TRANSCEIVERS

COMMUNICATION SYSTEMS

The ability of electromagnetic radiation to provide almost instantaneous communication without any interconnecting wires has been a major factor in the explosive growth of mobile communications, especially cellular and personal communications during the latter half of the twentieth century. The vacuum tube made the radio practical and affordable during the earlier half of the twentieth century. The invention of the transistor and highly complex yet cheap integrated circuits have allowed the development of ever complex digital communication systems that operate quite close to theoretical limits on channel capacity, spectrum efficiency, and so on.

The word *transceiver* is actually a combination of two words, *transmitter* and *receiver*. Figure 1 shows the block diagram of a general radio transceiver. Below each system block is a list of some of the popular techniques used. A modern digital transceiver uses most if not all of the system blocks in Fig. 1. Based on the direction of information transfer, radio systems are of three general types: simplex, half-duplex, and full-duplex systems. A *simplex* system transmits information in only one direction from a transmitter to a receiver. Examples of simplex systems are commercial audio and television broadcast systems organized in a *star* configuration. A *half-duplex* system is one where transmission is bidirectional but only one transmitter at a time can transmit. A *full-duplex* or duplex system is one where bidirectional communication can occur at any time. Although all communication systems are either some kind of simplex or duplex system, a variety of

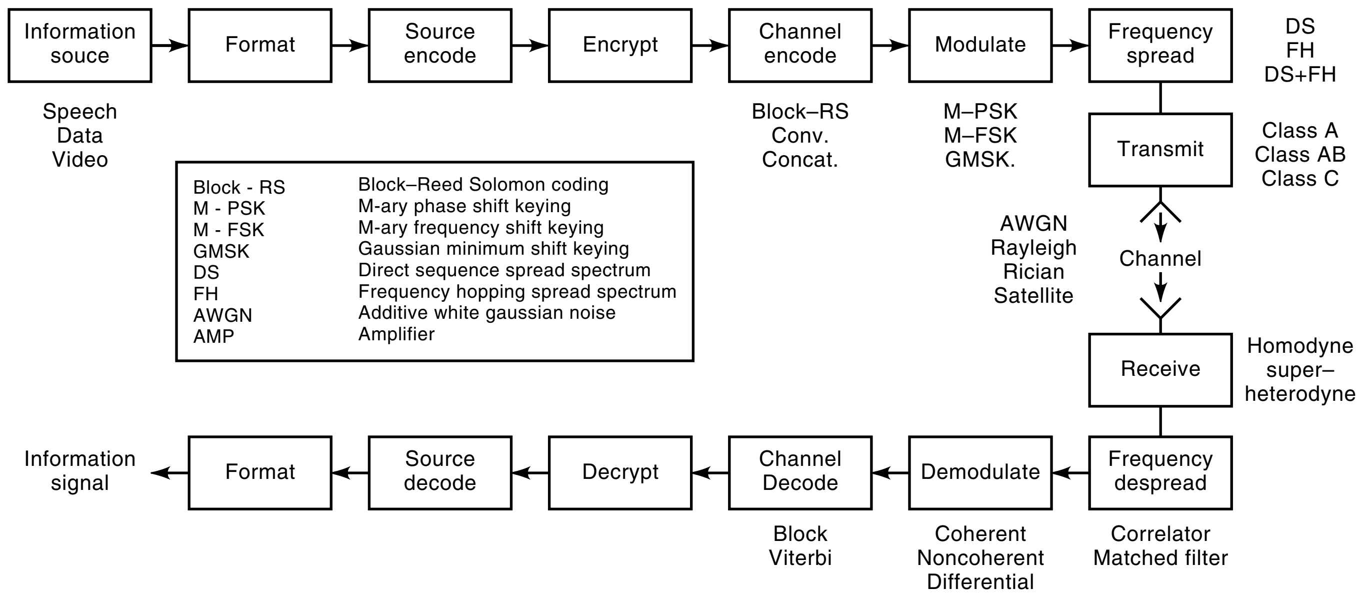

Figure 1. General block diagram of a modern transceiver. Abbreviations are listed at the end of the article.

communications architectures exist dependent on the end application.

The next section points out the important characteristics of radio transceivers. The section entitled "Radio Architectures" is devoted to the evolution of transceiver architecture. The characteristics of various different architectures are discussed in detail. The section following describes the implementation of important transceiver components with special emphasis on digital processing. The last section concludes by looking at the future.

TRANSCEIVER CHARACTERISTICS

Radio Transmitter Characteristics

This section details the important characteristics and performance issues for a transmitter.

Out-of-Band Emissions. There are stringent requirements as part of the standards for out-of-band power emissions. For example, the interim standard IS-54 for digital cellular radio specifies that power emission in adjacent and alternate channels must be 26 dB and 45 dB below the mean output power, respectively. Spectral shaping to maintain out-of-band emissions below those required by the standards is usually achieved through a combination of baseband pulse shaping, IF/RF filtering, and proper operation of the RF power amplifier.

Output Power, Distortion, and Efficiency of the Power Amplifier. In the early days of radio, power was amplified by vacuum tubes. With the invention of the transistor, solid state circuits have replaced vacuum tubes. The power amplifier should provide adequate output power with the least distortion and maximum possible efficiency (1). The gain, efficiency, distortion, and power of the transistors typically used in solid-state power amplifiers depend on the choice of the bias point and the drive mechanism through the familiar designation of the operating class (Class A, B, AB, C, or E). The distortion introduced by the amplifier is specified by its AM–AM and AM–PM transfer characteristics.

The two most important parameters of the power amplifier, namely, efficiency and distortion, are incompatible with one another. Linear power amplifiers provide the least distortion but have quite large quiescent currents, resulting in poor efficiency. Nonlinear amplifiers, like Class C amplifiers, are highly efficient but cannot be used for linear modulation because of the distortion introduced by the amplifier. The type of power amplifier used depends on the requirements of the modulation used. Constant envelope modulation like frequency modulation (FM) can be amplified with nonlinear Class C amplifiers. However, in recent years, increasing demand for RF spectrum usage is forcing the use of spectrally efficient linear modulation techniques. These signals have a fluctuating envelope, and nonlinear amplification results in spectral spreading and intermodulation products. Therefore highly linear power amplifiers are required. One way to achieve linear amplification is to back off the amplifier from saturation and operate in the linear region of its transfer function. However, such amplifiers have poor dc-to-RF conversion efficiency. A major challenge in designing a high power amplifier is to maintain linearity without compromising power efficiency (2).

Radio Receiver Characteristics

This section details the important characteristics and performance issues for a receiver.

Antenna and Input Characteristics. The main function of a radio receiver is to pick up electromagnetic radio energy at the antenna and transfer it efficiently to the processing section of the receiver. The important characteristics of an antenna are its efficiency, impedance, 3 dB beam width, bandwidth, and gain. Other characteristics are null-fill, upper sidelobe suppression, and performance versus specification.

All antenna characteristics are functions of the physical antenna dimensions relative to the operating wavelength. Another important component is the coupler between the antenna and the input circuit of the receiver which typically is a filter or an amplifier. Maximum energy is transferred if the impedance of the input circuit matches that of the antenna throughout the band of interest. Some of the antenna matching problems in a radio receiver are as follows:

1. The problem of matching the antennas at certain frequencies may be limited by component availability.
2. The impedance of antennas used in mobile applications or in locations where the environment changes with time due to foliage or traffic.

The problem of antenna matching is often solved by the system designer taking into account and compensating for a range of mismatch losses that might occur in practice. Other input characteristics that need to be taken into account are as follows:

1. The input RF circuits may be balanced, unbalanced, or both.
2. Protection from high voltage discharges due to lightning, etc.
3. Ability to handle high-power cochannel and adjacent channel transmissions

Gain and Sensitivity. Radio receivers typically process signals with a wide range of powers. The extent to which the signals can be received and processed usefully depends on the noise levels received at the antenna and those generated by the circuits within the receiver itself. The receiver is also required to produce a certain level of output power suitable for the application. Receivers are designed so that the gain is distributed among the various stages as required. Modern receivers are usually not gain-limited, and the weakest signal that can be processed is usually noise-limited. This signal level is known as the *sensitivity* of the receiver. A measure of sensitivity is the *minimum detectable signal* (MDS), which is the power of a sinusoidal signal that just equals the noise power at the intermediate frequency (IF) output of the receiver. MDS (1) can be expressed in dB as

$$\text{MDS} = KTB_nF$$

where K is Boltzmann's constant (1.38×10^{-20} mW/°K), T is the reference temperature (typically 290°K), B_n is the noise bandwidth of the receiver, and F is the noise figure.

