DIGITAL CLOCKS FOR SYNCHRONIZATION AND COMMUNICATIONS

MASAMI KIHARA
SADAYASU ONO
PEKKA ESKELINEN

Digital Clocks for Synchronization and Communications

For a listing of recent titles in the *Artech House Telecommunications Library,*
turn to the back of this book.

Digital Clocks for Synchronization and Communications

Masami Kihara
Sadayasu Ono
Pekka Eskelinen

Artech House
Boston • London
www.artechhouse.com

Library of Congress Cataloging-in-Publication Data
Kihara, Masami.
Digital clocks for synchronization and communications/Masami Kihara, Sadayasu Ono, Pekka Eskelinen.
p. cm.—(Artech House telecommunications library)
Includes bibliographical references and index.
ISBN 1-58053-506-2 (alk. paper)
1. Timing circuits. 2. Synchronization. 3. Telecommunication systems–Equipment and supplies. I. Ono, Sadayasu. II. Eskelinen, Pekka. III. Title. IV. Series.
TK7868.T5 K54 2002
621.382—dc21 2002032670

British Library Cataloguing in Publication Data
Kihara, Masami.
Digital clocks for synchronization and communications. — (Artech House telecommunications library)
1. Digital electronic clocks 2. Synchronization 3. Phase-locked loops 4. Synchronous data transmission systems
I. Title II. Ono, Sadayasu III. Eskelinen, Pekka
621.3'8215
ISBN 1-58053-506-2

Cover design by David Ford

Original Japanese edition published as Wakariyasui Digital Clock Gijutsu (Digital Clock Technologies Made Easy) by Masami KIHARA and Sadayasu ONO.

Published by Ohmsha, Ltd., 3-1 Kanda Nishikicho, Chiyodaku, Tokyo, Japan

International Standard Book Number: 1-58053-506-2
Library of Congress Catalog Card Number: 2002032670

10 9 8 7 6 5 4 3 2 1

To
Tomoko, Yuuka, and Yuuki
–Masami

Eiko, Satoko, Akiko, and Masanori
–Sadayasu

Tuula and Jussi
–Pekka

Contents

Preface

It was not until the 1970s that digital systems were first used for large-scale telecommunication networking. Performance enhancements since then have been strongly tied to advances in synchronization technology. Modern digital data transmission cannot work without clocks. Their rhythm—or oscillator frequency—controls the operation of devices and components, and thus it must be distributed throughout networks and shared by systems. Current communication arrangements such as different synchronous digital hierarchy (SDH)-based variants have the advantage of inherently sharing the same clock. In practice, high quality clocks are indispensable, as they simplify operational concepts and enhance the service level observed by the customers.

The key parameter of a clock is its frequency. Digital communication is likely to fail if this fundamental attribute is transferred or generated without regard to its purity. When designing a network, the time domain stability and spectral purity of the synchronization signal have to be assured for all network elements and components. A basic telecom clock system involves both synchronization and signal generation tasks. This book examines the use of various phase-locked loop (PLL) techniques for clock synchronization and the direct digital synthesizer (DDS) concept for clock generation. Although there are many ways of constructing and arranging clock assemblies, the fundamental issue is achieving adequate signal characteristics. The goal of this book is to show how this can be achieved.

The PLL is a common functional module in several engineering disciplines, as well as in telecom and electronics applications. Complete PLL modules having all the basic elements are available, but these ready-made devices may not meet the system requirements. It is often necessary to design a task-specific PLL if system performance has to be maximized, but this is seriously hampered by the multiple trade-offs involved. Moreover, the number of parameters that must be considered is high. Because a detailed explanation of the physical mechanisms is provided in this book rather than a calculation of precise values, the reader will be assisted in understanding the basic procedures and characteristics. Also, a coarse loop design method and associated useful estimation strategies are described.

The importance of the DDS concept almost equals that of the classical PLL. It can achieve characteristics that are impossible with the conventional analog approach. The DDS is a promising technology in that it simplifies previously complicated synthesizers. Applications in measurement instruments and cellular communication devices make use of its significant potential, and the enhanced speeds offered by digital circuits will only strengthen this trend. Sophisticated circuit designs are necessary in order to reduce the unnecessary or interfering frequency components generated due to the digital processing. This book provides a detailed review of the relevant mechanisms.

Guide to the Book

Chapters 1 and 2 present an introduction to clocks, clock synchronization, and clock generation. Digital signal processing, the basic digital operation, is described in Chapter 3. In Chapter 4, the characteristics of a digital PLL for frequency synchronization systems are described. Chapter 5 outlines the basic analysis tools that are necessary for designing digital PLLs. The movement toward the digital PLL is explained as thoroughly as possible, although the analysis is based on analog PLLs. Chapter 6 shows that simulations are effective for characteristic analysis. Chapter 7 presents a clock system for a network that fully utilizes PLLs. Chapters 8 and 9 present the design of the digital PLL, and the high-speed PLL LSIs used in networks, respectively. In Chapters 10 and 11, the operation and characteristics of DDS in signal generation are described. Frequency purity is examined in some detail. Chapter 12 describes DDS applications. Finally, several useful clock-related topics such as the measurement of noise, evaluation of frequency stability, and

spectrum analysis are discussed. Chapter 15 includes an overview of actual system characteristics with respect to noise and environmental effects.

This book is composed of five basic groups, as shown in Figure P.1. You may read the chapters in the order that best suits your purpose. To just obtain a basic understanding of PLL concepts, you can advance to Chapters 4, 5, and 6 after reading Chapters 2 and 3. If you simply desire an outline of the characteristics needed to use PLLs appropriately, Chapter 4 may

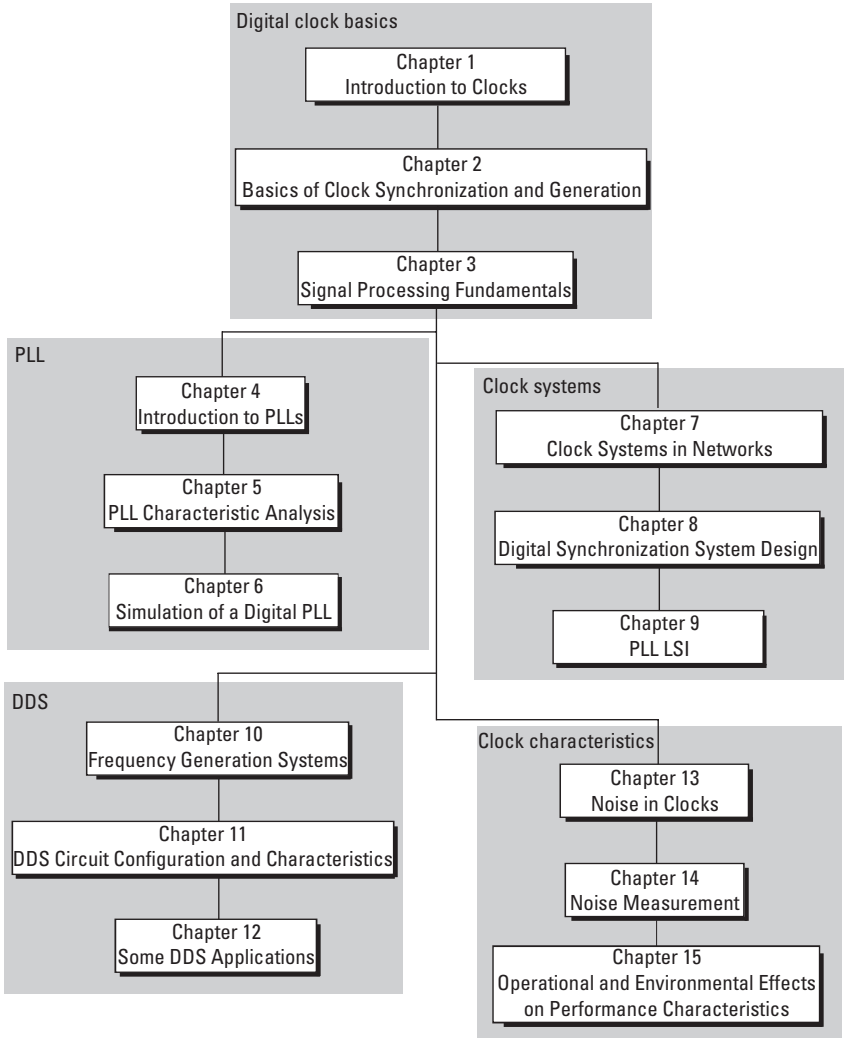


Figure P.1 Chapter configuration of this book.

be sufficient. If synchronization systems in networks are of interest, start at Chapter 7 and advance to Chapters 8 and 9. Of course, the basic PLL information needed can be acquired from Chapters 4 and 5. You can read Chapters 10, 11, and 12 to focus on the DDS issues. When designing a DDS, the basic information about signal processing in Chapter 3 might be useful. Frequency stability issues, signal impurities, and the relationships between noise characteristics and networks are covered in Chapters 13, 14, and 15, respectively. Since these three chapters are common elements of all clock frequency topics, you can refer to them as desired. Additionally, Chapter 15 deals with practical aspects such as some network and system configurations, various clock technologies, and their characteristics related to synchronization.

1

Introduction to Clocks

1.1 Clocks in Digital Systems

Clocks mean different things to different people. The authors of this book consider a clock to be a digital signal that consists of periodic pulses, or an electronic device or circuit used to create such signals. It is not, however, merely a question of a separate signal or an isolated piece of equipment. From a communication systems point of view, a clock carries a vital nutrient characteristic called timing. This clock timing is used to control digital telecommunications, further relying on subsystems that realize various logical operations. In addition, the functions of different digital subsystems can only be coordinated with each other if they share the same clock. Information can be processed reliably in the form of digital data when the associated clock performance and timing transfer are adequate. For us the clock is an indispensable element in digital communication systems [1].

The fundamental electronic clock signal is typically produced by an oscillator, which relies on the performance of a suitable resonator as a stabilizing element [2]. Most portable devices and a number of fixed systems utilize various types of quartz oscillators [3], and their output signal normally has just the same frequency as marked on the crystal device. When better intrinsic stability or smaller absolute frequency errors are needed, we may use atomic resonators [4]. Rubidium and cesium clocks show better aging characteristics than simple crystal oscillators but may be quite expensive. Additional technologies include H-masers and optical frequency standards, but

their use is currently more within the scientific synchronization tasks. A common feature of all these electronic clocks is their firmly fixed, practically nonadjustable output frequency, which in commercial units tends to be a combination of 1, 5, and 10 MHz.

1.2 Synchronization in Systems and Networks

The technical coordination within digital communication systems and subsystems is called synchronization. The main purpose of synchronization is to enable fast and reliable data transfer from one logical component to another and to ensure that individual logical components can be coordinated regardless of their physical distance. In both tasks, data interworking is a key requirement. When digital information is transferred between two components or devices, the receiver can accept data as true information only when both sites are mutually synchronized. That means that the digital clock used by the transmitter must be known, or regenerated in the receiver for successful communication. Digital clock synchronization is indispensable in all components and devices in order to establish a communication process between them at the two distant places.

There are two basic types of data transmission systems according to their level of synchronization: synchronous and asynchronous systems. In synchronous systems, all clocks are completely bound together, while asynchronous systems establish local clock synchronization. Asynchronous clocks are assumed to be independent and no effort is made to force them to synchronism. Of course, here the clocks are synchronized in practice to some extent. Asynchronous systems appear to be somewhat simpler, but the overhead of establishing local synchronization can be significant. Both technical approaches are being utilized in current systems and networks.

Clocks have to be shared in both synchronous and asynchronous systems. The uncertainties in this process are the key design parameters and are defined by the timing specifications. The target of any clock system is to provide those characteristics needed to achieve the required system and network performance, which are often expressed as data speed, jitter, time-to-synchronize, and error rates.

The other point in clock sharing is frequency itself. Not all devices in the transmitter and receiver work at the same frequency; each component operates at the best frequency for its specific function. For example, the processor of a cellular phone may run at 32 MHz but the RF oscillator may need to operate at 1,700 MHz. Alternatively, we have limitations from the

authorities, or the ultimate level of electronic component performance may set practical restrictions. In many cases we must use output frequencies that are too high in the microwave bands to be generated by any known precision oscillator alone. To allow all components of, say, a receiver to operate effectively as one subsystem, it must be possible to generate, from the shared clock, the different frequencies needed for device operation. The general requirements for frequency conversion call for minimal degradation in clock characteristics.

1.3 Clock Frequency Synchronization Technology

Synchronization is necessary to regenerate the original clock at the receiver, which may be physically far from the transmitter or just a few centimeters away inside the same device. Distribution generally reduces timing performance. In long-haul transmission, the clock is exposed to various disturbances such as noise, signal level variation, dispersion, and temperature effects. The receiver has to eliminate or reduce the influence of noise present in the clock signal and to regenerate a signal as close to the original as possible. This is most often done with a phase-locked loop (PLL). While the function initially required can simply be frequency synchronization, phase synchronization is applied most often since it has several benefits, which will be explained in Chapter 2. The original PLL circuit configuration was proposed in the 1930s, and its performance was considerably enhanced from the 1960s to the 1970s. This technology has allowed the digitalization of telecommunication networks; it has been the core method for modern cellular radio transceivers, and it is now vital for supporting the present Internet. Complete building blocks and integrated circuits for different types of PLLs are nowadays commercially available from many semiconductor vendors. A PLL can be easily created by combining a couple of circuits, but it is increasingly becoming a general function, available as a ready-made module.

1.4 Clock Generation Technology

Another important topic—besides data transmission—concerning clocks is the generation of precise yet arbitrary and freely adjustable frequencies. They are needed for timing purposes in other engineering fields like particle physics or radio astronomy. Radio communications are based on pure and very tightly controlled carrier frequencies. This is dictated by international

regulations but is also a technical background for radio navigation. Electronic measuring equipment uses synthesizers. The conventional frequency generation (synthesizing) circuit was originally completely analog. It has evolved into digital circuits to keep up with the rapid advances in system clock speeds and frequency synthesis demands. These include fast frequency setting or frequency hopping and complex modulations. The present end product of this trend is the direct digital synthesizer (DDS). The DDS is a comparatively new technology (it originated in the 1980s). It allows the desired frequency to be generated almost freely using a high-speed clock circuit. There is a natural upper frequency limit and the output frequency depends on the original clock frequency. The frequency resolution, however, is much better than would be possible with conventional synthesizers. One significant benefit is that a complete DDS can be integrated into other LSI devices. The superior characteristics of the DDS are offset to some extent by the noise present in the output signal. A mechanism is needed to suppress it.

References

[1] Smith, D. R., *Digital Transmission Systems*, New York: Van Norstrand Reinhold Company, Inc., 1985.

[2] Ball Efratom, *Precision Time and Frequency Handbook*, Ninth Edition, Irvine, CA 1993.

[3] Neubig, B., and W. Briese, *Das Grosse Quarzkochbuch*, Feldkirchen, Germany: Franzis Verlag, 1997.

[4] Coombs, C., *Electronic Instrument Handbook*, New York: McGraw-Hill, 1995.

2

Basics of Clock Synchronization and Generation

2.1 The General Concept of Synchronization

Let us first have a look at the concept of synchronization at the most general and perhaps familiar level. Although we may have a synchronized system within the smallest physical element (e.g., in a small integrated circuit), it is usual to define this process as happening between a local and another more distant system. Synchronization is defined as forcing a local system to track a remote target, either manually or automatically, in terms of some attribute such as speed or frequency.

Driving within urban traffic provides an example of synchronization, and the physical attribute is vehicle speed. In heavy traffic we are forced to match our speed to that of the car in front of us and should also try to avoid sudden braking so as not to frighten the driver behind us. Coordinating all the members of a symphony orchestra involves synchronization around a common beat. The reference is the conductor, and all musicians attempt to reproduce the beat as set by the conductor's baton. All the instruments, however, have to be tuned to follow the same scale of notes. Otherwise we would hear painful intermodulation products or the orchestra would seem to play out of tune. This situation is an elementary example of combined timing and frequency synchronization. Many low-cost radios still use simple manual tuning; the user captures the desired channel by adjusting the internal local oscillator frequency until the observed signal strength is maximized and the

best sound quality is achieved. This operation is a basic, yet primitive form of synchronization, or more precisely syntonization.

2.2 Clock Synchronization

Synchronizing a local system automatically requires several basic functional elements, as shown in Figure 2.1. First, we have to determine the difference in the attribute between the local system and the remote target. This is accomplished by the detector. The difference information output by the detector is taken as the data used to control the local system. The data is fed back to the local system through a feedback control mechanism, which usually involves some forms of "intelligent" filtering.

The physical quantity treated in clock synchronization is either frequency or, better still, timing, which can be thought of as phase over time (phase-time). Frequency and phase comparators can be used in clock synchronization to detect the difference between the attribute of the remote target and that of the local system. The local system can be a source—an oscillator—that generates a clock for a whole test instrument or for an equipment rack. Signal phase is determined by frequency, and frequency is an average value determined by the phase gradient over time. Accordingly, clock synchronization can be achieved by either frequency synchronization or phase synchronization. These two alternatives are in some way quite similar, but there is one very significant difference: phase synchronization is inherently more accurate—partly due to its predictive nature—which gives a far

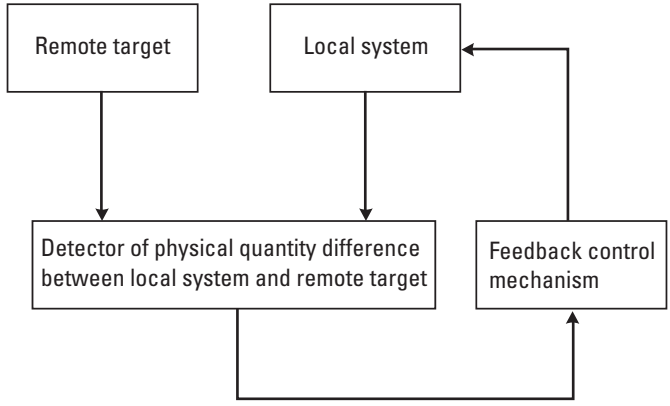


Figure 2.1 Elements essential to achieving synchronization.

better control system sensitivity. Because of this, some scientists consider phase-based processes as the only real synchronization schemes.

2.2.1 Frequency Synchronization

The typical frequency synchronizer operates as follows. First, the frequency difference is detected, and the output frequency of the controlled oscillator is altered appropriately (as shown in Figure 2.2) to reduce or remove this difference. Although the basic control mechanism is equivalent to that of a PLL, only the frequency is directly controlled here. When the frequency synchronization circuit reaches equilibrium, the output frequency of the controlled oscillator equals that of the reference signal. Perfect control will yield a perfect match in frequency. However, random and systematic errors are inherent when we attempt to measure any value. This yields control ambiguity. The fundamental problem is related to the definition of frequency, which basically has to be measured over at least one full cycle time of the input waveform. In practice, much longer periods are required in order to achieve a reasonable control resolution. This implies that the frequency comparator in Figure 2.2 can only produce information about the average situation. Therefore, frequency errors are sure to exist if only the frequency as such is controlled directly. In particular, an error in detecting the frequency difference is fatal. This is a systematic error and will be converted into a false output frequency of the controlled oscillator.

2.2.2 Phase Synchronization

The operation of a phase synchronization circuit (in reality called a PLL) appears to be very similar to the tuning process of a crystal radio, but here the frequency is controlled indirectly. The overall scheme of the synchronized entity is just the same as that seen in the frequency-related case. The controlled oscillator in the PLL (see Figure 2.3) accepts the momentary phase

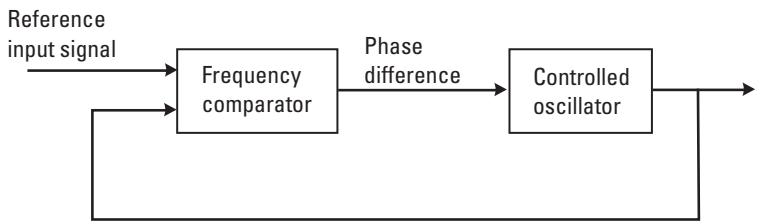


Figure 2.2 Example of a frequency synchronization system.

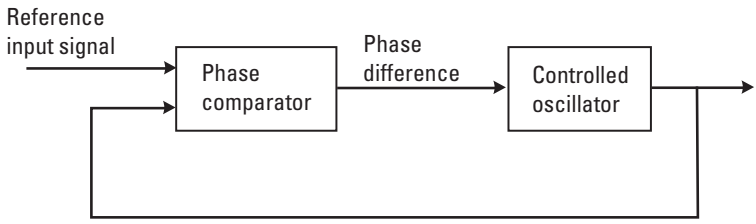


Figure 2.3 Example of a phase synchronization system.

difference, which can be expected to contain phase errors due to control chain imperfections, and outputs a signal whose frequency is determined by the phase gradient over time (see Chapter 4 for a full description). This means that if the phase error can be held fairly constant, the output waveform can be accurately synchronized to the reference frequency regardless of the size of this phase error. From the time domain perspective, the largest benefit of phase detection is its capability to utilize phenomena far shorter than one cycle of the input waveform. It gives the circuitry a much better chance to react to momentary disturbances, which will not simply be averaged out as is done in frequency detection. The process is quite different from the response seen in the simple frequency synchronization scheme illustrated in Figure 2.4. Thus, the PLL is superior for frequency synchronization, which explains its popularity in telecommunication clocks.

2.3 Clock Frequency Generation

It is rare that the frequency or frequencies required in our synchronized telecommunication systems are the same as the single internal reference frequency produced by, say, a crystal oscillator. In general, a frequency different from that of the reference is very often required. If just a single frequency is needed, a PLL can be used as a highly accurate frequency generation circuit. This is achieved quite simply by selecting suitable values for the PLL

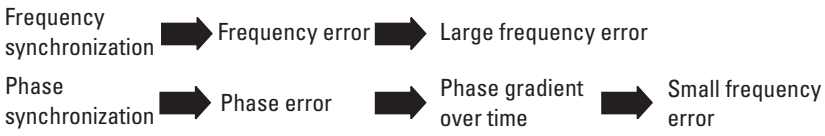


Figure 2.4 Difference between frequency and phase synchronization.

components. If many different frequencies are to be generated simultaneously, a more complicated architecture is needed that allows separate modules to be activated as necessary and which can freely synthesize any frequency. Frequency generation circuits are indispensable elements in wireless cellular phone systems. Their switching speed, frequency uncertainty, and spectral purity are vital design parameters. The situation is further complicated by the recently introduced frequency hopping base stations and handsets, which follow the technology once built for various military communication tasks. Though wired networks do not usually need the ability to change the clock frequency on the fly, synthesizers that can generate selectable frequencies are important as parts of measuring instrumentation in system development.

The analog component shown in Figure 2.5(a) can be combined with some additional modules (mainly a filter and an amplifier) to create a direct analog synthesizer. This synthesizer can process—within certain limitations—digital signals as well, but real logical operations are not performed. Instead, digital signals would be treated here as two square waves. This kind of a construction might be used in a microwave synthesizer. Note that a PLL is not an essential component of the direct analog synthesizers presented in Figure 2.5. They are, however, used in indirect synthesizers. While the direct analog synthesizer processes the input signal to create the output, the indirect synthesizer uses a separate controlled oscillator to generate a frequency different from the input. For these circuits, many of the simple analog modules of Figure 2.5 would still be usable, particularly the mixer and the divider.

2.4 Trends in Phase Synchronization and Clock Generation

The history of phase-locked loops follows quite closely that of industrial, military, and aerospace electronic circuits in general, with digitalization in the forefront. One of the key advances was the advent of the TTL-logic family in the late 1960s. The inputs and outputs of PLLs became digital when they started to be used as modules of digital circuits. At that time, the analog PLL changed to an analog-digital concept, which will be discussed in detail in Chapter 3. Though the analog-digital PLL retains some of the analog processing components, connecting it to other digital circuits is easy, and the device can utilize digital ICs. The key benefits of the digital loop concept are its excellent reproducibility (compared to analog PLLs), small physical size, and lowered manufacturing costs. These aspects have increased considerably the popularity of PLLs. A key issue is the reduction in the number of manual

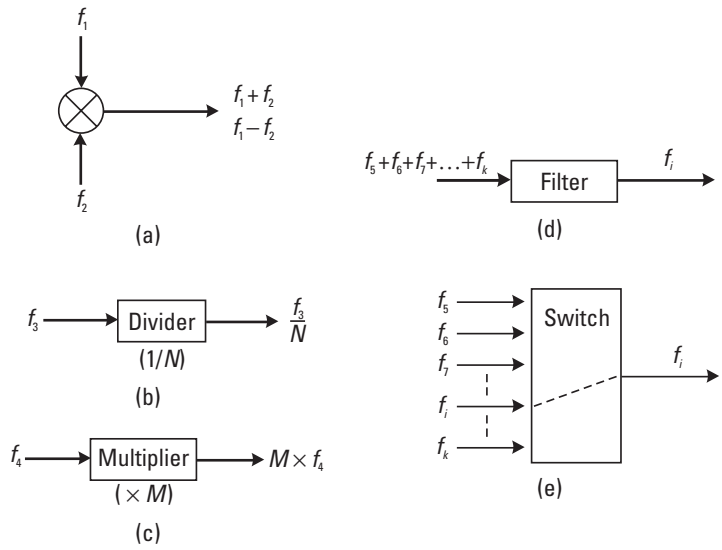


Figure 2.5 Components in direct analog synthesizers: (a) frequency addition and subtraction in a mixer, (b) frequency division, (c) frequency multiplier, (d) frequency selection by filtering, and (e) selection by switching.

adjustment operations due to the uniformity of the digital characteristics. Digital phase comparators and filters are now available. As a result, a PLL can be realized as a single-chip LSI device.

The second trend in current PLL development is the addition of new and sophisticated functions. The approach of using conventional technologies to improve loop characteristics cannot meet all the combined challenges of high switching speed, precise frequency control, and free output waveforms. Accordingly, the demand for clock generation circuits with improved performance figures is driving the adoption of the newer DDS circuit. The DDS eliminates the controlled oscillator used in PLLs and instead generates the output signal directly by digital or numerical processing. Its excellent performance and high level of functionality are driving its rapid penetration into many different fields. Although some DDS designs generate the required frequency by division and synthesis, this book uses this term to refer to devices that use only digital processing circuits.

3

Signal Processing Fundamentals

3.1 Signal Processing and Systems

3.1.1 Signal Processing

A signal provides information about the state of a system. The initial form or nature of a signal can be acoustic, thermal, ionizing radiation, or electronic. Familiar examples include circuit voltages and electric currents. Different system requirements may demand signal processing, which alters or extracts some or all of the characteristics of the signal. Analog signals can be described by their amplitude, frequency, phase, and spectral content. There is a wide variety of processing actions: unnecessary signals can be removed, just one specific signal component can be extracted, and one component can be converted into another form.

The most fundamental description of signal processing is that one obtains the output signal y with the required characteristics from the input signal x using a defined system S (see Figure 3.1). The result is shown in abstract form as a symbol $f(*)$. In many cases, because $f(*)$ is often a function of time t, Figure 3.1 can be rewritten using input signal x and output signal y as shown in Figure 3.2. The variables, however, cannot be assumed to be always bound to time. They often have the meaning of position, and a variable can have two or more simultaneous meanings as well.

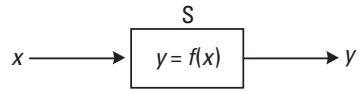


Figure 3.1 System with its input and output.

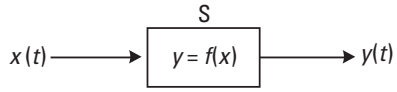


Figure 3.2 System with its input, output, and time.

3.1.2 Linear System

In this chapter, system S in Figure 3.2 is assumed to be linear and it thus satisfies the following conditions:

1. $f(\alpha x) = \alpha f(x)$ where α is a constant;
2. $f(x1+x2) = f(x1) + f(x2)$.

If the input signal in Figure 3.2 is a sine wave, these two conditions mean that only the amplitude value and the phase of the sine wave can be changed, since both the output signal $y(t)$ and the input signal $x(t)$ are sine waves. Strictly speaking, many analyses should be prefaced by the statement "only the amplitude and phase are considered to change." We do not explicitly make this statement hereafter. Three preconditions are typically set when solving engineering problems: they are linear systems, they satisfy causality, and they are time invariant. These preconditions are comparatively easy to satisfy in practice, but they are needed to make system analysis and synthesis feasible.

3.1.3 Fourier Transform

There are two ways of defining the desired $f(*)$, where the goal is to convert the input signal $x(t)$ into an output signal $y(t)$. We can either work in the time or frequency domain. The process in the time domain is already illustrated in Figure 3.2. The process in the frequency domain is described using $F_x(\omega)$ and $F_y(\omega)$, which can be converted through Fourier transformation

from $x(t)$ and $y(t)$. Mathematically, the Fourier transformation is an integral conversion of the function $x(t)$ with variable t as shown in (3.1).

$$F_x(\omega) = \int_{-\infty}^{\infty} x(t)\exp(-2\pi j\omega t)dt \tag{3.1}$$

where j is the common imaginary unit.

The result of this integration is naturally a complex number because $\exp(-2\pi j\omega t)$ is a complex number. This Fourier transform is an extension of the Fourier series expansion, and this conversion has a very important meaning in signal processing. The process in Figure 3.2 can be transformed into that of Figure 3.3 by Fourier transformation. The variable time t in the time domain of Figure 3.2 becomes variable ω, or angular frequency, in Figure 3.3, which deals with the frequency domain. Because this variable can be thought of as representing the angular frequency of the sine wave response, the signal is said to be described in the frequency domain. Theoretically, the time and frequency domains are of equal importance. Just as in engineering tasks, signal processing in the frequency domain is often, but not always, more convenient. For instance, the reduction of noise in the signal $x(t)$ is often equivalent to removing high-frequency components in the frequency domain. This, of course, is not always true. Nevertheless, trying to form $f(*)$ in the time domain to achieve a similar noise reduction is extremely difficult. The conclusion is that the challenge of creating the system S depends directly on the difficulty of forming $f(*)$.

3.2 Digital Signal Processing

3.2.1 Digital System

We say that a system is analog if the variable t, the input signal $x(t)$, and the output signal $y(t)$ take continuous values. Conversely, a system whose variable t, input signal $x(t)$, and the output signal $y(t)$ take only discrete values, is called a digital system. Time t is continuous, and most physical phenomena output continuous signal values. This suggests that digital systems are not

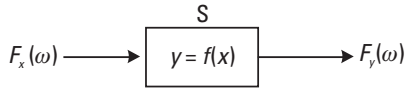


Figure 3.3 System with its input and output in the frequency domain.

spontaneous but instead represent the desire to sample a continuous signal at regular intervals. Such a system is often called a logic circuit or a digital system. The input signal and the output signal are generally expressed as series of binary (0 and 1) bits. For example, (0110111) is a 7-bit binary number and equals 55 in the common decimal notation. A system in which the input signal's interval is fixed to some reference clock is called a synchronous system; if no reference clock is used, the system is called an asynchronous system. Synchronous systems are used most often due to their operational stability.

There are two basic types of digital systems: the sequential system and the combinational system. In the former, the output is decided by the previous state (or states) of the circuit and the input signal. In the latter, the output is decided only by the instantaneous input signal. A sequential circuit is achieved by adding memory to a combinational circuit. Sequential circuits are needed for complex processing operations, and thus most digital circuits are of sequential character.

3.2.2 Analog and Digital Signals

In a digital system, the input signal should be also a digital signal. In order to digitize an analog signal, we need to determine two parameters: the sampling interval or frequency and the resolution of digitization (i.e., the number of separate analog levels and respective bits used to express the discrete sampled signal values).

To make the explanation of signal digitization easier to understand, we use Figures 3.4 and 3.5. The analog signal (which could be, for example, music or video) shown in Figure 3.4(a) is sampled at regular time intervals to create a series of discrete values [shown in Figure 3.4(b)]. Such plots are called pulse amplitude modulation (PAM) signals since the value that varies with time is the amplitude. The PAM signal is converted into a digital signal in Figure 3.5. In this instance, amplitude is converted into 7-bit digital data.

The digital signal obtained by this conversion becomes the input signal of the digital system introduced previously. Sampling is the precursor to digital signal processing in our inherently analog world. Note that we can reverse this conversion by turning a digital signal into a PAM signal by smoothing between the discrete digital values. The device that can create a PAM signal is called a sample and hold circuit. The device that quantizes a PAM signal into a digital signal is called an analog-to-digital (A/D) converter. The device that converts digital signals into PAM signals is called a D/A converter. The

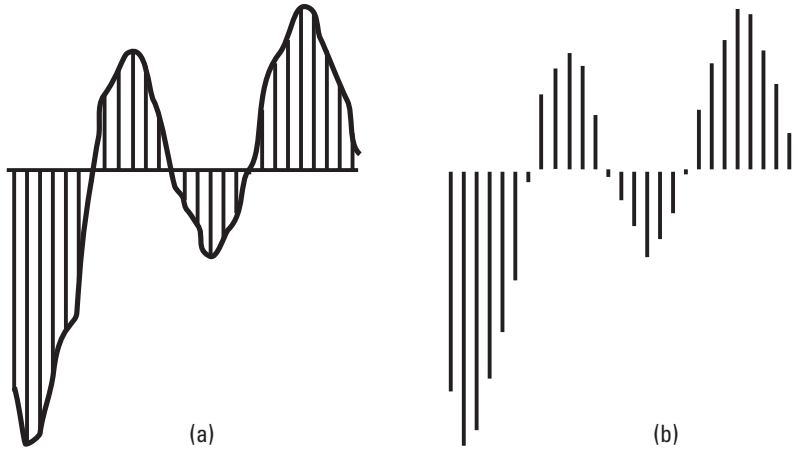


Figure 3.4 Sampling and PAM signals after the sampling process: (a) original analog signal, and (b) PAM signals.

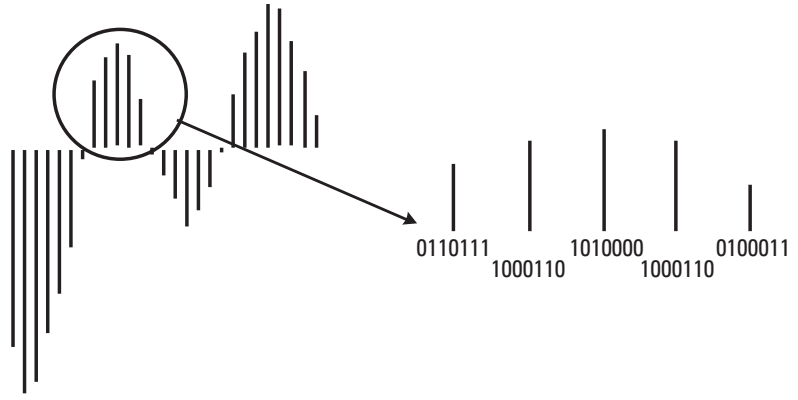


Figure 3.5 Quantization of a PAM signal.

device smoothing PAM signals is called a filter circuit. In practice, many commercial A/D converter chips contain a sample and hold function while most D/A converters contain a filter circuit.

If we want to convert an analog signal into a digital one, what should be the sampling interval? The answer depends on the goal of the process. Analog signals usually contain many different frequency components. The sampling interval should be set to capture the desired information in the original signal and this is the fundamental problem. Consider Figure 3.6(a),

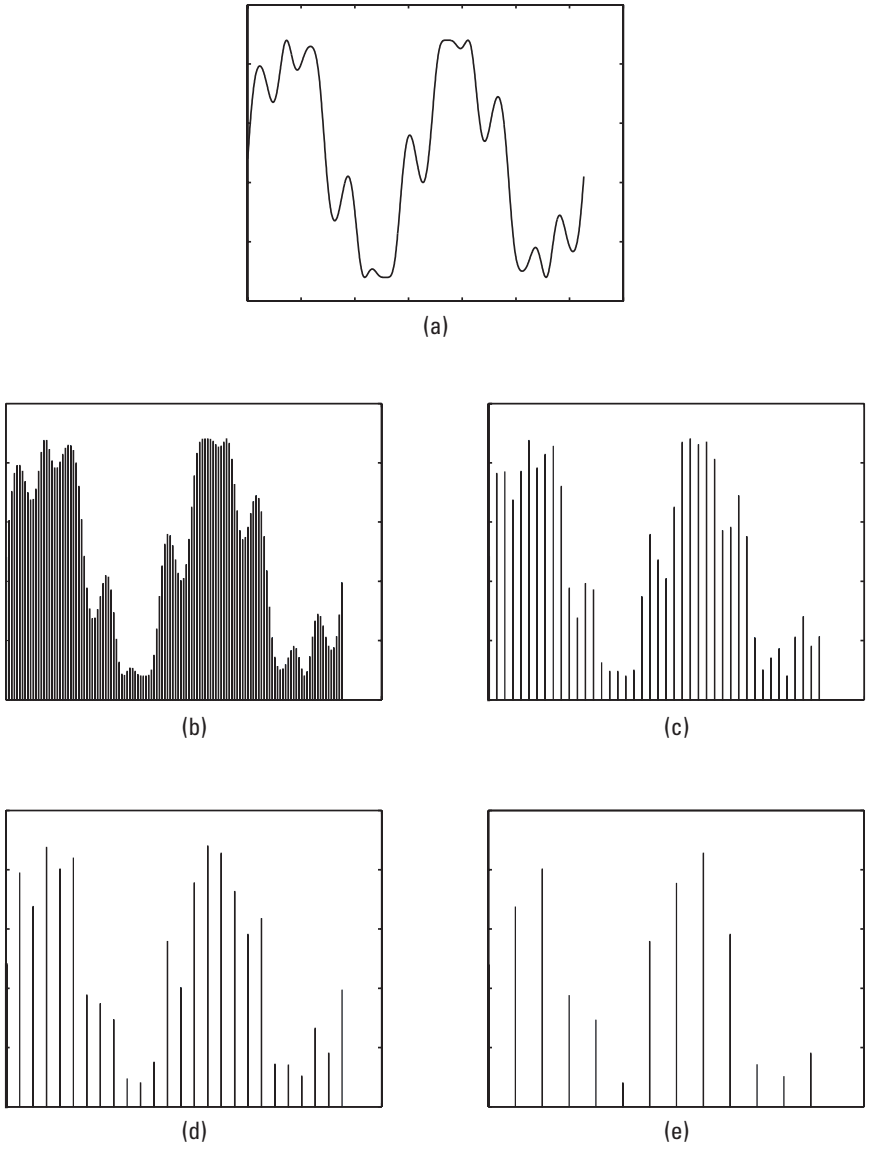


Figure 3.6 Time intervals in sampling: (a) original analog signal, (b) time interval T1, (c) time interval T2 (T2 >T1), (d) time interval T3 (T3 >T2), and (e) time interval T4 (T4>T3).

which shows an original analog signal, and Figure 3.6(b–e), which shows discrete signals sampled using different time intervals (these discrete signal levels

are not yet quantized). Figure 3.6(b, c) accurately records the level variation of the original analog signal along with elapsed time. Figure 3.6(d) retains most of the original signal, but some small peaks and valleys have already been lost. Figure 3.6(e) captures only relatively large peaks and valleys. The loss of signal details means that the original high-frequency components are no longer included in the sampled discrete signal. The sampling time interval shown is too large. On the other hand, the sampling time intervals in Figure 3.6(b, c) are sufficient to yield accurate digital signals. It is obvious, however, that short sampling time intervals greatly increase the volume of digital data produced. The trade-off between the sampling period and the level of detail is well captured by Shannon's sampling theorem, which provides a practical guide to the maximum effective sampling frequency.

Shannon's Sampling Theorem

> When the maximum frequency present in an analog signal is assumed to be f_m, this analog signal can be completely restored from a PAM signal, which is sampled by using a time interval shorter than $1/(2 \bullet f_m)$.

The above description is not strictly correct, since Shannon's goal was comprehensibility. This, however, does not cause many serious problems in practical use. The most notable exception is the additional requirement of a continuous process—theoretically one that goes on forever. Shannon's criterion does not work if the duration of the original signal is short.

Let us now turn to the other parameter, the number of bits used to express the signal amplitude. Once again there is a trade-off involved. Increasing the number of bits used increases the costs involved with regard to hardware scale and processing speed, but this allows more information in the original signal to be captured. Unfortunately, there is no guideline equivalent to Shannon's theorem. The problem related to the number of bits can be solved only empirically for each specific task. It depends on the application. Most image processors use a quantization depth of 8 to 14 bits while speech signals need 12 to 22 bits. We must note that the optimum quantization depth can be determined only by evaluating the characteristics of the application such as the subjective effects of image noise or clarity of sound.

3.2.3 Spectrum

When we perform a Fourier transformation on an analog signal, we create the functions $F_x(\omega)$, $Fy(\omega)$, which take complex values with variable ω (see

Section 3.1.3). The absolute value of these functions plotted against frequency is called a spectrum.

A similar Fourier transform, the discrete Fourier transform (DFT), can be defined to express the digitization process. The conversion result is a discrete numeric series of complex values, and the resulting absolute values also form a spectrum. The two spectra of the analog and corresponding digital signal agree to a certain extent. The degree of similarity involves extremely complex issues and we do not go into detail here. According to its history [1, 2], DFT was first used to obtain the spectra of statistical data and analog signals. DFT is here defined as (3.2). The PAM signal series $x(0)$, $x(1)$, $x(2)$, … is obtained by sampling an analog signal.

$$X(k) = \sum_{n=0}^{N-1} x(n) \exp\left(\frac{-2\pi jkn}{N}\right) \quad 0 \le k \le N-1 \tag{3.2}$$

The output $X(0)$, $X(1)$, $X(2)$ ... from this conversion is a series of complex numbers. The discrete spectrum of the PAM signal is obtained by taking the absolute values of this numeric series.

3.2.4 Method of Calculating DFT

An efficient DFT computation algorithm is available for cases where the involution values are powers of 2. Efficiency here means few multiplications and few additions. This does not mean, however, that such a method is the fastest for present processors. The most famous algorithm is the Cooley-Tukey fast-Fourier transform (FFT) [1]. Its design emphasized compound data permutation as a way to minimize the number of multiplications and additions. Compound permutations, however, take longer in current processors than does arithmetic processing. Current processors have special high-speed adders and can carry out multiplication and addition instructions at maximum speed. In addition, most processors offer pipelining so that arithmetic instructions can be continuously processed.

The Winograd DFT algorithm [3] is said to be the most efficient way to minimize the number of multiplications and additions. This method is also a variant of the FFT and employs compound array substitution. Therefore, it is not suitable for modern processors dedicated to digital signal processing including digital signal processors (DSPs). Given the attributes of current processors, it appears better to realize DFT by applying a simple format that can continuously pipeline data. Complex permutations should not be used.

3.2.5 Description of Signals and Systems in the Frequency Domain

The result of the DFT calculation is a description of the signal in the frequency domain. The two signals $x(t)$, $y(t)$ in Figure 3.2 can be described as a function of frequency ω. This description allows us to extract the amplitude and phase characteristics of the signal. Devising a system S that operates in the time domain is possible. This means, however, that it is necessary to understand which output signal is connected to a certain input signal. Such a process is indirect and troublesome. A method that more directly describes the operation of system S as a function of frequency ω is required.

The two candidates are the Laplace transform and the z-transform. It is natural that these are related to the Fourier transform. Moreover, when impulse signals or white noise are used as the input signal, these two methods produce equivalent results. Describing the operation of system S as a function of frequency is to describe the output given the input of an impulse signal or white noise. The reason for this is that such input contains all frequency components and so provides a comprehensive assessment of system operation.

3.2.6 Laplace Transform

The Laplace transform is widely used as a means to analyze the frequency characteristics of an analog system (see Figures 3.1 and 3.2). It represents the conversion of function $x(t)$ with variable t as shown in (3.3). It is a straightforward operational calculus that gives a systematic solution and is related to the Fourier transform as previously described.

$$F_x(s) = \int x(t)\exp(-st)dt \tag{3.3}$$

where s is a complex variable.

Its similarity with the Fourier transform seems clear. It is equivalent to the Fourier transform if s is assumed to be an imaginary number. For some time the Laplace transform and operational calculus were said to be identical even though no strict mathematical proofs had been provided. Such proofs have now been provided by Mikusinski [4] and Schwartz [5] in different ways.

Adopting the Laplace transform in Figure 3.2 yields Figure 3.7. However, this is a simple rewriting action of variable ω (a real number) to variable s (a complex number). The meaning of this is explained later in this section.

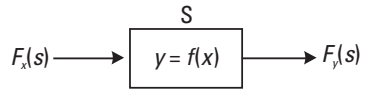


Figure 3.7 Laplace transform expression for a system.

The operation of a system S (see Figure 3.7) can be written as (3.4) under some assumptions such as linearity.

$$F_y(s) = T(s) \bullet F_x(s) \tag{3.4}$$

The reader is encouraged to consult [2] or [6] for details of such assumptions and for a more complete description. $T(s)$, called the transfer function, describes the operation of the system. It is assumed here that s is a complex number and can be described as shown by (3.5).

$$s = a + jb \tag{3.5}$$

where j is the common imaginary unit (written according to the general convention of electrical engineering where i is reserved for the electric current).

As an example, the case where $a > 0$ is the typical response of an attenuating sine wave; for an increasing sine wave the response can be described by $a < 0$. It is obvious that the response of a constant amplitude signal is described by $a = 0$; this case is equivalent to the Fourier transform.

In general, the transfer function $T(s)$ often becomes a polynomial type expression in which the coefficient s is a real number [see (3.6)].

$$\begin{aligned} T(s) &= \frac{q(s)}{p(s)} \\ p(s) &= a_1 s^{n-1} + a_2 s^{n-2} + \ldots + a_n s + a_{n+1} \\ q(s) &= b_1 s^{m-1} + b_2 s^{m-2} + \ldots + b_m s + b_{m+1} \\ & \qquad (n > m) \end{aligned} \tag{3.6}$$

Those values of s that satisfies $p(s) = 0$ are called poles, and those that yield $q(s) = 0$ are called zeros. In general, these two variables are complex numbers.

The absolute value of this transfer function *abs*($T(s)$), or $|T(s)|$ is called the frequency response of the system S (versus sine waves), and the argument *arg*($T(s)$) or $\angle T(s)$ shows the phase properties of $T(s)$. In addition, the derivative of this phase characteristic, $\frac{d\varphi(\omega)}{d\omega}$, expresses the group delay characteristic.

3.2.7 *z*-transform

Another conversion used to describe the behavior of a system in the frequency domain is the *z*-transform. A *z*-transform is defined against a discrete-time system (PAM signal series) $x(0)$, $x(1)$, $x(2)$, …, as shown in (3.7).

$$X(z) = \sum_{n=-\infty}^{\infty} x(n) z^{-n} \tag{3.7}$$

Although (3.7) and DFT yield similar conversion results, $X(*)$, confusing them is unlikely because DFT produces a complex series, while the *z*-transform yields a function with variable of z or $z-1$. This equation gives an abstract definition of a discrete-time system. The PAM signal values are continuous. However, $X(z)$ is not, in practice, calculated by applying this equation to actual PAM values. Therefore, we do not have to worry so much about a summation to infinity.

The transfer function of a discrete-time system can be similarly defined by using the *z*-transform as a polynomial expression—that is, a rational function of z or z^{-1} as shown in (3.8).

$$\begin{aligned} T(z) &= \frac{q(z)}{p(z)} \\ p(z) &= a_1 z^{n-1} + a_2 z^{n-2} + \ldots + a_n z + a_{n+1} \\ q(z) &= b_1 z^{m-1} + b_2 z^{m-2} + \ldots + b_m z + b_{m+1} \\ & \qquad (n > m) \end{aligned} \tag{3.8}$$

A rational function of z^{-1} is used here. Since z^{-1} obviously means delay, its use is very convenient. There is no difference between using either z or z^{-1} since a rational function is a multinomial expression. Here, z and z^{-1} only

indicate whether the numerator and denominator have been divided by z^{-1} or not. Those values of z that satisfy $q(z) = 0$ are called zeros, and those that give $p(z) = 0$ are called poles in this case, too.

3.2.8 Fundamental Property

If an analog system is to be described by the Laplace transform or a discrete-time system is to be described by the z-transform, they must have two fundamental properties: stability and causality. The first condition means that the system output will never increase to infinity even if a signal is added to its input. The second rule requires that there is no output if there is no input.

The stability of both analog and discrete systems is extremely easy to describe. For the analog case, stability is assured when the poles of the transfer function do not exist on the right half of the complex plane. For the discrete-time systems, the poles must lie within in the unit circle. The Routh-Hurwitz method (for analog systems) and the Shur-Cohn method (for discrete-time systems) are often used for stability evaluations [7, 8].

Whether a system is causal or not cannot be defined just from the coefficients of any transfer function. A key problem is the difficulty of finding a time-based function that well describes the system response. One condition of causality is that the transfer characteristics of the system can be defined using a weighting function. Basic system parameters other than stability and causality can be expressed as conditions of the transfer function. Refer to [2, 6] for details.

3.2.9 Spectrum Estimation

It seems reasonable to describe the input and the output as time-based variables. The reason for this is that we actually measure the physical quantity over time. This approach looks even more valid for systems that experience probabilistic variations. Even if the largest part of a system is deterministic, it very likely has at least some probabilistic noise.

While time-based measurements are intuitive, we find that an analysis in the frequency domain provides more information of the system's characteristics. Spectrum estimation was developed for this purpose at the end of the nineteenth century. Its first application was to extract simple sine waves buried in noise. The current status in signal processing technology was achieved in the 1950s. The fast Fourier transform is one of the key advances.

The basic method of calculating the spectrum from a time-based signal involves the use of the Wiener-Khinchine relation. It is an expression that links the correlation function to the spectral density. The correlation function is actually calculated from the time-based signal. The spectral density is convenient for grasping the response of the entire system. One can interpret the spectral density as a term that modifies the form of the spectrum.

The Wiener-Khinchine relation shows that an autocorrelation function $R(\tau)$ and the spectral density $S(\omega)$ are related to each other through the Fourier transform and its inverse transform, as shown in (3.9).

$$\begin{aligned} S(\omega) &= \int_{-\infty}^{\infty} R(\tau)\exp(2\pi j\omega\tau)d\tau \\ R(\tau) &= \int_{-\infty}^{\infty} S(\omega)\exp(2\pi j\omega\tau)d\omega \end{aligned} \tag{3.9}$$

This equation is cumbersome for defining the spectral density. More practical equations are available, but care must be taken in applying them. For example, (3.10) shows a more fluent form of the correlation function for discrete data.

$$R_k = E\left[x(n)x(n+k)\right] \tag{3.10}$$

$E[\ ]$ shows the expectation (average) of [ ].

Equation (3.11) associates the spectral density and the correlation function of a continuous system with the respective Fourier transform.

$$\begin{aligned} S(f) &= \Delta\sum_{k=-\infty}^{\infty} R_k \exp(-2\pi jfk\Delta) \\ R_k &= \int_{-2\Delta}^{2\Delta} S(f)\exp(-2\pi jfk\Delta)df \end{aligned} \tag{3.11}$$

The single-side spectral density $\phi(f)$, defined in the range of $0 < f < \infty$, is intuitively reasonable, although it is convenient to define $S(f)$ in the whole range of $-\infty < f < \infty$for purposes of computation.

$$\phi(f) = 2S(f) \tag{3.12}$$

where $0 < f < \infty$.

3.3 Digital Filters

3.3.1 What Is a Digital Filter?

A filter is a function or device that passes only the desired signal component(s). A digital filter basically emulates an analog filter. A digital filter can be used to process an analog signal, but this requires that the signal be first sampled and quantized. An analog signal can be digitally filtered in real time if hardware with sufficient speed is available. Such a filter offers the following benefits:

- Accurate signal values (within the available resolution);
- High reproducibility (theoretically limited only by numerical rounding errors);
- Processing flexibility (depending on the platform used for the realization);
- Stable system operation (particularly if the effects of any operating system are excluded).

The initial digital filters suffered from the following problems:

- Large size and physical complexity;
- Difficulty of processing high-frequency signals in real time;
- Significant arithmetic errors;
- High power consumption (partly due to size and complexity).

Recent advances in LSI technology have greatly reduced these problems, although they have not totally disappeared. The result is that digital filters are being widely used in many applications. Though the digital filter, as defined above, may seem complicated, in practice a comparatively simple linear device can be created by combining an adder, a multiplier, and unit delay elements. Its arithmetic processes mainly involve multiplications. Figure 3.8 shows the three elements that can be combined to create digital filters that give quite advanced responses. A digital filter is usually drawn as either a block diagram or as a signal flow graph. While it seems more intuitive to use a block diagram, flow graphs are more effective when comparing complicated structures.

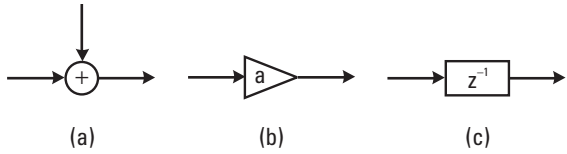


Figure 3.8 Components in a digital filter: (a) adder, (b) multiplier, and (c) unit delay.

Let us have a look at some of the fundamental properties of flow graph presentations. A signal flow graph consists of nodes and associated branches. Each branch has a definite direction and one of the two possible weights: $z-1$ or a_i. A node can have any number of input and output ports. Signals on more than one input port are combined or added, while signals on more than one output port are "split" from the same original signal and are therefore identical. If one or more of the constants of the filter's multipliers are changed while filtering is in progress, we end up with an adaptive linear digital filter. This type of filter, also called a nonlinear filter, is not discussed here.

3.3.2 Linear Digital Filter

There are two basic classes of linear digital filters:

1. Finite impulse response (FIR) filters (Figure 3.9);
2. Infinite impulse response (IIR) filters (Figure 3.10).

This classification is based on the impulse response of the filter treating it as a black box. A classification by internal structure is also possible and yields the following two types:

1. Nonrecursive type filters;
2. Recursive type filters.

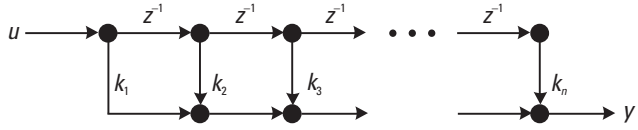


Figure 3.9 General FIR filter configuration.

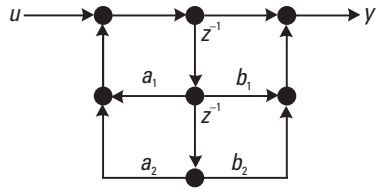


Figure 3.10 General IIR filter configuration.

Here we consider the possible existence of one or more closed loops inside the filter. These classes of either type can be combined since FIR filters are equivalent to nonrecursive type filters, and IIR filters are at the same time of recursive type.

Though a real-time FIR filter may require significant amounts of high-speed hardware, it can be used to realize filters with various perfect characteristics. A typical application might be an ideal linear phase filter with no distortion. On the other hand, the IIR filter offers generally lower hardware requirements and excellent selectivity at the cost of increased distortion. Both types are used widely, and it is impossible to definitively state that one is superior to the other.

3.3.3 Design Methods for Digital Filters

In principle, various combinations of adders, multipliers, and unit delay elements are possible. The structures shown in Figures 3.9 and 3.10, however, are the most common start-up arrangements for FIR and IIR filters. The IIR filter shown in Figure 3.10 is called the Biquad design. Cascading several of these blocks yields a higher-order filter just as happens in the analog world. Such chains posses a transfer function that has second-order terms (of $z-1$) in both the numerator and denominator. Cascading a transfer function corresponds to the conversion of $T_i(z)$ into the first- or second-order equivalents in terms of $z-1$, and connecting it to the next transfer function. The complete transfer function $T(z)$ is shown in (3.13).

$$T(z) = T_1(z) \bullet T_2(z) \ldots T_k(z) \tag{3.13}$$

The case of a first-order function corresponds to the situation in which the constants in Figure 3.10 are $a_2 = 0,\ b_2 = 0$.

Although there is a unique solution, which factorizes an arbitrary higher-order multinomial of types $p(s)$, $p(z)$ to a first- or second-order multinomial, there are various combinations of the first- or second-order multinomials $T_i(z)$ that are a rational function. This is called the problem of order pairing. The best combination should be found regarding dynamic range. In practice, this pairing is decided by the utility program of the selected DSP. When a higher-order transfer function is achieved directly without cascading Biquad filters, the system response is greatly influenced by even slight errors in $T(z)$. In practice, such errors in the coefficients a_i or b_i are unavoidable during quantization and in filter operation. Therefore, the cascaded Biquad configuration is the most commonly used form of digital filters.

3.3.4 Key Points in Designing a Digital Filter

The signal flow graph of a filter is merely an abstract, yet precise, description. It does not show the actual hardware. Filters can be roughly classified into wired logic devices and programmable logic systems. The former consists of various combinations of common logic gates such as AND, OR, NOT, NAND. The latter is formed by a program running on a processor. The term "wired logic" is reasonable since the primary function is defined by the arrangement of the "wires" used to join the logic elements. The main factor limiting the speed of wired logic systems is the number of consecutive gates. Other restrictions are the cost, power consumption, and manufacturing difficulties. Adding more functions increases geometrically the number of gates needed. This approach greatly increases system instability, and the required performance may not be achievable at all.

Since wired logic systems are ultimately fixed in nature and the complexity of filter functions will continue to increase, we need an alternative method that offers better flexibility. An effective reuse and sharing of subsystems involved in a device would limit the increase in the amount of hardware. This is possible with a programmable logic system. It places more emphasis on software operation and memory constraints. Because one subsystem, however, cannot simultaneously implement many different functions, sharing implies putting things in a queue as a function of time. Delays are an unavoidable trade-off.

The processor creates all necessary functions by running one or more programs to process the raw data in memory. Since software can be freely rewritten or invoked, programmable logic systems are far more flexible than their wired counterparts. A simple heuristic is "use wired logic systems to achieve high speed and programmable logic systems to achieve flexibility."

However, modern processor designs and semiconductor technology behind them are challenging this assertion. One example is the modern DSP with pipeline operation that can perform consecutive summations and multiplications at high speed. Accordingly, such DSPs are being used in various fields.

References

[1] Oppenheim, A. V., and R. W. Schfer, *Digital Signal Processing*, Englewood Cliff, NJ: Prentice-Hall, 1975.

[2] Oppenheim, A. V., and A. S. Willsky, *Signals and Systems*, Englewood Cliffs, NJ: Prentice-Hall, 1983.

[3] Blahut, R. E., *Fast Algorithm for Digital Signal Processing*, Reading, MA: Addison-Wesley, 1984.

[4] Mikusinski, J., *Operational Calculus*, London: Pergamon Press, 1959.

[5] Schwartz, L., *Theorie des Distribution*, Paris: Hermann, 1951.

[6] Roberts, R. A., and C. T. Mullis, *Digital Signal Processing*, Reading, MA: Addison-Wesley, 1987.

[7] Schwarz, R. J., and B. Friedland, *Linear Systems*, New York: McGraw-Hill, 1965.

[8] Jury, E. I., *Theory and Application of the z-Transform Method*, New York: Wiley, 1964.

4

Introduction to PLLs

4.1 Basic Mechanisms in Phase Synchronization

The characteristics of complete PLLs depend on their elements. The basic blocks are a phase comparator, a filter, and a controlled oscillator. Even complex PLLs share this configuration. Moreover, since the PLL output is synchronized to the input signal, the characteristics of the input signal also influence PLL performance.

Figure 4.1 shows the basic configuration of the PLL. The phase difference, measured with a phase comparator, is fed back to a controlled oscillator. The output of this oscillator is altered in order to hold the phase difference below a defined value. A filter is added between the phase comparator and the controlled oscillator (compare with Figure 2.3, which demonstrates a basic phase synchronization system). Figure 4.2 shows an example of the phase synchronization process. Here, the output frequency of the controlled oscillator is initially lower than that of the input signal, and the phase difference changes over time, as shown in area (a) of Figure 4.2. The output frequency is driven higher than the input frequency, and the phase difference changes as shown in area (b) of Figure 4.2. If this control action is performed well, the output frequency of the controlled oscillator coincides with the frequency of the input signal. This holds the phase difference within the required range. Phase synchronization is achieved in area (c) of Figure 4.2. Figure 4.2 depicts digital control. The oscillator frequency is discretely

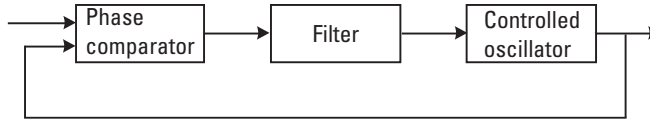


Figure 4.1 Basic PLL configuration showing the phase comparator, the loop filter, and the controlled oscillator.

