectrum of the input signal is involved. In consequence of this, a special approach to processing this kind of digital signal can be used. This approach leads to the complexity-reduced DFT not requiring multiplication of multidigit numbers.

As shown in Chapter 15, to be able to apply the standard DSP computer programs to process nonuniform digital signals obtained as a result of sine-wave crossings, sampling of these signals can often be regularized. When this is the case, the problem of effective processing of the sampled signals might be considered as resolved and nothing else needs to be done. The method used for the complexity-reduced DFT suggested here represents an alternative that can be used for conditions when the described regularization leads to large and unacceptable errors.

Whenever the sample values of the reference sine-wave function at the frequency f_r represent the sampled signal to be transformed, the expressions usually used for estimation of the Fourier coefficients at the ith frequency,

$$\begin{aligned} \hat{a}_i &= \frac{2}{N}\sum_{k=0}^{N-1} x_k \cos 2\pi f_i t_k, \\ \hat{b}_i &= \frac{2}{N}\sum_{k=0}^{N-1} x_k \sin 2\pi f_i t_k, \end{aligned} \tag{16.21}$$

assume the following form:

$$\begin{aligned} \hat{a}_i &= \frac{2}{N}\sum_{k=0}^{N-1} \sin 2\pi f_r t_k \cos 2\pi f_i t_k, \\ \hat{b}_i &= \frac{2}{N}\sum_{k=0}^{N-1} \sin 2\pi f_r t_k \sin 2\pi f_i t_k. \end{aligned} \tag{16.22}$$

Note that Equations (16.22) is given for $A_r = 1$, where A_r is the amplitude of the reference signal. The notations $\alpha = 2\pi f_r t_k$ and $\beta = 2\pi f_i t_k$ are now introduced. Then it can be written that

$$\begin{aligned} \hat{a}_i &= \frac{2}{N}\sum_{k=0}^{N-1} \sin\alpha \cos\beta, \\ \hat{b}_i &= \frac{2}{N}\sum_{k=0}^{N-1} \sin\alpha \sin\beta. \end{aligned} \tag{16.23}$$

Evidently Equations (16.23) can also be given as

$$\begin{aligned} \hat{a}_i &= \frac{1}{N}\sum_{k=0}^{N-1}[\sin(\alpha-\beta)+\sin(\alpha+\beta)], \\ \hat{b}_i &= \frac{1}{N}\sum_{k=0}^{N-1}[\cos(\alpha-\beta)-\cos(\alpha+\beta)], \end{aligned} \tag{16.24}$$

where

$$\begin{aligned} \alpha-\beta &= 2\pi(f_{\mathrm{r}}-f_i)\,t_k, \\ \alpha+\beta &= 2\pi(f_{\mathrm{r}}+f_i)\,t_k. \end{aligned}$$

Application of the definitions (16.24) is attractive as calculations carried out using them provide for fast DFTs as only a few multiplications need to be carried out, especially if calculations of $[\sin(\alpha-\beta)+\sin(\alpha+\beta)]$ and $[\cos(\alpha-\beta)-\cos(\alpha+\beta)]$ are carried out beforehand and the results are stored in a look-up table.

Therefore there are at least two options using fast algorithms to process the digital signals obtained as a result of sine-wave crossings. Firstly, the signals sampled in this way might often be regularized and then the fast DSP algorithms can be used. The second alternative is the approach discussed here. Application of the latter is preferable whenever the regularization errors are large.

Bibliography

Bilinskis, I. and Mikelsons, A. (1992) *Randomized Signal Processing*. Prentice-Hall International (UK) Ltd.

Bilinskis, I., Borovik, Yu. F. and Mikelsons, A. (1982) Use of rectangular periodic functions for computing discrete Fourier transforms (in Russian). *Autom. Control Comput. Sci*., **2**, 81–6.

Bilinskis, I., Borovik, Yu. F. and Mikelsons, A. (1982) Implementation of a DFT method based on the use of rectangular functions (in Russian). *Autom. Control Comput. Sci*., **6**, 81–8.

Bilinskis, I., Borovik, Yu. F. and Mikelsons, A. (1983) Complexity-reduced discrete Fourier transform. In *Signal Processing*, II: *Theories and Applications*. Amsterdam: Elsevier Science Publishers, pp. 734–46.

Bilinskis, I., Borovik, Yu. F. and Mikelsons, A. (1983) A digital Fourier transform method not requiring multiplication operations (in Russian). *Autom. Control Comput. Sci*., **1**, 79–86.

Harmuth, H.F. (1977) *Sequency Theory: Foundations and Application*. New York: Academic Press.

17

Spatial Data Acquisition and Processing

The topic of spatial signal processing is now approached to draw attention to the fact that the anti-aliasing signal processing technology described in this book was initially developed for handling signals in the time and frequency domains but is also applicable in the area of spatial signal processing vital for all types of radar systems. While various types of problems could be attacked on the basis of this technology, the following discussions are focused on the issue of the complexity reduction of large-aperture antenna arrays. Specifically, the potential of pseudo-randomization of large-aperture antenna arrays is considered and evaluated as an approach applicable for reducing the number of sensors in the array. To achieve good results in this area, it is suggested that the positive experience obtained at signal processing in the time domain can be adjusted to the specifics of spatial signal processing. The rationale of this approach is based on the fact that irregular taking of signal sample values in the time domain is equivalent to the irregular spacing of sensors in the array.

Although historically the problems of randomizing the temporal signal processing have received much more attention than the analogous problems of spatial data acquisition and processing, the potential usefulness of deliberate randomization in this area has not been overlooked. A number of publications provide evidence of this. However, at first glance, these publications have nothing to do with randomized processing of signals. They are mainly works dealing with radar antennae. The fact that a number of radar characteristics can be enhanced by randomizing their antenna designs had already been discovered in the early 1960s. It was found that when antenna sensor elements are spaced randomly rather than

Digital Alias-free Signal Processing I. Bilinskis

regularly, undesirable aliasing effects are reduced and other improvements in system functioning can be achieved.

A survey of the research results obtained in the field of array signal processing, the theory of which is relatively well developed, lies beyond the scope of this book. To establish the methodological, terminological and notation basis for the following discussions, a brief description of the principles and relationships relevant to the antenna arrays and array signal processing is given in Section 17.1. Typical spatial signal processing subsystems, usually included in the structures of these array systems, are considered.

Although antenna arrays are used for signal transmission or receiving, the central problem in array signal processing is beamforming. Proper beamforming is essential for concentrating signal energy in a required direction at their transmission and also allows an estimation to be made of the direction of signal arrival at their receiving. While transmission and receiving of signals seem to be distinctly different operations, the beamforming procedures in both cases have much in common. This allows just one of these cases to be considered and to extend the results of the analysis over both areas. Following this reasoning, the signal receiving case is selected as the basis for studies of the suggested specific beamforming techniques suitable for complexity reduction of the array signal processing systems. The discussions concern, first of all, various aspects of the signal direction of arrival (DOA) estimation.

17.1 Sensor Array Model

Consider a linear array of sensors shown in Figure 17.1(a). Suppose that the sensors are distortion-free, isotropic (capable of receiving signals from all directions equally well) and spaced equidistantly. To simplify consideration of the basic principles of such array functioning, suppose also that the signal impinged on the array from a far-field source is described as a narrowband signal with a common centre frequency f or as a nonmodulated sinusoidal carrier at the frequency f. Assume that the angle between the signal plane and the array axis is equal to Θ. In this situation, the signals received by sensors $2, 3, 4, \ldots$ are delayed with respect to the signal received by sensor 1 and these delays depend on Θ and the intervals $d, 2d, 3d, \ldots$ between the respective sensors. Indeed, the signal at sensor 2 is delayed for a time interval τ_2 during which the signal covers the distance $d \sin \Theta$ and as the signal propagation speed is equal to c, then $c\tau_2 = d \sin \Theta$ or $\tau_2 = (d/c) \sin \Theta$. For the kth sensor, the delay

$$\tau_k = \sin \Theta \frac{(k-1)d}{c}. \tag{17.1}$$

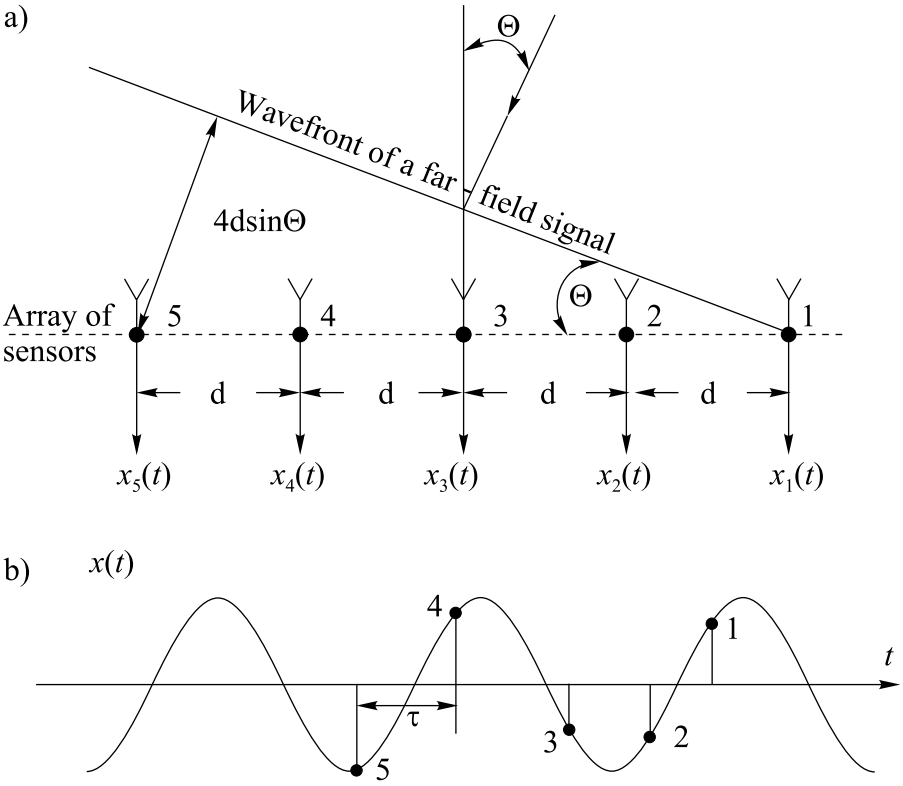


Figure 17.1 Basic array model: (a) delays in signal paths to various sensors depending on the signal arrival angle; (b) array signal formed from digital sample values taken off the sensors at their simultaneous sampling

As the phase shift between the output signals of sensors is $\varphi = -2\pi f\tau$ radians, the phase shift between the adjacent sensor outputs is given as

$$\varphi = -2\pi f \sin\Theta \frac{d}{c}. \tag{17.2}$$

Now suppose that all sensor outputs are sampled simultaneously and the respective signal values are read out in this way. Obviously, all these readings are from one and the same sinusoid and intervals between the sample values (Figure 17.1(b)) are equal to the delays mentioned. On the other hand, this sample sequence forms a discrete signal depending on the signal source frequency and location. This array signal

$$x_k = A \sin(2\pi\Omega d_k + \varphi), \quad k = 1, 2, 3, \ldots, K, \tag{17.3}$$

where A is the amplitude of the received signal, d_k is the coordinate of a sensor (with respect to sensor 1) and Ω and φ are the wavenumber (spatial frequency) and the array signal phase respectively. It can be seen from Figure 17.1 that $\Omega = \sin\Theta\,(f/c)$ and Equation (17.3) can be rewritten as

$$x_k = A\sin\left(\frac{2\pi f}{c}\sin\Theta d_k + \varphi\right), \quad k = 1, 2, 3, \ldots, K. \tag{17.4}$$

Now suppose that there are $m = 1, 2, 3, \ldots, M$ rather than one signal sources. In this case the received signals, arriving from various directions, add up so that the resulting sequence of the sensor output signal sample values, taken repeatedly at time instants t_n, can be considered as the following vector:

$$\boldsymbol{x}_n = \begin{bmatrix} \sum_{m=1}^{M} A_m \sin[2\pi\Omega_m d_{mn1} + \varphi_m + \Delta\varphi_m(\delta_n)] + \xi(d_{mn1}) \\ \sum_{m=1}^{M} A_m \sin[2\pi\Omega_m d_{mn2} + \varphi_m + \Delta\varphi_m(\delta_n)] + \xi(d_{mn2}) \\ \vdots \qquad\qquad \vdots \qquad\qquad \vdots \\ \sum_{m=1}^{M} A_m \sin[2\pi\Omega_m d_{mnN} + \varphi_m + \Delta\varphi_m(\delta_n)] + \xi(d_{mnN}) \end{bmatrix}, \tag{17.5}$$

where A_m, Ω_m and φ_m are amplitudes, wavenumbers and phases of the array signal components respectively, $\xi(d_{mnk})$ are the noise sample values and $\Delta\varphi_m(\delta_n)$ are the additional phase delays that appear in special cases when additional delays d_n are inserted. Note that when the sensor coordinates are denoted as d_{mnk}, it is presumed that this coordinate may change from one instant t_n to another. As that is usually not the case (the sensor positions are fixed), these coordinates can be denoted as d_{mk}.

In more general and specific cases, the signals impinged on an array should be considered as being wideband or narrowband rather than one sine function. Then, naturally, its definition is more complicated.

17.2 Temporal and Spatial Spectra of Array Signals

While the application area of arrays covers a large variety of tasks, solution of every one of them is based, in some way, on relationships of array signal temporal and/or spatial spectra. Basic relationships of this kind will be considered.

Suppose that the far-field signals impinged on a linear array of the type considered above are sufficiently well approximated by a finite number of complex

exponential functions. Suppose also that there are altogether m signal sources. Then the signal of the mth source can be written as

$$x_m(t) = \sum_{i=1}^{I_m} A_{mi} \sin(2\pi f_{mi} t + \varphi_{mi}), \tag{17.6}$$

where A_{mi}, f_{mi} and φ_{mi} are the amplitudes, frequencies and phases of the signal components respectively and I_m is the number of components in the signal emitted by the mth source.

The first sensor will be chosen as a reference point in time and in distance. Then, if the distance to the kth sensor is d_k and the arrival angle of the signal from the mth source is denoted by Θ_m, the component of the kth sensor output signal coming from the mth signal source can be given as

$$x_{mk}(t) = \sum_{i=1}^{I_m} A_{mi} \sin\left(2\pi f_{mi} t - 2\pi f_{mi} \frac{d_k \sin\Theta_m}{c} + \varphi_{mi}\right). \tag{17.7}$$

The complete output signal of the kth sensor, containing components coming from all signal sources, can then be defined as

$$\begin{aligned} x_k(t) &= \sum_{m=1}^{M} x_{mk}(t) \\ &= \sum_{m=1}^{M} \sum_{i=1}^{I_m} A_{mi} \sin\left(2\pi f_{mi} t - 2\pi f_{mi} \frac{d_k \sin\Theta_m}{c} + \varphi_{mi}\right) \\ &= \sum_{m=1}^{M} \sum_{i=1}^{I_m} \left[A_{mi} \sin\left(2\pi \frac{f_{mi} \sin\Theta_m}{c} d_k + \pi - \varphi_{mi}\right) \cos 2\pi f_{mi} t \right. \\ &\qquad \left. + A_{mi} \cos\left(2\pi \frac{f_{mi} \sin\Theta_m}{c} d_k - \varphi_{mi}\right) \sin 2\pi f_{mi} t \right]. \end{aligned} \tag{17.8}$$

The function $A \sin(2\pi f t + \varphi)$ can be represented in the following form:

$$A \sin(2\pi f t + \varphi) = a \cos 2\pi f t + b \sin 2\pi f t,$$

where $a = A \sin\varphi$ and $b = A \cos\varphi$.

The coefficients a and b are orthogonal projections of the sine function on $\cos 2\pi f t$ and $\sin 2\pi f t$. Suppose that a snapshot multitude $\{x_k(t_n)\}$, $n = \overline{1, N}$, has been obtained and that these data have been decomposed by applying some high-performance method and that parameters of the signal components as well as the orthogonal projections have been estimated as a result of temporal spectral

analysis with errors not exceeding some given margins. Then two essential cases can be distinguished.

17.2.1 When Signal Source Frequencies Do Not Overlap

Suppose that the ith frequency of the mth signal source does not overlap any other frequency of the other signal sources. Then the orthogonal projections corresponding to this frequency, which follows from Equation (17.8), are

$$\begin{aligned} a_{mi} &= A_{mi} \sin\left(2\pi \frac{f_{mi} \sin \Theta_m}{c} d_k + \pi - \varphi_{mi}\right), \\ b_{mi} &= A_{mi} \cos\left(2\pi \frac{f_{mi} \sin \Theta_m}{c} d_k - \varphi_{mi}\right). \end{aligned} \tag{17.9}$$

It can be seen from Equations (17.9) that these projections can be interpreted as spatial sinusoidal functions with sample values obtained at points d_k. To estimate the wavenumbers Ω_{mi}, spatial spectral analysis of data $\{a_{mi}, b_{mi}\}$ can be performed. According to Equation (17.8),

$$\Omega_{mi} = \frac{f_{mi} \sin |\Theta_m|}{c}. \tag{17.10}$$

It follows from Equation (17.10) that

$$|\Theta_m| = \arcsin \frac{c\Omega_{mi}}{f_{mi}}. \tag{17.11}$$

To determine the sign of Θ_m, Equations (17.9) are rewritten as

$$\begin{aligned} a_{mi} &= A_{mi} \sin \varphi_{mi} \cos 2\pi \frac{f_{mi} \sin \Theta_m}{c} d_k \\ &\quad + A_{mi} \cos(\pi - \varphi_{mi}) \operatorname{sgn}(\Theta_m) \sin 2\pi \frac{f_{mi} \sin|\Theta_m|}{c} d_k, \\ b_{mi} &= A_{mi} \cos \varphi_{mi} \cos 2\pi \frac{f_{mi} \sin \Theta_m}{c} d_k \\ &\quad + A_{mi} \sin \varphi_{mi} \operatorname{sgn}(\Theta_m) \sin 2\pi \frac{f_{mi} \sin |\Theta_m|}{c} d_k. \end{aligned} \tag{17.12}$$

Then the orthogonal projections a_{mi} and b_{mi} can be given in the following form:

$A_{mi} \sin \varphi_{mi}$	$A_{mi} \cos(\pi - \varphi_{mi}) \operatorname{sgn}(\Theta_m)$
$A_{mi} \cos \varphi_{mi}$	$A_{mi} \sin \varphi_{mi} \operatorname{sgn}(\Theta_m)$

(17.13)

Table 17.1 Bilinskis

$\Theta_m > 0$		$\Theta_m < 0$	
+	−	+	+
+	+	+	−
+	+	+	−
−	+	−	−
−	+	−	−
−	−	−	+
−	−	−	+
+	−	+	+

In order to determine the sign of Ω_m, Table 17.1 displays the signs of these projections at all possible positions of the phase angle φ_{mi} (it can be in any of the four quadrants). The amplitudes and phases of the respective sine functions can also be determined from the projections given in (17.13).

17.2.2 When Signal Source Frequencies Overlap

Suppose that some of frequencies in spectra of signals coming from different signal sources overlap. Assume that the frequency f_{mi} is common for all signal sources and denote it by f_1. Then the projections corresponding to this frequency, obtained in the course of the temporal spectral analysis as follows from Equation (17.8), can be written as

$$\begin{aligned} a_1 &= \sum_{m=1}^{M} A_{m1} \sin\left(2\pi \frac{f_1 \sin\Theta_m}{c} d_k + \pi - \varphi_{m1}\right), \\ b_1 &= \sum_{m=1}^{M} A_{m1} \cos\left(2\pi \frac{f_1 \sin\Theta_m}{c} d_k - \varphi_{m1}\right). \end{aligned} \tag{17.14}$$

As Θ_m differs for various signal sources, then projections a_1 and b_1 can be interpreted as M sums of sine functions at the following wavenumbers:

$$\Omega_m = \frac{f_1 \sin \Theta_m}{c}. \tag{17.15}$$

Now Equations (17.14) can be rewritten as

$$a_1 = \sum_{m=1}^{M} \left[A_{m1} \sin \varphi_{m1} \cos 2\pi \frac{f_1 \sin \Theta_m}{c} d_k + A_{m1} \cos(\pi - \varphi_{m1}) \operatorname{sgn}(\Theta_m) \sin 2\pi \frac{f_1 \sin |\Theta_m|}{c} d_k \right], \tag{17.16}$$

$$b_1 = \sum_{m=1}^{M} \left[A_{m1} \cos \varphi_{m1} \cos 2\pi \frac{f_1 \sin \Theta_m}{c} d_k + A_{m1} \sin \varphi_{m1} \operatorname{sgn}(\Theta_m) \sin 2\pi \frac{f_1 \sin |\Theta_m|}{c} d_k \right]. \tag{17.17}$$

To estimate wavenumbers Ω_m, $m = \overline{1, M}$, decomposition of a_1 and b_1 has to be carried out. As the frequency f_1 has been estimated at the stage of the temporal spectral analysis, then, by applying Equation (17.14),

$$|\Theta_m| = \arcsin \frac{c\Omega_m}{f_1}. \tag{17.18}$$

Amplitudes, phases and signs of the arrival angles Θ_m then can be estimated for all wavenumbers Ω_m in the same way as in the case considered above where frequencies do not overlap.

17.2.3 Aliasing in the Spatial Domain

Signal processing as described so far in this section is commonly used for spatial spectrum analysis of a signal received by an array of sensor elements. The temporal signal sampling is then periodic and the sensors in the array are placed equidistantly. Up to certain temporal and spatial frequencies this approach represents a good technical solution to the problem of spatial signal filtering. However, there are limitations. The upper frequency of the signal temporal spectrum is limited, as usual, by the sampling frequency used at the analog-to-digital conversions of the sensor output signals. Obviously this spectrum should not exceed a frequency equal to one-half of the sampling rate.

The minimal wavelength of array signal components is limited by the value d of the interval between the array elements. To guarantee that there is no aliasing,

at least two samples from a signal wavelength should be taken, i.e. $\lambda_{\min} \geq 2d$ and $\Omega_{\max} = 1/\lambda_{\min}$. If the wavenumber of an array signal component exceeds the given limit, aliasing occurs. Clearly, there is a row of indistinguishable wavenumbers that is analogous to the indistinguishable temporal frequencies. As the distance between the array signal digital values is equal to d, the following wavenumbers completely overlap and are accepted by regular arrays as one and the same:

$$\Omega_i, \frac{1}{d} \pm \Omega_i, \frac{2}{d} \pm \Omega_i, \frac{3}{d} \pm \Omega_i, \ldots. \tag{17.19}$$

The aliasing effect is well known and in array beam characteristics it is displayed as the so-called grating lobes. This effect has somehow to be avoided. Otherwise spectra of the signals and noise outside the alias-free range will overlap the spectra within this range and distort the results of signal processing.

The obvious way of enlarging the range free of spatial aliasing is to place sensors closer. However, if this approach is used then, depending on the number of sensors for a given aperture, the complexity of such arrays and of their signal processing systems is directly tied to the required wavenumber range of the spatial spectrum analysis to be performed. There are other problems as well. For instance, cross-interferences between sensors do not allow the sensors to be placed too close, thus limiting the achievable upper boundary of this range.

The fact that the inability to distinguish certain spatial signals is caused by the regularity of their sample value taking procedure was recognized a long time ago and there have been attempts to avoid such aliasing by spacing array elements randomly. However, so far they have not led to sufficiently good results. It seems that the main obstacle preventing the achievement of significant progress in this direction is lack of sufficiently well-developed special digital signal processing techniques that are successful in this area. That is the reason why this topic is included for consideration in this book. There is clearly an analogy between conditions for signal processing in the spatial and the time domains. This means that some part of the suggested advanced signal processing techniques based on nonuniform temporal signal sampling could also be exploited for alias-free processing of array signals. On the other hand, conditions for array signal processing are also quite specific. Therefore it is often not obvious how the positive experience gained in the area of processing nonuniformly sampled temporal signals could be transferred to the area of array signal processing. The discussions in the following sections should help to clarify this matter to some extent.

The potential possibility of array complexity reduction is based on the assumption that if the intervals between sensors in arrays are not equidistant and special anti-aliasing techniques are applied then these intervals between sensors do not

necessarily have to be less than half of the shortest wavelength of the array signal components. If that is the case, then the number of sensors in the array with the same aperture and complexity of the array, directly related to the quantity of used sensors, can be reduced. The central issue of this matter, of course, is elimination of aliasing.

Discussions of this issue will start by considering the spectrum analysis of spatial signals taken off a linear array of isotropic distortion-free sensor elements. Suppose that the obtained array signal reflects a signal impinged on the array that comes from a direction Θ and can be described as an unmodulated sinusoidal carrier. The estimate of the spatial spectrum for a single simultaneous reading of the sensor outputs can be written as

$$\hat{S}(\Omega_i) = \frac{2}{K} \sum_{k=1}^{K} x(d_k) \exp(-\mathrm{j}2\pi\Omega_i d_k), \tag{17.20}$$

where Ω_i is the considered wavenumber and K and d_k are the number and the coordinates of the sensors respectively. Suppose that the array signal is a sine function at one of the wavenumbers:

$$\Omega_l = r\,\Omega_\mathrm{s} \pm \Omega_i, \quad r = 0, 1, 2, \ldots, \tag{17.21}$$

where Ω_s is the array signal spatial sampling rate and is inversely proportional to the distance (mean distance) between the sensors. As the power spectrum does not depend on the signal phase, to simplify the expressions the signal phase can be omitted and Equation (17.20) can be rewritten as

$$\hat{S}(\Omega_i) = \frac{2}{K} \sum_{k=1}^{K} A_l \sin 2\pi(r\,\Omega_\mathrm{s} \pm \Omega_i) d_k \exp(-\mathrm{j}2\pi\Omega_i d_k). \tag{17.22}$$

Properties of this estimate depend to a large extent on the positions of the sensors. Consider the case where they are spaced equidistantly with the interval d between them. Then, apparently, $d_k = (k-1)d$, $\Omega_\mathrm{s} = 1/d$ and estimate (17.22) becomes

$$\hat{S}(\Omega_i) = \frac{2}{K} \sum_{k=1}^{K} A_l \sin 2\pi[r(k-1) \pm \Omega_i(k-1)d] \exp[-\mathrm{j}2\pi\Omega_i(k-1)d]. \tag{17.23}$$

This expression clearly shows that when the sensors are spaced equidistantly, the value of the power spectrum estimate at Ω_i is the same for all values of r. This means that they are the same for all array signal components at wavenumbers belonging to the row (17.21) if only the amplitudes Al of these components are the same.

The situation is quite different when the intervals between the sensors are not equal. Then the estimate (17.22) will differ at various wavenumbers (at various values of r), which actually means that under these conditions it should be possible to suppress the aliasing effects. This suggests that it should be possible to diminish the negative effect of aliasing by spacing the sensor elements irregularly. Consequently, by applying this technique of array irregularization, it should be possible to use a lesser number of sensors in arrays, because the upper boundary of the required wavenumber range is then no longer determined by the mean distance between the sensors. The computer simulation results given below confirm this presumption.

17.3 Beamforming

Beamforming for arrays is the most important function allowing signal energy to be concentrated in a specific direction or to tune the array in a given direction so that it becomes most sensitive to signals coming from that direction while suppressing all signals and noise received from other directions. Figure 17.2 illustrates the classical direct method for beamforming in the case of signal reception.

When a signal is impinged on a linear array from a direction characterized by the signal arrival angle Θ, as shown in Figure 17.2(a), this signal is received by separate sensors delayed for time intervals τ_2, τ_3, τ_4, τ_5 ($\tau_1 = 0$). If the delays in the sensor channels are equalized by inserting additional delays, as shown in Figure 17.2(b), and then a reading of the sensor output values is taken simultaneously, the same spatial signal sample value will be obtained at all sensors. Therefore, these outputs can be summed as shown and amplification of the signal, received from the direction to which the array has been tuned by inserting the indicated specific delays, will result while signals coming from other directions will be suppressed.

It can be shown that the signal at the output of the adder is given as

$$\begin{aligned} y(t) &= \frac{1}{K}\sum_{k=0}^{K-1} A\sin\left[2\pi f\left(t - \frac{d_k\sin\Theta'}{c} + \frac{d_k\sin\Theta}{c}\right) + \varphi\right] \\ &= \frac{1}{K}\sum_{k=0}^{K-1} A\sin\left(2\pi f t + \varphi + 4\pi\frac{f d_k}{c}\cos\frac{\Theta+\Theta'}{2}\sin\frac{\Theta-\Theta'}{2}\right) \\ &= A^*\sin(2\pi f t + \varphi^*), \end{aligned} \tag{17.24}$$

where

$$A^* = \sqrt{a^2 + b^2}, \qquad \varphi^* = \operatorname{arctg}\frac{a}{b},$$

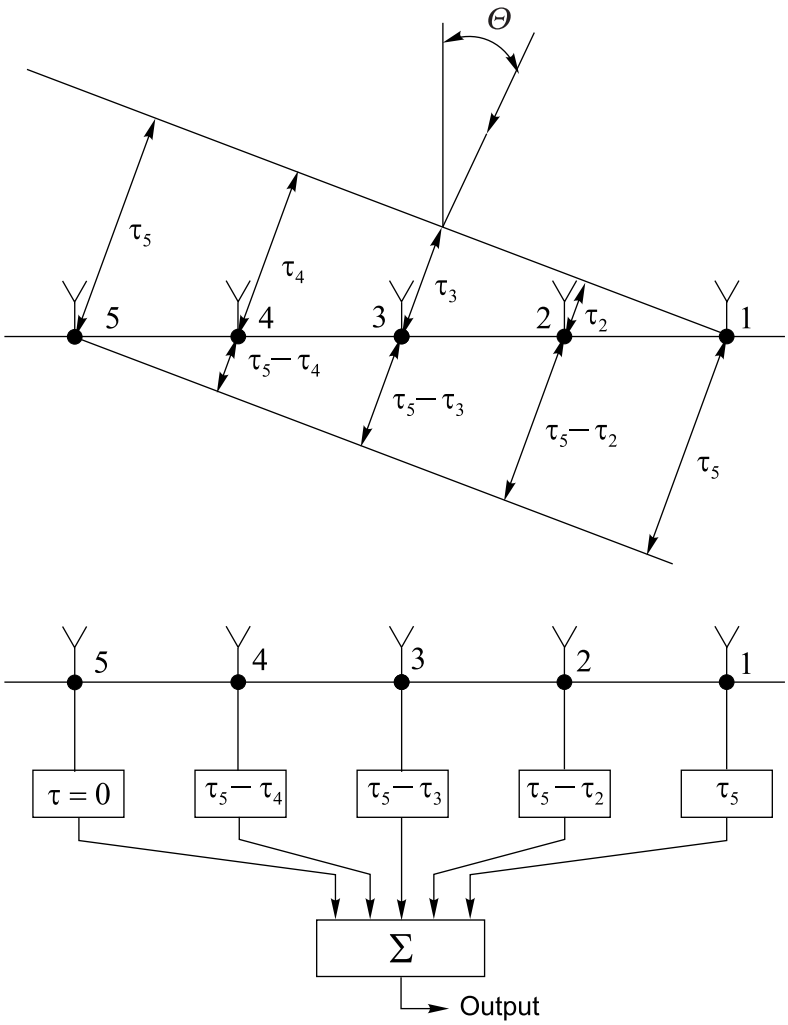


Figure 17.2 Illustration of the direct beamforming principle

and, in turn,

$$a = \frac{1}{K}\sum_{k=0}^{K-1} A \sin\left(4\pi\frac{f d_k}{c}\cos\frac{\Theta+\Theta'}{2}\sin\frac{\Theta-\Theta'}{2}+\varphi\right),$$

$$b = \frac{1}{K}\sum_{k=0}^{K-1} A \cos\left(4\pi\frac{f d_k}{c}\cos\frac{\Theta+\Theta'}{2}\sin\frac{\Theta-\Theta'}{2}+\varphi\right),$$

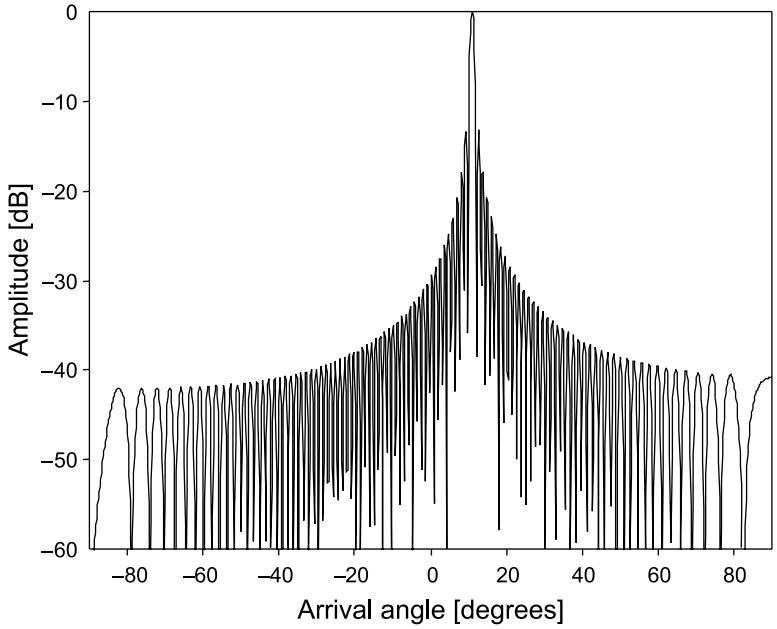


Figure 17.3 Directional pattern of an array of 128 equidistantly spaced sensors ($\Theta = 11^\circ$)

where Θ is the signal arrival angle and Θ' is the angle to which the array has been tuned. The directional pattern of such an array, containing 128 equidistantly spaced sensors, is shown in Figure 17.3. The array is tuned to $\Theta = 0^\circ$. The directional pattern is obtained on the basis of Equation (17.24) by varying the signal arriving angle. Signals arriving from angles differing from Θ are significantly suppressed.