Noise Figure. The noise figure compares the total receiver noise with the noise that would be present if the receiver generated no noise. This ratio is called the *noise factor F*, and when expressed in dB, the noise figure (NF). It is thus a measure of the amount of noise introduced by the circuits within the receiver itself.

$$F = \frac{(S/N)_{\text{input}}}{(S/N)_{\text{output}}}$$

where (S/N) is the signal-to-noise ratio.

Selectivity. Selectivity is the ability of a receiver to separate a signal at one frequency from signals at other frequencies. Selectivity is defined as the bandwidth for which a signal x dB stronger than the minimum acceptable signal at a nominal frequency is reduced to the level of that minimum acceptable signal. Two important characteristics are required in establishing the selectivity of a receiver. One is that the selective components of the receiver must be sufficiently sharp to suppress unwanted interference from adjacent channel transmissions and spurious responses. The other is that the components must be sufficiently broad to pass the highest frequency of interest with acceptable gain and phase distortion.

Dynamic Range. Dynamic range is used to indicate the ratio between the strongest and weakest signals that a receiver can handle with acceptable noise or distortion. The weakest signal commonly considered is the minimum detectable signal. This definition is of limited value especially when the desired signal is surrounded by other signals with varying signal power. The selectivity of a receiver provides protection against many of the unwanted signals. The strong unwanted signals, however, can still cause degradation because of nonlinearities in the receiver chain. Therefore it is important to consider the definition of the strongest signal component when determining the dynamic range of the receiver.

Characterization of Spurious Outputs. A modern receiver typically has a synthesizer and possibly several local oscillators, especially if superheterodyne architecture is used. It is possible for these frequencies to interact and produce spurious outputs without any inputs present. The following are other sources of spurious signals:

1. Parasitic oscillations in amplifiers because of parasitic feedback
2. Intermediate frequency subharmonics
3. Power supply harmonics

Frequency/Clock Generator Characteristics

Accuracy and Stability. Modern transceivers have a frequency synthesizer to which all other local oscillators are slaves. Earlier radios had free running oscillators that have largely been replaced by digital synthesizers because of the superior frequency accuracy, stability, flexibility, and cost performance of digital circuitry. Once the synthesizer has been set to operate at a specified frequency, its frequency must remain unchanged for a period sufficient for nominal operation despite temperature and environmental changes. Modern transceivers use temperature-compensated crystal oscillators as clocks for their digital circuitry. These oscillators typically are accurate to about 1 part per million. Higher accuracies are provided by oven-stabilized crystal oscillators and rubidium oscillators when sufficient power is available. In certain applications, such as mobile handsets where cost and power are at a premium, less expensive clocks with accuracies of approximately 3 to 10 parts per million are used.

Settling Time. Modern receivers typically span large frequency ranges and might be required to retune to a different frequency of operation. Because the frequency synthesizer used is typically based on a phase-locked loop, the loop goes out of lock for a short period whenever the receiver retunes.

The settling time of the loop is important as any loss of lock results in degraded receiver performance.

Digital Receiver Characteristics. In addition to the characteristics previously mentioned, these are other important characteristics useful for systems using digital modulation.

Eye Diagram. The eye diagram (3), the traditional way of displaying digital data, is obtained by displaying the received demodulated digital data signal in successive symbol intervals on top of each other. The eye pattern provides the following wealth of information:

1. The width of the eye opening defines the time interval over which the received signal can be sampled without error from intersymbol interference.
2. The sensitivity of the system to timing error is determined by the rate of closure of the eye, as the sampling time instant is varied.
3. The height of the eye opening at any specified time defines the margin over channel noise.

Bit Error Rate (BER). The BER (1) is the primary measure of the quality of a digital communication system. The BER is defined as

$$\mathrm{BER} = \frac{N_{\mathrm{E}}}{N}$$

where N_{E} is the number of bit errors and N is the total number of bits transmitted.

RADIO ARCHITECTURES

Evolution

Radio architectures have remained relatively unchanged since the invention of the homodyne and superheterodyne receivers in the early part of the twentieth century. With the advent of integrated processors during the 70s and 80s, there was a migration from analog to digital processing in almost every aspect of radio systems engineering. The only radio system block that survived this migration was the RF front end which by its function has to be analog. However, the basic radio architecture has remained the same. The word *digital* in digital radio has a double meaning. First, it refers to the fact that information is carried in digital form and secondly that the radio uses digital processing to recover the transmitted signal after it has been down-converted by an analog front end. The following are the advantages of using digital processing (1,4):

1. The repeatability and temperature stability of digital processing are substantially better than analog processing.
2. Certain functions that cannot be or are difficult to implement in analog hardware, such as sharp roll-off linear phase filters, can easily be implemented with digital processing.
3. Once engineered, digitally implemented system functions do not require the tuning or tweaking typically required in analog systems.

The software radio (5) can be thought of as the next logical evolution of the digital radio, where software control of radio functions is pushed as close as possible to the antenna in the conventional digital radio architecture. Here, the entire RF band of interest is digitized right at the operating RF band by high-speed analog-to-digital converters (ADCs). The rest of the radio functions, such as down-conversion, equalization, demodulation, and decoding, would be carried out by reprogrammable logic, typically digital signal processors (DSPs).

Software radio architecture has the following advantages over conventional hardwired digital radio architecture:

1. A highly flexible and reconfigurable transceiver can be implemented.
2. The transceiver can be easily adapted to any particular environment by changing the modulation, filtering, demodulation, and so on.
3. Because of open architecture, future upgrades can easily be made without reengineering the entire radio.
4. Software radio architecture benefits readily from concentration, where multiple radio channels share the same RF front end, whereas analog systems need a separate RF front end for each channel.

Applications which already use or are likely to use digital/software radio architectures include cellular and personal communications systems, satellite communications, digital television, digital audio broadcasting, navigation and position location systems, and test equipment.

Transmitter Architectures

The earliest analog radio transmitter architecture was the direct conversion transmitter. Figure 2 shows the general block diagram of an analog direct conversion architecture. Although this architecture is simple, it has the following disadvantages (6):

1. The analog implementation of precise modulators at the operating RF is difficult.
2. When used as an analog quadrature modulator, gain and phase imbalances between the mixers require compensation because unwanted sidebands are generated.
3. The filtering required to reduce out-of-band emissions to conform to government-mandated spectral masks must be carried out completely at the RF band of interest. Designing high roll-off RF filters that introduce minimal amplitude and phase distortion across the frequency band of interest is difficult.

The superheterodyne architecture shown in Fig. 3 was intended to overcome some of the disadvantages of the direct conversion architecture. Here, modulation is carried out at a low intermediate frequency (IF). Then the desired signal band is filtered to conform to the desired spectral mask, and the filtered signal is up-converted to the desired RF band. The main disadvantage of the superheterodyne architecture is that the one or more IF stages used increase power consumption, space, and cost.