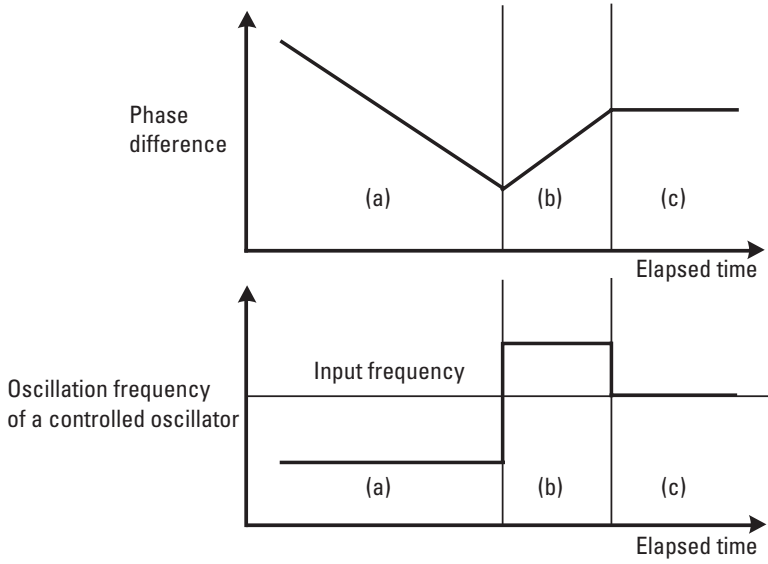


Figure 4.2 Phase difference and output frequency change over time during a phase synchronization process (digital control): (a) initial condition, (b) condition between the first and second digital controls, and (c) condition after the second control.

controlled. In analog control, the phase difference and the output frequency smoothly change as shown in Figure 4.3.

4.2 Element Operation

4.2.1 Phase Comparator Characteristics

A phase comparator is placed at the input of the PLL, and it is the first basic element to receive the input signal. In addition, the output of the controlled oscillator is connected to it. The comparator output corresponds to the phase difference between the two. The difference information in Figure 4.4 is

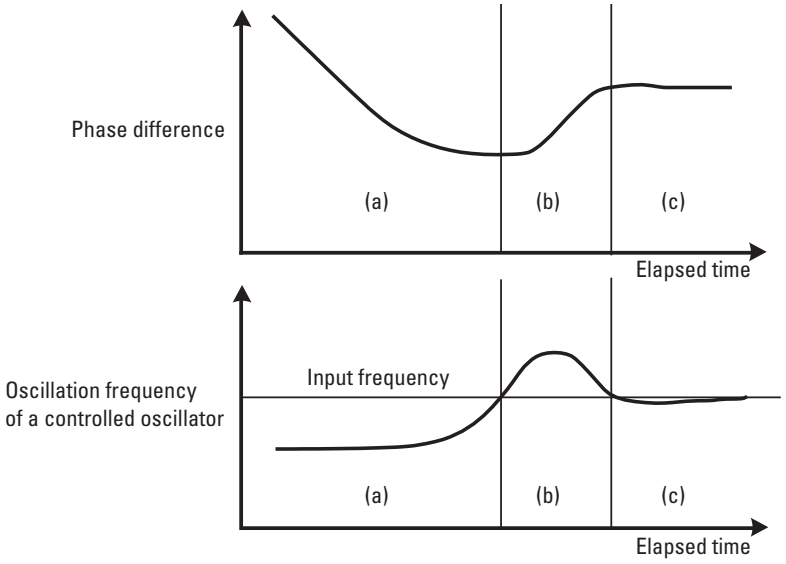


Figure 4.3 Phase difference and output frequency change over time during a phase synchronization process (analog control): (a) initial condition, (b) condition in analog control, and (c) after the control.

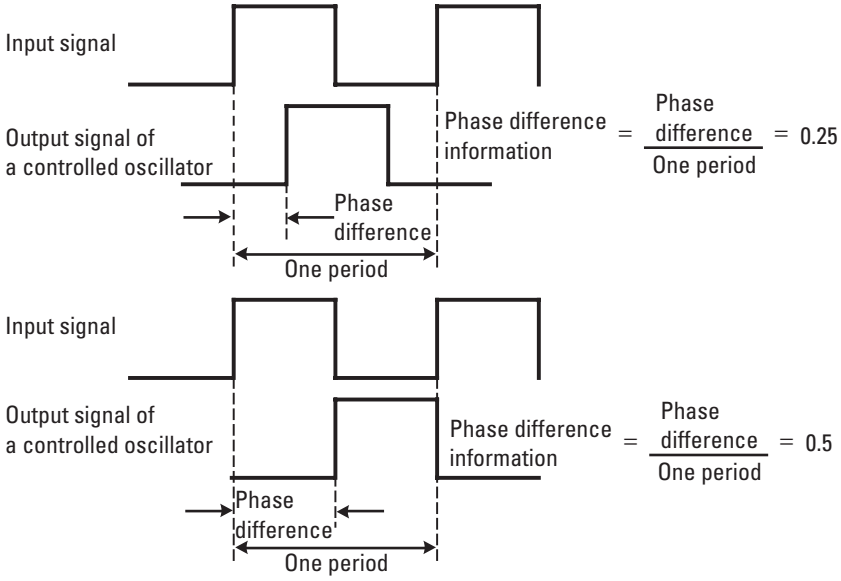


Figure 4.4 Phase difference information in the analog phase comparator output.

the ratio of the leading edge time difference to the period of the input signal, and it takes continuous values. This information is digitized as shown in Figure 4.5. Detection happens with a certain resolution, and we get a quantized value. The phase difference information is in this example an integer number from 1 to 128 since one period has been divided into 128 units in Figure 4.5. The detected phase difference has the value of 30 in Figure 4.5. The phase comparator characteristic can be expressed by the relationship between the practical phase difference and the output phase difference information. Since this information is reset every period, the typical relationship between the practical phase difference and the output phase difference information is as shown in Figure 4.6.

4.2.2 Controlled Oscillator Characteristics

The most significant difference between a PLL and passive devices such as filters is that the PLL has an active, internal signal source, the oscillator. Thus, the PLL does not directly output a processed form of the signal it receives but generates its own signal. This requires frequency and phase synchronization. The oscillator frequency has to be changed so as to synchronize it to the input signal of the PLL. The loop process of the PLL aims at phase synchronization, even if the process is explained in terms of frequency. The output phase of the controlled oscillator must be synchronized to the input signal

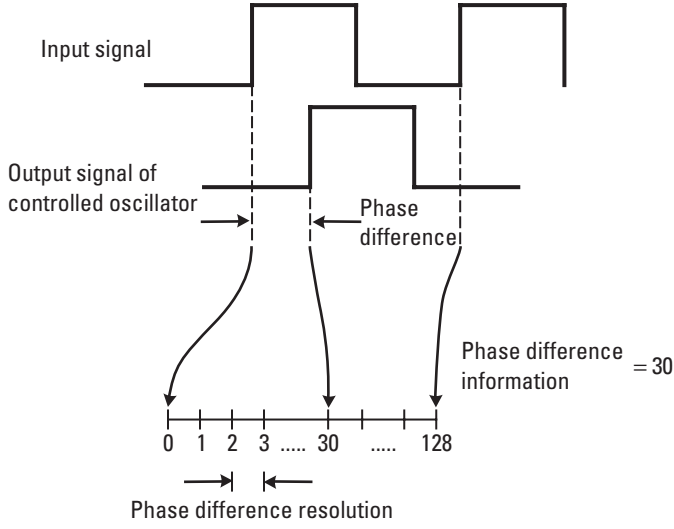


Figure 4.5 Phase difference information in the digital phase comparator output.

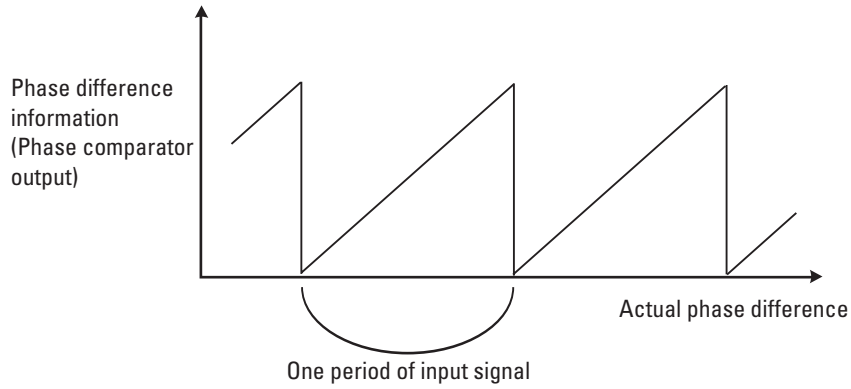


Figure 4.6 Phase comparator characteristic.

phase. The oscillator phase is generally controlled by changing the output frequency through a dedicated port. This input port receives the phase difference information coming from the phase comparator. The transfer function of the controlled oscillator can be specified by the relationship between the input control information and the output frequency.

One of the most important characteristics is the frequency range of the controlled oscillator. A wide frequency range offers two advantages: the PLL can synchronize to a wide range of input frequencies, and it can rapidly follow the fast phase changes caused by large frequency changes. These features are essential in synthesizers or when switching between frequencies, as is required in some frequency hopping radio systems. A typical controlled oscillator characteristic is shown in Figure 4.7.

4.2.3 Filter Characteristics

A filter is usually inserted between the phase comparator and the controlled oscillator. The phase synchronization function can work without the filter if the output of the phase comparator is directly brought to the controlled oscillator. The filter is used to precisely adjust the phase characteristics. Low-pass filters are generally used in PLLs, as shown in Figure 4.8. This configuration suppresses the undesirable high-frequency components not used in phase synchronization. Actual filter circuits often combine two different filters. The influence of filter characteristics on the overall PLL performance is described in Chapter 5.

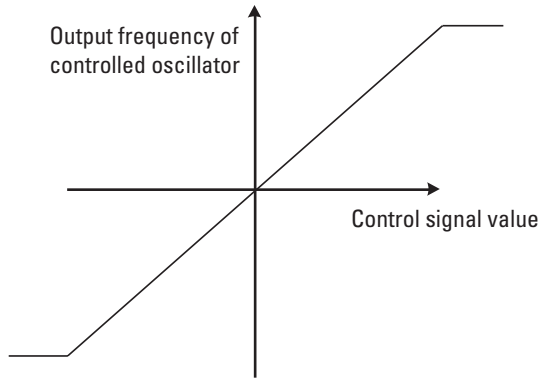


Figure 4.7 Relationship between the applied control signal and the output frequency of a controlled oscillator.

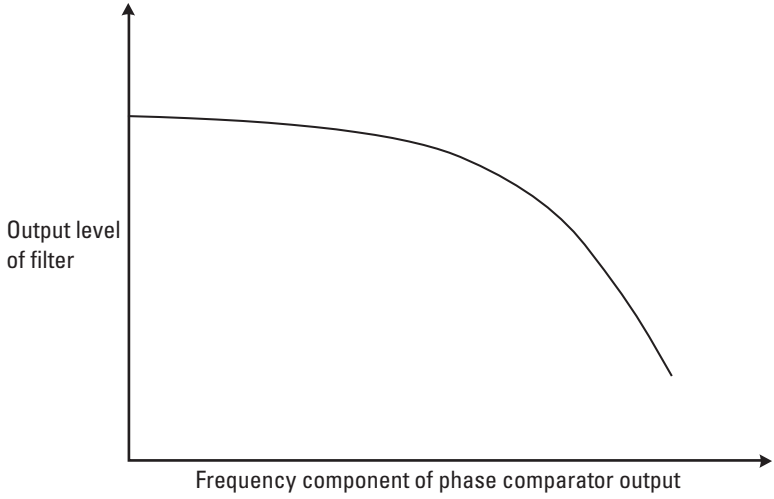


Figure 4.8 Lowpass filter frequency response.

4.3 Classification of PLLs

The PLL is characterized by the configurations of its phase comparator and controlled oscillator and by signal type. Analog and digital variants of the comparator and controlled oscillator are known. These elements accept analog or digital signals or a combination of them. This book recognizes four types of PLLs, as shown in Figure 4.9.

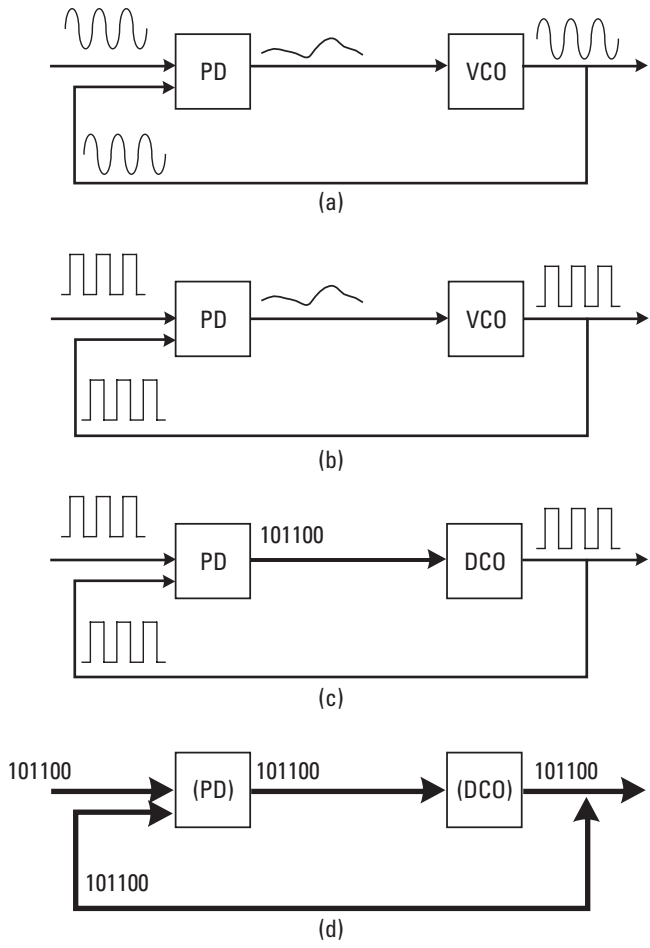


Figure 4.9 PLL classification: (a) analog PLL, (b) analog-digital PLL, (c) digital PLL, and (d) digital processing PLL.

4.3.1 Analog PLL

All signals and ports are of analog nature as indicated in Figure 4.9(a). Phase comparison results appear continuously at the output. The output frequency and phase show smooth and continuous changes since the controlled oscillator is continuously steered by the analog signal from the comparator. A

practical realization can be based on a double balanced RF mixer and a suitable voltage controlled oscillator (VCO). The main feature of the analog PLL is its excellent control precision.

4.3.2 Analog-Digital PLL

The input and output signals of this PLL are square waves [see Figure 4.9(b)]. It appears to be a kind of digital PLL. The output of the phase comparator, however, is analog and its internal processes are based on analog signals. The comparator does not process the digital information contained within the square wave, but instead uses the timing of the input signal. Such a comparator can be made of an XOR logic gate. This type of PLL is called the analog-digital PLL in this book. The analog-digital PLL can be used in digital circuits. It enables the controlled oscillator to achieve the precise phase control characteristics of the analog PLL. Note that the analog-digital PLL can not trace phase changes faster than one period of the square wave.

4.3.3 Digital PLL

The configuration of the digital PLL is shown in Figure 4.9(c). The timing differences between the transient changes in the input signal are usually detected in the phase comparator. It differs from the analog-digital PLL in Section 4.3.2 in that the phase comparison result is quantized. The output frequency of the oscillator is controlled digitally instead of by an analog signal. The entire PLL can be made of digital elements, which means that highly integrated PLL integrated circuits (ICs) can be achieved. Examples of this type of PLL are described in Chapter 8 of this book.

4.3.4 Digital Processing PLL

The digital processing PLL deals in purely arithmetical values in its internal circuits. The signal waveforms are meaningless at the input and output of the phase comparator, and phase information is expressed using the digital values extracted from the signal patterns. This differentiates it from the digital PLL, which uses the signal wave itself for synchronization. Its merit is that all functions can be realized as software. The trade-off, however, between bandwidth and flexibility must be accepted.

4.4 PLL Applications

The four major PLL applications are described below.

4.4.1 Synthesizer

The signal output by the controlled oscillator usually has the same frequency as the reference signal input to the phase comparator. This condition can be changed, however, if divider circuits are inserted between the controlled oscillator and the phase comparator and/or the input part of the PLL as shown in Figure 4.10.

The input frequency of the PLL and the output frequency of the controlled oscillator in the PLL of Figure 4.10 are related to the input frequency of the phase comparator as follows:

$$f_p = f_i \frac{n_i}{m_i} \quad f_p = f_o \frac{n_o}{m_o} \tag{4.1}$$

where n_i, m_i, n_o, and m_o are divider coefficients.

Equation (4.1) means that both the frequency of the PLL input signal f_i and the frequency of the controlled oscillator output f_o become the same frequency at the input of the phase comparator f_p after passing through their respective dividers. Therefore, the relationship between the input frequency f_i and the output frequency of the controlled oscillator f_o becomes

$$f_o = f_i \frac{n_i}{m_i} \bullet \frac{m_o}{n_o} \tag{4.2}$$

It is obvious that the input frequency of the PLL can be converted into a different output frequency by selecting adequate divider coefficients. This is the basic principle of a synthesizer, a frequency conversion system created around a PLL.

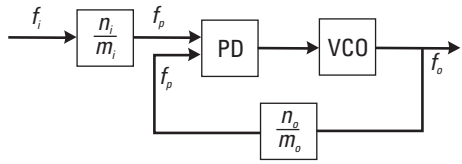


Figure 4.10 Basic configuration of a synthesizer.

4.4.2 Filter (Resonator)

The basic PLL configuration already includes a filter function. Only those frequency components that lie within a certain frequency region appear at the output of the PLL, since the PLL has a bandpass characteristic and cannot follow frequency components out of the bandpass frequency region. In the general resonator [see Figure 4.11(a)], the output level gradually decreases as the input frequency is detuned from the device's resonant frequency. On the other hand, the frequency characteristic of the PLL matches those of a bandpass filter whose passband is very narrow [see Figure 4.11(b)].

4.4.3 Demodulator

The PLL can work as a demodulator by combining the phase comparing function of the PLL with its lowpass filter (bandpass filter) characteristic (described in Section 4.3.2). When the PLL is used as a demodulator, the output point is different from that of the general PLL, as shown in Figure 4.12. While the general output is taken from the controlled oscillator, the output of a demodulator comes from the phase comparator. Consider a signal whose modulation frequency components f_m differ from the center frequency f_o. The frequency component f_o can be extracted at the output of the controlled oscillator by the configuration that suppresses the modulation frequency components by lowpass filtering (bandpass filter). The output of the phase

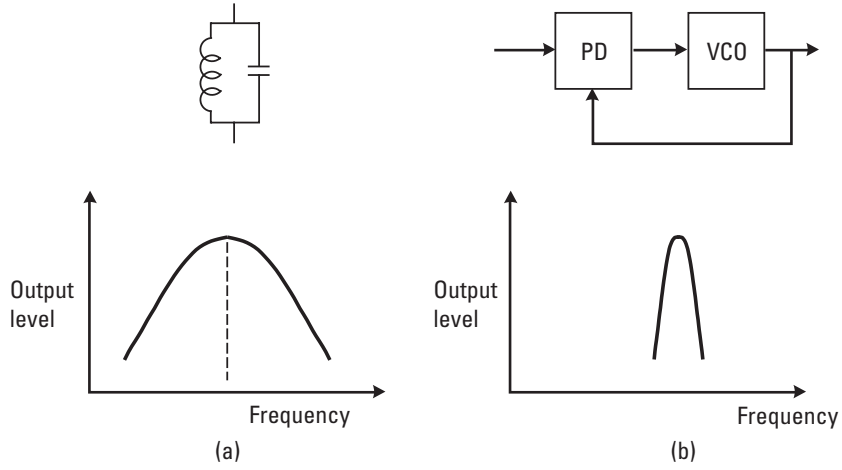


Figure 4.11 Filter constructions: (a) conventional resonator and (b) PLL-based filter.

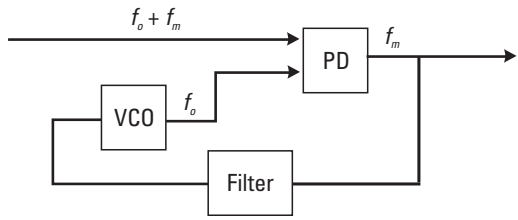


Figure 4.12 Demodulator made of a PLL.

comparator, therefore, is the modulation component f_m that corresponds to the difference between the input signal, which has f_o and f_m components, and the output of the controlled oscillator f_o.

4.4.4 Automatic Control System

The phase synchronization function of the PLL can be applied to realize an automatic control system. Various physical quantities can be controlled if the PLL is made to synchronize the phase of the controlled oscillator to the phase of a reference signal and dedicated transducers are used. The phase synchronization function of the PLL can be used if a cyclic signal proportional to phase difference information can be generated as the control target corresponding to the controlled oscillator as shown in Figure 4.13. The most typical example is the rotation control of motors. Transducer 1 converts the phase difference information to control the rotation speed of the motor, and transducer 2 converts the rotation speed of the motor into a cyclic signal. This enables the rotation speed of the motor to be synchronized to the reference oscillator.

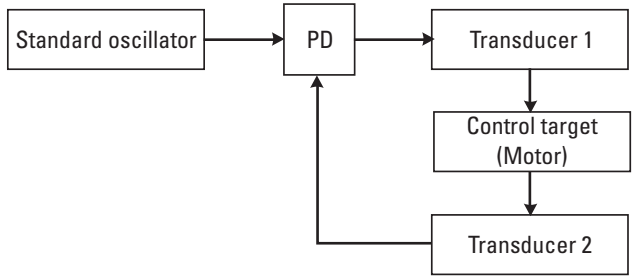


Figure 4.13 Automatic control involving a PLL structure.

4.5 PLL Characteristics [1]

Designing a high-grade PLL from scratch is extremely difficult since many characteristics must be considered. The key characteristics of the PLL are described below, and terms associated with the PLL are explained.

4.5.1 Cycle Slip

The phase synchronization of the controlled oscillator with the input signal can fail. Whether and when this occurs depends on the initial frequency and phase difference between the input and the output of the PLL. Let us assume a phase comparator that has the characteristics shown in Figure 4.14(a) and examine its phase behavior during the synchronization process.

The phase difference $\Delta\varphi_1$ of the two signals at time t_1 in Figure 4.14(b) is converted into output d_1 of the phase comparator. This is the input, which steers the frequency of the controlled oscillator. The new frequency is fed back to the phase comparator. Whether the phase difference decreases after this action depends on the relative relationship between the frequencies of the controlled oscillator output and the input signal. If the polarity of the frequency difference is inverted by the control based on the phase difference at time t_1, the phase difference could be decreased at the next phase comparison. The frequency of the controlled oscillator has to be higher than that of the input signal when the diminishing of $\Delta\varphi_1$ yields synchronization. If the amount fed back to the controlled oscillator, however, does not invert the polarity, the phase difference may increase as seen at time t_2 in Figure 4.14(b). If the polarity could be inverted at t_2, the phase difference diminishes. If, however, the polarity cannot be inverted after time t_2 again, the phase difference exceeds $+\pi$. If the phase difference is larger than $+\pi$, and the polarity of the frequency control is inverted together with the phase comparator output, the output frequency is driven in the direction that increases the frequency difference. Because the PLL decreases the phase difference, it is controlled toward $+2\pi$ at times t_4, t_5 (see Figure 4.15). Consider a phase comparator whose output is stable at the points of …, -2π, 0, $+2\pi$, …. In the phase pull-in process, if the output jumps over the nearest phase-stabilized point without synchronizing to it, we say that a cycle slip has occurred. Cycle slips can also occur in the negative direction of the phase difference domain.

4.5.2 Influence of External Disturbances on Synchronization

This section examines the effect of changes in the phase and frequency of the input signal separately to simplify the explanation. Of course, such changes

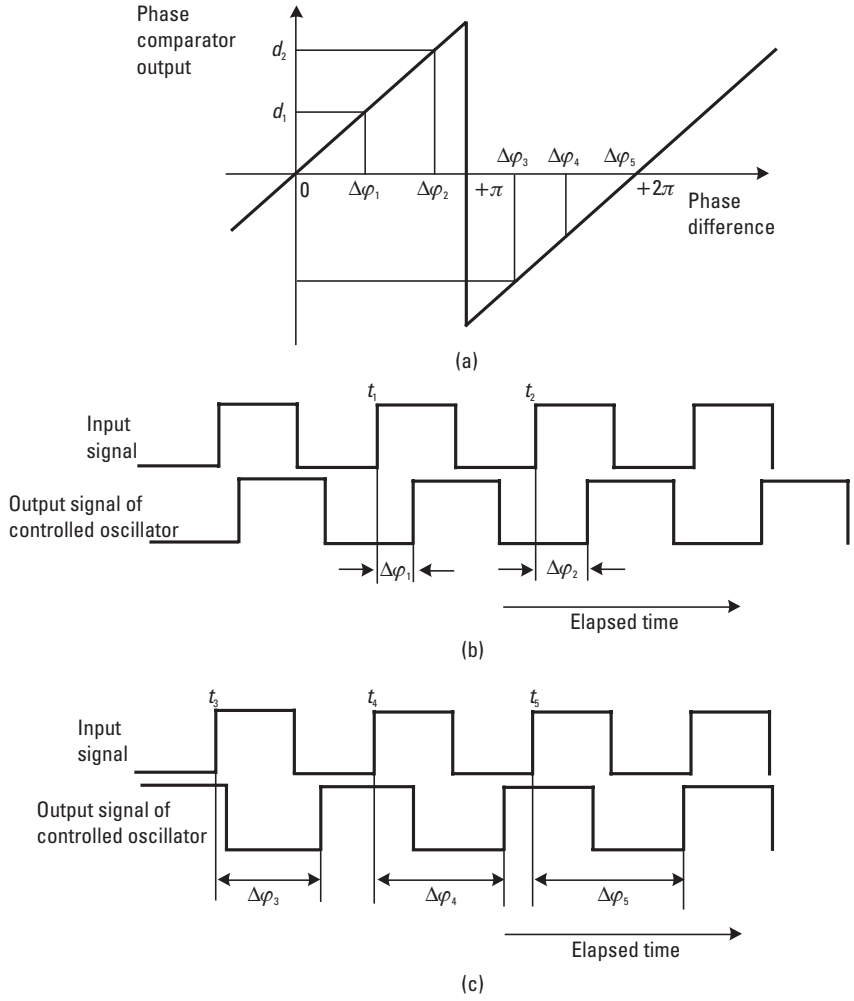


Figure 4.14 The process causing a cycle slip: (a) phase comparator characteristic, (b) phase difference change, and (c) phase difference change [continued from (b)].

can, and often do, occur together, but the basic response of the PLL is similar to what is described below.

4.5.2.1 Phase Jump

Once phase synchronization is achieved, the PLL tries to hold synchronization even if the input signal varies. If a phase jump $\Delta\varphi$ occurs in the input

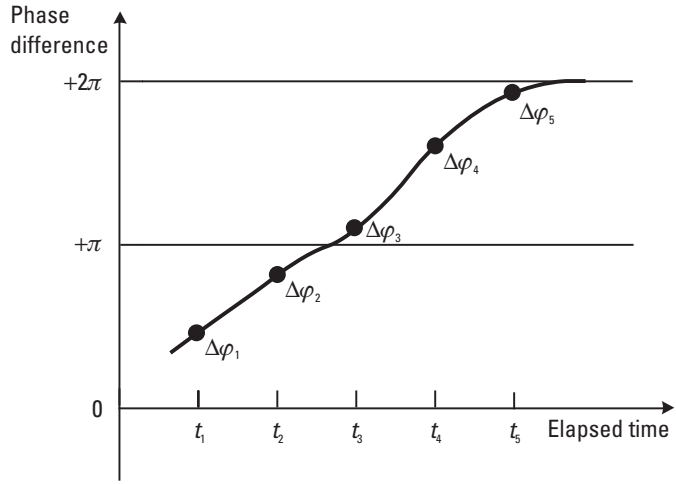


Figure 4.15 Phase difference change over time.

phase, φ_i at time t_1 in Figure 4.16, the phase difference information corresponding to $\Delta\varphi$ is generated at the output of the phase comparator. The PLL works to diminish the phase difference. The input phase and the output phase are depicted in Figure 4.17(a) where an external frequency standard having the same frequency as the input signal is used as a reference. The output of the PLL changes to track the input phase, and the amount of change is similar to the phase jump of the input signal. In this process, the output frequency deviates from the input frequency over time period a [see Figure 4.17(b)] to generate the phase change that corresponds to $\Delta\varphi$. This is equivalent to the phase synchronization process in which the frequency of the controlled oscillator is made to equal the frequency of the input; and this is a special case of the acquisition characteristics shown in Figures 4.2 and 4.3.

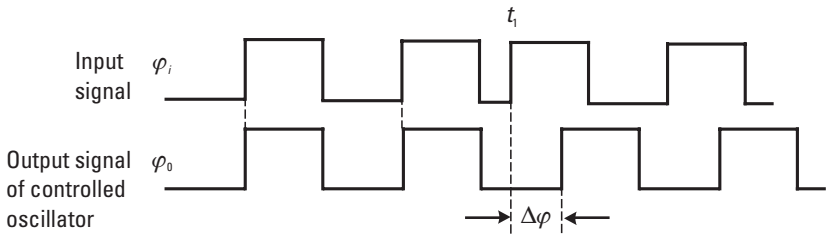


Figure 4.16 Phase jump in the input signal.

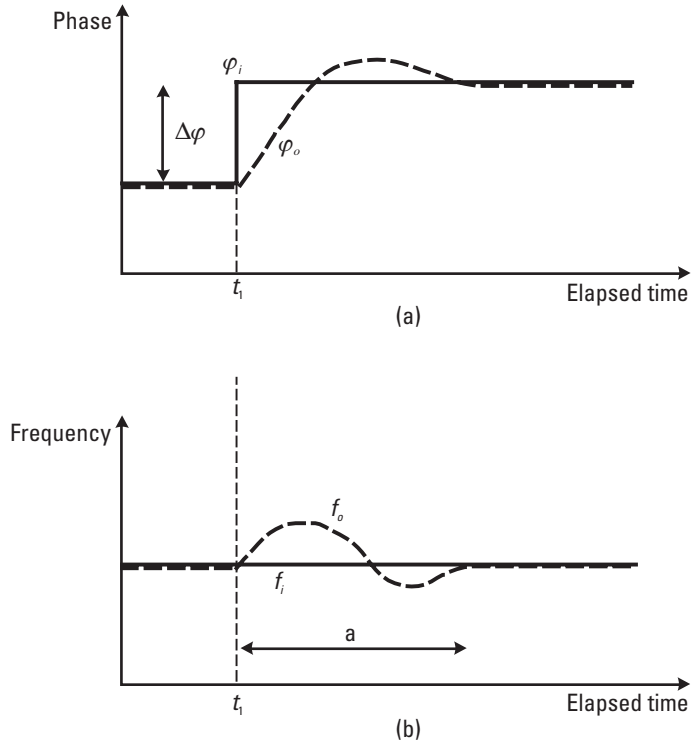


Figure 4.17 Phase difference and output frequency change of a controlled oscillator in response to a phase jump in the input signal: (a) input and output phase change, and (b) input and output frequency change.

Since Figure 4.16 shows a digital signal, the phase comparison is made at its leading or trailing edge. Although Figure 4.17 should strictly be plotted as a discrete graph due to this intermittent operation, continuous plots are shown to assist understanding.

4.5.2.2 Frequency Jump

Consider a frequency jump in which the frequency $1/T_1$ is changed to $1/T_2$ at time t_1 in a PLL that has been already synchronized, as illustrated in Figure 4.18. After the frequency jump has occurred, the phase difference gradually increases at the comparator output. Typical input and output phase changes are depicted in Figure 4.19(a). An external reference standard that has the same frequency as that before the frequency jump is used. If only a frequency offset occurs, the input phase change is linear over time

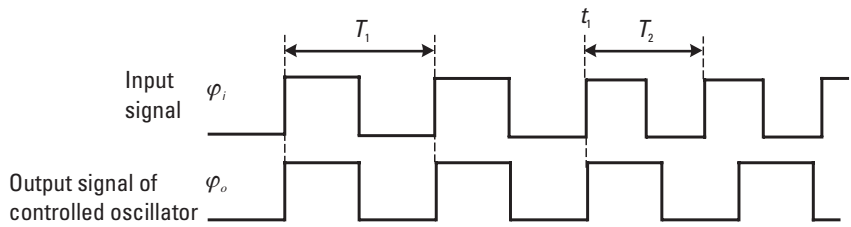


Figure 4.18 Frequency jump of the input signal.

after the jump [see the solid line in Figure 4.19(a)]. The output phase of the controlled oscillator tries to follow this input phase [the dotted line in Figure 4.19(a)]. The output frequency of the controlled oscillator then changes as shown in Figure 4.19(b). If input and output frequencies of the PLL are observed using an external standard, the input frequency is offset at time t_1 as shown in Figure 4.19(b), and the output frequency traces this offset and finally converges to it. The change in the phase difference between the input and the output signals, which eliminates any deviation in the reference, is shown in Figure 4.19(c). The phase difference converges to zero after the frequency jump.

4.5.3 Loop Gain

Since the PLL is synchronized by the feedback of phase difference information, its characteristics can be changed by altering the strength of the feedback signal. This is associated with the loop gain. It corresponds to the amount of feedback to the phase comparator and is determined by the originally detected phase difference information. Loop gain is an important parameter in that it determines the phase change characteristics. These include the time taken to acquire phase synchronization and the response to external disturbances.

The input phase change finally determines the output phase change as shown in Figure 4.20. In this process, loop gain determines the relationship between the output phase change $\Delta\varphi_o$ and the input phase change $\Delta\varphi_i$. This is similar to the amplification level of a simple amplifier. Increasing the loop gain raises the response speed of the PLL. Therefore, we can think of loop gain as representing the time constant of the PLL. Please remember that the loop gain of a PLL is not dimensionless. The input phase change alters the phase information at the output of the phase comparator. This information is transferred to the controlled oscillator and makes the output frequency

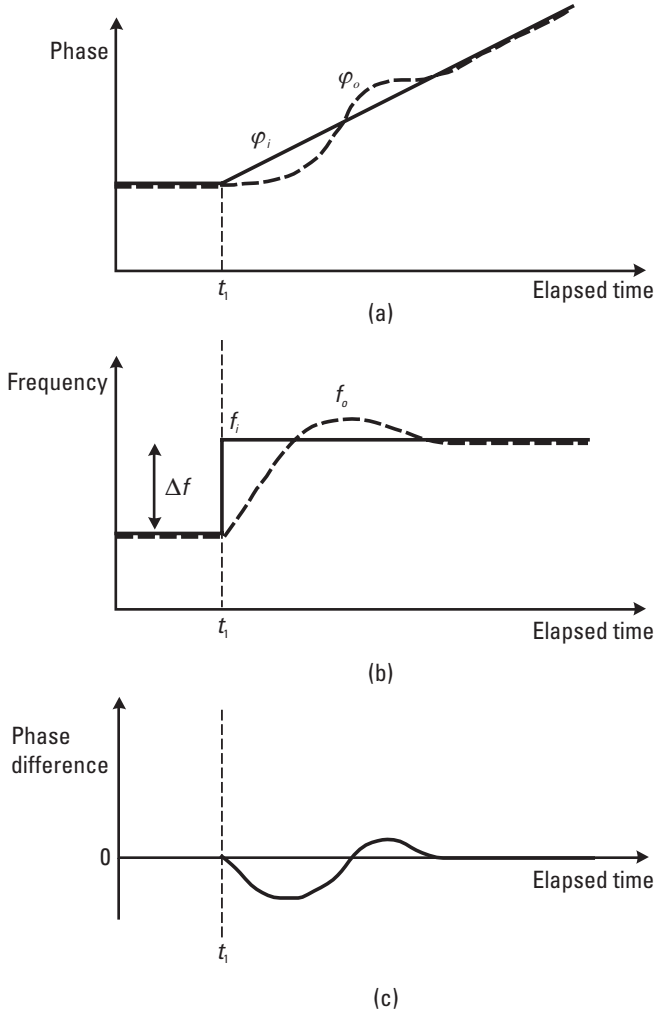


Figure 4.19 (a) Phase and (b) frequency change observed against an external standard. (c) Phase difference change of a phase comparator after a frequency jump.

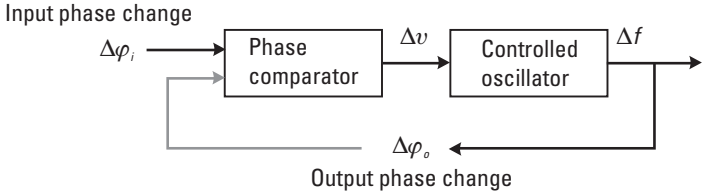


Figure 4.20 Relationship between input and output phase change and loop gain.

change, so the loop gain has the dimension of frequency or angular frequency.

4.5.4 Lowpass Characteristics

The ability of the PLL to track the input signal is determined by its frequency characteristics. When the input signal phase varies, the output phase of the PLL changes at the same rate as shown in Figure 4.21. If the variation rate is low as in Figure 4.21(a), the output phase tracks accurately the input phase. As the rate increases, some delay is created between the input and the output signals as shown in Figure 4.21(b), and the amount of phase change at the output gradually differs from that of the input. At much higher

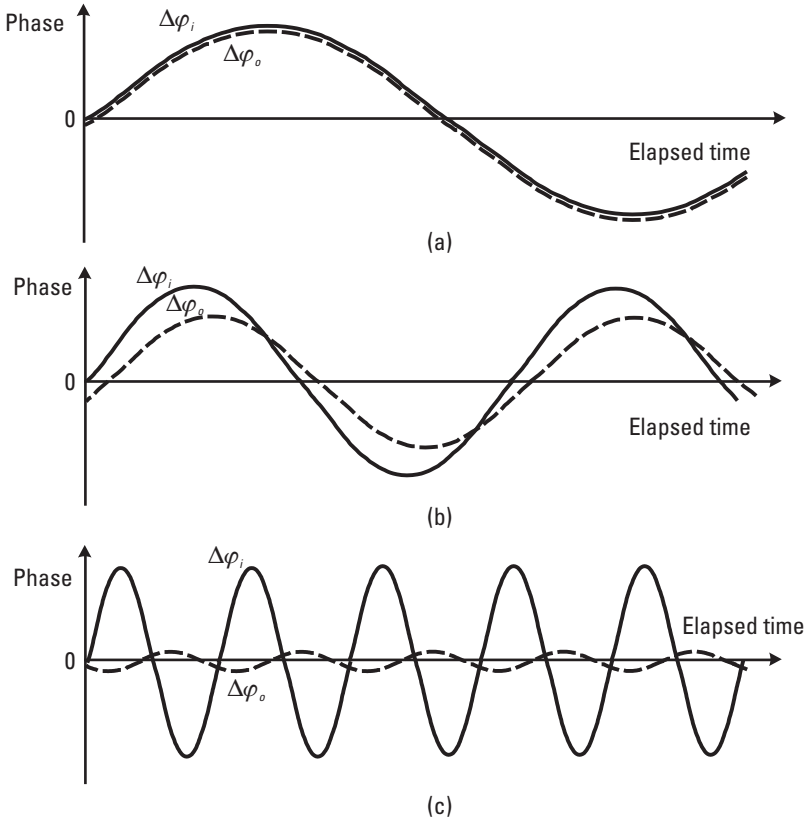


Figure 4.21 Output phase change for different variation frequencies: (a) low, (b) medium, and (c) high.

variation rates, the limited tracking capability of the PLL appears more clearly and the output phase changes only very slightly, as in Figure 4.21(c). These characteristics suggest that the PLL offers lowpass characteristics similar to those of a lowpass filter. Increasing the loop gain increases the cut-off frequency as shown in Figure 4.22.

The frequency of the phase change is expressed as deviation from the center frequency of the input signal. Figure 4.21 shows the case in which the center frequency of the input signal is modulated by the frequency of the phase change. The expression, "lowpass" indicates the frequency region close to this center frequency. It is better to think of the bandpass characteristics in the frequency region close to the center frequency of the input signal. We call these the bandpass (filter) characteristics. The bandpass characteristic of a PLL is given by the center frequency f_i and the cut-off frequency f_c. As shown in Figure 4.23(a), the frequency region that is passed by the PLL is $f_i \pm f_c$. If the bandpass characteristic is folded at the center frequency, you get Figure 4.23(b). The center frequency in Figure 4.23(a) becomes zero in Figure 4.23(b). While the variation signal is a sine wave in Figure 4.21, the same characteristics as Figure 4.23 are obtained by aperiodic signals. This is described in detail in Chapters 8 and 13.

As only frequencies within a certain band can pass through the PLL, there is a limit to the input phase change velocity that the circuit can track. If true phase synchronization is needed, the passband should be wide. On the other hand, if the input signal contains a lot of noise that should be suppressed, a narrow passband is better. It is obvious that there is a trade-off between phase tracking and noise reduction.

4.5.5 Steady-State Phase Error

The frequency control characteristic of the internal oscillator influences the final phase error in addition to the loop gain and the lowpass characteristics.

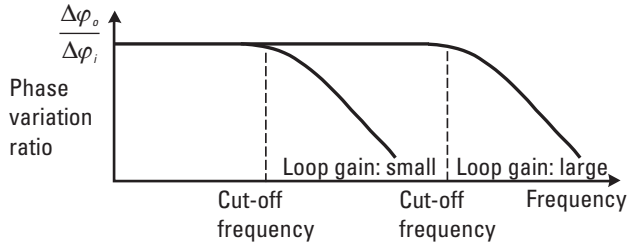


Figure 4.22 Relationship between the loop gain and cut-off frequency.

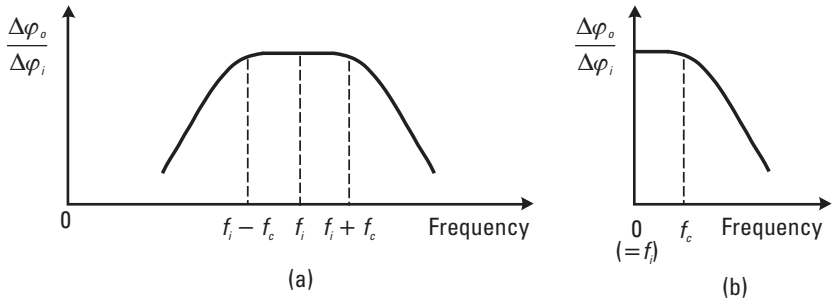


Figure 4.23 Cut-off frequency and bandpass characteristic: (a) actual PLL transfer characteristic and (b) PLL transfer characteristic with regard to PLL input phase variation frequency.

Consider an oscillator that has the frequency control characteristic of generating the same frequency, f_i, as the input frequency when the phase difference signal is zero as shown by plot (a) in Figure 4.24. In this plot, the output frequency equals the input frequency when the phase difference is zero. Plot (b) shows the effect of steady-state phase error; the output frequency is $f_i + \Delta f$ when the phase difference signal is zero. Plot (b) indicates that the free-running frequency of the PLL, the output frequency when there is no input

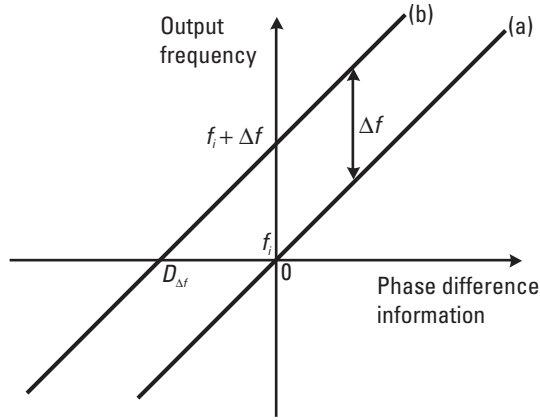


Figure 4.24 Frequency control characteristic of a controlled oscillator and the steady-state phase error: (a) frequency control characteristic of the internal oscillator when the free-running frequency of the internal oscillator is the same as the PLL input signal frequency and (b) frequency control characteristic of the internal oscillator when the free-running frequency of the PLL is offset from the design value.

signal, is offset from the design value. Since the output frequency of the PLL should equal the input frequency f_i in order to stabilize the phase difference, the phase difference should be $D_{\Delta f}$ [see plot (b)] when phase synchronization is established. After phase synchronization is established by compensating the free-running frequency in the feedback loop, the phase difference $D_{\Delta f}$ remains. This residual phase difference is called the steady-state phase error. The further the free-running frequency is from the input frequency, the larger the steady-state phase error becomes.

The free-running frequency of a controlled oscillator gradually shifts from its initial value over time. In crystal oscillators this is called aging, which may be highly nonlinear (e.g., logarithmic over time). This suggests that the amount of compensation performed in the feedback loop to achieve phase synchronization gradually increases over time. Note that this compensation can never be perfect due to limitations such as finite loop gain and delay in the feedback loop. Moreover, if the free-running frequency includes an unexpected frequency drift, the steady-state phase error may be increased. One key design factor is the accuracy with which the free-running frequency of the controlled oscillator can be initially set.

4.5.6 Synchronization Ranges

Whether the PLL can get synchronized at all or not depends on the initial frequency difference between the input signal and the output of the controlled oscillator. This implies that there is a synchronization range in terms of frequency. There are two categories: an acquisition range in which the synchronization process can be started, and a tracking range in which synchronization can be kept. The acquisition range can be classified into pull-in range (or capture range) in which cycle slips, described in Section 4.5.1, are generated, and lock-in range (or seize range) over which PLL synchronization can be completed within a time corresponding to the phase comparison period. The tracking range is also called the hold range and hold-in range. The lock range is used to express the range over which synchronization can happen within a time corresponding to the period of the beat signal generated by the frequency difference between the input signal and the free-running oscillator frequency. Pull-out range is used to express the range over which a PLL can normally lock again once synchronization is lost. We recommend that you check these definitions carefully before using them.

The pull-in range, lock-in range, and tracking range are used in this book. These three ranges are shown in Figure 4.25. The measure used to express range is the frequency difference between the input frequency of the

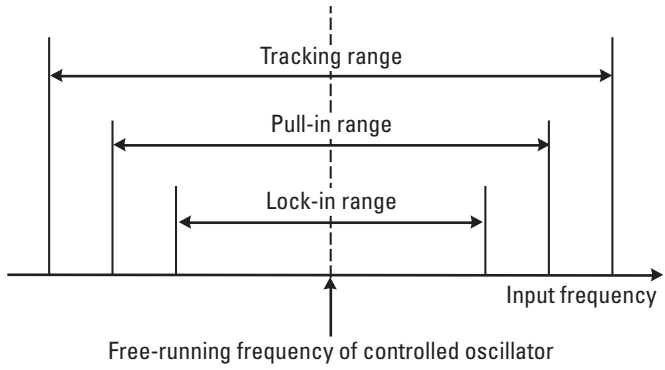


Figure 4.25 Definitions of various synchronization ranges in PLL circuits.

PLL and the free-running frequency of the controlled oscillator. To establish synchronization, this difference should be at least within the pull-in range. It is easily inferred that the lock-in range (synchronization without cycle slip) is narrower than the pull-in range. Once synchronization is established, it can be held up to the maximum frequency control range of the internal oscillator, provided the frequency difference changes slowly enough. This explains why the tracking range is generally wider than the other two ranges.

Reference

[1] Best, R. E., *Phase-Locked Loops*, New York: McGraw-Hill, 1997.

5

PLL Characteristic Analysis

5.1 Basic Analysis of PLL Characteristics

A clarified relationship between the input and output signals of a PLL is needed in order to analyze its characteristics quantitatively and in detail. This section describes the frequency and phase control mechanisms of a PLL, examines how it synchronizes to the input signal, and discusses the synchronization accuracy.

5.1.1 Issues Related to the Output Frequency

An understanding of the mechanism that determines the PLL output frequency is vital for elucidating its characteristics. The output frequency depends on the free-running frequency of the controlled oscillator and the frequency offset created by PLL operation, as shown in Figure 5.1. The free-running frequency is the frequency that is output by the controlled oscillator if no control signal is given to it. Although this frequency can be expected to be within a certain range, it is not guaranteed to equal the input frequency. It has to be corrected in a phase synchronization process. The frequency offset corresponds to the difference between the input frequency and the free-running frequency, and fit is output by the feedback circuit of the PLL. The frequency offset enables the output frequency of the PLL to be controlled in the phase synchronization process. The output frequency $f_o(t)$ is expressed as the sum of the free-running frequency $f_{free}(t)$ and the frequency offset $\Delta f_{contv}(t)$ that results from PLL control, as shown in (5.1).

$$f_o(t) = f_{free}(t) = \Delta f_{contv}(t) \tag{5.1}$$

In the analog PLL, the oscillator frequency is controlled by a voltage. The characteristic of the controlled oscillator can be taken as the relationship between input voltage and output frequency as shown in Figure 5.2. The controlled oscillator gain, which represents how much the output frequency changes when the input voltage changes by 1V, equals the gradient K_{of}. The frequency offset $\Delta f_{contv}(t)$ is defined by K_{of} and input voltage $v_{pd}(t)$. Since the free-running frequency is higher than the input frequency in the example of Figure 5.2, a negative voltage is required to move the frequency of the controlled oscillator closer to the input frequency. If there is no attenuation in the filter, placed at the output of the phase comparator, the input voltage to the controlled oscillator is the output voltage from the comparator.

Let us assume that the phase comparator has the characteristic in which its output voltage is proportional to the phase difference between the input signal phase and the output phase of the PLL as shown in Figure 5.3. The

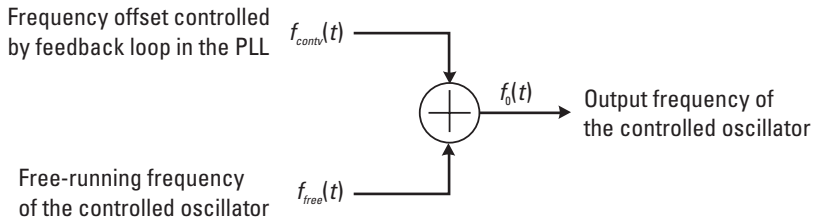


Figure 5.1 Output frequency decision mechanism.

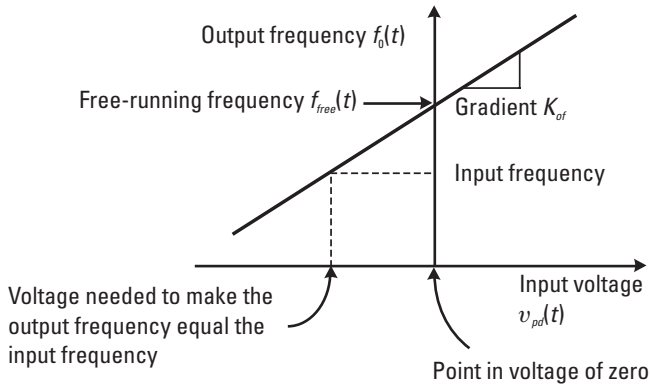


Figure 5.2 Frequency steering characteristic of an analog controlled oscillator.

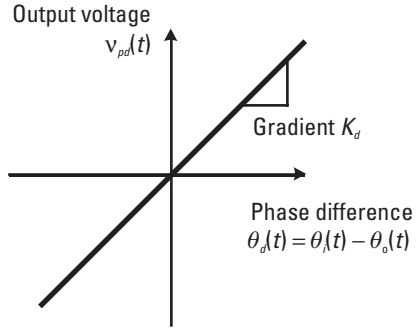


Figure 5.3 Phase comparator characteristic.

gain by which the phase difference is converted into voltage is given by the gradient K_d of the characteristic plot. The output voltage of the phase comparator can be expressed by the phase difference and K_d. Equation (5.1) can be transformed into that of (5.2) by assuming such a phase comparison characteristic:

$$\begin{aligned} f_o(t) &= f_{free}(t) + K_{of} \bullet v_{pd}(t) \\ &= f_{free}(t) + K_{of} \bullet K_d \theta_d(t) \\ &= f_{free}(t) + K_{of} \bullet K_d \left(\theta_i(t) - \theta_o(t)\right) \end{aligned} \tag{5.2}$$

5.1.2 Relationship Between Frequency and Phase

The PLL synchronization process involves both frequency and phase. First of all, to recast the approach of the PLL characteristic analysis, we consider the relationship between frequency and phase.

The signal phase advances with time. If the signal is a sine wave, the phase of the signal is given by:

$$\theta_o(t) = 2\pi f_o t \tag{5.3}$$

This phase change is shown in Figure 5.4 for two cases: high frequency and low frequency. The figure shows that the phase advances more rapidly at higher frequencies. Because the frequency determines the rate of phase change over time, the gradient of the phase change graph can be used to represent frequency yielding (5.4):

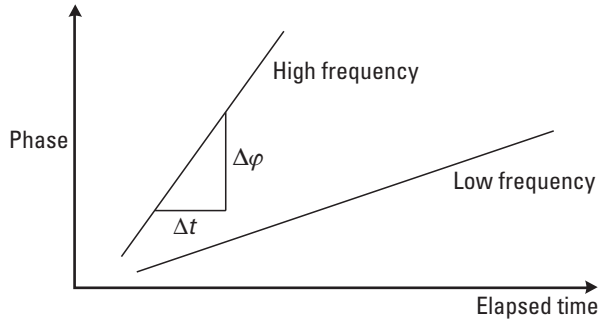


Figure 5.4 Signal phase change over time.

$$f_o = \frac{1}{2\pi}\frac{\Delta\varphi}{\Delta t} \tag{5.4}$$

The frequency can be obtained from (5.4) when the average frequency does not change. To generalize the relationship between phase and frequency, it is better to base the analysis on (5.5), since it yields a fractional frequency that can change over time.

$$f_o = \frac{1}{2\pi}\frac{d\theta_o(t)}{dt} \tag{5.5}$$

5.1.3 Relation Based on the Input and Output Phase Difference

The PLL has many characteristics specified by the signal phase and the phase difference between signals. The relationship between input signal phase and output signal phase is the most important. The control mechanism in the PLL can be taken as (5.6), which uses the phase difference between the input and output signals and the free-running frequency of the controlled oscillator:

$$\begin{aligned} \frac{1}{2\pi}\frac{d\theta_o(t)}{dt} &= f_{free}(t) + K_{of} \bullet K_d \bullet \theta_d(t) \\ \frac{d\theta_o(t)}{dt} &= \omega_{free}(t) + K_o \bullet K_d \bullet \theta_d(t) \end{aligned} \tag{5.6}$$

where $\omega_{free}(t)$ is the free-running angular frequency, and K_o is the angular frequency change ratio of the controlled oscillator.

The change over time in the input-output phase difference in the PLL synchronization process can be found by solving the differential (5.6). It uses the phase-frequency relationship clarified in Section 5.1.2. The influence of the frequency difference (angular frequency difference) between the free-running frequency of the controlled oscillator and the input frequency can be analyzed in terms of the input-output phase difference in the synchronization process.

5.2 Basic Characteristics [1–3]

5.2.1 Transfer Function

The steady-state characteristic in the frequency domain after phase synchronization is called a transfer function. It shows the amplitude changes of frequency components passing through the PLL. In general, the Laplace transform and its special case, the Fourier transform, are used for analysis in the frequency domain over continuous time. In particular, a differential equation can be converted into an algebraic one by using the Laplace transform because of its linearity and the conversion form for derivative functions. The first-order differential equation shown in (5.6) is converted by Laplace transformation into

$$s\Phi_o(s) = K_o \bullet K_d\left(\Phi_i(s) - \Phi_o(s)\right) \tag{5.7}$$

where $\Phi_o(s)$ and $\Phi_i(s)$ represent the Laplace transform of $\theta_o(t)$ and $\theta_i(t)$, respectively, and it is assumed that the free-running frequency of the controlled oscillator always equals the frequency of the input signal, and that $\omega_{free}(t) = 0$.

The ratio of the output phase change to the input phase change can be calculated as shown in (5.8), which is derived from that based on (5.7).

$$\frac{\Phi_o(s)}{\Phi_i(s)} = \frac{K_o \bullet K_d}{s + K_o \bullet K_d} \tag{5.8}$$

The response in the real frequency domain can be found by the Laplace transform (Fourier transform) along the $j\omega$ axis on the s plane by setting $s = j\omega$. After this, the response in (5.8) is becomes transfer function $H(j\omega)$ of

angular frequency ω and the amplitude or absolute value of $H(j\omega)$ showing the frequency response of the PLL can be obtained as

$$H(j\omega) = \frac{\Phi_o(j\omega)}{\Phi_i(j\omega)} = \frac{K_o \bullet K_d}{j\omega + K_o \bullet K_d}$$
$$|H(j\omega)| = \sqrt{\frac{(K_o \bullet K_d)^2}{\omega^2 + (K_o \bullet K_d)^2}} \tag{5.9}$$

This equation shows the frequency response of the PLL: how the output phase is influenced by the input phase. The frequency response $|H(j\omega)|$ when $\omega = 0$ is 1, and the input phase change is imposed on the output phase as it is. In the frequency region where the angular frequency is high, that is $\omega = \infty$, $|H(j\omega)|$ approaches 0, and the input phase change is not transferred to the output phase. The transfer function indicates a lowpass characteristic. The actual frequency of the PLL output, however, is not used in this analysis. $\omega = 0$ does not mean that the real frequency of the PLL output is 0. The frequency component in the variation of the output frequency is given by (5.9). The two frequencies, ω and ω_{real} are related as shown in (5.10):

$$\omega_{real} = \omega_o \pm |\omega| \tag{5.10}$$

The PLL actually represents a bandpass characteristic in the real frequency domain.

5.2.2 Steady-State Phase Error

The phase difference between the input and output as a function of time can be solved from (5.6). By Laplace transform and the results of Section 5.2.1, we get

$$s\Phi_d(s) = \Delta\Omega_{free-i}(s) - K_o \bullet K_d \bullet \Phi_d(s) \tag{5.11}$$

where $\Delta\Omega_{free-i}(s)$ and $\Phi_d(s)$ are the Laplace transforms of $\Delta\omega_{free-i}(t)$ and $\theta_d(t)$, respectively, and $\Delta\omega_{free-i}(t)$ is the angular frequency difference between the free-running angular frequency and the input angular frequency ($\Delta\omega_{free-i}(t) = \omega_{free}(t) - \omega_i(t)$).

It is assumed that $\Delta\omega_{free-i}(t)$ does not change within the time being considered. This represents the case where the free-running frequency of the

controlled oscillator has a constant offset against the input frequency of the PLL ($\Delta\omega = \Delta\omega_{free-i}(t)$).

$$s\Phi_d(s) = \frac{\Delta\omega}{s} - K_o \bullet K_d \bullet \Phi_d(s) \tag{5.12}$$

$$\Phi_d(s) = \frac{\Delta\omega}{s(s + K_o \bullet K_d)} \tag{5.13}$$

Equation (5.11) is converted into (5.12), and then rearranged to yield the phase difference as shown in (5.13). Because the steady-state phase error indicates the phase difference in the stationary state, as shown by its name, it is necessary to obtain $\Phi_d(s)$ when time is infinite. The Laplace transform exists if $\lim_{t\to\infty} \theta_d(t) = 0$.

$$\lim_{s\to 0} s\Phi_d(s) = \lim_{t\to\infty} \theta_d(t) = \theta_e \tag{5.14}$$

This is called the final value theorem [4].

When the relationship of (5.14) is applied to (5.13), the steady-state phase error becomes

$$\theta_d = \frac{\Delta\omega}{K_o \bullet K_d} \tag{5.15}$$

Equation (5.15) means that increasing the loop gain, $K_o \bullet K_d$, decreases the steady-state phase error. When the free-running frequency of the controlled oscillator has a constant offset against the PLL input frequency, the steady-state phase error reaches a constant value as well.

5.2.3 Synchronization Range

5.2.3.1 Tracking Range

The tracking range (i.e., the frequency range in which synchronization can be held after once established) is determined by two factors. One is the frequency control range of the internal oscillator. It determines whether the PLL can follow the input frequency even if the free-running frequency of the internal oscillator wanders from the input frequency. Of course, the frequency control range should basically be wider than the expected largest frequency difference between the input frequency and the free-running

frequency. The other factor is whether the control signal, which forces the controlled oscillator to output the required frequency, can be generated.