To see what happens under other conditions the sensor number in the same aperture is reduced to 32. In this case the array directional pattern is as shown in Figure 17.4. It can be seen that aliasing occurs and corresponding grating lobes appear. To avoid aliasing and suppress the grating lobes in this way, the sensors should be spaced in the array irregularly. The obtained directional pattern of such an array is given later in Figure 17.9 and is commented on in Section 17.5.

17.4 Signal Direction of Arrival Estimation

To estimate an array signal direction of arrival (DOA) and the received signal parameters, both temporal and spatial spectral analyses have to be carried out. When this kind of signal processing is considered, it soon becomes clear that there are several possible ways of approaching the task of the DOA.

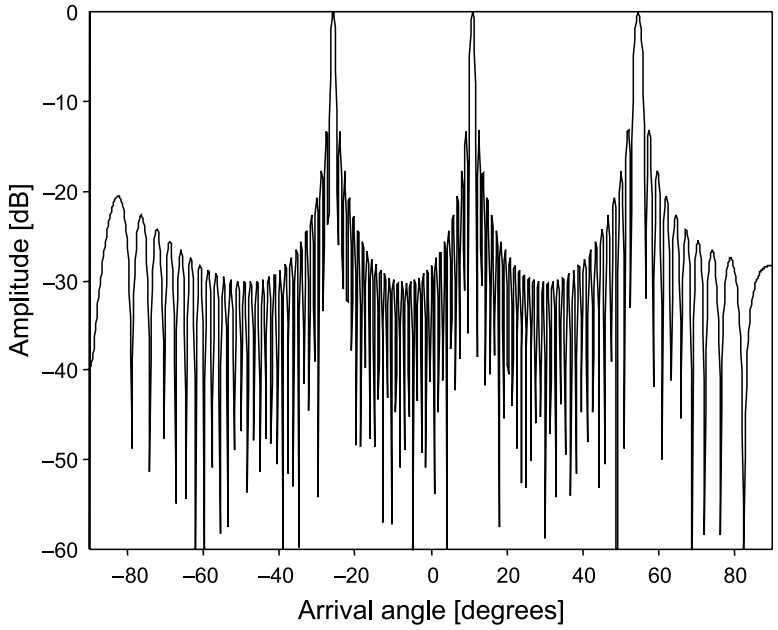


Figure 17.4 Directional pattern of a regular array of 32 sensors under aliasing conditions ($\Theta = 11^\circ$)

To simplify the following analysis and taking into account the fact that for DOA estimation a single sine function model of signals is applicable, the basic spectral analysis approaches are considered in the case where there are M signal sources and the signal $x(t)$ received by all sensors can be given as

$$x(t) = \sum_{m=1}^{M} A_m \sin(2\pi f_m t + \varphi). \tag{17.25}$$

Then the signal sample value x_{kn} taken off the kth sensor at the instant t_n is described by the following equation:

$$x_{kn} = \sum_{m=1}^{M} A_m \sin\left(2\pi f_m t_n + \varphi_m - 2\pi \frac{f_m \sin\Theta}{c} d_k\right). \tag{17.26}$$

It can be seen that there are two variables: t_n and Θ. One of them can be taken as fixed and the other can be varied. Hence there are two possible ways to deal with the analysis of this equation. They are illustrated by Figures 17.5(a) and (b). Both approaches will be considered.

Suppose that time is considered as a variable and Θ is fixed. In other words, the changing time sensor output signals are considered separately for each sensor.

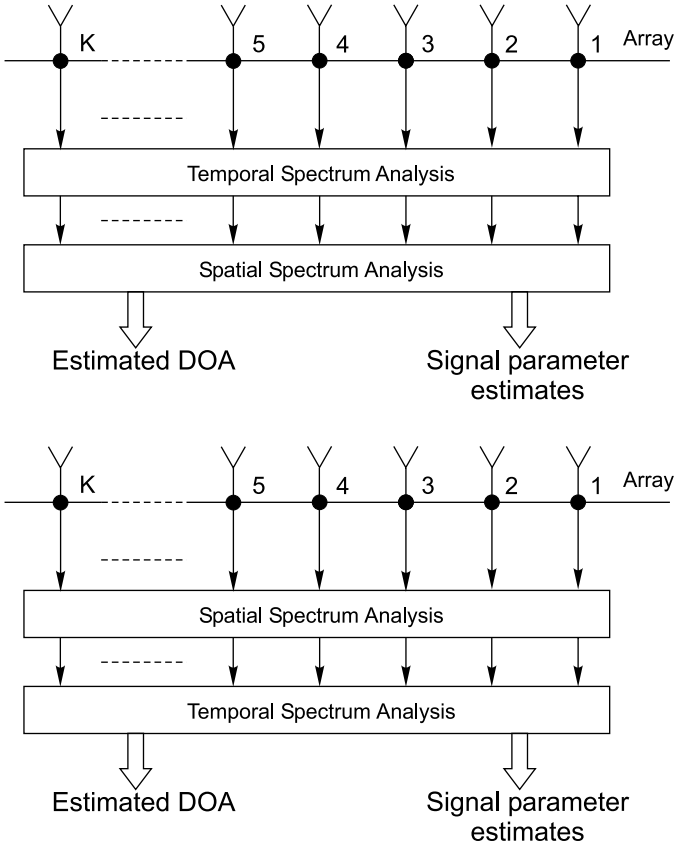


Figure 17.5 Two possible approaches to an estimation of the DOA

This approach corresponds to the scheme given in Figure 17.5(a) and means that the temporal spectral analysis is performed first and the spatial spectral analysis is carried out by processing the sequence of the Fourier coefficients obtained in the first step. Indeed, equation (17.26) can be rewritten as follows:

$$\begin{aligned} x_{kn} &= \sum_{m=1}^{M} \left[A_m \sin\left(2\pi \frac{f_m \sin\Theta_m}{c} d_k + \pi - \varphi_m\right) \cos 2\pi f_m t_n \right. \\ &\qquad \left. + A_m \cos\left(2\pi \frac{f_m \sin\Theta_m}{c} d_k - \varphi_m\right) \sin 2\pi f_m t_n \right] \\ &= \sum_{m=1}^{M} (a_{km} \cos 2\pi f_m t_n + b_{km} \sin 2\pi f_m t_n), \end{aligned} \tag{17.27}$$

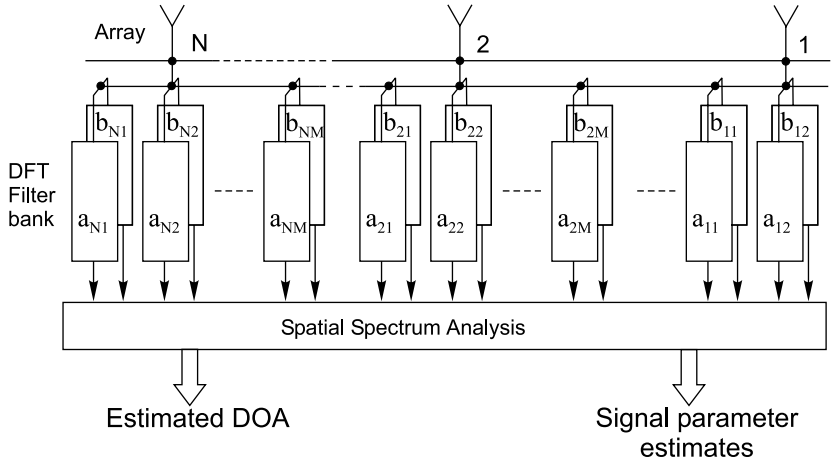


Figure 17.6 Implementation of the DOA estimation scheme on the basis of a DFT filter bank used at the temporal spectrum analysis stage

where

$$\begin{aligned} a_{km} &= A_m \sin\left(2\pi \frac{f_m \sin \Theta_m}{c} d_k + \pi - \varphi\right), \\ b_{km} &= A_m \cos\left(2\pi \frac{f_m \sin \Theta_m}{c} d_k - \varphi\right). \end{aligned} \tag{17.28}$$

Thus the coefficients a_{km} and b_{km} are time independent and for each frequency f_m these coefficients represent sample values of the corresponding sinusoidal spatial signal. This analysis approach can be directly implemented on the basis of the DFT by applying DFT filter banks as shown in Figure 17.6.

This approach can also be illustrated in a slightly different way, as shown in Figure 17.7. The interpretation is based on the sensor output signal sample value matrix

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{21} & x_{31} & \cdots & x_{K1} \\ x_{12} & x_{22} & x_{32} & \cdots & x_{K2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{1N} & x_{2N} & x_{3N} & \cdots & x_{KN} \end{pmatrix}, \tag{17.29}$$

which is formed from the sample value sequences obtained by repeated sampling of the sensor output signal multitude (by taking repeated snapshots). Then the data from this matrix are used to perform DFTs in each sensor signal processing

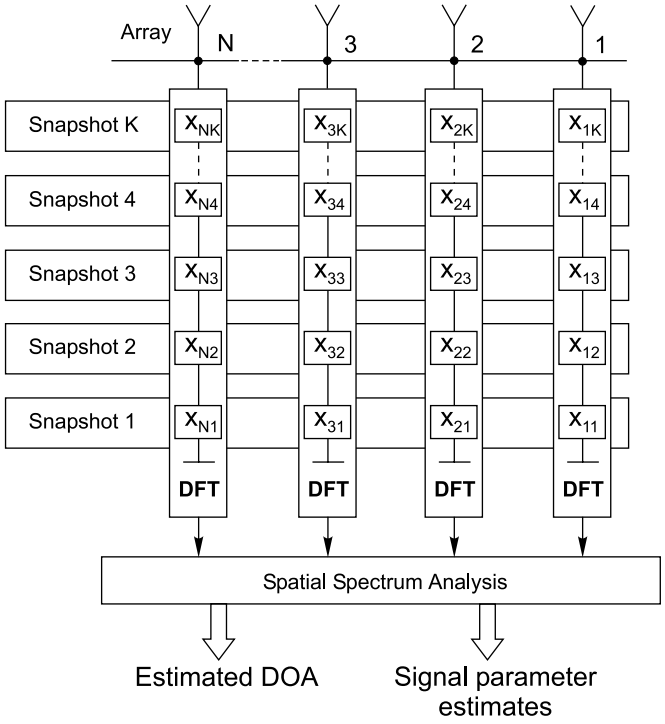


Figure 17.7 Generalized outline for the DOA estimation according to which the temporal spectrum analysis is carried out first

channel. The Fourier coefficients obtained in this way are later used as inputs for the spatial spectrum analysis.

Now the signal arrival angle Θ is considered as a variable and the time is fixed. In this case equation (17.26) can be rewritten as

$$\begin{aligned} x_{kn} &= \sum_{m=1}^{M} A_m \sin\left(2\pi \frac{f_m \sin\Theta_m}{c} d_k + \pi - 2 f_m t_n - \varphi_m\right) \\ &= \sum_{m=1}^{M} \left(a^*_{mn} \cos 2\pi\Omega_m d_k + b^*_{mn} \sin 2\pi\Omega_m d_k\right), \end{aligned} \tag{17.30}$$

where

$$\begin{aligned} a^*_{mn} &= A_m \sin(2\pi f_m t_n + \varphi_m), \\ b^*_{mn} &= A_m \cos(2\pi f_m t_n + \varphi_m - \pi. \end{aligned} \tag{17.31}$$

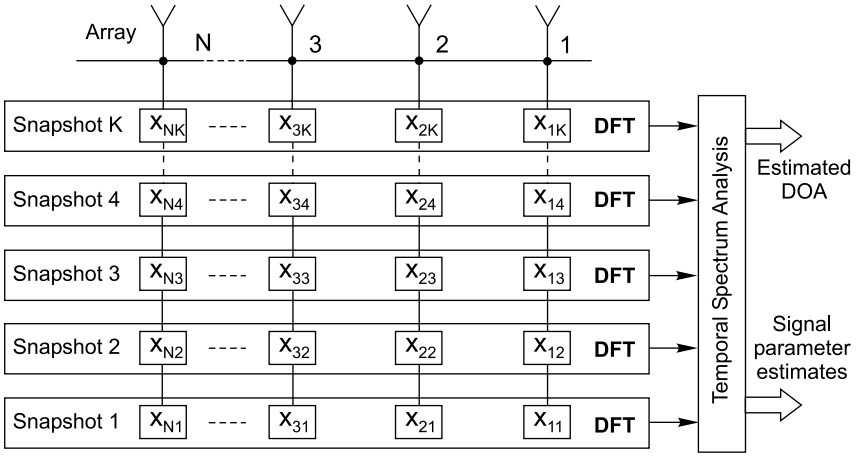


Figure 17.8 Generalized outline for the DOA estimation according to the which the DFT for the spatial spectrum analysis is calculated after taking each snapshot

It can be seen that now the coefficients a^*_{mn} and b^*_{mn} do not depend on the number of sensors. Therefore the spatial spectrum analysis, performed by processing the signal (17.30), will provide the estimates of $\{\Omega_m, a^*_{mn}, b^*_{mn}\}$, $m = 1, \ldots, M$. On the other hand, a^*_{mn} and b^*_{mn} can be considered as signals depending on the argument t_n. Note that these signals are sinusoidal. By performing a temporal spectrum analysis of them, the values of f_m and A_m, $m = 1, \ldots, M$, are estimated. This approach to the array signal analysis is illustrated by Figure 17.8.

Consider the schemes given in Figures 17.7 and 17.8 which correspond to the cases briefly described above. It can be seen that the input data block in both cases is the same. Therefore an impression may be found that the differences between them are purely formal. However, that is not true. While the end result is indeed the same, the sequence of procedures carried out in order to obtain that result differs and so do the algorithms developed on the basis of one or the other approach. For instance, in the case illustrated by Figure 17.8, the DFT can be performed by processing the data obtained at each snapshot sequentially in time (by applying only one filter bank) while, in the case of schema given in Figure 17.7, the performance of the DFT operations is delayed until all snapshots are taken. They are then either repeated sequentially in time by one device or N such DFT devices are required to perform these transforms parallel in time for all sensor signal processing channels. The preferable approach obviously depends on the conditions under which a specific array has to function.

17.5 Pseudo-randomization of Sensor Arrays

Typically there are many sensors in the sensor arrays used in radar systems. Each of them represents a signal source and the whole multitude of signals taken off the sensors has to be processed in real time. Therefore the sensor arrays and the attached systems fulfilling the task of array signal processing are quite complicated. The problem is that they cannot be simplified without changing the very essence of the array functioning. Indeed, if, for example, the classical one-dimensional array is examined, its spatial resolution is inversely proportional to the aperture. To achieve a sufficiently high resolution, this aperture has to be large. K sensors are placed within this aperture with the distance between the sensors equal to d so that the aperture is equal to $(K-1)d$. On the other hand, the minimal wavelength of array signal components is limited by the value d of the interval between the array elements. To guarantee that there is no aliasing, at least two samples from a signal wavelength should be taken, i.e. $\lambda_{\min} \geq 2d$ and $\Omega_{\max} = 1/\lambda_{\min}$. If the wavenumber of an array signal component exceeds the given limit, aliasing occurs. To satisfy both requirements of a sufficiently large aperture and small enough intervals between the sensors, the number of sensors in the array often has to be quite large and, consequently, the complexity of the array is then high.

17.5.1 Complexity Reduction of Arrays

Pseudo-randomization of the sensor array designs offers a way out of the described deadlock. Basically it is suggested that spatial aliasing can be avoided by placing sensors in the array nonuniformly. Assuming that this works, the mean distance between the sensors could then be enlarged and the number could then be significantly reduced.

This option for array complexity reduction certainly looks attractive. The question is how realistic it is. The directional pattern of a randomized array of 32 sensors is given in Figure 17.9. The first impression of the efficiency of sensor array randomization can be obtained by comparing Figures 17.4 and 17.9. It can be seen that in principle the grating lobes are indeed suppressed in the diagram in Figure 17.9. However, the obtained directional pattern of the nonuniform sensor array is clearly not very good. The question is: what is the meaning of this diagram and to what conclusions does it lead?

The well-defined aliasing taking place at functioning of the equidistant sensor arrays in the case of the nonuniform array is replaced by fuzzy spatial aliasing. The concept of fuzzy aliasing, observed at spectral analysis of signals in the time

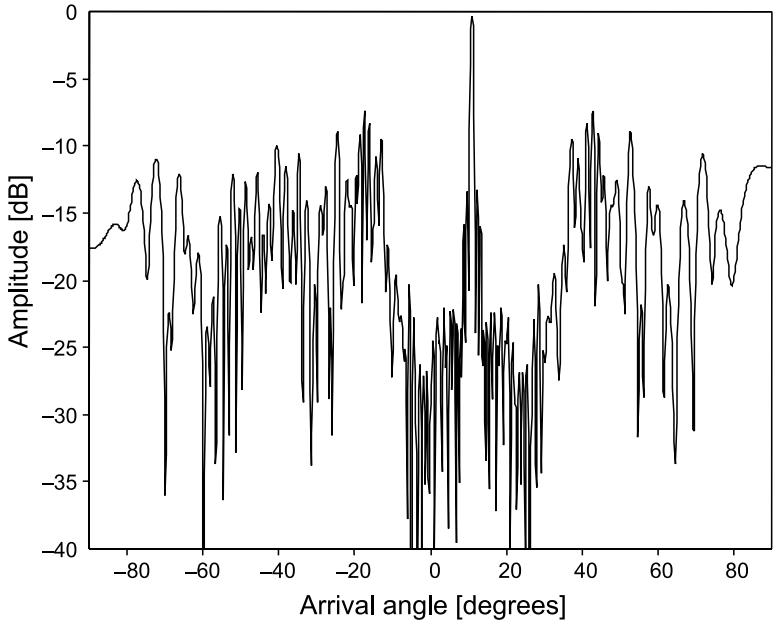


Figure 17.9 Directional pattern of a nonuniform array of 32 sensors ($\Theta = 11^{\circ}$).

domain, is explained in Chapter 9. To put this simply, the phenomenon is caused by the nonuniformities of the intervals between signal sample value taking instants. The mean sampling rate due to these irregularities then drift in time, which leads to aliasing occurring at varying frequencies. As a result, the aliasing pattern in the frequency domain becomes fuzzy. In other words, the sharp lines indicating aliasing at definite frequencies are spread out, showing that aliasing is suppressed but occurs in a certain frequency band.

Spatial fuzzy aliasing is analogous to fuzzy aliasing occurring at signal sampling in the time domain. The reason why fuzzy aliasing takes place in the spatial domain is the nonuniformity of the intervals between sensors in the array. The spatial sampling rate due to this nonuniformity is not constant, which leads to spreading out of the spatial aliasing over varying signal arrival angles.

Apparently that is not what is needed. Successful elimination of the aliasing effect really means avoiding aliasing rather than weakening it at certain frequencies and spreading the aliasing effect out into a broader frequency band. Therefore the displayed directional pattern of the pseudo-randomized array shown in Figure 17.9 should not be considered as an acceptable one.

The relatively bad results should not be surprising as they have been obtained simply by summing nonuniformly delayed spatial signal sample values. The problems encountered are similar to those observed at the DFT of nonuniformly sampled temporal signals and commented on in the introduction to Chapter 15. As soon as the sensors are placed in an array nonuniformly, the spatial signals taken off this array depend on the pattern of sensor coordinates in the array. This means that direct summing of nonuniformly delayed spatial signal sample values does not complete the process of beamforming. Obtaining this sum actually leads to the acquisition of intermediate signal processing results containing valuable information. Therefore this output signal $y(t)$ of the phased nonuniform array considered at various signal arriving angles under these conditions should not be automatically regarded as the directional pattern of the respective array. Additionally processing of $y(t)$ has to be undertaken with the specifics of the given sensor array nonuniformity taken into account.

A particular approach to resolution of this task is considered in Chapter 20. As shown there, adapting array signal processing to the nonuniformities of the sensor location in the array leads to much better beamforming. The directional pattern of the nonuniform array obtained as a result of such adaptation is incomparably better than the diagram displayed in Figure 17.9. This leads to the optimistic conclusion that it is possible to achieve significant complexity reduction of sensor arrays by pseudo-randomizing the array designs and performing appropriate array signal processing. The key to the success in this area is in performing effective alias-free signal processing.

17.5.2 Pseudo-randomization of Array Signal Processing

Pseudo-randomization of sensor arrays, in general, concerns both the array design (location of sensors in the array) and the array signal processing. To achieve good results at applications of the considered pseudo-randomized sensor arrays, it is crucial to ensure that processing of the signals taken off the nonuniform sensor arrays is carried out in an appropriate way. There is much in common in the approaches to temporal and spatial array signal processing.

At first glance it seems that only spatial signal processing requires development of special algorithms matched to the specifics of nonuniform sensor arrays and that processing of the temporal signals could be realized in the conventional manner. Basically this is true. On the other hand, various benefits can often be gained by also using special signal processing algorithms for handling the temporal signals. Of course, much depends on the characteristics of the signals transmitted or received by a particular sensor array. While periodic sampling

and the traditional algorithms for processing temporal signals is preferable when dealing with relatively narrowband signals, the technology for alias-free signal processing discussed in this book will lead to better results in cases where the signals are wideband and/or contain components at very high frequencies.

As shown in the block diagrams of the systems used to estimate the DOA, common techniques for digital signal processing are related to the DFT performed in both domains. As a lot of attention is paid in this book to realization of these transforms, a number of the discussed techniques and algorithms are also applicable in the area of array signal processing. The nonorthogonality appearing in the basis functions used for the DFT due to irregularities of the sampling process, both in the temporal and spatial domains, might be mentioned as an example. In fact, much of the knowledge accumulated in the area of temporal signal processing could be successfully exploited for enhancement of array signal processing in the spatial domain. This is demonstrated in the next chapter, where the algorithm for adapting processing of temporal signals to the nonuniformities of sampling is modified to extend the applicability of this approach so that it can also be used for adapting spatial signal processing to the specific nonuniform pattern of sensor locations in an array. As the results obtained and displayed there show, this approach, based on the experience gained from processing temporal signals, leads to significant improvement in pseudo-randomized array performance. An example showing what can be achieved is given in Chapter 20.

As the subject of pseudo-randomized arrays is really outside the scope of this book, the brief description given here is intended to draw the attention of readers to this exciting signal processing area where many of the special alias-free signal digitizing and processing techniques considered in this book can be successfully applied.

Bibliography

Bilinskis, I. and Lagunas, M.A. (1992) Randomizing of array element spacing and of processing array signals. Submitted to the Sixth European Signal Processing Conference, Brussels, 25–28 August 1992.

Bilinskis, I. and Mikelsons, A. (1992) *Randomized Signal Processing*. Prentice-Hall International (UK) Ltd.

Lo, Y.T. (1964) Mathematical theory of antenna arrays with randomly spaced elements. *IEEE Trans. Antennas Propag.*, **AP-12**, 257–68.

Lo, Y.T. and Simcoe, R.J. (1967) An experiment on antenna arrays with randomly spaced elements. *IEEE Trans. Antennas Propag.*, **AP-15**, 231–5.

Marvasti, F. (2001) *Nonuniform Sampling, Theory and Practice*. New York: Kluwer Academic/Plenum Publishers.

Steinberg, B.D. (1976) *Principles of Aperture and Array System Design*. New York: John Wiley & Sons, Inc.

18

Adapting Signal Processing to Sampling Nonuniformities

Introducing irregularities into a sampling process leads to the nonorthogonality of nonuniformly sampled discrete basis functions. If the DFT is performed on such a basis this nonorthogonality leads to significant errors in the estimation of signal parameters. Apparently the pattern of the nonuniform sampling point sequence defines this nonorthogonality and the errors related to it. However, at intentional pseudo-randomization of sampling this pattern is given *a priori*. Therefore it should be possible to use this information to suppress the errors caused by sampling nonuniformities. In other words, it should be possible to adapt processing of nonuniformly sampled signals to the involved specific sampling nonuniformity, which should lead to significantly better nonuniform signal processing results. However, it is not clear how to achieve this. One approach to this problem is suggested in Chapter 15 and is further considered here. This type of adapted signal processing is discussed for applications requiring processing of both the temporal and spatial signals.

18.1 Cross-interference Coefficients

As shown in Section 15.2, irregularities of the sampling point stream lead to cross-interference between the signal components. It is hard to overestimate the role this effect plays in processing nonuniformly sampled signals. There is no doubt that it is impossible to achieve high precision at processing this type of digital signal without taking this cross-interference into account in one way or another. Before discussing various options of how this could be done, the essence of this cross-interference will be considered.

Digital Alias-free Signal Processing I. Bilinskis

18.1.1 Definition

Suppose a signal $x(t_i)$ is sampled periodically with some quantity of the signal samples missing at random. If it is assumed that the signal sampled in this way contains M components, it can be described as follows:

$$x(t_k) = \sum_{m=0}^{M-1} [a_m \cos(2\pi f_m t_k) + b_m \sin(2\pi f_m t_k)] \tag{18.1}$$

where $t_k \in \{t_k\}$, $k = \overline{0, N-1}$, is the pseudo-randomly decimated periodic sampling instant stream. An estimation of the Fourier coefficients for signal $x(t_k)$ on frequency f_i might be performed on the basis of the usually applied formulae:

$$\begin{aligned} \hat{a}_i &= \frac{2}{N} \sum_{k=0}^{N-1} x(t_k) \cos(2\pi f_i t_k), \\ \hat{b}_i &= \frac{2}{N} \sum_{k=0}^{N-1} x(t_k) \sin(2\pi f_i t_k). \end{aligned} \tag{18.2}$$

The problem is that these estimates are corrupted by errors due to the irregularities of the sample taking process. Therefore the question is: what could be done to improve this kind of spectral estimation? An attempt will be made to find an answer to this question. Substitution of Equation (18.1) into Equations (18.2) leads to the following equations:

$$\begin{aligned} \hat{a}_i &= \frac{2}{N} \sum_{k=0}^{N-1} \sum_{m=0}^{M-1} [a_m \cos(2\pi f_m t_k) + b_m \sin(2\pi f_m t_k)] \cos(2\pi f_i t_k), \\ \hat{b}_i &= \frac{2}{N} \sum_{k=0}^{N-1} \sum_{m=0}^{M-1} [a_m \cos(2\pi f_m t_k) + b_m \sin(2\pi f_m t_k)] \sin(2\pi f_i t_k). \end{aligned} \tag{18.3}$$

Equations (18.3) might also be rewritten as follows:

$$\begin{aligned} \hat{a}_i &= \sum_{m=0}^{M-1} \left[a_m \frac{2}{N} \sum_{k=0}^{N-1} \cos(2\pi f_m t_k) \cos(2\pi f_i t_k) \right. \\ &\quad \left. + b_m \frac{2}{N} \sum_{k=0}^{N-1} \sin(2\pi f_m t_k) \cos(2\pi f_i t_k) \right], \\ \hat{b}_i &= \sum_{m=0}^{M-1} \left[a_m \frac{2}{N} \sum_{k=0}^{N-1} \cos(2\pi f_m t_k) \sin(2\pi f_i t_k) \right. \\ &\quad \left. + b_m \frac{2}{N} \sum_{k=0}^{N-1} \sin(2\pi f_m t_k) \sin(2\pi f_i t_k) \right]. \end{aligned} \tag{18.4}$$

These equations interpret the cross-interference coefficients characterizing interference between the signal components introduced in Chapter 15. Therefore expressions (18.4) can be given in the following form:

$$\begin{aligned}\hat{a}_i &= \sum_{m=0}^{M-1} [a_m(A_iC_m) + b_m(A_iS_m)], \quad i = \overline{0, M-1},\\ \hat{b}_i &= \sum_{m=0}^{M-1} [a_m(B_iC_m) + b_m(B_iS_m)], \quad i = \overline{0, M-1},\end{aligned} \tag{18.5}$$

where the coefficients

$$\begin{aligned}(A_iC_m) &= \frac{2}{N}\sum_{k=0}^{N-1} \cos(2\pi f_m t_k)\cos(2\pi f_i t_k),\\ (B_iC_m) &= \frac{2}{N}\sum_{k=0}^{N-1} \cos(2\pi f_m t_k)\sin(2\pi f_i t_k),\\ (A_iS_m) &= \frac{2}{N}\sum_{k=0}^{N-1} \sin(2\pi f_m t_k)\cos(2\pi f_i t_k),\\ (B_iS_m) &= \frac{2}{N}\sum_{k=0}^{N-1} \sin(2\pi f_m t_k)\sin(2\pi f_i t_k).\end{aligned} \tag{18.6}$$

These coefficients, reflecting the impact of the sampling imperfections, are actually the weights of the errors that corrupt the estimation of a Fourier coefficient a_i (or b_i) at frequency f_i and are related to the sampling nonuniformities of the sine (or cosine) component present in the signal at frequency f_m. Another set of cross-interference coefficients, specifically the coefficients A_mC_i, B_mC_i, A_mS_i and B_mS_i, characterize interference acting in the inverse direction from the signal component at frequency f_i to the component at frequency f_m. It follows from Equations (18.6) that

$$\begin{aligned}A_iC_m &= A_mC_i, \qquad B_iC_m = A_mS_i,\\ A_iS_m &= B_mC_i, \qquad B_iS_m = B_mS_i.\end{aligned} \tag{18.7}$$

Therefore it is not necessary to calculate the coefficients A_mC_i, B_mC_i, A_mS_i and B_mS_i on the basis of formulae similar to Equations (18.6), which is a great help.