The advent of digital modulation and digital integrated circuits has resulted in the ever increasing use of digital pro-

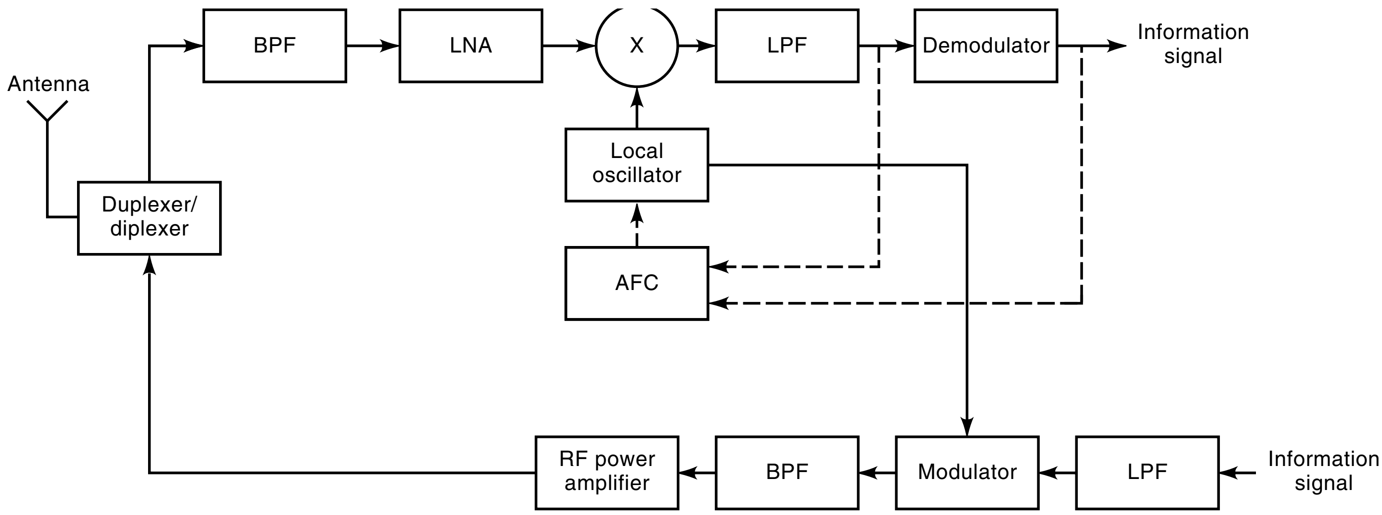

Figure 2. General block diagram of an analog homodyne transceiver.

cessing in the transmitter chain. Digital processing can alleviate most of the problems associated with analog direct conversion transmitters, and as a result there is a resurgence of interest, especially for low-power mobile handsets where space and power are at a premium. Following are the advantages of digital architecture (7):

1. Design and implementation are flexible.
2. Digital implementation overcomes the problems of gain and phase, dc offsets, and performance drifts in analog implementations.
3. Multichannel digital IFs especially at base stations eliminate the multiple analog IF chains required in analog architecture. With digital up-conversion, all digital IF signals can be combined in a digital summer and then transformed into a single analog signal by a digital-to-analog converter (DAC).
4. Manufacturing is reliable.

Some of these combined analog-digital architectures are shown in Fig. 4.

The advent of software radio architecture will result in almost complete replacement of the RF system by programmable digital processing. Figure 5 shows the block diagram of an ideal software radio. In this ideal architecture, except for the final RF power amplifier and filter, the analog up-conversion chain has been replaced by digital IF up-conversion. Limitations on the maximum sampling rate by currently available

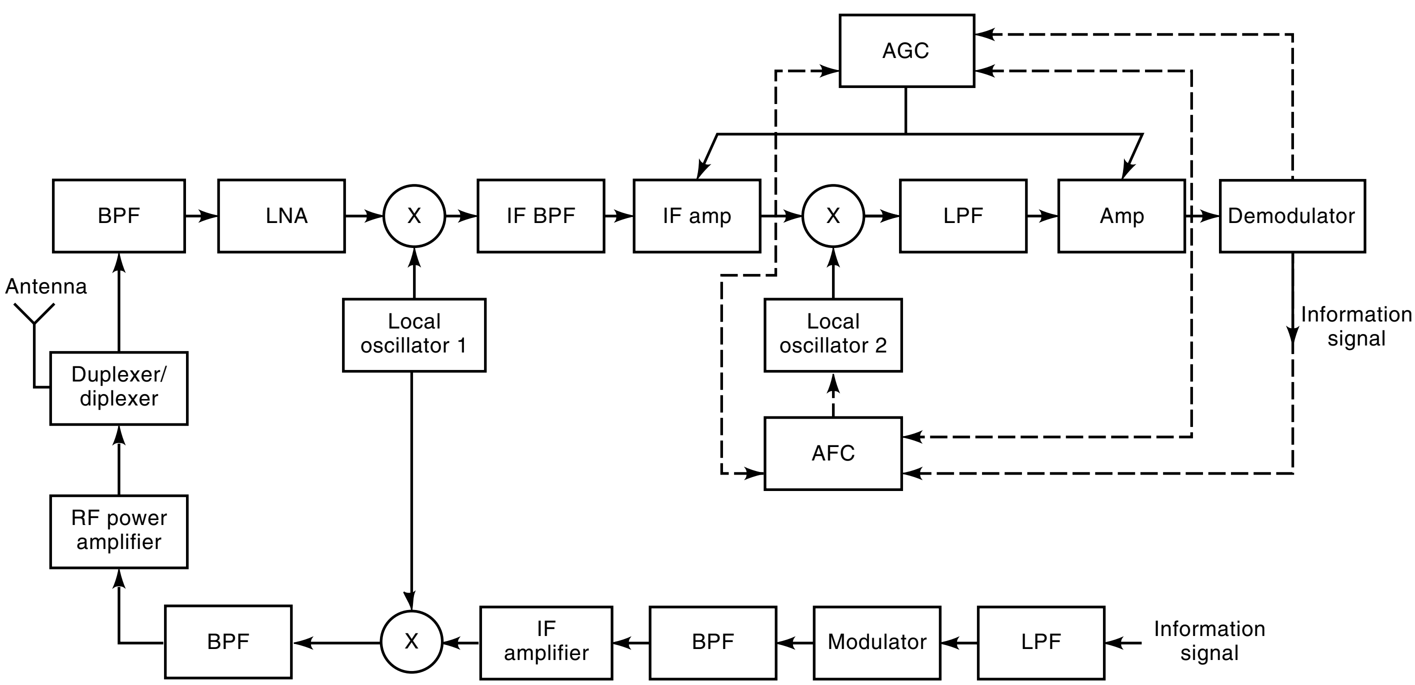

Figure 3. General block diagram of an analog superheterodyne transceiver.

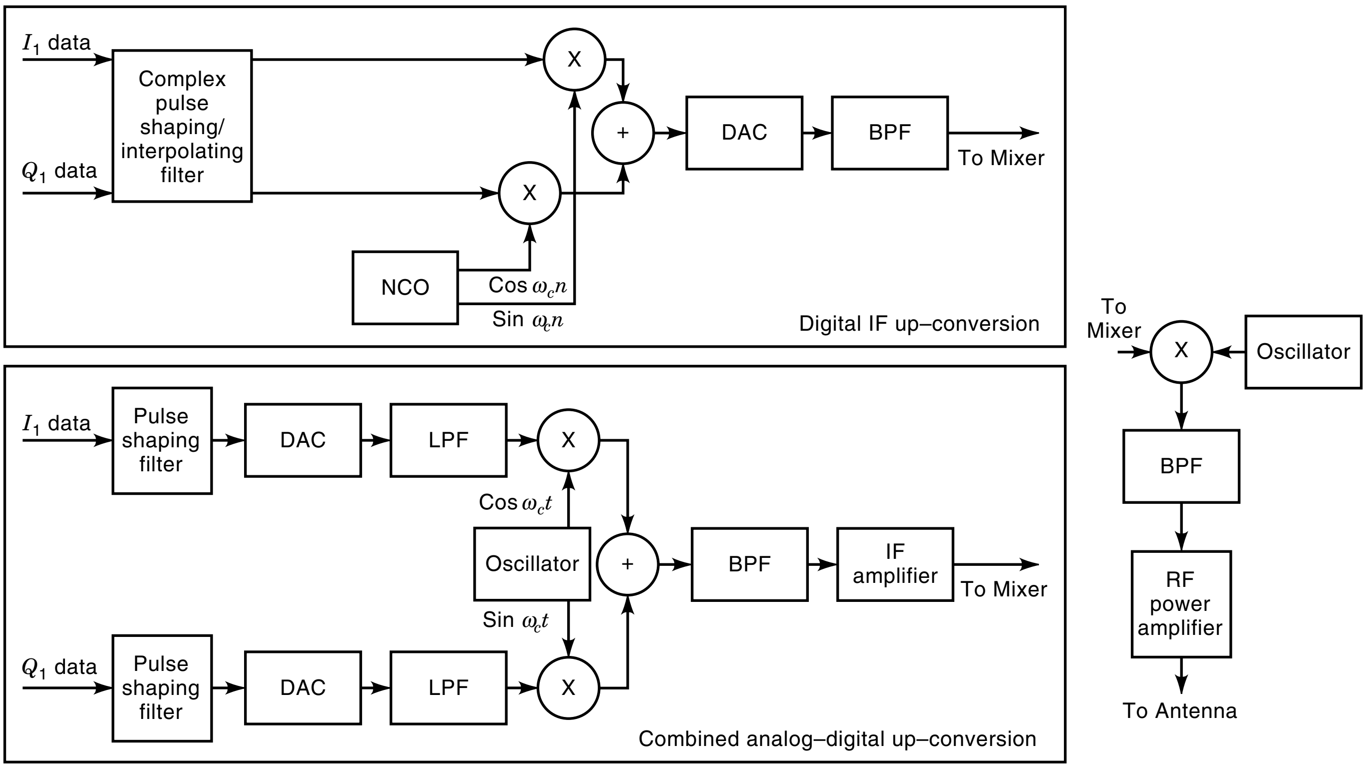

Figure 4. General block diagram of combined analog–digital up-conversion architectures.