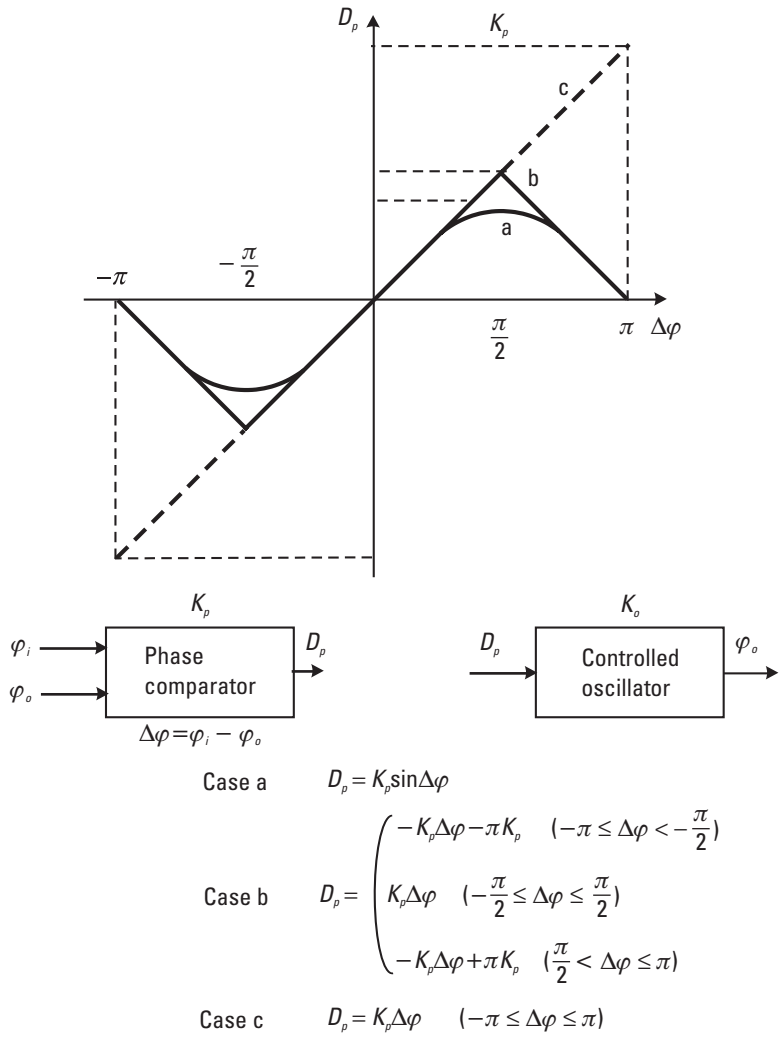
Assume that the frequency control range is wide enough, and that the frequency difference between the input frequency and the free-running frequency of the controlled oscillator changes slowly. In this case the tracking range corresponds to the frequency of the controlled oscillator output generated by the maximum control signal from the phase comparator output. If the maximum frequency difference that permits synchronization to be maintained is expressed by a phase change $\Delta\omega_{max}$, the maximum phase gradient of the controlled oscillator over time has to correspond to this $\Delta\omega_{max}$ at the maximum output of the phase comparator. For the phase comparison characteristics of a, b, and c in Figure 5.5, the output signal *Dp* becomes maximum when the phase difference $\Delta\varphi$ is $\pi/2$ for plots a and b, and when the phase difference is π in plot c. When the phase difference detection gain of the phase comparator is K_p, the maximum output $K_{p\max}$ is given by

$$\begin{aligned} K_{p\max} &= K_p && \text{for a} \\ K_{p\max} &= K_p \bullet \frac{\pi}{2} && \text{for b} \\ K_{p\max} &= K_p \bullet \pi && \text{for c} \end{aligned} \tag{5.16}$$

This output is passed to the controlled oscillator, and the maximum phase change $\Delta\varphi_{max}$ is generated. Therefore, $\Delta\omega_{max}$ is given by

$$\begin{aligned} \Delta\omega_{max} &= K_p K_o && \text{for a} \\ \Delta\omega_{max} &= K_p K_o \bullet \frac{\pi}{2} && \text{for b} \\ \Delta\omega_{max} &= K_p K_o \bullet \pi && \text{for c} \end{aligned} \tag{5.17}$$

This implies that the maximum phase change $\Delta\omega_{max}$ is calculated from (5.17) as a function of the loop gain. If K_p is given in [V/rad] and K_o in [rad/sec/V], $\Delta\omega_{max}$ is obtained in [rad/sec]. K_p, however, might be defined in [V] for curve a. In this case, it is necessary to note that this book treats the system of units differently from other texts (units are explained in Appendix A). Equation (5.17) indicates that the tracking range gets wider if the maximum output of the phase comparator increases. However, if the frequency demanded by the phase comparator output exceeds the frequency control range of the controlled oscillator, the PLL cannot synchronize to the input signal. Since the comparator output can be easily amplified, the practical



Case a $D_p = K_p \sin\Delta\varphi$

Case b
$$D_p = \begin{cases} -K_p\Delta\varphi - \pi K_p & (-\pi \le \Delta\varphi < -\frac{\pi}{2}) \\ K_p\Delta\varphi & (-\frac{\pi}{2} \le \Delta\varphi \le \frac{\pi}{2}) \\ -K_p\Delta\varphi + \pi K_p & (\frac{\pi}{2} < \Delta\varphi \le \pi) \end{cases}$$

Case c $D_p = K_p\Delta\varphi \quad (-\pi \le \Delta\varphi \le \pi)$

Figure 5.5 Phase comparator characteristic: (a) mixer type, (b) triangle type, and (c) sawtooth type.

tracking range is usually determined by the tuning range of the controlled oscillator.

5.2.3.2 Pull-In Range [4]

First of all, it is necessary to be able to synchronize to the input signal from the initial state so that the PLL synchronization process can run to

completion. The frequency range, which indicates whether the PLL can achieve synchronization from the initial state, is called the pull-in range. When the nominal value of the input frequency (to which the PLL has to synchronize) widely varies, the pull-in range is one of the most important PLL characteristics.

The pull-in range is determined by how close the frequency of the controlled oscillator can be made to approach the input frequency in the initial asynchronous state. Here, a small DC component is generated from the periodic signal that is observed at the output of the phase comparator. If this component can be added to the controlled oscillator, the PLL synchronization process advances towards completion. Figure 5.6 shows the signal state of the PLL when a large initial frequency difference exists between the input and the internal oscillator. The comparator produces a distorted signal whose cycle corresponds to the frequency difference. This signal is not a pure sine wave and includes some DC component, which is extracted by the lowpass filter preceding the controlled oscillator. If the system gain is now sufficient to make the frequency difference vanish, the synchronization is advanced.

The PLL pull-in range varies according to the filter characteristics. A first-order loop PLL without any filter (as described in Section 5.3.1) can synchronize at frequencies determined by the loop gain. The pull-in range $\Delta\omega_{pull}$ is given by

$$\Delta\omega_{pull} = K_o \bullet K_d \tag{5.18}$$

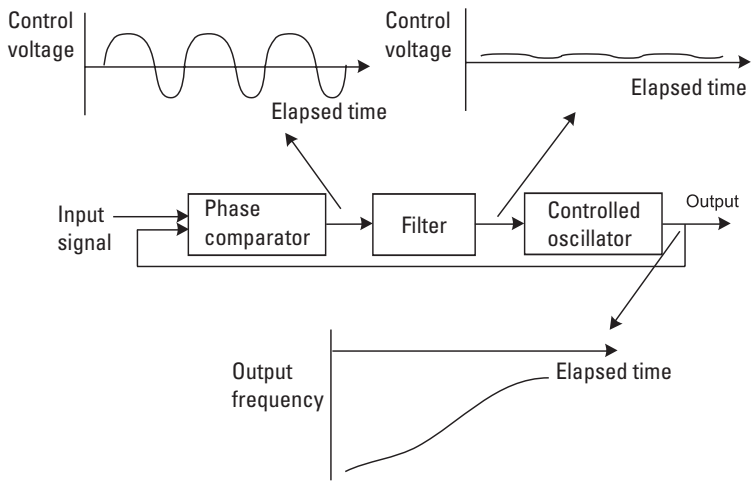


Figure 5.6 Initial pull-in process of a PLL.

In a second-order loop PLL (described in Section 5.4.1), the pull-in range $\Delta\omega_{pull}$ depends on the filter configuration. If passive filters are used, the coefficients of the filter (as discussed in Section 5.4.1) are included in $\Delta\omega_{pull}$ as shown in (5.19), since the loop gain in the initial state is determined by the filter gain at high frequencies.

$$\Delta\omega_{pull} = K_o \bullet K_d \sqrt{\frac{2\tau_2}{\tau_1 + \tau_2}} \tag{5.19}$$

This equation is valid if the value within the square root is much smaller than 1.

A PLL with an active filter (as described in Section 5.4.1) can synchronize at a maximum frequency determined by the maximum input level of the controlled oscillator, since even a very small DC output from the comparator can be integrated. Therefore, the pull-in range $\Delta\omega_{pull}$ becomes infinite.

$$\Delta\omega_{pull} = \infty \tag{5.20}$$

Of course, particularly here the oscillation limits of the controlled oscillator determine the final maximum pull-in range.

5.3 Transfer Function and Response Characteristic of the First-Order Loop PLL [1, 4]

5.3.1 Transfer Function in an Analog Circuit

The PLL transfer function can be used to evaluate the correlation between phase changes in the output and input. In Section 5.2.1, the transfer function was obtained for the simplest PLL, which consists of just a phase comparator and a controlled oscillator. This configuration represents a first-order loop since it has a linear function of *s*. The transfer function of the first-order loop is the ratio of the phase variation transferred from the input of the PLL to the output. In (5.21), the level of the phase variation ratio is obtained by calculating the absolute value of the transfer function.

$$\left|H(j\omega)\right| = \sqrt{\frac{(K_o \bullet K_d)^2}{\omega^2 + (K_o \bullet K_d)^2}} \tag{5.21}$$

In (5.21), the phase variation ratio approaches unity as ω approaches 0. Conversely, as ω is enlarged, the phase variation ratio becomes zero. It is shown that the PLL has a lowpass characteristic (see Figure 5.7), as only slow phase variations appear in the PLL output. The amplitude $|H(j\omega)|$ becomes $\sqrt{1/2}$ or drops by –3 dB when the angular frequency ω equals the loop gain $K_o \bullet K_d$. This can be considered as the PLL passband limit. In the frequency region where ω is sufficiently larger than the loop gain, the phase variation ratio decreases as ω increases. The rate of this decrease is –20 dB/decade, which is equivalent to –6 dB/octave.

5.3.2 Transfer Function in Digital PLLs

5.3.2.1 Transfer Function Without Considering Quantization

Let us study the digital PLL shown in Figure 5.8. Its transfer function is obtained by using the phase-time difference in seconds instead of radians. It is assumed that the phase comparator (PD) measures the phase-time difference and outputs the result as a digital value. There are advantages in utilizing the phase-time difference. The influence of the input and output frequencies is eliminated, and such units are suitable for practical measurements with, for example, oscilloscopes.

The output frequency of the controlled oscillator (DCO) is assumed to be changed by the digital value coming from the comparator. In practice, these values are quantized, but here, however, they are assumed to be continuous. This enables digital PLLs to be analyzed with the same analytical method as their analog counterparts.

The phase comparator characteristic is shown in Figure 5.9. Digital values are output according to the phase-time difference between the input phase of the PLL and the output phase of the controlled oscillator. In an

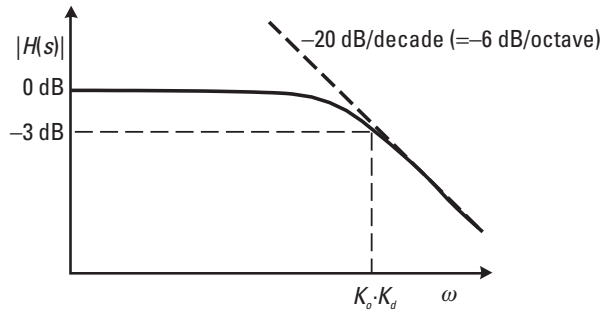


Figure 5.7 The frequency response of a first-order loop transfer function.

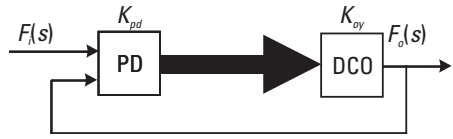


Figure 5.8 The basic digital PLL configuration.

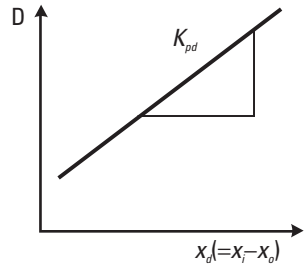


Figure 5.9 Phase comparator output as a function of phase-time difference.

actual phase comparator the values are generated at discrete time intervals based on the sampling period, but here continuous measurement is assumed. The change in phase comparator output is given by (5.22) using the phase-time difference measurement ratio K_{pd}.

$$D(t) = K_{pd} \bullet x_d(t) \tag{5.22}$$

Figure 5.10 shows the characteristic of the controlled oscillator. The output frequency usually varies within a minimal range compared to its nominal (center) value. We, therefore, prefer to use the relative frequency error *y*, which designates the normalized instantaneous frequency offset from nominal value f_0 as shown here:

$$y = \frac{f - f_0}{f_0} = \frac{\Delta f}{f_0} \tag{5.23}$$

where Δf is the difference between the nominal frequency f_0 and the output frequency *f*. The frequency deviation of the controlled oscillator output is determined by the control coefficient K_{oy} of the oscillator and by the phase information from the phase comparator. In addition, the controlled

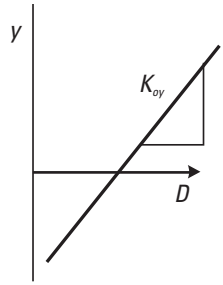


Figure 5.10 The controlled oscillator tuning characteristic.

oscillator output $x_o(t)$ can be related to the phase-time difference information $D(t)$ from the phase comparator based on the relationship between the frequency deviation and the phase-time in the controlled oscillator output.

$$\frac{dx_o(t)}{dt} = K_{oy} \bullet D(t) \tag{5.24}$$

The transfer function for the input and output phase of the digital PLL in units of phase-time can be obtained by (5.22) and (5.24).

$$H(s) = \frac{X_o(s)}{X_i(s)} = \frac{K_{pd} \bullet K_{oy}}{s + K_{pd} \bullet K_{oy}} \tag{5.25}$$

The phase variation ratio transferred from the input of the digital PLL to the output, calculated using (5.25), is equal to that given by (5.21). The transfer function in (5.25), however, should be used for a digital PLL, since it is in a general form by using phase-time x and frequency deviation y. Equation (5.25) is independent of oscillator and phase comparison frequencies.

5.3.2.2 Transfer Function in Multiple Level Quantization

When the phase-time difference of the phase comparator and the frequency control of the controlled oscillator are quantized with multiple levels and processed as digital values, the quantization resolution and the sampling frequency influence the transfer function. The internal information of the PLL should be treated in the discrete time domain; however, when quantization sampling is fast enough compared to the PLL operation (time constant), the

characteristics can be analyzed in the continuous time domain. This allows us to use the same equation expansion as in Section 5.3.2.1.

The phase-time difference detection resolution, which depends on the quantization steps of the phase comparator, influences the loop gain of the PLL. The size of the digital output for the phase-time difference is reflected in the loop gain as a coefficient of the phase comparator characteristic. The sensitivity of the comparator determines the output amplitude for a phase change. One step in the digital output value, which is an inverse number of the phase-time difference detection resolution, is thus K_{pd} in (5.9) as shown by

$$K_{pd} = \frac{\text{One step in a digital value at an output}}{\text{Phase-time detection resolution}} = \frac{1}{\text{Phase-time detection resolution}} \tag{5.26}$$

The oscillator frequency changes according to the digital value from the phase comparator. The frequency change for one digital step influences the loop gain through the coefficient of the controlled oscillator. Therefore, the frequency control resolution of the oscillator is the gradient of the line in Figure 5.10. K_{oy} is given by

$$K_{oy} = \frac{\text{Frequency deviation control step}}{\text{One step in a digital value at an input}} = \text{Frequency deviation control step} \tag{5.27}$$

The transfer function can be obtained by using these coefficients. The passband, within which the phase variation ratio of the digital PLL is –3 dB, is calculated from the frequency control resolution of the oscillator, Δy_{step}, and the phase-time difference detection resolution, ΔD_{step}, as shown in (5.28).

$$\omega_{cutoff} = \frac{\Delta y_{step}}{\Delta D_{step}} \tag{5.28}$$

This equation suggests that these two coefficients, both of which indicate quantization fineness, have opposite effects on the passband (loop gain). The better the phase-time difference resolution is, the wider the passband of

a digital PLL. The passband gets narrower as the frequency control resolution of the oscillator improves.

5.3.3 Phase and Frequency Response of the First-Order Loop

The PLL transfer function basically describes the phase variation in the steady state. In order to understand the loop response when both phase and frequency change, it is necessary to study transient phenomena as well. We have to solve the fundamental differential equation to get the time function, which gives the PLL's response to input phase-time and frequency jumps. The transient response is usually analyzed by examining the phase error at the comparator output. PLL output responses, observed by using an external reference signal, are shown here.

5.3.3.1 Phase-Time Step Response

The PLL output phase change over time is expressed by (5.29) when a phase-time step occurs at the input. The output phase-time, which does not change at $t = 0$, approaches the input phase-time step over time.

$$T_o = \Delta T\left(1 - e^{-Kt}\right) \tag{5.29}$$

Figure 5.11 plots (5.29). The larger the loop gain is, the faster the phase-time of the PLL output approaches that of the input. On the other hand, the process in a digitally controlled (discrete operation) PLL is shown in Figure 5.12. At time $t = 0$, the phase difference detected by the input phase comparator is equal to the phase-time step that is input [see point A in Figure 5.12(b)]. To cancel this phase difference, the oscillator frequency is controlled.

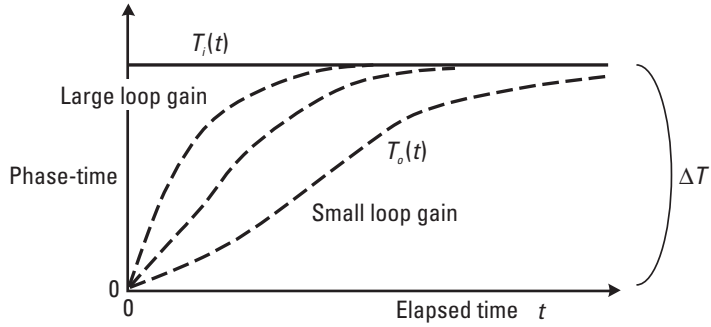


Figure 5.11 Output phase change over time after a stepwise phase-time jump.

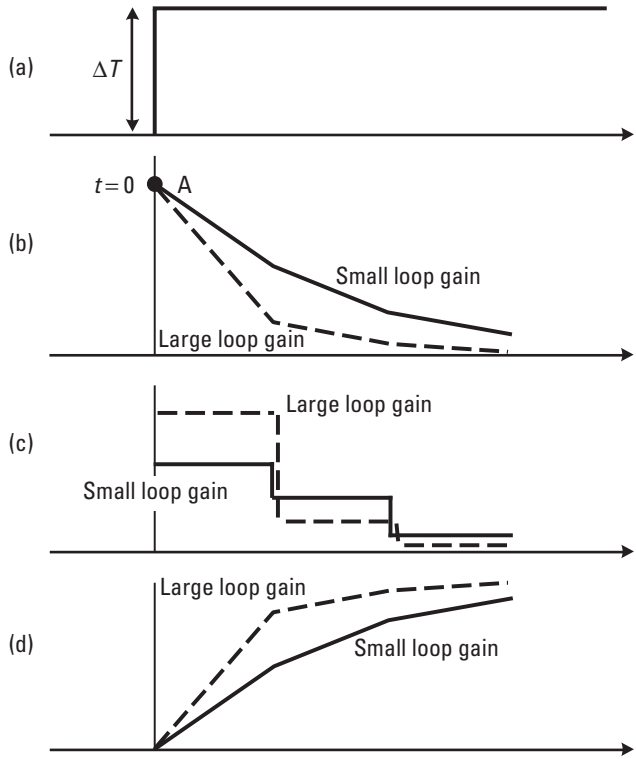


Figure 5.12 Digital PLL step response as the output phase change over time: (a) input phase-time change, (b) phase-time difference between input and output signals, (c) frequency difference between input and output signals, and (d) output phase-time change.

The output phase is changed by moving the controlled oscillator frequency far from the input frequency. The phase difference gradually decreases. At this time, increasing the frequency deviation of the controlled oscillator increases the output phase change [see the dotted line in Figure 5.12(c, d)]. This means that phase tracking is faster. The frequency change at the oscillator output can be increased by raising the frequency control sensitivity of the oscillator. The other way of realizing faster tracking is to improve the phase-time difference resolution of the phase comparator, and then to raise the sensitivity of the phase comparator. This allows very small phase changes to be detected, and prompt actions by the controlled oscillator will increase phase tracing speed.

In Figures 5.11 and 5.12(c, d) the PLL output is observed by using an external standard signal as a reference. These expressions correspond to the

case where the input and the output phase changes of the PLL are individually observed against an external standard with measurement equipment such as an oscilloscope. On the other hand, it is also possible to analyze the PLL through the phase difference between the input and output as shown in Figure 5.12(b). This corresponds to the observation of the internal response in the PLL. At the moment when this phase difference disappears, synchronization is established.

5.3.3.2 Frequency Step Response

When a frequency step occurs in the input signal, the input phase (expressed in terms of phase-time from an external standard) changes as shown in (5.30) if the original input frequency and phase equal those of the external reference.

$$T_i(t) = \frac{\Delta f}{f_o} t \tag{5.30}$$

The respective output phase-time change can be expressed using

$$T_o(t) = -\frac{\Delta f}{f_o}\frac{1}{K} + \frac{\Delta f}{f_o}\frac{1}{K}e^{-Kt} + \frac{\Delta f}{f_o}t \tag{5.31}$$

The first term on the right-hand side of (5.31) is the steady-state phase error caused by the frequency difference, which comes from the frequency step. It is measured between the input frequency and the free-running frequency of the controlled oscillator. The second term is the transient response in phase-time caused by the step. This term becomes small over time. The third term is the phase-time change caused by the frequency step. Since the output phase-time is expressed relative to the external standard, the phase-time change, which equals the input signal phase change shown in (5.30), is included as the third term. The output phase-time change over time is shown in Figure 5.13. If the output phase-time change of the PLL is measured without using an external standard, the third term does not appear. The input and output phase-time difference change over time can be expressed using

$$T_i(t) - T_o(t) = -\frac{\Delta f}{f_o}\frac{1}{K} + \frac{\Delta f}{f_o}\frac{1}{K}e^{-Kt} \tag{5.32}$$

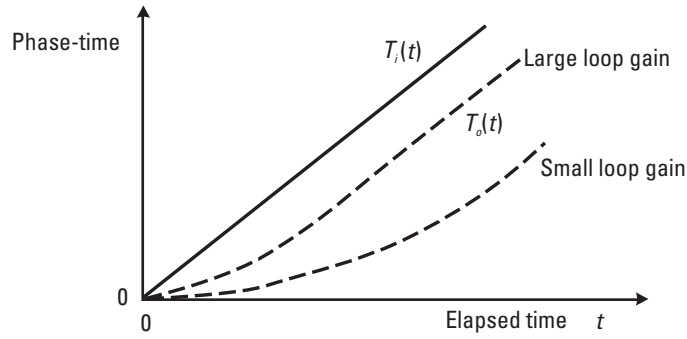


Figure 5.13 PLL output step response as a phase change over time.

Only the first term on the right-hand side of (5.32), which corresponds to steady-state phase error, remains after phase-time stabilizes. If the loop gain becomes small, the difference between phase-time before and after the frequency step increases. It is assumed here that the free-running frequency of the controlled oscillator is the same as the input frequency before the frequency step. If there already is an initial frequency difference between the input and output signals before the frequency step, such that steady-state phase error is present, it is necessary to add this to the frequency step in (5.32).

5.4 Transfer Function and Second-Order Loop Response [1, 4]

5.4.1 Transfer Function in Analog PLLs

The first-order loop has some drawbacks. If the PLL lowpass characteristic should suppress noise in the input signal, it is necessary to reduce the loop gain and to lower the cutoff frequency. However, the loop gain must be high to improve the response speed and the frequency synchronization range. These two requirements conflict and it is difficult to satisfy both at the same time in the first-order loop. Moreover, the first-order loop also generates a significant steady-state phase error if the input differs from the free-running frequency of the controlled oscillator. A filter circuit between the phase comparator and the controlled oscillator improves the situation. Three filter types are used (see Figure 5.14): the lag filter, the lag-lead filter, and the complete integral filter. The lag filter can be considered as a variant of the lag-lead filter. The lag-lead filter is a passive filter, while the complete integral filter is an

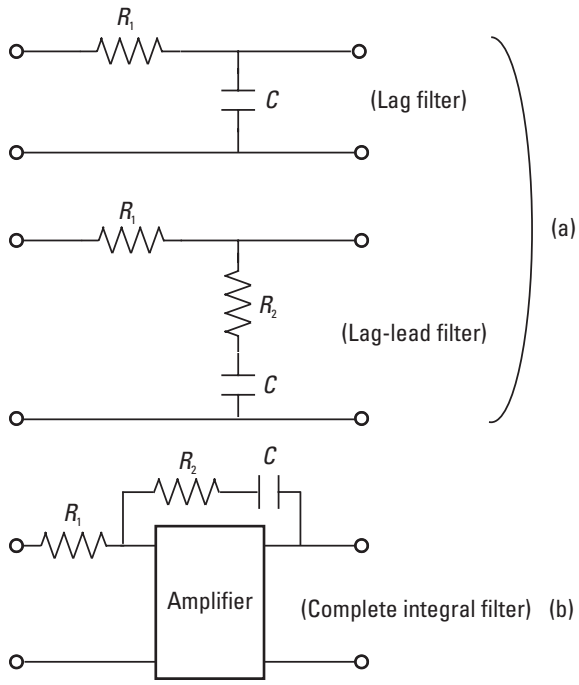


Figure 5.14 Analog filters used in PLLs: (a) passive filter and (b) active filter.

active filter. Transfer functions for the passive and active filters are given by (5.33) and (5.34), respectively.

$$\text{Passive filter: } F(s) = \frac{1 + s\tau_2}{1 + s(\tau_1 + \tau_2)} \tag{5.33}$$

where $\tau_1 = R_1 C,\ \tau_2 = R_2 C$.

$$\text{Active filter: } F(s) = \frac{R_2}{R_1} + \frac{1}{s\tau_1} = \frac{1 + s\tau_2}{s\tau_1} \tag{5.34}$$

where $\tau_1 = R_1 C,\ \tau_2 = R_2 C$, and the amplifier gain is assumed to be infinite.

Magnitude plots of these transfer functions are shown in Figure 5.15. Both passive filters have the same gain at low frequencies, but different characteristics at high frequencies. The gain of the active filter (ideal integrator)

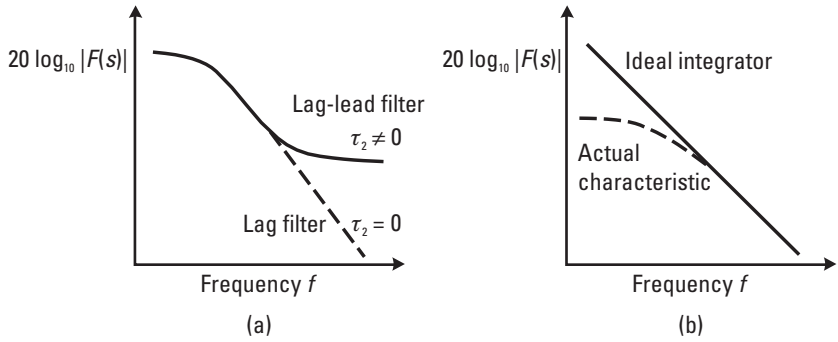


Figure 5.15 Magnitude of analog filter transfer functions in the frequency domain: (a) passive filter and (b) active filter.

continues to increase as the frequency falls. Actual gain plots saturate at low frequencies since a real amplifier has finite DC gain. In any case, the active filter offers much higher output magnitudes than passive filters. The total transfer function of the PLL using these filters is given by (5.35) and (5.36).

$$\text{Passive type: } H(s) = \frac{KF(s)}{s + KF(s)} = \frac{K + K\tau_2 s}{K + (1 + K\tau_2)s + (\tau_1 + \tau_2)s^2} \tag{5.35}$$

$$\text{Active type: } H(s) = \frac{KF(s)}{s + KF(s)} = \frac{K + K\tau_2 s}{K + K\tau_2 s + \tau_1 s^2} \tag{5.36}$$

where $K = K_o \bullet K_d$.

PLLs with these transfer functions are called second-order since the term with highest order has degree two in (5.35) and (5.36). These equations can be also converted as follows:

$$\text{Passive filter: } H(s) = \frac{\omega_n \left(2\xi - \frac{\omega_n}{K}\right)s + \omega_n^2}{s^2 + 2\xi\omega_n s + \omega_n^2} \tag{5.37}$$

$$\text{where } \omega_n = \sqrt{\frac{K}{\tau_1 + \tau_2}}, \xi = \frac{1}{2}\sqrt{\frac{K}{\tau_1 + \tau_2}}\left(\tau_2 + \frac{1}{K}\right)$$

$$\text{Active filter: } H(s) = \frac{2\zeta\omega_n s + \omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \tag{5.38}$$

where $\omega_n = \sqrt{\frac{K}{\tau_1}}, \zeta = \frac{\tau_2}{2}\sqrt{\frac{K}{\tau_1}}$.

Coefficients τ_1, τ_2, and K are replaced by the natural angular frequency ω_n (which has the dimension of angular frequency) and a dimensionless damping factor ζ. These parameters are advantageous in expressing the transient response characteristics of second-order PLLs and simplify the analysis. For the active filter, the angular frequency of the transfer function is normalized, and the characteristic can be analyzed with just ζ as a parameter.

$$H(j\omega) = \frac{1 + j2\zeta(\omega/\omega_n)}{1 - (\omega/\omega_n)^2 + j2\zeta(\omega/\omega_n)} \tag{5.39}$$

This equation is plotted in Figure 5.16. The magnitude characteristic of the transfer function in the second-order PLL does not change monotonously, and it peaks at certain frequencies. The magnitude of a first-order PLL never exceeds 0 dB as shown in Figure 5.7. In second-order PLLs, however, this is possible within some frequency region. This characteristic can degrade the total performance if several PLLs are connected in series. If unnecessary phase variation is added at the first stage of a PLL cascade, this variation can be amplified link by link. This phenomenon is described in Chapter 7.

A PLL using a passive filter exhibits the same characteristic. Its gain peaks in some frequency region. When the inequality of (5.40) holds, active and passive filters offer much the same characteristics.

$$\tau_2 K >> 1 \tag{5.40}$$

5.4.2 Transfer Function in Digital PLLs

The desired filter property can be directly realized by digital circuits using the method described in Chapter 3. Let us start by considering the transfer function of the digital PLL where the analog filter described in Section 5.4.1 is converted into a digital one. This is accomplished by using suitable digital

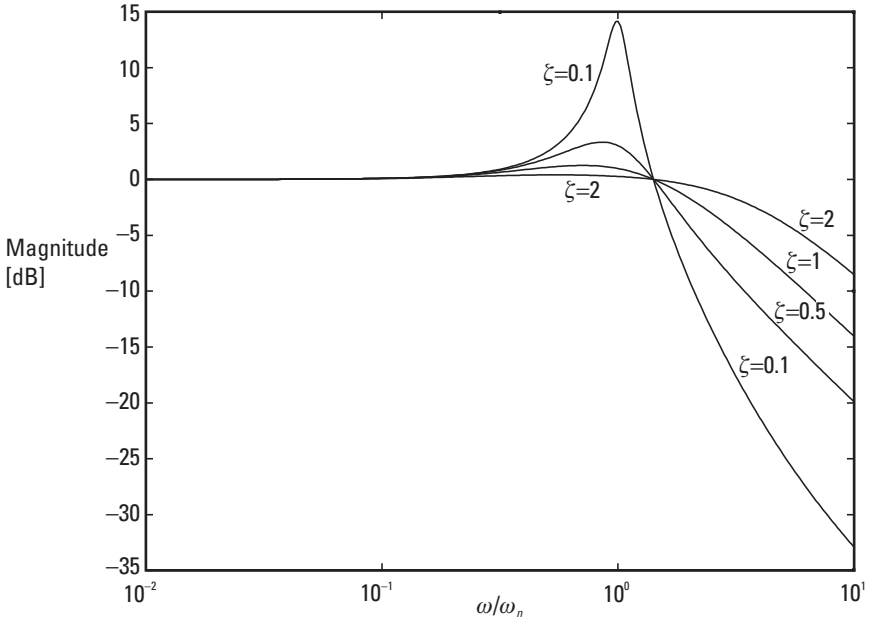


Figure 5.16 Transfer function of a second-order PLL with an active filter.

processing components such as adders, multipliers, and unit delay elements. The IIR filter is one such converted filter. For example, we can have

$$F(z) = a\frac{1 + b \bullet z^{-1}}{1 - c \bullet z^{-1}} \tag{5.41}$$

where

$$\text{passive: } a = \frac{T + 2\tau_2}{T + 2\tau_1 + 2\tau_2}, b = \frac{T - 2\tau_2}{T + 2\tau_2}, \text{ and } c = \frac{-T + 2\tau_1 + 2\tau_2}{T + 2\tau_1 + 2\tau_2}$$

$$\text{active: } a = \frac{T + 2\tau_2}{2\tau_1}, b = \frac{T - 2\tau_2}{T + 2\tau_2}, \text{ and } c = 1$$

Both passive and active filters have the same structure in (5.41). The actual realization of this transfer function is indicated in Figure 5.17. T in

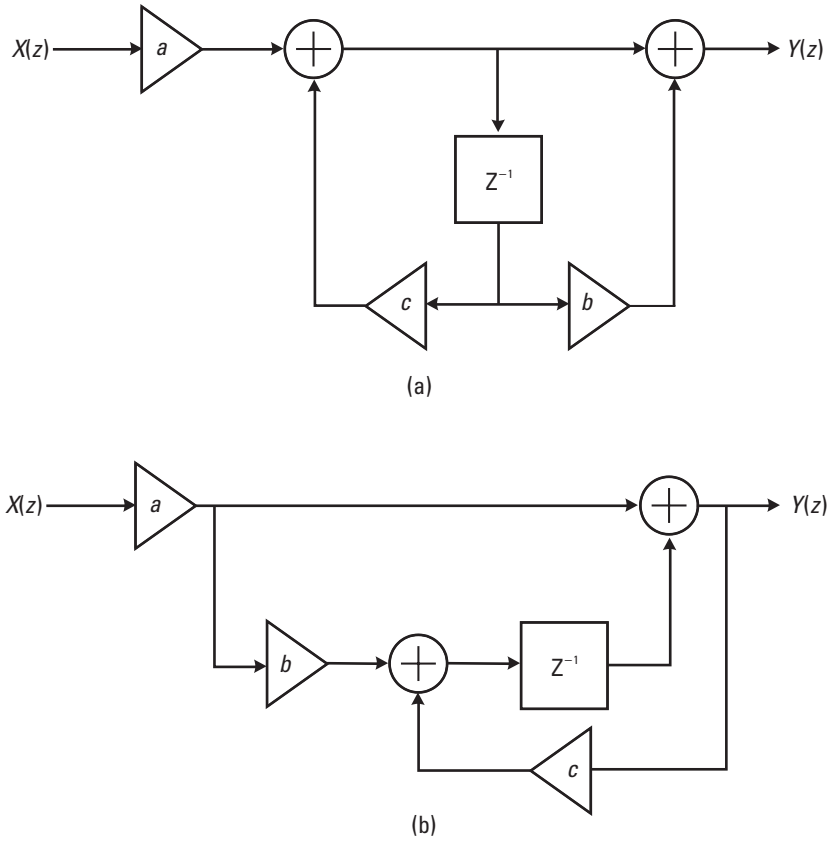


Figure 5.17 Examples of fundamental digital filter configurations: (a) example I and (b) example II.

(5.41) has the unit of time in digital processing. It can be considered as the data sampling time in an actual device. The transfer functions shown in Figure 5.17(a, b) are the same but with different components.

A second-order complete integral filter can be created using a different digital processing circuit. By separating the second-order complete integral filter [(5.34)] into proportional and integral parts and converting both into their digital equivalents, we obtain

$$F(z) = A\left(1 + \frac{B}{1 - z^{-1}}\right) \tag{5.42}$$

The digital processing described by (5.42) is realized by an adder, a multiplier, and unit delay elements as shown in Figure 5.18. The relationship between the coefficients of (5.41) and (5.42) is expressed as follows:

$$\tau_1 = \frac{T}{A \bullet B}, \tau_2 = \frac{T}{B}$$

$$a = \frac{A}{1+B}, b = -\frac{1}{1+B} \tag{5.43}$$

5.4.3 Phase and Frequency Response of a Second-Order Loop

Let us simulate the behavior of a second-order PLL with an active filter. The output phase changes as shown in Figure 5.19 when a phase jump occurs in the input signal. The *x* axis of this figure expresses the product of natural angular frequency and elapsed time *t*. The *y* axis expresses the phase variation normalized by the amount of the input phase jump. The parameter is the damping coefficient ζ. Figure 5.19 indicates that decreasing ζ yields larger overshooting, and more time is needed to converge towards the final value. In Figure 5.19, the phase change is observed in the input and output signals of the PLL using an external standard signal. Therefore, the output phase converges to 1 of the normalized output phase, which equals the amount of the phase jump of the input.

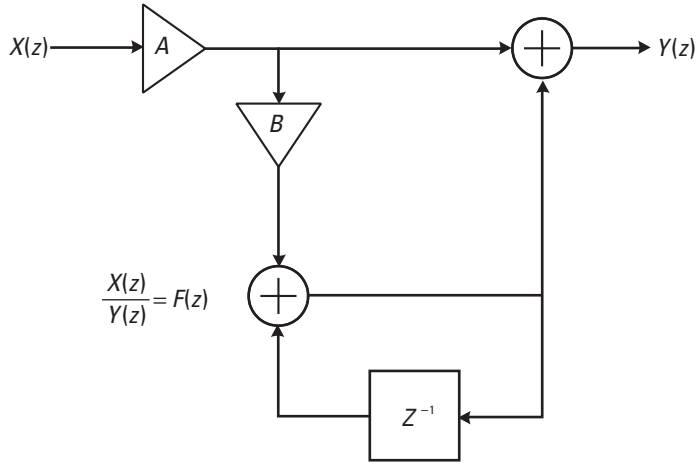


Figure 5.18 Digital processing in a second-order complete integral filter.

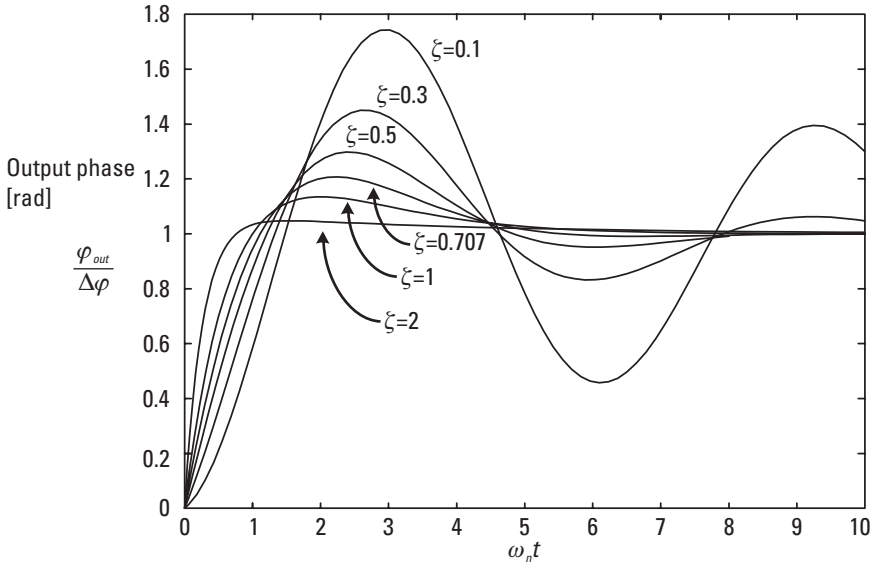


Figure 5.19 Phase step transient response of a PLL with an active filter.

When a frequency jump occurs in the input signal, the input phase changes after the jump, as per the dotted line in Figure 5.20, if measured against an external reference. The output phase change is plotted as the solid line in Figure 5.20. To better comprehend the tracing process, we should observe the phase difference between the input and output signals as shown in Figure 5.21. The tracing process varies with the natural angular frequency and the damping coefficient. As the damping coefficient ζ decreases, the overshoot increases and it takes more time to converge towards the input signal.

Let us simulate the difference in the transient response between second-order PLLs with either passive or active filters. The product $\tau_2 K$ in (5.35) and (5.36) is used as a parameter. Output phase changes follow the input phase jump and the input frequency jump as depicted in Figures 5.22 and 5.23, respectively. Both figures show that when $\tau_2 K$ is much larger than 1, the pair of curves (Figures 5.22 and 5.23) well duplicate the characteristic of a PLL with an active filter. When $\tau_2 K$ is 0, the corresponding pair of curves well mirror the characteristic of a PLL with a lag filter. As $\tau_2 K$ decreases, the overshoot is reduced as shown in Figure 5.22, and the output phase offset against the input signal phase (0 in Figure 5.23) increases as shown in Figure 5.23. The

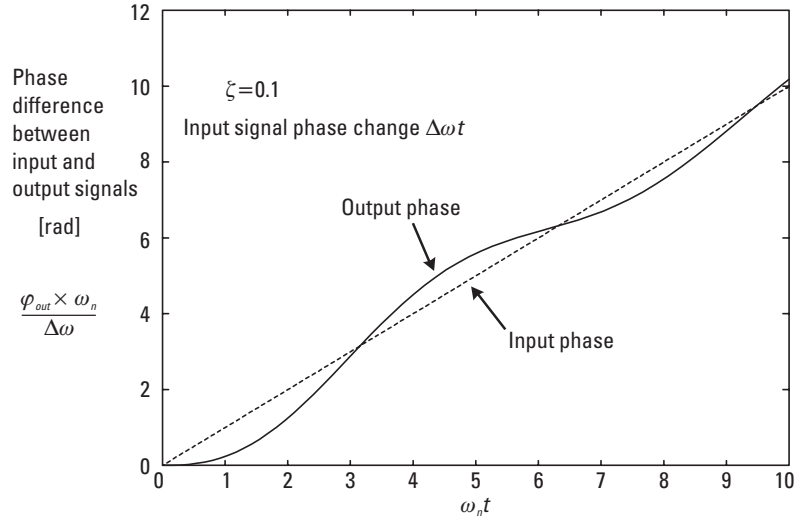


Figure 5.20 Frequency step response of a PLL with an active filter, as measured against an external standard.

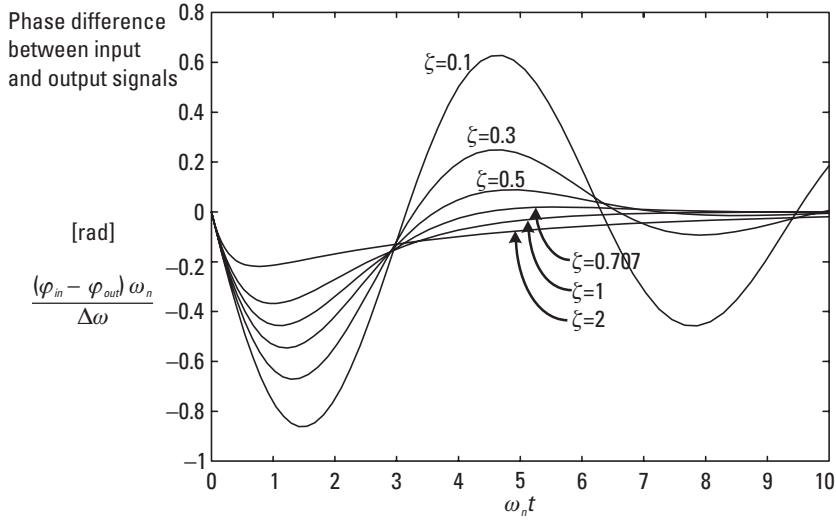


Figure 5.21 Frequency step response of a PLL with an active filter: the phase difference between input and output phase signals was observed.

phase offset in Figure 5.23 represents the steady-state phase error caused by the loop gain decrease that accompanies the decrease in $\tau_2 K$.

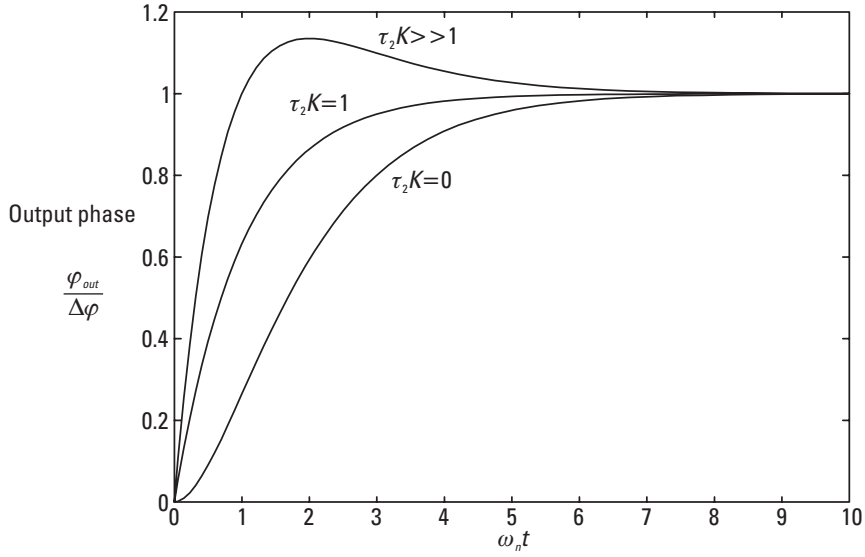


Figure 5.22 Phase step response of a PLL with a passive filter.

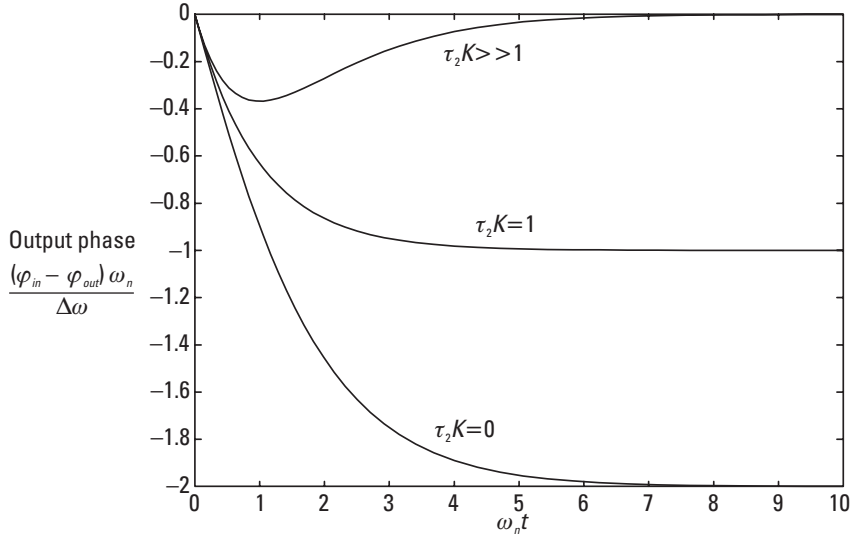


Figure 5.23 Frequency step response of a PLL with an active filter.

References

[1] Gardner, F. M., *Phaselock Techniques*, New York: John Wiley & Sons, 1966.

[2] Viterbi, A. J., *Principles of Coherent Communication*, New York: McGraw-Hill, 1966.

[3] Balanchard, A., *Phase-Locked Loops*, New York: John Wiley & Sons, 1976.

[4] Best, R. E., *Phase-Locked Loops*, New York: McGraw-Hill, 1997.

6

Simulation of a Digital PLL

Simulation is an effective tool for analyzing the characteristics of digital PLL constructions, since it is often difficult to find any closed-form arithmetic relationships in their behavior. In addition, because of increased computer power, simulation is an attractive and precise alternative if we are interested in a characteristic whose measurement is difficult or would take a lot of time. This chapter describes the validity of PLL simulation.

6.1 Transfer Function Analysis

Let us consider a basic digital PLL that is composed of a digital filter (described in Section 5.4.2), a digital phase comparator, and a digital controlled oscillator, as shown in Figure 6.1. A complete transfer function is hard to obtain, but a reasonable analysis is possible through an equivalent analog PLL in the region where the quantization effect can be ignored.

Let us compare the transfer function of an analog PLL calculated according to (5.36) with the measured data of the digital PLL, which uses the following coefficients:

1. Phase comparator:
   - Phase difference detection resolution: 6 ns;
   - Sampling: 8 kHz and 0th hold;

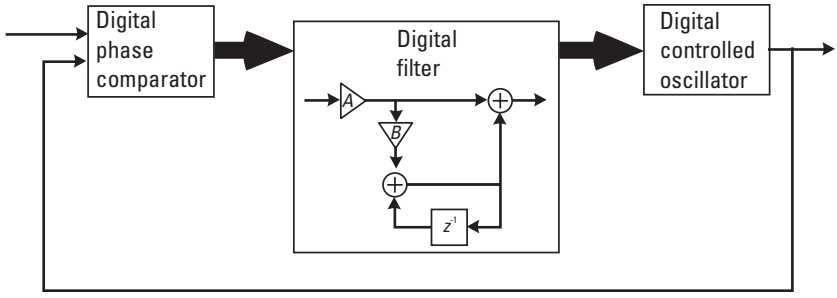


Figure 6.1 Example of a digital PLL.

2. Filter:
   - Type: complete second-order filter as expressed in (5.42);
   - Proportional coefficient of multiplication: A = 32;
   - Integration coefficient: $B = 6 \times 10^{-5}$;
   - Sampling period in integration: 8s;

3. Controlled oscillator:
   - Resolution of digital frequency control: 1×10^{-12}.

Figure 6.2(a) shows the transfer function for a variation in the input phase of the PLL. This is the characteristic of a general lowpass filter. Figure 6.2(b) shows the transfer function given a variation in the internal phase of the PLL loop. This is equivalent to a highpass filter characteristic as high-frequency components of the variation in the loop appear at the output. The most probable variations are phase noise and temperature dependence of the controlled oscillator. Solid lines in Figure 6.2 plot the calculated results based on the above parameters and the approximation of a digital PLL by an analog one. The approximated transfer function agrees well with the measured results of a digital PLL. The method is useful if the sampling period is smaller than the time constant of the digital PLL and quantization errors are insignificant [1].

6.2 Transient Response Simulation

The transient response of a digital PLL can be estimated by analyzing an equivalent analog design, as was done with the transfer function. The

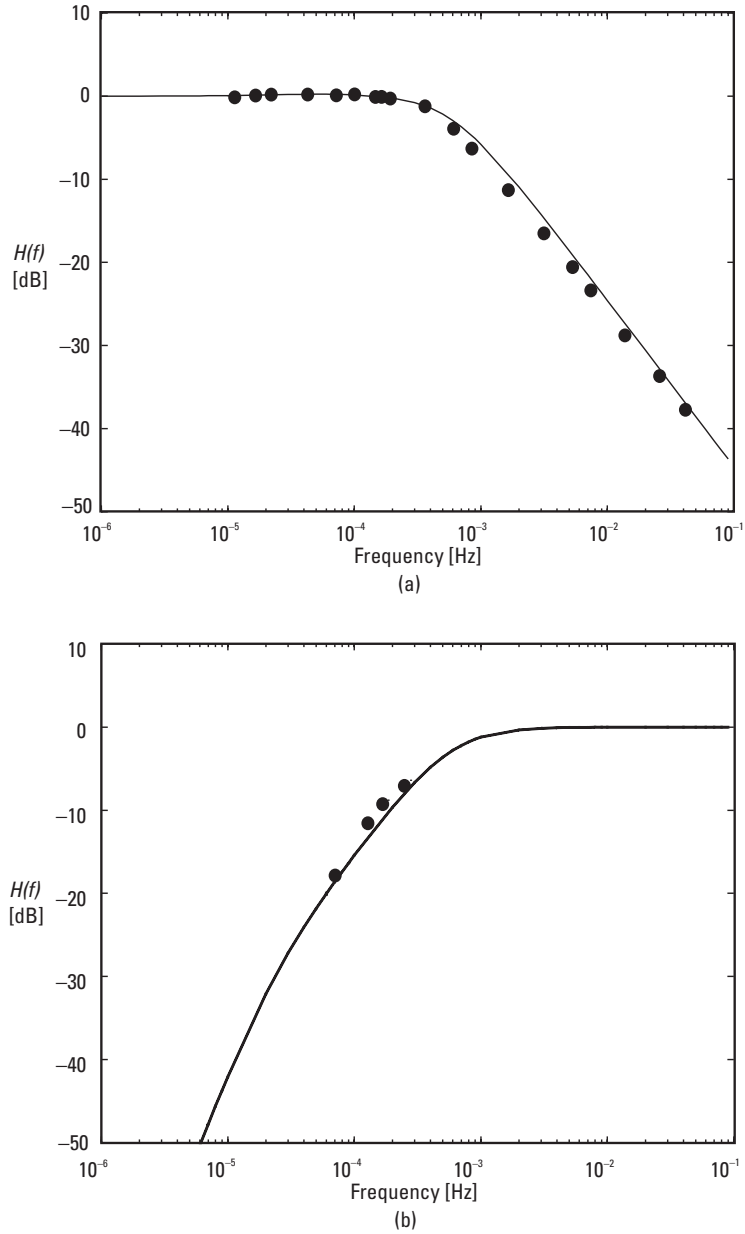


Figure 6.2 Calculation results of an analog approximation of the transfer function and respective measurement data for the digital PLL: (a) transfer function for input phase variation of the PLL, and (b) transfer function for internal phase variation of the PLL. (*After:* [1].)

influence, however, of the quantization effect is significant in the signal phase of practical digital PLLs. Figure 6.3 compares the measured and simulated values of the frequency step response in a digital PLL with the same configuration as was used in Figure 6.1. The phase starts to change (with the frequency step added at the beginning of observation) and gradually converges to the stationary state. Digital PLLs yield phase steps in this control process while analog designs behave smoothly. This is because there is a region where a digital PLL cannot detect a phase variation. Even if the phase change is continuous, the comparator output does not change until the difference reaches the next quantization level. The quantization step size corresponds to the phase difference detection resolution of the comparator. In the example of Figure 6.3, a step of 10 ns has been generated. The simulation results, which correspond very well to the measured data, are depicted by the solid line in Figure 6.3. A digital PLL mechanism was considered instead of the analog approximation. Computer simulations are thus very effective for analyzing digital PLLs in some detail.

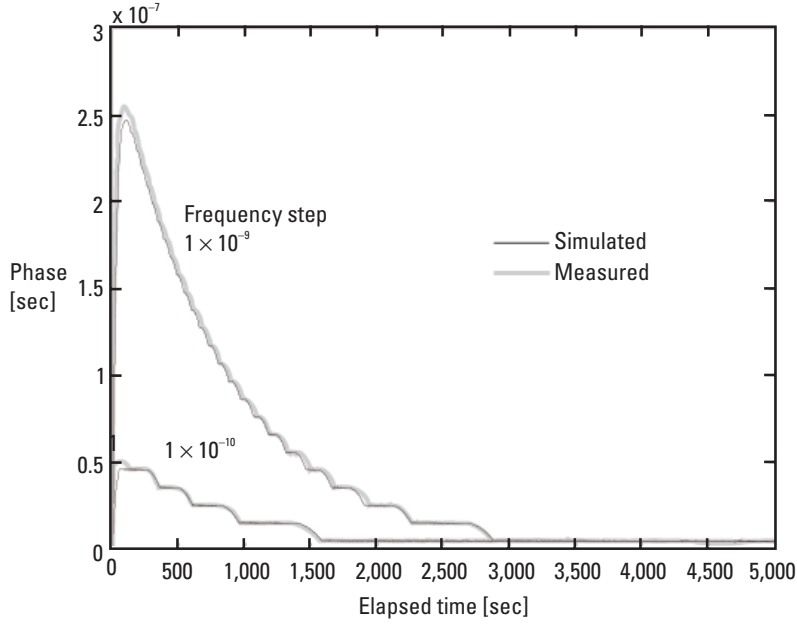


Figure 6.3 Frequency step response for an input signal in phase-time.

6.3 Steady-State Phase Error Simulation

Although a PLL synchronizes to the input signal phase, the output phase is not completely equal to the input phase due to the respective frequency difference between the input and output signals. The phase difference, which remains even after synchronization is achieved, is called the steady-state phase error. This phase error occurs when the free-running frequency is offset from the input frequency or vice versa. The steady-state phase error in a first-order loop was described in Section 5.2.2. The steady-state phase error in a second-order loop is given by

$$\Delta\theta_e = \frac{\Delta\omega}{K \bullet F(0)} \tag{6.1}$$

where $\Delta\omega$ is the angular frequency difference, K is the loop gain, and $F(0)$ is the DC gain of the filter. K is calculated from the frequency variable coefficient of the controlled oscillator and the phase difference detection coefficient of the phase comparator.

This equation indicates that the frequency difference between the input and the output signals shrinks according to the DC gain of the loop, which happened in the first-order loop as well. When using a passive filter as in Figure 5.15(a), there is a steady-state phase error due to the limited DC gain. In an ideal active filter this error can be eliminated since the DC gain is infinite. Actual physical active filters, however, do not provide this benefit. An ideal filter is possible, however, in software, and the steady-state phase error there can be completely eliminated.

There is a condition in which the input and output frequency difference increases over time. One example is the free-running frequency of a controlled oscillator, which can vary due to oscillator frequency aging or drift. The steady-state phase error is determined by the DC gain of the filter in a passive filter. If the frequency difference increases and we use a passive filter, (6.1) indicates that the steady-state phase error will eventually diverge to infinity. In practice, however, the steady-state phase error gradually increases rather than diverging, since the frequency difference changes very slowly compared to the PLL response time if the variation in the free-running frequency of the controlled oscillator is due to aging. Therefore, the maximum frequency variation of a controlled oscillator should be considered when estimating system life. For an active filter, the increase in the frequency difference can be canceled by adding an integration function. This holds the steady-state phase error constant as shown here:

$$\Delta\theta_e = \frac{D_{\Delta\omega} \bullet \tau_1}{K} \tag{6.2}$$

where $D_{\Delta\omega}$ is the angular frequency difference change per unit time and τ_1 is the time constant of the active filter.

6.3.1 Simulating the Steady-State Phase Error of a First-Order Loop

First, the synchronization process in a first-order loop digital PLL is simulated to observe the generation of the steady-state phase error. Figure 6.4 shows the phase-time error change over time from the initial condition of 2 μs to the stationary state. It is also assumed that there is a constant frequency difference between the input and output. While the phase error almost disappears after 5×10^4 seconds when there is no frequency difference, it is found that a phase error of 1 μs is present if there is a frequency difference. Equation (6.1), which uses units of [rad], can be converted into (6.3), which uses unit of phase-time (i.e., [s]).

$$\Delta T_e = \frac{\Delta y_s}{KF(0)} \tag{6.3}$$

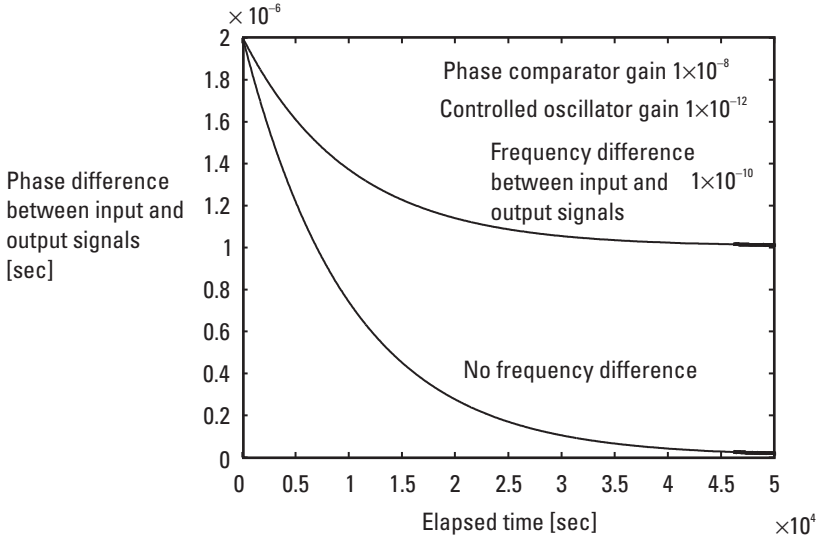


Figure 6.4 Steady-state phase error in the digital PLL with a first-order loop when there is a frequency difference between the input and output signals.

where ΔT_e is the steady-state phase error in seconds and Δy_s is the normalized frequency deviation. In the example shown in Figure 6.4, the DC gain (from the gains of the phase comparator and the controlled oscillator) is 10^{-4}, and the normalized frequency deviation is 1×10^{-10}. The steady-state phase error is found to be 1 μs ($=10^{-10}/10^{-4}$). The simulation result equals the calculated value.