18.1.2 Interpretation

To get a better idea of exactly how the sampling irregularities impact processing of the nonuniformly sampled signals, Equations (18.6) defining the

cross-interference coefficients could be converted and given as

$$\begin{aligned} A_i C_m &= D_\mathrm{c} + V_\mathrm{c}, \qquad & B_i C_m &= D_\mathrm{s} + V_\mathrm{s}, \\ A_i S_m &= V_\mathrm{s} - D_\mathrm{s}, \qquad & B_i S_m &= D_\mathrm{c} - V_\mathrm{c}, \end{aligned} \tag{18.8}$$

where

$$\begin{aligned} D_\mathrm{c} &= \frac{1}{N}\sum_{k=0}^{N-1} \cos 2\pi(f_i - f_m)t_k, \qquad & V_\mathrm{c} &= \frac{1}{N}\sum_{k=0}^{N-1} \cos 2\pi(f_i + f_m)t_k, \\ D_\mathrm{s} &= \frac{1}{N}\sum_{k=0}^{N-1} \sin 2\pi(f_i - f_m)t_k, \qquad & V_\mathrm{s} &= \frac{1}{N}\sum_{k=0}^{N-1} \sin 2\pi(f_i + f_m)t_k. \end{aligned} \tag{18.9}$$

Note that all the equations of (18.9) describe the mean values of nonuniformly sampled sinusoid limited realizations [0, Θ] either for frequency $(f_i - f_m)$ or $(f_i + f_m)$ and phase angles $\varphi = \pi/2$ or $\varphi = 0$. This is a significant fact. It means that the impact of random sampling irregularities on the essential properties of randomly sampled composite signals can be revealed by studying estimation specifics of the mean value of a sinusoid. The estimate $\hat{m}(f)$ of the mean value and the expected value of the squared estimate of the mean value $E[\hat{m}^2(f)]$, derived by Bilinskis and Mikelsons in 1992 for various types of random sampling point processes, prove to be very informative and convenient for describing effects caused by random sampling irregularities including those related to the cross-intereference. The values of D_c, D_s, V_c and V_s characterize deviations from the mean values of sine waves at frequency $(f_i - f_m)$ or $(f_i + f_m)$, which are due to the nonuniformity of the used sampling point process. This means that the considered cross-interference coefficients are related in this way to these deviations and are tied to specific realizations of a sampling point process.

18.1.3 Approximation

Consider the estimates first in a generalized form not related to specific sampling conditions. Assume that an analog signal $x(t) = \sin(2\pi f t + \varphi)$ is randomly sampled. It can then be represented in the following form:

$$x(t_k) = \sin(2\pi f t_k + \varphi). \tag{18.10}$$

The estimate of the mean value of this signal is given by

$$\hat{m}(f) = \frac{1}{N}\sum_{k=1}^{N} \sin(2\pi f t_k + \varphi). \tag{18.11}$$

The expected value of $\hat{m}(f)$ is obtained by averaging the right-hand side of Equation (18.11) over the sampling instants $\{t_k\}$. Then

$$E[\hat{m}(f)] = \frac{1}{N}\sum_{k=1}^{N}\int_{-\infty}^{\infty}\sin(2\pi f t+\varphi)p_k(t)\,\mathrm{d}t \tag{18.12}$$

where $p_k(t)$ is the probability density function of the time intervals $[0, t_k]$.

Thus the expected value of $\hat{m}(f)$ depends both on the sinusoidal signal parameters and on the sampling point process used when sampling this signal. In the case of periodic sampling with jitter,

$$p_k(t) = \begin{cases} 1/T & \text{for } t \in [(k-1)T, kT], \\ 0 & \text{for } t \notin [(k-1)T, kT]. \end{cases}$$

Substituting this function into Equation (18.12) leads to

$$\begin{aligned} E[\hat{m}(\omega)] &= \frac{1}{\Theta}\sum_{k=1}^{N}\int_{(k-1)T}^{\infty}\sin(\omega t+\varphi)\,\mathrm{d}t \\ &= \frac{1}{\Theta}\int_{0}^{\Theta}\sin(\omega t+\varphi)\,\mathrm{d}t \\ &= \frac{\sin\omega\Theta/2}{\omega\Theta/2}\sin\left(\frac{\omega\Theta}{2}+\varphi\right), \end{aligned} \tag{18.13}$$

where $\omega = 2\pi f$ and Θ is the time interval during which the signal is observed. In this particular case, $\Theta = NT$. This expected value of $E[\hat{m}(f)]$ of a sine-wave signal sampled periodically with jitter is shown in Figure 18.1 as a function of the normalized frequency $\nu = f\Theta = fNT$. The two diagrams (a) and (b) illustrate the cases where $\varphi = \pi/2$ and $\varphi = 0$ respectively.

This function is directly tied to the expected values of D_c, D_s, V_c and V_s and therefore also to the expected values of the cross-interference coefficients. While in the wide frequency range the expected values of these coefficients are meaningless as they are in a wide frequency range close to zero, the function describing the expected values of them in the low-frequency range is quite useful. The point is that the expected value of $E[\hat{m}(f)]$ represents a good approximation of the cross-interference function there.

Basically there are two reasons why the mean value of a sinusoid might differ from zero. Firstly, it happens if such a signal is observed and the mean value is calculated for a time interval not equal to an integer number of its periods and, secondly, the mentioned deviations occur as a result of nonuniform sampling. Both of these factors often act simultaneously, causing the cross-interference

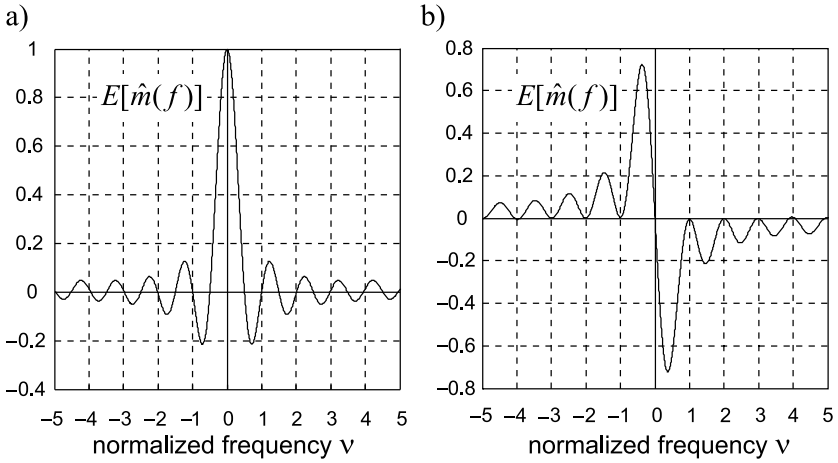


Figure 18.1 Expected mean value of a sinusoidal signal sampled periodically with jitter as a function of the normalized frequency

coefficient deviations from their expected mean values. The consequences of cutting off some part of a signal period prevail in the low-frequency range. They are much more powerful than the fluctuations due to nonuniform sampling, which dominate in a wide higher frequency range. Apparently these deviations of both types overlap.

Consider a signal component at frequency f_i. Suppose that the Fourier coefficient a_i is estimated in the presence of a cosine at frequency f_m. The impact of the frequency f_m on the estimation of the coefficient a_i is characterized by the cross-interference coefficient $A_i C_m$. Figure 18.2 illustrates how this coefficient depends on the distance between them under the given signal sampling conditions. While the frequency f_i is fixed in the indicated zero position, frequency f_m is varied and the value of $A_i C_m$ obtained for $N = 254$ is displayed as a function of f_m. This diagram, although characterizing the cross-interference coefficient $A_i C_m$, actually also reflects some typical tendencies observed at the estimation of other cross-interference coefficients. It is shown in more detail in Figure 18.3.

Equation (18.13) describing the expected value $E[\hat{m}(f)]$ of a sine-wave signal sampled periodically with jitter is used for approximation of the cross-interference coefficient $A_i C_m$ values in this case. Attention is drawn to two points. Firstly, this approximation is applicable for a wide variety of sampling process parameters. For example, although it has been calculated for the periodic sampling point process with jitter, it also fits well the curve obtained in the illustrated case under the conditions of additive sampling. Secondly, the suggested function approximates the cross-interference coefficient well in the f_m range close to frequency f_i. How

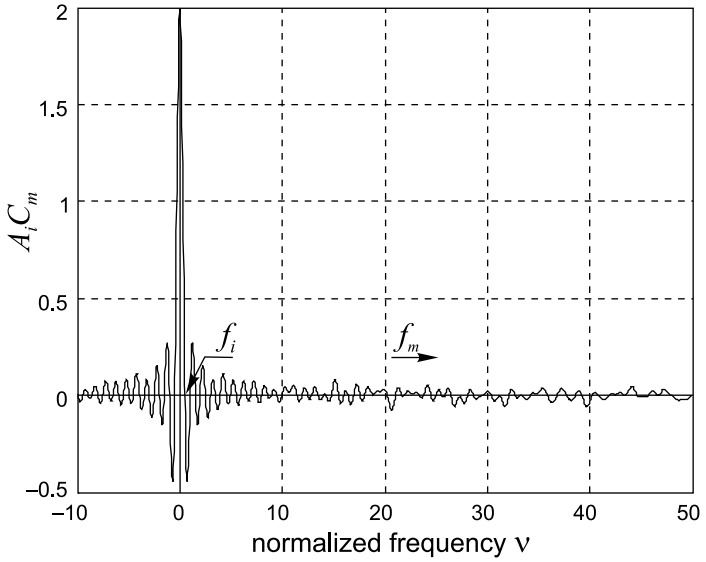


Figure 18.2 Dependence of the cross-interference coefficient A_iC_m on the distance between considered frequencies f_i and f_m

close depends on the particular parameters of the used sampling point process. Figure 18.4 illustrates this point.

18.2 Taking the Cross-interference into Account

When a spectral estimation process is described on the basis of the considered cross-interference coefficients (18.6), a system of equations can be composed showing the interference of all signal components. For example, in the case where there are three frequencies, the estimation of the signal spectral parameters is described by the following system of six linear equations:

$$\begin{aligned}
\hat{a}_1 &= a_1(A_1C_1) + b_1(A_1S_1) + a_2(A_1C_2) + b_2(A_1S_2) + a_3(A_1C_3) + b_3(A_1S_3),\\
\hat{b}_1 &= a_1(B_1C_1) + b_1(B_1S_1) + a_2(B_1C_2) + b_2(B_1S_2) + a_3(B_1C_3) + b_3(B_1S_3),\\
\hat{a}_2 &= a_1(A_2C_1) + b_1(A_2S_1) + a_2(A_2C_2) + b_2(A_2S_2) + a_3(A_2C_3) + b_3(A_2S_3),\\
\hat{b}_2 &= a_1(B_2C_1) + b_1(B_2S_1) + a_2(B_2C_2) + b_2(B_2S_2) + a_3(B_2C_3) + b_3(B_2S_3),\\
\hat{a}_3 &= a_1(A_3C_1) + b_1(A_3S_1) + a_2(A_3C_2) + b_2(A_3S_2) + a_3(A_3C_3) + b_3(A_3S_3),\\
\hat{b}_3 &= a_1(B_3C_1) + b_1(B_3S_1) + a_2(B_3C_2) + b_2(B_3S_2) + a_3(B_3C_3) + b_3(B_3S_3).
\end{aligned} \tag{18.14}$$

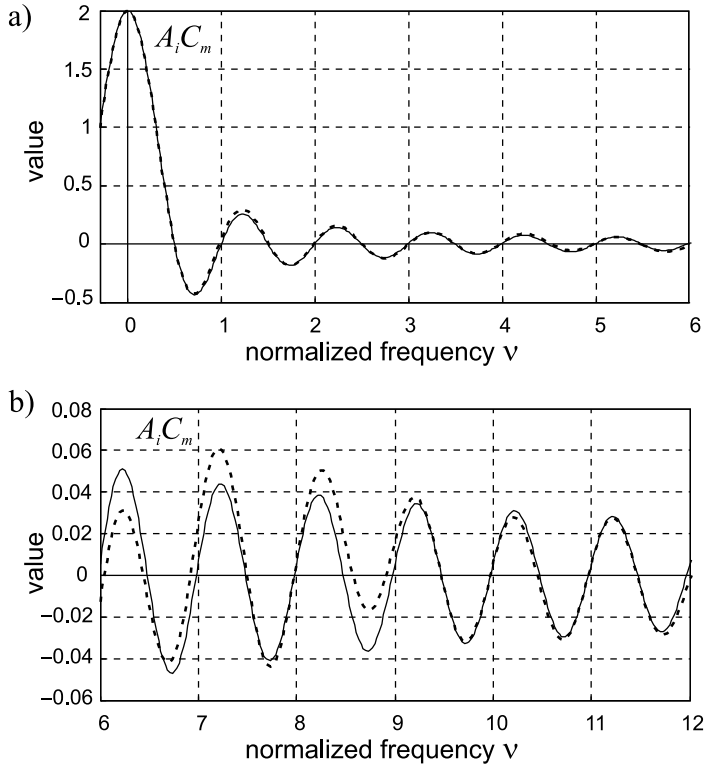


Figure 18.3 Actual and approximated values of the cross-interference coefficient A_iC_m: (a) in the f_m range close to frequency f_i; (b) in the f_m range relatively far from frequency f_i

Relationships (18.7) in this case lead to the substitutions given in Table 18.1. The equation system (18.14) can then be rewritten in a form symmetric to the diagonal and given as

$$\begin{aligned}
\hat{a}_1 &= a_1(A_1C_1) + b_1(A_1S_1) + a_2(A_1C_2) + b_2(A_1S_2) + a_3(A_1C_3) + b_3(A_1S_3),\\
\hat{b}_1 &= a_1(A_1S_1) + b_1(B_1S_1) + a_2(B_1C_2) + b_2(B_1S_2) + a_3(B_1C_3) + b_3(B_1S_3),\\
\hat{a}_2 &= a_1(A_1C_2) + b_1(B_1C_2) + a_2(A_2C_2) + b_2(A_2S_2) + a_3(A_2C_3) + b_3(A_2S_3),\\
\hat{b}_2 &= a_1(A_1S_2) + b_1(B_1S_2) + a_2(A_2S_2) + b_2(B_2S_2) + a_3(B_2C_3) + b_3(B_2S_3),\\
\hat{a}_3 &= a_1(A_1C_3) + b_1(B_1C_3) + a_2(A_2C_3) + b_2(B_2C_3) + a_3(A_3C_3) + b_3(A_3S_3),\\
\hat{b}_3 &= a_1(A_1S_3) + b_1(B_1S_3) + a_2(A_2S_3) + b_2(B_2S_3) + a_3(A_3S_3) + b_3(B_3S_3).
\end{aligned} \tag{18.15}$$

Table 18.1

		$A_iC_m = A_mC_i$	$A_iS_m = B_mC_i$	$B_iC_m = A_mS_i$	$B_iS_m = B_mS_i$
$i = 1$	$m = 1$	$A_1C_1 = A_1C_1$	$A_1S_1 = B_1C_1$	$B_1C_1 = A_1S_1$	$B_1S_1 = B_1S_1$
	$m = 2$	$A_1C_2 = A_2C_1$	$A_1S_2 = B_2C_1$	$B_1C_2 = A_2S_1$	$B_1S_2 = B_2S_1$
	$m = 3$	$A_1C_3 = A_3C_1$	$A_1S_3 = B_3C_1$	$B_1C_3 = A_3S_1$	$B_1S_3 = B_3S_1$
$i = 2$	$m = 1$	$A_2C_1 = A_1C_2$	$A_2S_1 = B_1C_2$	$B_2C_1 = A_1S_2$	$B_2S_1 = B_1S_2$
	$m = 2$	$A_2C_2 = A_2C_2$	$A_2S_2 = B_2C_2$	$B_2C_2 = A_2S_2$	$B_2S_2 = B_2S_2$
	$m = 3$	$A_2C_3 = A_3C_2$	$A_2S_3 = B_3C_2$	$B_2C_3 = A_3S_2$	$B_2S_3 = B_3S_2$
$i = 3$	$m = 1$	$A_3C_1 = A_1C_3$	$A_3S_1 = B_1C_3$	$B_3C_1 = A_1S_3$	$B_3S_1 = B_1S_3$
	$m = 2$	$A_3C_2 = A_2C_3$	$A_3S_2 = B_2C_3$	$B_3C_2 = A_2S_3$	$B_3S_2 = B_2S_3$
	$m = 3$	$A_3C_3 = A_3C_3$	$A_3S_3 = B_3C_3$	$B_3C_3 = A_3S_3$	$B_3S_3 = B_3S_3$

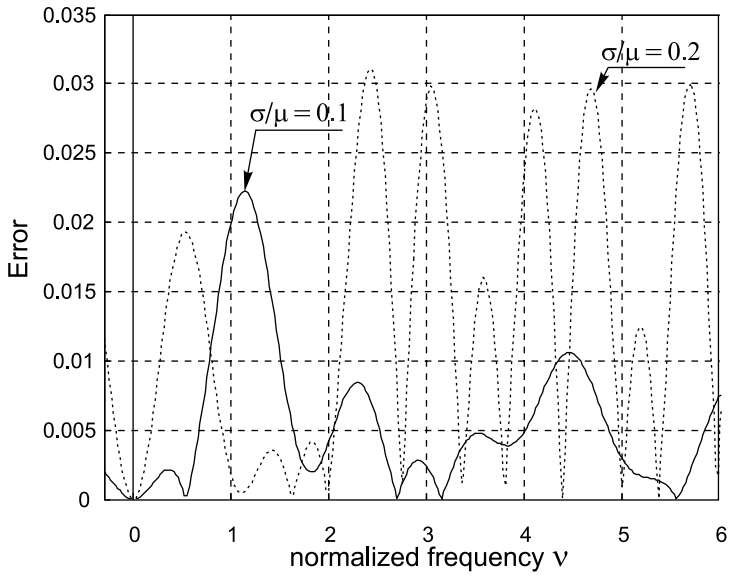


Figure 18.4 Approximation errors for the cross-interference coefficient A_iC_m values obtained in the case of additive sampling

In a general matrix form, this can be written as

$$\hat{\boldsymbol{p}} = \mathbf{C}\boldsymbol{p}, \tag{18.16}$$

where

$$\hat{\boldsymbol{p}} = [\hat{a}_1 \;\; \hat{b}_1 \;\; \hat{a}_2 \;\; \hat{b}_2 \;\; \cdots \;\; \hat{a}_M \;\; \hat{b}_M]^{\mathrm{T}},$$
$$\boldsymbol{p} = [a_1 \;\; b_1 \;\; a_2 \;\; b_2 \;\; \ldots \;\; a_M \;\; b_M]^{\mathrm{T}}$$

and the coefficient matrix is given as

$$\mathbf{C} = \begin{bmatrix} (A_1C_1) & (A_1S_1) & (A_1C_2) & (A_1S_2) & \cdots & (A_1S_M) & (A_1S_M) \\ (B_1C_1) & (B_1S_1) & (B_1C_2) & (B_1S_2) & \cdots & (B_1C_M) & (B_1S_M) \\ (A_2C_1) & (A_2S_1) & (A_2C_2) & (A_2S_2) & \cdots & (A_2S_M) & (A_2S_M) \\ (B_2C_1) & (B_2S_1) & (B_2C_2) & (B_2S_2) & \cdots & (B_2C_M) & (B_2S_M) \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ (A_iC_1) & (A_iS_1) & (A_iC_2) & (A_iS_2) & \cdots & (A_iC_M) & (A_iS_M) \\ (B_iC_1) & (B_iS_1) & (B_iC_2) & (B_iS_2) & \cdots & (B_iC_M) & (B_iS_M) \end{bmatrix}. \tag{18.17}$$

This matrix can be modified in accordance with the equation system version (18.15). Then the number of coefficients to be obtained is significantly reduced.

Solution of the given linear equation system provides corrected spectral parameter estimates. It can be written that

$$\boldsymbol{p} = \mathbf{C}^{-1}\hat{\boldsymbol{p}} \tag{18.18}$$

where $\mathbf{C}^{-1} = \text{inv}(C)$. Matrix (18.15) of the cross-interference coefficients is an essential characteristic of the respective nonuniform sampling point process. Once the pattern of sampling points is known, all coefficients of matrix $\mathbf{C}$, as well as inv($\mathbf{C}$), could be calculated. This means that when a signal is sampled in accordance with such a predetermined nonuniform sampling point process, solving the equation system (18.18) changes the estimation of the signal spectral estimates to the specific pattern of sampling points. That significantly reduces the errors due to the damaging cross-interference between the nonuniformly sampled signal components.

18.3 Achievable Improvement and Typical Problems

To illustrate the achievable improvement in signal parameter estimation by using the described concept of the cross-interference between nonuniformly sampled signal components, an example will be considered. Suppose that a nonuniformly sampled signal containing a number of components is decomposed and the parameters of the signal components are estimated. It is sampled according to the pseudo-randomized additive sampling model and τ/μ in this case is equal to 0.23. Fourier coefficients for the signal components are estimated on the basis of the DFT and by adapting the spectrum analysis to the specific irregularities of the used sampling point process. The true values of these Fourier coefficients and their estimates are given in Table 18.2.

It can be seen from the data given in Table 18.2 that adapting signal spectrum analysis to specific nonuniformities of the sampling process provides for

Table 18.2

i	Frequencies f/f_s	True values		DFT estimates		Estimates after iteration 1		Estimates after iteration 2		Estimates after iteration 4	
		a_i	b_i	a_i	b_i	a_i	b_i	a_i	b_i	a_i	b_i
1	0.3086	0	0.63	0.34	0.51	0.05	0.53	0.05	0.63	0	0.63
2	0.4629	1.41	0	1.63	−0.07	1.45	−0.02	1.39	0.06	1.41	0
3	0.7314	0	1.03	−0.24	1.28	−0.08	1.04	0	0.99	0	1.03
4	0.7600	2.92	0	3.39	0.08	2.96	−0.09	3	−0.09	2.92	0
5	0.8914	0.73	0	0.75	−0.22	0.69	0.18	0.7	−0.01	0.73	0
6	1.5714	0	3.32	0.36	3.38	0.09	3.37	0.05	3.21	0	3.32
7	1.6229	0	0.23	0.34	0.19	0.24	0.13	0.07	0.2	0	0.23
8	1.7419	0	2.44	0.33	2.52	0.2	2.52	0.11	2.5	0	2.44

significant improvement in signal parameter estimation accuracy. Spectrograms illustrating the achieved positive effect are given in Figure 18.5.

As the given example demonstrates, adaptation of signal processing to the nonuniformities of the sampling point processes used at digitization of the signal is very effective. Moreover, using this or some other still-not-discovered approach to the described adaptation is a compulsory precondition for achieving high-performance processing of signals sampled nonuniformly. The specific nonuniformities of the sampling process must be taken into account.

Of course there are some difficulties. For instance, the quality of such an adaptation depends on the involved frequency estimation precision as the cross-interference coefficient estimation is sensitive to the frequency estimation errors. The required frequency resolution is evaluated later in Section 18.5. Then there are problems related to the large quantity of cross-interference coefficients that characterize particular sampling point processes. They have to be either pre-calculated and kept in a memory or estimated in a real-time mode. Some approaches to the resolution of this problem are discussed in Sections 18.5 and 18.6. The fact that there is a simple approximation of the cross-interference coefficients, described in Section 18.1, certainly helps. Last but not least, there is a problem with the inevitable increase in the computational burden accompanying the introduction of the discussed adaptation. The calculations to be made are also relatively time consuming. An optional approach to implementation of the considered adaptation, making it possible to reduce the time needed for the required calculations, is based on exploitation of the parallel computing potential.

18.4 Parallel Computing Approach

The problem with direct adaptation of signal processing to the irregularities of the involved sampling process is the fact that it requires solving a very large equation

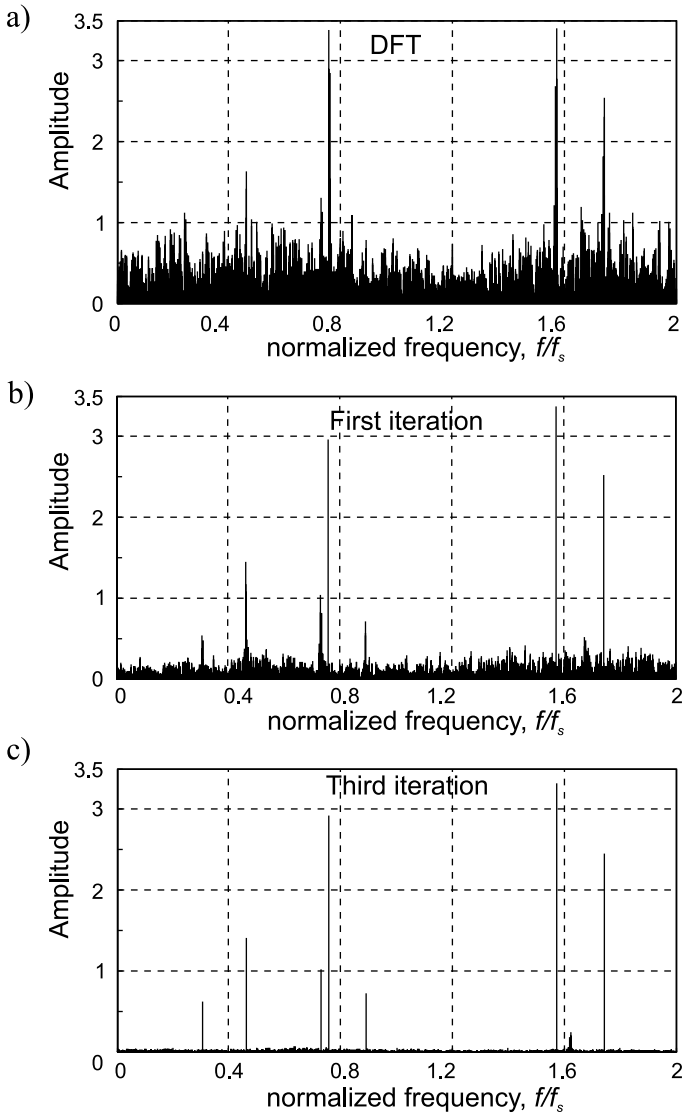


Figure 18.5 Illustration of the improvement in signal spectrum analysis achievable by adapting signal processing to the specific nonuniformities of the used sampling point process: (a) results of the direct application of the DFT; (b), (c) spectrograms obtained in the case of adapting after the first and third iterations respectively

system requiring operations with large dimension matrices. However, analysis of this matter shows that in many cases it is possible to subdivide this task into a number of subtasks that could all be resolved in parallel. This parallel computing approach is based on the following procedures:

1. The signal sample value sequence is subdivided into a number of periodic pseudo-randomly decimated subsequences.
2. Adapting the calculation of particular sets of Fourier coefficient estimates to the sampling nonuniformities for each signal sample value subset is performed separately and in parallel.
3. The obtained particular sets of Fourier coefficient estimates are aggregated in order to obtain the alias-free spectral estimates for the whole input signal.

Figure 18.6 illustrates this parallel computing approach to adapting signal processing to the irregularities of sampling. All these stages of parallel processing will be considered in some detail.

18.4.1 Decomposition of the Signal Sample Value Sequence

To rearrange the computing process as a parallel one, the input signal sample value sequence has to be decomposed. Suppose that the input signal is sampled according to the pseudo-randomized additive sampling model. Then subdividing the obtained signal sample value sequences into periodic pseudo-randomly decimated subsequences can be performed as described in Section 9.2. As explained there, it is essential to avoid fuzzy aliasing. To achieve that, the used additive nonuniform sampling instant process is decomposed into a number of periodic sampling pulse processes with pseudo-randomly skipped sampling instants. The sampling instant sequence is then considered as consisting of z periodic phase-shifted processes. All of these periodic processes have equal periods T and each of them is phase-shifted by a phase increment T/z. The time period T should be equal to or longer than the period of the highest sampling frequency for the ADC to be used. Each sampling instant in the initial sampling pulse sequence is marked as belonging to a specific periodic sampling point process.

Such subdivision of the total sampling pulse process into components, organized as shown in Figure 18.6, improves the quality of signal processing substantially. Firstly, aliasing for the periodic sampling pulse streams with pseudo-random skips is well defined. Secondly, these well-defined aliases could be taken out at spectral estimations by aggregating data obtained by calculating spectral estimates for all particular data streams.