technology, however, permits replacing only the first analog IF strip.

Receiver Architectures

The earliest analog receiver architecture was the homodyne or the direct conversion architecture, as seen in Fig. 2. The receiver consists of a band-pass filter (BPF) following the antenna for noise and interference rejection followed by a low-noise amplifier (LNA). Then the signal is down-converted to baseband by a pair of analog mixers. The in-phase and quadrature components are low-pass filtered to remove the mixer products and are demodulated. Following are the advantages of direct conversion architecture:

1. Reduced hardware complexity as there are no IF stages
2. No image frequencies

Following are the disadvantages of the direct conversion receiver, *especially* with an analog implementation:

1. Amplitude and phase mismatches between the mixers distort the signal.
2. Sharp roll-off analog low-pass filters (LPF) also distort the desired signal.
3. Carrier leakage, $1/f$ noise in the mixers, and bias in the filters all contribute to an unpredictable-time varying dc offset in the recovered signal.
4. All signal amplification has to be done at the carrier frequency. Building high-gain RF amplifiers at such high frequencies is difficult and expensive.
5. Carrier recovery for coherent reception has to be carried out at the carrier frequency. Precise control of high-frequency oscillators is difficult.

The superheterodyne receiver shown in Fig. 3 was developed to alleviate some of the disadvantages of the direct conversion receiver. In this architecture, the RF signal is down-converted to an intermediate frequency before being down-converted to baseband. This is known as the single IF stage superhetero-

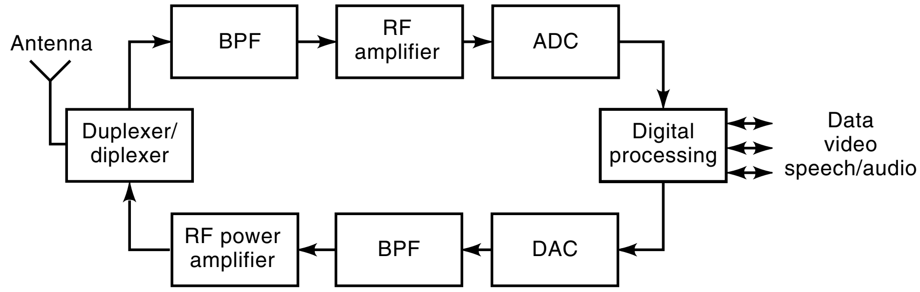

Figure 5. General block diagram of an ideal software radio architecture.

dyne receiver. Various versions of this general architecture with multiple IF stages have been developed. The most common version is the dual IF stage architecture. Following are the advantages of superheterodyne architecture compared to the direct conversion receiver:

1. Most of the signal amplification is done at relatively lower IFs, where it is easier to build high-gain amplifiers.
2. Automatic frequency control is usually carried out at the lower IFs and hence is easier to implement.

With the advent of digital modulation, high-speed ADCs, and digital integrated circuits, digital processing started to replace segments of the analog receiver architecture especially in demodulation and baseband processing. This led to the development of digital radio architecture. Both direct conversion and superheterodyne RF front end architectures have been used in digital radios. The direct conversion digital radio requires two ADCs to digitize the in-phase and quadrature components of the down-converted signal. Superheterodyne digital radios have used both low-pass and band-pass digitization. The advantage of band-pass digitization is that it can replace the last analog down-conversion stage. In addition, because the final down-conversion to center the chosen spectral image around dc is done digitally, all problems associated with quadrature analog down-conversion disappear. Some of the digital receiver architectures are shown in Fig. 6.

Although the digital radio provides superior performance, its architecture is still based on the direct conversion or the superheterodyne architectures. Recently, a more fundamental change in receiver architecture occurred with the advent of software radio architecture, made possible by technological advances in ADC technology, computing technology, and software engineering. The software radio architecture looks similar to that of the digital radio with one crucial difference. In software radio architecture, programmable digital processing is pushed as close to the antenna as technology permits. The block diagram of an ideal software radio architecture is shown in Fig. 5. In this architecture, the only analog RF components are the preselection band-pass RF filter and the low-noise RF amplifier. Then the RF signal is directly digitized using band-pass subsampling, and the rest of the receiver functions are carried out in embedded software modules running on high-speed DSPs. This architecture is still a few years away from commercial implementation because high-speed ADCs that operate with sufficient resolution at the desired RF band are unavailable. Current software radio architectures use at least one IF stage.

TRANSCEIVER IMPLEMENTATION

Transmitter

The design and implementation of transmitters involves the following:

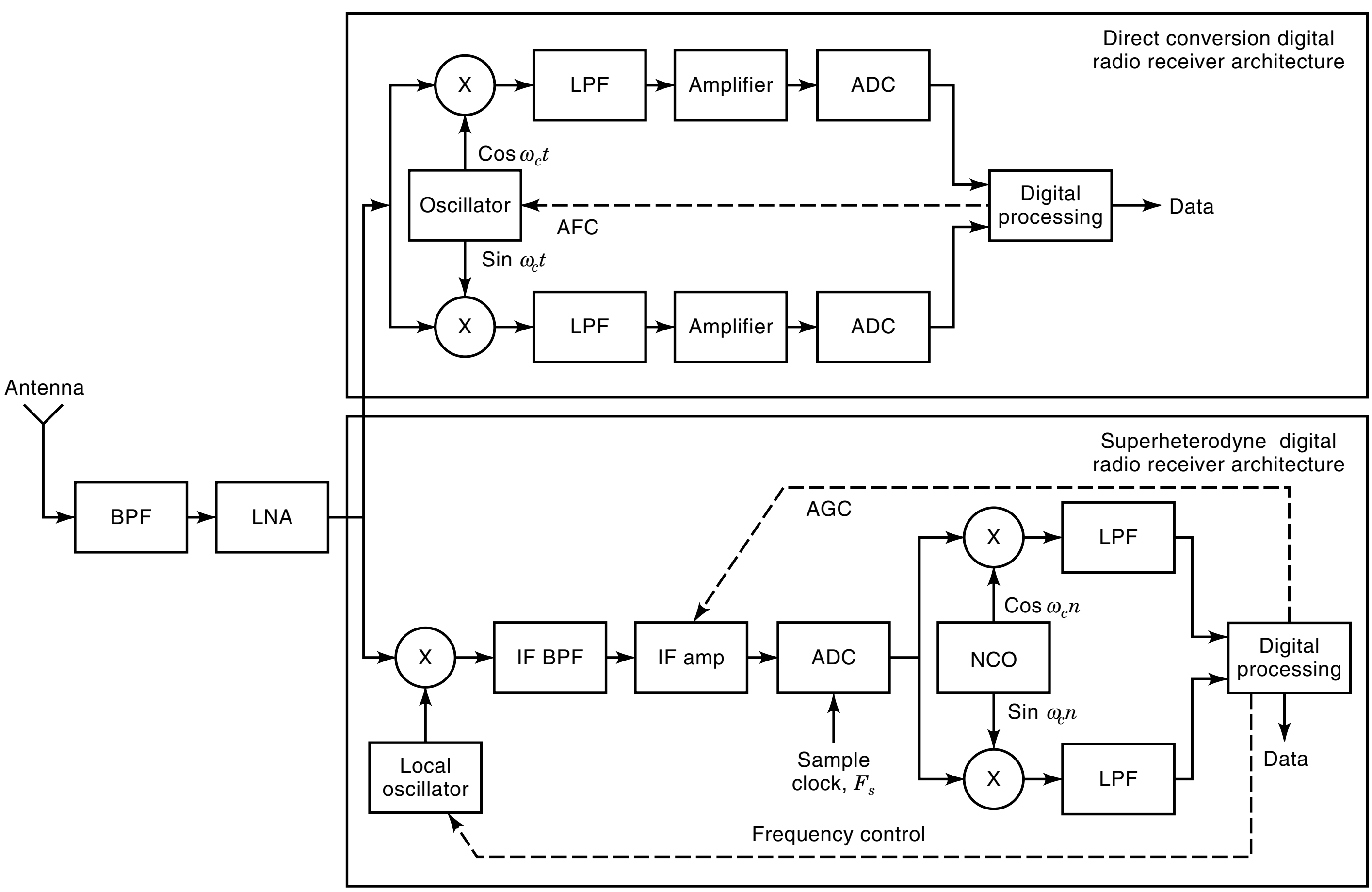

Figure 6. Some digital receiver architectures.

1. Filtering
2. Modulation and frequency up-conversion
3. Power amplification

Filtering. In analog radios, analog filters are the only way to achieve frequency selectivity and to limit out-of-band emissions. In digital radios, digital filters in addition to analog filters are used to shape the spectrum of the transmitted signal. Analog filters are discussed in detail later. Some of the popular digital filters used for pulse shaping to achieve a compact modulated spectrum are the raised cosine and Gaussian filters (8). The advantages of using digital filters are discussed in a later section.

Modulation. Modulation (8) is the process of encoding information from a message source onto a format suitable for transmission. The process involves translating the baseband message signal, called the modulating signal, onto a bandpass signal at a much higher frequency, called the carrier. This process is called modulation. Modulation is usually carried out by varying the amplitude, frequency, phase, or any combinations of these three parameters of a carrier signal. Based on whether the message information signal is analog or digital, modulation is classified as analog or digital modulation.

Analog Modulation. In analog modulation, the modulating signal is analog and can assume an infinite number of amplitude values. Analog modulation is broadly classified into two categories, amplitude modulation and angle modulation.

In amplitude modulation (AM), the amplitude of a high-frequency carrier signal is varied in accordance with the instantaneous amplitude of the modulating information signal (3,8). The amplitude-modulated signal is expressed as

$$S(t) = A_C[1 + m(t)]\cos 2\pi f_C t$$

where A_C is the amplitude of the carrier, f_C its frequency, $m(t)$ is the modulating information signal, and $S(t)$ is the modulated signal. The spectrum of an AM signal contains a component at the carrier frequency and two sidebands that replicate the original information spectrum. An AM signal is generated by a nonlinear device, such as a diode or transistor. Many variations of amplitude modulation exist based on what percentage of the sidebands is transmitted. Some of these variations are single-sideband AM (SSB-AM), pilot-tone-sideband AM, and vestigial-sideband AM (VSB-AM).

Angle modulation (3,8) varies the angle of the carrier signal according to the amplitude of the modulating signal. There are two important classes of angle modulation, frequency modulation (FM) and phase modulation (PM).

In FM, the instantaneous frequency of the carrier is varied with the information signal $m(t)$, as shown by the following equation:

$$S(t) = A_C \cos\left(2\pi f_C t + 2\pi k_f \int_{-\infty}^{t} m(l)\,dl\right)$$

where k_f is the frequency deviation constant measured in units of hertz per volt. There are two basic methods for generating an FM signal, the *direct method* and the *indirect method*. In the direct method, voltage-controlled oscillators vary the frequency of the carrier signal directly in accordance with the amplitude of the information signal. Such oscillators commonly use devices, such as varactor diodes, whose reactance can be varied in accordance with the modulating signal's voltage level. The indirect method is based on approximating a narrowband FM signal as the sum of a carrier signal and a single sideband signal where the sideband is 90° out of phase with the carrier.

PM is a form of angle modulation where the phase of the carrier is varied according to the information signal $m(t)$. A PM signal can be generated by first differentiating the information signal $m(t)$ and then outputting it to a frequency modulator.

Digital Modulation. Modern communication systems use digital modulation. In digital modulation, the modulating information signal is represented as a time sequence of symbols in which each symbol has m finite states. Each symbol represents n bits of information, where $n = \log_2 m$ bits/symbol. Digital modulation offers many advantages over analog modulation, including greater noise immunity, robustness to channel impairments, and easier multiplexing of various forms of information, such as voice, data, and video. Furthermore, digital transmissions use error-correcting codes and support complex signal conditioning and processing techniques, such as source coding/compression, encryption, and equalization to improve the performance of the communications. Advancements in very large scale integration and digital signal processing technology have made it possible to implement digital modulators and demodulators easily. The use of embedded software to do most of the signal processing allows alterations without having to replace the hardware.

Some of the widely used digital modulation techniques are m-ary amplitude-shift keying (ASK), m-ary frequency-shift keying (FSK), m-ary phase shift keying (PSK), combined amplitude and phase modulation, combined coding and modulation, and multicarrier modulation.

In ASK, the amplitude of the carrier signal is varied in accordance with the instantaneous amplitude of the modulating information signal, which can take one of several discrete amplitudes. In FSK, the frequency of the carrier signal is varied in accordance with the instantaneous discrete amplitude of the modulating information signal. In PSK, the phase of the carrier signal is varied in accordance with the instantaneous discrete amplitude of the modulating information signal. In a general sense, ASK, FSK, and PSK can be thought of as special cases of AM, FM, and PM, respectively, where the modulating signal is discrete in amplitude. In combined amplitude and phase modulation, both the amplitude and phase of the carrier are varied according to the amplitude of the modulating signal. Ungerboeck (9) realized that error-correction coding and modulation can be combined at a fundamental level to realize performance gains greater than with the conventional method of coding and modulating separately. Combined coding and modulating is more commonly known as *trellis-coded modulation.* The basic concept of multicarrier modulation is dividing a given RF bandwidth into many narrowband subchannels which are transmitted.

Frequency Up-Conversion. Frequency conversion is usually achieved by devices called *mixers.* A mixer is a component that acts as a frequency converter by mixing two input signals together to produce a desired signal. A mixer can be implemented by a variety of semiconductor devices, such as diodes (typically Schottky diodes), bipolar-junction transistors, and field-effect transistors. Mixers are generally of two types, single-balanced and double-balanced. The single-balanced mixer improves port-to-port isolation and reduces the magnitude of some of the spurious signals. It consists of two single-ended mixers interconnected in a configuration that greatly reduces some spurious products. A double-balanced mixer further improves port-to-port isolation and suppresses spurious intermodulation products even further.

Amplifiers. The triode vacuum tube amplifier was the mainstay for many communication systems before the transistor was invented. Because of transit time limitations, triodes were limited to VHF and UHF. For operation at higher microwave frequencies, the magnetron was invented in the early 1940s. The invention of the transistor was a boon for developing low cost, reliable, hand-held, low-power mobile communication systems. Since then, solid-state amplifiers have replaced vacuum tube amplifiers in almost all communication systems. Systems requiring extremely high power at microwave frequencies, such as deep space and radar systems, continue to use tube amplifiers.

All commercial communication systems use solid-state transistor amplifiers. Solid-state amplifiers come in four main types: discrete, hybrid, integrated (ICs), and application-specific IC (ASIC). A discrete amplifier is one built with discrete transistors and passive components. Hybrid modules, also known as microwave integrated circuits (MICs), have a substrate and discrete devices, including RF matching and bias circuitry. RF ICs, also known as microwave monolithic ICs (MMICs), have all bias and RF matching circuitry on the same substrate, whereas MICs use different materials to achieve optimum matching. The advantage of using a single substrate is that components can be closely matched in value. The difference between an IC and an ASIC is that an IC typically integrates only a few transistors together with transmission line filters and inductors, whereas ASICs, on the other hand, contain several hundred or thousands of transistors.