6.3.2 Simulating the Steady-State Phase Error of a Second-Order Loop

The steady-state phase error in a second-order PLL with an active filter can be simulated assuming that the frequency of the controlled oscillator increases at constant rate over time. The parameters of the PLL are the same as those used in Section 6.3.1. The rate at which the normalized frequency deviation increases was set to 1×10^{-14}/s, or about 1×10^{-9}/day. This is quite typical if a controlled crystal oscillator is aging. We assume that the input frequency continuously equals that of the external standard.

The simulation result is plotted in Figure 6.5. The frequency difference starts to increase at time 0. The steady-state phase error of a second-order PLL is found to settle down to a constant value, even if the frequency difference increases. The final steady-state phase error depends on the amount of integration in the active filter. The more integration, the smaller the phase error is. It is possible to theoretically calculate the steady-state phase error from (6.2).

Equation (6.2) (units of [rad]) can be transformed into (6.4) (units of phase-time).

$$\Delta T_e = \frac{D_{\Delta y} \bullet \tau_1}{K} \tag{6.4}$$

where ΔT_e is the steady-state phase error in seconds, $D_{\Delta y}$ is the normalized frequency deviation difference between the input and the output (unit [1/s]), and τ_1 is the coefficient given by (6.5), which has been derived from (5.43).

$$\tau_1 = \frac{T}{A \bullet B} \tag{6.5}$$

The steady-state phase error given by this relationship is 1 μs when the integration coefficient is 1×10^{-3}. The simulation plot for the same integration coefficient indicates the same value.

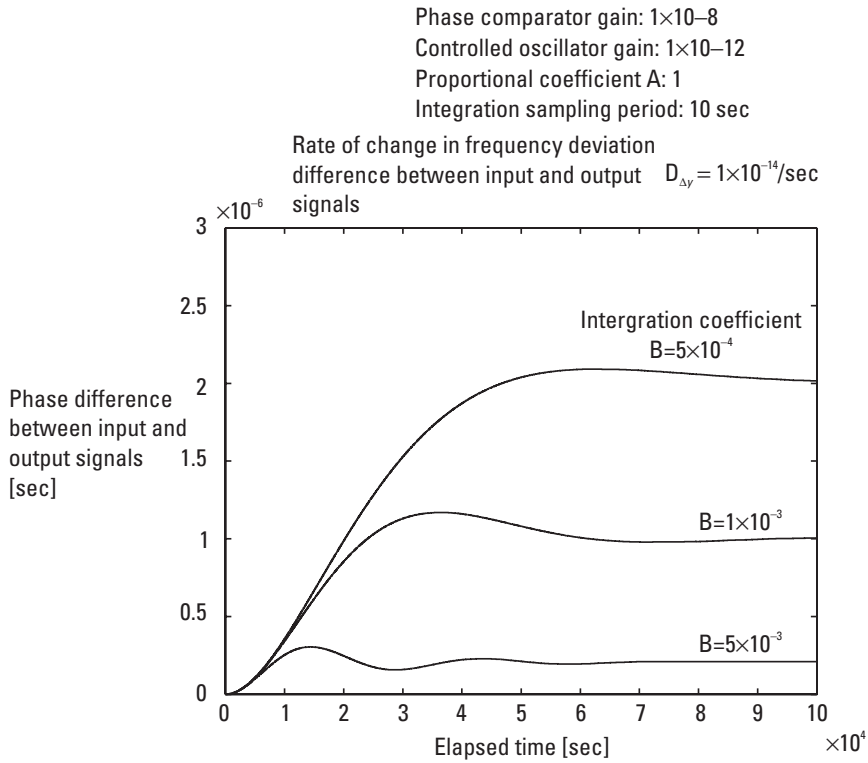


Figure 6.5 Simulated steady-state phase error in the digital PLL with a second-order loop when there is normalized frequency deviation between input and output signals. The digital filter is composed of proportional and integration parts.

When designing a digital PLL we have to estimate its long-term characteristics such as the final steady-state phase error at the end of system life and the frequency changes of its controlled oscillator. Since it is difficult in practice to measure the long-term characteristics of a controlled oscillator, the above theoretical calculation or simulation will be very useful in designing a PLL.

Reference

[1] Kihara M., "Performance Aspects of Reference Clock Distribution for Evolving Digital Networks," *IEEE Communication Magazine,* Vol. 4, No. 4, 1989, pp. 24–34.

7

Clock Systems in Networks

7.1 Wired Systems

7.1.1 Clocks in Digital Systems

Clock signals are containers that carry the physical quantity of frequency. Frequency is transferred across digital systems by the clock in a way that is different from the one in analog systems. Clocks are a new part of digital systems.

Digital systems are either synchronous or asynchronous. In the former case, all system elements utilize the same reference clock, while in asynchronous systems they do not. Current telecommunication networks, especially long-haul transmission networks, are basically synchronous. In each such network, the clock is synchronized and shared. It conveys the common frequency used to synchronize the network elements. In this configuration, the clock distribution area constitutes the whole network. Each asynchronous network, on the other hand, may have many clock distribution areas, each of which is usually formed between neighboring functional blocks. The physical separation between functional blocks varies from neighboring integrated circuits to several or several tens of kilometers.

Synchronous and asynchronous communication both require that the clock be transferred to the intended destination. The transfer process is a basic item for any system. In a synchronous system, the clock is distributed from the master node, which acts as a reference. This reference clock is regenerated at all other nodes and further transmitted to all systems in each node.

Since all circuits in the network use the same clock (i.e., same frequency accuracy) this method makes it easy to multiplex/demultiplex and add/drop information using registers. If the clock frequency deviates from the reference clock, information loss can occur, and in the worst case, communication might fail. In asynchronous systems, since the frequency is not synchronized, the individual elements are insensitive to clock frequency deviation. However, these systems require additional mechanisms that can prevent information loss, which would otherwise result from the frequency difference.

A network that transfers information packets, such as the Internet, can be realized either as a synchronous or as an asynchronous system. Asynchronous transfer mode (ATM), which uses labeled packets called cells, has been deployed into telecommunication networks. An ATM transmission system is normally realized using the frame format of synchronous digital hierarchy (SDH), which is basically a synchronous method. Note that an asynchronous configuration was considered in the original development phase. Ethernet, originally developed as a local area network (LAN) standard, is a completely asynchronous system. Even the more recent Gigabit Ethernet is asynchronous. Designing a synchronous system demands some assumptions as to network scale and the type and number network elements used. Designing an asynchronous system appears to be easier in that each element can be designed more independently. Most internal circuits, however, use synchronous mechanisms. Since synchronous systems support higher clock frequencies, which means faster data rates, they have become more popular over the last few years.

7.1.2 Clocks in Transmission Systems

In basic digital communications, the clock of the sending site should be used on the receiving site as well. The sending site generally transmits the common clock together with the information [1]. This enables the receiving site to regenerate the information. The clock is distributed independently of associated information on the circuit level, but these two share the same medium (cable, radio carrier) in the transmission system. This is natural because the transmission medium cost dominates in long-haul systems. This sharing, however, leads to a problem. If the receiving site loses the clock, information too will be lost. Each transmission code has its own merits and weaknesses. The non-return to zero (NRZ) code, adopted in the SDH system, is the most popular transmission scheme in current networks [2]. If the data stream consists of a succession of either all 0s or all 1s, the receiving site loses the clock. To prevent this, the final circuit at the sending site scrambles

the signal. This evens out the occurrence frequencies of 0s and 1s [3]. The receiving site descrambles the received data stream to get the original data.

Figure 7.1 shows a transmission clock system. Information on the sending site is converted into the data stream according to the transmission clock generator. It also hosts the clock changing circuit from which the converted data is released to the transmission path. The converted data holds both information and clock components. The receiving site needs a clock recovery circuit and an information regeneration and discrimination decision circuit. The transmission clock recovery circuit receives the data stream from the transmission path, and extracts only the clock component. The information regeneration and discrimination decision circuit uses the extracted clock to regenerate the original information. Noise added in the transmission path is removed by the transmission clock recovery circuit, and the regenerated information is output by the information regeneration and discrimination decision circuit

7.1.2.1 Transmission Clock Generation Circuit

The transmission clock generation circuit must generate the clock with the specified frequency accuracy. If the frequency of the data stream does not lie within the specified range, the clock cannot be properly extracted because it is outside the synchronization range of the clock recovery circuit at the receiving site. Receiving errors or information loss consequently occur. Long-haul SDH transmission systems demand frequency accuracies of 20 ppm (20×10^{-6}) or better. Small quartz oscillators are suitable for achieving such frequency accuracy. Unfortunately, such oscillators do not cover all frequency ranges. Figure 7.2 shows how we can generate a clock for the gigahertz frequency region using a quartz oscillator that directly generates 155 MHz. The example in Figure 7.2 is actually used in SDH systems [2]. The output of the 155.52-MHz quartz oscillator is converted into 2.48832 GHz in the PLL.

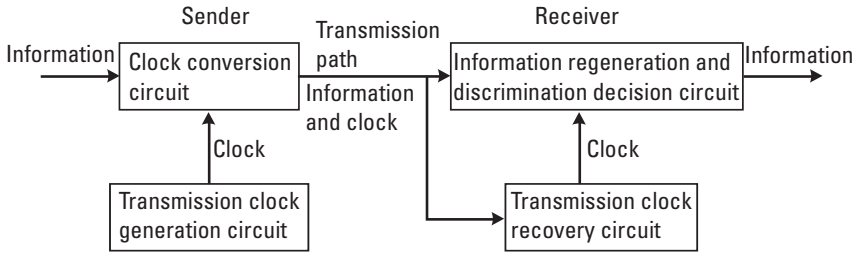


Figure 7.1 Typical clock arrangements in a transmission system.

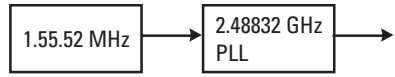


Figure 7.2 Clock generation in the gigahertz frequency region.

The converted clock corresponds to the clock frequency of an STM-16 transmission system. Figure 7.3 shows a schematic of a commercial clock generation IC [4]. The clock for STM-16 is generated from the frequency of 155.52 MHz by the PLL. The multiplexer is installed in this IC. Sixteen data signals, each 155.52 Mbps (STM-1), are multiplexed, and the data stream of 2.48832 Gbps is output through the data discrimination decision circuit. A 10-Gbps clock generation circuit (STM-64) will enter the market soon [5].

7.1.2.2 Transmission Clock Recovery Circuit

The transmission clock recovery circuit must reliably extract just the clock component from the transmitted signal. The first task is to suppress the effect of noise coming from the transmission path. In addition, the clock must be recoverable from any transmitted signal. The direct approach is to use a filter as a clock recovery circuit. The most generic clock regeneration approach consists of an LCR circuit and a piezoelectric quartz resonator. Recent demands for higher resonant frequencies, better frequency stabilization, and lower costs have forced the clock recovery circuits to employ PLLs as shown in Figure 7.4. A transmission clock recovery circuit with a PLL (which offers phase synchronization) generates the same frequency as the initial frequency component in the transmitted signal. At the same time, transmission path noise is suppressed by the lowpass characteristic of the PLL.

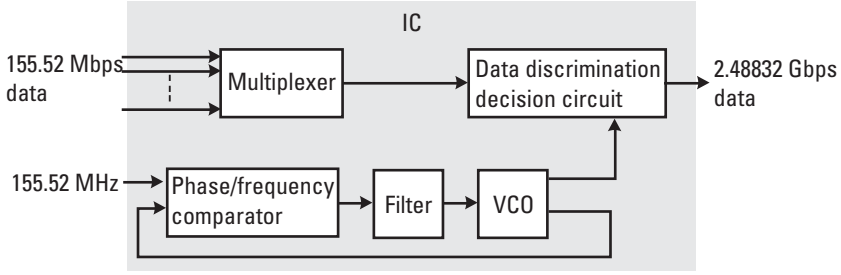


Figure 7.3 An example of a clock generation IC suitable for transmission systems.

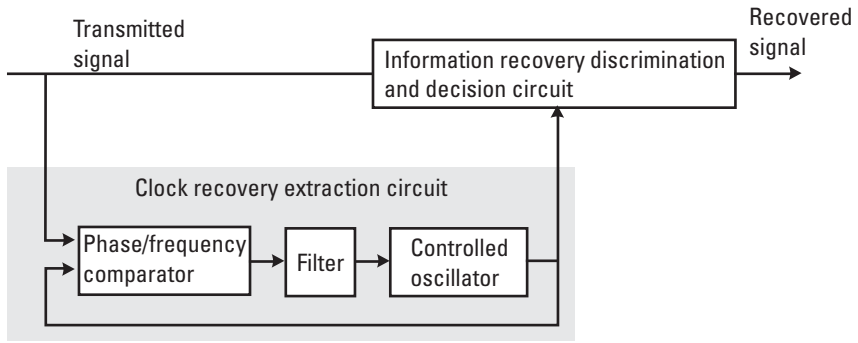


Figure 7.4 Signal receiver with a recovery circuit using a PLL.

A resonator-based clock extraction circuit needs a high Q value to remove noise. The output phase of a resonator, however, will change when the center frequency shifts from the clock frequency of the sending node. Thus, the phase margin in the clock recovery circuit becomes small, which raises the error probability. Since this tendency is proportional to the Q value, there is an upper limit for it in resonator-based clock recovery circuits. Also, the noise reduction possibilities are limited. On the other hand, in PLL clock recovery circuits, noise can be sufficiently suppressed by selecting the cutoff frequency of the lowpass characteristic. The steady-state phase error in PLLs can limit the phase margin in the information regeneration discrimination decision circuit. Second-order PLLs with passive filters are often used to solve this. Unfortunately, such PLLs exhibit a lowpass characteristic with positive gain. This means that a series connection leads to noise accumulation. It is necessary to decide PLL parameters such as the filter configuration considering the noise characteristics of the whole network.

7.1.3 Clock System for Network Synchronization

All elements in a synchronized network must use the same clock. At any healthy node in the network, the total amount of information coming in always equals the total amount going out. This network type has a significant advantage: multiplexing/demultiplexing and adding/dropping of information can be simply and easily achieved by digital processing. If the clock frequency shifts from the nominal value, the transmission quality deteriorates. Clock generators must meet strict specifications such as frequency accuracy and stability [6–8].

The reference clock is distributed from the master node to all other nodes using star and/or hierarchical structures. The components forming the clock distribution network that links the master node to the slave nodes are shown in Figure 7.5. The reference clock at the master node becomes the unique master clock for the network. A cesium atom standard is usually used for this purpose due to the demanding requirements [6]. Alternatively, a GPS disciplined rubidium unit may be sufficient. The slave clock in each node must satisfy three points:

1. Receive and regenerate the reference clock transmitted from the master node, which requires that the noise caused by the transmission process be reduced.
2. Supply derived clocks that have the appropriate frequencies to the node's systems.
3. Continue clock supply if the reference clock fails for whatever reason.

Slave clocks that can achieve these points generally have the configuration shown in Figure 7.6. The PLL has the frequency keeping (holdover) function to maintain the required clock frequency accuracy when the reference clock fails [7]. A redundant system configuration is generally adopted to

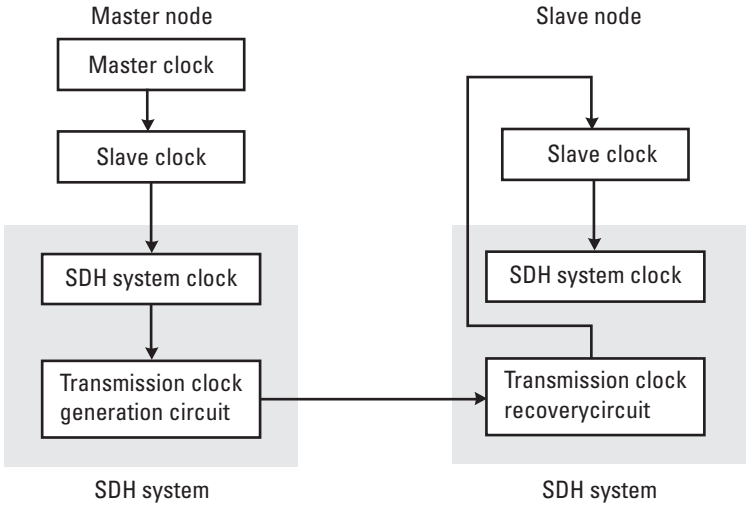


Figure 7.5 Clock arrangement for SDH network synchronization.

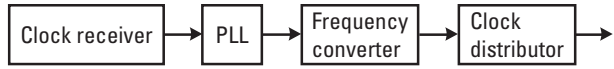


Figure 7.6 Slave clock configuration.

ensure reliability. The PLL design methodology for slave clocks is described in Chapter 8.

The SDH system lies within the shaded area in Figure 7.5. It consists of SDH system clocks, transmission clock generation circuits, and transmission clock recovery circuits. The SDH system clock receives its reference from the slave clock and generates a clock with the specific frequency used in the SDH system. In addition, the SDH system has a function to hold the required frequency accuracy to avoid the impact of slave clock failure. Transmission clock generation/regeneration is a basic feature of transmission systems such as the SDH system as described in Section 7.1.2. The signal generated in the transmission clock generation circuit is transmitted to the slave node (the right side in Figure 7.5) and recovered in the clock recovery circuit of the respective SDH receiver circuit. The transmission clock from the clock recovery circuit is directly passed to the slave clock and it becomes one input of the clock receiver (see Figure 7.6). By repeating this process, the reference clock is transmitted from node to node, and the reference clock is distributed to each element in each node. All clocks throughout the network can be synchronized by this process.

7.1.4 Clock System Across Multiple Links

In a synchronous system, the clock is transferred by passing it through the components shown in Figure 7.5. This configuration connects many PLLs in series. The phase variation that can occur as the clock transits a path can be propagated and often amplified. Even a transient response, which is not a severe problem in a single PLL, might become a significant cause of phase variation. If the damping coefficient of a second-order PLL (see Section 5.4.1) is small, and there are phase and frequency variations in its characteristics, a lot of time will be needed to stabilize the phase. This phenomenon is equivalent to the PLL transfer function having positive gain over the frequency range, as shown in Figure 5.16. This gain occurs in each link of the PLL chain, and the phase overshoot is magnified at the last node of the chain. There are two elementary solutions to this problem: restrict the number of linked PLLs or completely separate the transfer functions of the

individual PLLs. These two tricks tend to be impractical. The more practical remedies are as follows:

1. Reduce the transfer function gain.
2. Make the transfer functions of the PLLs (the slave clock, the SDH system clock, and the transmission clock generation circuit) different.
3. Select a larger damping factor in the slave clock and the SDH system clock since they have a larger margin in terms of the maximum synchronization time.

7.2 Wireless Systems

7.2.1 The Effects of Network Topologies on Clock and Frequency Synchronization

Modern technology has enabled very flexible cable-based digital communication networks in which a strictly predefined communication hierarchy is no longer necessary. The ultimate requirement of unlimited mobility [9], however, has pushed the respective solutions of some radio communication networks even further, and it is thus appropriate to consider some the typical topologies separately. The associated frequency and synchronization issues, related difficulties, and selected solutions will be pointed out with an emphasis on the different systems used to transport a digital baseband signal.

7.2.1.1 Point-to-Point Communications

Two distinct classes of point-to-point radio frequency communication arrangements can be found. The first, and perhaps the more straightforward one, is a microwave link, which forms the communication path between two physically separated but stationary nodes of a larger network. This solution is selected either due to geographical constraints, which would make a fiber optic connection impractical or due to synergies with other transmission systems (shared use of antenna towers, etc.). We usually find symmetric capacity in both directions, although this is not necessary. In these cases the problems of synchronization are normally not very severe, particularly if an all-digital link chain is in use. The radio frequency dimensioning provides an adequate carrier-to-noise ratio, there is seldom any measurable multipath due to the very high gain antennas, and timing jitter should only become a real problem in extreme environmental conditions and at very high data rates. Such

circumstances include tropical rainstorms, heavy-wet snow, and wind speeds above 30 m/s. Water or snow particles will simply cause a huge additional attenuation whereby the connection will be temporarily lost. Winds, however, might cause large enough distortions of the tower installation itself to disturb the path delay profile. In fact, the receiving site would see a moving transmitter, and depending on the associated time constants, some of the synchronization (either carrier frequency or timing) might be temporarily lost.

Some mobile point-to-point communication systems exist as well. They are very often established for a dedicated mission of limited duration, and some can make use of radio interfaces, which are intended for a broader scope too. Sometimes both ends of the link are moving, but more often only one. A highly asymmetric capacity profile is commonly seen. Less conventional examples include an image transmission system for helicopter-based ice-monitoring to assist an ice-breaker or a military unmanned airborne vehicle (UAV) light airplane used for reconnaissance tasks. Most commercially available systems utilize frequencies at or above the ultra high frequency (UHF, generally from 300 MHz) band due to size constraints. The radio frequency scenario is challenging both due to the continuously changing multipath and due to the Doppler processes involved. A typical detected synchronization signal is illustrated in Figure 7.7. The electronics on board the moving platform will face an adverse environment, which will be discussed in Section 15.2.2. Despite data rates often being very high and the reliability requirements exhaustive, a link between two users is still relatively

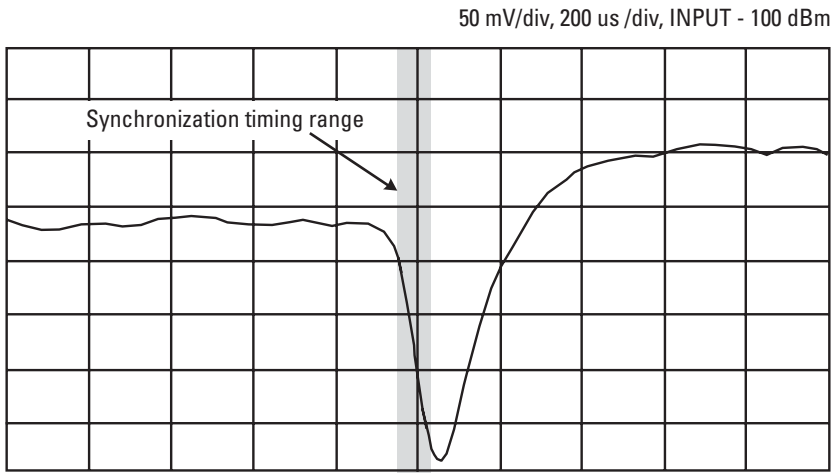


Figure 7.7 An example of the received synchronization signal in a mobile radio network.

easy to handle. Both the up and downlinks can use frequency steering and adaptive oscillator control without any concern for additional receivers or transmitters in the same system—because there are none! The whole time domain is available just for this traffic. If necessary, the two sites can be pre-synchronized prior to the mission, and very often there is some reserve bandwidth available whereby the system is not pushed against its theoretical capacity limits—as opposed to many dense multipoint solutions.

In addition to the two above-mentioned cases, many multipoint networks (like military high-frequency communication setups between 10 kHz and 30 MHz) can be operated as separate simple point-to-point systems, but even so, their technical arrangements tend to follow the original concepts.

7.2.1.2 Point-to-Multipoint Networks

Strictly speaking, this kind of a tree-like topology is not a true communication arrangement, except in a very few special cases. The topic must be included, though, because its most well-known application—broadcasting in all its forms—is currently being merged with the traditional communication functions. Technologically this network type is relatively easy to control if no or very limited uplink traffic is needed [10]. The arrangement is a typical master-slave system in which the frequency and timing control flows directly from the node towards the (normally) receive-only terminals, which is also the direction of data flow.

Some of the technical issues in this network category are quite interesting. The required data rates for live video broadcasting are among the highest, and the system may include both stationary and highly mobile users. Although human vision can be fooled to a certain extent—as has been done for decades in the old analog television systems—many transmission-related defects will turn up immediately. The transmitting node has typically no information about the behavior or state of the receivers (e.g., clock frequencies), and even if this were technically arranged, there would be hardly any means available to handle the amount of incoming data. So the communication must follow a blind transmission scheme and practically no means are at hand to recover lost data at a receiving site. A further technical complication is caused by the requirement for low-cost terminals, which should provide almost superior broadcasting quality even in adverse radio frequency conditions. Because normal radio and television programs include human voices, any temporary interruptions are very annoying and are thus considered unacceptable.

Simple carrier frequency tracking is easy to implement in the receiver units. An example of the performance of such an arrangement is shown in Figure 7.8. One of the greatest threats is still—after the introduction of

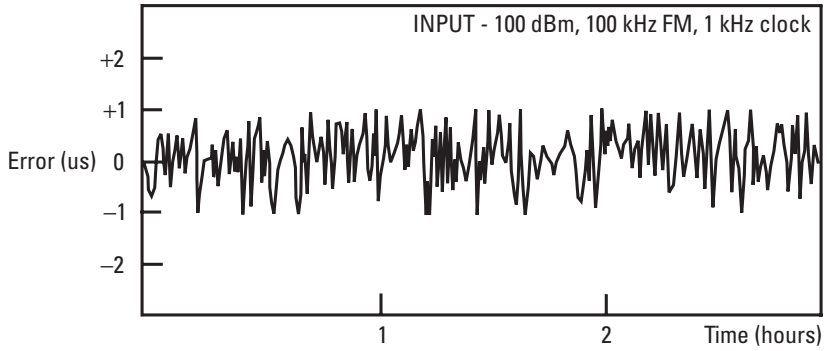


Figure 7.8 The measured performance of a frequency modulated carrier-tracking receiver.

orthogonal frequency division multiplexing (OFDM)—multipath propagation, which may cause the pattern synchronization to collapse [11]. This is the case particularly if a simultaneous combination of a high data rate with a high physical velocity is needed. A resynchronization action will often take too much time to be hidden from the end user of the communication terminal.

Other users of the point-to-multipoint concept are, for example, networks, which provide alarm messages to rescue forces. Here, however, the transmission data rates are normally very low and adequate service can be provided without dedicated radio frequency tricks.

7.2.1.3 Cellular Networks

Cellular communication networks were originally designed to transmit and receive human conversations. The first designs utilized an all-analog concept, and the main challenges were in providing a comprehensive geographical coverage at a minimum acceptable signal quality level. The requirement for frequency characteristics was just the maximum allowed frequency deviation including the clock frequency aging. The present digital networks follow the centralized control concept and must follow a fully synchronized principle in addition to real-time carrier frequency control.

Although the cellular network topology is based on a connection between the terminal and a base station or cell, the synchronization arrangement must take into account the other mobile units as well as the adjacent cells. In practice, all cells share a common data clock and preferably a common carrier frequency reference too. This can be easily derived from the first clock whereby phase continuity is assured also during modulation.

In terms of synchronization management, a cellular network is somewhere between a simple point-to-point system and a totally "free" ad hoc network (which is described next in Section 7.2.1.4). The individual cells, the base stations, and base station controllers form a fixed configuration having predefined delays and clock characteristics as has been outlined in Section 7.1. The mobile terminals (or, simply put, cellular phones), on the contrary, take arbitrary positions within the coverage area. They also have a relatively wide velocity range. Thus, there is a need to adjust for the varying time differences and to take into account possible frequency shifts. At present, most cellular networks use a form of feedback control, which is feasible due to the symmetric nature of the communication path. However, the time constants of the loop, which consists of the mobile terminal, the base station, and the communication paths, become longer. Unless better local clocks are used, the inherent synchronization quality will be deteriorated.

The challenges for synchronization are partly related to multipath propagation, which may be severe and highly fluctuating in a dense urban environment, and partly related to the technical realization of the oscillators in the mobile unit. Also, the evolution of data transmission needs from the simple human discussion towards Internet and real-time video has necessitated much higher clock speeds. Naturally, this leads to more and more stringent stability requirements. Similarly to the broadcasting environment, a cellular system is badly hampered by any retransmissions or resynchronization actions, as opposed to other wired digital communication.

7.2.1.4 Ad Hoc Networks, Multipoint-to-Multipoint Networks

Both modern commercial and military communication systems nowadays make use of an ad hoc concept in which no predefined radio network topology exists. The individual units have to follow a common standard but in addition to their completely arbitrary geographical location, their logical placement is freely arranged. Some of the networks may have limited forms of centralized control but this is not necessarily implemented. Typical examples are some short-range office systems operating in the higher microwave bands, tactical military UHF networks, and other systems intended for government authorities.

Clock and synchronization arrangements are extremely complicated in these radio networks if high data rates and availability factors are the goal. This is due to the unknown and constantly changing path delays and the variety of individual units. Each mobile terminal can at the same time be a node for several others, and frequent rerouting may occur. It is also possible that a device, which has initially been stationary and has served as a transfer

node for a multitude of other terminals, suddenly starts to move. In such a case we have to cope with the relative frequency changes as well.

7.2.1.5 Mixed Communication, Navigation, and Broadcasting

A somewhat different synchronization task is met with the new systems where several radio-related functions have been combined inside a single mobile framework. For example, devices capable of simultaneously performing broadcast reception, personal navigation, and data communication have been introduced. Instead of thinking of such concepts as a simple physical merging of individual technologies, the efficient use of spectral resources and hardware and the wish for add-on services suggests that the whole operational idea be harmonized. This is a challenge to synchronization and frequency control but does provide new opportunities too.

Within any reasonable time span, we must assume that the original source networks used, for example, for broadcast data generation, will use their independent clock systems. A similar situation will apparently prevail inside the navigation community. Thus, the new network's designs and its individual mobile terminals must be able to resynchronize themselves "on-the-fly" as defined either by user preferences or technical constraints. On the other hand, we can assume that the increase in the available background synchronization references as seen by the terminal devices can be utilized to enhance the local clock performance. For example, low-cost mobile communication devices can reap the benefit of navigation-based clock stability to provide an error-free data link performance for their users.

7.2.2 Synchronization and Frequency Control Arrangements for Radio Networks

Many mobile communication systems follow a simplified step-wise scheme of synchronization. The radio frequency parts are first aligned either by using simple crystal oscillators, phase locked loops, or DDS circuits. After that, the baseband synchronization can start in the desired order. At the radio frequency interface, the frequency control activity is needed not only because of natural requirements of functionality but also because of the ultimate need of improved spectral efficiency. The smaller the uncertainties in the carrier frequency, the smaller the guard intervals need to be. The technological approach is somewhat different depending on the communication task, although the low-cost digital processing blocks have already started to take over almost everywhere. For historical reasons it is interesting to note that the

first practical radio transmitter networks based on frequency synchronization were designed according to a German patent granted in 1933 [12].

7.2.2.1 Fixed Frequency, Single Carrier Systems

If the communication system has only one carrier frequency, which will stay constant both in terms of elapsed time and coverage volume, the frequency-generating technology has traditionally been the simplest one. If the operating band is low enough, direct crystal oscillators are preferred. In the past, when equipment density was very low, simple LC resonant circuits could be utilized whereby any arbitrary frequency value could be tuned but, of course, at the expense of stability. Higher frequency microwave signals are produced by PLL circuits and multipliers, which impair the stability in proportion to the order of multiplication. Harmonic mixers must be used to cover the highest microwave and millimeter wave bands to support the PLL units. As these systems have no need for frequency changes, all individual elements can be tuned to give the best possible performance.

7.2.2.2 Multiple Carrier Systems

Most radio communication networks have several carriers simultaneously in use, and at least the terminal devices must be able to make use of all of them. It means that they have to be able to change their frequency rapidly and accurately. Thus, simple crystals are not suitable but we need PLL-based synthesizers or, more recently, DSP generators. The frequency range is often not the most severe constraint because amplitude-based network planning normally favors carrier spans of only some tens of megahertz (at typical UHF frequencies). It is more challenging to provide the necessary frequency switching speed and the respective settling time to achieve the necessary frequency accuracy. This is particularly true for cellular networks during a hand-over action, which might be requested by a mobile device traveling at 100 m/s. In addition to this, the spectral purity is a major issue because of the heavy crowding of the available channels. Spurious emissions at one transmitting frequency may be seen as interference in the nearby receivers, which operate at a close-by frequency.

7.2.2.3 Frequency Hopping Networks

Slow or fast frequency hopping has been initially introduced to the military communication networks to provide some low-probability-of-intercept/low-probability-of-detection (LPI/LPD) characteristics. The idea is to change the transmitting frequency so rapidly that the opponent cannot find the transmitter by surveillance receivers nor can he try to jam its message.

The commercial communication networks started to use frequency hopping to reduce the effects of multipath, as will be suggested later by (15.2). The technical challenges are first to design analog circuitry (amplifiers, filters, antennas, etc.) that has a wide enough bandwidth and then to construct oscillators that are able to perform the very fast frequency switching action synchronously with the receiving installation. Initially, the devices have used special PLL circuits in which some compromises have been allowed for faster settling, but modern designs rely on the DSP concept as long as the frequency range permits. Here the difficulties in maintaining required spectral purity near the momentary carrier are even worse than in the more decent synthesizers because of the restrictions in filtering.

7.2.2.4 Spread Spectrum Communications

Despite frequency hopping networks often being seen as spread spectrum networks, real spread spectrum communications systems are implemented by extending the information (data) content of a signal over a much wider radio frequency range than required by modulation theory. So it need not be frequency hopping; nor are all hopping systems real spread spectrum.

The practical spreading action requires that the data synchronization be tightly bound to the RF generation process. This is best accomplished by using modern DSP circuits. If the desired frequency range is much higher than the capabilities of current digital blocks, it is always possible to use an additional upconverter after the DSP unit. Care should be taken to use adequate filtering in order to remove any local oscillator leakage from the output.

If needed, the spread spectrum scheme can be added to a frequency hopping system. This improves the protection of the content and will give additional multipath rejection in those systems where the baseband rate is low to medium. Normally the individual radio terminals have very high quality oscillators [either an ovenized quartz oscillator (OCXO) or rubidium] to enable autonomous operation. This means that, for example, an airborne communication link is completely locked before the mission and will stay so as long as the respective oscillators are able to maintain the required timing error. Synchronization over the air interface is strictly forbidden. It would unavoidably reveal the vital system parameters to the external observer.

7.2.2.5 Impulse Radio, Ultra Wide Band (UWB) Transmission

The impulse radio concept, now in use by some military units for short-range communications [13], can be thought of as a special case of spread

spectrum. Each individual bit can, in theory, occupy a bandwidth of several gigahertz. This makes a hostile detection very complicated and gives very efficient multipath rejection characteristics. The synchronization requirements are very severe—not in terms of carrier frequency, because there is none—but regarding the baseband timing. Again, there is no reason to separately transfer the clock pattern across the air interface. Some reports of scientific experiments suggest pulse widths down to about 50 ps. A wideband data communication capability seems to require timing uncertainties below this, which can be quite difficult to achieve in a mobile radio device regardless of technology.

7.2.2.6 Frequency Tracking Through the Air Interface

The invention of practical PLL circuits made it possible to start experiments with a real-time frequency and timing control over the air interface whereby radio terminals can theoretically show performance equal to their cable-connected counterparts, (see, for example, [14]). The first widespread implementations of this method were introduced as automatic frequency correction (AFC) circuits in simple FM receivers. Currently the need for frequency and code tracking is partly based on the Doppler cancellation and partly on the requirement of highest data rates. Intelligent DSP blocks can be utilized to combine both functions on single chip.

The problems of synchronization through the mobile radio interface are twofold. First of all, the carrier-to-noise ratio is typically far less than in cable networks. This sets considerable requirements to the adaptive filtering functions, normally installed in the DSP chip. Second, the delay stability and pulse train integrity of the detected base band signal is often questionable. This can often not be overcome just by algorithmic processing but might require supporting hardware in the form of local high stability oscillators.

References

[1] ITU-T Recommendation, "Physical/Electrical Characteristics of Hierarchical Digital Interfaces," G.703, 2001.

[2] ITU-T Recommendation, "Optical Interfaces for Equipments and Systems Relating to the Synchronous Digital Hierarchy," G.957, 1999.

[3] ITU-T Recommendation, "Network Node Interface for the Synchronous Digital Hierarchy (SDH)," G.707, 2000.

[4] Clock synthesizer, Lucent Technologies, TTRN012G5 (2.5Gbit/s).

[5] Clock synthesizer, Lucent Technologies, TTRN01110G.

[6] ITU-T Recommendation, "Timing Characteristics of Primary Reference Clocks," G.811, 1997.

[7] ITU-T Recommendation, "Timing Requirements of Slave Clocks Suitable for Use as Node Clocks in Synchronization Networks," G.812, 1998.

[8] ITU-T Recommendation, "Timing Characteristics of SDH Equipment Slave Clocks (SEC)," G.813, 1996.

[9] Norbert, G., and W. Wiesbeck, *Planungsmethoden für die Mobilkommunikation*, Heidelberg: Springer Verlag, 1998.

[10] Kadish, J., and T. East, *Satellite Communication Fundamentals*, Norwood, MA: Artech House, 2000.

[11] Dambacher, P., *Digital Terrestrial Television Broadcasting*, Heidelberg: Springer Verlag, 1998.

[12] Trenkle, F., *Die deutschen Funkstörverfahren bis 1945*, Ulm: AEG-Telefunken, 1984.

[13] Vakin, S., L. Shustov, and R. Dunwell, *Fundamentals of Electronic Warfare*, Norwood, MA: Artech House, 2001.

[14] Eskelinen, P., "Experiments with a Caesium-Locked UHF Time and Frequency Reference Transmitter," *Proc. 14th European Time and Frequency Forum*, Turin, 2000, pp. 204–208.

8

Digital Synchronization System Design

This chapter deals with the necessary design methods for the network synchronization clock system described in Section 7.1.3. A digital PLL is selected as a design example, since it is often used in a practical clock system. Design parameters of various digital PLL elements (a digital-processing phase comparator, a digital-controlled oscillator, and an active integrating digital filter) are explained. The total PLL characteristics that vary with those parameters are described as well.

8.1 Basic Design

The most important component of a PLL is the controlled oscillator. PLL performance can be improved to some extent by optimizing its design parameters and by using additional circuits such as a variable time constant circuit and a high-order filter. The improvements that are possible with these techniques, however, pale in comparison with those achieved by using a highly stable oscillator. A fundamental requirement is a controlled oscillator with good frequency stability. Oscillator selection requires us to balance performance, cost, and size. The size and cost often dominate the selection process. If the maximum size and cost are given, the upper stability limit—and hence overall PLL performance—can be approximated. Generally speaking, the smaller an oscillator is, the worse its frequency stability is. In a controlled oscillator with poor stability, the output frequency can exceed

the specified value if the frequency control is slow compared to the frequency variation of the oscillator itself. The best solution is to set the PLL cutoff frequency high whereby the control loop time constant gets shorter. This improves the frequency stability of the controlled oscillator because the PLL output is more tightly coupled to the input signal.

8.2 Basic Configuration

Let us design a digital PLL as shown in Figure 8.1, which is composed of a digital processing phase comparator (DPC), a digital-controlled oscillator (DCO) and an active integrating digital filter. The target in this example is to achieve a circuit that can synchronize as accurately as possible to an input signal used as the reference for the PLL. The DPC outputs the phase difference as digital data. A highly stable controlled oscillator is assumed. The active integrating filter can be achieved only by digital signal processing.

8.3 DCO Design

8.3.1 Oscillators and Their Selection

Oscillator selection should be carefully made given its impact on PLL performance. The important point is to determine the priority of the oscillator characteristics, since no oscillator suits all applications. Table 8.1 shows some currently used PLL oscillators and their characteristics. The most popular in PLLs is the quartz crystal oscillator. Probably the most important factors in stationary applications are its temperature characteristics (frequency changes with environmental temperature) and frequency drift (aging over time). The OCXO, which holds the crystal in a thermostatic oven, offers high frequency stability against temperature change. Since such ovens may be large and consume power, the temperature compensated quartz oscillator (TCXO), which

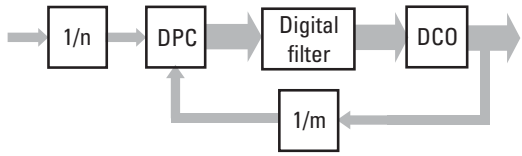


Figure 8.1 Digital PLL configuration.

Table 8.1
Oscillators and Their Characteristics

Oscillators	Quartz crystal oscillator			Rubidium standard	Cesium standard
		TCXO	OCXO		
Frequency accuracy	1×10^{-4}	2×10^{-6}	5×10^{-8}	$\sim 5 \times 10^{-10}$	5×10^{-13} $\sim 1 \times 10^{-11}$
Frequency aging	$2 \sim 5 \times 10^{-6}$ /year	$0.5 \sim 1 \times 10^{-6}$ /year	5×10^{-9} $\sim 5 \times 10^{-11}$ /day	1×10^{-11} $\sim 5 \times 10^{-11}$ /month	(Within frequency accuracy to longevity)
Frequency temperature dependence	$10 \sim 50 \times 10^{-6}$	5×10^{-7} $\sim 3 \times 10^{-6}$	$1 \sim 5 \times 10^{-9}$	$1 \sim 5 \times 10^{-10}$	1×10^{-13} $\sim 1 \times 10^{-12}$
Frequency short-term stability $\tau = 1$ sec	$\sim 1 \times 10^{-9}$	$\sim 1 \times 10^{-9}$	1×10^{-12} $\sim 5 \times 10^{-11}$	5×10^{-12} $\sim 5 \times 10^{-11}$	5×10^{-11}
Consumption power	10~250 mW	10 mW	0.1~5W	5~20W	30~50W
Size	0.02~10 cm^3	0.02~10 cm^3	5~200 cm^3	100 ~1,000 cm^3	5,000 ~30,000 cm^3
Price	~$100	~$100	$500 ~$10,000	~$10,000	~$50,000

is equipped with temperature detection and frequency control circuits, is preferred when the size limit is severe.

Large oscillators are basically advantageous in terms of lower aging because the quartz resonator stability tends to be directly related to its size. The quartz oscillator is a device with a long and glorious history. Its frequency stability has been improved by decades of research in such areas as the crystal structure itself and the electric circuits used. Improvements in the manufacturing technology have also been continuous. A recent trend is miniaturization. TCXO progress has been accelerated by the popularity of cellular phones. TCXO packages that occupy 3×5 mm have been developed. They offer frequency accuracy within some ppm including their temperature characteristics [1]. OCXOs smaller than 5 cm^3 have been developed as well [2].

Atomic oscillators, traditionally used as network standards, are far more stable than quartz oscillators. They have been thought to be too large and expensive for most applications. Due to constant research and development activities, extremely small rubidium oscillators are now available. They are only slightly more expensive than OCXOs and can now be used in place of crystals in many applications. An ultra miniature rubidium atomic standard with a volume less than 100 cm^3 has been developed [3]. A unit-price comparison is shown in Table 8.1. A modern rubidium oscillator is considered hereafter because of its frequency stability. This greatly simplifies PLL parameter selection.

8.3.2 Digital Control System

Rubidium oscillators generally produce a constant frequency. There are two approaches to using the rubidium atomic standard as a controlled oscillator:

1. *Control by digital/analog conversion:* The output frequency of the rubidium atomic standard can be slightly adjusted via its internal magnetic field due to its operating principle. The control information (a digital value) must be converted into an analog voltage or current with a digital/analog converter, and the output frequency is changed via the internal magnetic field. This method can also be applied to general quartz oscillators that have an analog control input. The advantage is that the frequency control resolution can be freely set by selecting the converter binary width. In practice, noise in the rubidium control input sets an ultimate limit for resolution improvement.
2. *Control by synthesizer:* The output frequency of the rubidium atomic standard can be adjusted by controlling its internal synthesizer. The synthesizer can be an analog one, composed of frequency dividers and mixers, which generate the rubidium resonant frequency or a digital design based on the digital PLL technique or the frequency generation technique to be described in Chapter 10. The final output frequency can often be adjusted by controlling this internal synthesizer. The advantage is that the final frequency can be steered directly without a digital/analog converter. There is a risk, however, of upsetting the rubidium physics process itself.

This chapter provides a PLL design example that assumes the use of digital/analog conversion and a digitally controlled rubidium oscillator.

8.4 Digital Processing Phase Comparator

The function of the phase comparator is to detect the phase difference between the input signals. In an analog-digital PLL, the comparator generates a pulse width generally corresponding to the phase difference. Of course, the phase difference must be expressed as a digital value in a digital PLL. The most basic circuit that can do this is a counter. The phase difference is measured as the number of counts within the period corresponding to the phase-time difference. In this measurement technique, the counter clock period determines the time resolution.

8.5 Digital Control

8.5.1 Gain Adjustment in the Phase Comparator

The phase difference is quantized in the digital phase comparator as shown in Figure 8.2. The gain is determined by the ratio between the phase difference detection resolution and the digital output of the comparator. A comparison against the analog phase comparator is possible by considering the average gain characteristic, the straight line connecting the top points of the step-wise characteristic. Figure 8.3 shows that there are two ways of controlling the gain of the phase comparator:

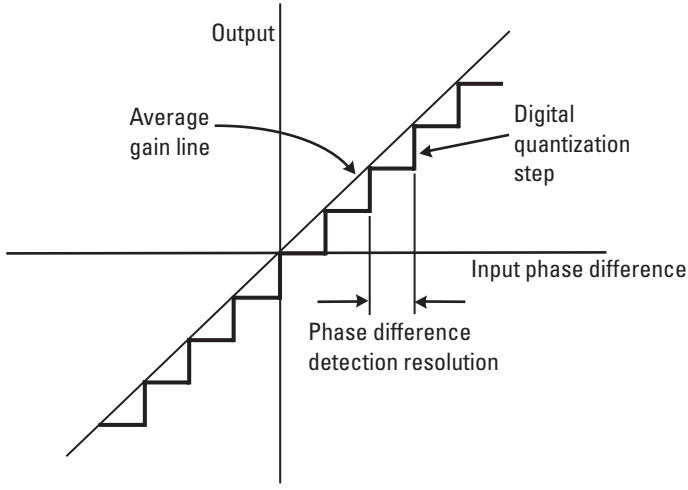


Figure 8.2 Digital phase comparator characteristic showing the binary output as a function of input phase difference.

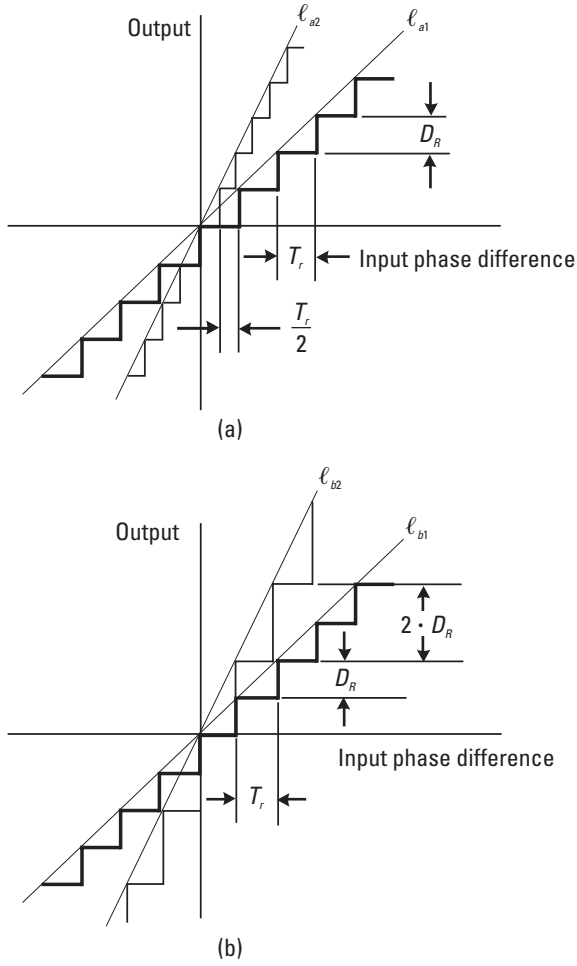


Figure 8.3 Phase comparator gain characteristics: (a) gain characteristic example, and (b) gain characteristic in which phase difference detection resolution changes.

1. Control phase-time difference detection resolution;
2. Amplify/attenuate the output.

Since the phase difference detection resolution T_r is improved to 1/2 in the first method [shown in Figure 8.3(a)], the gain changes from ℓ_{a1} to ℓ_{a2} and therefore doubles. Though the detection resolution T_r is not changed in the second method [shown in Figure 8.3(b)], the output corresponding to

one step of detection resolution doubles to $2 \bullet D_r$. This corresponds to allocating one step of detection resolution to two or more bits. Thus, the gain changes from ℓ_{b1} to ℓ_{b2}, and therefore doubles. There are, of course, differences in the quantization error and phase comparison range between these two methods.

In the digital processing phase comparator the output is converted into a digital value. The output changes by 1 bit when the phase difference exceeds one step of the detection resolution. To cover a range from 0 to 2π with a resolution of T_r and phase comparison frequency f_p, the number of steps N_s is

$$N_s = \frac{1}{f_p \bullet T_r} \tag{8.1}$$

and the number of required bits, N_B, is

$$N_B = \left[\log_{10} N_s \,/\, \log_{10} 2\right] + 1 = \left[\log_{10}\left(\frac{1}{f_p \bullet T_r}\right) \Big/ \log_{10} 2\right] + 1 \tag{8.2}$$

where $[a]$ represents the Gauss operator and yields the maximum integer that does not exceed a.

If the phase comparison frequency is 8 kHz (T = 125 μs) and the detection resolution is 100 ns, N_s becomes 1,250. Accordingly, the output bit number N_B, must be at least 11-bits-long. An example of a phase comparator whose output is an 8-bit digital value is shown in Figure 8.4. This comparator can detect a phase difference from 0 to 2π. The output changes by 1 bit (i.e., by one step of the phase detection resolution T_{r1}) when the phase changes by $2\pi/2^8$, that is, $2\pi/256$, since the phase comparison width of 2π is replaced by 8-bit data. The full detection range cannot be achieved if the resolution is T_{r2} (see Figure 8.4) and N_s equals 256. It is necessary to increase the number of output steps.

8.5.2 Gain Adjustment in the Controlled Oscillator

The gain of the controlled oscillator is determined by the amount the output frequency changes for a 1-bit change of the digital control signal. The larger the output frequency changes, the larger the gain is. The gain of ℓ_2 is larger than that of ℓ_1 in the output frequency change characteristic ($\Delta f \,/ \Delta d$) for digital input values in Figure 8.5(a). Figure 8.5(b, c) shows the frequency

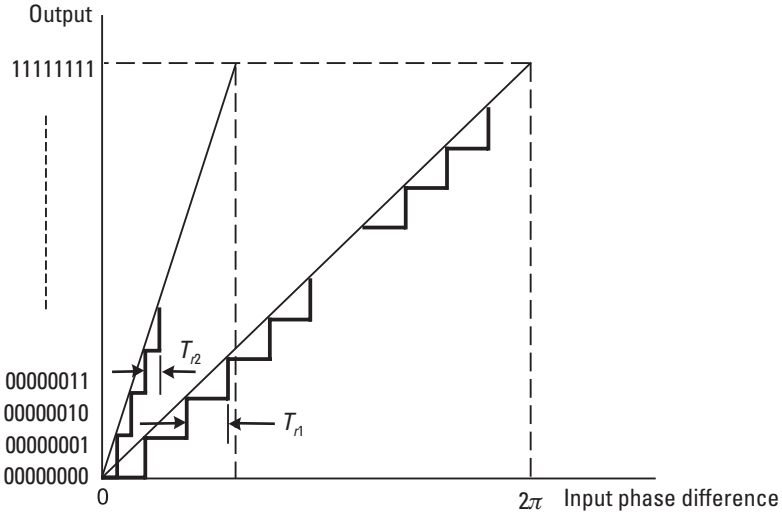


Figure 8.4 Relationship between phase difference detection resolution and digital output.

and phase changes over time as the input value changes with the parameters of ℓ_1 and ℓ_2. It is found that ℓ_2 yields a larger frequency change and a faster phase response. This means that the detected phase difference can be more rapidly reduced by using ℓ_2.

The output frequency of the digital-controlled oscillator initially takes discrete values such as f_1 and f_2 as indicated in Figure 8.6. The oscillator can, however, generate average frequencies such as f_a, f_b, and f_c, and can synchronize to input signals having these intermediate frequencies. The output frequency and phase changes are shown for these in Figures 8.7, 8.8, and 8.9, respectively. Frequencies between f_1 and f_2 can be generated by alternating between input values of d_1 and d_2, and by controlling the duration of each stage. In Figure 8.7, frequency f_a, which is close to f_2, is generated by setting the duration time of d_2 longer than that of d_1. The mean value at the speed of the phase change (amount of phase change/observation time) is equivalent to f_a. In Figure 8.8, the intermediate frequency f_b between f_1 and f_2 is generated by setting equal duration times for d_1 and d_2. In Figure 8.9, frequency f_c, which is close to f_1, is generated by using a longer d_1 duration time than that of d_2. When instantaneous frequency and phase are observed, we find a fluctuating frequency offset and phase difference between the input and output. If the phase change is observed over a long enough time, the average output frequency and phase correspond to those of the input signal of the PLL. The

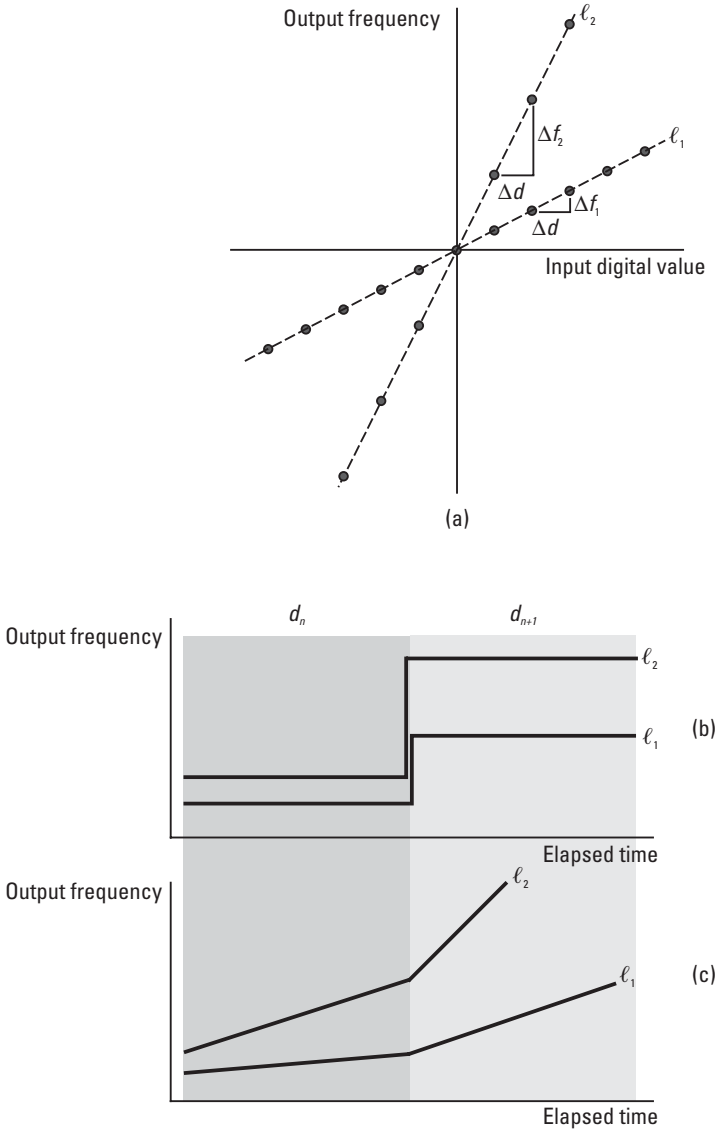


Figure 8.5 Controlled oscillator characteristics: (a) frequency tuning characteristic, (b) output frequency change, and (c) output phase change.

described method cannot always be applied to systems that involve a radio interface. This is because of the unwanted modulation sidebands caused by the temporary changes in the binary control input.

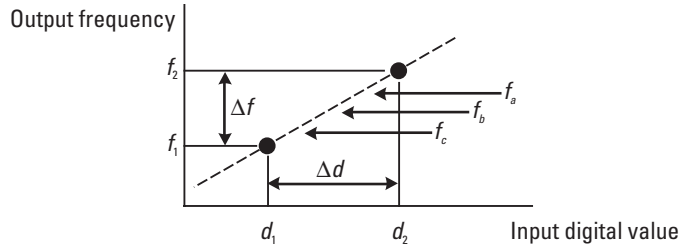


Figure 8.6 Output frequency tuning characteristic and average output frequency in a digitally controlled oscillator.

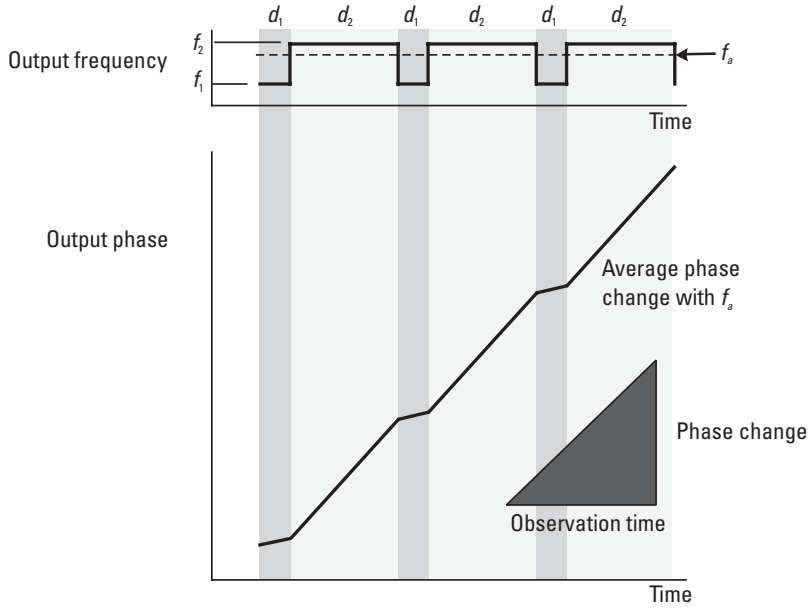


Figure 8.7 Digital control needed to obtain an average frequency of f_a.

8.6 Frequency Range of the Controlled Oscillator

The controlled oscillator of the PLL shown in Figure 8.10 must be able to generate the same frequency as that of the input signal. To counter the changes in the frequency difference expected over time, the controlled oscillator must have some frequency range. In the frequency relationship shown in Figure 8.11, where the center frequency of the input signal corresponds to that of the controlled oscillator, the PLL can basically synchronize to the

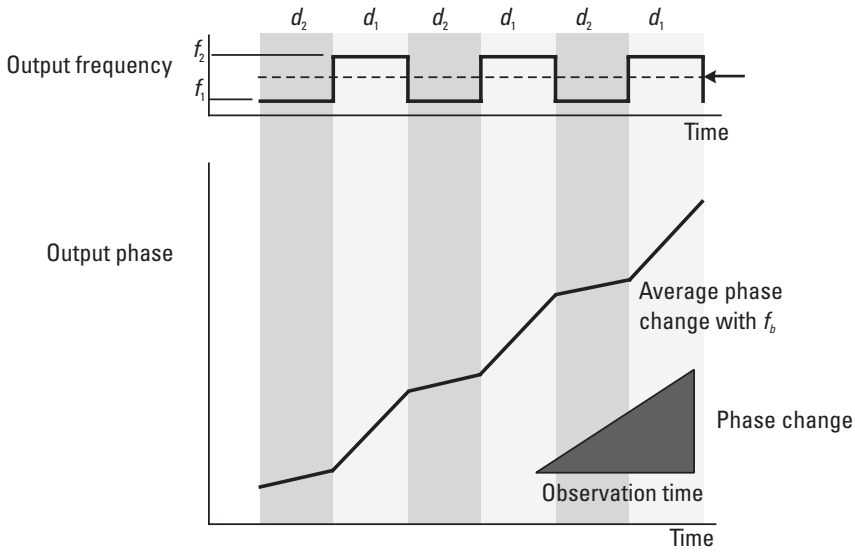


Figure 8.8 Digital control needed to obtain an average frequency of f_b.