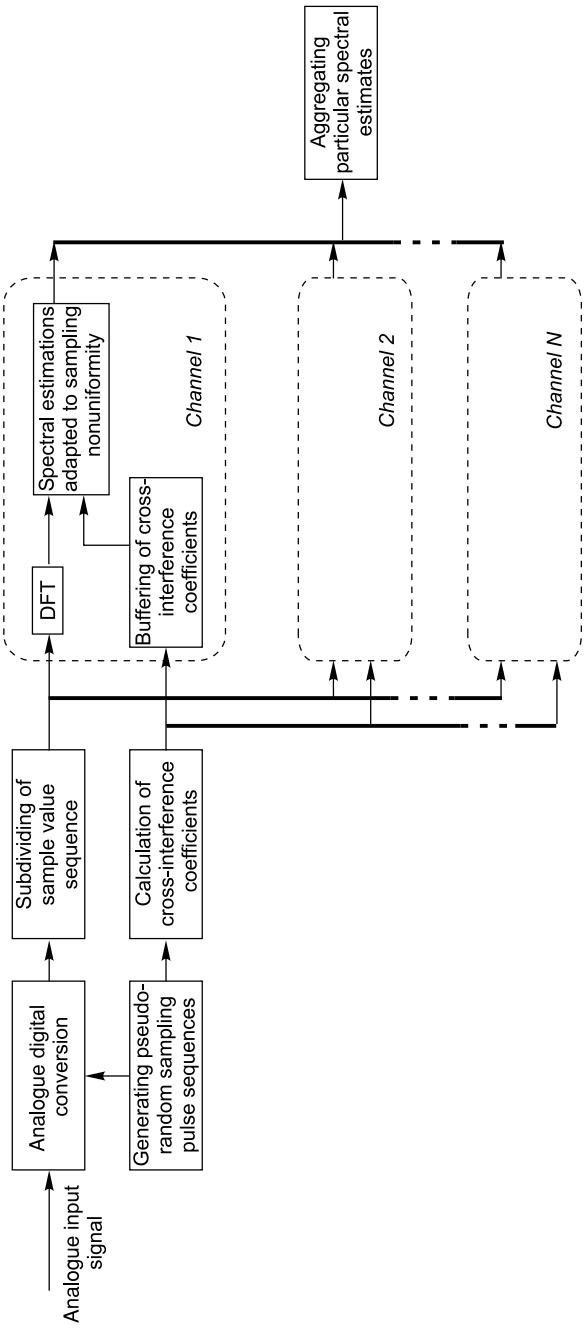


Figure 18.6 Block diagram of a system for spectrum analysis adapted to the sampling nonuniformity and based on parallel computing of the Fourier coefficient estimates

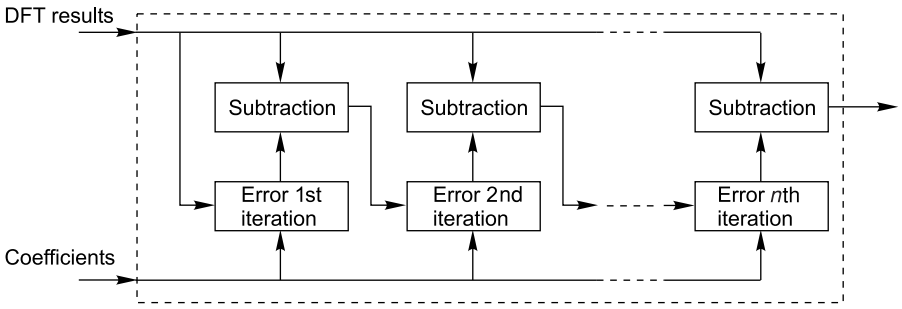


Figure 18.7 Detailed block diagram of the functional unit for adapting signal spectrum estimations to the specific sampling nonuniformities of a signal sample value subset

To realize how this could be done, the impact of phase-shifting of the sampling pulse sequences on the aliasing conditions has to be taken into account. The point is that conditions for frequency overlapping or aliasing essentially depend not only on the sampling frequency but also on the sampling phase. Therefore the estimates of signal components in the frequency domain differ depending on the phase shift of the corresponding periodic sampling instant process. The fact that the phase angle of the signal, downconverted due to aliasing, depends on the phase of the sampling point process is exploited in order to eliminate the aliases at the stage of aggregating the data. The particular z sets of the Fourier coefficient estimates, obtained by processing the signal sample value z subsets, are used to calculate the final alias-free characteristics of the input signal spectrum.

18.4.2 Adapting the Estimation for Each Signal Sample Value Subset

Adapting the estimation of the Fourier coefficients for each signal sample value subset is performed as described in Section 18.2. Once the pattern of the missing sampling points determined by the pseudo-random number generator used to form the sampling pulse sequence is known, all coefficients of matrix **C** are calculated for each signal sample value subset. They are then used to eliminate the errors due to the described cross-interference caused by sampling point skipping.

In general, in this approach to the considered adaptation process, the number of signal components might exceed the number of signal sample values of a particular periodic sampling subset. Then an iterative procedure has to be used to adapt the calculation of particular sets of Fourier coefficient estimates to sampling nonuniformities for each signal sample value subset, as shown in Figure 18.7.

Apparently the differences between the true and estimated values of Fourier coefficients are errors. While there are various error components caused by

quantization errors, external noise, etc., the errors due to the sampling nonuniformities considered here are dominating. The error of estimating the Fourier coefficients at a frequency i is proportional to the amplitude of the interfering quadratic components at the frequency j multiplied by the corresponding cross-interference coefficient. In general, it might be written that these errors of estimating coefficients a_i and b_i depend on the amplitudes of all other quadratic components a_j and b_j of the signal at their respective frequencies f_j:

$$\varepsilon_{ij}(a_i, b_i) = \left\{ \begin{array}{ll} A_i C_j a_j, & A_i S_j a_j \\ B_i C_j b_j, & B_i S_j b_j \end{array} \right\}. \tag{18.19}$$

However, the true values of a_i and b_i are not known. Only their DFT estimates $\hat{a}_i$ and $\hat{b}_i$ are given. Nevertheless, they make it possible to calculate the approximation of the estimation errors:

$$\varepsilon_{ij}^{(1)}(a_i, b_i) = \left\{ \begin{array}{ll} A_i C_j \hat{a}_j, & A_i S_j \hat{a}_j \\ B_i C_j \hat{b}_j, & B_i S_j \hat{b}_j \end{array} \right\}. \tag{18.20}$$

These errors are subtracted from the DFT estimates $\hat{a}_i$ and $\hat{b}_i$ and the first set of the improved estimates $a_i^{(1)}$ and $b_i^{(1)}$ are obtained as shown in Figure 18.7.

For the next iteration step, the estimation error

$$\varepsilon_{ij}^{(2)}(a_i, b_i) = \left\{ \begin{array}{ll} A_i C_j a_j^{(1)}, & A_i S_j a_j^{(1)} \\ B_i C_j b_j^{(1)}, & B_i S_j b_j^{(1)} \end{array} \right\} \tag{18.21}$$

is calculated and then the iteration is made. This iterative process is continued for a number of times. For the $(k+1)$th iteration step,

$$\varepsilon_{ij}^{(k+1)}(a_i, b_i) = \left\{ \begin{array}{ll} A_i C_j a_j^{(k)}, & A_i S_j a_j^{(k)} \\ B_i C_j b_j^{(k)}, & B_i S_j b_j^{(k)} \end{array} \right\}, \tag{18.22}$$

where the improved estimates $a_i^{(k)}$ and $b_i^{(k)}$ are obtained at the previous kth iteration step. The obtained estimates of the Fourier coefficients for each and all sampling pulse substreams are adapted to the sampling nonuniformities in this way. They are significantly more precise than the initially used raw estimates obtained directly as a result of application of the DFT. The improved estimates represent data for further signal processing.

18.4.3 Data Aggregation

Data representing estimates of Fourier coefficients, calculated in the frequency range 0–0.5 f_s for each multitude of signal sample values under conditions of

100 % aliasing, are to be aggregated to obtain alias-free estimates of Fourier coefficients in the whole frequency range defined by the phase-shifted value of the sampling pulse substreams.

Without loss of the generality, the specific equations are given below that are used for spectral estimations in the case where the common sampling point stream is subdivided into four substreams. The phase shift between separate sampling pulse substreams is then equal to $\frac{1}{4}T$. Therefore the highest frequency of the signal might exceed two times the frequency of the periodic sampling point processes used. After the signal sample values are subdivided into four subsets, each of them is processed separately in the Nyquist frequency band (0–0.5f_s). Each of the signal basic frequencies f_0 within the frequency band (0–0.5f_s) has three potentially aliasing frequencies f_1, f_2 and f_3 respectively. Spectral estimations of the signal components in the whole frequency range (0–2f_s) have to be made. Accuracy of the estimation of the Fourier coefficients for all signal components in the whole frequency range depends on the accuracy of estimating of $a_0^{[m]}$ and $b_0^{[m]}$ for all four ($m = 0, 1, 2, 3$) sampling subroutines. The values of $a_i^{[0]}$, $b_i^{[0]}$, $a_i^{[1]}$, $b_i^{[1]}$, $a_i^{[2]}$, $b_i^{[2]}$ and $a_i^{[3]}$, $b_i^{[3]}$ are calculated by applying the DFT procedure and then these estimates are improved by adapting them to the specific sampling nonuniformities, as explained above. After that, the obtained data have to be used to obtain the improved estimates of the signal components in the whole bandwidth.

The following system of equations is used:

$$\begin{aligned}
a(f_i) &= \tfrac{1}{4}\left(a_i^{[0]} + a_i^{[1]} + a_i^{[2]} + a_i^{[3]}\right),\\
b(f_i) &= \tfrac{1}{4}\left(b_i^{[0]} + b_i^{[1]} + b_i^{[2]} + b_i^{[3]}\right),\\
a(f_s - f_i) &= \tfrac{1}{4}\left(a_i^{[0]} + b_i^{[1]} - a_i^{[2]} - b_i^{[3]}\right),\\
b(f_s - f_i) &= \tfrac{1}{4}\left(-b_i^{[0]} + a_i^{[1]} + b_i^{[2]} - a_i^{[3]}\right),\\
a(f_s + f_i) &= \tfrac{1}{4}\left(a_i^{[0]} - b_i^{[1]} - a_i^{[2]} + b_i^{[3]}\right),\\
b(f_s + f_i) &= \tfrac{1}{4}\left(b_i^{[0]} + a_i^{[1]} - b_i^{[2]} - a_i^{[3]}\right),\\
a(2f_s - f_i) &= \tfrac{1}{4}\left(a_i^{[0]} - a_i^{[1]} + a_i^{[2]} - a_i^{[3]}\right),\\
b(2f_s - f_i) &= \tfrac{1}{4}\left(-b_i^{[0]} + b_i^{[1]} - b_i^{[2]} + b_i^{[3]}\right).
\end{aligned} \tag{18.23}$$

After that the Fourier coefficients of all signal components at frequencies within the frequency bands (0–0.5f_s), (0.5f_s–f_s), (f_s–1.5f_s) and (1.5f_s–2f_s) are explicitly expressed through the amplitudes of the quadratic components of the sine functions filtered out within the frequency range (0–0.5f_s).

18.5 Mapping of the Cross-interference Coefficients

The first step to be undertaken for adapting digital signal processing to the nonuniformity of a given sampling point sequence should apparently be mapping of the cross-interference coefficients linking them to this specific sampling point sequence. Matrix (18.13) has to be composed for that. However, it is not easy to fulfil this task. A lot of calculations have to be carried out and there are some questions that have to be answered before a start can be made. One of the most essential is the question: what is the required dimension of this matrix? The answer obviously depends on the frequency of the step-size Δf on which these calculations need to be based. The next question arising is: how large should this step-size be? The answer to this question is not as obvious as it may seem.

18.5.1 Required Frequency Resolution

The frequency resolution Δf usually used in a spectrum analysis of a signal is inversely proportional to the time interval Θ during which the signal has been observed. This step-size Δf more often than not proves to be too large for mapping the cross-interference coefficients. Adaptation of signal processing to the nonuniformities of the used sampling point process is sensitive to deviations from signal frequencies. Therefore to achieve acceptable results a frequency resolution higher than $\Delta f = 1/\Theta$ typically has to be ensured. The exact required frequency resolution depends on a number of factors that have to be taken into account.

The errors in estimating $A_i C_m$ sharply increase when the distance between the frequencies f_i and f_m (for a particular position of frequency f_i) becomes small. That happens in cases of high-resolution DFTs where some signal components are located very close to each other. Then their frequencies need to be estimated with very high precision. How the error $\Delta A_i C_m$ in estimating the cross-interference coefficient $A_i C_m$ depends on the error in estimating the frequency f_i is shown in Figure 18.8.

Suppose that the frequency f_i of a signal component is estimated with an error equal to $0.2\Delta f$. This error then leads to obtaining incorrect values of the corresponding cross-interference coefficient $A_i C_m$. The dependence of the error in estimating this coefficient on the distance between considered frequencies f_i and f_m for a particular position of frequency f_i is illustrated by Figure 18.9.

Note that these diagrams have been obtained by processing $N = 256$ signal sample values. The error in estimating the cross-interference coefficient is less

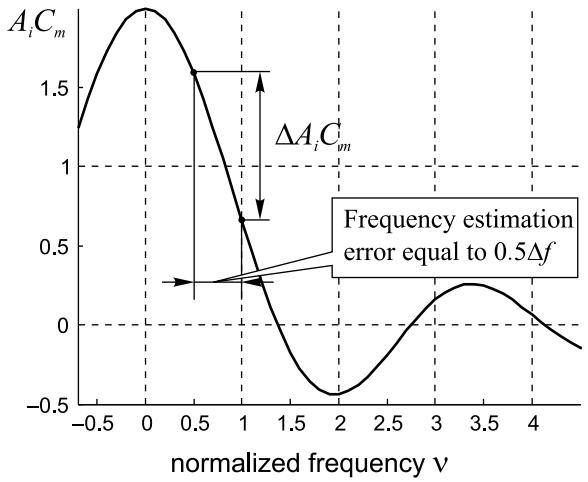


Figure 18.8 Illustration of how the error in estimating the cross-interference coefficient A_iC_m depends on the error in estimating the frequency f_i

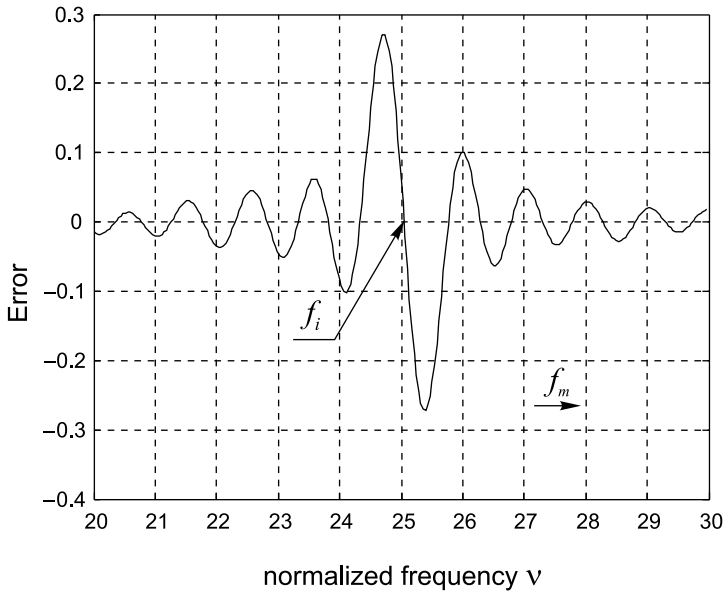


Figure 18.9 Dependence of the maximal error in estimating the cross-interference coefficient A_iC_m on the distance between the frequencies f_i and f_m

sensitive to error than in estimating the frequency f_i in the cases where more signal samples are processed.

18.5.2 Coefficient Mapping Versus On-line Calculations

As each realization of a sampling point process is exclusively characterized by a matrix of cross-interference coefficients (18.17), it is tempting to compose and use such a matrix. However, the size of it represents a problem. So far, to simplify this discussion, various conditions for estimating just one of the cross-interference coefficients have been considered. This is acceptable as basically the same conclusions would be made if estimating other cross-interference coefficients is studied as well. At this stage of the analysis of conditions for adapting signal processing to the sampling nonuniformities, it is clear that signal component frequencies for successful adapation typically need to be estimated with errors not exceeding $0.1–0.5\Delta f$. Therefore the size of matrix (18.17), containing four sets of four different cross-interference coefficients for a frequency band from 0 to $M\Delta f$, would be up to $10M \times 10M \times 4$. Although it is possible to have and use a memory of this size, it certainly is large.

The size of the memory used to store information about the nonuniformities of the sampling process could be significantly reduced if the mean values of nonuniformly sampled sinusoids for variable frequencies are calculated and kept rather than the values of the cross-interference coefficients. As these coefficients could be easily calculated on the basis of $m(f)$ values and Equations (18.8) and (18.9), it makes sense to store them in the memory and then use them for simple calculations of the cross-interference coefficients as the mean values of specifically sampled sinusoids at phase angles $\varphi = \pi 2$ and $\varphi = 0$. This approach reduces the required size of the memory two times. Although this reduction is significant, the dimensions of the memory even after such a reduction might actually be considered as excessive.

Further reduction of the size of this memory could be achieved on the basis of a trade-off between the complications related to the large size and complications caused by additional on-line calculations.

19

Estimation of Object Parameters

To achieve the best results, algorithms for processing digital signals should take into account the specific features of the digital signals to be processed. These features to a large extent depend on the chosen and used sampling and quantizing methods. The latter, in turn, depend on the given task to be solved. Thus the task actually dictates the requirements that have to be satisfied when choosing and using signal digitizing and processing techniques.

In the previous chapters an attempt has been made to show that signals could be digitized and processed in various ways and that it is essential to use appropriate methods and techniques. This helps to meet requirements of many different applications and the discussed methods for signal digitizing and processing actually cover a wide application field where good results can be achieved. On the other hand, in some areas it is easier to be successful than in others. One such application area where it is relatively easy to get good results at solving challenging tasks is considered in this chapter. This is the area where signal processing is carried out to identify objects and to evaluate their parameters by analysing the signals reflecting the reaction of these objects to some excitation signals. The parameters and characteristics of the excitation signals are usually known. Access to this valuable information makes it easier to keep algorithms relatively simple and to achieve high performance.

19.1 Measuring the Frequency Response of Objects

Identification of objects by observing and analysis of their reaction to specific test signals is a task often met. Many kinds of signal processing systems fulfil

Digital Alias-free Signal Processing I. Bilinskis

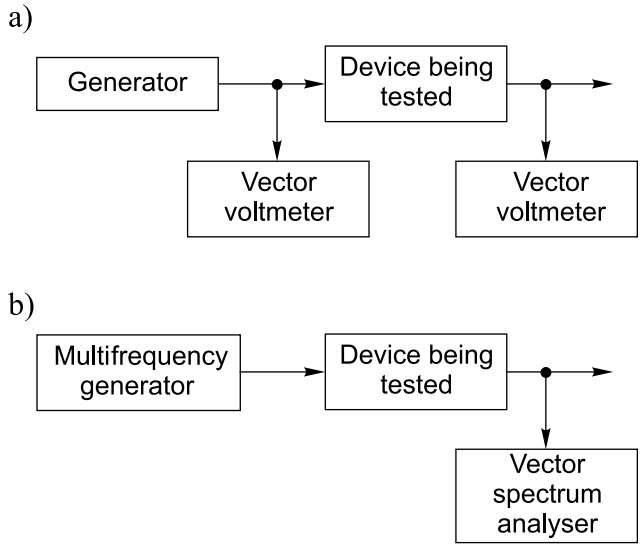


Figure 19.1 Block diagrams of systems for measuring the frequency response of objects. (a) prototype; (b) improved version

tasks in this category. The existing systems of this type in many cases are good enough. Nevertheless, with the advancement of high technologies, the requirements are growing in this area. The nontraditional signal processing techniques discussed in this book might prove to be beneficial for achieving significant widening of the frequency range where these systems could operate, simplification of the system designs and/or increasing their operational speed. Simplified block diagrams of basic and suggested typical systems of this kind are given in Figure 19.1.

The basic structure of the most popular system for measuring the frequency response, often considered as a network analyser and shown in Figure 19.1(a), is well known. Typically, it contains a variable frequency single tone generator and two vector voltmeters connected to the input and output of the electronic device to be tested by measuring its frequency response and other derivative parameters. The frequency response is measured by changing the frequency of the generated input signal step by step over the whole frequency range of interest. At each frequency setting, amplitudes and phase angles of the signals at the input and output of the device under test are measured and the data obtained in this way are used to calculate all needed parameters.

An improved version of this type of system is given in Figure 19.1(b). The structure and performance of it differs in the following ways:

1. A multifrequency signal source is used to speed up the whole measurement process by avoiding a repetition of the measurement process step by step for each separate frequency.
2. A spectrum analyser rather than a vector voltmeter is used for high-speed estimation of the frequency response of the electronic device being tested by processing a single shot of the output signal.
3. The multifrequency signal source generates the test signal using specified parameters, including the phase angles for all frequency components, leading to the possibility of using a single rather than two spectrum analysers and thus reducing the complexity and production costs of the system.
4. Pseudo-randomized sampling techniques are used to achieve and provide for a high-resolution vector spectrum analysis of wideband signals in the frequency range up to a few GHz.

While all the elements of the system determine its performance, the multifrequency test signal generator plays a special role. In an ideal case, the generated excitation signal contains discrete frequencies covering the required density for the whole frequency range within which the object behaviour is to be observed. It is not easy to cover all of it at one go if this frequency range is indeed wide. It is more realistic to generate multifrequency test signals covering some part of the whole frequency range of interest. Then the test measurement procedure has to be repeated several times. Obviously, much depends on the frequency band that could be covered by a single realization of the test signal.

There is some risk when just one channel spectrum analysis is used. Then most of the responsibility for phase-locking of the generated frequency components of the excitation signal in the predetermined positions is put on the excitation signal synthesizer. There may also be some hidden problems as various impedances may impact on the measurement accuracy. It is necessary to check whether this effect can be kept under control and whether one spectrum analyser measurement scheme can be applied in all cases.

All these considerations show that the performance of the whole test system depends to a large degree on the capabilities of the multifrequency test signal synthesizer. An approach to such signal synthesis, based on exploitation of the alias-free signal sampling techniques, is discussed in the next section.

19.2 Test Signal Synthesis from a Sparsely Periodically Sampled Basis Function

The task of analog signal synthesis from discrete data comes up frequently and the ways used to resolve it by using a DAC are well known. However, that is true only if

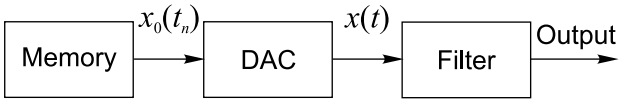


Figure 19.2 Generalized structure for continuous signal $x(t)$ synthesis from digital data

and when the classical requirements are met, specifically when the corresponding discrete data block to be transformed into an analog signal is actually represented by a sample value sequence taken from a certain basis function under conditions where at least two sample values are taken in the time interval equal to the period of the highest frequency present in this function. Then no aliasing takes place and there are no problems. Therefore this is the preferred way to perform analog signal synthesis whenever the mentioned conditions can be satisfied. However, the operational speed of DACs is limited and, consequently, the highest frequency of the signal to be synthesized in this way cannot exceed a certain boundary determined by the switching speed of the sufficiently high bit rate DAC available. The accumulated experience of successful suppression of various negative effects due to aliasing suggests that it is worth looking for a way to avoid this synthesis limitation.

The generalized model of the continuous signal $x(t)$ synthesis from digital data is given in Figure 19.2. To obtain a continuous or analog signal $x(t)$, a related basis function $x_0(t)$ is formed. Typically that is done by a computer, forming it as a sum of a finite number of sinusoidal components. As parameters (frequencies, amplitudes and phase angles) depend in some way on the analog signal to be synthesized, their values, as well as the sampling frequency f_s, should be calculated on the basis of the algorithm used for synthesis. The approach to such analog signal synthesis will be discussed further. At this stage it should simply be noted that these signal component parameters are fixed and, consequently, sample values $x_0(t_n)$, $n = 1, 2, \ldots, N$, of the basis function can be calculated and stored in a memory. These digital values are than passed at the sampling rate f_s to a DAC performing their conversion into a kind of analog signal, which changes its values in an idealized stepwise way, as shown in Figure 19.3. The synthesized signal is taken off the output of the filter.

It can be seen from the diagrams in Figure 19.3 that the values of the basis function change continuously while the DAC output signal remains at some level for relatively long time intervals, which are longer than the period of the basis function. The DAC output signal certainly does not look similar to the given basis function. The question arises as to what kind of properties this staircase signal prossesses and whether it is possible to obtain such a signal with the required properties.

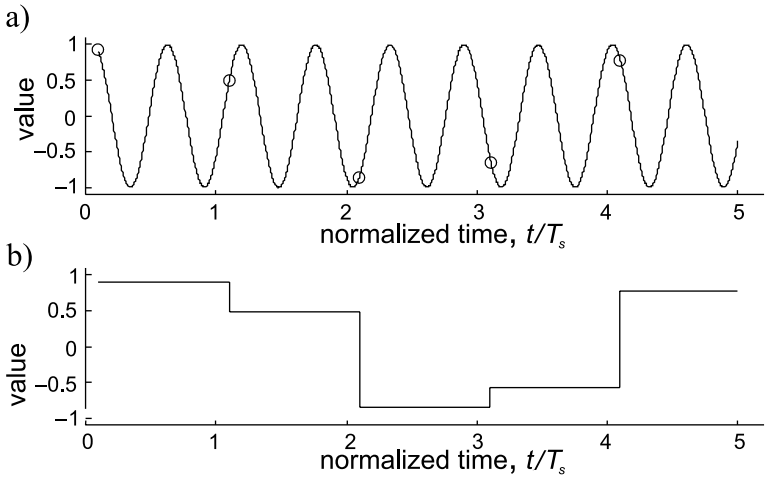


Figure 19.3 Illustration of analog signal synthesis in the case of periodic sampling; (a) basis function $x_0(t)$ with its sample values $x_0(t_n)$ indicated; (b) signal at the output of the DAC

The spectrum of the DAC output signal is obviously unlimited, even in the case of the low-frequency basis function $x_0(t)$. On the other hand, this spectrum clearly depends both on the basis function and the sampling frequency f_s. This relationship determines the properties of the synthesized signal and analysis of it should provide the answer to the question of how feasible it is to achieve the possibility of developing a high-performance signal synthesizer. To find the answer to the question of how feasible it is to synthesize analog signals under conditions where the requirements of the sampling theorem are not met, it is necessary to establish how feasible it is to transfer the frequencies of the basis function to the DAC output signal, even when those frequencies exceed the sampling rate, and whether it is possible to synthesize a signal having the required frequencies not corrupted by aliases.

In order to get answers to these and other similar questions, there will be a focus on the estimation of the spectra of the stepwise DAC output signal that clearly depends on the sampled basis function and on sampling conditions.

19.2.1 Synthesis in the Case of Monoharmonic Basis

Assume that the basis function $x_0(t)$ is monoharmonic:

$$x_0(t) = A_0 \sin(2\pi f_0 t + \varphi_0). \tag{19.1}$$

It follows from the explained synthesis approach that the DAC output signal is given as

$$x(t) = x_0(t_n), \quad \text{if } t_n \le t < t_{n+1}, -\infty < n < \infty. \tag{19.2}$$

As sampling is periodic, $t_n = n\mu$, where $\mu = 1/f_s$ is the sampling time interval. The spectrum of the signal $x(t)$ in this case can be defined in the following way:

$$S_x(f) = \lim_{N\to\infty} \frac{2}{(2N+1)\,\mu} \sum_{n=-N}^{N} \int_{n\mu}^{(n+1)\mu} A_0 \sin(2\pi f_0 n\mu + \varphi_0)\, \mathrm{e}^{-\mathrm{j}2\pi f t} \mathrm{d}t. \tag{19.3}$$

Consider the real $a_x(f)$ and the imaginary $b_x(f)$ parts of this Fourier transform separately. Then

$$\begin{aligned} a_x(f) &= \lim_{N\to\infty} \frac{2}{(2N+1)\,\mu} \sum_{n=-N}^{N} A_0 \sin(2\pi f_0 n\mu + \varphi_0) \int_{n\mu}^{(n+1)\mu} \cos 2\pi f t \ \mathrm{d}t. \\ &= \lim_{N\to\infty} \frac{A_0 \operatorname{sinc}(\pi f \mu)}{2N+1} \sum_{n=-N}^{N} \{\sin[2\pi(f_0 - f)n\mu - \pi f\mu + \varphi_0] \\ &\quad + \sin[2\pi(f_0 + f)n\mu + \pi f\mu + \varphi_0]\} \\ &= \lim_{N\to\infty} A_0 \operatorname{sinc}(\pi f\mu) \left\{ \frac{\operatorname{sinc}[\pi(f_0 - f)\Theta]}{\operatorname{sinc}[\pi(f_0 - f)\mu]} \sin(\varphi_0 - \pi f\mu) \right. \\ &\quad \left. + \frac{\operatorname{sinc}[\pi(f_0 + f)\Theta]}{\operatorname{sinc}[\pi(f_0 + f)\mu]} \sin(\varphi_0 + \pi f\mu) \right\}, \end{aligned} \tag{19.4}$$

where $\Theta = (2N+1)\,\mu$ and $\operatorname{sinc}(x) = \sin x/x$. The coefficient $b_x(f)$ is obtained in a similar way:

$$\begin{aligned} b_x(f) &= \lim_{N\to\infty} \frac{2}{(2N+1)\,\mu} \sum_{n=-N}^{N} A_0 \sin(2\pi f_0 n\mu + \varphi_0) \int_{n\mu}^{(n+1)\mu} \sin 2\pi f t \ \mathrm{d}t \\ &= \lim_{N\to\infty} \frac{A_0 \operatorname{sinc}(\pi f \mu)}{2N+1} \sum_{n=-N}^{N} \{\cos[2\pi(f_0 - f)n\mu - \pi f\mu + \varphi_0] \\ &\quad - \cos[2\pi(f_0 + f)n\mu + \pi f\mu + \varphi_0]\} \\ &= \lim_{N\to\infty} A_0 \operatorname{sinc}(\pi f\mu) \left\{ \frac{\operatorname{sinc}[\pi(f_0 - f)\Theta]}{\operatorname{sinc}[\pi(f_0 - f)\mu]} \cos(\varphi_0 - \pi f\mu) \right. \\ &\quad \left. - \frac{\operatorname{sinc}[\pi(f_0 + f)\Theta]}{\operatorname{sinc}[\pi(f_0 + f)\mu]} \cos(\varphi_0 + \pi f\mu) \right\}. \end{aligned} \tag{19.5}$$

The transition $N \to \infty$ is performed in Equations (19.4) and (19.5). Then

$$a_x(f) = \begin{cases} A_0 \operatorname{sinc}(\pi f \mu) \sin(\varphi_0 - \pi f \mu), & \text{if } 0 < f = f_0 \pm r f_s, \\ A_0 \operatorname{sinc}(\pi f \mu) \sin(\varphi_0 + \pi f \mu), & \text{if } 0 < f = -f_0 \pm r f_s, \\ 0, & \text{if } 0 < f \neq \pm f_0 \pm r f, \end{cases} \quad (19.6)$$

$$b_x(f) = \begin{cases} A_0 \sin c(\pi f \mu) \cos(\varphi_0 - \pi f \mu), & \text{if } 0 < f = f_0 \pm r f_s, \\ -A_0 \operatorname{sinc}(\pi f \mu) \cos(\varphi_0 + \pi f \mu), & \text{if } 0 < f = -f_0 \pm r f_s, \\ 0, & \text{if } 0 < f \neq \pm f_0 \pm r f_s, \end{cases} \quad (19.7)$$

where $r = 0, 1, 2, 3, \ldots$.