There are a number of amplifier configurations based on operating classes A, B, AB, C, D, E, or F. Silicon (Si) bipolar, heterojunction-bipolar, and field-effect transistors (FET) have been used most often to date in RF circuits. Gallium arsenide (GaAs) and complementary metal oxide semiconductor transistors have been gaining interest. GaAs amplifiers offer simple functionality with some biasing and matching components around a chain of transistors. One of the advantages of GaAs versus silicon is that GaAs is an insulator, whereas silicon is a conductor at RF frequencies. As a result, GaAs can integrate a number of RF components monolithically, facilitating closer components and better matching.

High power amplification at microwave frequencies, especially those used for satellite communications are usually handled by traveling-wave tube amplifiers (TWTAs). Other tube amplifiers are the magnetron, coupled-cavity TWT, continuous-wave TWT, helix TWT, klystron, and crossed-field amplifier (CFA).

Receiver

The main function of the receiver is to pick up the RF energy transmitted at its antenna and efficiently and effectively recover the original information signal transmitted. The design and implementation of receivers involves the following:

1. Antennas
2. Amplification
3. Filtering
4. Down-conversion
5. Automatic gain control
6. Demodulation and other associated signal processing functions

Antennas, amplification, analog filtering, and down-conversion are common to both analog and digital/software radios. The rest of the receiver functionality depends on its implementation, whether analog or digital. Because these implementations are quite different, they are discussed in different sections with special emphasis on digital implementation.

Antennas. The type of antenna used in a transceiver depends on its application. Hand-held or backpacked transceivers require integral antenna structures. Vehicular transceivers must use antennas of limited size and relatively short wavelengths. Antennas that operate at frequencies substantially below their first resonance are called small antennas. Large point-to-point systems use large antennas. Some of the most popular antennas used in commercial communication systems are the vertical whip antenna, the loop antenna, and the dipole array.

Whip Antenna. For operating frequencies below the quarter-wave resonance of the antenna, the whip input impedance appears as a small capacitance in series with a resistance. The radiation resistance R_R of a short vertical whip is given by (1)

$$R_R = 40\pi^2 \left(\frac{h}{\lambda}\right)^2$$

where h is the antenna height and λ is the operating wavelength. The whip is also used as a quarter-wavelength monopole antenna for applications, such as cellular and PCS handsets. The whip antenna has an omnidirectional antenna pattern in azimuth.

Loop Antennas. Loop antennas have been used in portable broadcast receivers and radio direction finders. When the dimensions of the loop are small compared to the wavelength, the loop is said to be small and its impedance is an inductance in series with a resistance. The radiation resistance R_R for a loop with N turns is given by

$$R_R = 320\pi^4 \left(\frac{AN}{\lambda^2}\right)^2$$

where A is the area of the loop. The loop antenna responds as the cosine of the angle between its face and the arrival direction of the electromagnetic wave. This results in a figure eight antenna pattern with the null for waves arriving perpendicularly to the loop face.

Amplification. The RF signal picked up by the antenna is very weak and has to be amplified before it can be processed. Typically the very first amplifier used is a specially designed, low-distortion, low-noise amplifier. The operating characteristics of the amplifier are important, as its noise performance dominates the noise figure of the receiver. Both bipolar-junction and FET amplifiers have been used. In recent years, the use of GaAs instead of Si has been gaining interest.

Analog Filters. The selectivity of an analog radio receiver is achieved solely by band-pass and low-pass analog filters. Even in digital and software radio architectures, the analog band-pass filter is an important component and is used both for preselection and antialiasing. Digital radios also use analog filters in their RF front end to implement some selectivity.

The most important characteristics of a filter are its amplitude and phase response. The various characteristics of a filter are interrelated because they are completely determined by the poles and zeros of the transfer function of the filter. Following are some of the common filter families:

1. Butterworth
2. Chebychev
3. Elliptic
4. Equiripple

Analog filters are implemented with a number of different resonators. Following are the available technologies:

1. Inductor–capacitor (*LC*) resonators
2. Mechanical resonators
3. Quartz crystal resonators

Another important filter implementation is the surface acoustic wave (SAW) filter, of interest because it can be implemented with integrated circuit techniques and can use finite impulse response designs, similar to those for digital filters.

Down-Conversion. Down-conversion is the process of shifting the received RF signal to baseband. Both direct conversion and superheterodyne architectures are used. The basic components of down-conversion are mixers, band-pass and low-pass filters, and oscillators.

Automatic Gain Control (AGC). The large dynamic range of signals that must be handled by radio receivers requires gain adjustment to prevent overload or intermodulation of the stages to adjust the demodulator input level for optimum operation. Gain control is generally distributed over a number of stages throughout the receiver architecture. AGC typically measures the signal level into the demodulator and tries to keep the level in the desired range by a feedback control loop. The control should be smooth and cause a generally logarithmic variation with the input variable.

Other Analog Radio Receiver Functions. The remaining functionality of the analog radio receiver is demodulation. Double-sideband AM signals are usually detected by an envelope detector. An envelope detector is any rectifier circuit that produces a component at the modulating frequency which is then recovered by a low-pass filter. The rectifier is generally implemented by diodes and by bipolar and field effect transistors. Other AM transmissions, such as SSB-AM and VSB-AM, are demodulated by a coherent demodulator. The coherent demodulator uses a mixer circuit with a local oscillator signal synchronized to the AM input carrier. Carrier synchronization is achieved through a carrier recovery circuit, such as the Costas loop. Common FM demodulators are the slope detector, quadrature detector, phase-locked loop (PLL), demodulator, and zero-crossing detector. The slope detector uses linear circuits to convert the frequency variations to envelope variations which can then be detected by an envelope detector. The quadrature detector consists of a network that shifts the phase of the FM signal by an amount proportional to its instantaneous frequency and uses a phase detector to detect the phase difference between the original FM signal and the signal at the output of the phase-shift network. The output of the phase detector is proportional to the instantaneous frequency of the FM signal. In this manner, a frequency-to-amplitude conversion is achieved and the FM signal is demodulated. Phase detectors are generally implemented by diode-based mixer circuits. Because a PM signal can be modeled as an FM signal where the modulating signal has first been differentiated, PM demodulation is achieved by passing the PM signal through an FM demodulator and integrating its output.

Other Digital/Software Radio Receiver Functions. The rest of the digital/software radio receiver can be split into two distinct segments: signal digitization and signal processing.

Signal Digitization. Signal digitization, implemented by ADCs, is a two-step process (10), signal sampling followed by quantization. The sampling process is critical in signal digitization. There are two types of sampling, uniform and nonuniform sampling. In uniform sampling, signal samples are taken at uniform intervals, whereas in nonuniform sampling the samples are nonuniformly spaced. The ADCs in communication systems use uniform sampling and so the rest of the discussion concentrates on uniform sampling ADCs. The sampling methods for uniform sampling are Nyquist sampling, oversampling, quadrature sampling, and band-pass sampling.

The general sampling theorem for a band-limited analog signal with no spectral components above f_M Hz requires that the sampling rate F_S satisfies

$$F_S \geq 2f_M$$

$F_S = 2f_M$ is known as Nyquist sampling, and at this rate the replicas of the spectrum of the original analog signal do not overlap. Two practical problems arise when implementing Nyquist sampling. The first is defining what a truly band-limited signal is, and the second is antialiasing filtering before the ADC. In general, a RF signal has components at all frequencies. It is desirable that the distortion of the desired signal be dominated by ADC nonlinearities, not by spectral overlap. This requires that signals higher in frequency than $F_S/2$ be lower in power than the largest spurious response of the ADC. Band-limiting is usually carried out by the analog antialiasing filter before the ADC. Unfortunately, practical analog filters cannot provide the kind of "brickwall" filter response required. Also, as the steepness of the filter roll-off increases, the phase response of the filter becomes more nonlinear, introducing more distortion.

Sampling the signal at a rate higher than the Nyquist rate is called oversampling. The benefit of oversampling is that the spectral replicas of the original analog signal in the sampled signal spectrum become increasingly separated as the sampling rate is increased beyond the Nyquist rate. Hence, a simpler antialiasing analog filter with a more gradual transition band can be used.