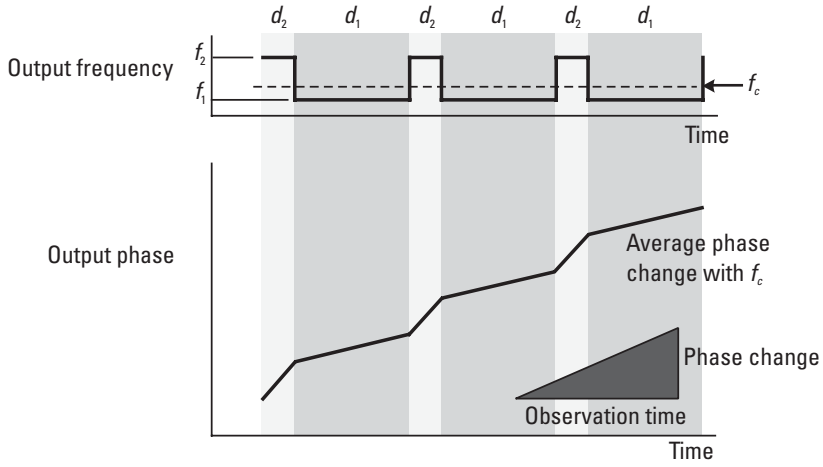


Figure 8.9 Digital control needed to obtain an average frequency of f_c.

input signal if the frequency range of the controlled oscillator matches the frequency range of the input signal.

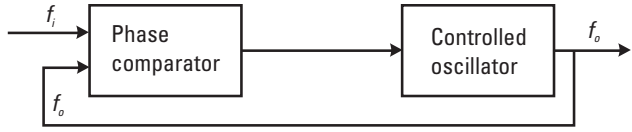


Figure 8.10 PLL in which the input and output frequencies are the same.

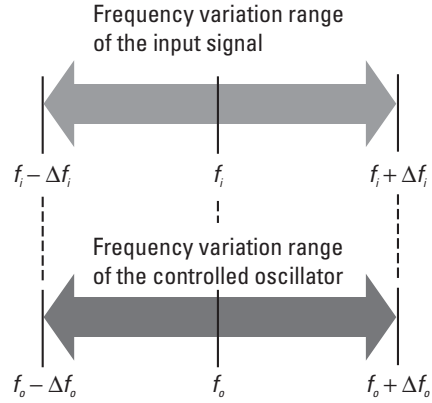


Figure 8.11 Relationship between the frequency variation range of the input signal and the frequency tuning range of the controlled oscillator (in the case where the center frequencies of the input and output signals are the same).

The free-running frequency of a controlled oscillator usually changes over time. This is called frequency drift. PLL synchronization may fail if the frequency range of the controlled oscillator is simply equal to that of the input signal. In frequency region A in Figure 8.12, synchronization cannot be achieved. It is necessary to expand the frequency range of the controlled oscillator to eliminate this region. The range has to be so wide as to include the frequency offset due to drift. The frequency range $\Delta f_o'$ shown in Figure 8.12 must satisfy

$$\Delta f_o' \geq \Delta f_i + |f_D| \tag{8.3}$$

where, f_D is the frequency offset due to drift.

The frequency range in the example of Figure 8.10 was determined from absolute frequency values. The PLL shown in Figure 8.13 has an input frequency, which differs from the output. Here a relative frequency error normalized by the nominal frequency is convenient for calculating the range.

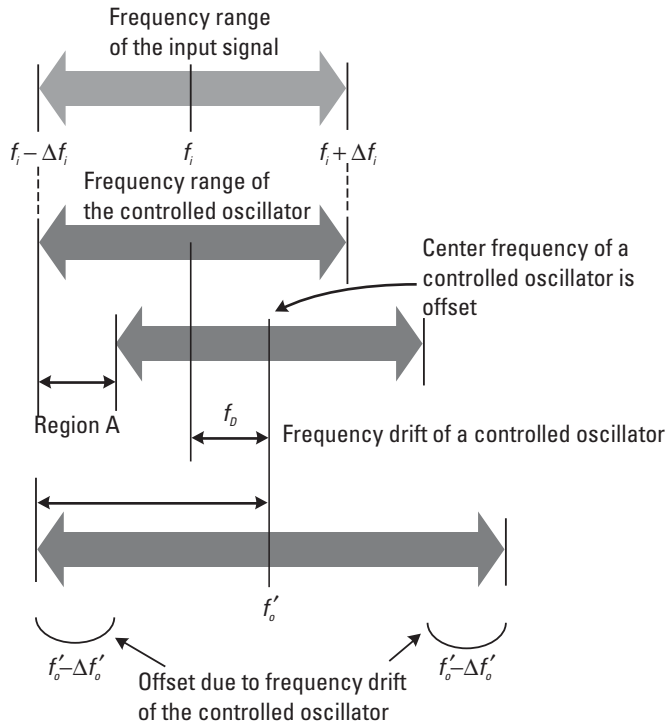


Figure 8.12 Setting the frequency range of the controlled oscillator: the impact of frequency drift.

In Figure 8.13, both the input signal and the oscillator output are individually divided and fed to the comparator. The frequency range Δf_v required for phase comparison is

$$|\Delta f_v| = \left|\frac{\Delta f_i}{N}\right| + \left|\frac{\Delta f_o}{M}\right| \tag{8.4}$$

Converting (8.4) into relative frequency error yields

$$\begin{aligned} |y_v| &= \frac{|\Delta f_i / N| + |\Delta f_o / M|}{f_p} \\ &= \left|\frac{\Delta f_i}{N \bullet f_p}\right| + \left|\frac{\Delta f_o}{M \bullet f_p}\right| = \left|\frac{\Delta f_i}{f_i}\right| + \left|\frac{\Delta f_o}{f_o}\right| = |y_i| + |y_o| \end{aligned} \tag{8.5}$$

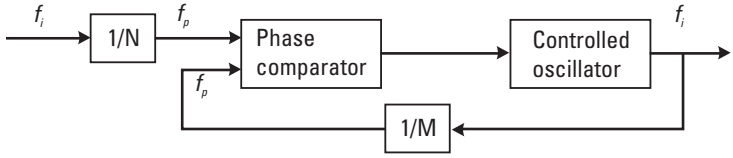


Figure 8.13 PLL whose input and output frequencies are not the same.

where y_i is the relative error of the input frequency, and y_o is relative error of the output.

It is found from (8.5) that using the normalized frequency deviation (relative error) allows the required frequency range to be obtained independent of the frequency difference between input and output signals and the phase comparison frequency. If the input frequency changes by $\pm 1 \times 10^{-8}$, and the frequency drift of the digital-control rubidium atomic oscillator is $\pm 1 \times 10^{-8}$, the required frequency range for this controlled oscillator is $\pm 2 \times 10^{-8}$. The output frequency of the controlled oscillator has to be changed by up to ± 0.1 Hz if an oscillator having a 5-MHz output is used. This estimation is quite rough. The PLL may not be able to synchronize to the input signal due to input noise and loop response even if the frequency range is correct. The estimation described here, however, is an effective step towards the initial design of the PLL.

8.7 Time Constant and Synchronization Range

The cutoff frequency characteristic of a digital PLL can be adjusted comparatively easily by changing the time constant. This is possible because an ideal filter can be achieved by digital processing unlike in an analog circuit. The time constant can be enlarged and the cutoff frequency lowered by the following three methods.

1. Coarsen the phase difference detection resolution of the phase comparator.
2. Refine the frequency control step of the controlled oscillator.
3. Lengthen the control cycle of the controlled oscillator.

Generally, the first and third methods are feasible while the second often faces severe hardware constraints. A reduction of the time constant is

best done by enlarging the frequency control step of the oscillator. This makes the phase of the oscillator output move fast and the response speeds up, which reduces the time constant. Of course, the instantaneous frequency changes more coarsely.

Figure 8.14 shows the relationship between the digital phase comparator and the digital-controlled oscillator in a digital PLL. The comparator and the oscillator both deal with 12-bit digital values. If one cycle of the comparison frequency in the comparator is set to the output value of 4,096, the digital controlled oscillator can be controlled in any phase difference condition, as shown in Figure 8.15.

Let us increase the frequency control step of the controlled oscillator to reduce the time constant of this PLL configuration. To achieve this in a practical way, the phase comparator output is amplified and the frequency control step is increased. When the comparator output changes by one step, the output digital value is doubled (see Figure 8.16). The output phase of the controlled oscillator changes rapidly and the time constant consequentially falls. Since the step number of the full digital value does not increase even if the output is doubled, the comparator output reaches its maximum value and remains constant thereafter as the phase difference increases. In this case, amplification has no impact for phase differences from $+\pi$ to $+2\pi$, though the time constant becomes small for phase differences from 0 to $+\pi$. The

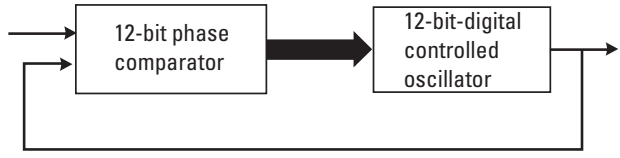


Figure 8.14 Relationship between a digital phase comparator and a digital-controlled oscillator.

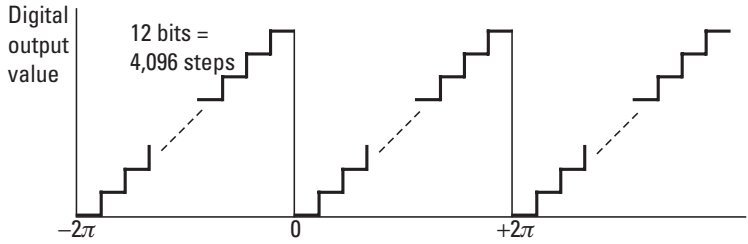


Figure 8.15 Digital phase comparator transfer characteristics.

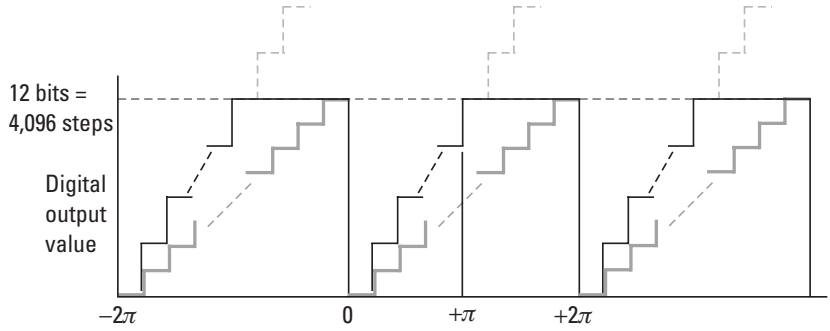


Figure 8.16 Amplified digital phase comparator characteristic.

output frequency of the controlled oscillator does not change in the region where the phase comparator output is constant. This means that phase synchronization is not possible. The phase difference, however, exists and increases because of the frequency difference between the input signal and the controlled oscillator output signal of the PLL. When the phase difference reenters the range from 0 to $+\pi$, the PLL will begin the phase synchronization operation.

8.8 Loop Filter Configuration and Transfer Function [4]

This section assumes that a loop filter is used to realize an active integrating digital filter. The digital filter is composed of proportional and integrating parts (refer to Section 5.4.2). First, the comparator output, $X(z)$ is fed to the digital filter and the magnitude is adjusted by the proportional part. The result is transferred to the integrating part and added to the previous result after magnitude adjustment. The sum $Y(z)$ of the proportional and integrating outputs becomes the input of the controlled oscillator. Figure 8.17 illustrates this process. The actual software must use wide enough registers for storing the calculation results. Otherwise the processing may not give the expected filter characteristic due to a cancellation of significant digits. The transfer function of a digital PLL with the digital processing filter of Figure 8.17 is

$$H(s) = \frac{\alpha s + \alpha\beta}{s^2 + \alpha s + \alpha\beta} \tag{8.6}$$

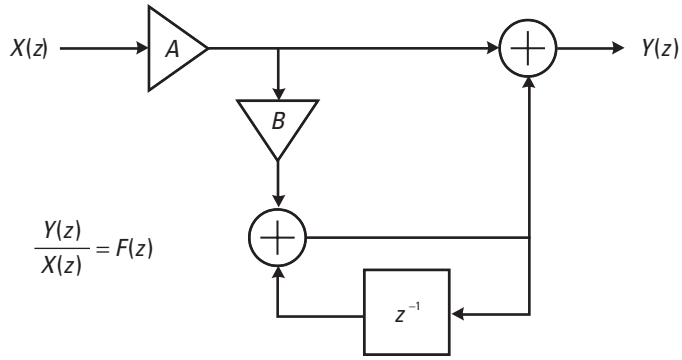


Figure 8.17 Fundamental signal processing elements in a digital filter.

where $\alpha = K_d K_o A$ and $\beta = \dfrac{B}{T}$.

The 3-dB cutoff frequency of this transfer function is

$$f_{-3\,dB} = \frac{1}{2\pi}\sqrt{\frac{\alpha^2 + 2\alpha\beta + \sqrt{\left(\alpha^2 + 2\alpha\beta\right)^2 + 4\alpha^2\beta^2}}{2}} \tag{8.7}$$

This transfer function exceeds unity gain in a certain frequency region as described in Chapter 5. The frequency of maximum gain and the respective gain value are

$$f_{peak} = \frac{1}{2\pi}\sqrt{-\beta^2 + \beta\sqrt{\beta^2 + 2\alpha\beta}} \tag{8.8}$$

$$G_{peak} = \sqrt{\frac{\alpha^2\sqrt{\beta^2 + 2\alpha\beta}}{\left(\alpha^2 - 2\alpha\beta - 2\beta^2\right)\sqrt{\beta^2 + 2\alpha\beta + 2\beta^3 + 4\alpha\beta^2}}} \tag{8.9}$$

8.9 Noise Performance [4]

8.9.1 Jitter Accumulation in Digital PLLs

Digital PLLs have been used for slave clocks in synchronous networks as described in Chapter 7. They offer additional functions such as noise

reduction and frequency holdover. While the digital PLL is the most suitable choice for meeting the synchronization requirements, it should be considered that many PLL types may be installed in a synchronous system, and that serial PLL connections, which can degrade the synchronization characteristics, do exist. Digital PLLs in slave clock applications have to receive their input signals from the transmission path. These signals include phase variations called jitter and wander (see Chapter 13). Generally, jitter is phase variation whose frequency components are above 10 Hz, and wander is slower than 10 Hz. The source of jitter is the entire network. Every element in the network contributes to jitter. Wander is mainly generated by the transmission paths. The momentary PLL phase response can be seen as a kind of wander since its time constant can vary from several seconds to thousands of seconds (i.e., down to very low frequencies). In a synchronous system, the master clock, slave clocks, and transmission paths are sources of phase variation.

The digital PLL can reduce frequency components above its cutoff frequency, but lower ones can appear at the output. A PLL might even amplify phase variations within a certain frequency region. As a worst case, let us assume that the frequency of the input signal equals the one at which the PLL transfer function offers maximum gain. The PLL output phase variation J_1, is

$$J_1^2 = \left(J_0^2 + J_p^2\right)G_{peak}^2 \tag{8.10}$$

where J_0 is the phase variation in the master clock, J_p is phase variation in the transmission path, and G_{peak} is the PLL peak gain from (8.9).

Equation (8.10) expresses the phase variation in the PLL output if it is directly connected to the master clock via a transmission path. If a series of slave clocks is considered, the phase variation in the last digital PLL is

$$J_n^2 = \left(J_0^2 + J_p^2\right)G_{peak}^{2n} + \sum_{k=2}^{n} J_p^2 \prod_{\ell=k}^{n} G_{peak}^2 \tag{8.11}$$

8.9.2 Short-Term Frequency Stability of a PLL

The method in Section 8.9.1 is a rough estimation, but it still provides a good idea of the amount of phase variation. If the optimum PLL time constant (cutoff frequency) has to be defined, however, it is necessary to analyze the noise spectrum in terms of the input and output phase variation of the

PLL. The power spectral density in the frequency domain and the short-term frequency stability $\sigma_y(\tau)$ in the time domain (see Chapter 13) are effective in this task. Let us apply the short-term frequency stability to obtain the optimum time constant.

The short-term frequency characteristics of the input signal and that of the controlled oscillator are necessary data for a time constant optimization. The practical ranges of these two are shown in Figure 8.18 as gray areas. The stability of the input signal includes the influence of the transmission path and the characteristics of the higher layer slave clock and the master clock. The effect of transmission path, which includes the SDH system elements, dominates over observation times from 10^0 to 10^5 seconds. The master clock characteristic determines the frequency stability over observation times from 10^5 to 10^7 seconds. For observation times longer than 10^7 seconds, the region is determined by the phase variation of transmission lines, (i.e., wander). The short-term performance of the controlled oscillator is determined by the oscillator type. For observation times from 10^4 to 10^7 seconds the region is influenced by the oscillator temperature characteristic. Since this characteristic varies due to the environment, in practice it shows some spread. Figure 8.18 indicates a typical rubidium oscillator characteristic.

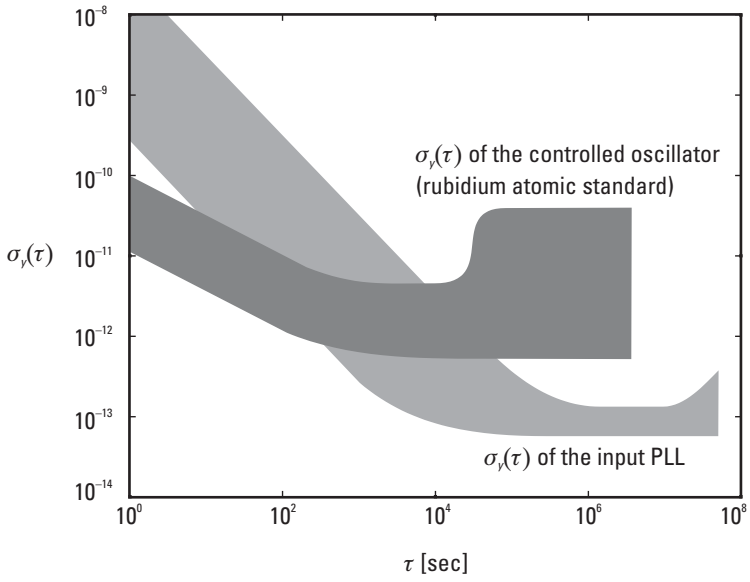


Figure 8.18 Comparison of the short-term frequency stability of the PLL input and the controlled oscillator output.

8.9.3 Noise Influence on the PLL Transfer Function

Figure 8.18 shows examples of the short-term frequency stability of the input signal and that of the controlled oscillator. These characteristics are further formed due to PLL feedback, which is illustrated in Figures 8.19 and 8.20. Figure 8.19 shows the short-term frequency stability of the input signal after it is affected by the PLL transfer function. High-frequency noise components, which appear in the short-term frequency stability profile for short observation times, are suppressed by the PLL lowpass characteristic. Figure 8.20 shows the short-term frequency stability of the controlled oscillator output after this is influenced by the PLL transfer function. In the internal loop, phase variations faster than the PLL time constant appear in the oscillator output since the PLL cannot follow them. The noise impact of the controlled oscillator depends on the highpass characteristic of the PLL. The short-term frequency stability shown in Figure 8.20 is reduced with large observation times.

8.9.4 Optimum Time Constant

The total PLL short-term frequency stability is obtained by adding the respective characteristics of the input signal, the controlled oscillator, and the

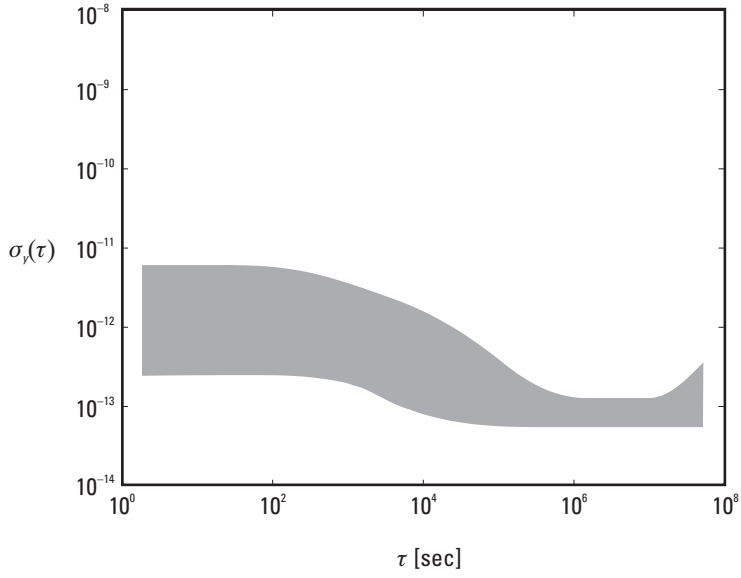


Figure 8.19 Short-term frequency stability of the PLL input as affected by the PLL transfer function.

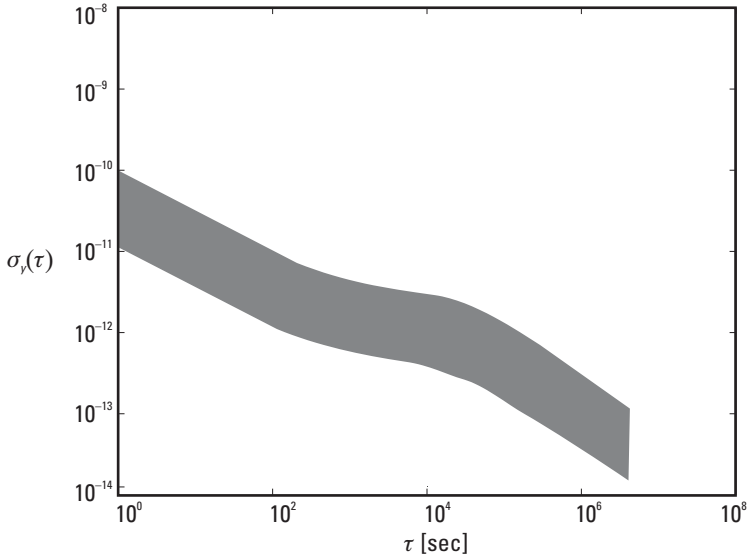


Figure 8.20 Short-term frequency stability of the controlled oscillator (rubidium atomic standard) as affected by the PLL transfer function.

feedback loop itself. It varies with the PLL time constant, which must be carefully defined for optimum performance. The best time constant is the point at which the two curves cross. That is because the better of the two characteristics can be selected on either side of the crossing point. For example, Figure 8.18 indicates that observation times of 10^2 to 10^4 seconds yield the optimum time constant. Of course, this value depends on the short-term frequency stabilities of the input and the controlled oscillator. The characteristics shown in Figure 8.18 are typical of current SDH systems. The short-term frequency stability of the controlled oscillator depends on the selected type. If a small quartz crystal oscillator such as a TCXO is chosen, its temperature characteristics are more important than the steady state short-term stability. The temperature-related frequency error of most rubidium clocks is on the order of 10^{-10} and around 10^{-6} in TCXOs for a temperature range of 0 to 50° C. When such a TCXO is used, the short-term frequency stability of the PLL output can be optimized by loop time constants shorter than those described earlier. Figure 8.21 shows the short-term frequency stability of a PLL output when the time constant is 10^3 seconds. Figure 8.21 illustrates that the total characteristic is a combination of the controlled oscillator output characteristic (region I) and the PLL input characteristic (region II). The

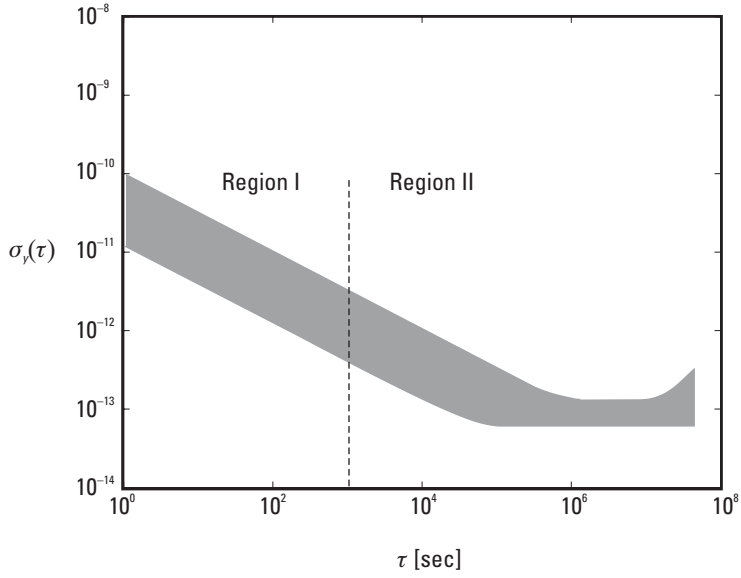


Figure 8.21 Short-term frequency stability of the PLL output as a combination of the controlled oscillator output characteristic (region I) and the PLL input characteristic (region II).

crossing point is also influenced by the PLL transient response. When the loop gain is too high, the short-term frequency stability is degraded by switching between the two regions. Selecting a ζ-value greater than 10 generally improves switching performance.

8.10 Actions Against Input Signal Problems

8.10.1 Input Phase Jump

The general role of a PLL is to follow the input phase accurately, but this does not always give the best result. The phase movements of the reference clock (i.e., the PLL output phase variations) influence the performance of synchronous systems. A synchronous digital processing system having memory assumes that the transmitted signal phase variation lies within some limits. Phase movements exceeding these limits cause errors due to the deletion or duplication of data. Minimal output phase variations in the slave clock PLLs help to avoid this problem.

For instance, when a transmission line fails, the transmission path might be changed. The transmission delay, and thus the input phase, might change accordingly as shown in Figure 8.22(a). In this case, the frequency of

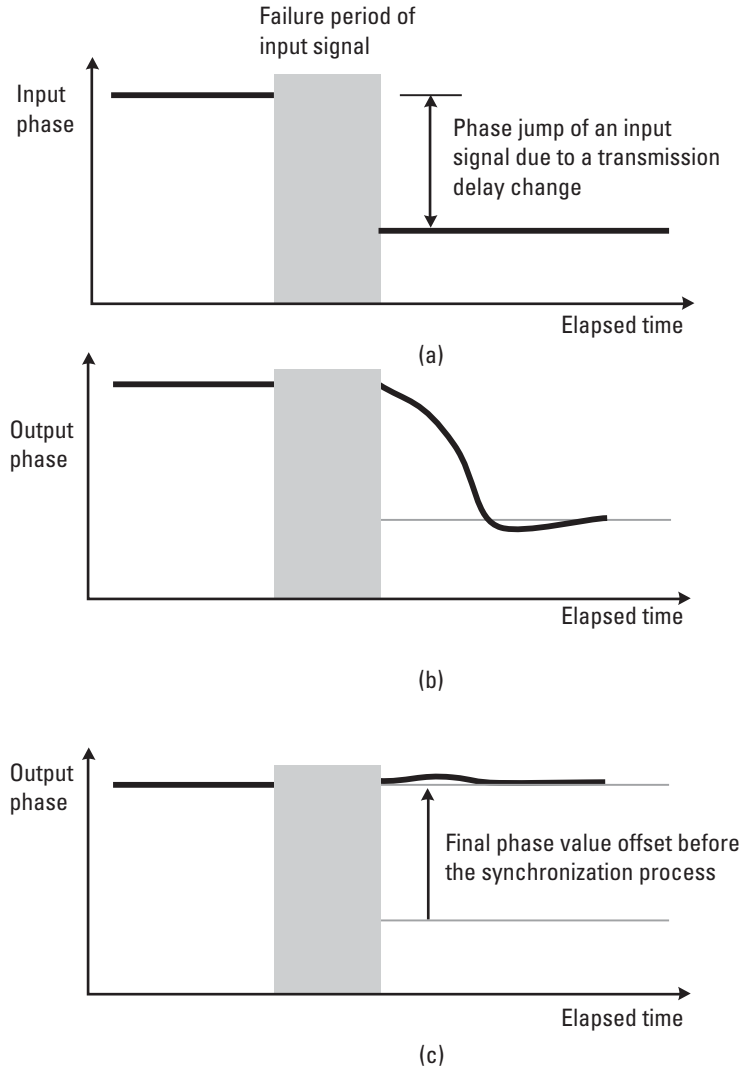


Figure 8.22 Output phase movement with/without the final phase value offset method after input signal failure: (a) input signal change, (b) conventional phase synchronization process, and (c) phase synchronization process with the final phase value offset method.

the recovered input signal is usually the same as that of the original reference clock, assuming that the higher layer slave clock (which generates the reference clock) does not fail. Examples of the input phase jump and the related transient response in the PLL output are shown in Figure 8.22(b). In analog PLLs, the output phase is usually stabilized to the point at which there is no phase difference. The phase difference between the input signal and the controlled oscillator output is held to zero in the usual phase difference comparison characteristic as shown in Figure 8.23(a). That is a reason why the phase jump occurs at the PLL output as illustrated in Figure 8.22(b).

When the PLL is used to synchronize frequency, however, it is meaningless to say that the final value in the phase synchronization process is set to zero. The phase difference detection characteristic of the phase comparator can be shifted as shown in Figure 8.23(b) so that the final phase is synchronized to a point other than zero. In digital PLLs, the final stabilized phase can be set freely to any point within the input phase difference domain by the processing software. In particular, the digital processing shown in Figure 8.17 enables this within one phase comparison period. This operation enables the digital PLL to synchronize to the input signal without any phase jump as shown in Figure 8.22(c). In Figure 8.23, the final phase stabilized value is offset in the negative direction. This reduces the phase transient time and minimizes the output phase variation.

The other approach for absorbing the initial phase difference is to directly control signal phase. In the example illustrated in Figure 8.24, the

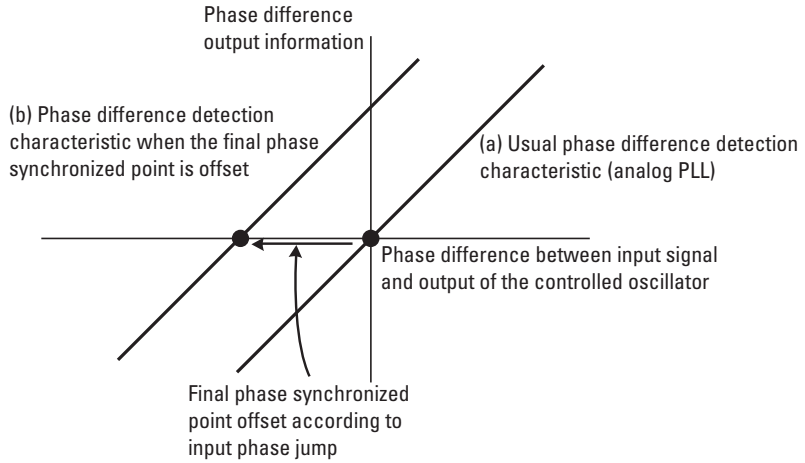


Figure 8.23 Effect of offsetting the final phase synchronized point: (a) usual and (b) offsetting phase difference detection characteristics.

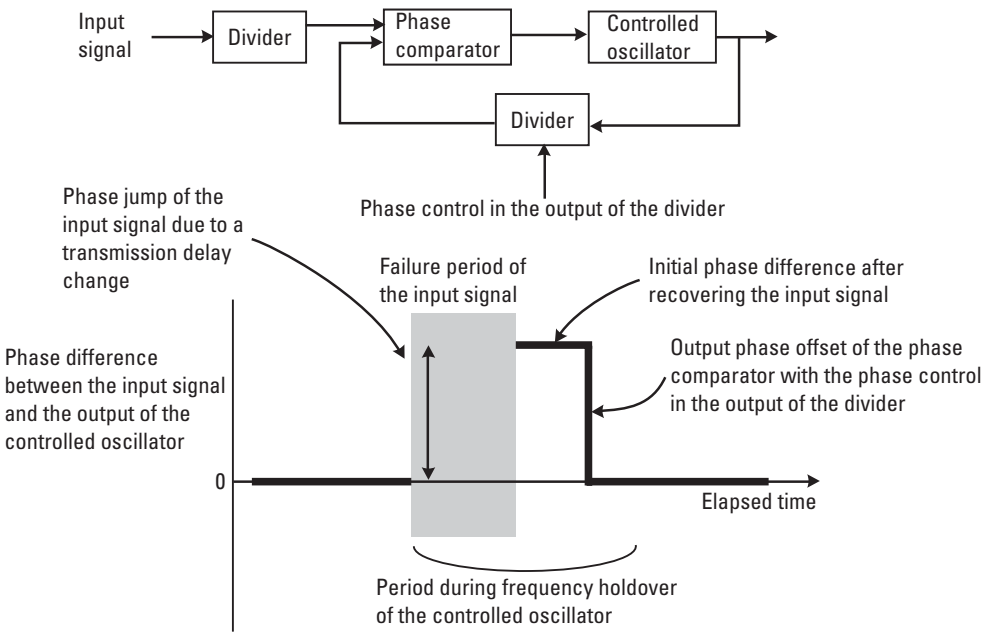


Figure 8.24 Absorbing a phase jump through phase control in the divider output.

feedback signal from the controlled oscillator is steered. The divider reset action adjusts its signal phase. The resolution is determined by the input frequency of the divider. This method is suitable for a PLL whose phase comparison frequency is lower than the frequencies of its input and/or output. The initial phase difference in the comparator is set directly to zero by controlling the output phase of the frequency divider. The comparator output phase can be offset. If the oscillator output frequency can be maintained after link failure by holdover (see Section 8.10.2) until the divider output phase is adjusted, the phase variation during signal recovery can be masked. The phase change outside the transient period can be made negligible by frequency holdover. Phase offset processing might cause unstable PLL operation if the phase jump is accompanied by an undetected frequency change. Although this method is difficult to implement in networks and complicated systems, it is effective if the PLL has a long time constant as shown in Section 8.9.4.

8.10.2 Frequency Holdover

In analog PLLs, the output voltage of the loop filter strongly changes when the input signal fails. The voltage can be held for a period that corresponds to the PLL time constant, which might be 1 to 10 seconds or less. In a digital PLL, on the other hand, the oscillator frequency can be held by freezing the control until input recovery. During this time interval the output frequency follows the oscillator characteristics. The output phase-time change over time $T_o(t)$ is

$$T_o(t) = T_0 + y_{DCO} \bullet t + \frac{1}{2} y_{drift} \bullet t^2 \tag{8.12}$$

where T_0 is the phase difference at holdover start, y_{DCO} is the oscillator frequency control error, and y_{drift} is the oscillator frequency drift in units of 1/sec. In a digital PLL that uses a rubidium oscillator with a control resolution of 10^{-12}, the output phase is within 1 μs after 24 hours.

References

[1] TOYOCOM Co., Ltd.: TCO-5828.

[2] Vectron International: CO-6000 Series.

[3] FUJITSU Ltd.: FC-6041A.

[4] Kihara M., "Performance Aspects of Reference Clock Distribution for Evolving Digital Networks," *IEEE Communication Magazine*, Vol. 4, No. 4, 1989, pp. 24–34.

9

PLL LSI

9.1 PLL in Transmission Interfaces

One network may use many clocks at different frequencies. PLLs appear in clocks, components, and elements of the network. They synchronize the network and supply synchronized frequencies and timing to individual elements. While networks can still use conventional PLLs, the application of PLL LSIs at transmission interfaces has begun. They offer many benefits including low cost in the hyper-competitive telecommunications field. Since the number of transmission interfaces is large, LSIs used in the interfaces are essential for reducing the cost of the whole network, which stimulates development. The transmission interface can, except for the optical devices, be realized as a small number of LSIs if we use a PLL LSI. The clock generation and recovery functions (see Chapter 7) are parts of the transmission interface. LSIs aiming towards these functions are typical research goals, although the current devices are still immature. The configurations of the fully integrated clock recovery circuit and clock generation circuit are described in this chapter.

9.2 Basic Configuration of Transmission Clock Recovery Circuit

When a digital signal is transmitted to another place, it is necessary to minimize the waveform deterioration that occurs during transmission and then regenerate the original signal waveform. These functions are placed at the

input of the receiver system. An example of the receiver circuit configuration is shown in Figure 9.1. The transmission information regeneration circuit accurately reconstructs the information even if transmitted over long distances. The receiving node must use a clock that is identical to the one in the sending node. The clock recovery circuit extracts the clock component from the transmitted signal. This clock is used by the signal regeneration discrimination decision circuit, which is usually composed of a flip-flop, to output a digital signal identical to the original one. The clock transmitted with the data is converted into a separate, extracted clock. The noise accumulated during transmission is reduced.

The role of the clock recovery circuit is to recover just the clock (i.e., no noise) to allow data discrimination decision and regeneration. Although noise suppression is needed, it is also necessary to extract the clock component regardless of the data pattern of 0s and 1s. The PLL shown in Figure 9.2 has been utilized as a clock recovery circuit since it can regenerate the clock even if the pattern includes consecutive 0s or 1s. The steady increase in the data rate and the pressure to lower the cost of the transmission interface also encourage the use of a PLL instead of a resonator with discrete components. The PLL of Figure 9.2 is usually placed in the signal regeneration discrimination decision circuit and is called the clock data recovery circuit (CDR).

9.3 Signal Receiving and Regeneration Circuit Using PLL

9.3.1 PLL Input Signal

The clock signal extraction circuit in Figure 9.2 is composed of general PLL components. When a random NRZ signal is transmitted, the receiving circuit needs an additional function. Since here the basic clock frequency has low power, it is difficult for the PLL to synchronize to it. The easiest way to get the clock component is to double the NRZ signal. In Figure 9.3 the

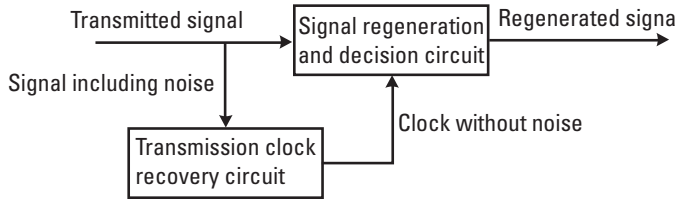


Figure 9.1 Signal regeneration circuit at the receiving site.

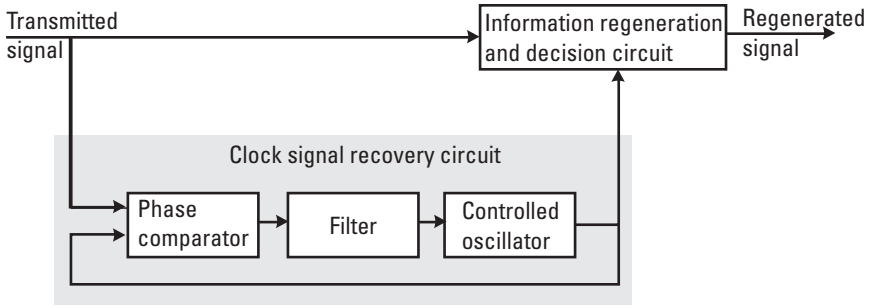


Figure 9.2 Details of the regeneration circuit in the receiver.

input signal is first passed through a circuit to detect the signal level change point, and then the signal, which has been modified by the detector, is input to the PLL. This clock recovery circuit shown in Figure 9.3 combines a phase-locked loop with a frequency-locked loop to ensure synchronization. The controlled oscillator is steered by phase error information obtained from both the phase and the frequency loops. The frequency stability of a controlled oscillator implemented as an LSI circuit is quite bad compared to a quartz oscillator. The output frequency of the controlled oscillator offsets far from the nominal value due to circuit condition, frequency drift and temperature. Combining the phase and frequency loops is a very effective way of ensuring synchronization [1–3].

A typical signal level transient point detector is shown in Figure 9.4 [1]. This detector combines an exclusive OR (XOR) circuit with a delay circuit.

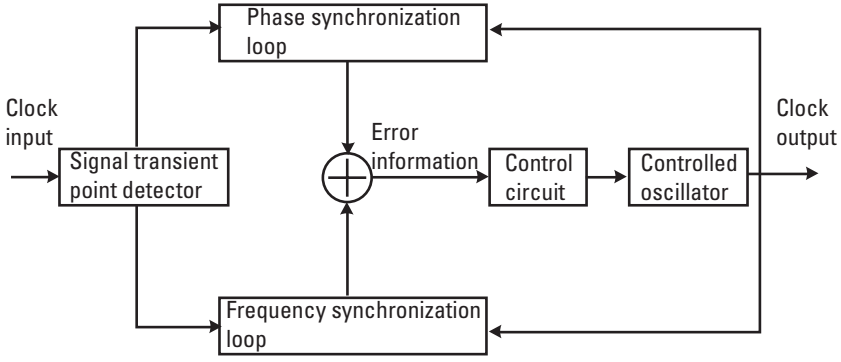


Figure 9.3 Clock signal recovery circuit.

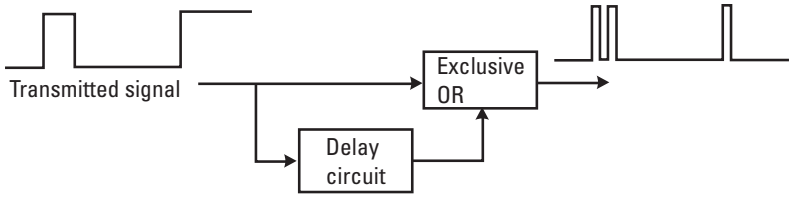


Figure 9.4 Signal transient detection circuit. (*After:* [1].)

When the signal level changes, a pulse is generated. This yields a clock whose period corresponds to that of the NRZ signal.

9.3.2 Phase and Frequency Comparison Circuits

Figure 9.5 shows a clock signal recovery circuit that uses both phase and frequency comparison circuits [2]. Orthogonal output signals are generated in the controlled oscillator. These signals are fed to mixers and differential frequencies are obtained at the mixer outputs. The differential frequency signal for 0 degrees passes a gradient detection circuit, and then goes to another mixer with the differential frequency signal for 90 degrees. A differential frequency signal that is the combination of the two differential signals is finally generated. This differential frequency signal clearly expresses the difference between the input frequency and the frequency of the controlled oscillator. The function of the frequency lock loop can be achieved by returning the differential signal to the controlled oscillator. The differential frequency signal for 90 degrees is added to the signal forming the frequency lock loop and

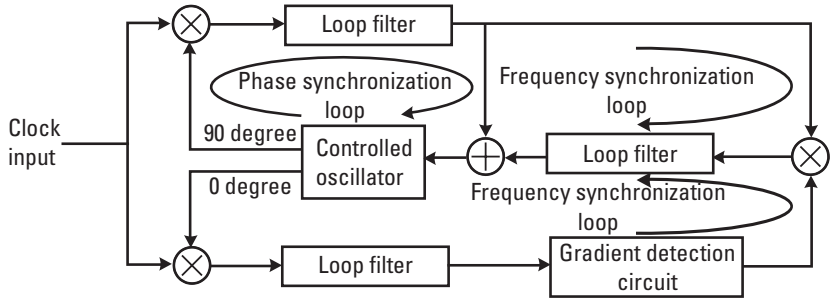


Figure 9.5 Clock signal recovery circuit with phase and frequency comparison blocks. (*After:* [2].)

is returned to the controlled oscillator. This completes the phase-frequency locking loop.

9.3.3 Frequency Holding Function

The clock recovery circuit does not operate if the transmitted signal contains a continuous series of 0s or 1s. There are two methods to eliminate this problem. One is to choose a PLL with a large time constant. Even if the PLL input signal disappears, the output phase does not change rapidly. It is advantageous to suppress the noise included in the input signal since this makes the time constant longer. This, however, increases the time to synchronize. If the clock recovery circuit must respond quickly, this method is not so suitable. The other approach is to add a function to the PLL that holds the frequency of the controlled oscillator as shown in Figure 9.6. This function is called holdover. The control information stored either in digital or analog form immediately before losing clock is used. This information is input to the controlled oscillator until the clock is reacquired. This condition is a type of free-running state [4].

9.3.4 Expansion of Synchronization Range

Figure 9.7 shows a PLL-based clock generation circuit whose maximum output frequency is 50 MHz [5]. This PLL consists of three different function blocks. One is a low loop gain phase-locked loop that reduces input noise. Another is a frequency-locked loop that expands the capture range. The other is a stabilization loop for the controlled oscillator that suppresses the frequency variation induced by environmental changes such as temperature.

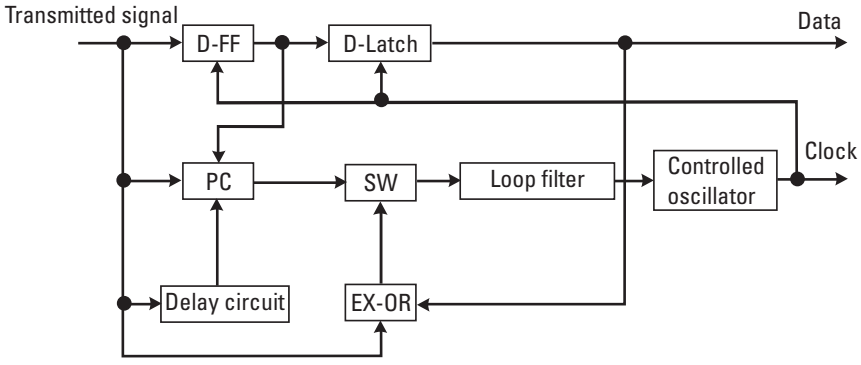


Figure 9.6 Clock signal recovery circuit with a control voltage hold circuit. (*After:* [4].)

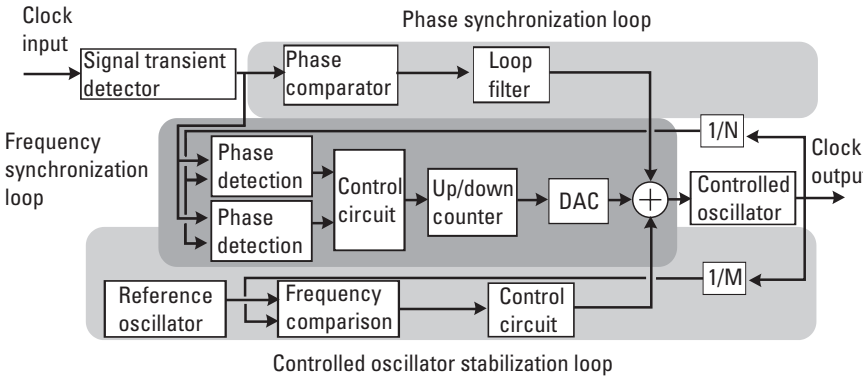


Figure 9.7 Clock signal recovery circuit with a controlled oscillator stabilization circuit. (*After:* [5].)

This configuration enables the PLL to synchronize to input signals from 1 to 50 MHz. Most parts of the PLL can be miniaturized and fabricated using the CMOS process. If the controlled oscillator is stable enough, the respective stabilization loop is not necessary. However, this loop assists synchronization also when the oscillator frequency changes or drifts.

9.4 CDR Based on the PLL

The most popular system, the SDH system, used in the current network offers the transmission path frequencies of STM-1 (155.52 Mbps), STM-4 (622.08 Mbps), STM-16 (2,488.32 Mbps), and STM-64 (9,953.28 Mbps). Commercial ICs that realize clock recovery for each of these transmission frequencies are available. CDRs for 2,488.32 Mbps and for 9,953.28 Mbps have been released by several companies [6–8].

9.5 Transmission Clock Generation Circuit

The circuit configuration of the transmission clock generation circuit is basically the same as the clock recovery circuit. The transmission path clock can be generated by using a PLL circuit. The difference is that the input comes from the internal clock generated within the system. In SDH systems, this clock is synchronized to a reference clock distributed across the network and is used as the input clock of the transmission path clock. Commercial clock

generation ICs have output frequencies of 2,488.32 MHz and 9,953.28 MHz for the input frequencies of 155.52 MHz and 622.08 MHz, respectively [9, 10]. Since the input signal jitter is small in the transmission clock generation circuit compared to that of the CDR, reducing the time constant of the PLL below that of the CDR improves the PLL short-term frequency stability. However, there is a lower limit to the time constant since the PLL loop gain falls if the input frequency is too low. As a consequence of this, the time constant of the PLL increases.

9.6 PLL Circuits in Wireless Systems

The basic features of common PLL concepts are also applicable to circuits used in different radio devices. The emphasis, however, is partly on different aspects. Besides frequency accuracy we have to consider spectral purity and settling time as well as frequency step size. The spectrum is evaluated both close to the wanted carrier and at its harmonics and the applied limits are very tight, down to –90 dBc due to interference rejection. In portable devices the overall power consumption is also of importance. These elements have an impact on the practical PLL realization as discrete circuits or as LSIs.

Up to 2.5 GHz, it is feasible to combine the controlled oscillator, phase comparator, and divider circuits as a single IC chip. Generally, the active parts of the loop filter are included as well, and the user only needs to add some filter capacitors, which are difficult to implement on the conventional silicon or GaAs process. Crystal oscillators are often added separately but certain LSIs incorporate their active elements as well so that the user only has to connect a suitable quartz crystal. These commercial PLL LSIs, however, can often offer a restricted frequency resolution because of the fact that they have been customized for particular networks that have a predefined channel spacing. The other significant drawback is the low stability offered by the on-chip controlled oscillator. This must be overcome through a relatively tight loop feedback.

Both all-analog and analog-digital PLL designs can be used in radio frequency circuits. The first concept mainly utilizes double balanced mixers as phase comparators. This implies that the frequency response of the loop can be very wide if no filter is installed. Of course, such a characteristic enables very fast control actions that might be of benefit in hopping synthesizers. The mixer output is unfortunately of bipolar nature whereby most controlled oscillators require an additional DC bias. If not properly dimensioned and designed, this bias may severely deteriorate the final output spectrum by

spurious and noise. Many commercial RF PLL chips include the mixer, but the user must often provide the filter, which can be, for example, a SAW device.

Analog-digital PLLs have been the most common choice in VHF/UHF devices. Here, the controlled oscillator often is of the conventional analog type, and a simple level detector connected after a power divider is used to convert its output into a square wave. The required frequency step is obtained by dividing the square wave frequency in a simple digital flip-flop chain. The difficulty in this approach is the leakage of the final phase comparison frequency and its harmonic multiples to the RF output. These semi-digital PLLs are available as ready-made LSI packages.

The very high microwave and millimeter wave communication devices often use some prescaling in front of or before the PLL chips. The scaling can be accomplished through harmonic downconversion or through very fast dividers. Although the requirements of spectral purity are basically important here too, the designer is often forced to allow some relief in specifications in order to secure the PLL locking. One example of PLL usage at millimeter wave frequencies is illustrated in Figure 9.8.

Frequency hopping radio systems offer a special challenge to PLL ICs. Two main applications exist currently. Some mobile networks include the slow frequency hopping (SFH) feature in which the hop rate is typically much below 1 kHz. This would be possible with a well-designed single

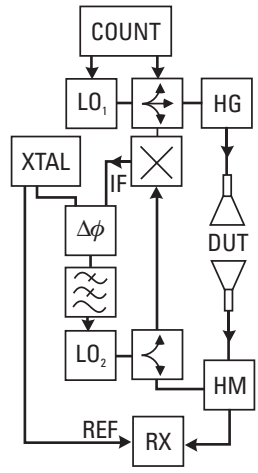


Figure 9.8 A millimeter wave test setup, which utilizes PLLs and harmonic mixing to produce very precise signals at 900 GHz.

oscillator loop, but many devices actually use a multicarrier transmission and switch the data between them. Thus the PLL IC outputs a dedicated comb spectrum. Military fast frequency hopping (FFH) tactical radios must have hop rates above 1 kHz but they can not employ the multicarrier concept due to low probability of intercept/low probability of detection (LPI/LPD) requirements. Thus, their electronics are highly complicated. Because of this they are currently being replaced by DDS units, which can already cope alone with the relatively low VHF frequencies.

References

[1] Ransijn, H., and P. O'Connor, "A PLL-Based 2.5-Gb/s GaAs Clock and Data Regenerator IC," *IEEE Journal of Solid-State Circuits*, Vol. 26, No. 10, 1991, pp.1345–1353.

[2] Razavi, B., "A 2.5-Gb/s 15-mW Clock Recovery Circuit," *IEEE Journal of Solid-State Circuits*, Vol. 31, No. 4, 1996, pp. 472–480.

[3] Bellisio, J. A., "A New Phase-Locked Timing Recovery Method for Digital Regenerators," *IEEE Int. Cont. Rec.*, Vol. 1, 1976, pp. 10-17–10-20.

[4] Ishihara, N., and Y. Akazawa, "A Monolithic 156 Mb/s Clock and Data Recovery PLL Circuit Using the Sample-and-Hold Technique," *IEEE Journal of Solid-State Circuits*, Vol. 29, No. 12, 1994, pp. 1566–1571.

[5] Pritchett, R. L., et al., "A Monolithic 1 to 50 MHz CMOS Clock Recovery and Retiming," *ESSCIRC'86*, Delft, 1986, pp. 122–124.

[6] Lucent Technologies, Clock Recovery, Data Demultiplexer, TRCV012G5.

[7] Lucent Technologies, Clock Recovery, Data Demultiplexer, TRCV0110G.

[8] NTT Electronics Corp., CDR/DEMUX Module, MOS41M.

[9] Lucent Technologies, Clock Synthesizer, Data Multiplexer, TTRN012G5.

[10] Lucent Technologies, Clock Synthesizer, Data Multiplexer, TTRN0110G.

10

Frequency Generation Systems

10.1 The Need for Frequency Generation

Internal clocks of synchronous systems often have different frequencies to the reference clock. It is necessary to convert the reference clock into new frequencies without losing accuracy. A PLL is suitable for this because it performs the conversion and synchronization. A function for freely synthesizing frequencies can be implemented by a mechanism that changes the PLL configuration. Such a circuit is an indispensable element of a wireless system. Frequency accuracy, clock purity, and switching speed are key design parameters in cellular phone networks. In a cable or fiber system, on the other hand, frequency switching is not so common. A frequency generation system provides an important measurement function for design and performance verification. These devices are called synthesizers.

10.2 The Evolution of Frequency Synchronization and Generation Systems

The history of the PLL reflects that of other electronics with a steady move towards complete digitalization. The analog-digital PLL was the first type to be used in digital circuits. They are still found in many systems since they can be treated as if they were fully digital. The main goal of PLL design has been miniaturization. The phase comparator and the loop filter have been digitalized. The modern quartz crystal oscillator is so small that such a device for

cellular phones occupies less than 1 cm^3. Ring-type oscillators have also been digitalized. Modern PLLs are being realized as single-chip ICs, as described in Chapter 9.

The second trend is performance improvement. The demand for rapid and accurate phase or frequency control and for the flexible generation of waveforms is usually satisfied by sophisticated controlled oscillators. They are realized by extending and combining conventional technologies. The latest demands placed on clocks, however, are accelerating the development of new generation mechanisms. The trend is to fully control the signal by using digital circuits rather than trying to improve the performance of conventional control methods. This new technique is called direct digital synthesizer (DDS). DDS allows us to build synthesizers that can generate variable and multiple frequencies and arbitrary waveforms. The term DDS has also been used to refer to a method of generating the required frequency by dividing and adding. In this book, however, this term refers to the technique in which the signal is synthesized by completely digital circuits.

10.3 Ideal Clock Generation Using Digital Signal Processing

The idea of using just digital signal processing to generate arbitrary signals was first introduced in the 1960s when books on signal processing began to be published. Only two circuits are needed for this job, as indicated in Figure 10.1. The digital-to-analog conversion (DAC) part generates an equivalent analog waveform according to a digital value, which comes from the processing part. If the two parts function ideally under a high frequency clock, the controlled oscillator becomes unnecessary.

The digital modeling of the arbitrary signal happens in the processing part. Data examples of sine and sawtooth waves are shown in Figure 10.2. The sampling theorem (described in Section 3.2.2) indicates that the highest frequency that can be generated is one half of the quantization frequency, the upper limit of which is set by the processing speed. The DAC conversion part is composed of the converter and a filter that smoothes the analog signal.

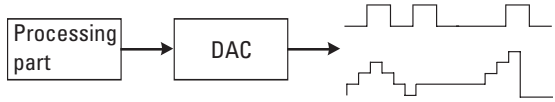


Figure 10.1 Basic circuit configuration for generating arbitrary wave signals through digital signal processing.

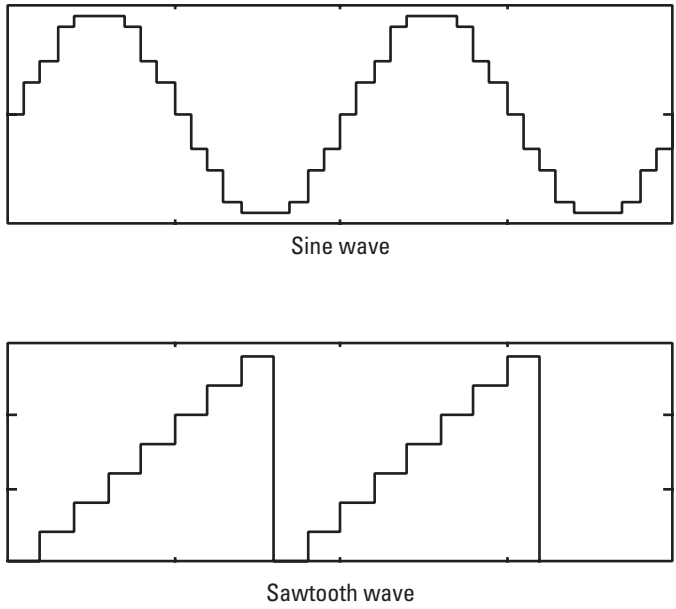


Figure 10.2 Examples of quantized signals.

The speed of the DAC determines the maximum frequency of the DDS, which is one promising application for high-speed DACs.

DDS devices fall into two classes. Both use adders and registers for processing. One type is shown in Figure 10.3 where the processed result is directly used as the DDS output. This is a suitable method if a rectangular output waveform is required. It can generate high frequency signals since it does not have a DAC. A drawback is that the level of spurious frequency components is also high. The second DDS type outputs a sine wave by placing a DAC after the processing part as shown in Figure 10.4. This is

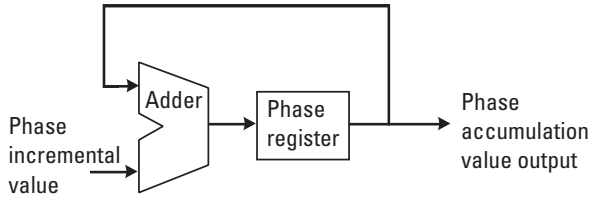


Figure 10.3 DDS configuration that directly uses the output of the processing part.

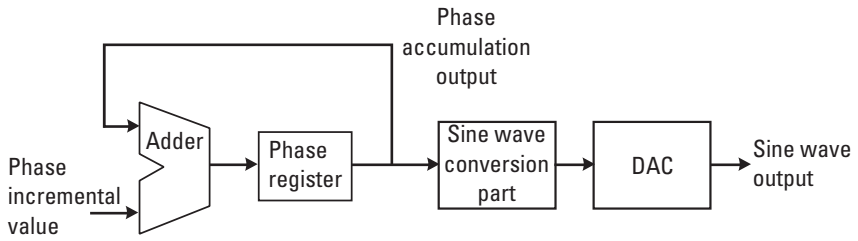


Figure 10.4 Basic DDS configuration for generating sine waves.

currently the most common DDS configuration. The first DDS with sine wave output was proposed by J. Tierney [1]. A look-up table in the conversion part memory is used to specify the amplitude values of the sine wave. The data is converted into a sinusoidal signal by the DAC.

10.4 DDS Characteristics

Because a DDS can be composed of all-digital logic circuits, its configuration is very suitable for miniaturization. DDS ICs (without a DAC) and complete single-chip ICs have been released. Several DDS variations, including those that offer frequency or phase modulation and chirping are available. Some are also used in the radio frequency part of wireless systems [2, 3].