The Fourier transform module $A_x(f)$, actually representing the amplitude of the signal $x(t)$ at the frequency f, is derived from Equations (19.6) and (19.7) as

$$A_x(f) = \begin{cases} A_0 |\operatorname{sinc}(\pi f \mu)|, & \text{if } 0 < f = \pm f_0 \pm r f_s, \\ 0, & \text{if } 0 < f \neq \pm f_0 \pm r f_s. \end{cases} \quad (19.8)$$

It follows from Equation (19.8) that the signal $x(t)$ has discrete components: it contains the component at frequency f_0 and all aliasing frequencies corresponding to it and to the sampling frequency f_s. Equation (19.8) defines the amplitudes of these components. They decline in accordance with $|\operatorname{sinc}(\pi f \mu)|$. The component at the lowest frequency from the frequency row $0 < f = \pm f_0 \pm r f_s$, rather than the component at the frequency f_0, has the maximal value amplitude. However, it should be noted that, in the case of periodic sampling, any frequency from the row $0 < f = \pm f_0 \pm r f_s, r = 0, 1, 2, \ldots$, is represented by one and the same data set at the time instants $t_n, n = 1, 2, \ldots$. It follows from Equations (19.6) and (19.7) that the phase spectrum of the DAC output signal is given as

$$\varphi_x(f) = \begin{cases} \varphi_0 - \pi f / f_s = \varphi_0 - \pi f_0 / f_s \pm \pi r, & \text{if } 0 < f = f_0 \pm r f_s, \\ \pi - \varphi_0 - \pi f / f_s = -\varphi_0 + \pi f_0 / f_s + \pi(1 \mp r), & \text{if } 0 < f = -f_0 \pm r f_s, \\ 0, & \text{if } 0 < f \neq \pm f_0 \pm r f_s. \end{cases} \quad (19.9)$$

19.2.2 Synthesis in the Case of Multifrequency Basis

Consider the case where the basis function $x_0(t)$ is actually a sum of a certain number of sinusoidal components:

$$x_0(t) = \sum_{k=1}^{K} A_k \sin(2\pi f_k t + \varphi_k). \tag{19.10}$$

As the Fourier transform is a linear operation, the Fourier coefficients $a_x(f)$ and $b_x(f)$ for the DAC output signal $x(t)$ will differ from zero only at frequencies f_k, $k = 1, 2, \ldots$, and at all corresponding aliasing frequencies:

$$0 < f_{kr} = \pm f_k \pm r f_s, \quad k = 1, 2, \ldots, K; r = 0, 1, 2, \ldots. \tag{19.11}$$

Therefore the described synthesis transforms every sinusoidal component of the basis function into an infinite sum of sinusoidal components at frequencies defined by (19.11) and with amplitudes and phase angles determined by Equation (19.8) and (19.9) if f_0, A_0 and φ_0 are substituted by f_k, A_k and φ_k respectively. The component at the zero frequency or at frequencies equal to an integer number of f_s, i.e. if $f_k = r f_s, r = 0, 1, 2, \ldots$, represents an exception. In this case, all discrete values of that component are the same and it is accepted that this is a constant level and, consequently, the DAC output signal is also constant. Therefore if the basis function contains a component at any of the frequencies $f_k = r f_s, r = 0, 1, 2, \ldots$, it transforms into a constant level equal to $A_k \sin \varphi_k$.

All of the basis function frequencies clearly have to be chosen so that they do not overlap the frequencies of the rest of the basis function components or their aliases. Therefore any frequency of a basis function component $f_m, m = 0, 1, 2, \ldots, K$, has to meet the following condition:

$$f_m \bigcap \left(\bigcup_{\substack{k=1 \\ k \neq m}}^{K} \bigcup_{r=1}^{\infty} f_{kr} \right) = \phi, \quad m = 1, 2, \ldots, K, \tag{19.12}$$

where ϕ is an empty multitude while f_{kr} is determined by (19.11). The amplitude spectrum of the DAC output signal, for the case where the basis function is given by

$$x_0(t) = \sum_{k=1}^{5} \sin \left[2\pi \frac{(2k-1) f_s}{20} t + \varphi_k \right], \tag{19.13}$$

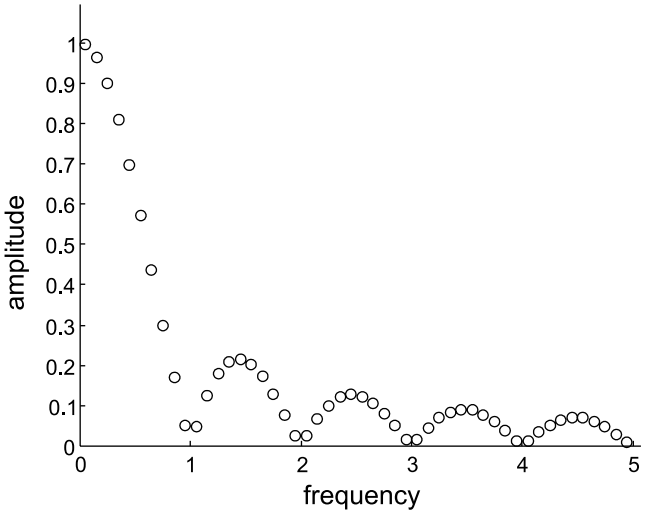


Figure 19.4 Amplitude spectrum of the DAC output signal in the case of a multifrequency basis function

is shown in Figure 19.4. The spectrum corresponds to the normalized frequency $f_s = 1$.

It can be seen from this analysis that when the frequencies of the basis function components cover the Nyquist range uniformly, the components of the DAC output signal cover uniformly the whole frequency axis. However, their amplitudes decline rather rapidly. If the basis function is represented by a sum of sinusoidal components, the DAC output signal has a discrete spectrum. Each of the basis function components at the frequency f_k, if $f_k \neq rf_s, r = 0, 1, 2, \ldots,$ is transformed into a sum of an infinite number of sinusoidal components with precisely predetermined frequencies, amplitudes and phase angles. Even when the basis function is relatively simple, the obtained synthesized signal can have a rich spectrum with precisely predetermined parameters of all components.

19.3 Test Signal Synthesis from a Nonuniformly Sampled Basis Function

Consider the case where a monoharmonic basis function $x_0(t)$ is sampled nonuniformly. The idealized DAC output signal $x(t)$ formed in this case is shown in Figure 19.5.

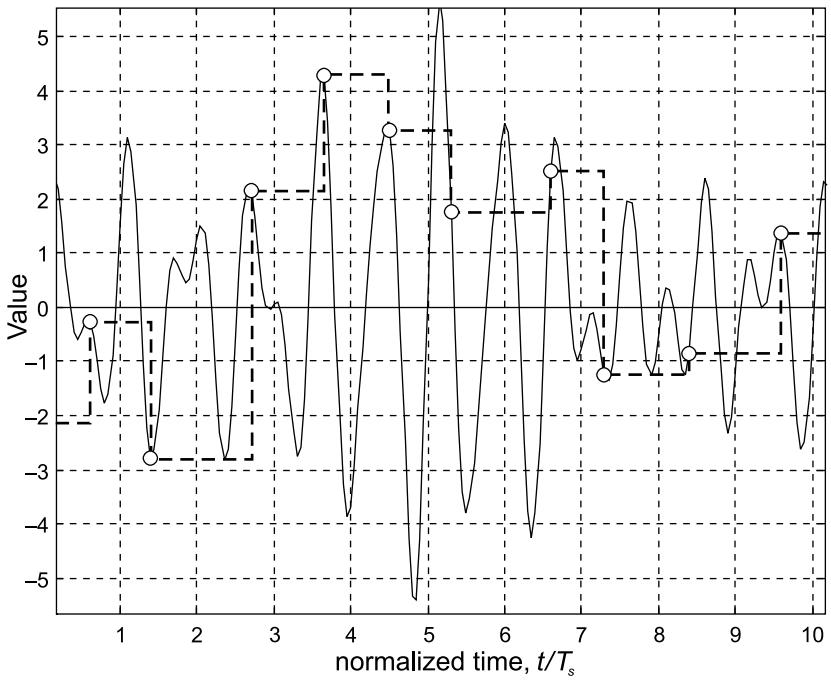


Figure 19.5 Illustration of forming the DAC output signal in the case of nonuniform sampling of the basis function

The basis function $x_0(t)$ with its sampled values marked at time instants t_k are shown in Figure 19.5 by a continuous line while the signal $x(t)$ at the DAC output is given by a dashed line. The aim is to find what are the properties of the analog signal obtained in this case and how they differ from the properties of a similar signal formed in the case of sparse periodic sampling.

19.3.1 Spectrum of the Synthesized Signal

The spectrum of the analog signal $x(t)$ at the DAC output in this case is defined as follows:

$$S_x(f) = \lim_{N\to\infty} \frac{2}{(2N+1)\mu} \sum_{n=-N}^{N} A_0 \sin(2\pi f_0 t_n + \varphi_0) \int_{t_n}^{t_{n+1}} \mathrm{e}^{-\mathrm{j}2\pi f t}\, \mathrm{d}t. \tag{19.14}$$

The notation $\mu_n = t_{n+1} - t_n$ is introduced. The Fourier coefficient can be defined as

$$\begin{aligned} a_x(f) &= \lim_{N\to\infty} \frac{2A_0}{(2N+1)\mu} \sum_{n=-N}^{N} \sin(2\pi f_0 t_n + \varphi_0) \int_{t_n}^{t_n+\mu_n} \cos 2\pi f \, \mathrm{d}t \\ &= \lim_{N\to\infty} \frac{A_0}{(2N+1)2\pi f \mu} \sum_{n=-N}^{N} \{\cos[2\pi(f - f_0)t_n + 2\pi f\mu - \varphi_0] \\ &\quad - \cos[2\pi(f_0 + f)t_n + 2\pi f\mu + \phi_0] - \cos[2\pi(f_0 - f)t_n - \varphi_0] \\ &\quad + \cos[2\pi(f_0 + f)t_n + \phi_0]\}. \end{aligned} \tag{19.15}$$

Suppose that the sampling intervals are formed as follows:

$$\mu_n = \mu + r_n\delta,$$

where the mean value of these intervals is $\mu = M\delta$, M is a positive integer, δ is the digit for the intervals $m_n, r_n = 0, \pm1, \pm2, \ldots, \pm m$, and $f_0 < \frac{1}{2}\delta$. Therefore there are $(2m+1)$ various intervals m_n. It is assumed that all of them are repeated for an equal number of times. The time instant t_n is defined as

$$t_n = (N + n + 1)\mu + \sum_{i=1}^{N+n+1} r_i\delta = \left[M(N + n + 1) + \sum_{i=1}^{N+n+1} r_i \right] \delta. \tag{19.16}$$

This kind of sampling point process is statistically equivalent to a periodic point stream with random skips obtained in the case where the sampling frequency is equal to $1/\delta$. From Equation (19.15), by taking into account (19.16), it is found that

$$a_x(f) = \begin{cases} \lim\limits_{x\to\infty} \dfrac{A_0}{(2N+1)2\pi\,(i/\delta + f_0)\mu} \displaystyle\sum_{n=-N}^{N} [\cos(2\pi f_0\mu_n - \varphi_0) - \cos\varphi_0] \\ \qquad \text{if } f = i/\delta + f_0,\ i = 0, 1, 2, \ldots, \\ \lim\limits_{x\to\infty} \dfrac{A_0}{(2N+1)2\pi\,(i/\delta - f_0)\mu} \displaystyle\sum_{n=-N}^{N} [\cos\varphi_0 - \cos(2\pi f_0\mu_n - \varphi_0)] \\ \qquad \text{if } f = i/\delta - f_0,\ i = 1, 2, \ldots, \\ 0 \qquad \text{for other } f. \end{cases} \tag{19.17}$$

By taking into account the properties of the considered sampling point process, the following sum can be calculated:

$$\begin{aligned}\sum_{n=-N}^{N}\cos(2\pi f_0\mu_n-\varphi_0) &= \frac{2N+1}{2m+1}\sum_{r=-m}^{m}\cos[2\pi f_0(\mu+r\delta)-\varphi]\\ &= \frac{(2N+1)\cos(2\pi f_0\mu-\varphi_0)\sin\pi f_0\delta(2m+1)}{(2m+1)\sin\pi f_0\delta}\\ &= (2N+1)\frac{\operatorname{sinc}[\pi f_0\delta(2m+1)]}{\operatorname{sinc}(\pi f_0\delta)}\cos(2\pi f_0\mu-\varphi_0)\end{aligned} \tag{19.18}$$

Substitution of Equation (19.18) into Equations (19.17) yields

$$a_x(f)=\begin{cases}\pm\dfrac{A_0}{2\pi(i/\delta\pm f_0)\mu}\left\{\dfrac{\operatorname{sinc}[\pi f_0\delta(2m+1)]}{\operatorname{sinc}(\pi f_0\delta)}\cos(2\pi f_0\mu-\varphi_0)-\cos\varphi_0\right\} & \\ \qquad\qquad\qquad\qquad \text{if } f=i/\delta+f_0, i=0,1,2,\ldots, & (19.19)\\ 0 \qquad\qquad\qquad\qquad \text{for other } f.\end{cases}$$

The second Fourier coefficient $b_x(f)$ is obtained in a similar way:

$$b_x(f)=\begin{cases}\dfrac{A_0}{2\pi(i/\delta\pm f_0)\mu}\left\{\dfrac{\operatorname{sinc}[\pi f_0\delta(2m+1)]}{\operatorname{sinc}(\pi f_0\delta)}\sin(2\pi f_0\mu-\varphi_0)-\sin\varphi_0\right\} & \\ \qquad\qquad\qquad\qquad \text{if } f=i/\delta+f_0, i=0,1,2,\ldots, & (19.20)\\ 0 \qquad\qquad\qquad\qquad \text{for other } f.\end{cases}$$

Equations (19.19) and (19.20) lead to the module of the Fourier transform:

$$A_x(f)=\begin{cases}\dfrac{A_0}{2\pi(i/\delta\pm f_0)\mu}\left(1+\left\{\dfrac{\operatorname{sinc}[\pi f_0\delta(2m+1)]}{\operatorname{sinc}(\pi f_0\delta)}\right\}^2-2\dfrac{\operatorname{sinc}[\pi f_0\delta(2m+1)]}{\operatorname{sinc}(\pi f_0\delta)}\cos 2\pi f_0\mu\right)^{1/2} & \\ \qquad\qquad\qquad\qquad \text{if } f=1/\delta\pm f_0, i=0,1,2,\ldots, & (19.21)\\ 0 \qquad\qquad\qquad\qquad \text{for other } f.\end{cases}$$

It can be seen from Equations (19.21) that the analog signal has a discrete component at frequency f_0 of the basis function and also has all aliasing frequencies corresponding to frequencies f_0 and $1/\delta$. Now it should be checked whether there are only discrete components present in the synthesized analog signal $x(t)$ or whether there are also additional noisy components with a continuous spectrum. The frequencies of the discrete components in increasing order are the

following:

$$\{f_0, 1/\delta - f_0, 1/\delta + f_0, 2/\delta - f_0, 2/\delta + f_0, \ldots\}. \tag{19.22}$$

Their power, on the basis of Equations (19.21), can be calculated as

$$P_{xk} = \frac{A_x^2(f_k)}{2}, \tag{19.23}$$

where f_k is the kth frequency of the sequence (19.22).

If the pseudo-random sampling process is properly arranged, the power of the aliasing components is usually negligible. Therefore, almost all of the total power of all signal $x(t)$ discrete components is concentrated within the first component at frequency f_0. However, the signal $x(t)$ contains a component with a continuous spectrum, in other words, noise.

19.3.2 Multifrequency Signal Synthesis

The properties will be established of the DAC output signal in the case of a nonuniformly sampled basis function represented by a sum of a certain number of sinusoidal components, when the basis function is described by Equation (19.13) at $K > 1$. It has been already shown that when the basis function is sampled nonuniformly with the smallest time digit δ, any sinusoidal component of it transforms into a sum of an infinite number of sinusoidal components with given frequencies, amplitudes and phases, plus noise due to the cross-interference caused by sampling irregularities. If the value of the time digit is small enough, the aliases relative to the frequency $1/\delta$ could be neglected as their amplitudes are very small. Therefore, in all practically significant cases, it could be assumed that any component of the basis function is transformed into a sinusoidal component at the same frequency and noise. The amplitude and phase angle of this discrete component in the DAC output signal is determined by Equations (19.19) and (19.20) if f_0, A_0 and φ_0 are replaced by f_k, A_k and φ_k, respectively. The noise has declining spectral density and its power is equal to $(A_k^2 - A_{xk}^2)/2$, where A_{xk} is the amplitude of the DAC output signal discrete component.

Thus, when the basis function $x_0(t)$ is a sum of several sinusoidal components, the obtained analog signal at the DAC output, due to the linearity of the Fourier transforms, will have components at all those frequencies with precisely predetermined amplitudes and phases (defined by Equations (19.19) and (19.20)) in a mixture with the cross-interference induced noise. The amplitude spectrum of the DAC output signal, corresponding to the nonuniformly sampled multifrequency basis function, is given in Figure 19.6(a) for the same basis function (19.13)

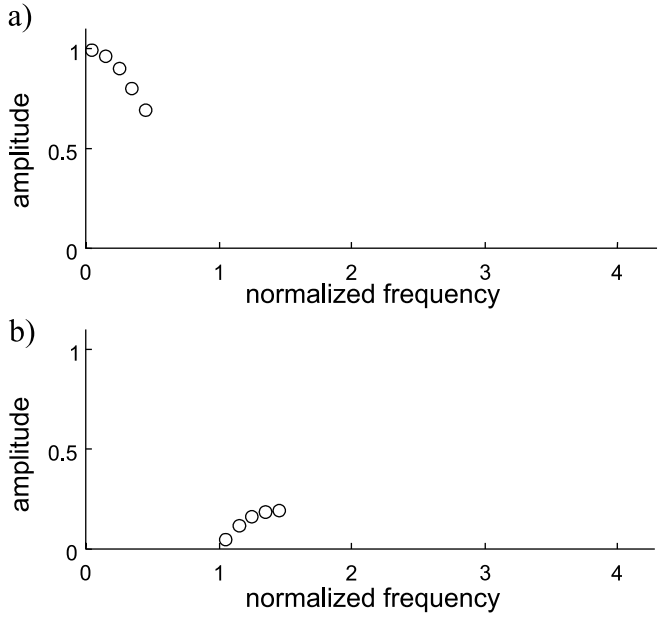


Figure 19.6 Amplitude spectra of the DAC output signal in the case of nonuniformly sampled multifrequency basis function: (a) in the case of basis function (19.13); (b) in the case of basis function (19.24)

used before as an example at the discussion of the periodic sampling case. The amplitude spectrum of the analog signal at the output of the DAC, shown in Figure 19.6(b), corresponds to the case of nonuniform sampling of the following basis:

$$x_0(t) = \sum_{k=1}^{5} \sin\left\{2\pi\left[f_s + \frac{(2k-1)f_s}{20}\right]t + \varphi_k\right\}. \tag{19.24}$$

As the same frequencies in Equations (19.13) and (19.24) (at the same k) are overlapping, these functions when sampled periodically are equivalent; therefore the corresponding analog signals at the DAC output including their amplitude and phase spectra are also equivalent.

That is not the case when the basis functions are sampled nonuniformly. Then the obtained analog signal contains only those frequencies that are in the spectrum of the corresponding basis function. The amplitudes of the analog signal components are lower for higher frequencies and, with frequencies growing, the power of the additional noise, due to the cross-interference, increases.

19.3.3 Amplitude Equalization

As shown in the previous section, it is possible to synthesize an analog signal from scarce data obtained under conditions when the well-known requirements of the sampling theorem are not met and when the frequencies of the signal to be synthesized exceed half of the DAC switching rate. However, there are also some problems and limitations. First of all, the amplitude of the synthesized sinusoidal signal mainly depends on its frequency. That might cause problems when the analog signal to be synthesized contains a number of frequencies. As that is exactly what is typically needed for the generation of multifrequency test signals, the feasibility of synthesizing such a multifrequency signal will be considered.

Apparently various types of multifrequency test signal are needed. In the context of the considered application, a group of frequencies typically has to be generated around a certain central frequency. Even in this case, there are a number of possible variations. Amplitude spectra of some of them are given in Figure 19.7.

Probably the most popular structure of a multifrequency test signal is the one shown in Figure 19.7(a). All frequencies of it are equidistanced on the frequency axis and amplitudes of all components are equal. However, as discussed further, the price of achieving the equality of amplitudes may sometimes be too high. Quite often it might be better to accept some amplitude variation. Such amplitude variations, if predetermined, should not represent a problem, especially if all component amplitudes are given with sufficiently high precision. This kind of multifrequency signal is given in Figure 19.7(b). The third type of multifrequency signals, illustrated in Figure 19.7(c), differs from the others by the location of the frequencies. In this case the signal components are placed on the frequency axis irregularly at arbitrary places. How beneficial this possibility could be is another question; here the fact is simply demonstrated that the suggested synthesis method would permit that possibility. When the traditional methods for generating such multifrequency test signals are used, it certainly is not simple to provide for this functional feature.

While there are no problems at synthesizing signals with specified frequencies, except that they cannot be placed too closely together, it is quite another situation with the synthesis of multifrequency signals with equal or specified amplitudes of their components. Direct application of the suggested synthesis approach leads to the synthesis of signals with component amplitudes defined by Equation (19.8) and illustrated by Figure 19.4. It follows from Equation (19.8) that when the frequency of a component grows its amplitude declines according to $|\mathrm{sinc}(\pi f/f_s)|$, where f_s is the sampling frequency. In addition, the amplitudes of components close to frequencies $f_s, 2f_s, f_s, \ldots$ are very small (frequencies $f_s, 2f_s, \ldots$ cannot

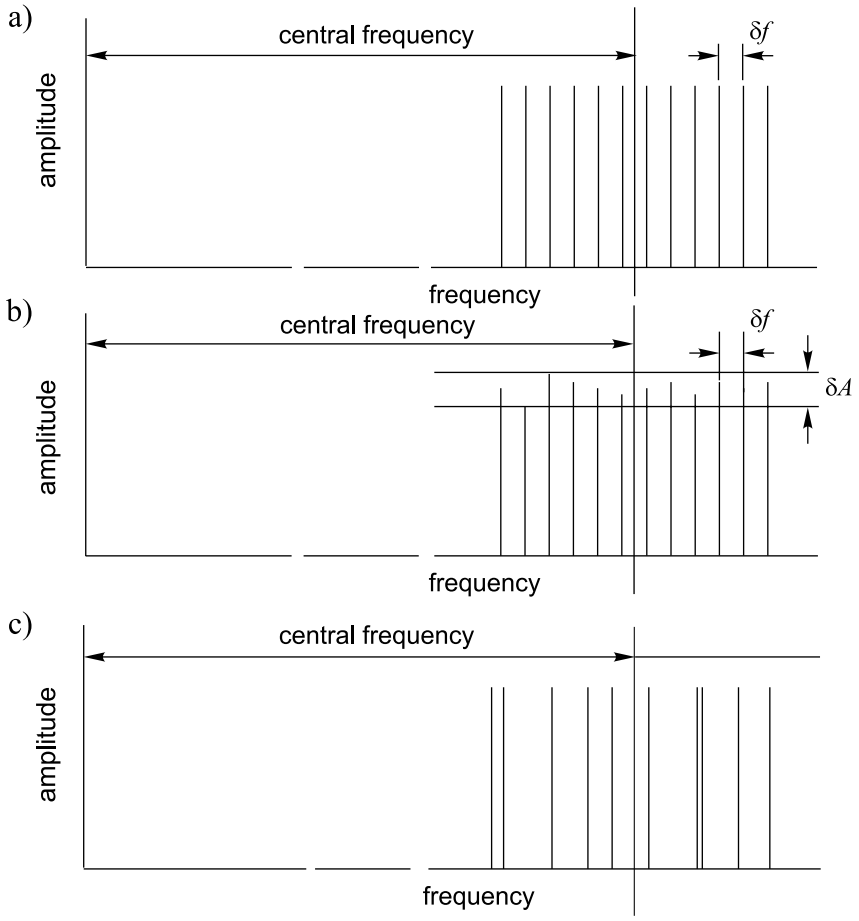


Figure 19.7 Amplitude spectra of various types of multifrequency signals to be synthesized

be used as they are transformed into a constant level). Therefore there may be considerable measurement problems at those frequencies where the components are small.

This problem of how 'small amplitudes' might be resolved are considered. There may be at least two distinctly different situations.

19.4 Synthesis of Narrowband and Wideband Signals

Suppose a passband signal $z(t)$ has to be synthesized so that it has a passband in the frequency range ($f_{\min}$, $f_{\max}$) and that it is possible to chose a realizable sampling

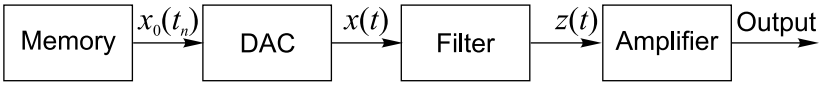


Figure 19.8 An optional structure for synthesis of a passband signal

frequency f_s to sample the basis function so that the frequency band (f_{min}, f_{max}) can be completely inserted within the interval $(if_s, (i + 0.5)f_s)$, or within the interval $((i + 0.5)f_s, (i + 1)f_s))$, where i is a positive integer. To obtain such a signal, the basis function is formed as explained above so that the components of it $(f_1, f_2, \ldots, f_K)$ cover the frequency band (f_{min}, f_{max}) with the frequency step δf. This basis function is sampled and a DAC is used to convert the obtained digital signal into an analog signal. Then the DAC output signal is passed through an analog bandpass filter having the same passband (f_{min}, f_{max}). The structure of such a synthesizer is illustrated by Figure 19.8.

In the case of an ideal filter, the amplitudes of the signal $z(t)$ components are determined by Equation (19.8). Therefore if the amplitudes of the basis function components are taken as

$$A_K = \frac{A_0}{|\text{sinc}(\pi f_k f_s)|}, \tag{19.25}$$

where $k = 1, 2, \ldots, K$, the components of the synthesized signal $z(t)$ will have equal amplitudes A_0. The dynamic range of the signal $z(t)$ is naturally narrower than that of the DAC output signal (t). It is essential to know their ratio. Let D_x and D_z be the dynamic ranges of signals $x(t)$ and $z(t)$. Then their ratio $\rho_{xz} = D_x/D_z$.

Figure 19.9 illustrates this approach to signal synthesis where the amplitudes of the basis function have been calculated on the basis of Equation (19.25). Therefore they differ for all components, which leads to equalization of the synthesized signal $z(t)$ component amplitudes, as shown in Figure 19.9(a). The dynamic ranges of signals $x(t)$ and $z(t)$ are obviously different in comparison with the respective dynamic ranges when the output signal component amplitudes are not equalized. In this case $\rho_{xz} = 4.4$.

This effect of ρ_{xz} enlargement rapidly increases if the frequency range (f_{min}, f_{max}) approaches any of the frequency points if_s. For instance, if (f_{min}, f_{max}) is defined as $(1.01f_s, 1.41f_s)$ and again nine components are placed within it with a frequency step of $0.05f_s$ then the named ratios ρ_{xz} for the cases without and with output signal component amplitude equalization are 6.0 and 27.0 respectively. This is due to the fact that the frequency f_{min} has approached frequency $f_{s.}$ This can also be seen from Equation (19.25) as near to if_s the

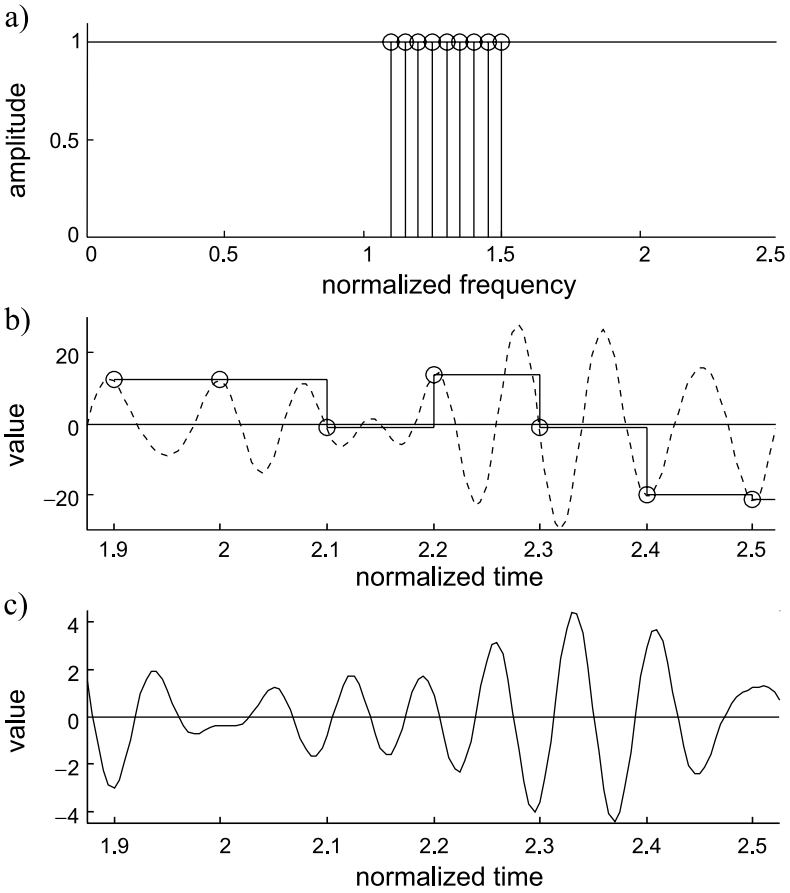


Figure 19.9 Synthesis of a narrowband multifrequency signal with component amplitude equalization: (a) amplitude spectrum of the equalized synthesized signal $z(t)$; (b) basis function $f_0(t)$ (dashed line) and DAC output signal $x(t)$ (continuous stepwise line); (c) the obtained signal $z(t)$ at the filter output

function $|\text{sinc}(\pi f_k/f_s)|$ approaches zero and therefore the corresponding amplitudes of the basis function components have to be large.

Figure 19.10 shows how the amplification coefficient of the basis signal components depends on the normalized frequency (f_k/f_s). The curve is given in the logarithmic scale.

Suppose that the signal $z(t)$ to be synthesized is wideband, the available DAC satisfies the condition $f_{max} - f_{min} > f_s/2$ and the amplitudes of the signal components have to be equalized. In this case it makes sense to subdivide the whole band

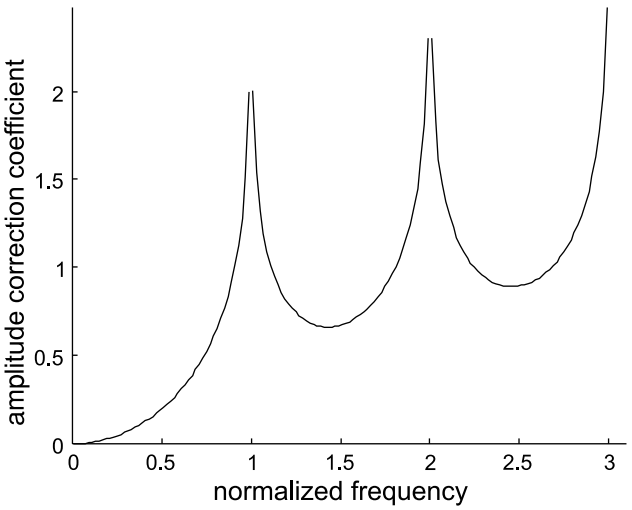


Figure 19.10 Amplification coefficient of the basis signal components versus the normalized frequency (f_k/f_s)

($f_{\min}$, $f_{\max}$) into sub-bands. After that proper sampling frequency has to be chosen for each one of them so that every frequency subinterval is completely placed within one of the following intervals: $(if_s, (i + 0.5)f_s)$ or $((i + 0.5)f_s, (i + 1)f_s)$. Therefore this approach to the multifrequency signal synthesis is based on sequential in the time synthesis of a number of signals containing components at frequencies from the corresponding frequency interval.

19.5 Measuring Small Delays and Switching Times

As shown in Chapter 6, the density of sampling points at additive randomized sampling tends to a constant level inversely proportional to the mean sampling interval $\bar{q}$. When a signal is sampled according to such a sampling model all of its instantaneous values are sampled with equal and constant probability. This valuable feature of the additive sampling model could be beneficially exploited to achieve a high resolution when measuring various signal parameters, such as delays in the reaction of objects to some excitation signals, switching times and differences in signal phase angles. Tasks of this type are apparently based on time-interval measurements.