In quadrature sampling, the signal to be digitized is split into two signals. One of the signals is multiplied by a sinusoid to down-convert the signal to a zero-center frequency and then filtered to form the in-phase component of the analog signal. The other signal is multiplied by a 90° phase-shifted version of the sinusoid and filtered to form the quadrature component. Because each of these two signals occupies only one-half the bandwidth of the original RF signal, the sampling rate can be reduced by one-half at the expense of requiring two ADCs.

Band-pass sampling is based on the band-pass sampling theorem which states that a band-pass signal with no frequency components below f_L Hz and none above f_H Hz can be determined uniquely by sampling the signal at a rate F_S Hz, where

$$\frac{2f_H}{k} \le F_S \le \frac{2f_L}{k-1}$$

where k is restricted to integral values that satisfy

$$2 \le k \le \frac{f_H}{f_H - f_L}$$

and

$$(f_H - f_L) \le f_L$$

Band-pass sampling provides an image of the desired signal at multiples of the sampling frequency, and the spectral replica of the original analog band-pass signal closest to dc is usually chosen for further processing.

Once sampling is over, the sampled analog signal with its infinite range of amplitudes has to be converted to a finite set of discrete amplitudes. This is known as quantization. There are two general quantization methods, uniform and nonuniform quantization. In uniform quantization, the voltage difference between each quantization level is the same. In nonuniform quantization, the quantization levels are nonlinearly spaced in voltage. The ADCs used in RF and IF digitization typically use uniform quantization. In uniform quantization some error is introduced into the quantized signal because the analog signal cannot be represented exactly by a finite number of discrete amplitude levels. Statistically, it can be assumed that the error signal is uniformly distributed within a quantization level.

Signal Processing. Signal processing is the core of the radio receiver and is the segment where the original transmitted information signal is recovered. Many operations are carried out by this system. These operations are quite application specific and may include some or all of the following:

1. Down-conversion
2. Filtering, either spatial or temporal or both
3. Equalization
4. Despreading
5. Synchronization
6. Demodulation
7. Automatic gain control
8. Carrier recovery
9. Error-correction decoding
10. Source decoding
11. Decryption
12. Timing recovery

Some of the most important signal processing functions of a receiver are examined here in further detail.

Down-Conversion. When band-pass subharmonic digitization is used, spectral replicas of the original analog signal are found at multiples of the sampling frequency. However, there is generally no spectral replica centered around the zero-center frequency. To generate the complex baseband signal centered around the zero-center frequency, the output of the ADC is sent to a pair of digital multipliers. The reference inputs for the digital multipliers come from a quadrature-output, numerically controlled oscillator (NCO). The multipliers shift the spectral replica to the zero-center frequency. Then the outputs of the multipliers are sent to low-pass digital filters, which are typically finite impulse response (FIR) filters, to recover the baseband signal and filter out the other mixer products.

Filtering. Digital filters (11) are widely used in communication signal processing for tasks, such as digital down-conversion, equalization, interference suppression, and pulse shaping. Following are the advantages of digital filters:

1. Exact linear phase filters can be implemented easily.
2. Filters with almost any desired frequency and phase response can be designed and implemented easily.
3. Changes in filter responses due to component variations caused by aging are eliminated.
4. Changes to the filters can be carried out easily because most of the filtering is implemented in software running on programmable processors.

Digital filters are of two general types, finite impulse response (FIR) filters and infinite impulse response (IIR) filters. Finite impulse response filters are the most common digital filters in radio receivers mainly because of the following advantages:

1. Filters with exactly linear phases can be easily designed.
2. There are efficient recursive and nonrecursive realizations of FIR filters.
3. FIR filters realized nonrecursively are always stable.
4. Round-off noise inherent in finite precision arithmetic implementations are easily made small for nonrecursive realizations.

Following are the disadvantages of FIR filters:

1. A large filter order is required for sharp cutoff filters.
2. The delay of linear phase FIR filters need not always be an integral number of samples.

The main advantage of IIR filters is that sharp cutoff filters can be realized in relatively small filter orders. Following are the main disadvantages:

1. IIR filters generally do not possess linear phases.
2. IIR filters are more prone to be unstable because of quantization and round-off noise.

Carrier Recovery. Coherent demodulation requires that the phase and frequency of the transmitted carrier be known. Carrier recovery is the process of estimating the phase and frequency of the carrier to establish a reference for demodulation at the receiver. Any error in estimating the phase and frequency of the carrier causes significantly degraded performance. The information signal may be modulated onto the RF carrier so that a residual component at the RF exists in the overall transmitted signal spectrum. This residual RF component can be easily tracked by a narrow-band, PLL and provides the desired reference signal. However, this residual component represents power unavailable to transmit the information. Techniques that conserve power are of interest especially in mobile applications where power is at a premium because batteries supply power to the radio. As such, suppressed carrier transmissions are widely used. Following are some of the popular suppressed carrier recovery techniques:

1. Squaring loop
2. Costas loop
3. Decision feedback loop

Demodulation. The ultimate function of the radio receiver is to recover the original information signal that modulated the transmitted carrier. This process is known as *demodulation.* The portion of the receiver system that implements demodulation is known as the *demodulator.*

Some of the popular FSK demodulators are the limiter-discriminator, the PLL, and noncoherent and coherent demodulators. Noncoherent demodulation is carried out by a bank of band-pass filters whose outputs are envelope-detected. The largest output is selected as the transmitted symbol. The band-pass filters used to detect the tones are implemented either as FIR or IIR filters. Coherent demodulation compares the received signal to all of the reference frequencies. The comparison is done by multiplying the received signal by all the reference signals and then low-pass filtering the outputs of the bank of multipliers. The largest output is selected as the demodulated symbol. The locally generated reference signals must be synchronized in phase and frequency to the transmitted signal states.

Common PSK demodulators are the coherent demodulator and the differential demodulator. PSK can also be demodulated by using a frequency demodulator, such as a limiter-discriminator or a PLL, and integrating the output before the decision stage. The block diagram of a coherent PSK demodulator is shown in Fig. 7. Differential demodulators determine the cosine and sine of the phase difference and then decide on the phase difference accordingly. Differential demodulation shown in Fig. 7 is implemented by taking the product of the signal and a delayed version of the same signal. The output of the multiplier is low-pass filtered, usually by a FIR filter, to recover the information symbol transmitted. The transmitted symbols need to be encoded differentially to use differential demodulation. Differential demodulators are often used in highly mobile applications where fading in the channel makes it impossible to get a robust coherent estimate of the transmitted carrier.

Processing Implementation. The processing elements that implement the functions in a radio are crucial, especially in software radio architecture, as they implement virtually all of the functions of the radio transceiver except the frequency

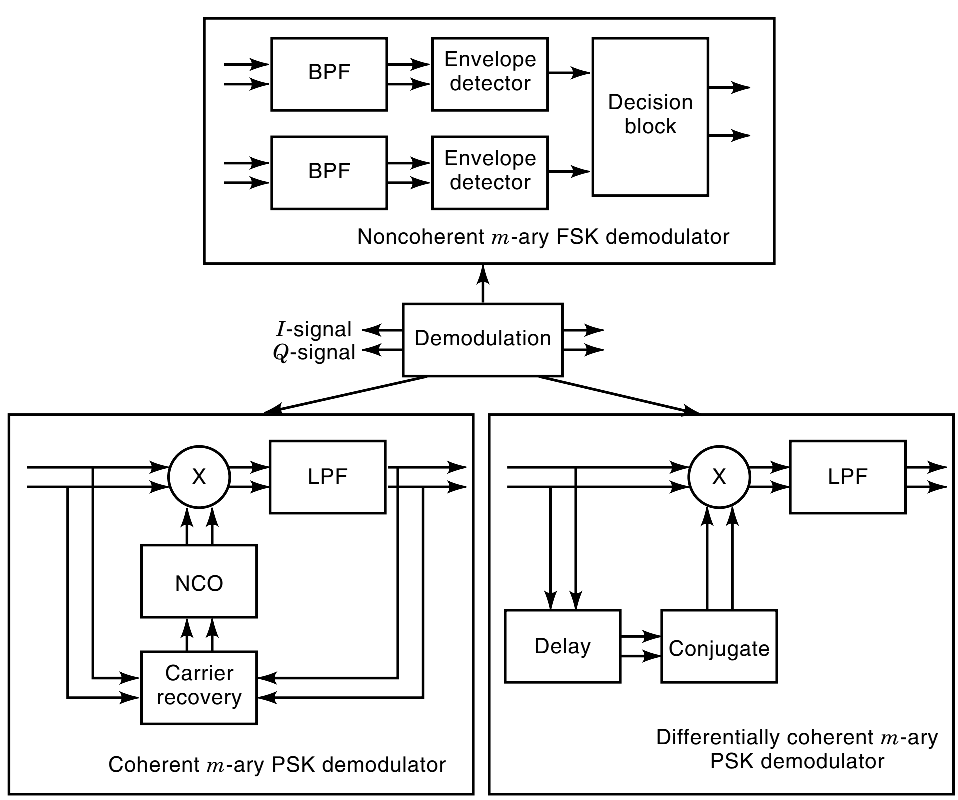

Figure 7. Some general demodulator architectures.

