3.3.2. Dipoles for Circular Polarization. For applications that require a circularly polarized antenna such as TV and FM broadcasts and space communications, at least two dipoles, each of which has a linear polarization, must be combined in an array, often referred to as crossed dipoles. In a crossed dipole configuration, dipoles are mounted perpendicular to each other for circular polarization or at other angles for elliptical polarization. Currents are fed $90^\circ$ out of phase between the two dipoles. These can also be used as probes for sensing vector fields to isolate individual components of the electric field. Adaptations of the crossed dipole are shown in Figs. 16a and 16b. Dipole arrays such as the Yagi–Uda can also be combined to provide circular polarization, as shown in Fig. 16c.

**Figure 15.** Typical $E$- and $H$-plane patterns of a Yagi–Uda array; total number of elements $= 27$, number of directors $= 25$, number of reflectors $= 1$, number of driven elements $= 1$, total length of reflector $= 0.5\lambda$, total length of feeder $= 0.47\lambda$, total length of each director $= 0.406\lambda$, spacing between reflector and feeder $= 0.125\lambda$, spacing between adjacent directors $= 0.34\lambda$, radius of wires $= 0.003\lambda$ (from Ref. 7).

**Figure 16.** Cross-dipole applications for circular or elliptical polarization (from Ref. 6): (a) two shunt-fed slanted V dipoles; (b) series-fed slanted dipoles; (c) circularly polarized Yagi–Uda array.

**BIBLIOGRAPHY**


**DIRECT SATELLITE TELEVISION BROADCASTING**

JOHN P. GODWIN
DirecTV

Direct-to-home (DTH) satellite television broadcasting has no strict technical or legal definition. Since the late 1970s the term has been used to delineate commercial systems that deliver television directly to consumer homes using communication satellites in geosynchronous orbit. Systems originally intended for DTH applications have operated at downlink frequencies above 11 GHz and with antennas of 1 m or less. Certain systems operated at 4 GHz were planned for cable television distribution and became, secondarily, DTH systems with customer parabolic antennas in the 2.5–3.0 m range. Most systems have been supported primarily from subscription and pay-per-view revenues rather than advertising revenues. In the
various direct-to-home systems deployed worldwide, a variety of technologies has been used, including analog and digital modulations and both standard and high-definition television formats. Certain systems have been entirely national in scope, while others have broadcast on a regional basis. From a regulatory viewpoint, both fixed satellite service (FSS) bands and broadcasting satellite service (BSS) bands have been used. Direct-to-home systems are sometimes also referred to as direct broadcast satellite (DBS) systems. This article describes the broadcasting and reception systems of a typical digital DTH broadcasting system, but does not cover the substantial infrastructure necessary for customer service and billing.

1. EVOLUTION AND EXISTING SYSTEMS

Although DTH satellite television was a dream of satellite engineers since the early 1960s, little progress was made until the early 1980s. Satellite technology steadily improved in generating high-radiofrequency (RF) power levels, and ground electronics improved by the introduction of low-cost, low-noise microwave transistors. Through 1994 these early systems used analog frequency modulation.

During the 1980s in the Americas, the earliest major system was the Satellite Technology Corporation project in the United States. This plan intended to deliver five channels to each time zone with a dedicated satellite for each. The user terminals were to employ parabolic dishes of 85 cm diameter. This project was abandoned, primarily for economic reasons. Also, in the United States during the 1980s, home reception began of satellite transmissions intended for delivery to cable television systems. The transmissions were at C band in the frequency range 3.7–4.2 GHz. These early home dishes were 2.5–3.0 m in diameter and cost several thousands of dollars, but increased satellite power permitted new C-band home dishes to drop in size to about 1.5 m by the early 1990s. This United States C-band DTH marketplace peaked at about 3.9 million homes in 1994. In Japan the quasigovernment broadcaster NHK utilized satellite delivery to 45-cm dish antennas for both standard National Television Systems Committee (NTSC) and multiple sub-Nyquist encoding (MUSE) high-definition television. By 1993 this service, called BS for broadcasting satellite, was received by 4.5 million homes. The inexpensive analog BS receivers also became a typical feature of new television sets for the Japanese marketplace. In Europe the early use of satellites was for delivery of state-owned television networks. In the early 1990s, the Astra satellites became a major vehicle for DTH delivery of private, commercial channels. Multiple television broadcasters utilized Astra, including British Sky Broadcasting (BSkyB), which was providing over 40 analog channels to 6.4 million homes at the end of 1997 [1]. Other European satellites are also providing DTH services, including Eutelsat and Hispasat.

During 1994 the era of multichannel, all-digital DTH satellite delivery began with two systems in the United States, the Primestar system owned by a consortium of cable firms, and a system operated primarily by DirecTV, Inc., a unit of Hughes Electronics. The Primestar system used “medium-power” satellites and approximately 0.75–1.0 m dishes; the DirecTV broadcast used “high-power” satellites and 45-cm dishes. By late 1997 the Primestar system delivered more than 160 channels to nearly 2 million homes in the United States by year end 1997. The DirecTV service delivered more than 175 channels to 3.3 million homes in the United States by year end 1997. In 1995 another DTH business using “high-power” satellites entered this marketplace; this new entrant, EchoStar, reached approximately 1 million homes by year end 1997. AlphaStar, a short-lived DTH service, acquired only about 51,000 subscribers in the United States before filing for bankruptcy in 1997. Elsewhere in the Americas, three DTH services to Latin America were initiated in the early 1990s. One of these ventures, Galaxy Latin America, began broadcast operations in June 1995. Galaxy Latin America is a joint venture of Hughes and major media firms from Mexico, Venezuela, and Brazil. (The general company information given above was found at the World Wide Web sites listed in the Further Reading list.)