According to the traditional and most often used approach to digital time-interval measurements, short clock pulses are generated at a high and stabilized frequency and are then passed through a gate that is opened for the duration of

the time interval to be measured. The pulses that pass through are registered by a counter. All is well as long as the clock period is much smaller than the time interval being measured because the ± 1 count quantization error, characterizing this measurement method, is acceptable only as long as the number of pulses counted is sufficiently large. Measuring shorter time intervals is more difficult. A fairly efficient solution to this problem is based on the use of randomized pulse sequences mathematically described identically to the description of the additive sampling point processes. Although the randomization techniques concern time-interval quantizing rather than sampling, the statistical relationships describing both of those procedures are in fact the same.

Block diagrams of two implementation schemes of the suggested randomized time measurement method are given in Figure 19.11. Short pulses generated by random pulse generators represent the quantizing thresholds in this case. These pulse sequences are generated in exactly the same way as in the case of the additive random sampling, i.e. these pulses are formed at time instants given by Equation (6.4). The performance of the schemes shown in Figure 19.11 depends first of all on the perfection of the voltage comparators used to generate the inputs. The comparators should be sensitive wideband devices, capable of detecting small input and reference voltage U_0 differences. Their aperture time should be in the picosecond range.

The time intervals to be quantized are given as durations of the input pulses to be measured at some level determined by the reference voltage U_0. The number of time intervals to be averaged is preset and counter 1 in both schemes is used for counting these intervals in order to stop the quantizing and averaging process after a given number of pulses have been processed. Note that in the case of the scheme in Figure 19.11(b), the number of these time intervals is determined indirectly by actually counting the pulses that fall within time intervals of a constant and known duration, generated by the single-shot time base generator triggered by the appearance of each time interval to be measured. The quantization and averaging results obtained are registered in both schemes by counters 2. The time intervals at the input can be repeated either randomly or periodically.

When the first of these two schemes (Figure 19.11(a)) is used, the result of quantizing a single time interval is given as

$$\widehat{\Delta t}_i = n_i \bar{q}, \tag{19.26}$$

where $\bar{q}$ is the mean value of the time intervals between the quantizing pulses. The estimate $\widehat{\Delta t}$ of the unknown duration Δt of the time intervals is obtained by

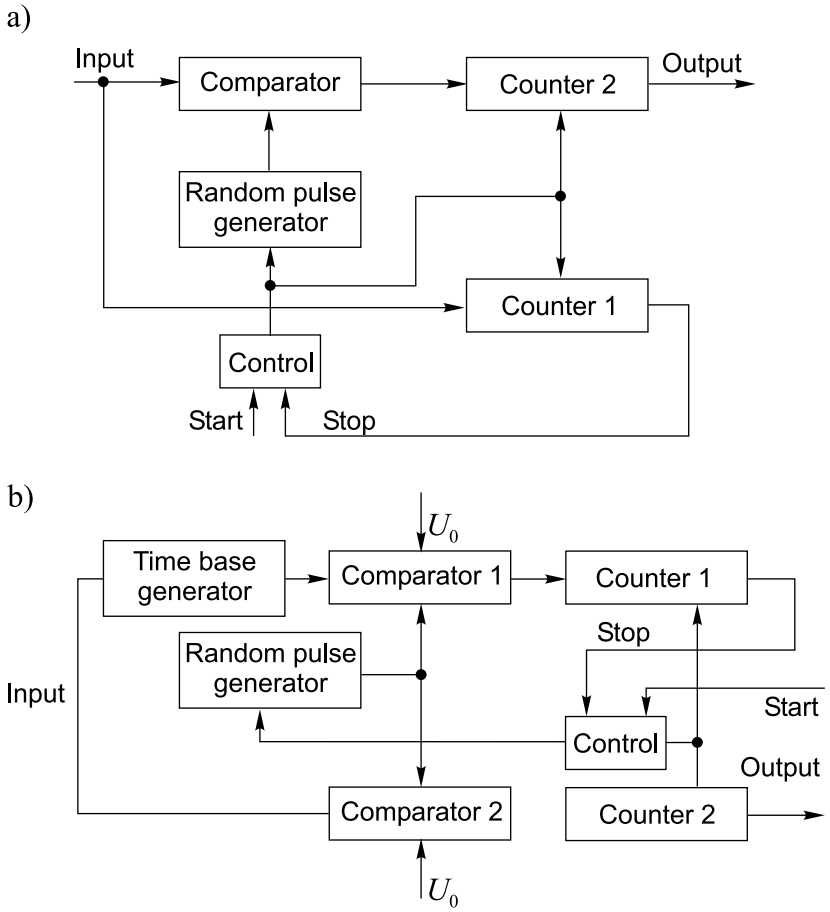


Figure 19.11 Block diagrams of two systems for measuring short time intervals

averaging N particular results $\widehat{\Delta}t_i$ as follows:

$$\widehat{\Delta}t = \frac{1}{N}\sum_{i=1}^{N} n_i \bar{q}. \tag{19.27}$$

If the pauses between the time intervals to be quantized are large enough, the subsequent particular quantization results can be considered to be statistically independent. Then it can be written that

$$\mathrm{Var}[\widehat{\Delta}t] = \frac{\bar{q}^2}{N}\sum_{i=1}^{N} \mathrm{Var}[n_i], \tag{19.28}$$

where $\mathrm{Var}[n_i]$ is the variation of the number of quantizing pulses falling within the time interval being measured. In the case of uncorrelated Δt_i, the random error can be written as

$$\varepsilon_r = t_\beta (\mathrm{Var}[\widehat{\Delta} t])^{1/2}. \tag{19.29}$$

Although this time-interval estimation scheme is simple and practical, the mean value $\bar{q}$ of the time intervals between the quantizing pulses or, in other words, the mean repetition rate of these pulses, should be stabilized, which is certainly a disadvantage.

The second time-interval estimation scheme, shown in Figure 19.11(b), does not have this disadvantage. This improvement is achieved by organizing the estimation process in such a way that the unknown input time intervals are compared with some constant time intervals T, rather than with the parameter $\bar{q}$ of the pulse sequence used for quantizing. Therefore, in this case, it is the time reference pulse duration T that has to be stabilized. These time reference pulses are generated whenever a time interval to be quantized appears at the input. The reference and input time intervals are quantized in parallel by means of the same quantizing pulse sequence. The quantization results m_i and n_i are entered into the respective counters 1 and 2. It can be written that

$$E[n_i] = \frac{\Delta t_i}{q} \qquad \text{and} \qquad E[m_i] = \frac{T}{\bar{q}}.$$

Therefore

$$\Delta t = \frac{E\,[n_i]}{E\,[m_i]} T$$

and the estimate $\widehat{\Delta} t$ of the time interval Δt, obtained by averaging N quantization results, is given as

$$\widehat{\Delta} t = \frac{T \sum_{i=1}^{N} n_i}{\sum_{i=1}^{N} m_i}. \tag{19.30}$$

An estimation of Δt can be carried out until either N or the sum of m_i reaches some preset value m. The second approach is preferable and, if it is applied,

$$\Delta t = \frac{T}{m} \sum_{i=1}^{N} n_i. \tag{19.31}$$

In order to carry out the averaged quantizing of time intervals according to Equation (19.31), the capacity of counter 1 should be set equal to m and quantizing

Table 19.1

Time interval (ns)		Relative error (%)	
True values	Measurement results	Expected values	Estimated experimentally
2.00	1.85	2.06	1.65
2.65	2.63	1.65	2.04
5.00	5.18	1.02	0.84
10.00	10.07	0	0.30

should be stopped when this counter is filled up. Then the readout from counter 2 will represent $\widehat{\Delta t}$ in time units.

The advantage of this method for averaged measuring of time intervals, unlike the classical deterministic measurement method, is that it can be applied for evaluating time intervals shorter than $\bar{q}$ when no more than one pulse can fall within the time interval to be quantized and n_i and m_i can assume only the values of 0 or 1. When a quantizing pulse coincides with the time reference interval, the probability that it will also fall within the interval Δt is equal to $\Delta t/T$. It can therefore be written that

$$E[\widehat{\Delta t}] = \frac{T}{m} m \frac{\Delta T}{T} = \Delta t \qquad \text{and} \qquad \text{Var}[\widehat{\Delta t}] = \frac{\Delta t T}{m}\left(1 - \frac{\Delta t}{T}\right). \quad (19.32)$$

Hence the random estimation error

$$\varepsilon_r = t_\beta \left[\frac{\Delta t T}{m}\left(1 - \frac{\Delta t}{T}\right)\right]^{1/2}, \quad (19.33)$$

where t_β is half of the confidence interval corresponding to the confidence probability β.

Note that under the described measurement conditions the random estimation error does not depend on the probabilistic characteristics of the quantizing pulse sequence used. This random quantizing pulse sequence is used in this case as a tool for comparing Δt with the reference value T.

The capabilities of this digital time-interval measurement approach are to some extent illustrated by the experimental measurement results given in Table 19.1. They are obtained under the following conditions: Δt varies in the range 2–10 ns and $m = 2^6 \times 10^3$ calculations were made for the 99 % confidence level. The measurements of each Δt value were repeated 20 times.

The described digital short time-interval measurement techniques can be used for various applications based on the delay, phase angle or pulse rise and fall time measurements. This approach has been proved to be of practical value whenever

it is essential that the designs of systems are extremely simple. The fact that the time intervals to be measured in this way may be repeated either randomly or periodically at relatively high frequencies might be considered as an additional advantage. It is also relatively easy to achieve high precision. Substantial practical experience has been gained in this field. A number of test and measurement systems used to measure dynamic parameters of various microelectronic elements in the process of their manufacture have been developed on this basis and successfully used. This experience confirms the reality of achieving, in a simple way, subnanosecond resolution usually needed for digital measurements of short time intervals.

19.6 Bioimpedance Signal Demodulation in Real-time

It has been discovered that various processes going on in a human body change the bioimpedance of it and that these changes carry valuable diagnostic information. To obtain this information, a test signal in the form of a constant voltage specific frequency is applied to the body being examined and the bioimpedance signal, reflecting the reaction of the biological object to this excitation signal, is extracted and analysed. To do this, the picked-up signal needs to be demodulated. Real-time demodulation of bioimpedance signals is considered here as an example showing what could be gained by exploiting the digital alias-free signal processing technology.

19.6.1 Typical Conditions for Bioimpedance Signal Forming

Variations of the impedance of a human body due to cardiac activities and breathing lead to modulation of the test signal. Amplitude as well as phase modulation of the carrier takes place. Therefore the modulated carrier has to be processed in a way that leads to the discovery of both the amplitude and phase changes. Demodulation of bioimpedance signals can be performed in various ways. A particular universal solution of this demodulation task could be based on the estimation of both Fourier coefficients at the carrier frequency or at all involved carrier frequencies when a number of test signals at differing frequencies are generated and used. The Fourier transforms in this case need to be performed in real-time.

More often than not, the structure of bioimpedance analysers is multichannel. In addition, bioimpedance signals sometimes have to be obtained at a number of test frequencies simultaneously. This means that a large amount of data typically needs to be acquired and handled. This represents a problem, especially when the

bioimpedance signals have to be observed for a long time. Then it is desirable to preprocess the data in a cost-effective way and to compress them before these data are transferred to the host computer. The technical characteristics of this signal preprocessing unit in terms of complexity, volume, power consumption and operational speed depend on the method used for bioimpedance signal demodulation and its hardware and software implementation. Therefore it is essential to develop and use a good enough method for bioimpedance signal demodulation.

It has been found that impedance signals at various frequencies carry different information. This means that the test signal frequencies could vary in a wide frequency range, from relatively low frequencies of about 100 kHz up to 1 GHz. This is one of the conditions that makes this application area well suited for the specifics of the digital alias-free signal processing technology.

The next essential condition that complicates the bioimpedance signal on-line demodulation task considerably is the relatively large number of input signals that need to be processed in parallel. Typically there are 12 inputs, but sometimes even more inputs are needed.

The demodulation task is simplified to some extent by the fact that the modulations are typically within a range of frequencies that is much lower than the carrier frequencies. The cardiac related and breathing processes that basically modulate the test signal are typically in the frequency range not exceeding 1 kHz. Therefore relatively few readings of the Fourier coefficients need to be obtained in a second to present the demodulated signal sufficiently well. That helps to process a number of bioimpedance signals in parallel.

19.6.2 Complexity Reduction of Bioimpedance Signal Demodulation

Analysis of the mentioned considerations has led to the conclusion that the method developed earlier and discussed in Chapter 16 for the complexity-reduced DFT represents a promising option for demodulation of the bioimpedance signals under the given conditions. Using this approach for development of such a system has provided results close to those expected.

The outline of the suggested approach is as follows. In general, the Fourier analysis is accomplished in two stages. In the first stage, the signal to be analysed is formally decomposed on the basis of some rectangular functions assuming only the values -1, 0 and $+1$, so this decomposition can be carried out without performing the multiplication operations. In the second stage a spectral conversion is performed. The set of coefficients α_i and β_i obtained in the first stage are then converted into a set of Fourier coefficients.

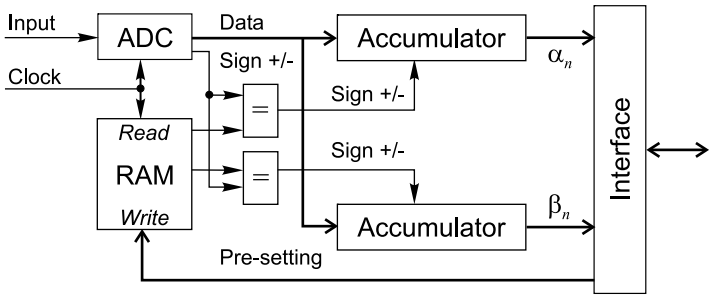


Figure 19.12 Block scheme of the bioimpedance signal preprocessor performing the signal demodulation and data compression functions

It follows from the description of the complexity-reduced DFT method given in Chapter 16 that this is especially attractive for applications where the spectrum of the signal to be analysed does not contain harmonics above the third (when system 1 is used) or the fifth (in the case of using system 2). Then the spectral estimations are actually single stage and no multiplication operations are executed. The conditions for bioimpedance analysis seem to fit this criterium. Multitone bioimpedance analysis might sometimes represent an exception, but at least the single carrier frequency analysis could be well implemented on the basis of this approach.

A number of benefits could be gained in this way. Firstly, the hardware for demodulation is simple. The basic scheme for estimation of Fourier coefficients is basically built on data accumulators. Secondly, real-time demodulation could be easily implemented on this basis by making estimation of the Fourier coefficients continuously on-line. Thirdly, due to the simplicity of the basic structure, multichannel demodulation could also be realized, which could be done in a sufficiently simple way. Fourthly, in addition, the approach is especially attractive for embedded bioimpedance analysers as they are characterized by low power consumption.

All of the mentioned considerations lead to the conclusion that demodulation of the bioimpedance signals can be arranged in accordance with the scheme given in Figure 19.12. Although Figure 19.12 illustrates the structure of a single-channel bioimpedance signal demodulator, the simple signal preprocessor shown there actually would be exactly the same in the case of a multi-channel demodulator. The ADC then has simply to be replaced by a multiplexer.

Note that estimates of the coefficients α_n and β_n rather than sample values of the modulated carrier are passed through the interface block to the computer. In

this case these coefficients are equal to the corresponding Fourier coefficients a_n and b_n. Thus this very simple signal preprocessor performs the DFT and substantial data compression is achieved. This makes the task of the computer receiving these data much easier. It simply has to collect the preprocessed data and display the demodulated signals.

Bibliography

Artjuhs, J. and Bilinskis, I. (2006) Method and apparatus for alias suppressed digitizing of high frequency analog signals. EP 1 330 036 B1, European Patent Specification, Bulletin 2006/26, 28.06.2006.

Artyukh, Yu., Bilinskis, I., Greitans, M. and Vedin, V. (1997) Signal digitizing and recording in the DASP-Lab System. In Proceedings of the 1997 International Workshop on *Sampling Theory and Application*, Aveiro, Portugal, June 1997, pp. 357–60.

Artyukh, Yu., Bilinskis, I. and Vedin, V. (1999) Hardware core of the family of digital RF signal PC-based analyzers. In Proceeding of the 1999 International Workshop on *Sampling Theory and Application*, Loen, Norway, 11–14 August 1999, pp. 177–9.

Artyukh, J., Bilinskis, I., Boole, E., Rybakov, A. and Vedin, V. (2005) Wideband RF signal digititising for high purity spectral analysis. In Proceedings of the International Workshop on *Spectral Methods and Multirate Signal Processing (SMMSP 2005)*, Riga, Latvia, 20–22 June 2005, pp. 123–8.

Artyukh, Yu., Boole, E. and Vedin, V. (2006) Digital synchronous demodulator for measurement of complex amplitude deviation. electronics and electrical engineering. *Kaunas: Technologija*, **5**(69), 29–32.

Bilinskis, I. and Artjuhs, J. (2006) Method and apparatus for alias suppressed digitizing of high frequency analog signals. United States Patent US 7,046,183 B2, 16 May 2006.

Bilinskis, I. and Sadovskis, G. (1974) Statistical measurement of phase angles (in Russian). In Proceedings of Conference on *Phasometric Systems and Devices*, Tomsk, pp. 170–4.

Bilinskis, I., Trejs, P.P. and Nemirovski, R.F. (1972) Verfarhen zur digitalen Messung von Zeitintervallen und Einrichtung zu dessen Realisierung. Patentscrift 93960, DDR.

Bilinskis, I., Trejs, P.P. and Nemirovski, R.F. (1974) Zpusob digitalniho mereni casovych intervalu a zarizeni k provadeni tohoto zpusobu. Aut. osved. 163383, CSSR.

Min, M., Ollmar, S. and Gersing, E. (2003) Electrical impedance and cardiac monitoring – technology, potential and applications. *International Journal of Bioelectromagnetism*, **5**(1), 53–6.

20

Encapsulating DASP Technology

At present there are no external obstacles preventing beneficial exploitation of the digital alias-free signal processing technology on a relatively large scale. The existing microelectronic elements could be used for developing the hardware needed for that and there does not seems to be any problems with creating software. Actually this is not exactly correct. There are problems that complicate widespread application of the signal processing technology discussed in this book. Basically these are related to the fact that considerable expertise is required to be successful in this area and it is not very easy to acquire this. Substantial investments in terms of effort, time and money are needed to gain knowledge and experience sufficient to achieve really significant positive results. It is not realistic to expect that building such expertise can be done in many places quickly. A more rational and practical approach to widening the application field for this technology is based on the idea that the benefits in this area could be gained more easily by developing and exploiting application-specific subsystems for digital alias-free signal processing. That clearly has to be done by embedding them in various larger IT systems operating as usual. This approach is discussed in this chapter. Among the gains immediately achievable in this way, widening of the digital domain in the direction of higher frequencies, simplification of data compression, complexity reduction of sensor arrays and achieving fault tolerance for signal processing could be mentioned. Issues essential for the development of the embedded systems capable of providing these benefits are discussed in this chapter in some detail.

Digital Alias-free Signal Processing I. Bilinskis

20.1 Linking Digital Alias-free Signal Processing with Traditional Methods

To embed specific digital alias-free signal processing subsystems into larger conventional systems operating on the basis of equidistant data representation in the time domain, all specific features of the subsystems to be embedded should be encapsulated within them while their inputs and outputs are defined in a way fitting the standards of the respective embedding system. The advantage of this approach is obvious: there is no need to pay any attention to what happens inside the particular embedded system.

Various embedded DASP systems might be needed to cover a sufficiently wide application field. There could be universal and application-specific modifications. However, the variety of the embedded systems should be minimized while the application field covered by them is maximized. The complexity of the universal embedded systems capable of performing flexible signal processing in a wide application range could be much higher than the design complexity of the embedded systems applicable for fulfilling specific functions. These are much simpler by definition. Consequently, the expected operational speed of the specific embedded systems is typically much higher than the data processing rate achievable by the first class of systems. This fact increases the attractiveness of the specific embedded systems focused on fulfilling special functions. On the other hand, orientation towards specialized embedded systems leads to the necessity of having a longer list. To focus this discussion on the basic concepts of such embedded systems, the emphasis is put on a universal generic model.

20.1.1 Generic Model of the Embedded DASP Systems

Particular designs of the embedded DASP systems depend both on the microelectronic technology used for their implementation and on concepts underpinning them. Figure 20.1 illustrates a universal generic model of such an embedded system. It reflects the principles according to which this type of embedded system should be built in order to achieve their wide applicability for digital alias-free processing of various signals.

This version of the embedded DASP systems basically performs input signal encoding in an alias-free way and performs processing of the obtained digital signal in accordance with the algorithm used for programming the computer included in this scheme. The functional blocks inserted into the structure of this model are needed to fulfil a number of essential functions. One of the basic

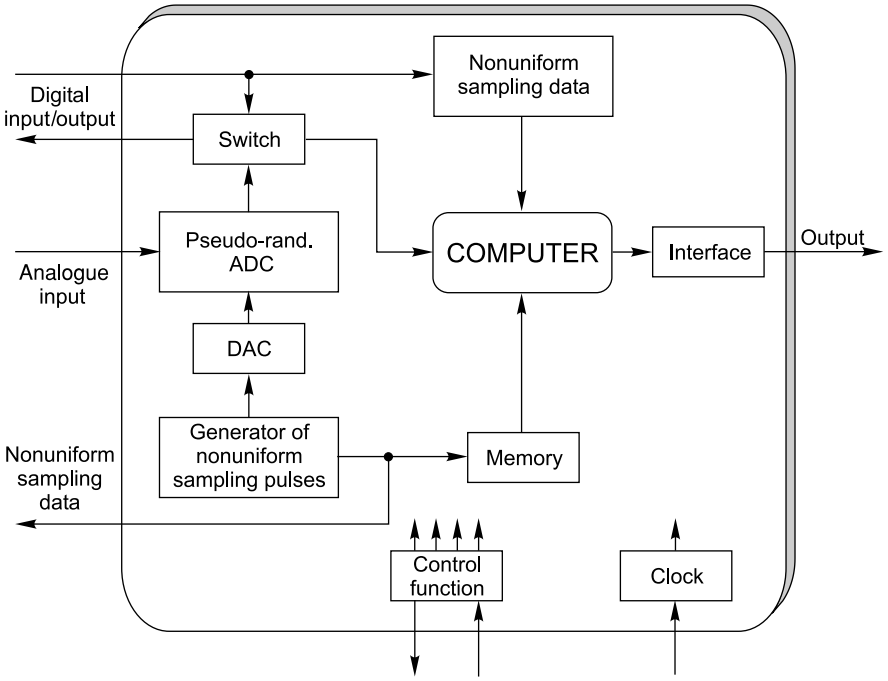


Figure 20.1 Structure of a generic embedded system for digital alias-free signal processing

functions of the considered embedded DASP systems is to provide a dedicated service of converting input signals into their digital counterparts and representing them in a format meeting the requirements of the algorithms for their effective alias-free processing. Therefore they should have an analog input and a special pseudo-randomized ADC for performing nontraditional signal sampling and/or the quantizing procedure.

The next block of particular significance is the computer. Capabilities of the embedded systems to a large extent depend on both the technical characteristics of the used computer and on the signal processing algorithms. It should also be realized that successful use of such embedded systems is possible only if signal processing is properly matched to the specifics of digitizing the signal. To ensure that, it is crucial to adapt processing of signals to the irregularities of sampling carried out at the stage of their digitizing. The information of the specific nonuniformities of the involved sampling process is kept in a memory and is used in the process of adapting signal processing operations to the signal sampling conditions.

While the embedded systems realized according to this generic model can be used for processing the input signals in many different ways, the following two operational modes can be described as the basic ones:

1. *Signal analysis/synthesis*. This operational mode is universal and very flexible. The essence of it is signal spectrum analysis with a following reconstruction of the input signal waveform and periodic resampling of it at a significantly increased frequency. This kind of signal processing requires relatively complicated iterative computations described in some detail in the following section. The problem is that the required calculations take a relatively long time and therefore it is not easy to achieve functioning of this system in real-time.
2. *Signal vital parameter estimation*. This second operational mode is also applicable in many cases. The point is that while waveforms of signals contain full information carried by them, only some parameters characterizing these waveforms often have to be estimated and used. Whenever this is the case, signal processing can be carried out in a way that is typically less complicated. This means that the programmable processors used for realizing this operational mode could provide the results substantially faster. In those cases, it is easier to achieve operation of this type of embedded system in real-time.

It may seem that these two operational modes differ substantially. In fact, this is often not true. Actually technical realization of the second operational mode might be simpler only in cases where some application-specific parameters characterizing some specific features of the signal are estimated. Whenever the parameters to be estimated are more universal, like the Fourier coefficients, amplitude and power, the complexity of both operational modes is approximately the same. This is so because it is crucial to adapt processing of nonuniformly sampled signals to the respective irregularities of the sampling process and the involved iterative procedure, basically determining the complexity of calculations, is based on direct and inverse Fourier transforms repeated several times in both cases.

In general, the role of the embedded systems for digital alias-free signal processing is to provide the dedicated service of digital processing of the input analog signal in a significantly widened frequency range, with representation of the output signal in a digital standard format that is normally and widely used to execute the standard instructions of traditional DSP algorithms. The DSP technology, well-developed over a long period of time, and the wealth of classic DSP algorithms, including algorithms for versatile digital filtering, could be fully used for further processing the signal digital waveforms and obtained signal parameter estimates whenever the described embedded systems could be and were used.

Incorporation of these universal embedded systems into traditional IT systems apparently leads to significant widening of the frequency range where signals could be processed fully digitally. However, other types of various significant benefits could be gained in this way as well.

20.1.2 Various DASP System Embedding Conditions

The conditions for matching inputs and outputs of the embedded DASP systems to the specifics of the traditional signal processing master systems obviously differ. Some of the variations most often met in conditions for application of the discussed embedded systems are considered.

Using Embedded Systems for Processing Pseudo-randomly Represented Signals of the Considered Type in Parallel

To connect specifically programmed embedded systems in parallel, a digital input/output is neeeded in addition to the analog input shown in Figure 20.1. According to this concept, the digital signal can be passed to the input of the processor either from the ADC output or from the digital input. In other applications, the digital signal formed by the ADC in an embedded system could be passed to other embedded systems used to fulfil additional signal processing tasks in parallel. A digital switch is included in the structures of these embedded systems in order to carry out these functions.

Encoding and Processing Several analog Signals in Parallel

So far encoding of only a single analog signal source has been considered. In many real-life situations there are a number of such sources. If the quantity of them is not too large, then the discussed universal embedded system is still applicable. The required number of these systems simply has to be used then or, in some other cases, the signal sources could be connected to a single embedded system of the considered type through a multiplexer.

Using Embedded Systems for Alias-Free Processing of a Large Quantity of analog Signals

Massive data acquisition and processing is not covered by the universal embedded system described above. A special approach to embedding systems performing the nonuniform procedures of signal encoding and preprocessing clearly has to be used in cases where data are acquired from a large number of signal sources. Although specific techniques have to be used in those cases, it still makes sense to use embedded systems as confined areas where the specifics of the digital

alias-free signal processing are enclosed. The system illustrated by Figure 11.7 could be considered as an example of the embedded systems used to handle a large cluster of remotely sampled signals.

Achieving Complexity Reduction of Sensor Arrays

Using pseudo-randomized spatial and temporal signal encoding and processing is essential for achieving high performance at the array signal processing at a reduced quantity of sensors in the array. While it is of course essential to use embedded systems for encapsulation of specific nonuniform procedures for spatial and temporal spectrum analysis in this case, the fact has to be taken into account that a large amount of data is acquired from the sensors in the array and has to be processed as quickly as possible. Although, in principle, universal embedded systems built according to the above model could be used for pseudo-randomized spatial and temporal signal encoding and processing, better results could be expected if special types of embedded systems matched to the specifics of handling the array signals are developed and used.

Using Embedded Systems in Special Cases of Processing Distorted Periodic Sample Value Sequences

As shown in Sections 2.3 and 2.5, the algorithms and techniques developed specially for pseudo-randomized digitizing and processing of signals digitized in this special way could also be successfully used for reconstruction of impaired periodically sampled signals, achieving considerable fault tolerance, and for data compression executed in a simple manner. At first glance it seems that again specific embedded systems for digital alias-free signal processing need to be developed for applications in this area. However, more careful consideration of this matter has led to the conclusion that if only some minor changes are made in the structure of the universal embedded system then it could be used for these applications as well. Actually only a single functional block, specifically an analyser of sampling irregularities extracting the nonuniform sampling data, needs to be added to the structure shown in Figure 20.1. The task of this functional block is to analyse the digital input signals in these cases and to supply the processor with the information characterizing the specific sampling irregularities. This information can then be used to adapt signal processing to these specific sampling irregularities, as explained in the following sections. If this approach is used, no special embedded systems have to be developed and used for reconstruction of the compressed data and for reconstruction of data corrupted by interference or some other kind of functional fault.

20.2 Algorithm Options in the Development of Firmware

The essential functions of the considered embedded systems are signal alias-free digitizing, their representation in a digital format and waveform reconstruction. They are supported by hardware and firmware. Therefore the effectiveness of their application to a large extent depends on the methods and algorithms used as the basis for development of the embedded system firmware.

While there is little to be added to what has been said about signal pseudo-randomized digitizing in previous chapters, less clear is the subject of quality achievable at signal representation in the frequency and time domains and at transforms of one particular type of signal representation into another one. The problem is that the classical approach to periodically sampled signal spectrum analysis and waveform reconstruction, based on direct and inverse Fourier transforms, is not fully applicable in the cases where the digital signals are given as sequences of sample values taken at pseudo-random time instants. The results of the DFT in those cases are distorted by errors due to sampling irregularities and therefore they do not represent the signal in the frequency domain sufficiently well. These errors have to be somehow filtered out, which is a difficult task. Discussions of the problems related to nonuniformly sampled signal representation in the frequency and time domains and a comparison of various approaches used to resolve them follow.

20.2.1 Sequential Exclusion of Signal Components

The basic problem with the spectrum analysis of nonuniformly sampled signals is the fact that errors in estimating parameters of a signal component in the frequency domain depend on the power of all other signal components. This dependence, of course, was noticed a long time ago but it took some time to develop a special algorithm for nonuniformly sampled signal processing that takes this fact into account. A successful approach to this problem was finally found. It is based on the use of the so-called sequential component extraction method, or SECOEX. Historically, development of the SECOEX method could be considered as a serious achievement marking progress gradually made in this area. The spectrum analysis of nonuniformly sampled signals performed on the basis of this algorithm demonstrated the fact that it is possible to improve significantly the results obtained at this stage of the DFT.

This method is based on estimating the most powerful of the signal components, taking it out of the original signal, then estimating the next most powerful component, subtracting it from the previously calculated difference and repeating

these cyclic operations until the power of the reminder becomes smaller than a given threshold. More specifically, cyclic computations are carried out to optimize the following equation:

$$\sum_{k=1}^{N} (a_i \cos 2\pi f_i t_k + b_i \sin 2\pi f_i t_k - x_{ik})^2 = \min, \tag{20.1}$$

where the sample value set corresponding to the ith cycle is denoted by $\{x_{ik}\}$, $i = 1, 2, 3, \ldots, k = 1, \ldots, N$. To obtain $\{x_{1k}\}$, the most powerful signal component $a_1 = A_1 \sin \varphi_1; b_1 = A_1 \cos \varphi_1$ (A_i, φ_i and f_i are the amplitude, phase and frequency of the ith component respectively) has to be estimated and then subtracted from the digital input signal $x(t_k) = x_k, k = 1, \ldots, N$. In general,

$$x_{(i+1)k} = x_{ik} - A_i \sin(2\pi f_i t_k + \varphi_i). \tag{20.2}$$

These operations are performed by estimating and varying all three parameters A_i, φ_i and f_i in accordance with Equation (20.1).