conversion and RF amplification. The main attraction of the software radio concept, namely, its flexibility and ease of adaptation, is possible because signal processing is implemented in software modules. The software requires a hardware platform to run on, and the capability of the hardware architecture of the processing platform is critical.

Typically digital signal processing functions are implemented on special digital processors called *digital signal processors* (DSPs) (12). Although DSPs are also microprocessors, there are a number of crucial differences between general multipurpose microprocessors and DSPs. General microprocessors are typically built for a range of general functions and normally run large blocks of software. The DSP, on the other hand, is built for a small dedicated group of tasks, the most important being the multiply-accumulate arithmetic operation which forms the core of any digital filter. DSPs contain large, high-speed data busses and use direct memory access (DMA) to transfer large amounts of data, thereby avoiding communication bottlenecks. In addition, DSPs contain dedicated hardware blocks, such as multipliers, to speed up the arithmetic-intensive signal processing steps.

Some of the signal processing functions are so complex that parallel and sequential partitioning of algorithms is required to get the required processing power. DSPs are getting faster but are currently incapable of implementing everything possible on a single chip. One approach has been to use multiprocessing to share the computational burden. The traditional approach to multiprocessing has been to integrate various DSPs on a board. Modern DSPs contain various hooks to simplify multiprocessing, such as simplified addressing across processors. A recent approach has been to integrate multiple processors within a single chip. This within-the-chip approach benefits from having closely coupled memory and cache which improves communication efficiency.

To keep the flexibility of a programmable solution and the efficiency of a dedicated solution, field-programmable gate arrays (FPGAs) are increasingly becoming another viable option to implement highly complex signal processing functions. FPGAs are logic devices whose hardware architecture can be programmed before use. Techniques, such as distributed arithmetic for array multiplication, can increase the data bandwidth and throughput of an FPGA-based solution by orders of magnitude beyond those possible with general purpose DSPs. It is projected that DSP cores will have on-chip FPGA sections to provide configurable accelerators.

CONCLUSION

This article has presented an overview of the radio transceiver, its architecture, and the implementation of its most important system blocks. A number of other specialized functions are not present in all transceivers and have not been discussed here. Some of these circuits are noise limiting and blanking, squelch, diversity reception, and adaptive antenna array processing.

Traditionally, transceivers have used analog circuits for implementation. The capabilities and advantages of digital processing have allowed replacing many of these analog functions. The movement of the digital portions of the processing closer to the antenna has resulted in the development of software radio architecture. The software radio is a powerful architectural framework that helps to deliver advanced radio services by leveraging the economics of contemporary microelectronics and software technologies. Though much technological progress has been made in the field of digital processing, technology is not currently available to implement the ideal software radio. Following are some of the challenges and issues that face radio designers today:

1. To engineer low-cost, low-loss, and low-distortion wideband antennas
2. To engineer low-cost, low-loss, and low-distortion wideband RF front ends
3. To develop high-efficiency linear power amplifiers
4. To develop low-power integrated RF front ends
5. To develop low-cost, high-resolution ($>$14 bits) and high-speed ADCs
6. To develop low-cost, high-speed reconfigurable digital processors

The development of the software radio transceiver is by no means over. Further technological advances are required, especially in the hardware implementation of ADCs and reconfigurable processors. There will be further development toward integrating the analog RF front end into a single integrated circuit. The development of low power RF and digital circuits is another challenge. The ultimate goal of implementing a radio on a chip, although not yet a practical reality, is not far away.

ABBREVIATIONS

ADC	Analog-to-digital converter
AFC	Automatic frequency control
AGC	Automatic gain control
AM	Amplitude modulation
BER	Bit error rate
BPF	Band-pass filter
DAC	Digital-to-analog converter
FIR	Finite impulse response
FM	Frequency modulation
FPGA	Field-programmable gate array
FSK	Frequency-shift keying
Hz	Hertz
IF	Intermediate frequency
IIR	Infinite impulse response
LNA	Low-noise amplifier
LPF	Low-pass filter
MDS	Minimum detectable signal
NCO	Numerically controlled oscillator
NF	Noise figure
PLL	Phase-locked loop
PM	Phase modulation
PSK	Phase-shift keying
RF	Radio frequency

BIBLIOGRAPHY

1. U. L. Rohde, J. Whitaker, and T. T. Bucher, *Communications Receivers,* 2nd ed., New York: McGraw-Hill, 1996.

2. M. Faulkner and M. A. Briffa, Amplifier linearisation using RF feedback and feedforward techniques, *Proc. Veh. Technol. Conf. (VTC),* 1995, pp. 525–529.
3. S. Haykin, *An Introduction to Analog and Digital Communications,* New York: Wiley, 1994.
4. H. Meyr, M. Moeneclaey, and S. A. Fechtel, *Digital Communication Receivers-Synchronization, Channel Estimation, and Signal Processing,* New York: Wiley, 1998.
5. J. Mitola, The software radio architecture, *IEEE Commun. Mag.,* **33**: 26–38, 1995.
6. J. K. Cavers and S. P. Stapleton, A DSP-based alternative to direct conversion receiver for digital mobile communications, *Proc. Veh. Technol. Conf. (VTC),* 1990.
7. D. B. Chester and J. Fakatselis, The architecture and benefits of digital IF up conversion implementations, *Proc. Int. Conf. Signal Process. Appl. Technol. (ICSPAT),* 1995, pp. 356–360.
8. T. S. Rappaport, *Wireless Communications-Principles and Practice,* Upper Saddle River, NJ: Prentice Hall, 1996.
9. G. Ungerboeck, Trellis coded modulation with redundant signal sets, part 1: Introduction, *IEEE Commun. Mag.,* **25**: 5–21, 1987.
10. J. A. Wepman, Analog-to-digital converters and their applications in radio receivers, *IEEE Commun. Mag.,* **33**: 39–45, 1995.
11. J. G. Proakis and D. G. Manolakis, *Digital Signal Processing,* 2nd ed., New York: Macmillan, 1992.
12. R. Baines, The DSP bottleneck, *IEEE Commun. Mag.,* **33**: 46–54, 1995.

FRANCIS DOMINIQUE

TRANSCONDUCTANCE-C INTEGRATOR. See INTEGRATING CIRCUITS.

TRANSCONDUCTOR-CAPACITOR FILTERS. See ANALOG INTEGRATED CIRCUITS.

TRANSCUTANEOUS INDUCTIVE LINKS. See PROSTHETIC POWER SUPPLIES.

TRANSDUCERS. See UNDERWATER SOUND PROJECTORS.

TRANSDUCERS, BIOMEDICAL. See BIOMEDICAL SENSORS.

TRANSDUCERS, MAGNETORESTRICTIVE. See MAGNETOSTRICTIVE DEVICES.

TRANSDUCERS, PRESSURE. See PRESSURE SENSORS.

TRANSDUCERS, UNDERWATER SOUND. See UNDERWATER SOUND PROJECTORS.

TRANSFER FUNCTION. See FILTER APPROXIMATION METHODS.