Though the DDS seems to be an ideal frequency generation circuit, a conventional analog-digital PLL offers output signals with superior frequency purity. Because analog-digital PLLs can use quartz crystal oscillators, the required spectral purity can be achieved by selecting the appropriate oscillator. The DDS circuit configuration greatly influences its output spectrum due to the quantizing action and due to the amplitude transients in the processed output. Reduction methods are thus very important. A circuit parameter selection method and a noise rejection method have been proposed. The fundamental solution, however, is the high frequency DDS, which requires a high-speed DAC. Frequency setting resolution and switching speed are important design parameters in the DDS as well. They are being gradually improved by the continuous increase in DDS clock frequency.

References

[1] Tierney, J., M. Charles, and B. Gold, "A Digital Frequency Synthesizer," *IEEE Transactions on Audio and Electroacoustic*, Vol. AU-19, No. 1, 1971, pp. 48–57.

[2] Analog Devices, CMOS 300 MHz Complete-DDS, AD9852.

[3] Texas Instruments, Single-Chip RF Transceiver, TRF6900 10-5.

11

DDS Circuit Configuration and Characteristics

DDS allows us to build synthesizers that can generate variable and multiple frequencies and arbitrary waveforms as mentioned in Chapter 10. Theoretically all signals can be generated by the DDS concept—only limited by the circuit processing speed and the sampling theorem. However, practical circuit configurations and design parameters determine the DDS output frequency and signal characteristics. In particular, output signal purity should be considered sufficiently. Undesired frequency components are inevitable in digital processing and they degrade signal purity. This chapter describes the relationships between DDS circuit configurations and their output characteristics such as the spurious generation mechanism.

11.1 Basic Parameters

11.1.1 Output Frequency

The DDS output is determined by the digital data generated by the processing part based on a system clock. According to Shannon's sampling theorem, the maximum output frequency of the DDS cannot exceed 50% of the system clock, but the practical limit is empirically supposed to be around 33% of the system clock or even lower. Here, the spurious characteristics in the output are a major concern.

Figure 11.1 shows the circuit that directly uses the output of the processing part to generate the sawtooth waveform shown in Figure 11.2(a). The output is accumulated by the adder over each clock and is reset when the accumulated value reaches the maximum of the adder. The phase accumulation speed for one clock is determined by the phase increment θ_i of the adder input. Assuming a 4-bit adder and phase register and a phase increment value of 3, each bit pattern changes according to the clock count as shown in Table 11.1. In Table 11.1, the binary value of the phase accumulation value ranges from 0000 to 1111 (0 to 15 in decimal notation), since the processing part uses 4-bit data. If the clock count starts from 0, the maximum phase accumulation value is reached when the count is 5, the adder is reset once when the count reaches 6, and then the phase accumulation value becomes 0010. The sawtooth is output by repeating this operation as shown in Figure 11.3. If the reset signal due to adder overflow can be used, a square waveform can be generated as shown in Figure 11.2(b).

The phase accumulation value output in decimal value changes as shown in Figure 11.4. One cycle of this sawtooth waveform is

$$N_p = 2^A / \theta_i \tag{11.1}$$

where A is the adder binary width. Since the sawtooth wave repeats through every clock count N_p, the frequency f_o of this signal is

$$f_o = \frac{\theta_i}{2^A} \bullet f_{clk} \tag{11.2}$$

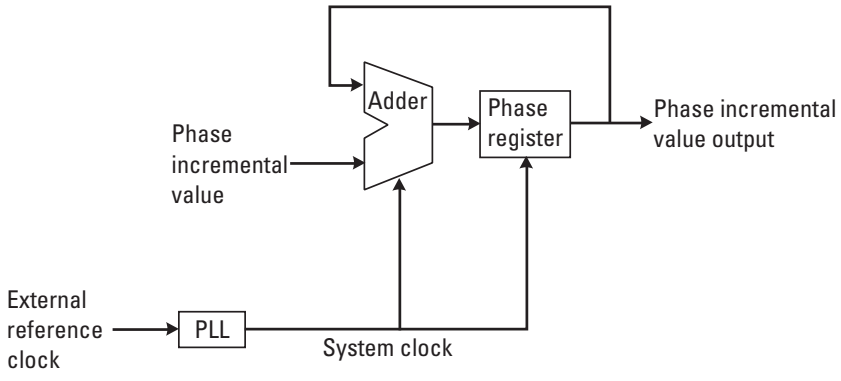


Figure 11.1 DDS configuration that directly uses the output of processing part.

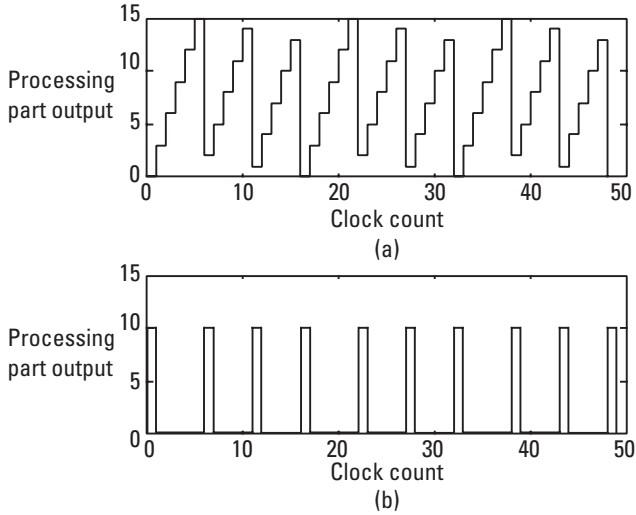


Figure 11.2 Output waves in the processing part: (a) direct output and (b) square wave output using the overflow signal.

Table 11.1
Phase Accumulation Value Change

	Phase Accumulation Value Output				
Clock Count	**Bit 3**	**Bit 2**	**Bit 1**	**Bit 0**	**Decimal**
0	0	0	0	0	0
1	0	0	1	1	3
2	0	1	1	0	6
3	1	0	0	1	9
4	1	1	0	0	12
5	1	1	1	1	15
6	0	0	1	0	2
7	0	1	0	1	5
8	1	0	0	0	8
9	1	0	1	1	11

Table 11.1 (continued).

	Phase Accumulation Value Output				
Clock Count	**Bit 3**	**Bit 2**	**Bit 1**	**Bit 0**	**Decimal**
10	1	1	1	0	14
11	0	0	0	1	1
12	0	1	0	0	4
13	0	1	1	1	7
14	1	0	1	0	10
15	1	1	0	1	13
16	0	0	0	0	0
17	0	0	1	1	3
18	0	1	1	0	6
19	1	0	0	1	9

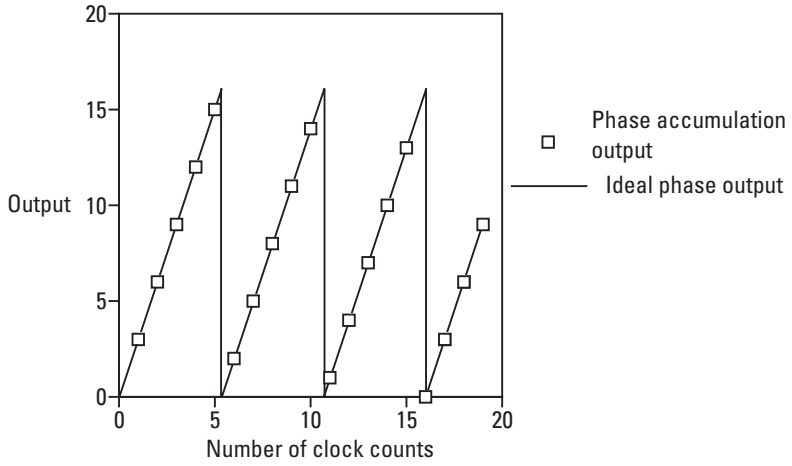


Figure 11.3 Example of the processing part output.

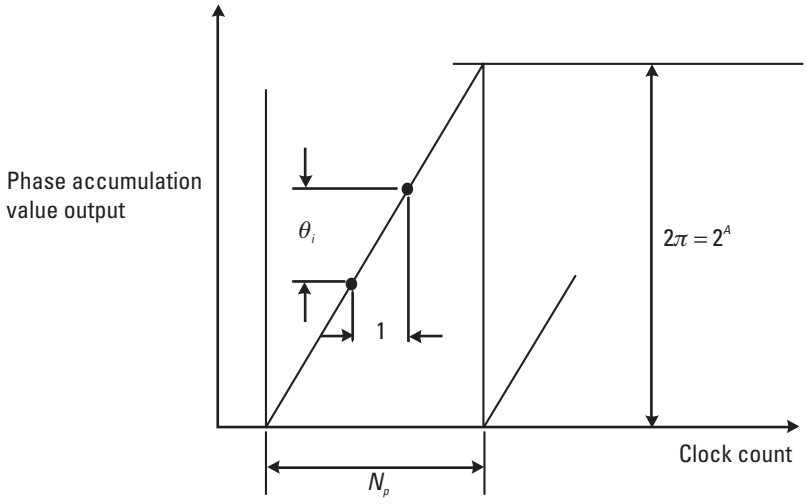


Figure 11.4 The rate of change in the phase accumulating value output.

For a high output frequency we have to raise the processing speed and the clock frequency f_{clk} if the phase increment value of the adder input and the adder binary width are fixed. When the phase increment value of the adder input exceeds the maximum value that can be handled by the adder, the value basically equals $\theta_i - 2^A$. The number of setting frequencies in the sawtooth waveform is thus $2^A - 1$.

11.1.2 Frequency Control Resolution

The frequency is changed by θ_i if the clock frequency f_{clk} and the adder binary width A are fixed. The frequency resolution Δf is obtained from (11.2) as

$$\Delta f = \frac{f_{clk}}{2^A} \tag{11.3}$$

The frequency control resolution depends on the adder binary width A and can be increased comparatively freely compared to PLLs. If we assume a clock frequency of 1 MHz and a 12-bit adder, the DDS frequency control resolution is 244 Hz.

11.2 Spectrum

DDS offers flexible output waveform control, high-speed variable frequency generation, and good frequency resolution. Since the wave is generated from digital data that attempts to emulate the target in each clock cycle, deviations from the ideal waveform exist [1]. The highest priority must be on minimizing these errors if the circuit is used to generate a sine wave. The results are evaluated in the frequency domain since the defects usually appear in the form of spurious components.

Figure 11.5 shows an example of a sine wave DDS configuration. The phase accumulation output is converted into a sine wave through the combination of the sine wave conversion part and a DAC. The sine wave memory holds a conversion table, which describes the correspondence between the

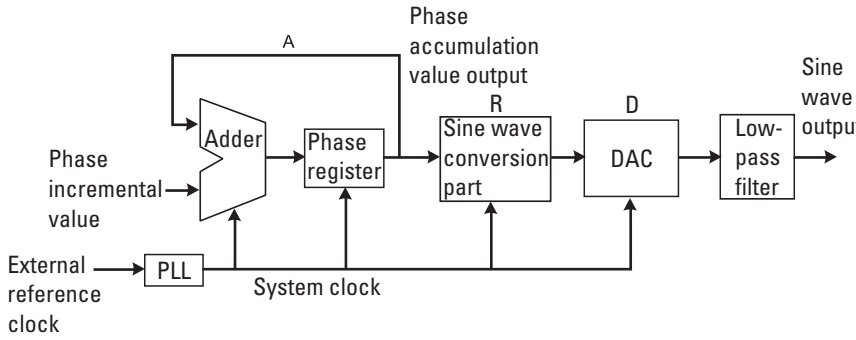


Figure 11.5 Basic DDS configuration that generates sine waves.

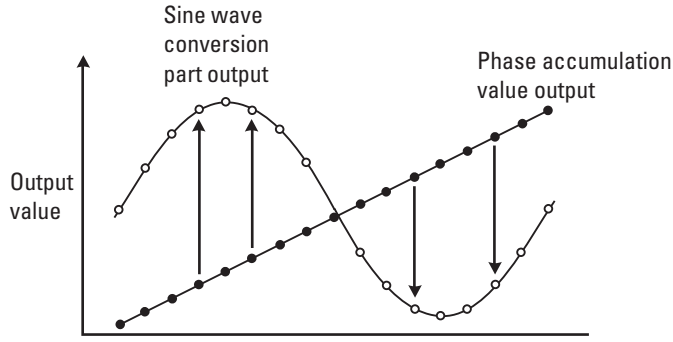


Figure 11.6 Conversion of a sawtooth wave into a sine wave in the sine wave conversion part.

phase and amplitude of the target sine wave. The sine wave conversion part can convert the phase accumulation output, which is a sawtooth as shown in Figure 11.6, into a digital value that corresponds to the amplitude of the sine wave, based on this conversion table. Assuming a 4-bit adder and a 4-bit phase register, and a phase increment of 3, the actual phase accumulation output and the sine wave conversion part output change are as shown in Figure 11.7. Here, the phase accumulation value having 16 levels is converted into a sine wave. Since the binary width of the adder and the phase register are small and they have a relatively high frequency compared to the clock, the sine wave conversion part output considerably differs from the ideal sine wave. This creates spurious components, which become parts of the output spectrum. The characteristics of the spurious components greatly vary with the phase increment value and the binary processing width. They are examined further by using *A*, *R*, and *D* as the binary widths for the adder, the conversion part, and the DAC, respectively.

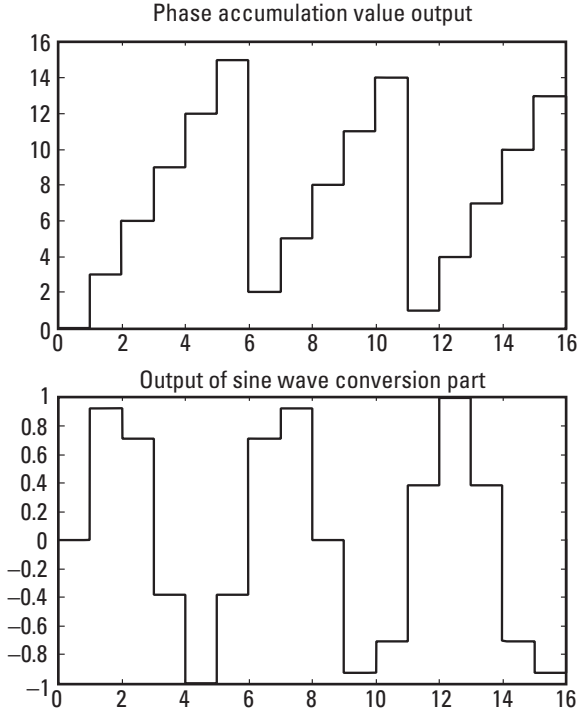


Figure 11.7 Examples of the phase accumulation value and sine wave conversion part outputs.

11.2.1 Adder Output Waveform

Since the adder output is reset after reaching some preset maximum, its operation can be expressed as

$$\begin{aligned} &\theta_a(n) = \theta_a(n-1) + \theta_i \\ &\quad \text{if } \theta_a(n) > 2^A \\ &\quad \text{then } \theta_a(n) = \theta_a(n) - 2^A \\ &\quad \text{end if} \end{aligned} \tag{11.4}$$

When θ_i is a power of 2 (2^m) in (11.4), the adder output becomes a periodic sawtooth waveform. The repetition rate P_i of the output is

$$P_i = \frac{2^A}{\theta_i} = \frac{2^A}{2^m} = 2^{A-m} \quad m = 1, 2, \ldots \tag{11.5}$$

This relationship is obvious from (11.1) and (11.2) as well. The frequency of the square wave at the adder output equals the clock frequency divided by 2^{A-m}. When an 8-bit adder is used and the phase increment value is 8, the relative frequency normalized by the clock frequency is 0.03125 (= 8/256). The output spectrum of this sawtooth wave has the frequency component A in Figure 11.8. The frequency component B in Figure 11.8 is an example of the spectrum when the phase increment value is 16. Higher

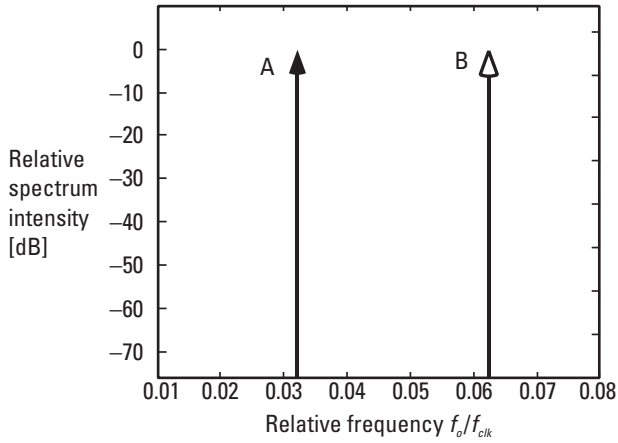


Figure 11.8 Frequency spectrum of a sawtooth wave having a precise period.

harmonics are generated since the output is not a sine wave. No spurious components other than the higher harmonics appear as shown in Figure 11.8, since these two examples have an ideal sawtooth waveform at the adder output.

Figure 11.3 indicates that the practical maximum phase accumulation value differs from the peak value of the ideal sawtooth waveform, when θ_i is not a power of 2. Even though the mean frequency is the same as the value expressed by (11.2), the instantaneous signal period shifts from $1/f_0$. The distribution range Δe_a of this timing error is

$$\Delta e_a : \frac{1}{f_{clk}} \sim \frac{(\theta_i - 1)}{f_{clk}} \tag{11.6}$$

Figure 11.9 shows the spectrum influenced by the timing error in the case when the adder is 8 bits and the phase increment value is 11. A lot of spurious components appear when the desired signal has the relative frequency of 0.043 (= 11/256). The position and amplitude of these components change with the repetition rate of the adder output. The spectrum thus varies greatly with the phase increment value.

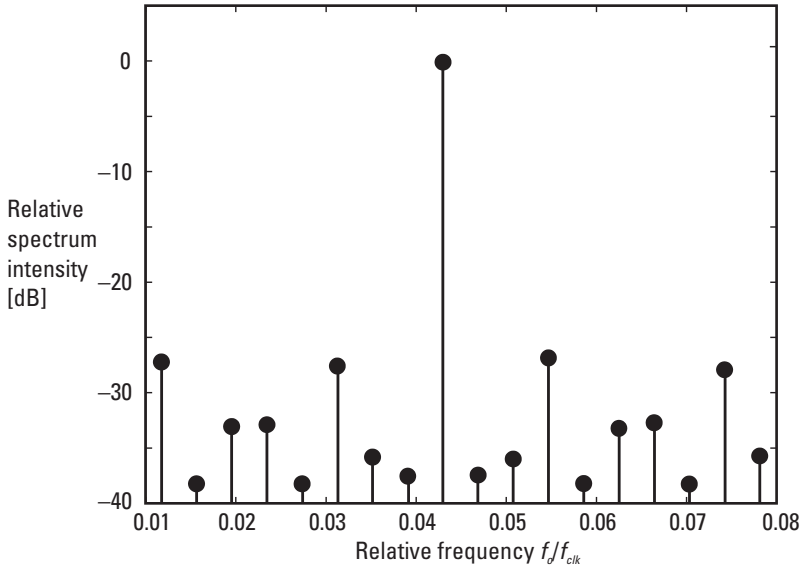


Figure 11.9 Frequency spectrum of a sawtooth wave generated with an 8-bit adder and a phase accumulation value of 11.

The repetition rate in the phase increment value of the adder output [2, 3] can be obtained from

$$P_a = \frac{2^A}{\mathrm{GCD}\left[\theta_i, 2^A\right]} \tag{11.7}$$

where GCD[n_1, n_2] represents the greatest common divisor of n_1 and n_2.

Equation (11.7) implies that a lot of other spurious components are generated around the fundamental frequency, which is represented by the mean value from (11.2). In Figure 11.9, the adder is 8 bits and the phase increment value is 11, so P_a becomes 256. This means that 128 spurious components appear within the relative frequency range from 0 to 0.5. Figure 11.10 shows the spectrum of this relative frequency domain.

11.2.2 Phase Error Due to Binary Width Limitation in the Sine Wave Conversion Part

The sine wave conversion part processes the digital value representing the sawtooth wave of the adder output into another digital value that

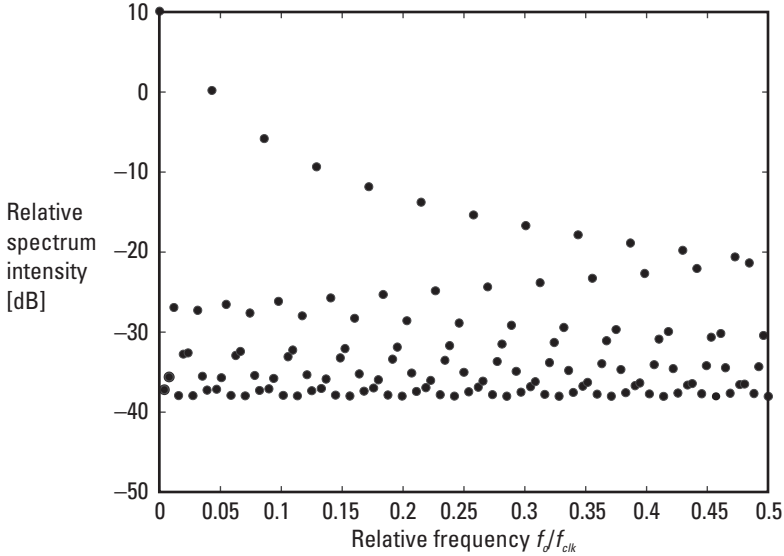


Figure 11.10 Frequency spectrum of a sawtooth wave generated with an 8-bit adder and a phase accumulation value of 11.

corresponds to the amplitude of the sine wave. The numerical precision of the conversion table in the sine wave conversion part is one of the parameters determining the DDS spurious characteristics together with the DAC binary width. Figure 11.11 shows how the adder output changes into a sine wave at the output of the sine wave conversion part. The digital adder output $\theta_a(n)$ defines the memory address of the sine wave conversion table, which corresponds to sine wave phase. The sine amplitude $S_r(n)$ at that memory address is read out.

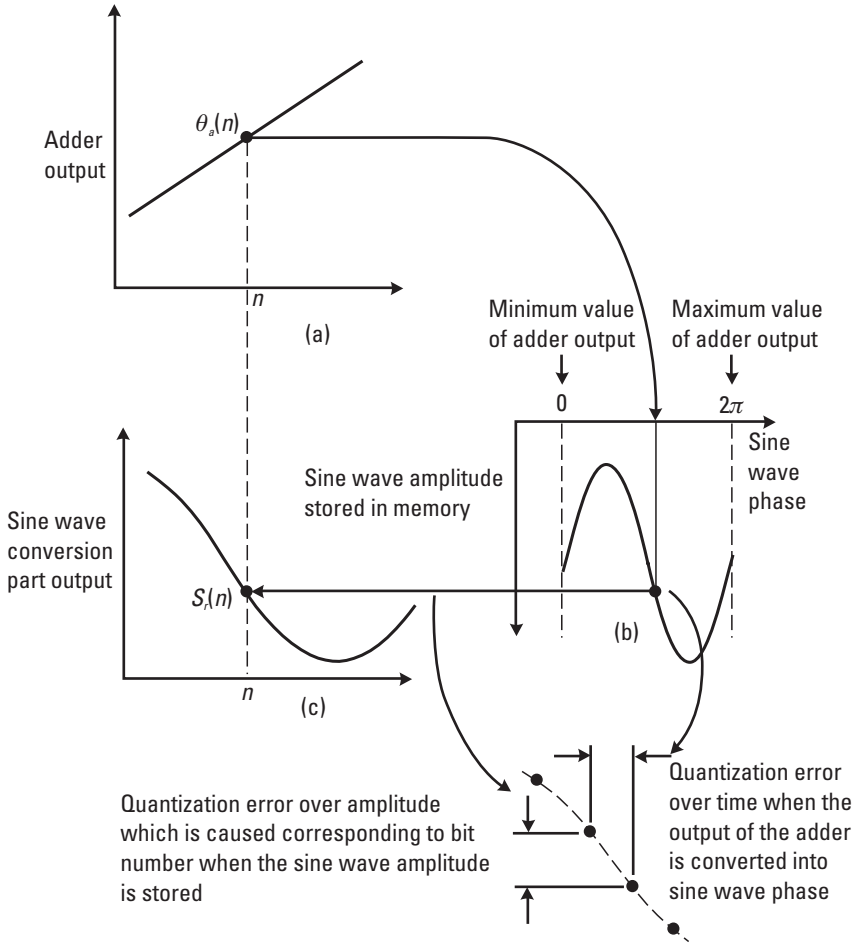


Figure 11.11 The conversion process of adder output into a sine wave: (a) adder output, (b) operation of the sine wave conversion part, and (c) DAC output.

There is no error at this stage if the conversion table can have a sufficient number of addresses—one for each output values—since the number represents sine phase resolution. If the ROM address width is limited by design requirements such as a circuit scale, a quantizing error is generated when the adder output is converted into the phase of the sine wave. Figure 11.12 illustrates examples where the binary width of 4 bits is reduced to 3 bits (dropping the LSB), and then to 2 bits. For instance, sine amplitude values of 0010 and 0011 in the adder output are both rounded to 001 when using 3 bits, and to 00 if the binary width is 2 bits.

When $\theta_a(n)$ in (11.4) is converted into sine wave data without any binary width limitation ($A = R$), the sine wave conversion part output is given by

$$S_r(n) = \sin\left(\frac{2\pi}{2^A}\theta_a(n)\right) = \sin\left(2\pi\frac{\theta_i}{2^A}n\right) \tag{11.8}$$

where $n = 0, 1, 2, \ldots$ and one step of n corresponds to time $1/f_{clk}$.

Equation (11.8) specifies the sine wave conversion where the phase data binary width is R, and the amplitude quantization resolution is infinite. The sine wave conversion part output becomes a pure sine wave with a single

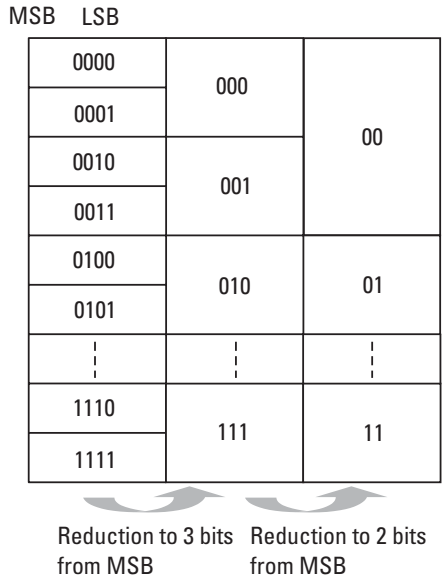


Figure 11.12 Changes in the digital value due to a reduction in the number of bits.

spectral component at the relative frequency of $\theta_i/2^A$. If the phase data in the sine wave conversion part is stored using finite binary width of R, a phase error is generated. Since the error occurs when $A > R$, if the adder binary width A is rounded to the sine wave conversion part binary width R, the output [4] can be expressed as

$$S_r(n) = \sin\left(2\pi \frac{\theta_i}{2^A} n + e_r(n)\right) + e_d(n) \tag{11.9}$$

where $e_r(n)$ is the error due to the phase data address limitation and $e_d(n)$ is the error due to the amplitude quantization. If the phase error $e_r(n)$ is small enough, (11.9) can be modified to yield

$$S_r(n) = \sin\left(2\pi \frac{\theta_i}{2^A} n\right) + e_r(n)\cos\left(2\pi \frac{\theta_i}{2^A} n\right) + e_d(n) \tag{11.10}$$

When the adder binary width is 8, the sine wave conversion part binary width is 4, and the phase increment value is 11, the digital values of each part and the deviation from the ideal waveform are as indicated in Figure 11.13. Figure 11.13(a) shows the adder output and Figure 11.13(b) shows the digital value change of the sine wave conversion part input. Since the binary width shrinks from 8 to 4, the sine wave conversion part input shown in Figure 11.13(b) is limited to the range from 0 to 15, unlike the adder output range (0 to 255) shown in Figure 11.13(a). Figure 11.13(c) shows the deviation of the limited wave from the original sawtooth. Figure 11.13(c) corresponds to the phase error $e_r(n)$ in (11.9) but as a digital value—not in radians. Note that the error is in digital values and the unit of the y axis in Figure 11.13(c) is not radians. Figure 11.13(d) indicates the digital value change in the sine wave conversion part output and shows the sine wave output from the 4-bit DAC as controlled by the conversion table.

11.2.3 Spectrum of the Sine Wave Conversion Part

The difference $e_r(n)$ between the digital value, which is truncated by the limited bit number in the input of the sine wave conversion part, and the ideal sawtooth waveform becomes a cyclic signal as found in Figure 11.13(c). Its period P_r can be obtained by the same method used to determine that of the adder output [see (11.7)], and will be

$$P_r = \frac{2^{A-R}}{\mathrm{GCD}\left[\theta_i, 2^{A-R}\right]} \tag{11.11}$$

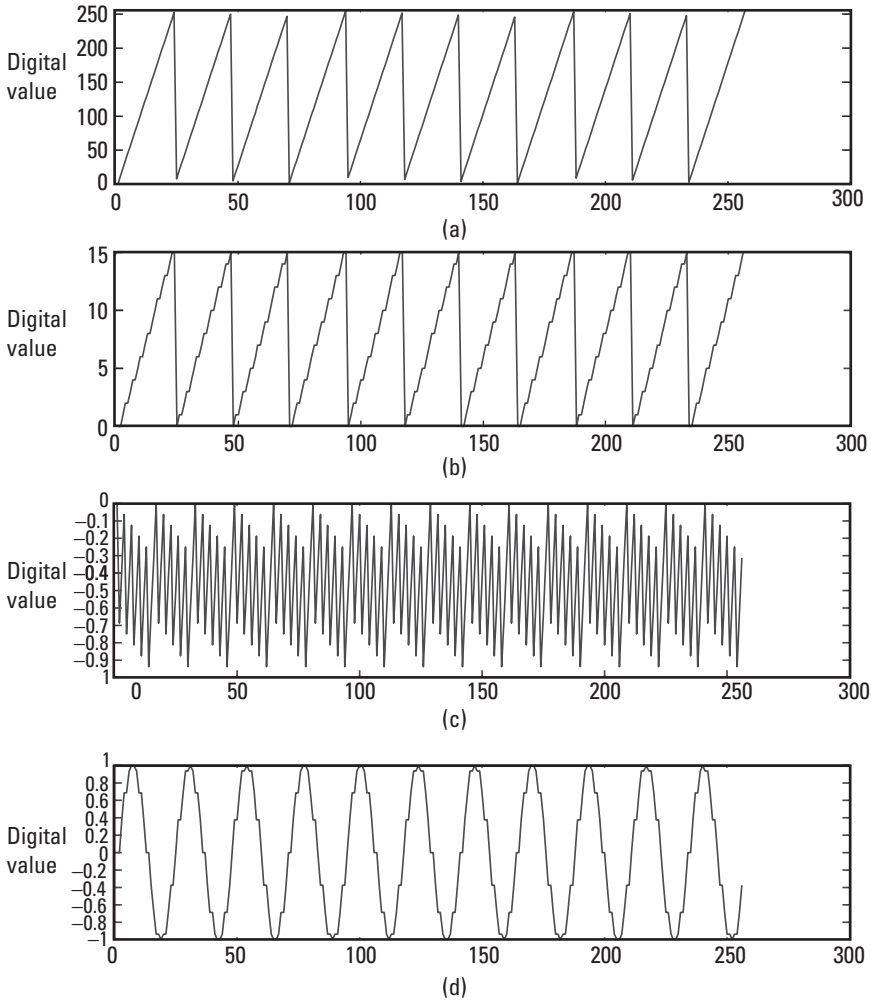


Figure 11.13 Digital value changes and bit reduction error (bit number of adder: 8, bit number of the sine wave conversion part: 4, and phase incremental value: 11). (a) adder output, (b) adder output after bit reduction in input of the sine wave conversion part, (c) error against the ideal sine wave in the sine wave conversion part, and (d) the sine wave conversion part output.

If the adder binary width is 8, the sine conversion binary width is 4, and the phase increment value is 11, P_r will be 16. The spectrum of the phase error $e_r(n)$ has 8 spurious components within the relative frequency range of 0 to 0.5. This spectrum is modified again at the output of the sine wave conversion part.

The second term of (11.10) suggests that with carrier suppression, the phase error is amplitude modulated at a frequency of $\theta_i/2^A$. If the phase error spectrum has a spurious component at a relative frequency F_i as shown in Figure 11.14(a), it is changed in the sine wave conversion part as indicated in Figure 11.14(b). The original component disappears and two new ones appear at relative frequencies $F_i - \theta_i/2^A$ and $F_i + \theta_i/2^A$. Figure 11.15 illustrates the case where the adder binary width is 8, the sine conversion

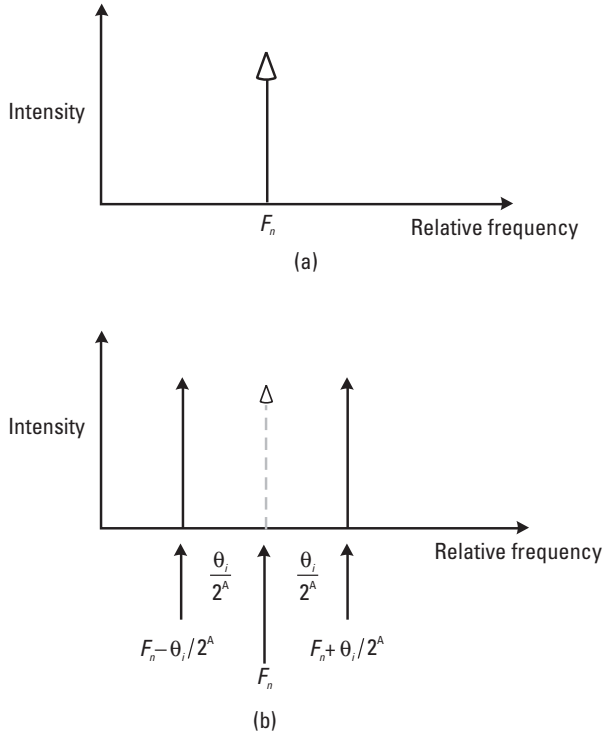


Figure 11.14 Transformation of spurious components due to phase error in the sine wave conversion part: (a) spurious signals included in the phase error frequency spectrum before sine wave conversion, and (b) spurious components included in the frequency spectrum of the sine wave output.

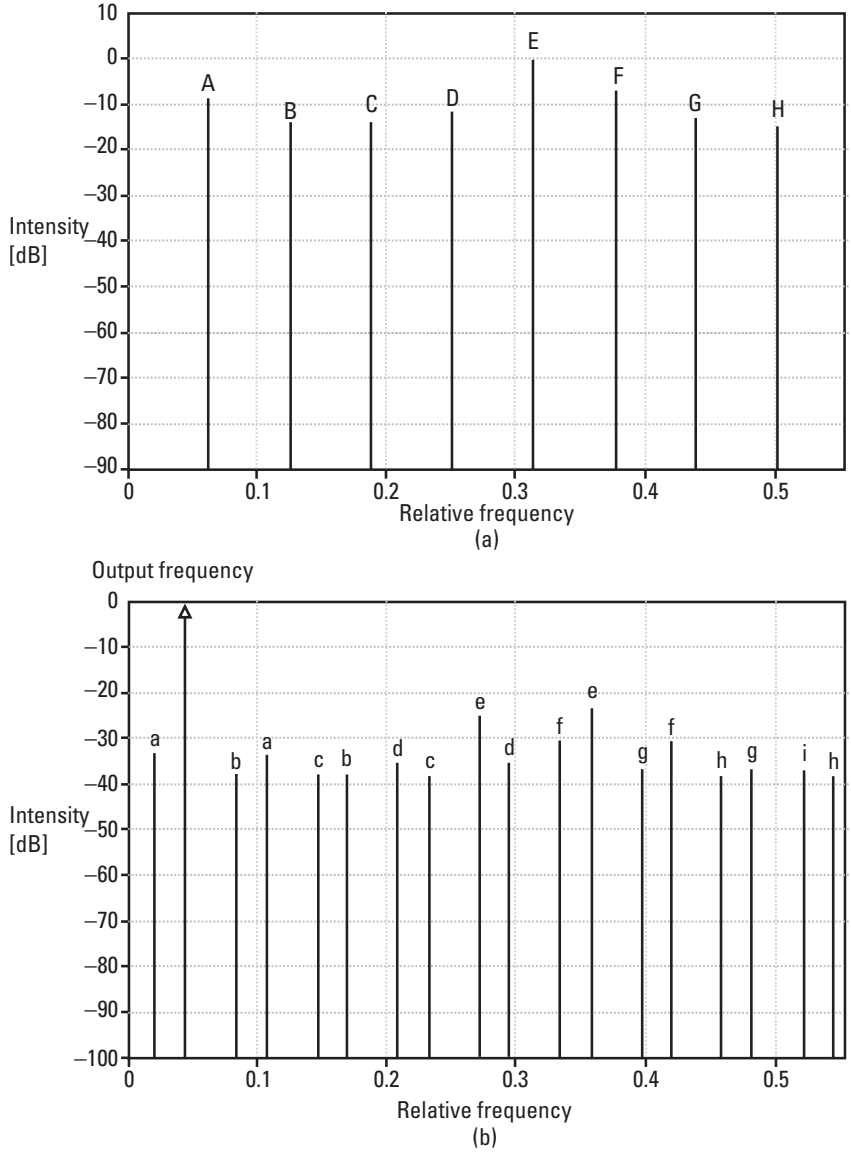


Figure 11.15 Frequency spectrum in the input and output of the sine wave conversion part: (a) frequency spectrum due to phase error in the input, and (b) frequency spectrum of sine wave in the output.

binary width is 4, and the phase increment value is 11. The 8 spurious components from A to H in relative frequency range of 0 to 0.5, change into 16

spurious components from a to h [Figure 11.15(b)]. Note that although the error due to the amplitude quantization binary width limitation is actually included in the output of the sine wave conversion part, it is not shown in Figure 11.15(b).

The signal purity is usually determined by maximum spurious amplitude and this is thus an important point in DDS design. The largest relative spurious amplitude N_{smax} is from [2] and is given by (11.12). The amplitude of the sine wave output is 1.

$$N_{smax} = 2^{-R} \frac{\frac{\pi}{P_r}}{\sin\left(\frac{\pi}{P_r}\right)} \tag{11.12}$$

If the adder binary width is 8, the sine conversion binary width is 4, and the phase increment value is 11, the spurious level will be –24 dB. This agrees well with the spurious amplitude in Figure 11.15(b). The maximum amplitude N_{smax} varies with the phase increment value θ_i and the binary widths A and R. The true maximum is reached when the coefficient P_r from (11.11) is 2. The worst level of the maximum spurious component is found to be

$$\begin{aligned} \max[N_{smax}] &= 2^{-R} \bullet \frac{\pi}{2} \\ &= -6.02R + 3.99[\text{dB}] \end{aligned} \tag{11.13}$$

and is determined by the sine wave conversion binary width R, as shown graphically by Figure 11.16. For instance, the sine wave conversion binary width must be 14 or more in order to suppress the spurious components below –80 dBc. Because the wanted sine amplitude is chosen to be one, $1/N_{smax}$ can also be thought to represent the relationship between the fundamental and unnecessary spurious components, the signal-to-noise ratio. The unit of spurious levels in (11.12) and (11.13) is dBc.

11.2.4 Spectrum in the DAC Output

The DAC at the output of the sine wave conversion part creates the final DDS output. Its binary width usually corresponds to that of the sine amplitude data in the conversion part. Figure 11.17 shows the relationship

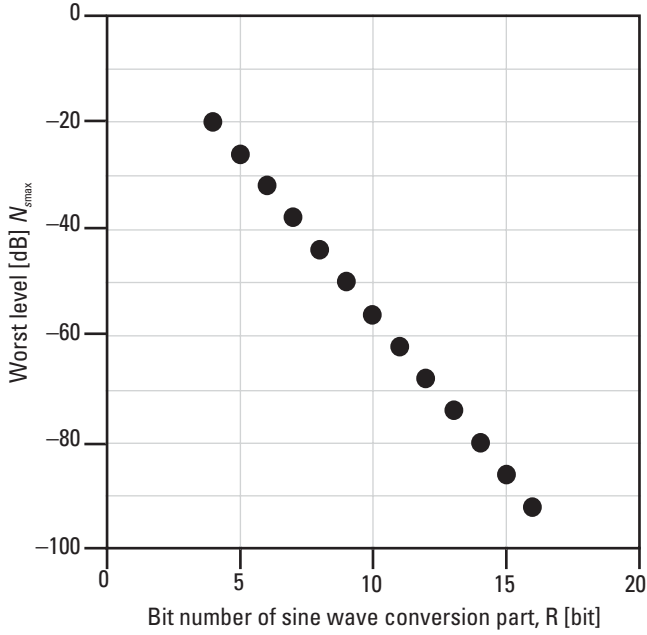


Figure 11.16 Worst change in the maximum spurious component level by the bit number of the sine wave conversion part.

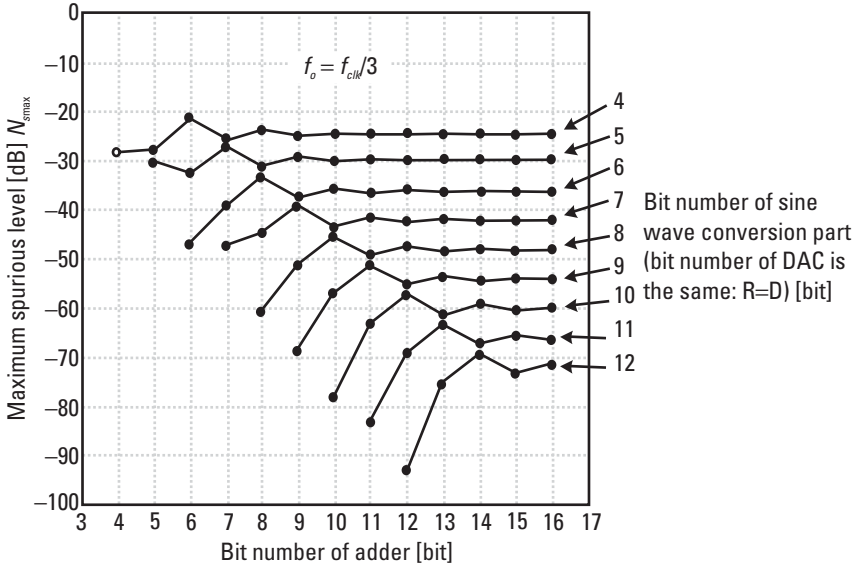


Figure 11.17 Maximum spurious component level change by bit numbers of the adder and the sine wave conversion part. (*After:* [5].)

between the adder binary width and the maximum spurious level in the case where the sine wave conversion binary width equals that of the DAC. This simulation was presented by Joseph F. Garvey and Daniel Babitch [5]. The spurious level is lowest when the binary widths of the adder, the sine conversion part, and the DAC are the same. The adder binary width is usually set larger than the others to obtain the highest resolution possible at the frequency setting. Under such conditions the spurious level is constant even if adder binary width is increased. This can be also understood from the other characteristic in which the worst value of N_{smax} is determined by the sine wave conversion binary width R.

Figure 11.18 shows the relationship between DAC binary width and the maximum spurious level with the sine wave conversion binary width as a parameter. The adder binary width is 18 [5]. If the binary widths of the sine wave conversion and DAC are equal, the line that connects the maximum spurious levels has a gradient of –6 dB/bit. This corresponds to the case where the coefficient R [see (11.13)] is 6.02 dB. A line with –8.5 dB/bit can also be observed. This indicates the limit imposed by the DAC binary width. This figure implies that the maximum spurious level is not degraded even if the DAC binary width is decreased up to the amount that corresponds to the

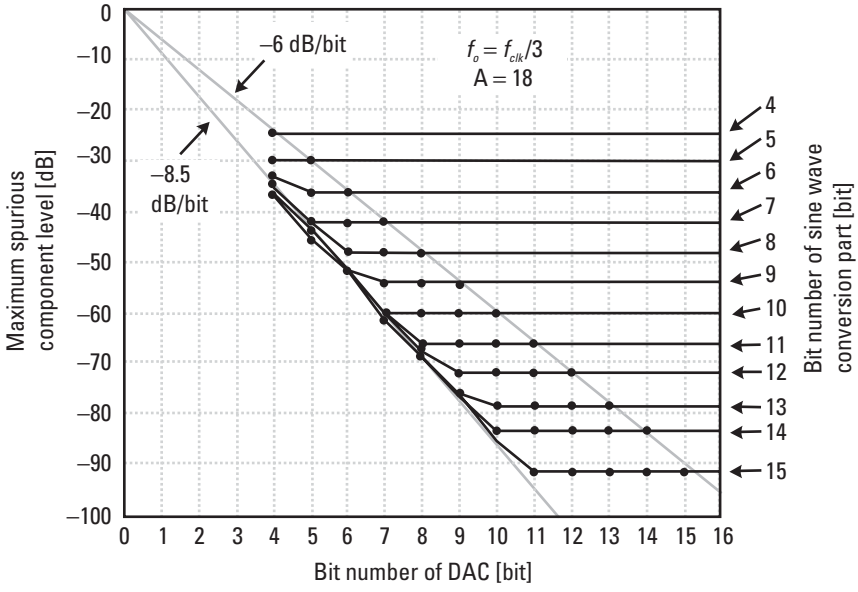


Figure 11.18 Maximum spurious component level change versus the DAC bit number. The bit number of sine wave conversion part is used as a parameter. (*After:* [5].)

area between the two lines. This is a very important point when designing a high-speed DDS. For instance, even if the DAC binary width is decreased to 9 bits, the spurious characteristics are maintained if the sine wave conversion binary width is 12.

The influence of the DAC binary width in Figures 11.17 and 11.18 is waveform distortion in the amplitude of the sine wave output. The distortion results from the quantization in the DAC even if the binary width expressing the sine wave data in the memory is large enough. When the output range of the DAC is equally divided by 2^{D} and a sine wave is generated, the actual noise power P_N which can be found by using the sine wave power P_S and the variance σ_q^2 of the quantization noise as

$$P_N = \frac{1}{SNR} = \frac{\sigma_q^2}{P_s} = \frac{1}{12 \times 2^{2D-3}} = -6.02D - 1.76 \text{ [dB]} \quad (11.14)$$

There are two noise sources related to DDS output. They are the spurious components determined by the sine wave conversion binary width, and the quantization noise defined by the the DAC binary width. A reasonable approach to achieve the best output performance is to harmonize these two noise levels. Equations (11.13) and (11.14) suggest that a good compromise is

$$R \geq D + 1 \quad (11.15)$$

This equation is consistent with the condition obtained from Figure 11.18 where, even if the DAC binary width decreases by several bits, the maximum spurious level does not grow. The process used to obtain (11.15) assumed that the DAC quantization noise [see (11.14)] defines the lowest noise level over the whole DDS output frequency band. Actually, since the quantization noise power is dispersed over all frequency bands, increasing the sine wave conversion binary width above the value obtained from (11.15) will improve the signal-to-noise ratio of DDS output.

11.2.5 Output Filter

The DAC output includes unnecessary high-frequency components generated by the sampling process. Their density depends on the sampling rate. Figure 11.19 shows waveforms and their spectrums in two cases where the clock frequency is close to and far from the output frequency. The clock

frequency is 20 Hz in Figure 11.19(a) and 50 Hz in Figure 11.19(b), and the output frequency is 2 Hz. After sampling we have unnecessary frequency components f_s, which are

$$f_s = n \bullet f_{clk} \pm f_o \quad n = 1, 2, \ldots \tag{11.16}$$

The higher the clock frequency is compared to the output frequency, the larger is the difference between the wanted and unnecessary frequency components. The condition of Figure 11.19(b) is more advantageous than that of Figure 11.19(a). Unnecessary frequency components can be suppressed by a filter whose passband follows the shaded area in Figure 11.19. The filter must have a fast response to able to follow signal changes. For sine wave output a filter having sufficient attenuation should be selected, even if its time domain response is not perfect. On the other hand, when aiming at arbitrary waveforms, the time domain response should be emphasized.

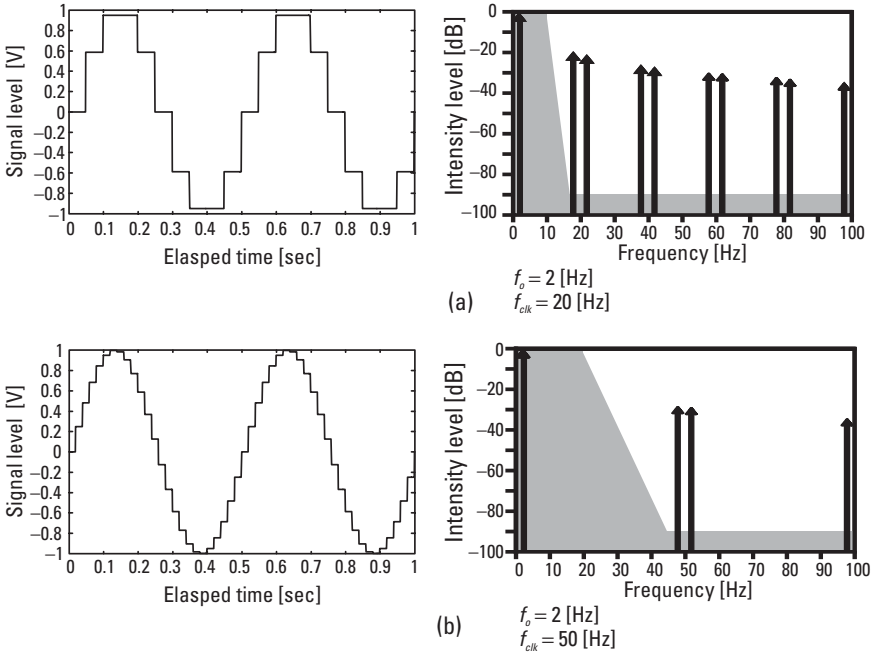


Figure 11.19 Waveforms of the DAC output signal and their frequency spectra: (a) case in which the clock frequency is close to the output frequency; (b) case in which the clock frequency is far from the output frequency.

11.3 New Design Methodology

11.3.1 Spurious Component Reduction

Although spurious components can basically be attenuated by optimizing the DDS parameters, it is necessary to change the whole configuration in order to rigorously suppress them [6]. We explain two typical methods in this paragraph.

11.3.1.1 Interpolation Method [6]

The internal digital data of the DDS and the analog output signal have another periodic component that generates spurious signals besides the main signal. If the spurious power is concentrated in a narrow frequency band, its level becomes high. It is thus necessary to disperse this power. An interpolation method can be used to randomly fluctuate the power of the spurious components and various randomization methods have been proposed:

1. Add a random signal to the phase increment value.
2. Add a random signal to the adder output.
3. Add a random signal to the sine wave conversion part output.
4. Fluctuate the clock frequency of the DAC.

These methods aim to disperse, not suppress, the power of the spurious components. Let us consider the approach of adding a random signal to the adder output as shown in Figure 11.20. The output spectra without and with the random signal are illustrated in Figure 11.21(a, b), respectively. Adding this random signal suppresses many detailed spectra. On the other hand, the lower limit of the noise (noise floor) rises to around –50 dBc. This method is ineffective against comparatively large spurious components such as A, B, and C in Figure 11.21 but works well with weak unwanted products around the carrier output. The two spurious components, D and E, are suppressed by 15 dB or more, which allows an output signal-to-noise ratio of 40 dB.

11.3.1.2 Newly Proposed Adder [4, 7]

DDS configurations that suppress the spurious signals by attenuating the error component have been proposed. They differ from the previous dispersing approach in any one particular frequency band. Figure 11.22 shows the configuration of the new DDS with two adders as was reported by Paul O'Leary and Franco Maloberti [4]. The first adder outputs the usual phase accumulation value. The second adder, shown as the gray area, adds the

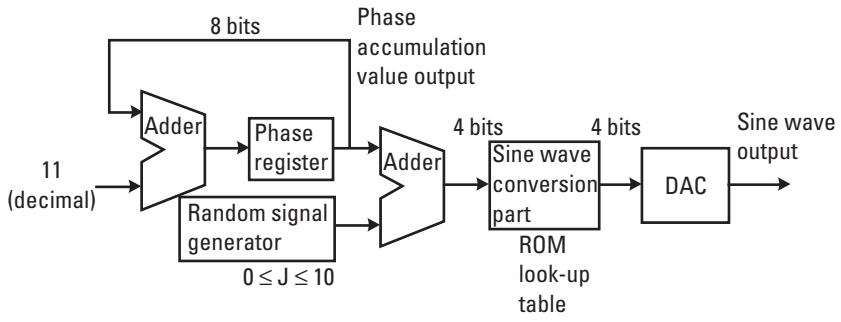


Figure 11.20 DDS configuration with a random signal generator in the phase accumulation value output.

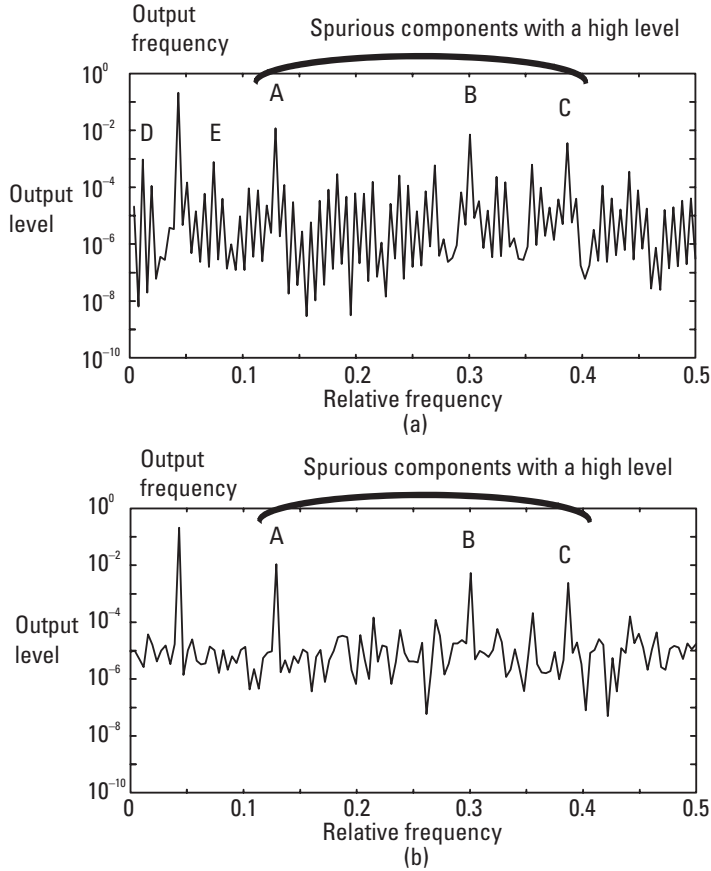


Figure 11.21 Output frequency spectrum with the random signal in the phase accumulation value output: (a) no random signal and (b) with random signal.

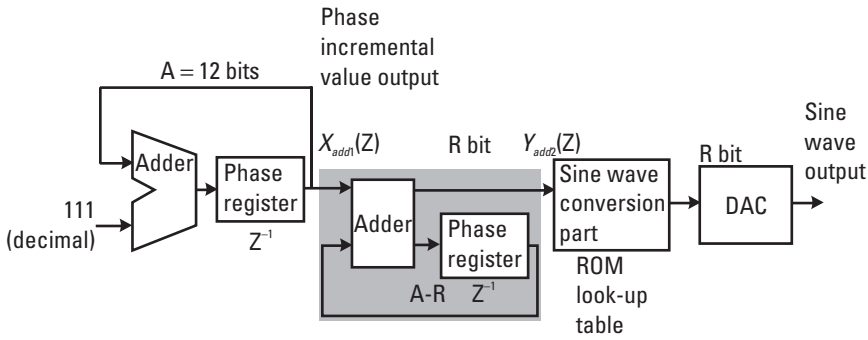


Figure 11.22 DDS configuration with the second adder. (*After:* [4].)

output of the first adder to the data that is truncated by the sine wave conversion part (A–R). The relationship between the input and output of the second adder is

$$Y_{add\,2}(z) = \left(1 - z^{-1}\right) \bullet X_{add\,1}(z) \tag{11.17}$$

The transfer function of the second adder shows a highpass characteristic as indicated in Figure 11.23. This characteristic decreases the level of the spurious component in the low frequency region. When the first adder binary width A is 12 and the address binary width R of the sine wave conversion part ranges from 3 to 6, the spectra become as illustrated in Figure 11.24. The configuration shown in Figure 11.22 suppresses spurious components by 10 to 20 dB over the comparatively low relative frequency region around the carrier of 0.027 (relative frequency). However, since the noise floor increases because of the second adder as shown in Figure 11.23, this configuration is not effective when the address binary width R of the sine wave conversion part is large.

11.3.2 Circuit Scale Reduction

Circuit scale is a key DDS design parameter. In particular, for a DDS with sine wave output, the circuit scale of the sine wave conversion part is very important. The sawtooth wave is converted into a sine wave under the control of the conversion table stored in the memory (ROM) of the sine wave conversion part. The scale of this conversion table is determined by the binary widths of the input phase accumulation, sine wave amplitude

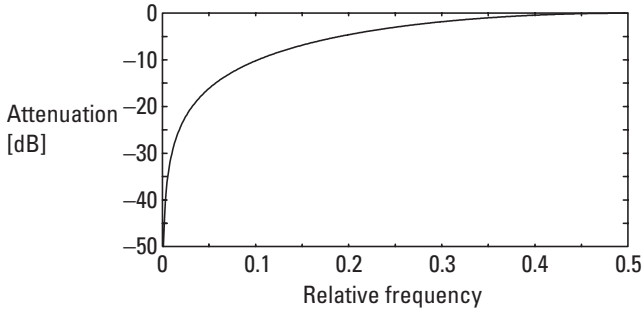


Figure 11.23 Transfer function of the second adder.

quantization, and DAC. The quantization step of the sine wave amplitude after conversion can be calculated from (11.18). When the phase accumulation binary width is 14, the required bit number is 2^{11} or more.

$$\frac{1}{\sin\left(\frac{2\pi}{2^{14}}\right)} = 2.6\times 10^{3} \approx 2^{11} \tag{11.18}$$

The phase quantization step must be the same as the input phase accumulation value. However, if the sine wave data from 0 to $\pi/2$ can be reused four times by changing its polarity and phase direction, the required quantization step can be decreased to one-fourth. If the phase accumulation binary width is 14, the modified step can be achieved by using 12 bits. Therefore, the required capacity in the conversion table is

$$2^{12}\times 11 = 45[\text{kbit}] \tag{11.19}$$

Another method that uses a different conversion technique has been proposed by Jouko Vankka [8]. Here, the conversion table is made by using the relation

$$\begin{aligned}&\sin(A+B+C) = \sin(A+B) + \cos(A)\cos(B)\cos(C)\\ &-\sin(A)\sin(B)\sin(C) \approx \sin(A+B) + \cos(A)\sin(C)\end{aligned} \tag{11.20}$$

This equation yields, as an approximate expression of the sine wave conversion, the digits of the phase accumulation value. Assuming that the

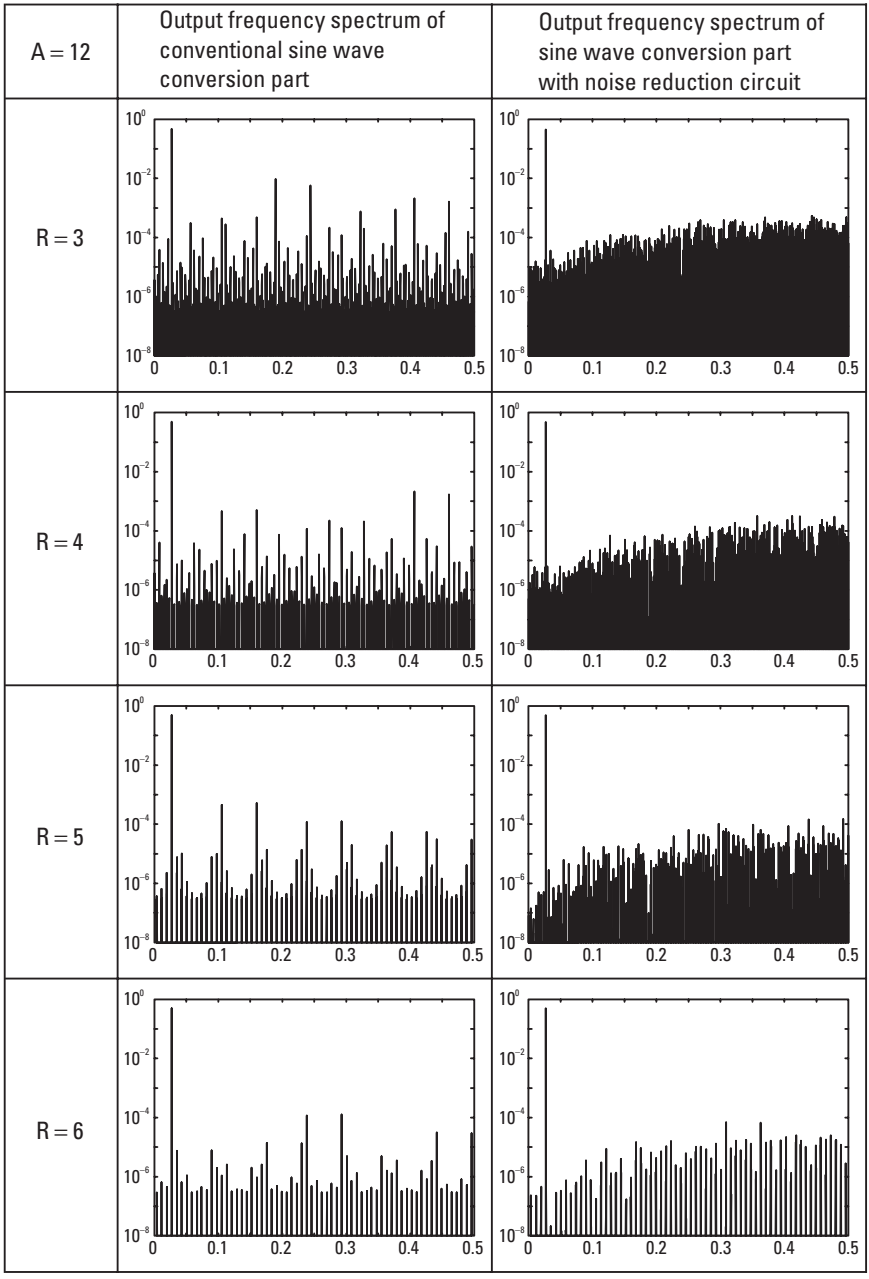


Figure 11.24 Frequency spectrum of the sine wave conversion part output with a noise reduction circuit.

phase quantization step is 12 bits and it is divided into 4 bits for each parameter, the respective digital value ranges and minimum phase resolutions are as listed in Table 11.2. In this case, it is necessary that the maximum value, which is the first term on the right-hand side of (11.20) be $\pi/2$, the phase step is 2^8, and the quantization step of the sine wave amplitude is 2^{11}. Since the maximum value, indicated by the second term, is approximately $\pi/2 \times 2^{-8}$ and the phase step is 2^8, the sine wave amplitude quantization step is

$$\frac{\sin\left(\frac{\pi}{2}\times 2^{-8}\right)}{2^{-11}} \approx 2^4 \tag{11.21}$$

Consequently, the sine conversion table here occupies just 3.8 kb, which is 1/12 times that of the similar table that expresses data from 0 to $\pi/2$. Since memory compression can influence the power consumption as well as the circuit scale, it is an important design item even in LSIs.