The algorithm developed on the basis of the SECOEX method suppresses relatively well the aliases present in the signal and the errors related to the cross-interference between the signal components. For example, the spectrogram shown in Figure 20.2(b) has been obtained using it. Compare this with the spectrogram displayed in Figure 20.2(a), which represents the results of the DFT performed for the same signal. It can be seen that application of SECOEX has improved the spectrogram significantly.

However, after a period of time it became clear that this method had some noticeable drawbacks and that it is possible to achieve more accurate results. Firstly, this method does not take into account the specific irregularities of the sampling instants at which the sample values of the respective signal are taken. Therefore it cannot be used as the basis for development of the processing procedures adaptable to the nonuniformities of the sampling process. Secondly, errors due to the cross-interference effect are not taken out completely. While taking out the larger components of the signal one by one effectively reduces the impact of the more powerful components on the estimation of the smaller components, the errors in the parameter estimation of the more powerful components are still enlarged as a result of the additive impact of the smaller components due to cross-interference. Thirdly, execution of the calculations arranged according to SECOEX is computationally burdensome. According to this approach, because the signal components are taken out sequentially, the Fourier coefficient estimates are calculated repeatedly in the whole frequency range of the respective signal for a number of times equal to the number of the most significant signal components.

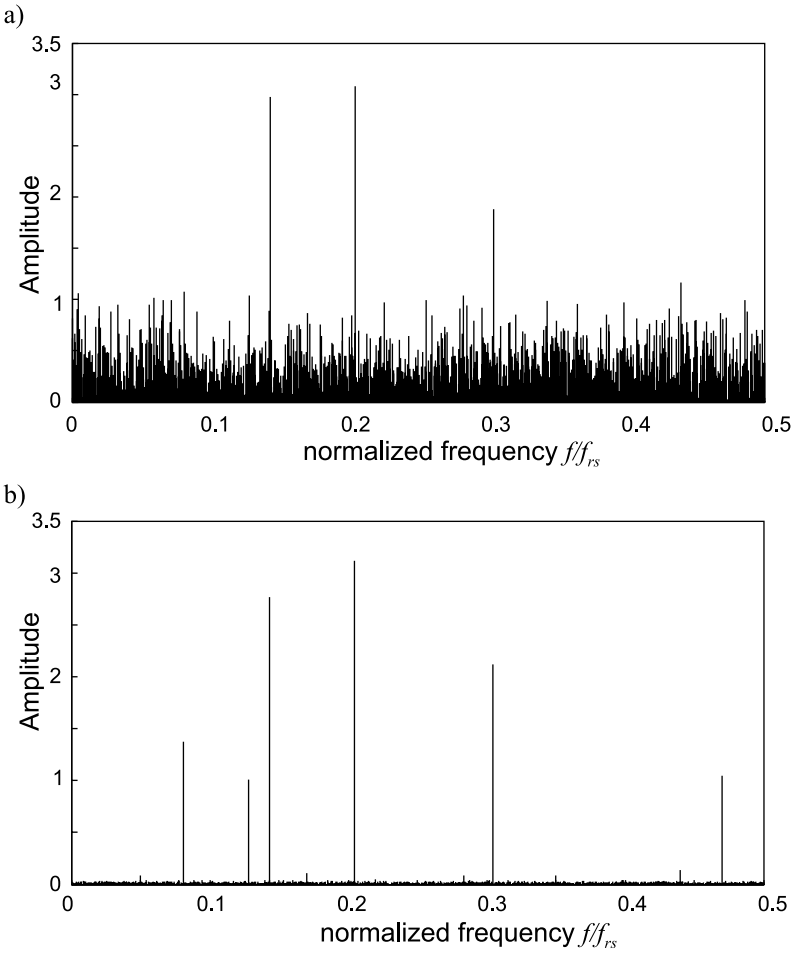


Figure 20.2 Spectrograms characterizing a nonuniformly sampled multitone signal obtained (a) as a result of the DFT and (b) by calculations made on the basis of the SECOEX method

20.2.2 Iterative Variable Threshold Calculations of DFT and IDFT

Another approach to improving the results of DFT could be used. Although it is also an iterative one, fewer iteration cycles are needed for realization of it. A group of signal components are estimated at each cycle rather than a single component, as in the case of SECOEX. Such an approach naturally leads to a reduction in the computational burden and to a saving of time.

A particular realization of this iterative method, based on using the FFT, could be used whenever the original is sampled pseudo-randomly with sampling instants located nonuniformly but on a well-defined time grid. Then the so-called zero-padding method could be used at the first cycle to transform the nonuniform sampling event stream into a periodic one, as briefly described in Section 15.3. Once that is done, the signal could be formally processed on the basis of the FFT.

The idea of using the FFT to obtain estimates of the Fourier coefficients is of course attractive as application of this fast algorithm drastically reduces the amount of calculations. However, using this approach evidently leads to large errors in estimating the Fourier coefficients. The achievable accuracy of these estimates is typically not acceptable. Therefore this procedure, if and when used, should be regarded as preliminary calculations providing only raw intermediate results. The degree of their usefulness depends on the used specific algorithm for signal analysis.

To suppress the errors due to sampling irregularities down to an acceptable level, a step-by-step approach should be used. The more powerful signal components are estimated at the first cycle. After that the second group of less powerful components are estimated and so on. Such an approach makes sense as the relative errors are smaller for the more powerful components. To realize it, a few threshold levels are introduced. The signal components above the upper threshold are estimated at the first cycle. Then the inverse DFT (IDFT) is carried out and the missing sample values are substituted by the corresponding instantaneous values of the roughly reconstructed waveform. After that the obtained signal sample value sequence is used for the repeated DFT. Significantly more accurate estimates of Fourier coefficients are obtained. At the next step, the threshold level is lowered and the components above it are estimated again. The process is continued in this way for a given number of cycles.

At first the DFT is performed for frequencies located on the frequency grid, with the interval between the frequencies determined by the signal observation time as usual. If all signal components are at frequencies located exactly on this grid, the DFT provides spectral estimates that are sufficiently accurate. However, the frequencies of real signal components are often shifted with regard to this grid. Then the positions of these components on the frequency axis have to be estimated more precisely. Otherwise the IDFT will result in unacceptable waveform reconstruction errors.

The IDFT is performed for all frequencies exceeding the mentioned thresholds. The waveform obtained as a result of the IDFT, performed for a number of the

most powerful signal components, is given as

$$y_{(0)}(i\,\Delta T) = \sum_{m=0}^{L-1} \hat{a}_{(0)}(f_m)\cos(2\pi f_m i\,\Delta T) + \sum_{m=0}^{L-1} \hat{b}_{(0)}(f_m)\sin(2\pi f_m i\,\Delta T), \quad i = \overline{0, M-1}, \qquad (20.3)$$

where $L-1$ is the number of frequencies considered at the particular iteration cycle.

This waveform is periodically sampled with the sampling intervals equal to the step of the time grid. The initially used zeroes are then replaced by the taken sample values of this waveform. A periodic sequence of signal sample values is formed in this way. All actually taken signal sample values are in the right places and the roughly estimated sample values are inserted in the places initially occupied by zeroes. The precision of the signal digital representation can be significantly improved in this way. Therefore much more accurate estimates of the Fourier coefficients are obtained when the DFT is performed for this digital signal for the second time.

The following iteration cycles are repeated in an analogous way. Each time the threshold is put at a lower level and a group of previously not-estimated signal components, characterized by powers above this level, are estimated, this information is used for reducing the errors in signal sample value estimates at the time instants where the zeroes were placed at the beginning of the signal analysis process.

Relatively good results are obtained in this way. More will be said about this in Subsection 20.2.4. The main advantage of this iterative algorithm, in comparison with SECOEX, is the achieved possibility of using a significantly reduced number of iteration cycles. However, it is still not perfect. The basic drawback is related to the fact that the cross-interference strongly impacting the results of the DFT is not taken into account.

20.2.3 Algorithms Adapted to the Sampling Irregularities

There is no doubt that signal processing needs to be adapted to the nonuniformities of the involved sampling process whenever it is possible. As shown in Figure 20.1, such adapting is normally performed. The algorithms for this actually combine the adapting procedures with the iterative approach to cyclically repeated execution of the DFT and IDFT. In general, algorithms for signal processing adapted to the

sampling irregularities are similar to the iterative algorithm discussed above. The basic difference is in the estimation of the Fourier coefficients for a group of signal components. While standard DFTs are used for this in the case of the described iterative algorithm, cross-interference coefficients are calculated and the effect of the cross-interference is taken into account in the case of the algorithm adapted to sampling irregularities. That changes the situation significantly. The precision of Fourier coefficient estimation is substantially increased at each iteration cycle, which leads to faster convergence to the final spectral estimates. Figure 20.3 illustrates this kind of adapted iterative signal processing.

It can be seen from Figure 20.3 that the estimation process develops quickly in this case. A more detailed comparison of this type of algorithm with both previously considered ones follows. Note that the illustrated spectrum analysis is actually only a part of the whole signal processing process. Each adaptation cycle contains calculations of direct and inverse DFTs. Therefore the signal waveform is also repeatedly estimated with growing precision.

For calculations carried out during the process of this adaptation, at each iteration cycle they typically cover a relatively small number of signal components at arbitrary frequencies. Under these conditions, the cross-interference coefficients, at separate adaptation cycles, are usually calculated for the particular group of peaks in the signal spectrum that is processed for this cycle. This means that it is then not necessary to calculate and use the matrix of the cross-interference coefficients characterizing the respective nonuniform sampling point process used to digitize the signal as described in Chapter 18. Direct on-line calculations of the cross-interference coefficients are then much more productive and the embedded systems shown in Figure 20.1 are based on this concept.

In cases where it is essential to achieve high operational speed, the discussed adaptation process could be realized in accordance with the scheme given in Figure 18.6. The used nonuniform sampling point process then has to be decomposed into a number of periodic processes with pseudo-randomly skipped sampling points. The signal sample values obtained at time instants defined by each particular sampling point substream can then be processed separately. In this way the sequential performance of a large number of required computational operations could be replaced by parallel calculations of reduced complexity carried out in parallel.

20.2.4 Comparison of Algorithm Performance

While all three types of the algorithms discussed above are applicable for adapting signal waveform reconstruction to the irregularities of the sampling process,

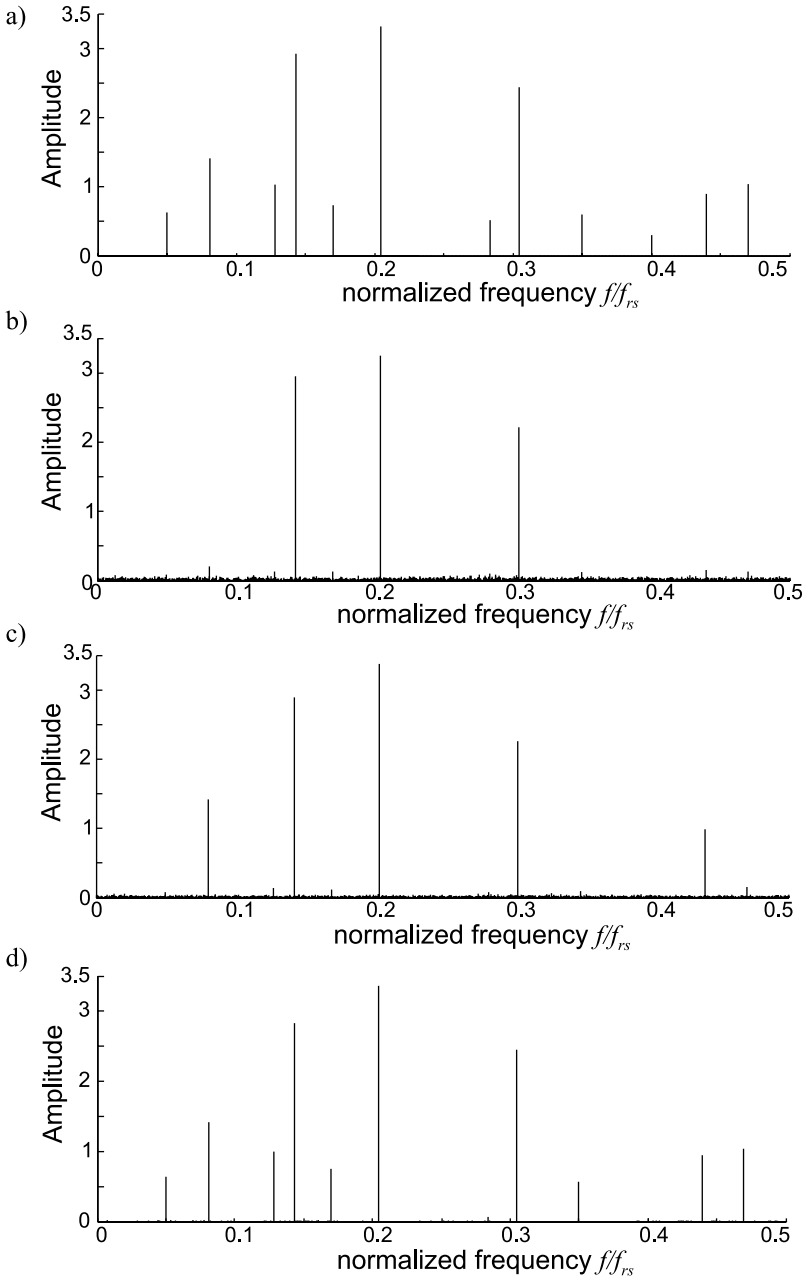


Figure 20.3 Results of spectrum analysis performed according to the iterative spectrum analysis adapted to the nonuniformities of the sampling process used when digitizing the signal: (a) true spectrogram; (b) spectrogram obtained at the end of the first cycle; (c) spectrogram after the second iteration; (d) spectrogram after the third iteration

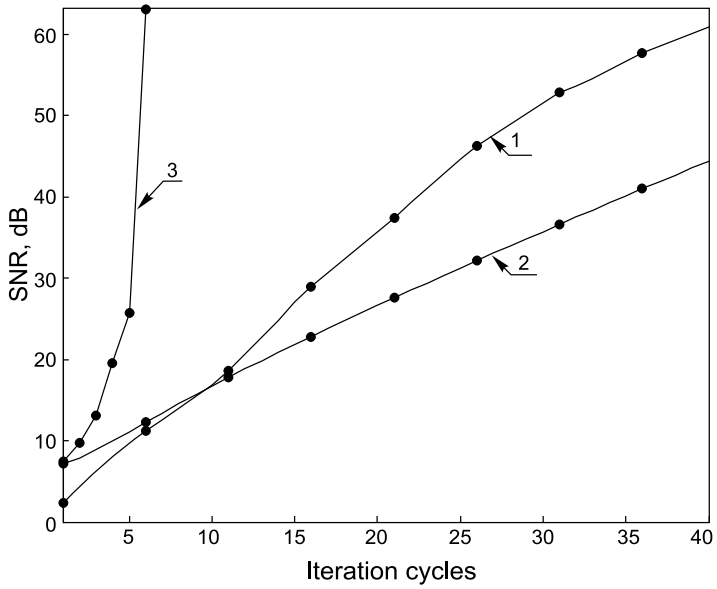


Figure 20.4 SNR versus the iteration cycle numbers characterizing application of the three compared algorithms for signal waveform reconstruction

their performance characteristics are different. To gain a first impression of their comparative qualities, a multitone test signal was digitized and its waveform was reconstructed in the presence of noise by applying all three algorithms. Diagrams of the SNR versus the iteration cycle numbers characterizing application of the three compared algorithms for signal waveform reconstruction are shown Figure 20.4. It can be seen from them that the waveform reconstruction errors, in the case where the reconstruction is performed on the basis of the SECOEX algorithm, decrease from one iterative cycle to the next relatively slowly and the iterative process is long. The second algorithm, the iterative algorithm, provides better results. However, the best results are obtained in the case of the third algorithm based on iteratively adapting the waveform reconstruction process to the sampling irregularities. When a signal waveform is reconstructed in accordance with this, suppression of the reconstruction error to their minimal value is achieved in a reduced number of iteration cycles.

20.3 Dedicated Services of the Embedded DASP Systems

According to the definition, embedded systems provide dedicated services to embedding systems. The services that should be offered by the embedded DASP

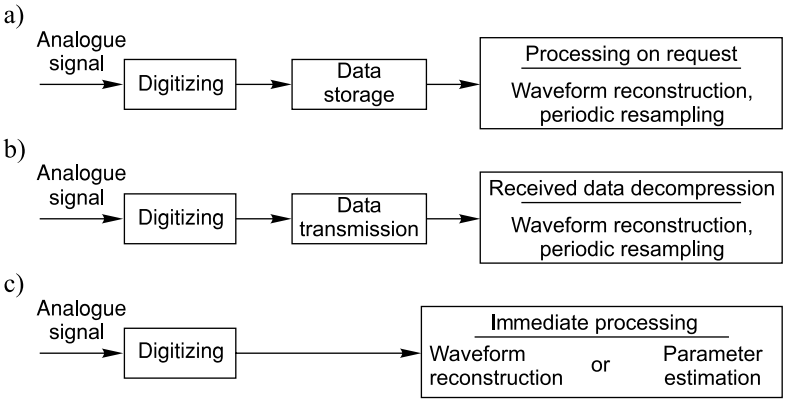


Figure 20.5 Typical dedicated services of the embedded DASP systems provided for digital processing of wideband analog signals

systems are unusual, not paralleled by other types of embedded systems. Basically they need to widen significantly the frequency band within which the analog input signals can be fully processed digitally. More specifically, the dedicated services provided by them are based on the following functions:

- digitizing analog input signals in a way making it possible to avoid aliasing-induced errors at the following digital processing of these signals;
- performing signal digital preprocessing matched to the nonuniformities of the sampling process;
- presenting the results of digital preliminary processing of the signals into a form acceptable for the classic DSP hardware and software tools, which are used for further additional signal processing.

Some application hints are given in Figure 20.5 for the cases where the signals to be processed by the discussed embedded systems are analog. Firstly, these signals could be digitized, the digital versions of them stored in a memory and then processed when requested (Figure 20.5(a)). Note that the digital signal is typically encoded in a compressed form. Secondly, the digitized signals are transmitted over wire or wireless channels, received, demodulated and processed (Figure 20.5(b)). At least two embedded systems are obviously needed in that case. Thirdly, processing of the digitized signal is carried out without delay, preferably in the real-time processing mode, as illustrated by Figure 20.5(c).

Consider the suggested structure of this type of embedded system given in Figure 20.1. It is assumed that signal sample values are taken at time instants dictated by the generated pseudo-random sampling point process. This means that

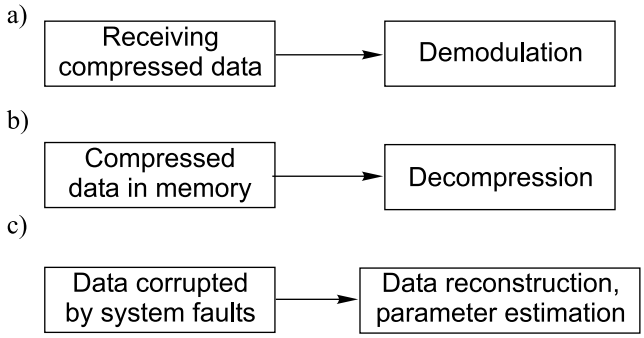


Figure 20.6 Typical dedicated services of the embedded DASP systems provided for processing of randomly decimated periodic digital signals

the signal sample value sequences processed by the computer are randomly decimated periodic sequences of numbers or, in other words, are periodic sequences of numbers with random skips. More often than not, the quantity of the missing numbers in these sequences is much larger than the quantity of the present numerical signal sample values.

Now suppose that the computer processes the data given in the above form, reconstructs the signal waveform and presents it as a periodic sample value sequence with the period equal to the period characterizing the input data sequence. No signal sample values will then be missing. Compare the data flows at the computer input and output. Evidently the computer has decompressed the data given in a specifically compressed form. The performance of this function is representative of the kind of embedded system already discussed, so the role of the computer in the considered type of embedded system can often be defined in this way.

As the reconstruction of compressed data is a function of considerable practical value, it makes sense to use the embedded DASP systems, in addition to exploitation of them for dealing with analog signals, to process specifically presented digital signals. This is the reason why the digital inputs and some additional functional blocks are included in the structural scheme shown in Figure 20.1. The embedded systems with these digital inputs added are capable of fulfilling a number of useful digital signal processing functions. Some embedded system services of this type are indicated in Figure 20.6.

Firstly, the pseudo-randomly digitized analog signals are demodulated and processed (Figure 20.6(a)). The multichannel bioimpedance signal demodulation described in Section 19.6 represents an example of this type of embedded

system application. Secondly, the digital data compressed in a special way could be decompressed by reconstructing and resampling the original analog signal (Figure 20.6(b)). The third type of service that could be given by the discussed embedded systems is providing for fault tolerance (Figure 20.6(c)). This particular embedded DASP system application is considered in some detail in the following section.

Apparently the considered embedded systems could be used for many different applications in a very wide frequency range. While analog signal processing on this basis is more interesting for frequencies ranging from hundreds of MHz up to several GHz, the data compression function is in demand for applications covering a much wider frequency range starting from low frequencies and extending up to the microwave frequencies. Discussion of some typical applications of this kind follows.

20.4 Dedicated Services Related to Processing of Digital Inputs

In general, using embedded systems to handle specifically represented digital signals is based on reconstruction of data compressed purposefully or damaged as a result of system faults. The approach to this task is specific and differs from various techniques usually exploited for compression of data.

20.4.1 Approach to Data Compression

Data can be compressed if they contain some redundancy. This means that the achievable data compression rate is limited. Various data compression techniques, capable of compressing data to the limit, are known and are used. In general, application of the pseudo-randomized data compression approach does not lead to more intensive data compression than the data compression rate achievable by traditional methods. The advantage of the suggested approach is different and is illustrated in Figure 20.7.

Figure 20.7(a) illustrates the typical approach to data compression. According to it, data obtained as a result of signal digitizing are processed to achieve their compression. The compressed data are then either stored or transmitted over some communication lines and after that the compressed data are processed again to restore and convert them into the initial form. In this case, according to this scheme, data are processed twice. The first time resources are used to compress data and the second time to decompress them. Both of these operations typically are computationally burdensome and prosecution of them takes time.

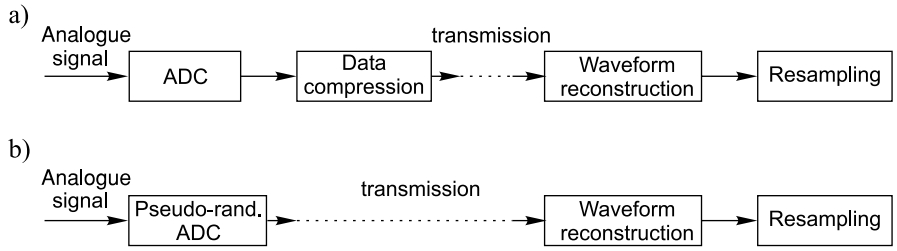


Figure 20.7 Typical difference between the (a) traditional and (b) suggested data compression/decompression schemes

The advantage of the considered pseudo-randomized data compression approach over the conventionally used data compression techniques is illustrated in Figure 20.7(b). The functional block for data compression is not included in the second scheme. It is not needed as data in the second case are actually presented in a compressed form at the output of the digitizer. Nothing special has to be done to compress the acquired data. Therefore it is evident that designs of the data acquisition subsystems exploiting this approach to data compression could be much simpler with reduced power consumption. The full computational burden in this case is placed on the functional block performing data decompression. For that, the waveform of the original signal usually has to be reconstructed. Once that is accomplished, this waveform could be represented by its equidistant sample values taken at a sufficiently high sampling rate to meet the requirement of the sampling theorem.

The embedded systems if they are built according to the scheme given in Figure 20.1 would support this data compression and reconstruction scheme. When they are used for dealing with analog signals, the data representing them are compressed at the stage of signal digitizing. In cases where these data are immediately processed, the computer performs the processing as required. If the given data have to be converted into a digital periodic signal sample value sequence, the computer performs data decompression and the signal waveform is reconstructed for that.

The computer of a particular embedded system apparently can reconstruct signal waveforms either by processing the digital signal taken from the output of the pseudo-randomized ADC included into the structure of this embedded system or by processing some other external digital signal encoded in a similar way. Figure 20.8 illustrates a few versions of the digital signals represented in various forms. The basic digital signal taken off the digitizer output is a sequence of the input signal sample values obtained at nonuniformly spaced time instants

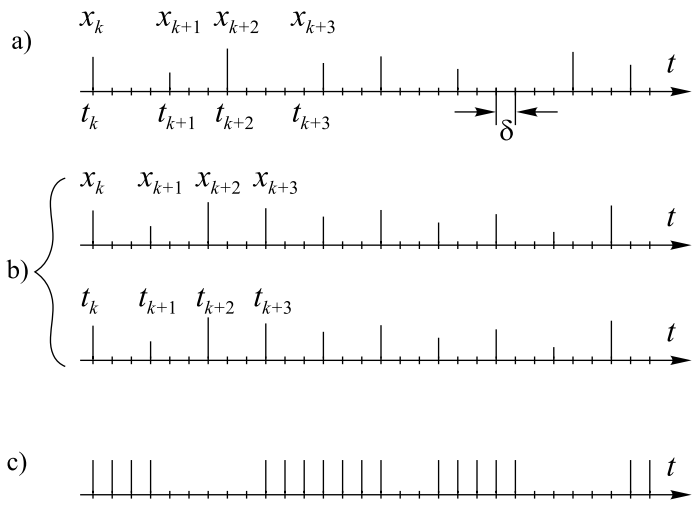


Figure 20.8 Various digital representations of the (a, b) compressed and (c) distorted data

$\{t_k\}$. Both the numerical value and the sampling instant need to be given for each signal sample.

This information can be encoded in various ways. Firstly, this kind of digital signal could be and often is presented in the form shown in Figure 20.8(a). In these cases, the digital time intervals $(t_{k+1} - t_k)$ between the sample values x_{k+1} and x_k are equal to the time intervals between the respective sampling time instants. They are equal to some quantity of the smallest sampling time digits δ. The time diagram of a typical digital output signal of the pseudo-randomized ADC looks like this.

Figure 20.8(b) illustrates the second possible approach to the compressed data representation. It is shown that the signal sample values $\{x_k\}$ and the respective sampling instants $\{t_k\}$ could also be given by two periodic sequences of numerical values. However, it is much better to sample signals at precisely predetermined time instants and to use this sampling instant information for a reconstruction of the respective signal waveform. This approach should be used whenever possible. Then the compressed signal is simply a periodic sequence of the signal sample values and the information of the sampling process nonuniformity is kept in a memory as shown in Figure 20.1.

The third variety of the digital signals that could be processed by the considered embedded systems is given in Figure 20.8(c). It is a periodic signal with randomly skipped values. This type of signal has to be processed when a periodically

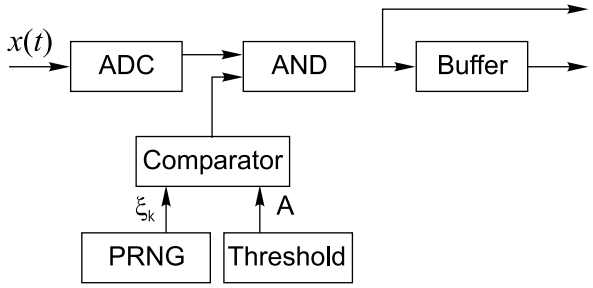


Figure 20.9 Block diagram of a device performing data compression by pseudo-random decimation of the digital signal present at the ADC output

sampled signal is distorted by some faults in functioning of the respective system. A typical task that has to be resolved is the reconstruction of the distorted and excluded signal sample values rather than decompression of some compressed data block. However, the waveform of the original signal apparently also needs to be restored in this case and when this reconstructed waveform is obtained there are no problems in reconstruction of the missing sample values.

20.4.2 Data Compression for One-Dimensional Signals

The basic advantage of the considered data compression/reconstruction approach is the extreme simplicity of the compression. Actually no special data processing functions need to be performed to compress data in this case. Indeed, to compress the data representing, for instance, one-dimensional signals given as periodic sequences of their sample values, the sample value sequences simply have to be pseudo-randomly decimated, with the remaining sample values packed and transmitted or stored as shorter periodic digital signals having a reduced number of discrete values. Forming of the compressed data blocks could be carried out using the scheme given in Figure 20.9.

Data representing the input signal are given as a uniform digital signal at the ADC output. The simple logic circuitry of this scheme dictates which of the signal sample values x_k are passed to the output and which are excluded. The PRNG is used for generation of pseudo-random numbers ξ_k within the range (0, 1). The AND gate is opened when the pseudo-random number ξ_k exceeds the threshold A. The signal sample values taken at these time instants are included in the output signal, all other sample values being omitted. Thus the signal at the output of the AND gate is periodic with pseudo-randomly skipped sample values. The

threshold level A defines the probability of a signal sample value being missing. This signal representing the compressed data could be transmitted in this form with zeroes replacing the excluded sample values or the sample values could be stored in a buffer memory and then transmitted as a shorter periodic sequence.

Evidently this decimation of the signal sample values is performed for data compression in a way predetermined by the design of the PRNG and its synchronization to the signal. If the same type of properly synchronized PRNG is used at reconstruction of the compressed data, it is easy to replace the missing sample values by zeroes. Once that is done, any of the algorithms considered in Section 20.2 could be used for recovery of the original signal or, in other words, for decompression of data. The PRNG, of course, could be replaced by a memory containing a code prescribing the time instances at which the signal sample values should be eliminated.

Figure 20.10 illustrates a signal reconstructed from compressed data. The signal given in Figure 20.10(a) was reconstructed by using the SECOEX algorithm. The next signal shown in Figure 20.10(b) was reconstructed on the basis of the iterative reconstruction algorithm adapted to the pattern of the missing sample values. To obtain data for this demonstration under well-controlled conditions, a periodically sampled (sampling frequency 44.1 KS/s, 16-bit quantizing) signal, specifically recorded music, was used. These data were compressed three times in the described way and then reconstructed. The reconstructed sample values can be compared with the waveform of the original signal.

This example is given to show how data representing a signal could be compressed in a very simple way and then reconstructed. Although these data compression techniques can be used universally, they were actually developed for applications based on processing and transmission of RF and microwave signals.

20.4.3 Data Compression for Two-Dimensional Signals

Various techniques are used for compression of data representing two-dimensional images. Typically they are relatively complicated. Therefore it is tempting to use the data compression approach described above for compressing this kind of two-dimensional data. A particular image of 'Lena', given in Figure 20.11(a), was reconstructed from data compressed four times. The compressed image with 75 % of pixels missing is shown in Figure 20.11(b). The restored image is given in Figure 20.11(c). The reconstruction was again based on the iterative reconstruction algorithm adapted to the pattern of the missing pixels.