Table 11.2
Digital Value Range and Minimum Phase Resolution

Digital Value Range	Miniumum Phase Resolution
$A < \frac{\pi}{2}$	$\frac{\pi}{2}2^{-4}$
$B < \frac{\pi}{2}2^{-4}$	$\frac{\pi}{2}2^{-8}$
$C < \frac{\pi}{2}2^{-8}$	$\frac{\pi}{2}2^{-12}$

References

[1] Kroupa, V. F., "Spectral Properties of DDS Computer Simulations and Experimental Verifications," *IEEE International Frequency Control Symposium*, 1994, pp. 613–623.

[2] Nicholas, H., and H. Samueli, "An Analysis of the Output Spectrum of Direct Digital Frequency Synthesizers in the Presence of Phase-Accumulator Truncation," *41st Annual Frequency Control Symposium*, 1987, pp. 495–502.

[3] Troxell, W. C., "Performance Analysis of the Numerically Controlled Oscillator," *40th Annual Frequency Control Symposium*, 1986, pp. 373–378.

[4] O'Leary, P., and F. Maloberti, "A Direct-Digital Synthesize with Improved Spectral Performance," *IEEE Transactions on Communications*, Vol. 39, No. 7, 1991, pp. 1046–1048.

[5] Garvey, J. F., and D. Babitch, "An Exact Spectral Analysis of a Number Controlled Oscillator Based Synthesizer," *44th Annual Symposium on Frequency Control*, 1990, pp. 511–521.

[6] Reinhardt, V., et al., "A Short Survey of Frequency Synthesizer Techniques," *40th Annual Frequency Control Symposium*, 1986, pp. 355–365.

[7] Vankka, J., "Spur Reduction Techniques in Sine Output Direct Digital Synthesis," *IEEE International Frequency Control Smyposium*, 1996, pp. 951–959.

[8] Vankka, J., "Methods of Mapping from Phase to Sine Amplitude in Direct Digital Synthesis," *IEEE International Frequency Control Symposium*, 1996, pp. 942–950.

12

Some DDS Applications

12.1 DDS Trends

The DDS concept is difficult in designs that require high spectral purity because of the unacceptable level of spurious components. This results from the basic DDS configuration and from noise, which is generated, for example, by layout defects. These phenomena are of a fundamental nature and are thus hard to eliminate. For an adequate suppression of spurious components it is better to increase the internal DDS frequency rather than the processing width. An optimization of the relationship between the reference and the output frequency is inferior too. Recent progress in LSI technologies has, however, enabled fast DACs, which are used in high-frequency DDS circuits. They replace analog systems and conventional synthesizers and generate pure signals. Arbitrary waveform generators, which mostly output the signal according to programmed data, and frequency converters, which are required to generate complicated frequency ratios, have been realized from the latest DDS. These generators utilize the DDS feature that enables a free and precise control of the output frequency and phase.

12.2 Synthesizer

Figure 12.1 illustrates a new synthesizer that can generate a single frequency with various waveforms, arbitrary waveforms from internal memory, and that provides several modulations. This synthesizer can generate sine and

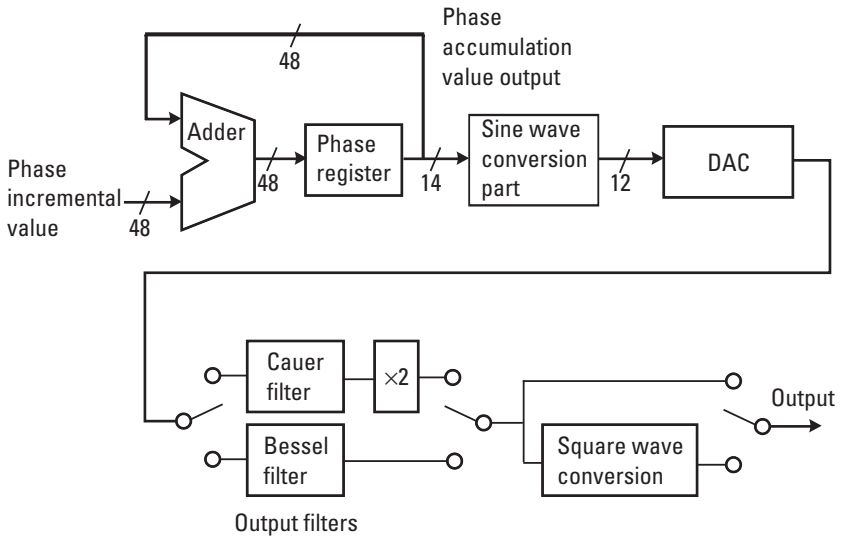


Figure 12.1 Example of a synthesizer configuration. (*Source:* [1]. Courtesy of Stanford Research Systems Co. and Seki Technotron Co.)

square waves up to 30.2 MHz with a frequency resolution of 1 μHz. The spurious level is below –50 dBc over this frequency range. The DDS part is composed of a 48-bit adder (implemented as an LSI), a wave conversion part, whose input is the 14-bit output of the LSI, a 12-bit DAC, the output part, which has two filters, and a square wave conversion part. A ninth-order Cauer (Elliptic) filter is used for sine waves, while a seventh-order Bessel filter is used for ramp, triangular, and arbitrary waveforms. The square wave generation circuit is different from the basic DDS. A general DDS can generate square waves but spurious (the level of which is determined by the internal bit combination) and jitter exist. To avoid these degradations, the square wave conversion part modifies the sine wave back to a square wave, which has the same characteristics (such as frequency resolution and noise) as the sine wave.

12.3 Combination of DDS and PLL

DDS output frequency has been rising because of the development of high-speed DACs and high-speed LSIs. Clock frequencies from 100 MHz [2] to around 1 GHz [3] have already been developed. However, existing DDS fail

in the gigahertz region. To break through this limitation and achieve high purity signal sources, a microwave synthesizer that consists of a combination of DDS and PLL has been proposed [4]. Figure 12.2 shows examples of

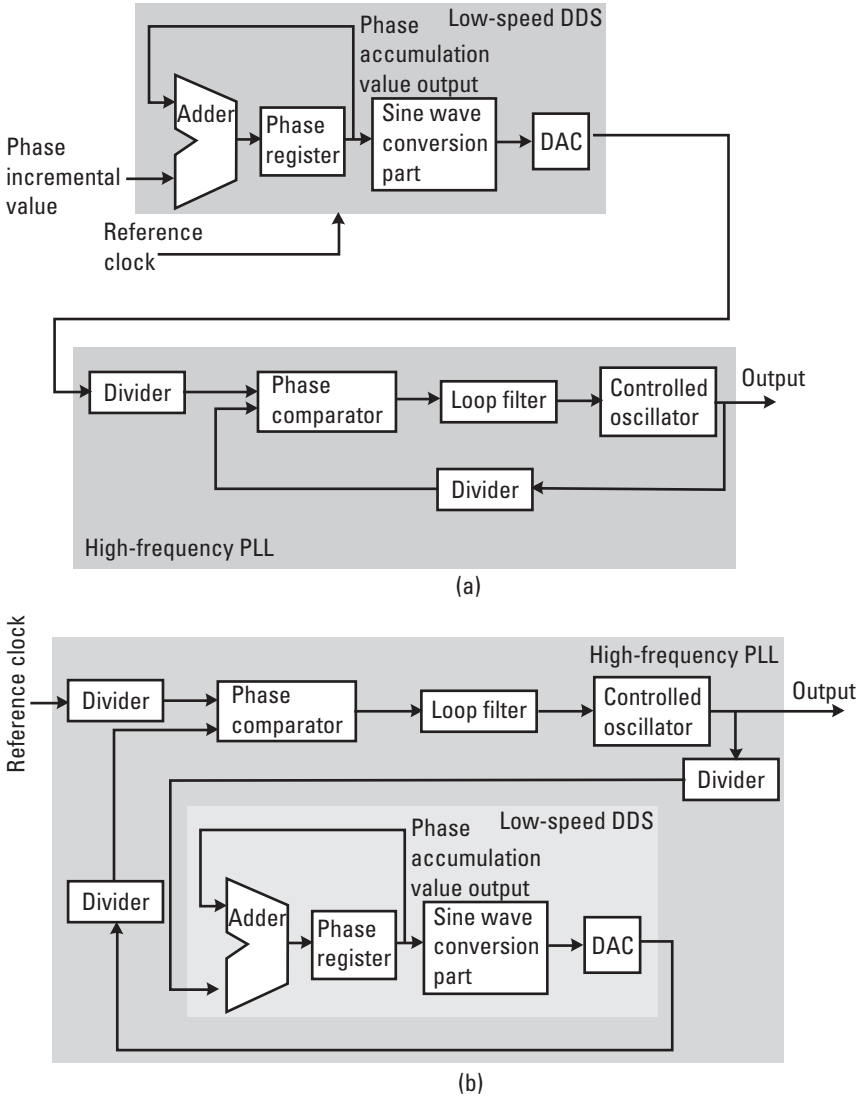


Figure 12.2 High frequency generation using a low-speed DDS applied for synthesizers: (a) synthesizer configuration with an external DDS, and (b) synthesizer configuration with an internal DDS.

synthesizer designs. There are two configurations. In Figure 12.2(a), the DDS is connected to the input of the PLL. In a general PLL synthesizer, the division ratio becomes large if frequency resolution is increased. This decreases the phase comparison frequency. Consequently, the PLL output phase stability is degraded. Because here the reference clock is converted once by the DDS, it is possible to make the PLL phase comparison frequency large enough. It is obvious that spurious components exist in the DDS output, but they can be suppressed by the PLL lowpass characteristic. In Figure 12.2(b) the DDS is installed in the PLL feedback circuit. In this case, too, the PLL phase comparison frequency can be raised by adjusting the DDS output frequency. This also improves the PLL phase stability.

12.4 Atomic Frequency Standards

An atomic frequency standard is an oscillator based on some atomic resonance. We can create a resonator in the microwave region by using the quantum mechanics phenomena of, for example, cesium, hydrogen, or rubidium atoms. Because such resonance frequencies are very stable, we can design highly accurate quartz oscillators by synchronizing to them. Figure 12.3 shows the configuration of a conventional atomic standard. The microwave resonator is fed by a combination of a synthesizer, a multiplier, and a mixer based on the output of the controlled 10-MHz quartz oscillator. The error voltage corresponding to the frequency difference between the inherent atomic resonance and the quartz oscillator output can be used to synchronize the controlled oscillator to the atomic frequency. The frequency accuracy,

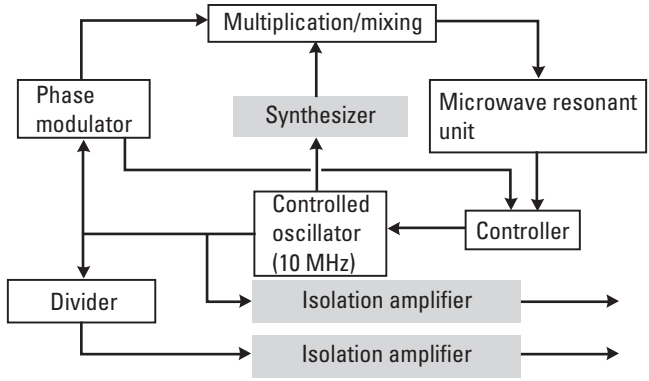


Figure 12.3 Conventional atomic standard configuration.

determined by the atomic resonance characteristic, can be obtained at the output of the oscillator.

The oscillator should work at 5 or 10 MHz, which are the most convenient frequencies for measurement systems and among the most stable quartz frequencies. The frequency of the controlled oscillator is irrelevant to the inherent resonance frequency of the standard. Therefore, the synthesizer must generate the specific microwave frequency. A configuration that uses two DDSs to achieve this is shown in Figure 12.4. They allow us to select a frequency that, given the atomic frequency, optimizes the overall performance. This configuration can improve frequency stability and reduce unit size. The DDS can also generate the 5-MHz and 10-MHz outputs [5].

12.5 DDS Circuits in Wireless Systems

As already mentioned, two inherent challenges exist regarding the use of DDS circuits at the air interface circuits of wireless systems. The method itself is known for its poor spectral purity, and commercial DDS chips cannot yet enter the microwave bands economically. Attempts to improve the spectrum through a spreading of the spurious power across wider bandwidths may not be suitable unless the so-created noise-like power falls much below the natural, thermal noise. Note that the noise limit must be set at the level prevailing in the input of an ideal, noiseless receiver.

Three interesting features are available. The DDS signal can be modulated very precisely without any analog intervention. This enables a straightforward method for creating complicated and combined modulations

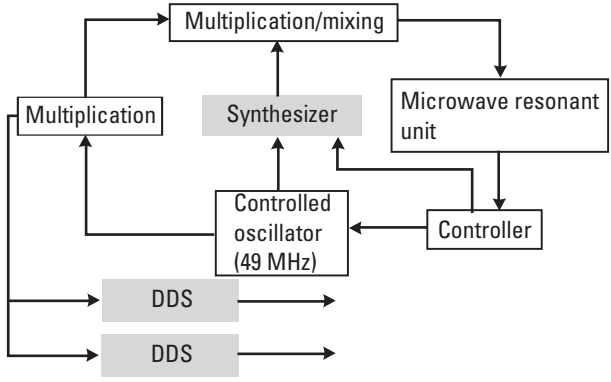


Figure 12.4 Atomic standard configuration using DDS. (*After:* [5].)

schemes with well-defined characteristics. Additionally, the frequency switching speed is theoretically superior to most other oscillators. Frequency hopping and spread spectrum radios can efficiently make use of this characteristic provided that the required final output frequency is well within the limits of the DDS chip. Another limitation or trade-off is set by the total hopping width because an efficient filtering of the DDS spurious components may turn impossible. Measuring equipment and radar systems, in particular, benefit from the improved frequency resolution, which is possible by using a DDS unit as an HF fundamental oscillator. After it, the required microwave signals can be generated either by PLL circuits or through mixing and analog multiplying. Unfortunately, the final spectrum is often far from acceptable and heavy filtering is necessary. This may not be simple in the upper microwave or millimeter wave bands.

References

[1] Stanford Research Systems, Model DS345 Synthesized Function Generator.

[2] Kent, G. W., and Sheng, N.-H., "A High Purity, High Speed Direct Digital Synthesizer," *IEEE International Frequency Control Symposium*, 1995, pp. 207–216.

[3] Kushner, L. J., "The Composite DDS—A New Direct Digital Synthesizer Architecture," *IEEE International Frequency Control Symposium*, 1993, pp. 255–260.

[4] United States Patent, "Direct Digital Synthesizer Driven Phase Lock Loop Frequency Synthesizer," 4965533.

[5] Koyama, Y., et al., "An Ultra-Miniature Rubidium Frequency Standard," *IEEE Frequency Control Symposium*, 2000.

13

Noise in Clocks

13.1 Frequency Stability and Noise

When frequency stability is used to specify clock performance, it is generally understood as referring to frequency change. This corresponds to drift or aging over time in oscillators. Frequency stability also includes the concept of frequency variation. For telecommunications systems, frequency and phase stability are specified using statistical definitions.

Noise must be analyzed if we are to assess the statistical clock performance. More detailed specifications of frequency stability are needed for this goal. The basic PLL analysis described in Chapter 5 did not consider the noise of signals and circuits. Spurious characteristics were introduced in the DDS analysis in Chapter 11. Only digital signal analysis was used, however, and the concept of noise was not considered. This chapter describes the influence of random electrical noise as well as measures to evaluate and specify noise in clocks.

13.2 Relationship Between Sine Wave and Noise

13.2.1 Noise in the Frequency Domain

A common ideal sine wave generated by a signal source has a single frequency component. We must accept, however, that the amplitude, frequency, and

phase of a practical sine will include noise. The power of a sine wave is concentrated on just its fundamental frequency f_o if the wave can be expressed as

$$v(t) = v_o \sin(2\pi f_o t) \tag{13.1}$$

If this signal is analyzed with a spectrum analyzer, the measured frequency domain spectrum is as shown in Figure 13.1. A single peak exists at f_o and its magnitude is related to the effective power. The practical spectrum looks more like that in Figure 13.2, where unwanted frequency components lie around the center frequency. These components around frequency f_o indicate the existence of noise, which is one of the main factors degrading frequency stability.

Peaks at frequencies that are harmonic multiples of the fundamental wave can be observed as shown in Figure 13.3. Such signals are called harmonic spurious components. Although the suppression of spurious components is an important point in DDS design, it is not discussed here. This chapter focuses instead on unwanted random peaks around the center frequency.

13.2.2 Noise in Time Domain

One approach to noise analysis is to observe signal variations in the time domain. This approach is akin to an analysis in the frequency domain and corresponds to the observation of the wave shape with an oscilloscope. An ideal noiseless sine wave is shown in Figure 13.4. One cycle T_o of the signal is the time interval that accurately corresponds to $1/f_o$.

If a sine wave contains phase noise, its timing varies as shown in Figure 13.5. The variation depends on the characteristics of the noise and is generally random. Square waves exhibit the same characteristic, but here the

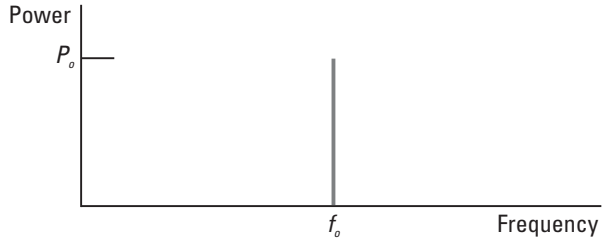


Figure 13.1 The single frequency component of an ideal sine wave.

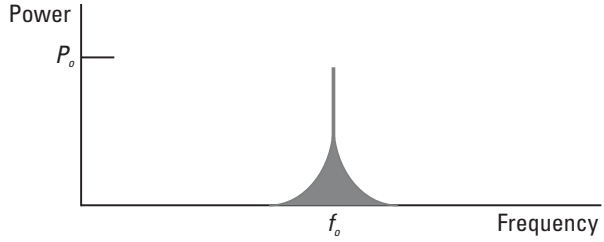


Figure 13.2 Actual frequency components of a realistic sine wave.

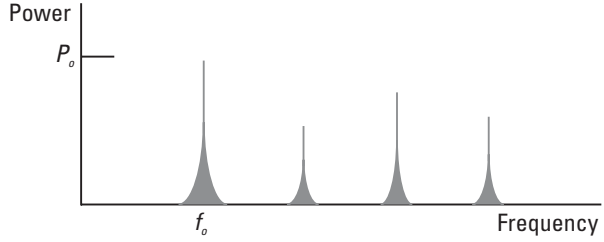


Figure 13.3 Sine wave with many harmonic components (distortion).

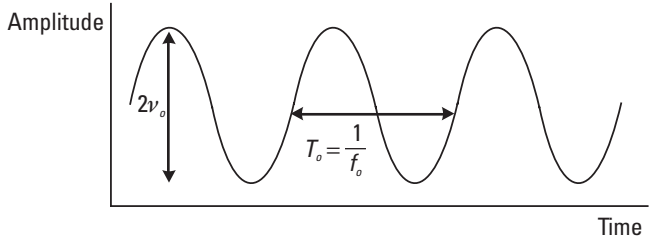


Figure 13.4 Sine wave in the time domain showing the momentary amplitude over time.

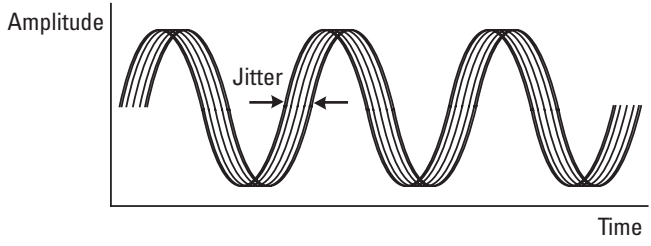


Figure 13.5 Sine wave including noise, presented over time.

position of the rising and falling edges varies as illustrated in Figure 13.6. This variation is called jitter, and it is conveniently used to specify digital network characteristics. Jitter is a time-based measure. It has units of [sec] in phase-time and [rad] in phase. In addition, unit interval (UI) is frequently used in recent network analyses. It uses one clock cycle as one UI, which normalizes values to the clock cycle.

13.3 Noise Processing

Noise is present in the amplitude, frequency, and phase of clock signals. The presence of noise in a sine signal can be expressed as

$$v(t) = \left[v_o + v_n(t)\right]\sin\left[2\pi f_o t + \theta_n(t)\right] \tag{13.2}$$

The noise components related to frequency and phase are generally much above amplitude noise. Conventional methods developed for general signals can be used to analyze the noise of PLLs and DDSs.

13.3.1 Noise Characteristics in PLL

The frequency response of the PLL transfer function defines how phase noise in the loop input propagates to the output. Thus, the functions of Chapter 5 can also be used for phase noise analysis. Figure 13.7 shows the relationship between input and output noise of a PLL. Figure 13.7(a) shows the spectrum of the input signal, which includes the carrier and noise sidebands. Let us assume that the PLL has the transfer function shown in Figure 13.7(b), and that the passband of the transfer function is narrower than the frequency band of the input signal noise. Under these conditions, the output spectrum will be like that in Figure 13.7(c). Some of the input noise components are

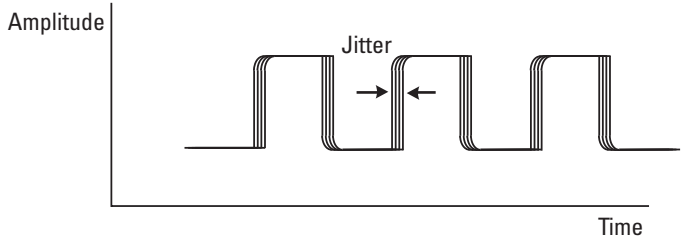


Figure 13.6 Square wave with noise, presented as a function of time.

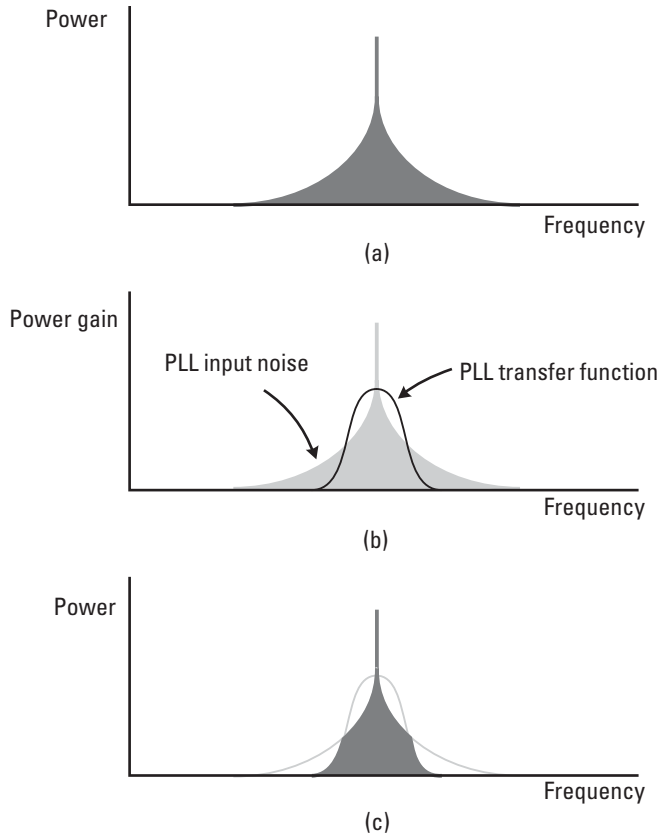


Figure 13.7 Input and output noise of a PLL: (a) PLL input noise, (b) input noise and transfer function of a PLL, and (c) PLL output noise.

suppressed by the lowpass characteristic of the transfer function, which is actually a bandpass characteristic as shown in Figure 13.7(b).

13.3.2 Noise in Frequency Conversion

Input noise of frequency conversion systems, such as a DDS, is converted together with the wanted frequency component, and appears in the output. For example, the noise in the system clock [see Figure 13.8(a)] influences the DDS output as indicated in Figure 13.8(b). The frequency domain operation is halving (i.e., one of two pulses from the system clock is dropped). In this case, the phase variation of the rising and falling edges mostly equals the original noise in the system clock. With frequency multiplication, the system

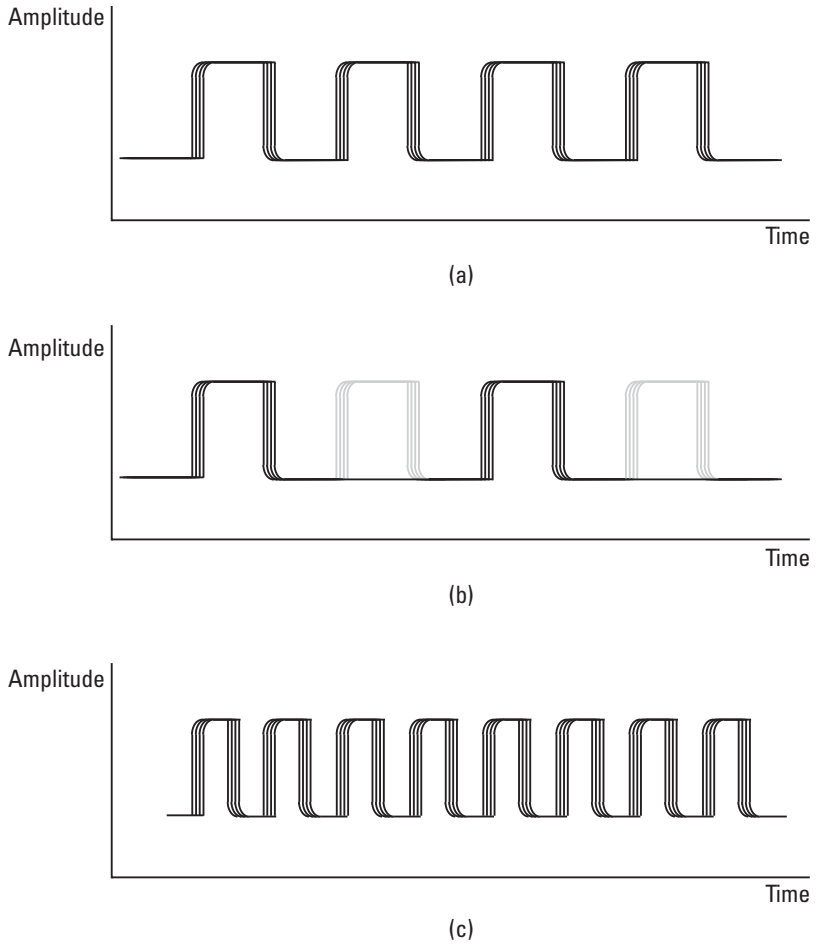


Figure 13.8 Influence of noise in frequency conversion: (a) noise present in the system clock, (b) DDS output noise (output frequency is 1/2 of the system clock), and (c) influence of noise with frequency multiplication.

clock is converted as illustrated in Figure 13.8(c). The timing variation does not change although the number of pulses has increased. The noise, which is included in the synthesized output signal, is processed in roughly the same way as the system clock. If the influence of circuit noise can be ignored, the synthesized signal noise is equal to the original noise characteristic of the system clock.

Note that while the amount of noise does not change if it is evaluated by using units of phase-time [sec], the synthesized signal noise is converted

by the frequency synthesis ratio if the units are phase [rad]. Since the fundamental frequency of the signal decreases with frequency division, the noise level decreases. When the frequency is increased, the noise will appear to have increased.

13.3.3 Noise Power

The power of a sine signal is concentrated on the fundamental frequency. It can be defined regardless of the width of the observation bandwidth. On the other hand, noise power changes according to the frequency band in which it is measured, since noise usually consists of many components spread over a wide frequency range. Measurement instruments such as spectrum analyzers can measure signal power over different frequency regions by changing the frequency passband of the instruments' filter. The measurement result is the same as the noise power when the filter passband is wider than the noise bandwidth as shown in Figure 13.9(a). However, if the filter is narrower, the result is below the true noise power. This is because the filter passes only a part of the noise. Note that the measuring bandwidth must thus be specified.

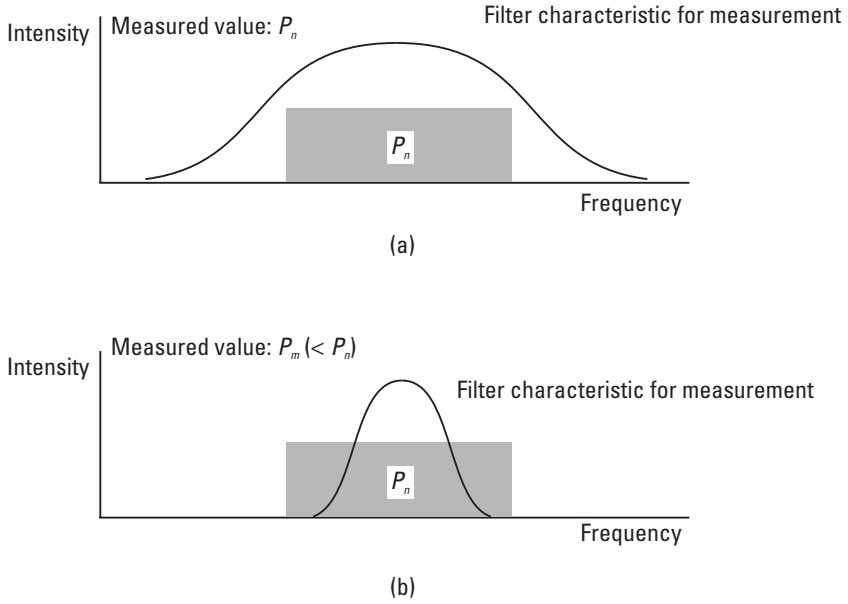


Figure 13.9 Noise measurement: (a) noise measurement using a wider frequency band than that of noise, and (b) noise measurement using a narrower frequency band than that of noise.

A general way of expressing the true noise power while eliminating the influence of measurement conditions is to use normalized noise power where the normalization involves a unit frequency band. In Figure 13.10, the noise, whose power is P_n, is distributed from frequency f_l to f_h over its distributed frequency region of Δf, and the normalized power is $P_n/\Delta f$. This measure enables an easy comparison of noise powers that are measured by using different filters.

13.4 Phase Noise and Frequency Noise [1]

Noise has been observed to modulate the frequency and/or phase of the clock signal, even if it is simply added to the observed signal in a circuit. To clarify the noise characteristic generated from this mechanism, let us consider the case where the observed signal is modulated by a sine signal

$$v_m(t) = m_0 \cos(2\pi f_m t) \tag{13.3}$$

When the sine wave, whose frequency is f_0, is modulated in phase and frequency by the sine signal of (13.3), the result will be

$$v(t) = v_o \sin\left[2\pi f_o t + m_0 \cos(2\pi f_m t)\right] \tag{13.4}$$

$$v(t) = v_o \sin\left[2\pi f_o t + \frac{m_0}{f_m}\sin(2\pi f_m t)\right] \tag{13.5}$$

Modulated signal spectra are seen if (13.4) and (13.5) are separated into terms that involve the fundamental and the modulation frequency.

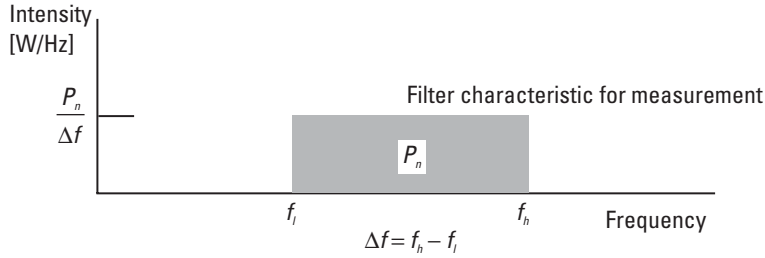


Figure 13.10 Noise power expression in the frequency domain.

Modulation components are on both sides of the fundamental component. Modulation appears at frequencies $f_o + f_m$ and $f_o - f_m$ as shown in Figure 13.11 when the magnitude of the modulation signal is small. The magnitudes of the components are proportional to the magnitude of the modulation signal. These features of phase and frequency modulation are inherited by noise, which contains numerous frequency components.

In the case where noise causes phase modulation, the noise characteristic appears as it is in signal spectrum. Therefore, the noise frequency component $N_\varphi(f)$ is independent of Fourier frequency, and can be found as

$$N_\varphi(f) \propto |f|^0 \tag{13.6}$$

In frequency modulation, the influence of noise can be described as the inverse of a Fourier frequency. Its frequency component is

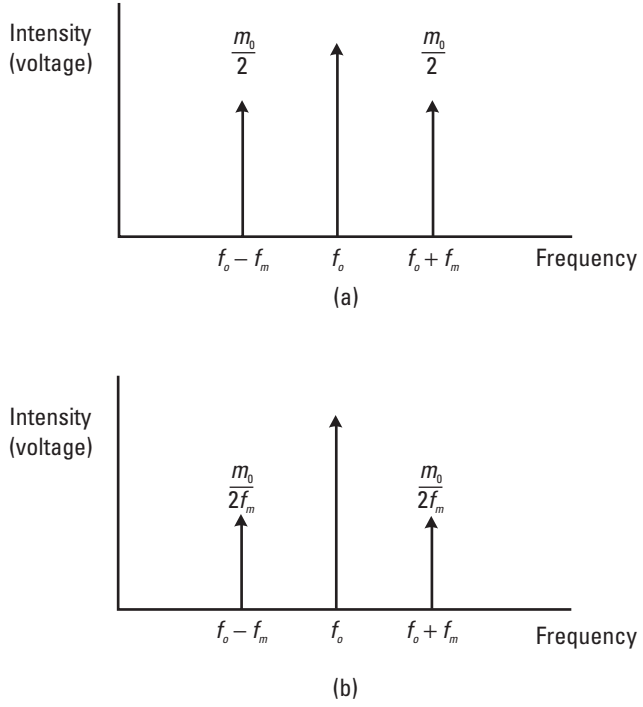


Figure 13.11 Modulation component amplitudes: (a) modulation component amplitude in phase modulation, and (b) modulation component amplitude in frequency modulation.

$$N_{\varphi}(f) \propto |f|^{-1} \tag{13.7}$$

The frequency f used in (13.6) and (13.7) lies in the range $0 \leq f \leq \infty$.

Let us use power instead of voltage to represent the magnitude of a frequency component. This helps subsequent noise analyses in which probabilistic quantities are handled. If the frequency component of a signal $n_{\varphi}(t)$ at time $0 \leq t \leq T$ is $N_{\varphi}(f)$, then $|N_{\varphi}(f)|^2$ represents the energy magnitude of the original signal as a mean square value. In addition, if $|N_{\varphi}(f)|^2$ is a finite function, the average energy per unit time is given by

$$S_n(f) = \lim_{T \to \infty} \frac{1}{T} |N_{\varphi}(f)|^2 \tag{13.8}$$

Since (13.8) shows the average energy per unit time (unit frequency), it can be considered as yielding power spectral density. Equations (13.6) and (13.7) can be converted into the form of power spectral density by using (13.8), whereby we get

$$\text{Phase modulations: } S_p(f) = h_p \bullet |f|^0 \tag{13.9}$$

$$\text{Frequency modulations: } S_f(f) = h_f \bullet |f|^{-2} \tag{13.10}$$

where h_p and h_f are the power spectral density magnitudes.

13.5 Frequency Components of Noise [1]

It is well known that noise exhibits an interesting characteristic in terms of its frequency components. The observed trend is that the amplitude of frequency components increases with falling frequency. Noise can be classified based on this characteristic into white, flicker, random walk, flicker walk, and random run noises [1]. These noise spectra are compared in Figure 13.12. White noise, the most well-known type, has constant amplitude regardless of the frequency (Fourier frequency), f. The other types show an increasing emphasis on low frequency weighting in the order flicker noise, random walk noise, flicker walk noise, and random run noise. The most commonly seen noise types are white, flicker, and random walk noises.

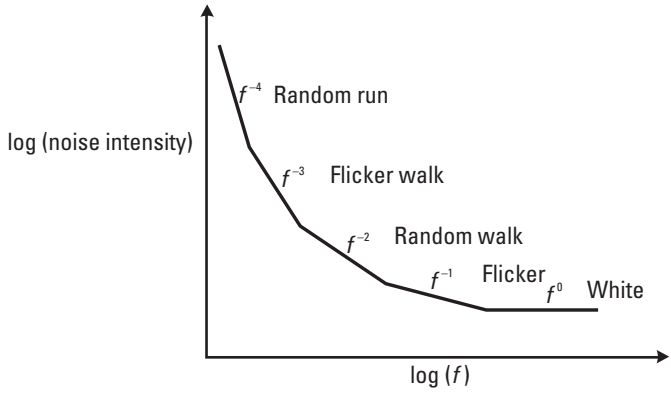


Figure 13.12 Frequency spectra of noise.

Let us consider the cases in which these three types are present in a clock that undergoes phase and frequency modulation. Six noise types are created by the combination of the three original noise processes and two modulations. Their spectra are drawn in Figure 13.13. The influence on the power spectrum with phase and frequency modulation can be determined from (13.9) and (13.10) by using

$$S_m(f) \propto |f|^m \tag{13.11}$$

where $m = 0$ for phase modulation and $m = -2$ for frequency modulation. The tendency of the noise can be expressed also as

$$S_n(f) \propto |f|^n \tag{13.12}$$

where $n = 0$ for white noise, $n = -1$ for flicker noise, and $n = -2$ for random walk noise. The total power spectrum density is determined by the combined influence of the modulation and frequency characteristics of the noise. It is given by

$$S_\varphi(f) \propto |f|^m \bullet |f|^n = |f|^\alpha \tag{13.13}$$

where, $\alpha = 0, -1, \ldots, -4$.

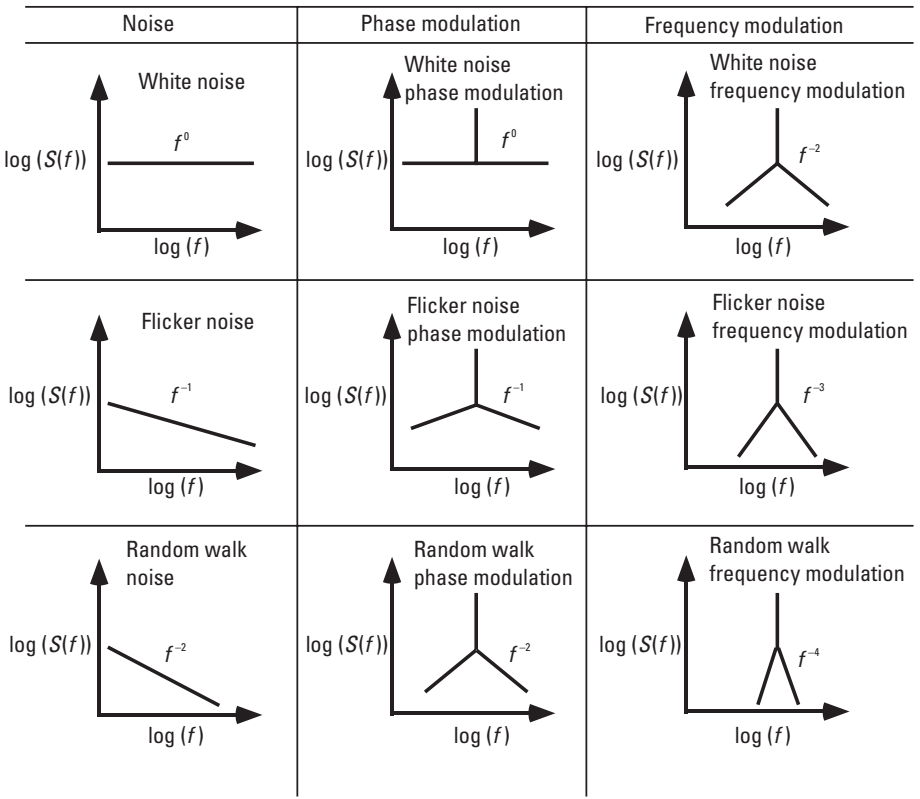


Figure 13.13 Frequency spectra in a combination of noise and modulation.

While the six responses can be calculated as shown in Figure 13.13, the following five responses actually exist:

1. White noise phase modulation;
2. Flicker noise phase modulation;
3. White noise frequency modulation;
4. Flicker noise frequency modulation;
5. Random walk frequency modulation.

13.6 Other Measures for Spectrum Analysis [1–3]

Clock noise can be expressed by measures other than those involving the phase modulation domain. Power spectral density terms can be created for frequency variation, normalized frequency deviation, and for the phase-time. These measures are related to phase as follows.

Frequency Variation Power Spectral Density

From the relationship between phase and frequency we obtain

$$n_{\Delta f}(t) = \frac{1}{2\pi}\frac{dn_{\varphi}(t)}{dt}[\text{Hz}] \tag{13.14}$$

Based on (13.14) we have the frequency variation power spectral density as

$$S_{\Delta f}(f) = f^2 \bullet S_{\varphi}(f)[\text{Hz}^2/\text{Hz}] \tag{13.15}$$

Frequency Deviation Power Spectral Density

The relation between the normalized frequency deviation, normalized by the nominal frequency, and the phase of the signal is

$$n_y(t) = \frac{1}{2\pi f_o}\frac{dn_{\varphi}(t)}{dt}[1/\text{Hz}] \tag{13.16}$$

From (13.16) we get the normalized frequency deviation power spectral density as

$$S_y(f) = \frac{f^2}{f_o^2} \bullet S_\varphi(f) \quad [1/\text{Hz}] \tag{13.17}$$

Phase-Time Power Spectral Density

The relationship between time and phase can be determined from the signal frequency, which is the target of noise analysis, as

$$n_x(t) = \frac{n_\varphi(t)}{2\pi f_o}[s] \tag{13.18}$$

Similarly to previous discussion, the phase-time power spectral density will be

$$S_x(f) = \frac{1}{(2\pi f_o)^2} \bullet S_\varphi(f)[s^2/Hz] \tag{13.19}$$

Since time refers to elapsed time and phase is assumed to have units of radians, phase-time is used in this book to indicate phase with units of seconds.

Normalized Power Spectral Density

Ratio $\mathcal{L}(f)$ of the noise power to the total signal power is used to show the relative magnitude of the noise power spectral density against carrier power. This normalized ratio is defined by

$$\mathcal{L}(f) = \frac{\text{Phase power spectral density}}{\text{Total signal power}} \tag{13.20}$$

When the noise power satisfies the condition of (13.21), $\mathcal{L}(f)$ can be related to phase power spectral density as shown in (13.22).

$$\text{Noise power} << 1[rad] \tag{13.21}$$

$$\mathcal{L}(f) = \frac{1}{2} S_\varphi(f) \tag{13.22}$$

13.7 Variance and Power Spectral Density [1–3]

13.7.1 Sample Variance

As explained in Section 13.2.2, the fluctuation of noise over time can be indicated by jitter. Jitter is a very useful measure for white phase noise since it makes dispersion amplitude easier to understand. If the main component is white phase noise, jitter expresses the magnitude of the dispersion; note that the magnitude of the other noise types vary with the measurement parameters. This is because jitter expresses only the total phase dispersion. This restriction can be partly overcome by using sample variance. It well expresses the dispersion characteristics as its magnitude varies with the observation time.

The N sample variance can be calculated from the mean frequency deviation over observation time τ by using the time interval T, as in Figure 13.14. The frequency deviation is given by

$$\bar{y}_i = \frac{1}{\tau}\int_{t_i}^{t_i+\tau} y(t)dt = \frac{1}{\tau}\left[x(t_i+\tau) - x(t_i)\right] \tag{13.23}$$

The N sample variance is defined by using (13.23) as

$$\sigma_y^2(N,T,\tau,f_h) = \frac{1}{N-1}\sum_{n=1}^{N}\left(\bar{y}_n - \frac{1}{N}\sum_{i=1}^{N}\bar{y}_i\right)^2 \tag{13.24}$$

This equation is almost equal to the well-known expression of variance. The upper frequency limit of the measurement arrangement is f_h. The measurement system cannot measure over an infinite frequency range, which influences the observed noise magnitude. The definition given by (13.24)

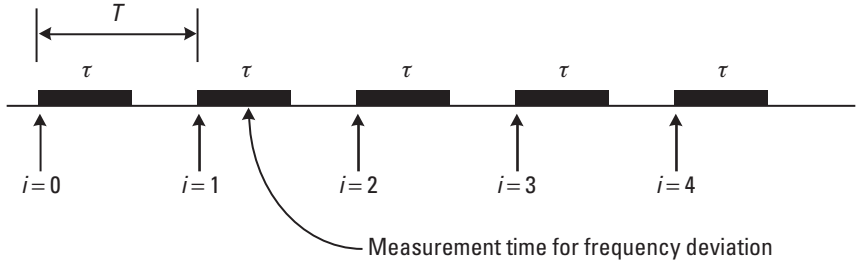


Figure 13.14 Measurement method for sample variance.

has a defect in that the variance cannot be calculated if the mean value of frequency deviation $\bar{y}_i$ is not fixed. If the frequency drifts over time, N sample variance diverges as the number of samples grows. To overcome this defect of the general variance, two-sample variance, also called the Allan variance, was proposed. Two-sample variance is given as

$$\sigma_y^2(\tau) = \frac{\left\langle \left(\bar{y}_{i+1} - \bar{y}_i\right)^2 \right\rangle}{2} \tag{13.25}$$

where $\langle\ \rangle$ shows an infinite time average.

One of the features of two-sample variance is that when the normalized frequency deviation $\bar{y}_i$ constantly changes over time, it is canceled out. This is a very effective measure for frequency offset signals like a clock. Equation (13.25) can be related to N sample variance through (13.26) and (13.27). Two-sample variance varies with the dead time, $T - \tau$, which must be defined in practical measurements.

$$\sigma_y^2(\tau) = \sigma_y^2(2, T, \tau, f_h) \tag{13.26}$$

$$\sigma_y^2(\tau) = \sigma_y^2(2, \tau, \tau, f_h) \tag{13.27}$$

13.7.2 Sample Variance and Power Spectral Density

Measures for noise analysis, power spectral density (frequency domain), and sample variance (time domain) were introduced in this chapter. These two physical measures well complement each other as measuring devices can handle noise in different areas. Spectrum analyzers are good for high frequencies. The practical lower limit is approximately 0.01 Hz, even for the digital-processing based spectrum analyzer with best resolution. On the other hand, systems based on a counter for variance measurement (time domain) offer continuous recordings over long time periods such as a week or a month. Here the limits imposed by the instrument speed and tolerable input clock frequency prevent the measurements over short time periods.

It is useful to combine these measures in frequency and time domains as described above and also to confirm measured data. Time series data measured in the time domain can be converted into power spectral density via the autocorrelation function. Equation (13.28), also called the theorem of Wiener-Khintchine, is an important relational expression. It shows that the

autocorrelation function and the power spectral density are related through the Fourier transform (refer to Chapter 3) [4].

$$R_y(\tau) = \int_0^\infty S_y(f)\cos(2\pi f\tau)df$$
$$S_y(f) = 4\int_0^\infty R_y(\tau)\cos(2\pi f\tau)d\tau \tag{13.28}$$

The N sample variance defined by (13.24) can be converted into an expression involving the power spectral density $S_y(f)$ by using (13.28), and we get

$$\sigma_y^2(N,T,\tau,f_h) = \frac{N}{N-1}\int_0^\infty S_y(f)\left|H(f)\right|^2 df \tag{13.29}$$

where, $\left|H(f)\right|^2 = \dfrac{\sin^2(\pi f\tau)}{(\pi f\tau)^2}\left[1-\dfrac{\sin^2(\pi f\tau rN)}{N^2\sin^2(\pi f\tau r)}\right]$, and $r = \dfrac{T}{\tau}$.

Equation (13.29) indicates that the sample variance corresponds to the measurements obtained when power spectral density $S_y(f)$ is observed via window function $H(f)$. The N sample variance can be calculated from power spectral density. In general, the opposite process is difficult except for a special case. When the tendency described in Figure 13.15 is observed by using the two-sample variance [(13.25)], the original power spectral density can be

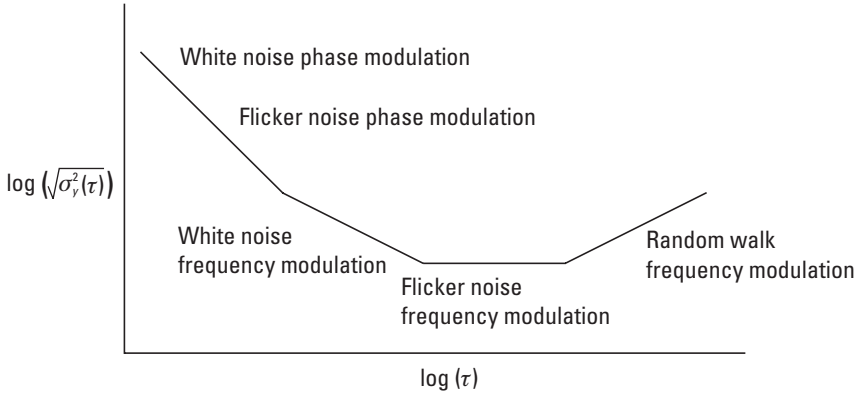


Figure 13.15 Two-sample variance characteristics depending on noise.

estimated. The relationship between power spectral density $S_y(f)$ and $S_\varphi(f)$ and the two-sample variance is listed in Table 13.1. The kinds of noise specified by the two-sample variance become straight lines in a graph when both axes use logarithmic scales. The noise type can be identified from the gradient. Note that white noise phase modulation and flicker noise phase modulation have the same gradient in the two-sample variance and therefore cannot be differentiated. This is said to be the only defect of the two-sample variance.

Table 13.1
Relationship Between Power Spectral Density and the Two-Sample Variance

Power-law spectra	$S_y(f)$	$S_\varphi(f)$	$\sqrt{\sigma_y^2(\tau)}$
White noise phase modulation	$h_2 f^2$	h_2	$\sqrt{\frac{3f_h h_2}{(2\pi)^2}} \bullet \tau^{-1}$
Flicker noise phase modulation	$h_1 f$	$h_1 f^{-1}$	$\sqrt{\frac{\left(1.038 + 3\ln\left(2\pi f_h \tau\right)\right)h_1}{(2\pi)^2 \tau^2}} \approx \sigma_1 \times \tau^{-1}$
White noise frequency modulation	h_0	$h_0 f^{-2}$	$\sqrt{\frac{h_0}{2}} \bullet \tau^{-\frac{1}{2}}$
Flicker noise frequency modulation	$h_{-1} f^{-1}$	$h_{-1} f^{-3}$	$\sqrt{2\ln 2 h_{-1}} = \sigma_2 \tau^0$
Random walk frequency modulation	$h_{-2} f^{-2}$	$h_{-2} f^{-4}$	$\sqrt{\frac{2\pi^2 h_{-2}}{3}} \bullet \tau^{\frac{1}{2}}$

References

[1] Barnes, J. A., et al., "Characterization of Frequency Stability," *IEEE Trans. IM,* Vol. IM-20, No. 2, 1987.

[2] NBS, *Time and Frequency: Theory and Fundamentals,* NBS 1974.

[3] Kartaschoff, P., *Frequency and Time,* London: Academic Press, Inc., 1978.

[4] Bendat, J. S., and A. G. Piersol, *Random Data: Analysis and Measurement Procedure,* New York: John Wiley & Sons, Inc., 1971.

14

Noise Measurement

When clock specifications are discussed, noise analysis of clock signals is inevitable. Parameters described in Chapter 13 are used to specify those clock characteristics that are influenced by noise. While practical data can be measured with commercial measuring equipment, some understanding of the physical mechanism in the measurement process is required to get accurate data and to process it. This chapter explains the measurement mechanisms and shows the connection between variance in time domain and power spectrum in frequency domain.

14.1 Variance Measurement in the Time Domain

14.1.1 Measurement of the Two-Sample Variance with a Frequency Counter

Frequency counters directly measure the signal frequency from which the two-sample variance can be calculated. A frequency counter measures, however, the mean frequency within its gate time from t_i to $t_i + \tau$ and cannot record the instantaneous value. Naturally, τ corresponds to the time over which the frequency averaged. Therefore, the normalized frequency deviation can be expressed as

$$\bar{y}_i = \frac{\bar{f}_i - f_o}{f_o} \tag{14.1}$$

where f_o is the nominal frequency of the signal to be measured and $\bar{f}_i$ is the actual measured value. The two-sample variance can be calculated by repeating the frequency measurement M times, based on (14.2). $M-1$ differential data is obtained by these measurements.

$$\sigma_y^2(\tau) = \frac{1}{2(M-1)} \sum_{i=1}^{M-1} \left(\bar{y}_{i+1} - \bar{y}_i\right)^2 \tag{14.2}$$

Here the number of measured samples is finite and is limited to $M-1$ different values, and thus, (14.2) differs from the two-sample variance defined in (13.25). In practice, of course, we cannot help but use (14.2). A practical two-sample variance calculation based on (14.2) incurs some deviation from the true two-sample variance. Therefore, if a calculated two-sample variance is given, the number of data points should also be provided. The other problem is that a frequency counter normally needs some processing time for each single measurement whereby some dead time is introduced between two samples. This means that the measurement period T and the electronic measuring time τ (which corresponds to averaging time for frequency) are not equal. Again, some additional error is generally created [1]. In this case, the deviation from the true value depends on the noise characteristics and on the τ/T ratio. We should thus also specify the respective measurement conditions such as dead time.

14.1.1.1 Limitation of Direct Measurements with a Frequency Counter

Consider the case of measuring a clock directly with a frequency counter as illustrated in Figure 14.1. The measured points are used to calculate the two-sample variance of the clock. It is obvious that the counter resolution imposes one limit. If it can display 10 digits at a nominal frequency of 10 MHz, the display will be limited to 2 decimal places (i.e., 10,000,000.00 Hz). In this case, the display error is the least significant digit (0.01 Hz), and a relative frequency deviation of 10^{-9} can be measured. If two-sample variance is calculated from such data, the measurement limit is $5 \times 10^{-10}/\tau$ as shown in Figure 14.2. White PM noise and flicker PM noise in a clock signal cannot be measured if they fall below this because their characteristics match the trend exhibited by the measurement limit. If τ is large, it might be possible to measure white FM noise and flicker FM noise since their characteristic has a gentler slope.

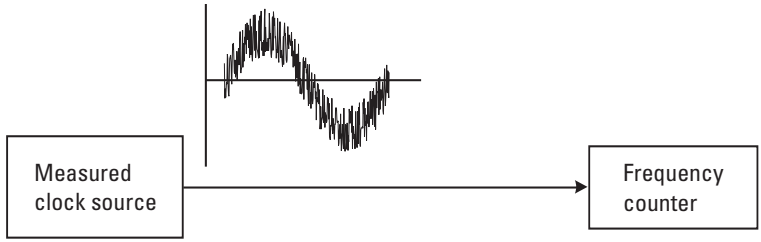


Figure 14.1 Direct frequency measurement with a frequency counter.

14.1.1.2 Heterodyne Frequency Measurement

Unfortunately, the resolution is automatically set when the nominal frequency is selected in a counter with a fixed number of digits. The heterodyne frequency measurement method (see Figure 14.3) overcomes this limit. Here a beat signal corresponding to the frequency difference between the clock to be tested and the reference is first generated with a mixer and a lowpass filter and then measured with a counter. As the frequency of the beat signal is low,

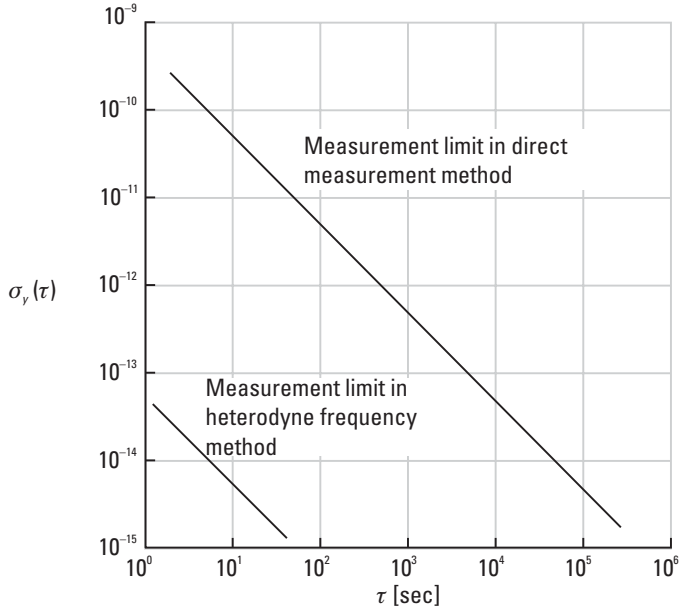


Figure 14.2 Comparison of measurement limits.

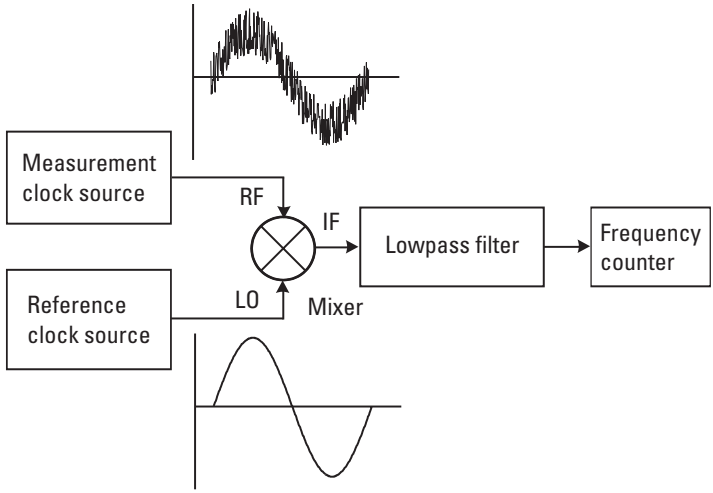


Figure 14.3 Heterodyne frequency measurement method.

the frequency resolution is better. Because the variation of the signal frequency f_o can be expanded to the beat frequency f_b the resolution will be improved by a factor of f_b/f_o. Therefore, it is effective to select the lowest beat frequency possible in order to improve the measurement resolution. However, there are two limitations for this as well:

1. The beat frequency cannot be less than the frequency variation of the initial signal.
2. We cannot measure variance over shorter time intervals than the period of the beat signal.

The beat frequency is often set to about 1 kHz. In this case, if the signal frequency to be measured is 10 MHz, the measurement resolution is improved by a factor of $1/10^4$.

When the two-sample variance of a clock is measured with these methods, the frequency reference of the measurement is, in the direct frequency measurement method, the internal oscillator in the counter. In the heterodyne frequency measurement approach, the frequency reference is provided by a separate source connected as a terminal LO (see Figure 14.3). Measurement results are influenced by the noise present in the reference source. The two-sample variance includes the degradation due to these. The two-sample variance measurement result, $\sigma^2_{yMEAS}(\tau)$, accurately follows

$$\sigma^2_{yDUT}(\tau) = \sigma^2_{yMEAS}(\tau) \tag{14.3}$$

if the variance of the reference, $\sigma^2_{yREF}(\tau)$, and that of the clock, $\sigma^2_{yDUT}(\tau)$, satisfy

$$\sigma^2_{yDUT}(\tau) >> \sigma^2_{yREF}(\tau) \tag{14.4}$$

If $\sigma^2_{yREF}(\tau)$ is equal to $\sigma^2_{yDUT}(\tau)$, halving $\sigma^2_{yMEAS}(\tau)$ gives the clock variance as

$$\sigma^2_{yDUT}(\tau) = \frac{\sigma^2_{yMEAS}(\tau)}{2} \tag{14.5}$$

The true variance can be easily obtained from the measurement results when the variance of the reference source and the variance of the measured clock either differ widely or are equal. If these two are slightly different, the variance of the reference source must be evaluated separately.

14.1.2 Two-Sample Measurement with a Time Interval Counter

14.1.2.1 Direct Measurement

A time interval counter can measure phase and phase-time. It has two input ports for reference and measured signals and can measure phase-time between selected triggering points in the two signals. While the two-sample variance is calculated from frequency deviation, it can be obtained from phase-time measurements as well. Frequency deviation calculated based on frequency data equals phase-time gradient over time as

$$\bar{y}_i = \frac{\Delta f}{f_o} = \frac{\Delta x}{x} \tag{14.6}$$

where Δx is the phase-time change during interval time x.

In the phase-time measurement approach, the measuring time can equal the evaluation period as shown in Figure 14.4. The two-sample variance can be obtained from its definition. In this method, the gaps between time interval measurements are set to be the same as the measurement period, T. The two-sample variance values at different τ can be obtained by changing the

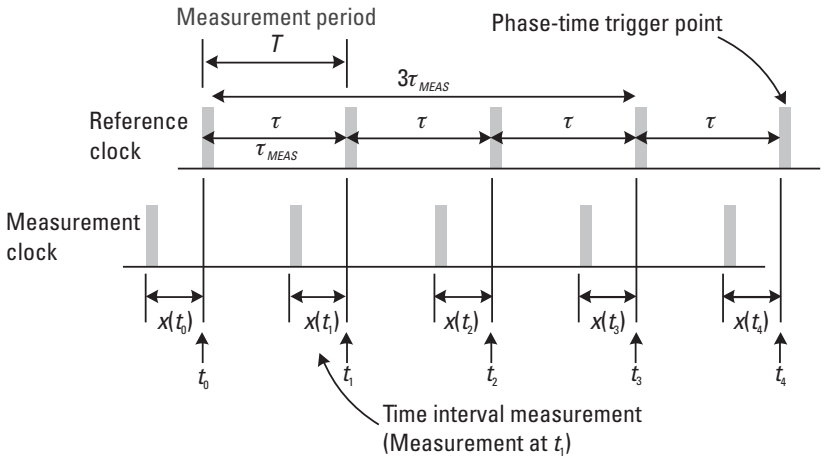


Figure 14.4 Two-sample variance measurement using phase-time data.

measurement interval. If τ is a multiple of the smallest measurement interval τ_{MEAS}, the two-sample variance can be calculated from data collected in each measurement period. Other measurement periods are possible; for example, over three times the smallest interval as illustrated in Figure 14.4 for measurements of $x(t_o)$ and $x(t_3)$. Note that the time interval counter must be configured for single shot operation. Usually a mode for averaging is available. This effectively reduces random variation of the measurement data, but this gives false results from two-sample variance measurements.

14.1.2.2 Dual Mixer Method

The heterodyne frequency measurement method described in Section 14.1.1.2 requires that the frequency of the reference source be freely variable in order to generate suitable beat frequencies. The frequency stability of the measuring instrument generally imposes a limit. This is one reason why this method is not often used when testing highly stable clocks. The technique that solves this problem is the dual mixer method. The phase-time of two beat signals at the outputs of respective mixers is measured by a time interval counter as indicated by Figure 14.5. No offset reference source is needed. The reference frequency can equal that of the measured signal. The relative short-term frequency stability between independent and identical clocks can be measured based on the relationship shown in (14.5). This is very important when highly stable clocks are needed. A reference, which has higher

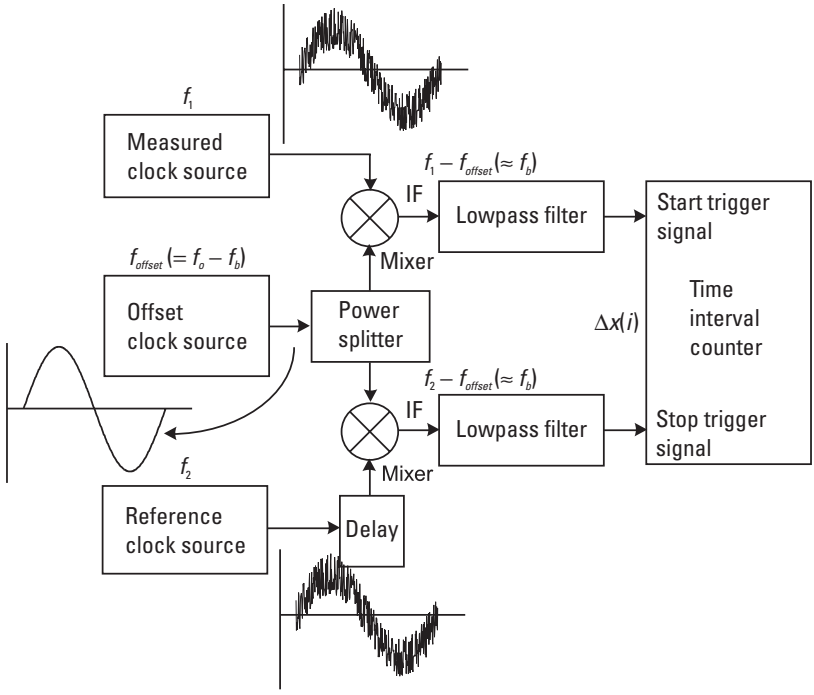


Figure 14.5 Dual mixer method for frequency measurement.

stability than the measured clock, seems impossible since the clock under test is often the most stable frequency source available.

In the dual mixer method, two beat signals (their frequencies are approximately f_b) appear at the mixer outputs. One is the frequency difference between the clock under test (having a frequency of f_1) and the offset source (frequency f_{offset} (=f_o–f_b)). Then f_o is the nominal value of f_1 and f_2. Since the short-term frequency stability of the offset source has an equal influence on the two beat signals, it will be theoretically canceled. Therefore, their frequency deviation is calculated from the measured time interval values $\Delta(i)$ as

$$y_{f1f2}\left(i, \frac{n}{f_b}\right) = \frac{\Delta x(i+n) - \Delta x(i)}{n} f_b \tag{14.7}$$

Because the time interval counter cannot measure phase-time difference during a period that is shorter than the beat signal period, the

normalized frequency deviation y_{f1f2} can be obtained only as the average of n measurements as indicated by (14.7).

14.2 Spectrum Measurement in the Frequency Domain

14.2.1 Direct Measurement of Frequency Spectrum

A clock, which includes amplitude and phase noise, can be treated as

$$v(t) = [v_o + v_n(t)]\sin[2\pi f_o t + \theta_n(t)] \tag{14.8}$$

This equation is the same as (13.2). When this signal is observed with a spectrum analyzer, which can directly analyze frequency components, the spectrum of the signal voltage can be obtained.

The signal voltage spectrum indicates the phase noise itself if $v_n(t)$ in (14.8) is very small compared to the signal voltage and if we can write for the phase noise

$$\theta_n(t) << 1\,\text{rad} \tag{14.9}$$

When the clock spectrum is measured (see Figure 14.6), $\mathcal{L}(f)$ over frequency, f_{rf}, is expressed by

$$\mathcal{L}(f_{rf}) = \frac{P_{rf}}{P_c} \tag{14.10}$$

Then $S_\varphi(f)$ can be calculated based on the relationship of (13.22) as

$$S_\varphi(f_{rf}) = 2\frac{P_{rf}}{P_c} \tag{14.11}$$

Note that when $S_\varphi(f)$ is calculated from the measured values as illustrated in Figure 14.6, it is necessary to consider the frequency band from which the spectrum is obtained. $S_\varphi(f)$ has units of dBc/Hz (that is, power per unit frequency). While some digital signal spectrum analyzers can automatically display the result, analog type spectrum analyzers usually require manual computations from the bandwidth in which the noise power is measured. If an analog spectrum analyzer is used for noise measurements, its IF filter

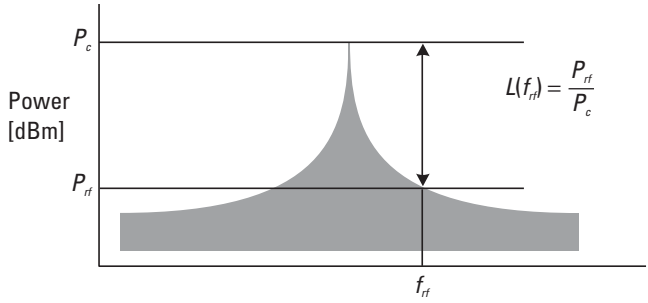


Figure 14.6 Frequency spectrum of a clock signal.

should be evaluated first. If we assume a Gaussian-type voltage response, the equivalent noise bandwidth is generally about 1.2 times this.

14.2.2 Phase Synchronization Method for Phase Noise Frequency Spectrum

The dynamic range of spectrum analyzers is generally around 100 dB. In the direct measurement of a frequency spectrum, this sets the limit at –100 dB/Hz for measurements with 1-Hz bandwidth. A phase synchronization method can be used to improve the situation. This technique relies on another clock that is synchronized (in phase) to the clock under test as illustrated in Figure 14.7. The phase difference between the clock under test and the reference is first detected with a mixer. Its output is fed back to the reference clock, and phase synchronization is thus achieved. This mechanism is basically a PLL. The feature of this technique is that signal power and noise can be independently measured, and the limit due to the dynamic range of the spectrum analyzer can be eliminated. Signal power corresponds to the power of the beat signal that

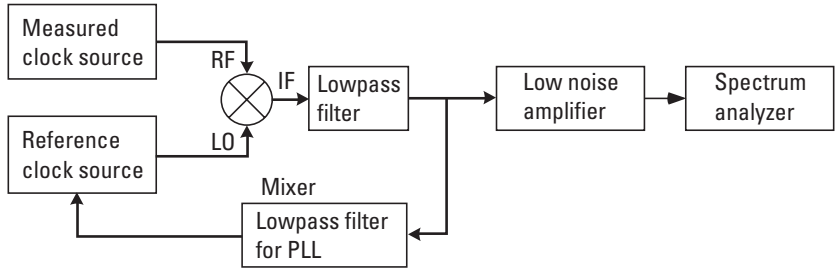


Figure 14.7 Phase noise measurement with the phase synchronization method.

appears at the mixer output when the phase-locked loop is opened. The frequency of the reference can be synchronized to that of the source under test by the phase lock mechanism, so the noise differential between two sources appears in the mixer output. If the noise characteristics of the two sources are identical, the measured noise will double.

In the noise power measurement, the input sensitivity of the spectrum analyzer can be set to suit the noise level. As an example, a commercial instrument with a sensitivity of –160 dBc/Hz has been released. The drawback of this method is that noise below the cutoff frequency allowed by the phase lock mechanism is measured as being lower than the true value of the measured spectrum shown in Figure 14.8. The measured characteristic of Figure 14.8 differs from the directly measured characteristics of Figure 14.6. The carrier, the center frequency of the clock under test, disappears, and the noise power is shown in the baseband. The phase lock mechanism enables the reference clock to be completely synchronized to the clock under test at frequencies below the cutoff frequency. Consequently, the noise differential becomes very small. There are commercial instruments that can compensate for the cutoff characteristic and can display the true value. We also note that this method often fails to measure the noise characteristics of clock sources whose purity is low, since such sources make phase synchronization difficult.

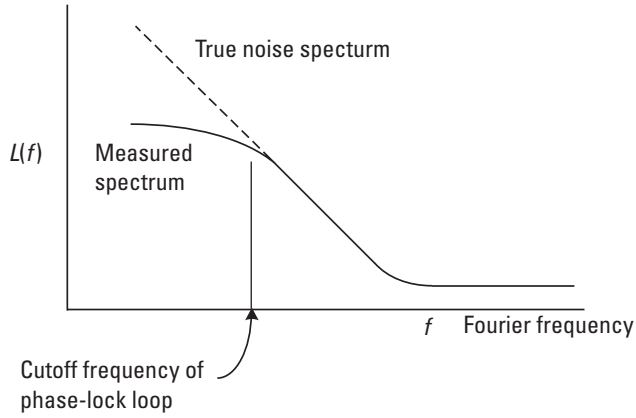


Figure 14.8 Frequency spectrum measured by the phase synchronization method.

Reference

[1] Howe, D. A., D. W. Allan, and J. A. Barnes, "Properties of Signal Sources and Measurement Method," *Proc. 35th Annual Frequency Control Symposium,* 1981, pp. A1–A47.

15

Operational and Environmental Effects on Performance Characteristics

Clock and synchronization technologies are related to every individual system in a communication network. When we design such a network function, it is very important to understand the mutual relationships and to investigate the influence of the selected synchronization scheme on the performance of network elements, or vice versa. In this chapter, we describe some network and system configurations and their characteristics related to the synchronization function and various clock technologies.

15.1 Performance Characteristics in Wired Transmission Systems [1]

15.1.1 Network Configurations

First we have to consider some network elements in order to configure an information transmission system. A transmission medium must exist between an information server and the client as shown in Figure 15.1. This medium does not merely mean cables for wiring or air for propagating radio signals. It includes various components such as transmission and switching devices. Optical transmission and multiplexing functions such as SDH systems provide broadband services. Moreover, switching and cross-connect systems such as ATM and add and drop multiplexer (ADM) contribute to

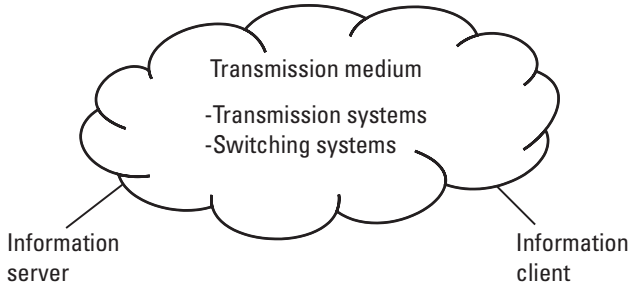


Figure 15.1 Network elements in information transmission.

the reduction of network cost for carriers and consequently to a reduction in the charges made to customers. Figure 15.2 gives an example of an actual network configuration where computers transmit and receive information. An access network connects the customer premises with a carrier building. An access system and a router are included in this network. The information delivered to the carrier building is transferred to a transmission and switching system, and it travels over a long-distance network. Domestic and international networks are composed of rings, which are now one of the most popular topologies.

15.1.2 Relationship Between System and Synchronization Characteristics

Network performance is now mainly specified by its error characteristics. It is difficult to find why and where errors occur. They are either extemporaneous or stochastic phenomena. In this chapter, the stochastic phenomenon is described. It is related to clock characteristics in a synchronization mechanism. Signal-to-noise ratio, which can be found from the clock characteristics, usually determines the error performance. In particular, phase variation, which is called jitter (see Section 13.2) in a digital system, influences error performance in clock characteristics. The term "jitter" fits the specifications of digital systems in the time domain. However, power spectral density in the frequency domain and variance in a time domain described in Chapter 13 should be used to design and analyze communication networks. These measures can accurately define synchronization characteristics. The current communication network consists of many systems as shown in the example of Figure 15.2. The total characteristics of a composite network can be found by using the respective values of each element.

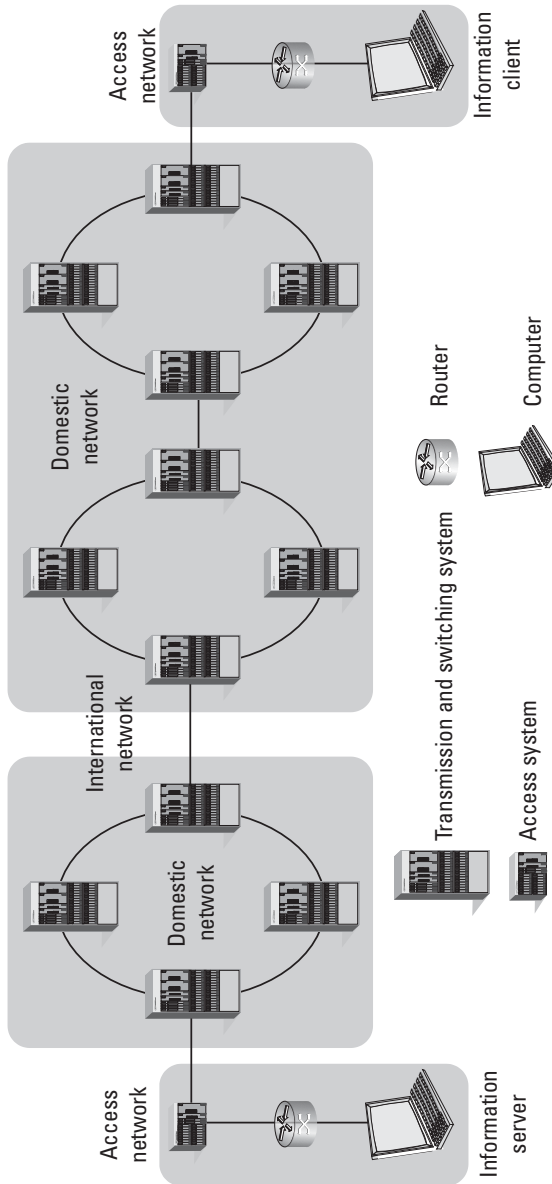


Figure 15.2 Example of an actual network transmitting information between computers.

Both asynchronous and synchronous systems exist in the current network. An independent system clock determines synchronization performance in the asynchronous system. In the synchronous system, the common clock (see Chapter 7) has an important role as well as the system clock. The network performance is influenced by the following items:

1. System clock characteristics at the transmitting site;
2. Transmitted system clock as received at the receiving site;
3. Independent system clock at the receiving site;
4. Common clock at the transmitting and receiving sites.

Figure 15.3 indicates the relationship between these clocks.

In transmission systems, the system clock usually determines the output signal characteristics. Independent clock characteristics in an asynchronous system appear in the output. Since the system clock synchronizes to the common clock in the synchronous system, the output clock characteristics are equal to those of the common clock.

An actual network example using synchronous technology is described in Figure 15.4. It consists of four synchronous systems, S1, S2, S3, and S4, in the ring configuration. Systems S1 and S2 directly receive the common clock from a master node over the respective transfer network. System S3 receives the common clock via S2. The common clock of S3 is transferred to the

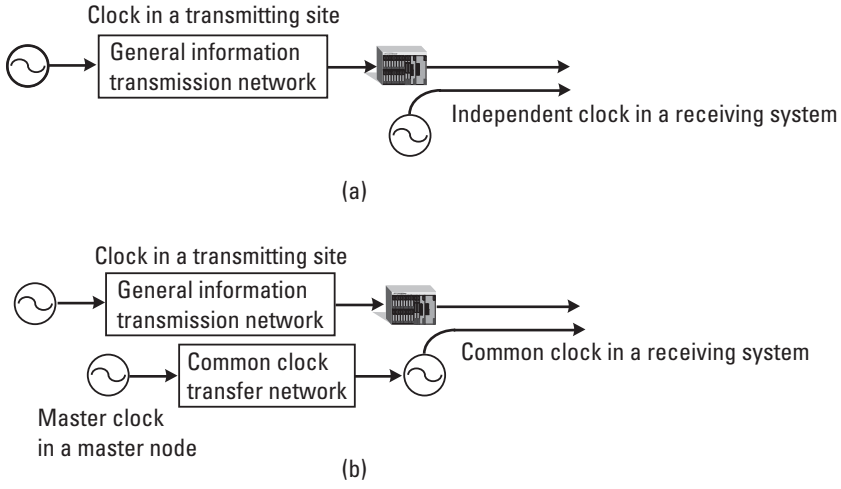


Figure 15.3 Clock relationships: (a) asynchronous system and (b) synchronous system.

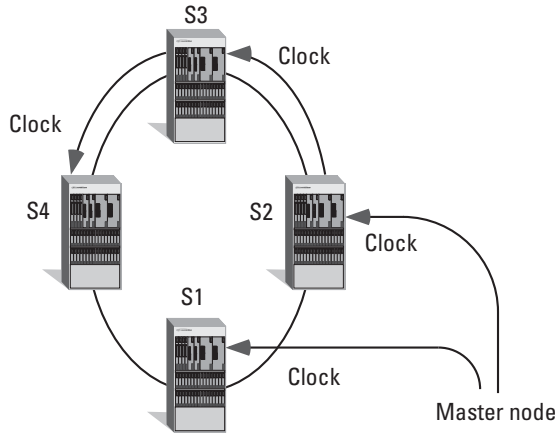


Figure 15.4 Common clock in synchronous systems.

synchronous system S4. The synchronization performance in S3 and S4 is influenced by both the common clock transfer network and other synchronous systems. Thus we have to consider clock characteristics of various elements to determine the synchronization performance in real networks.

15.1.3 Wired Transmission Characteristics

Clock performance in transmission can be specified by phase variation in the medium and the clock regeneration capability, which is described in Chapter 7, at the receiving site. If the system clock is generated by a stable oscillator, whose phase variation is small enough, and if the transmission paths do not degrade the clock characteristics, theoretically transmission errors will not occur. The reason to investigate network performance is to know how degraded the clock signal is at the receiving site. In particular, a firm grasp of transmission characteristics is very important for network design and management.

Phase variation is a useful physical quantity for analyzing clock performance in networks, since the error performance is directly influenced by clock phase changes and overall signal phase as generated from the clock. Phase variations can be expressed by

$$\Delta T(t) = at + \frac{1}{2}bt^2 + c + d(t) \tag{15.1}$$

where a represents an initial frequency offset, b is the phase change that results from a frequency change over elapsed time such as aging and drift, c is the phase offset during a transition such as a system switching to a redundant function, and $d(t)$ is phase variation due to noise.

The clock regeneration function at the receiving site can synchronize to the transmitted clock, which is initially generated by the system clock at the transmitting site, so phase changes corresponding to a and b in (15.1) do not theoretically occur. When a clock regeneration LSI based on the PLL concept of Chapter 9 is used, the clock phase slightly changes over time due to the PLL steady-state phase error (see Section 4.5). However, since this corresponds to the aging rate of oscillator in the LSI, it is very slow compared to the observation time. The phase offset c can be ignored under normal conditions, so only the phase variation due to noise can be consequently observed in one link transmission such as a transmitter-receiver combination.

While phase variation is a useful quantity in digital systems, the power spectral density is better for stochastic noise in networks. We can specify each characteristic clock element by it and can also estimate the total performance of a cascaded clock system. Since the power spectral density can be theoretically related to the time domain parameters (see Chapter 13), data based on it can be translated into phase variation whenever required.

Figure 15.5 shows an example of the power spectral density of clock characteristics in a transmission system. Here, $\mathcal{L}(f)$ (see Chapter 13) is used to show the relative magnitude of noise power spectral density against the carrier power. Flicker noise phase modulation (see Chapter 13) is observed below approximately 1 kHz. Most current digital networks show the same tendency. The noise power rapidly decreases above 1 kHz. This is caused by the PLL-based clock regeneration function. The threshold frequency varies with the PLL lowpass characteristic (see Chapter 4). Since a multilink network configuration, in which a combination of transceivers is repeated, is often adopted in actual networks, the total noise can be calculated by multiplying the noise power shown in Figure 15.5. The noise power basically goes down with an increase in the transmission bit rate. Jitter can be calculated by integrating the noise power over the observation frequency band. Actual jitter in the current optical transmission systems is typically much smaller than 1 ns.

In Figure 15.5, the characteristics of asynchronous multiplexing systems, which are using a pulse-justification mechanism, are depicted as well. The number of these systems is gradually decreasing in current networks. If a network includes such elements, however, the configuration and its performance should be carefully designed. The power spectral density is much larger than in other optical transmission systems.

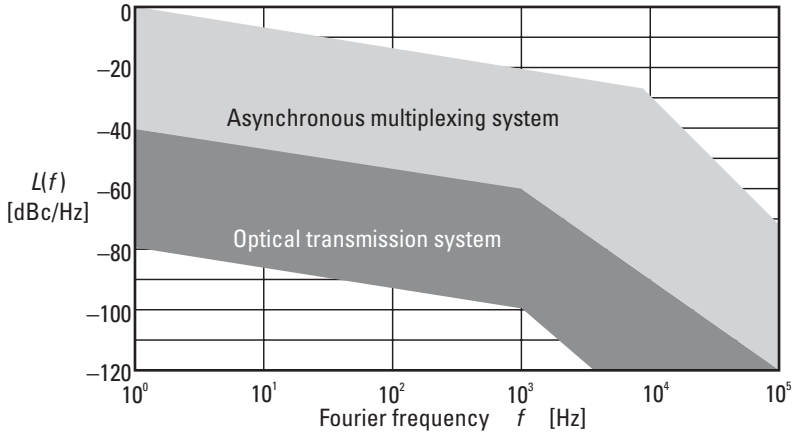


Figure 15.5 Noise power to the total signal power in different transmission systems.

15.1.4 Synchronization Characteristics

The common clock characteristics can be found by combining those of the clock transfer network and of the clock devices themselves. They are called the digital synchronization system as described in Chapter 8. The transfer network performance is just that of the wired transmission path. The common clock system functions as a lowpass filter and is typically constructed by digital processing technologies as was indicated in Chapter 8. The cutoff frequency can be set as low as 10^{-5} to 10^{-3} Hz. Therefore the noise power in Figure 15.5 is mostly eliminated in the output of the digital synchronization system, and the noise characteristics of the internal oscillator dominate.

15.2 Performance Characteristics in Radio Systems

This section gives the reader an introduction to the more specialized topics of frequency control, synchronization, and timing arrangements in various radio systems. A general description of the fundamental differences between radio and cable communication networks will be given. The reader is encouraged to use Chapter 7 as a point of reference. As opposed to networks, which can solely utilize either optical or copper cables as their transmission media, radio systems are characterized by a relatively uncontrollable path structure between the transmitter and the receiver. This includes spatial and temporal variations in the complex frequency response and a multitude of external signals, against which only limited measures are available. A very

specific feature of the communications air interface is its transmission bandwidth. If no limitations were set by the terminal electronics or other hardware, the media itself would allow an almost endless frequency range for the user. Based on the more realistic performance of the individual manmade elements, however, we can estimate the available communication spectrum to be a linear function of the selected center frequency. Particularly at the higher UHF bands dedicated for cellular communication networks, the term "radio channel" or simply "channel" is used to collect most of the performance-related data. Thus, we speak about channel prediction, channel delay, and channel models, just to name a few examples.

Radio networks are essential for mobile users of telecommunications services although a huge number of fixed microwave links currently exist. This implies that we are not able to define exact locations of terminal equipment nor can we accurately find their motional state. Much of the user electronics like mobile phones or airborne communication devices are subject to very harsh handling and various environmental effects whereby their technical performance may face complications. Another problem is the requirement for very cost-effective units because of the huge manufacturing volumes involved when aiming at the general consumer sector. All these factors have an unavoidable impact on the frequency generation and synchronization schemes, which have to be adopted in radio communication. Their scope extends itself from the individual circuits up to the highest hierarchical network levels.

15.2.1 Some Characteristics of the Air Interface Regarding Synchronization

A selected part of the atmosphere—or in many cases respective parts of the adjacent near-Earth space—replaces the cable connection between the transmitter and receiver of the communication link. Unfortunately its fundamental electrical properties cannot be defined as precisely as the performance of a coaxial or optical cable. Particularly noteworthy is the poor stability of the propagation path as a function of elapsed time. In addition to this, the wide range of operating frequencies and associated physical phenomena complicates the process.

Typically, the analysis of a radio communication path relies on a combination of some elementary calculus based on idealized—or so-called free space—conditions followed by semiempirical evaluations specifically valid for the case at hand. These may include the effects of the signal environment (urban, rural, space, etc.) and often use additional information about the

temporal characteristics of the medium (rain rate, electron density of the ionosphere). Naturally, this phase takes into account the specific frequency or frequencies in use.

15.2.1.1 Frequency Allocations for Radio Communication Systems

Almost the entire technically available electrical spectrum from about 10 kHz up to 100 GHz and more can be and is used for communication over the air interface. The specific frequency selection is still partly based on historical reasons but has gradually evolved towards the more feasible approach. This means that the lower part of the spectrum, typically below 400 MHz or so, is reserved for LANs with a relatively sparse topology, and the higher microwave bands are more used for limited point-to-point solutions or for the recently introduced short-range high-capacity systems. Naturally, the unavoidable size constraints, set by the requirement of sensible antennas, partly steers this process. Thus, although the low VHF bands would allow good penetration characteristics, for example, in rural areas, their use in mobile radio is limited by the clumsy radiating elements. Another issue is the frequency-specific propagation scheme, which generally favors the higher frequencies for more precise targeting. The final but very important guiding factor in the frequency selection chain is the available communication bandwidth or data rate, which obviously is directly related to the respective carrier frequency. A brief summary of the currently used radio frequency bands, available baseband bandwidths, and some examples of present communications-related users is shown in Table 15.1. The reader should note that recently introduced digital processing algorithms seem to be able to multiply the baseband performance of HF communication networks.

15.2.1.2 Free Space Propagation and Thermal Noise

The case of a single transmitter and its receiving counterpart hanging up in deep space is relatively easy to handle in terms of radio frequency propagation. Let us first assume the following parameters for the two communication sites:

- Transmitter output power: P_T;
- Transmitting antenna gain towards the receiving site: G_T;
- Receiving antenna gain towards the transmitting site: G_R;
- Received signal power: P_R;

Table 15.1
Radio Communication Bands, Typical Installed Baseband Capacities, and Examples of Current Communication Users

Frequency Range (MHz)	Baseband Bandwidth (MHz)	Some Users
0.01–0.15	< 0.01	Long-range military (global)
0.15–1.5	< 0.01	Broadcasting, military (continental)
1.5–30	< 0.01	Broadcasting, military, amateurs, public short-range
30–140	< 0.02	Broadcasting, military, aviation
140–180	< 0.02	Land mobile, maritime
180–400	< 0.2	Broadcasting, aviation,
400–1,000	< 1	Cellular
1,000–2,500	< 2	Cellular
2,500–15,000	< 20	Short-range cellular, point-to-point, satellites
15,000+	< 200	Very short-range cellular, satellites

- Operating wavelength, defined simply as c/f: λ;
- Path distance along a straight line: *r*.

Based on these, we can write the equation that connects the received signal power P_R to the other prevailing conditions as

$$P_R = \frac{G_T P_T}{4\pi r^2} \bullet \frac{\lambda^2}{4\pi} G_R \tag{15.2}$$

This formulation is valid in the optimum situation when no other attenuating factors appear and when, for example, the respective receiving and transmitting bandwidths agree. The derivation of the carrier-to-noise ratio (*C/N*) is quite straightforward as well when we limit the discussion to thermal noise sources or to those that can be conveniently converted into such for computational purposes. If the equivalent noise temperature of the radio receiver is T_N and the respective noise bandwidth is B_N, we get the noise power at the antenna interface as

$$P_N = kT_N B_N \tag{15.3}$$

Following this, the *C/N* is

$$\frac{C}{N} = \frac{G_T G_R P_T \lambda^2}{(4\pi r)^2 k T_N B_N} \tag{15.4}$$

There is an inherent frequency response in the radio channel, as indicated by (15.2). Almost always, however, the relative bandwidths occupied by the modulation and the intrinsic baseband signal are only differential compared to the carrier frequency. We can thus neglect the amplitude response for all practical radio communication arrangements excluding UWB or impulse radio, in which the bandwidth relations are completely reversed. This topic is further discussed in Section 7.2.2.5.

The other, sometimes most important, issue for synchronization in the radio communication link is its time delay. For the elementary case of free space propagation along a straight line, there is no dispersion—the signal will travel at the speed of light in vacuum regardless of its frequency. In other words, the group delay τ_g over the air interface is constant and defined only by the straight line distance as

$$\tau_g = \frac{c}{f} \tag{15.5}$$

Unfortunately, (15.5) is not valid in most practical radio communication channels for the reasons that are outlined below.

15.2.1.3 The Nature and Effects of Multipath Propagation

In most terrestrial radio communication links the signal propagates along several nonparallel paths from the transmitting antenna to the receiving site. Basically, this is true regardless of carrier frequency. For example, the common fading characteristics of short wave reception are actually caused by multiple signal paths, but traditionally this case is considered as ionospheric propagation. In radio engineering literature, multipath phenomena are normally restricted to frequencies above about 30 MHz. This means that most manmade objects of adequate size, and particularly the surface of the Earth, can be considered to be nearly perfectly conducting. In its simplest form, the situation can be discussed following the case of Figure 15.6. If the propagation environment has only one indefinitely large conducting plane, above which the transmitting and receiving sites are located, we will have two signal components. The first goes along the direct, shortest path and the second

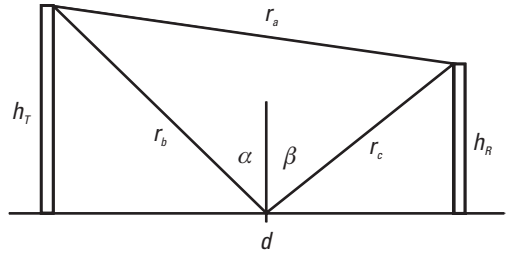


Figure 15.6 A simplified multipath scenario above a conducting plane (e.g., the surface of the Earth).

gets reflected from the conducting surface and follows Snell's law, whereby $\alpha = \beta$. Thus, the discussion is often referred to as geometrical optics.

The transmitting and receiving antennas are elevated above the plane and the direct and reflected signal paths are established.

The first task is to find out the resulting amplitude of the received signal. A very good approximation is that the reflection at the Earth's boundary is lossless. After some elementary vector manipulations, the field strength at the receiving site can be expressed as

$$E_R = 2E_0 \sin\left[2\pi \frac{h_T h_R}{\lambda d}\right] \tag{15.6}$$

where E_0 is the respective free space signal strength as calculated from (15.2). In normal communication networks the two individual signal paths are of almost equal length, whereby the attenuation that they have experienced will be the same. As indicated by (15.6), the worst-case result would be a complete loss of signal, and the best situation would provide double the level compared to the undisturbed scenario. The field strength is a function of instantaneous frequency but again the fractional differences caused by conventional baseband signals are quite small and we cannot expect any remarkable amplitude changes due to modulation only. A completely different situation will be at hand, however, if the carrier frequency is intentionally shifted during the communication process across a considerably larger span. This scheme is called frequency hopping or spread spectrum (for detailed discussions, see Sections 7.2.2.3 and 7.2.2.4).

In the time domain much depends on the difference of the path lengths divided by the propagation velocity, which simply gives us

$$\Delta\tau = \frac{\left[r_a - (r_b + r_c)\right]}{c} \tag{15.7}$$

Distortion starts to appear when the difference in the arrival times becomes a substantial fraction of the actual baseband cycle. It is obvious that high-speed data transmission over a mobile radio interface is one of the most vulnerable processes regarding this defect. Channel sounder measurements have indicated delay spread values up to about 3 μs with an average around 1 μs in urban environments and a maximum of 20 μs in rural areas. This means that communication systems with air interface clock periods on the order of 10 μs or less need special arrangements in built-up areas. A typical deterioration is called inter symbol interference (ISI). Naturally, the adverse effects of a very large differential delay will be mostly—but unfortunately not always—canceled out by a high attenuation of the reflected signal.

15.2.1.4 Statistical Propagation Parameters

The general radio propagation scenario is too complicated and unstable to be handled analytically. Statistical methods, either computerized or simply tabulated, are utilized instead. The tasks to be taken care of include, but are not limited to, the following cases:

- General temporal and spatial variations of the propagation environment;
- Numerous and moving reflecting bodies and related models;
- Effects of the ionosphere;
- Channeling;
- Rain and atmospheric particle attenuation.

First of all, the typical multipath scenario is comprised of a multitude of individual signal components. Reflections from buildings and terrain obstacles create even in the steady state a vector field in which computational efforts to define the amplitude or timing relations tend to be frustrating. The clear benefit, however, is that we get rid of the deep minima and are generally able to provide sufficient coverage also in non-line-of-sight (NLOS) conditions.

The real-life multipath environment is far from stable. If we are facing ionospheric reflections at HF frequencies, we must take into account the continuous fluctuations of electron density and equivalent height of the

reflecting layer. Here the time constants are on the order of some seconds or some tens of seconds. Algorithmic corrections, supported by dedicated hardware, have been found feasible to counteract this phenomenon. On the other hand, the UHF multipath problem in an urban environment suffers from the rapid changes of the relative positions of the communication terminals. In addition to this, the reflecting background is changing as well, for example, because of moving vehicles. This means that there is no straightforward way of defining the timing relations as seen by the cellular system. Many currently employed digital radio communication arrangements use a special training sequence to probe the channel conditions and to perform the required reconstructive operations.

Very often the real wave propagation problem is approached in terms of an availability factor or its complement. They can be expressed as percentages or as absolute values. A distinct difference exists between mobile communication networks and fixed, normally point-to-point systems. The propagation path of cellular systems is typically defined as a percentage of time and place, but for tower-mounted microwave links, time is the only relevant parameter, of course. In addition to the purely statistical data, some quite effective semiempirical formulas have been created for the estimation of attenuation from water vapor, which can be written as

$$L_{atm} = \left[0.1 + \frac{(f - 30)}{30} 0.15 \right] \frac{dB}{km} \tag{15.8}$$

where the frequency f is to be given in gigahertz.

15.2.1.5 Synchronization and Frequency Planning in a Radio Network

The freely propagating radio frequency spectrum is a delicate natural resource and should not be wasted. In fact, many governmental communication authorities nowadays treat megahertz as a trade item that has financial value. When the number of radio transmitters was low and their mutual geographical distances were large, little to no coordination was necessary for a proper, undisturbed operational system. The path loss simply took care of any interference issues. Today, the main focus of radio network planning is to create an adequate coverage in terms of service quality while taking care of interference rejection. That is, the plan not only includes specific antenna sites and transmitter powers but also gives a frequency list, which limits both internal and external interference to a minimum. At the same time, a good

network plan aims at the most efficient frequency reuse inside it, whereby the overall communications capacity can be increased.

The frequency and timing stability issues have a surprisingly tight connection to modern radio network planning through a number of ways. First, the better we are able to control the individual carrier frequencies as such, the less guard spectrum is needed around each spectral channel. Second, in simple TDMA networks the uncertainty of synchronization in the time domain defines, to a certain extent, the necessary guard time between successive transmission intervals. This naturally influences the capacity of the system. Further benefits are available if the network plan can be based on completely coherent transmitters, whereby so-called orthogonality can be used to increase capacity or to reduce mutual interference compared to the free-running, noncoherent arrangement. The challenges of maintaining tight coherence are much higher in a mobile radio network compared, for example, to a physically stationary link system.

The timing, synchronization, and frequency control issues of many radio communication networks depend to a significant extent on the operational difficulties under which the mobile terminals operate. The mobility itself, the three-dimensional motion, and the positional uncertainty together constitute one factor. Unlike most fixed cable networks, the user of a radio terminal will face problems related to the power supply, and the communications capacity requirements hardly match the link possibilities. Naturally, very complicated questions are related to physical stresses like temperature variations and vibration, which are seldom a significant problem for permanently installed devices.

15.2.2 External Effects on Terminal Clock Generation

A typical radio communications user is mobile. It may be a person with a cellular phone or a fighter pilot shooting across the sky at twice the velocity of sound. Movements might be predominantly in one dimension or they might fill the whole available space. A mobile user seldom is able to run the communication device continuously but has to switch it on and off several times per day or mission. Additionally, extremes of temperature and vibration are often encountered in practice. This creates quite a different scenario compared to a fixed cable communication network.

15.2.2.1 Typical Oscillators Used in Mobile Clocks

Typical internal oscillators in radio communication terminals currently use either quartz crystals, SAW devices, or rubidium cells as their ultimate

frequency-defining component. From these, the final frequencies are derived either through simple multipliers, heterodyne mixing, phase locking, or more recently through DSP blocks. Regardless of the processing scheme, the environmental effects facing the fundamental oscillator have the greatest overall impact.

Crystal oscillators are normally divided into four groups according to their thermal regulation method. Room temperature units (RTXOs) are generally not suitable for mobile radio terminals due to their very large temperature dependence. No means to counteract the frequency changes with temperature have been implemented excluding a possible quartz cut optimization process. Temperature compensated oscillators (TCXOs) are an older analog concept where a crystal temperature measurement is used to provide a control voltage that adjusts the parallel capacitance of a varactor diode. The achievable frequency profile is sufficient for some basic mobile units up to an uncertainty level of 10^{-7}.

The modern DTCXO (where D stands for "digital") uses a similar compensating scheme but has an advanced look-up-table where the precise correction factors for each specific temperature can be found. Thus, the units can be made to take into account the nonlinear and nonmonotonic frequency characteristics of the fundamental oscillator block. The available frequency uncertainty against temperature changes is approximately 10^{-8} or even better. Ovenized crystal oscillators take a different approach because heating is used to adjust the temperature. Modern miniaturized double oven constructions have demonstrated uncertainties in the 10^{-10} range, which can be further enhanced through an internal DTCXO algorithm for the remaining temperature changes [2].

Rubidium oscillators are mainly used in military communication devices due to their less cost-effective construction. Actually, a rubidium unit must already contain an internal frequency conversion because the atomic resonance happens at approximately 6 GHz. Typically the commercially available designs incorporate special low-noise PLL circuits, the output frequency of which is often 5 or 10 MHz. As is indicated by the construction, most environmental constraints found in crystal oscillators will be valid for rubidium devices as well. If, however, the characteristics of the built-in crystal unit allow, a wideband PLL will transfer more of the better environmental stability of the atomic resonance to the final output frequency.

15.2.2.2 Effects of Position and Velocity

If we evaluate the situation only at discrete time intervals, we see the mobile user as a quasi-stationary case but always having a different geographical

location. This has three main consequences regarding synchronization and timing. First of all, the transmission delays will be different from the previous moment of time. Second, the signal amplitude will be different, which will cause the carrier-to-noise ratio to change. This might just lead to a deteriorated clock jitter but the signal might be totally lost too. Finally, the multipath scenario will be different from the previous one whereby the channel response is changed. Adding a real-time feature means that we have to face the Doppler effect in addition to all the previous features.

Different platforms have different challenges. A personal land mobile terminal will generally suffer from very severe multipath and fluctuations but the Doppler shifts are minimal. General land mobile units may already have bad enough Doppler and have the same multipath as pedestrians. An airborne radio, on the other hand, can often work in a much more decent multipath environment but must be able to handle very large Doppler frequencies. The same holds generally for satellite communications.

These factors cause the synchronization process to have a statistical nature. Besides this, the system should utilize closed loop measurements of radio channel characteristics whenever the technical features permit. The electronics and algorithms must provide the shortest possible time-to-resynchronize (TTRS). Naturally, the autonomous clock running time should be long enough to cover a temporary loss of carrier. Such arrangements are currently implemented in many cellular networks. A key feature is the channel equalizer, which is nowadays a DSP implementation. The means to create these characteristics are discussed further in Chapters 9 and 10.

15.2.2.3 Effects of User Profile

Most modern digital radio communication terminals must be able to comply with a very challenging combination of user wishes and practices, which altogether have an impact on the synchronization schemes. Although the mobile devices were initially developed for slow speed voice and perhaps related add-on data, we now must provide very high-speed data communications with the same radio interface. Thus the temporal variations of communications capacity are tremendous and the synchronization scheme must be able to adopt in almost real-time. Many users of the previous generation were satisfied with a modest availability of service, but now a quality and reliability similar to fixed lines is anticipated.

A very specific feature of many radio communication terminals is their operation on battery power and—partly related to this—the continuous on/off switching of the devices either to conserve energy or to allow for user privacy and comfort. This means that the internal oscillators must have very

rapid turn-on characteristics with the minimum of warm-up time, and the time-to-synchronize (TTS) must be extremely short despite the desired high baseband bandwidth. Many of the problems can be partly solved with the closed loop control of timing and frequency in those systems (including many current cellular networks) where an adequate uplink capacity is available. The mobile oscillators and synchronization circuits can mostly operate in slave mode. If not properly dimensioned, however, a "rubber-band" effect may appear as in any set of unstable control functions.

Depending on the platform where the mobile communication terminal is to be used, we must design it to withstand the prevailing temperature profile, shocks, and mechanical vibration [3] spectra without any loss of synchronization or carrier stability performance. The task is particularly challenging if we have to guarantee autonomous operation in a hostile environment. Table 15.2 shows some typical requirements both for commercial devices and military electronics [4, 5]. As can be seen, the largest differences appear for mechanical stresses where the requirements of military use are greater by at least an order of magnitude.

15.2.2.4 Temperature Effects

Most of the adverse effects in mobile synchronization circuits related to the external physical temperature are due to the change in dimensions. Frequency shifts are mainly caused by differences in the quartz size whereas phase or delay will change due to alterations in transmission line lengths. Two distinct types of temperature effects are normally indicated. First of all, there is a characteristic curve, which relates frequency to the steady-state temperature. Very often this function is far from linear and does not show any

Table 15.2
Some Environmental Design Parameters for Commercial and Military Mobile Radio Communication Oscillators

Parameter	Consumer-Grade Oscillator	Military-Grade Oscillator
Temperature range	–25 ... +65°C	–55 ...+75°C
Vibration	1.5 G/500-Hz sine	6 G / 2,000-Hz sine
Shocks	6 G/11-ms half-sine	50 G/11-ms half-sine
Magnetic field	3×10^{-11}/Gauss	3×10^{-11}/Gauss
Barometric pressure	3×10^{-11} @ 12 km	3×10^{-11} @ 12 km

monotonic nature. In medium-grade crystal oscillators the measured relative frequency offset due to temperature may be around 1×10^{-8} at +55°C. Another effect is the fast temperature transient, which may be encountered when, for example, a mobile radio terminal is moved from a heated cabin to an outdoor environment. Here the question is whether the temperature control and phase-locked loops can remain stable and avoid an exaggerated correction.

15.2.2.5 Shocks and Vibration

Oscillators in a fixed cable communication network seldom have to face mechanical shocks or vibration. In mobile radio communication terminals the situation is completely different. Physical movements have a direct influence on oscillator performance as long as the output frequency is somehow created by similar, yet microscopic vibrations. As we know, this is just the case in quartz oscillators. Also, the various resonating circuits in filters may face small changes in their shape or size due to external mechanical forces. This can cause a change in their Q value or detune them slightly. In a PLL circuit the mechanical vibration can sometimes get coupled through the filter, which might even make the loop unstable and break the lock.

A constant acceleration as found, for example, in a banking fighter aircraft will generally induce a frequency bias, which would try to push the clock system out of sync. However, a continuous banking for several minutes is highly unlikely due to physiological and even structural limitations. Maritime and space born oscillators may face such heavy shocks that their fundamental frequency will never show a perfect retrace. Very high vibration levels will naturally destroy the mechanical crystal mounting.

15.2.2.6 Aging and Retrace

The internal oscillators of mobile radio communication devices may be stored for long periods of time and then be taken into operation with a very short notice. As was discussed in Section 15.2.2.3, the user often wants to switch off the device and reactivate it later. These two facts will create another challenge for the network synchronization process. An individual terminal may have experienced a relatively large shift in its fundamental frequency, maybe on the order of 10^{-6}, without any prior information to the network management system. Alternatively, the frequency retrace will never be perfect even after a short off-period, but the equipment user might assume immediate service after activating the unit again.

References

[1] Bregni, S., *Synchronization of Digital Telecommunications Networks,* New York: John Wiley & Sons, 2002.

[2] Neubig, B., and W. Riese, *Das grosse Quarzkochbuch,* Feldkirchen, Germany: Franzis' Verlag, 1997.

[3] Eskelinen, H., and P. Eskelinen, "Shock Tests with a Phase-Locked Crystal Oscillator," *Proc. European Frequency and Time Forum,* Turin, Italy, 2000, pp. 324–328.

[4] Ball Efratom Corporation, *Precision Time and Frequency Handbook,* Ninth Edition, Irvine, CA, 1993.

[5] Underhill, M., "The Adiabatic Anti-Jitter Circuit," *Proc. 1999 Joint Meeting of the EFTF and the IEEE FCS*, April 13–16, 1999, Besancon, pp. 611–614.

Appendix
Units Related to PLL Design

Complicated units are not normally used in the analysis of PLL characteristics. The most common units are frequency [Hz], phase [rad], phase-time [s], and voltage [V]. However, which of these units are used is very important. The equations described in Chapter 5 are reconsidered in terms of these units.

A.1 Units in the Basic Equation

Let us consider the units of the coefficients used in the basic equation that is based on the phase difference between the input and output signals as per (5.6) of Section 5.1.3.

$$\begin{aligned} \frac{1}{2\pi}\frac{d\theta_d(t)}{dt} &= \Delta f_{free-i}(t) - K_{of} \bullet K_d \bullet \theta_d(t) \\ \frac{d\theta_d(t)}{dt} &= \Delta\omega_{free-i}(t) - K_o \bullet K_d \bullet \theta_d(t) \end{aligned} \tag{A.1}$$

The units on both sides of (A.1) are angular frequency [rad/s]. The units of the second term in the right must be [rad/s] as well. $\theta_d(t)$, which is

included in the second term on the right, is the phase difference between the input and output signals as shown in (A.2), and its units are [rad/s].

$$\theta_d(t) = \left[\theta_i(t) - \theta_o(t)\right] \tag{A.2}$$

This phase difference information is converted into voltage in the phase comparator. K_d is the total sensitivity of the phase comparator and represents its output voltage. The unit of sensitivity is [V/rad]. K_o is the sensitivity of the controlled oscillator. Since the output angular frequency is determined by the voltage of the phase comparator output, its unit of sensitivity is $[\frac{\text{rad/sec}}{\text{V}}]$. Therefore, the second term on the right of (A.1) has units as given by

$$\left[\frac{\text{rad}}{\text{sec} \times \text{V}}\right] \bullet \left[\frac{\text{V}}{\text{rad}}\right] \bullet [\text{rad}] = \left[\frac{\text{rad}}{\text{sec}}\right] \tag{A.3}$$

This unit system can be used in the analysis of multiplier-type phase comparators as well (which are not described in detail in this book) if a linear approximation is feasible (i.e., small phase differences).

A.2 Units in the Transfer Function

When a transfer function is calculated by using practical PLL parameters, the system of units is not as simple as suggested by (A.1). Consider the units of the first order PLL, which is

$$\left|H(j\omega)\right| = \sqrt{\frac{\left(K_o \bullet K_d\right)^2}{\omega^2 + \left(K_o \bullet K_d\right)^2}} \tag{A.4}$$

Here the angular frequency ω is in [rad/s] and the units of $K_o \bullet K_d$ are [1/s] as

$$\left[\frac{\text{rad}}{\text{sec} \times \text{V}}\right] \bullet \left[\frac{\text{V}}{\text{rad}}\right] = \left[\frac{1}{\text{sec}}\right] \tag{A.5}$$

Since (A.4) is not altered in the calculation process in Chapter 5, the original units of the coefficients ought to appear in the transfer function. Though the units of [1/s] (i.e., [Hz]) are often converted into the units of [rad/s], multiplied by 2π in practical calculations, the unit [rad] is physically dimensionless. When both units, [rad] and [1/s], appear in an equation, you should reconsider the original units. It is necessary to determine whether [1/s] should be converted into [rad/s], when assigning values to it.

If K_d has units of [V/rad], $K_o \bullet K_d$ has units of [1/s]. Note that $K_o \bullet K_d$ should not be multiplied by 2π, even if its unit is inconsistent with that of ω in (A.4).

A.3 Units in Digital PLLs

Let us consider the units of the digital PLL in which a phase difference is detected as a digital value, and a controlled oscillator is steered by this. It is assumed that the phase comparator is a digital type that can detect the phase-time difference over time in [s] with no quantization effect and with infinite sampling frequency (see Figure A.1). The controlled oscillator characteristic is assumed to be that shown in Figure A.2 with no digital control effect. Since the phase comparator measures phase-time difference between the input and output signals, its own output is

$$D(t) = K_{td}\left[T_i(t) - T_o(t)\right] \tag{A.6}$$

K_{td} has units of [1/s], because the phase-time difference generates the digital output $D(t)$. The output phase-time characteristic of the controlled oscillator is found to be

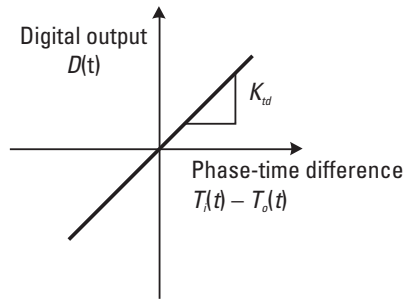


Figure A.1 Digital phase comparator characteristic.

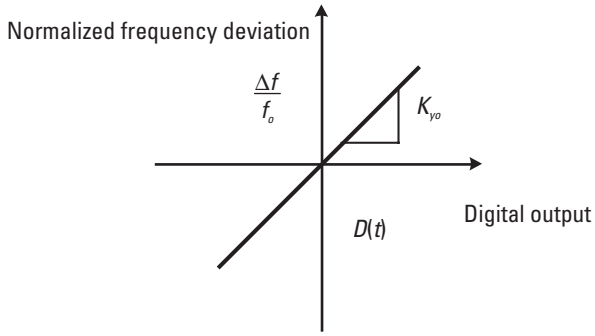


Figure A.2 Controlled oscillator characteristic.

$$\frac{dT_o(t)}{dt} = K_{yo} \bullet D(t) \tag{A.7}$$

This equation is induced from the relationship in which the time differential of phase-time corresponds to normalized frequency deviation. Normalized frequency deviation is equal to frequency difference normalized by nominal frequency, which is naturally

$$\frac{f_{out} - f_o}{f_o} = \frac{\Delta f}{f_o} \tag{A.8}$$

K_{yo} is dimensionless, [1/1], since K_{yo} is the ratio of normalized frequency deviation to a digital value. The transfer function whose coefficients have a unit of phase-time can be obtained as shown in (A.9) by using the Laplace transform relationships of (A.6) and (A.7).

$$\left|H(j\omega)\right| = \sqrt{\frac{\left(K_{yo} \bullet K_{td}\right)^2}{\omega^2 + \left(K_{yo} \bullet K_{td}\right)^2}} \tag{A.9}$$

$K_{yo} \bullet K_{td}$ has units of [1/1][1/s], which consequently becomes [1/s]. There are two units, [rad/s] in ω and [1/s] in $K_{yo} \bullet K_{td}$, on the right-hand side of (A.9) as well as (A.4). In this case, $K_{yo} \bullet K_{td}$ has not been converted into another unit system during the calculation process. It is not necessary to convert $K_{yo} \bullet K_{td}$ into units of [rad/s].

About the Authors

Masami Kihara received his B.S. and Ph.D. in electrical engineering from Nihon University, Tokyo, Japan, in 1976 and 1985, respectively. He joined Nippon Telegraph and Telephone Corporation (NTT) in 1976. He was engaged in the development of the cesium atomic beam standard with magnetic state selection, the reference frequency standards system for evolving digital networks, and other highly stable oscillators for general use. He has also conducted research on synchronization technologies in digital networks and has developed synchronization systems. His present research interests are time synchronization and jitter/wander performance in networks. He is currently a senior research engineer and supervisor of NTT Microsystem Integration Laboratories, Atsugi, Japan. He was also a visiting professor at Helsinki University of Technology, Finland, from 2001 to 2002. Dr. Kihara is a member of the Scientific Committee of the European Frequency and Time Forum, Switzerland. He is a member of the Institute of Electrical and Electronic Engineers (IEEE); the Institute of Electronics, Information, and Communication Engineers (IEICE); the Institute of Electrical Engineers (IEE) of Japan; and the Geodetic Society of Japan.

Sadayasu Ono received his B.S., M.S., and Ph.D. in electrical engineering from Keio University, Yokohama, Japan, in 1971, 1973, and 1976, respectively. In 1976, he joined the Yokosuka Electrical Communication Laboratories of NTT. He was a member of the design team that developed the NTT digital filter CAD system and the first floating-point digital signal processing

(DSP) large-scale intergration (LSI), and he became a project leader of the first C compiler development group for the DSP. During the last several years, he has been responsible for research and development on the ultra-high-speed parallel signal processor for SHD image signals. Currently, he is a research fellow at the Ono Research Laboratory at NTT Network Innovation Laboratories, Yokosuka, Japan.

Dr. Ono is a coauthor of two books, *Software on Signal Processors* (Corona Co., 1989) and *Digital Signal Processors* (Ohm Co., 1990).

He is a member of the IEEE, IEICE, and IPSJ, and is the associate editor of *IEICE Transactions*. He is also the vice chairman of the Digital Signal Processing Study Group of the IEICE.

Pekka Eskelinen is a professor in the radio laboratory and the head of the Institute of Digital Communications at the Helsinki University of Technology, Finland. He recently resigned his position as the head of the Electronics and Information Technology Department at the Finnish Defense Forces' Technical Research Center, prior to which he was the head of the Microwave Laboratory and a professor of electronics at the Lappeenranta University of Technology, Finland.

Professor Eskelinen received his M.Sc. (1979) and Dr.Sc. (EE) (1992) from the Helsinki University of Technology. He worked for 5 years at the Technical Research Center of Finland, designing subsystems for satellite ground stations and military radar equipment. In 1984 Professor Eskelinen joined the National Board of Aviation, where he was chief of the avionics section. Since 1988, he has been involved in accident investigation and is a technical expert in the Planning Commission for Investigation of Major Accidents.

Professor Eskelinen has also been a member of the board of governors in the IEEE Aerospace and Electronic Systems (AES) Society, and he continues as an associate editor of the *IEE AES Systems* magazine and as a member of the Scientific Committee of the European Frequency and Time Forum, Switzerland. He is currently the chairman of the Microwave Theory and Techniques/Antennas and Propagation/Electronic Design (MTT/AP/ED) chapter of the IEEE Finland section. Professor Eskelinen has been nominated to join the International Advisory Committee of the International Joining of Materials (JOM) Institute and has gained a docent's status of radio frequency engineering both at the Lappeenranta University of Technology and at the Tampere University of Technology, Finland. He is also a member of the communications section in the Scientific Advisory Council for Defense and has been a representative of Finland in the Western European Armament Group's Panel 6.

Index