The advantage of this approach to image compression is again in the extreme simplicity of the data compression procedure. Reconstruction of the images

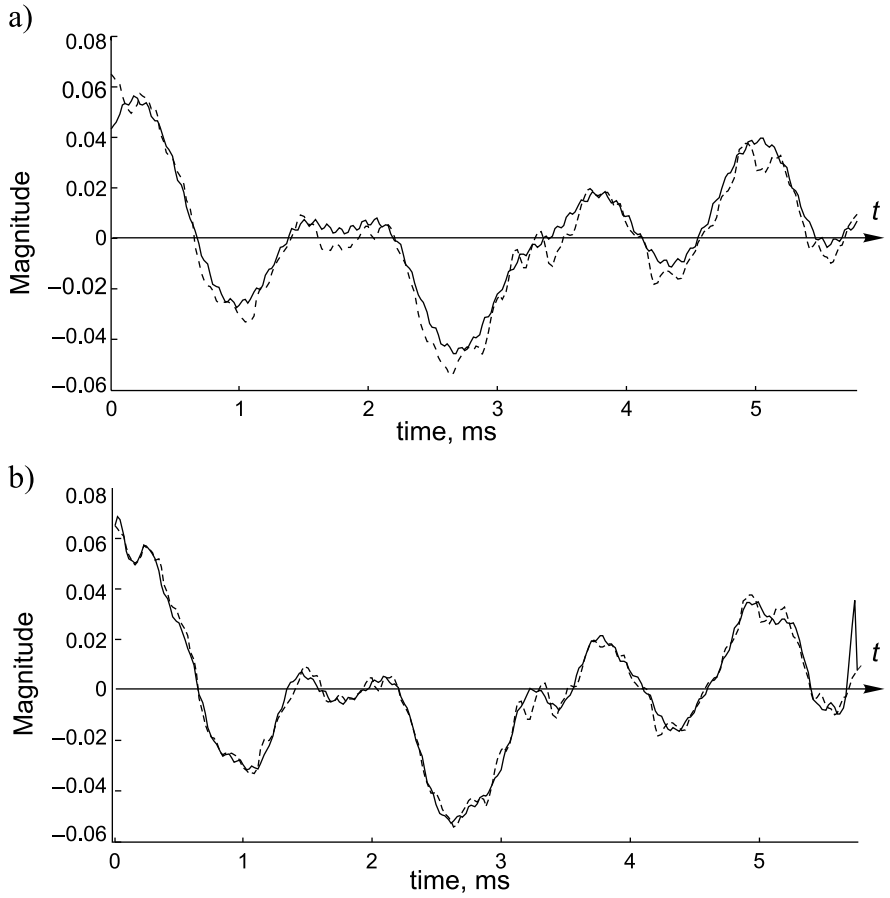


Figure 20.10 Signal waveform reconstructed (solid line) from data compressed three times. The original signal is given by dashed lines

encoded in this way is computationally burdensome. However, the reconstruction quality is relatively good.

20.4.4 Providing for Fault Tolerance

Fault-tolerant sensor systems represent yet another area in which application of the discussed embedded systems could turn out to be quite useful. The point is that the digital alias-free signal processing technology is well suited for performing under conditions when data are presented in a nonuniform way, which is exactly

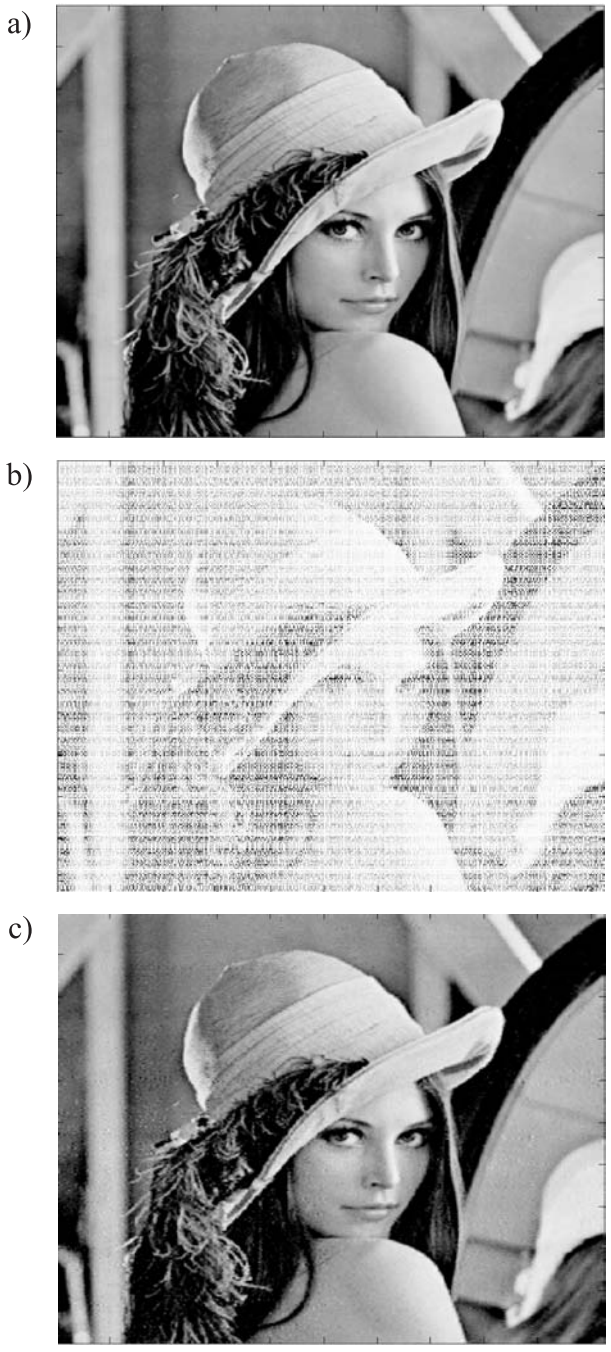


Figure 20.11 Illustration of two-dimensional data compressed four times and reconstruction of the respective image: (a) original image; (b) compressed image with 75 % of the pixels taken out; (c) reconstructed image

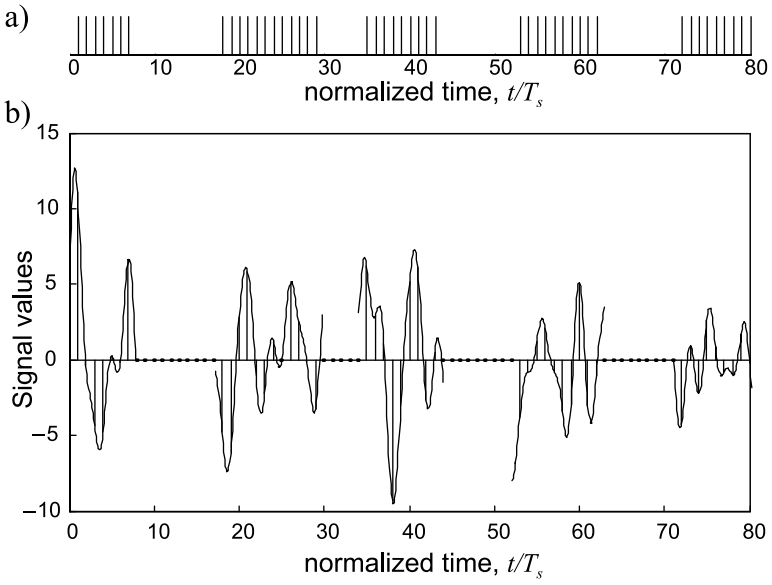


Figure 20.12 Illustration of a signal waveform distorted by faults

what is needed for reconstruction of data damaged as a result of system faults. Therefore application of this technology for the development of sensor signal processing systems tolerant to faulty signal sample values or to sensor failures is well worth considering.

To demonstrate the potential of the algorithms discussed above for reconstruction of signals heavily distorted by fault bursts, a typical case of signal distortion by faults was simulated and the obtained results follow. It was assumed that as a digital sensor signal is corrupted by a powerful interference and many of the signal sample values are completely distorted they had to be taken out of this signal at time instants indicated in Figure 20.12(a) by zeroes. The signal waveform distorted by these fault bursts is shown in Figure 20.12(b).

The iterative signal reconstruction algorithm adapted to the specific fault sequence was used for recovery of the original signal. A zoomed segment of the reconstructed signal sample value sequence is given in Figure 20.13. It can be seen that the quality of the distorted signal reconstruction in this case is good as the sample values of the reconstructed signal overlap the original signal. As expected, the algorithms for reconstruction of signals from their nonuniform sample value sequences adapted to the sampling nonuniformities could be successfully used also for recovery of signals distorted by system faults.

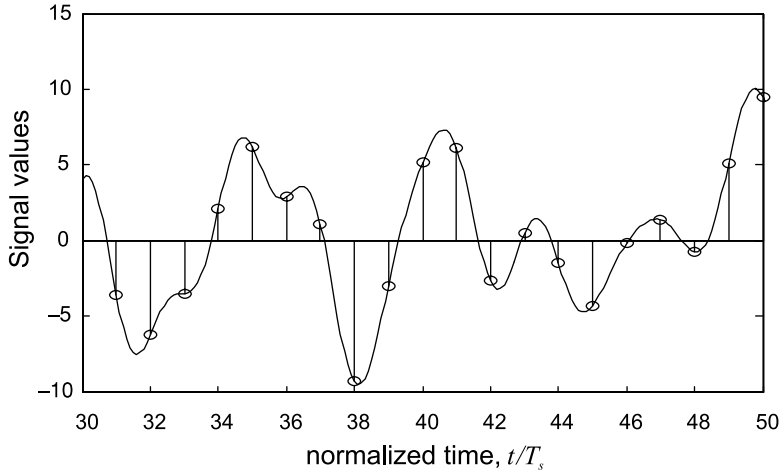


Figure 20.13 Segment (zoomed) of the reconstructed signal sample value sequence overlapping the waveform of the original signal

Note that while the approach to data compression–signal reconstruction described above is applicable for a wide variety of signals, the achievable reconstruction quality to a certain extent depends on spectra of them. To achieve good signal reconstruction adaptability to specific sampling irregularities, it should be possible to decompose the respective signals into their components and the frequencies of these components should be estimated with sufficiently high resolution and precision. That can usually be done but there might be problems.

20.5 Reducing the Quantity of Sensors in Large-aperture Arrays

The subject of high-performance complexity-reduced array signal processing is rather complicated and is actually beyond the scope of this book. Therefore only a few points on this subject are considered here, simply to draw attention to the fact that the discussed digital alias-free signal processing technology is quite competitive and has a high application potential in this area.

As explained in Chapter 17, pseudo-randomization of the distances between sensors in arrays helps in the suppression of the aliasing effect and under certain conditions might lead to a significant reduction in the number of sensors in arrays. However, to succeed, it is crucial to use appropriate nonuniform signal processing techniques for handling array signal processing, both in the time and spatial

domains. While it is relatively easy to achieve suppression of spatial aliasing, there are problems with obtaining sufficiently precise estimates of the spatial spectrum not corrupted by noise related to the fuzzy aliasing effect. Figure 17.9 illustrates this fact. To achieve better results, special anti-aliasing array signal processing techniques and tools have to be used to take into account the cross-interference between components of nonuniformly sampled spatial signals. Although the required systems for processing spatial signals could be built in various ways, the development and use of application-specific embedded systems adaptable to the pattern of sensor positions in the array seems to be the best approach. The feasibility of the sensor quantity reduction achieved by such adaptation is confirmed by the results of computer simulations given below.

20.5.1 Adapting Signal Processing to Pseudo-random Positions of Sensors

As soon as the sensors are placed in an array nonuniformly, the spatial signals taken off this array strongly depend on the pattern of the sensor coordinates in the array. In particular, the basis functions for spatial spectrum analysis under these conditions become nonorthogonal, which leads to cross-interference between spatial signal components. The impact of this cross-interference is observed as background noise corrupting the results of spatial spectrum analysis and beamforming. This noise depends both on the irregularities of sensor positions in the array and on the received or transmitted signal.

The impact of cross-interference due to non-uniformities of sensor positions in the array could be characterized by cross-interference coefficients defined in a way similar to the definition (15.9). Specifically, these coefficients characterizing the cross-interference between the spatial signal components might be introduced as follows:

$$\begin{aligned} \hat{a}(\Omega_m) &= \sum_{n=1}^{M} [a_m(A_m C_n) + b_m(A_m S_n)], \qquad m = \overline{1, M}, \\ \hat{b}(\Omega_m) &= \sum_{n=1}^{M} [a_m(B_m C_n) + b_m(B_m S_n)], \qquad m = \overline{1,\ M}. \end{aligned} \tag{20.4}$$

These coefficients are actually the weights of the errors introduced into the spatial signal in-phase component (or the quadrature component) at frequency Ω_n by nonuniform distancing of sensors. These errors corrupt the estimation of a Fourier coefficient a_m (or b_m) at spatial frequency Ω_m and therefore the coefficients (20.4)

describe the interference between spatial frequency components of the array signal. Their definition is similar to the definition of the cross-interference coefficients (18.6) derived in Chapter 18 for the temporal spectrum analysis. In the case of the spatial spectrum analysis they are given as

$$\begin{aligned}
(A_m C_n) &= \frac{2}{K}\sum_{k=1}^{K}\cos(2\pi\Omega_m d_k)\cos(2\pi\Omega_n d_k),\\
(A_m S_n) &= \frac{2}{K}\sum_{k=1}^{K}\cos(2\pi\Omega_m d_k)\sin(2\pi\Omega_n d_k),\\
(B_m C_n) &= \frac{2}{K}\sum_{k=1}^{K}\sin(2\pi\Omega_m d_k)\cos(2\pi\Omega_n d_k),\\
(B_m S_n) &= \frac{2}{K}\sum_{k=1}^{K}\sin(2\pi\Omega_m d_k)\sin(2\pi\Omega_n d_k).
\end{aligned} \tag{20.5}$$

Another set of cross-interference coefficients, specifically coefficients A_nC_m, B_nC_m, A_nS_m and B_nS_m, characterize interference acting in the inverse direction from the spatial signal component Ω_m to the component Ω_n. It follows from (20.5) that

$$\begin{aligned}
A_nC_m &= A_mC_n, \qquad B_nC_m = A_mS_n,\\
A_nS_m &= B_mC_n, \qquad B_nS_m = B_mS_n.
\end{aligned} \tag{20.6}$$

Thus the cross-interference effect, impacting spatial spectrum analysis in the case where sensors in the array are placed nonuniformly, is reflected by the following matrix of the cross-interference coefficients:

$$\mathbf{Z} = \begin{bmatrix}
(A_1C_1) & (A_1S_1) & (A_1C_2) & (A_1S_2) & \cdots & (A_1C_M) & (A_1S_M)\\
(B_1C_1) & (B_1S_1) & (B_1C_2) & (B_1S_2) & \cdots & (B_1C_M) & (B_1S_M)\\
(A_2C_1) & (A_2S_1) & (A_2C_2) & (A_2S_2) & \cdots & (A_2C_M) & (A_2S_M)\\
(B_2C_1) & (B_2S_1) & (B_2C_2) & (B_2S_2) & \cdots & (B_2C_M) & (B_2S_M)\\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots\\
(A_MC_1) & (A_MS_1) & (A_MC_2) & (A_MS_2) & \cdots & (A_MC_M) & (A_MS_M)\\
(B_MC_1) & (B_MS_1) & (B_MC_2) & (B_MS_2) & \cdots & (B_MC_M) & (B_MS_M)
\end{bmatrix}. \tag{20.7}$$

This matrix of the cross-interference coefficients is an essential characteristic of the arrays of sensors with pseudo-random distances between them. Once the pattern of sensors in the array is known, all coefficients of matrix $\mathbf{Z}$, as well as of the matrix inv($\mathbf{Z}$), can be calculated. This means that when a signal is

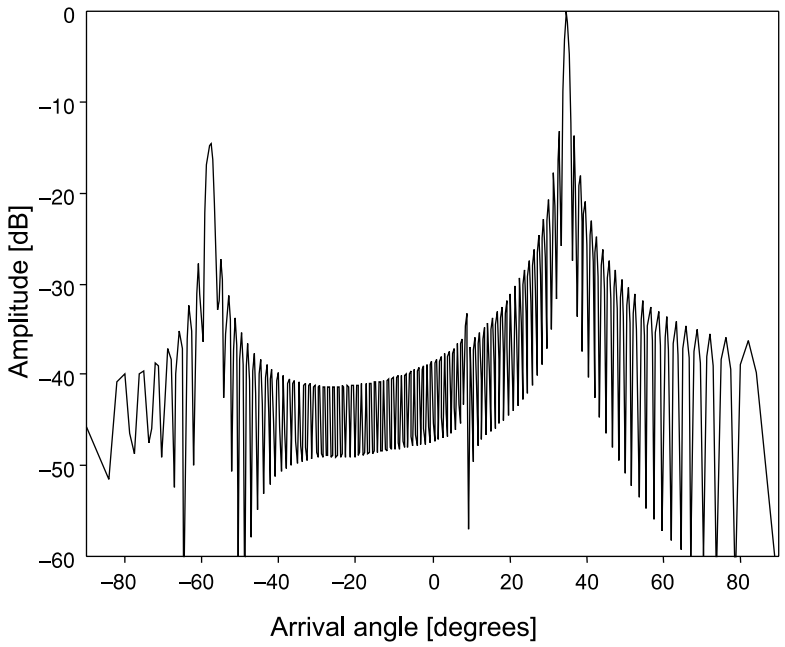


Figure 20.14 Estimated spatial spectrum of the array signal taken off the array of 128 equidistantly spaced sensors

sampled in accordance with such a predetermined nonuniform sampling point process, an equation system similar to (18.19) could be solved, which adapts the estimation of the array signal parameters to the specific positions of the sensors in the array. That significantly reduces the impact of cross-interference between the estimates of the spatial spectrum parameters caused by pseudo-randomization of the array.

This extension of the positive experience accumulated in the area of temporal signal processing for enhancement of array signal processing in the spatial domain seems to be quite successful. It is demonstrated by the following computer simulations. The spatial spectrum analysis of an array signal, containing three components with arrival angles and amplitudes $\Theta = 34.75^\circ$ (0 dB), $\Theta = 8.96^\circ$ (-42 dB) and $\Theta = -57.37^\circ$ (-15 dB), is simulated. Figure 20.14 illustrates the results of such an analysis obtained in the case where the array consists of 128 equidistantly placed sensors. To reduce the number of sensors in the array with the same aperture, the distances between the sensors need to be irregular otherwise there will be spatial aliasing. The results of the spatial spectrum analysis of

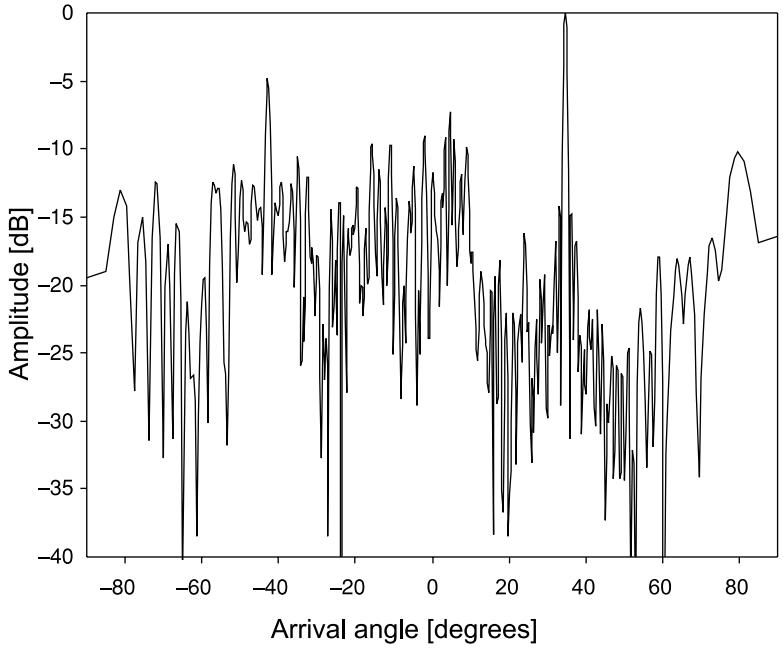


Figure 20.15 Results of the spatial DFT in the case where there are 32 nonuniformly placed sensors in the array

the array signal formed by a nonuniform array containing a four times smaller amount of sensors (only 32) are given in Figure 20.15.

A comparison of the spectrograms given in Figures 20.14 and 20.15 leads to the conclusion that pseudo-randomization of the sensor array design alone does not lead to an acceptable array performance. To achieve better results, the spatial spectrum analysis has to be adapted to the specific irregularities of the sensor positions in the array, as explained above. Then the spectrograms are considerably less corrupted by background noise and are obtained in a significantly wider dynamic range. Figure 20.16 illustrates such adapting of spatial signal processing to the specific nonuniform pattern of the sensor locations in an array.

It can be seen from the obtained and displayed spectrograms in Figure 20.16 that deliberate randomization of sensor array designs does indeed make it possible to reduce the number of sensors in the arrays of the same aperture significantly. However, to benefit from this approach, it is crucial to use effective spatial signal processing techniques well matched to the specifics of this kind of array. The point is that the techniques used for adapting spatial signal processing to the

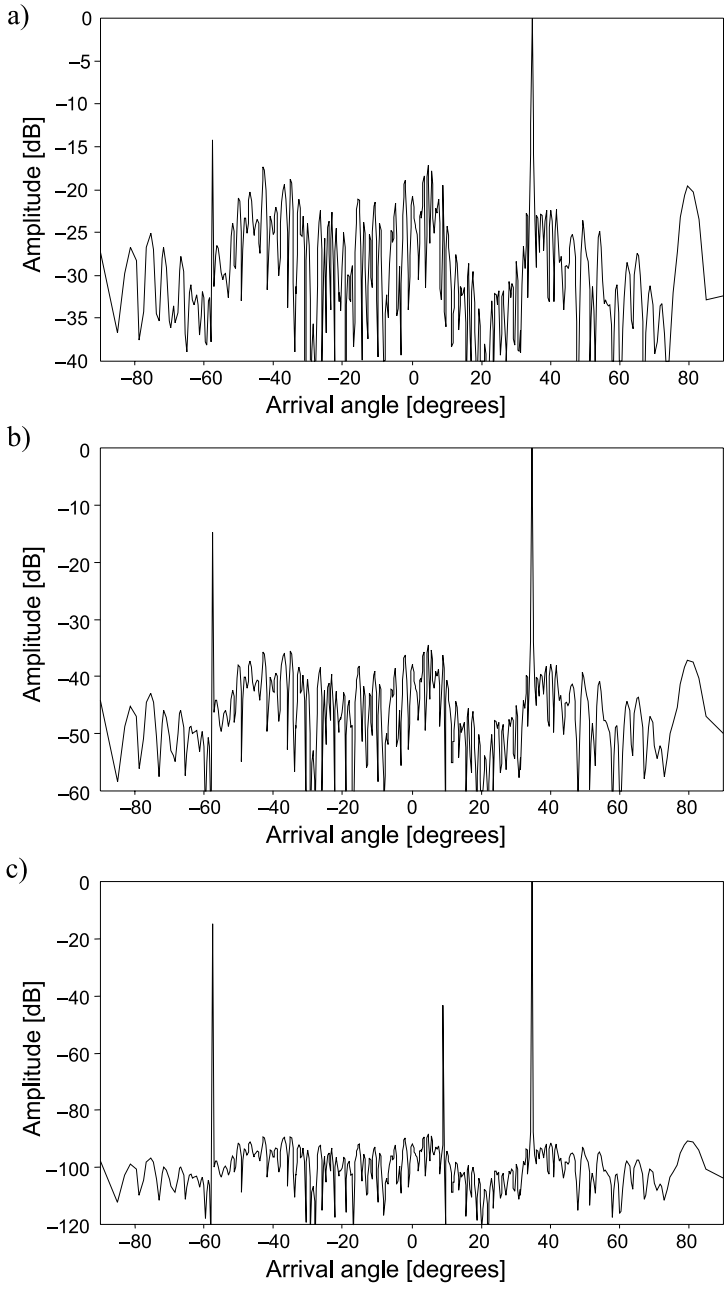


Figure 20.16 Spatial spectrum of a signal taken off an array of 32 pseudo-randomly spaced sensors obtained by adapting the spatial spectrum analysis to the nonuniformity of the sensor positions in the array: (a) after the second iteration; (b) after the fifth iteration; (c) after the tenth iteration

irregularities of the sensor positions in arrays meet these requirements. They are capable of improving the performance of this type of sensor array significantly and therefore application of them makes it possible to reduce the number of sensors in the array.

The achievable complexity reduction of this kind of irregular sensor array depends on the specific conditions of their applications. For instance, the array complexity and array signal processing time often needs to be traded-off. Nevertheless, it should be possible to reduce the number of sensors in arrays substantially in this way. As the research results so far obtained show, it is feasible to reduce in many cases the number of sensors in one-dimensional arrays from 4 to 5 times and from 16 to 25 times in two-dimensional arrays. It is evident that appropriate embedded systems for array signal processing, adapted to the irregularities of the respective sensor arrays, are needed and have to be used for that.

Bibliography

Allay, N. and Tarczynski, A. (2005) Direct digital recovery of modulated signals from their nonuniformly distributed samples. In Colloque International, TELECOM'2005 and 4 èmes *Journées Fnanco-Maghré' bines des Micro-ondes it leurs Applications* (*JFMMA*), INPT, Rabat, Moroc, March, pp. 535–8.

Artjuhs, J. and Bilinskis, I. (2006) Method and apparatus for alias suppressed digitizing of high frequency analog signals. EP 1 330 036 B1, European Patent Specification, Bulletin 2006/26, 28.06.2006.

Artyukh, Yu., Bilinskis, I., Greitans, M. and Vedin, V. (1997) Signal digitizing and recording in the DASP-Lab System. In Proceedings of the 1997 International Workshop on *Sampling Theory and Application*, Aveiro, Portugal, June 1997, pp. 357–60.

Artyukh, Yu., Bilinskis, I. and Vedin, V. (1999) Hardware core of the family of digital RF signal PC-based analyzers. In Proceedings of the 1999 International Workshop on *Sampling Theory and Application*, Loen, Norway, 11–14 August 1999, pp. 177–9.

Artyukh, J., Bilinskis, I., Boole, E., Rybakov, A. and Vedin, V. (2005) Wideband RF signal digititising for high purity spectral analysis. In Proceedings of the International Workshop on *Spectral Methods and Multirate Signal Processing* (*SMMSP 2005*), Riga, Latvia, 20–22 June 2005, pp. 123–8.

Bilinskis, I. and Artjuhs, J. (2006) Method and apparatus for alias suppressed digitizing of high frequency analog signals. United States Patent US 7,046,183 B2, 16 May 2006.

Bilinskis, I. and Cain, G. (1996) Digital alias-free signal processing in the GHz frequency range. In Digest of IEE Colloquium on *Advanced Signal Processing for Microwave Applications*, 29 November 1996.

Bilinskis, I. and Cain, G. (1996) Digital alias-free signal processing tolerance to data and sensor faults. In IEE Colloquium on *Intelligent Sensors*, Leicester, 1996.

Bilinskis, I. and Mikelsons, A. (1990) Application of randomized or irregular sampling as an anti-aliasing technique. In *Signal Processing*, V: *Theories and Application*. Amsterdam: Elsevier Science Publishers, pp. 505–8.

Bilinskis, I. and Rybakov, A. (2006) Iterative spectrum analysis of nonuniformly undersampled wideband signals. *Electronics and Electrical Engineering*, **4**(68), 5–8.

Brannon, B. (1996) Wide-dynamic-range A/D converters pave the way for wideband digital-radio receivers. In EDN, 7 November 1996, pp. 187–205.

Brannon, B. Overcoming converter nonlinearities with dither. *Analog Devices*, **AN-410**, Application Note.

Harrington, R.F. (1961) Sidelobe reduction by nonuniform element spacing. *IRE Trans. Antennas Propag.*, **AP-9**(2), 187–92.

Laakso, T.I., Murphy, N.P. and Tarczynski, A. (1996) Reconstruction of nonuniformly sampled signals using polynomial filtering. In Proceedings of the IEEE Nordic Signal Processing Symposium (*NORSIG'96*), Vol. 1, Helsinki, 24–27 September 1996, pp. 61–4.

Laakso, T.I., Tarczynski, A., Murphy, N.P. and Välimäki, V. (2000) Polynomial filtering approach to reconstruction and noise reduction of nonuniformly sampled signals. *Signal Processing*, **80**(4), April, 567–75.

Lo, Y.T. (1964) Mathematical theory of antenna arrays with randomly spaced elements. *IEEE Trans. Antennas Propag.*, **AP-12**, 257–68.

Lo, Y.T. and Simcoe, R.J. (1967) An experiment on antenna arrays with randomly spaced elements. *IEEE Trans. Antennas Propag.*, **AP-15**, 231–5.

Marvasti, F. (2001) *Nonuniform Sampling, Theory and Practice*. New York: Kluwer Academic/Plenum Publishers.

Mednieks, I. and Mikelsons, A. (1990) Estimation of true components of wide-band quasi period signals. In *Signal Processing*, V: *Theories and Application*. Amsterdam: Elsevier Science Publishers, pp. 233–6.

Mikelsons, A. (1994) Alias-free spectral estimation of signals with components of arbitrary frequencies. In *Adaptive Methods and Emergent Techniques for Signal Processing and Communications*, Ljubljana, Slovenia, April 1994, pp. 105–8.

Mikelsons, A. and Greitans, M. (1996) Fault detection based on spectral estimation of nonuniformly sampled signals. In IEE Colloquium on Intelligent Sensors.

Tarczynski, A. (1997) Sensitivity of signal reconstruction. *IEEE Signal Processing Letters*, **4**(7), July, pp. 192–4.

Tarczynski, A. (2002) Signal reconstruction from finite sets of arbitrarily distributed samples. In Proceedings of the 8th Biennial Baltic Electronics Conference (*BEC2002*), Tallinn, Estonia, 6–9 October 2002, pp. 233–6.

Tarczynski, A., Allay, N. and Qu, D. (2003) Reconstruction of nonuniformly sampled periodic signals. In 2003 International Workshop on *Sampling Theory and Applications* (*SAMPTA'03*), Strobl, Austria, 26–29 May 2003, pp. 94–9.

Tarczynski, A. and Cain, G.D. (1997) Reliability of signal reconstruction from finite sets of samples. In 1997 Workshop on *Sampling Theory and Applications* (*SAMPTA'97*), Vol. 1, Aveiro, 16–19 June 1997, pp. 181–6.

Tarczynski, A. and Qu, D. (2006) Reliability of signal reconstruction from arbitrarily distributed noisy samples. *WSEAS Trans. on Signal Processing*, **2**(7), July, 925–32.

Tarczynski, A. and Qu, D. (2006) Quality assessment of reconstructing signals from arbitrarily distributed samples. In Proceedings of 10th International Conference on *Systems*, Vouliagmeni, Athens, Greece, 10–12 July 2006, pp. 7–13.

Index

Digital Alias-free Signal Processing I. Bilinskis