In Japan in 1996, the joint venture PerfecTV started multichannel, all-digital broadcasting with approximately a half-million subscribers by year-end 1997 [2]. This firm was joined in the marketplace by DIRECTV JAPAN in December 1997. A third entrant, Japan Sky Broadcasting (JSkyB), announced in 1997 that it would merge with the first broadcaster, PerfecTV. All three firms use a Japanese industry variant of the digital videobroadcasting (DVB) format, and all three use medium-power FSS satellites. Within Japan the category of service provided by these three competitors is called digital communications satellite, or digital CS, in contrast to the high-power broadcasting satellite or BS service by NHK.

By early 1998 in Europe there were plans underway to convert existing analog systems, for example, BSkyB in the United Kingdom, and to launch new digital satellite platforms. New digital satellite systems in operation include DF1 in Germany; Telepiu in Italy; Via Digital (Hispasat) and Canal Satellite in Spain; TPS, AB-sat, and Canal Satellite Numerique in France [3].

2. REFERENCE ARCHITECTURE

Figure 1 shows a simplified diagram of an all-digital multichannel satellite DTH system. Figure 2 shows the exterior of a typical DTH broadcasting site including four 13-m uplink antennas.

2.1. Broadcasting Facility

Most existing DTH systems have been used as delivery systems for existing programs, for example, broadening the market exposure of existing programming or delivering the programming with improved quality or convenience. As a delivery or rebroadcast system, a substantial portion of programming typically arrives at the DTH broadcasting or uplink facility via other “backhaul” satellites or terrestrial fiber. Programming, such as theatrical films, arrives at the facility as prerecorded
digital tapes. In a limited number of systems, the broadcasting facility also includes studios for the creation of unique programming.

The broadcasting facility provides a number of functions common to any broadcasting facility, such as incoming signal monitoring, adjustment, and resynchronization, signal routing within the facility, and for prerecorded material, quality control, cloning, and playback. For playback, broadcast-quality tape players are utilized or, more recently, the material is stored on and played from video file servers using redundant arrays of independent disks (RAID) technology.

Large, multichannel “pay” DTH broadcasting also requires that the broadcast site provide conditional access equipment, service information/electronic program guide (SI/EPG) equipment, compression encoders, and multiplexing, error control, and modulation equipment. The conditional access system, which includes equipment within the home, permits customer access to programming services only when certain conditions have been met—for example, the customer account is in good standing or the customer is located outside a program blackout area. The SI/EPG equipment prepares specialized broadcast streams that provide the consumer equipment with technical attributes of each view channel (the service information), along with program content information for display by the home receivers. The EPG data typically include program title, start and stop time, synopsis, parental rating, etc. The signal compression equipment performs redundancy reduction processing on the television video and separately the audio to reduce the total information rate. A typical digital studio signal at 270 Mbps (megabits per second) is reduced to the range of 2–10 Mbps before broadcast. This dynamically reduces the investment needed to put the transmission path in service (i.e., the satellites) and, conversely, greatly increases the number of available viewer channels for a given satellite investment. Most operational digital DTH systems have utilized the Motion Picture Experts Group (MPEG) MPEG1 or MPEG2 compression standards [4,5], or proprietary systems with similar characteristics. (See Section 4). The compressed streams from multiple channels are typically multiplexed into a single high-speed stream. This multiplexing process may be “fixed” in that peak bit rates are allocated to each video channel or, in certain systems, the individual channel rates may vary dynamically depending on their instantaneous bit rate need—the latter approach is called statistical multiplexing. The composite bitstream is then coded by error control to add

Figure 1. Simplified diagram of an all-digital multichannel satellite DTH system. Major broadcasting and transmission equipment groups are shown but not customer service and billing systems.

Figure 2. This DTH site in Colorado uses four 13-m antennas for uplink operations and numerous smaller dishes for programming reception. Courtesy of DirecTV, Inc.
selective redundancy for error detection and correction. The error control coding permits systems to be designed that offer high-quality operation with a threshold level lower than that possible in previous analog systems. The modulation utilized is commonly a constant envelope modulation such as quadrature phase shift keying (QPSK), which is typical of a satellite system for which the satellite repeater has a limiting final output stage.

2.2. Transmission Path

The transmission path includes the error control coding and modulation described above, the uplink site's upconverters, transmitters, and antennas, the uplink propagation path, the relay satellite, and the downlink transmission path including the subscriber antenna and receiver front end. In all existing DTH systems, the satellite has been a frequency translating microwave repeater. The expense of generating high satellite RF transmitter levels has caused these systems to be “downlink limited,” meaning that the composite uplink and downlink carrier-to-noise ratio (CNR) is dominated by the downlink CNR. The downlink CNR is determined primarily by the satellite effective isotropic radiated power (EIRP) per transponder, carrier attenuation by rain along the line of sight, and the subscriber antenna gain. The subscriber electronics equipment completing the transmission path consists of a small-aperture antenna, a low-noise block downconverter, tuner, demodulator, and error control decoder. The “error corrected” information stream out of the error control decoder is passed to the remainder of the digital circuitry within the receiver. (See Section 5.)

2.3. Home Electronics

The home electronics in a typical all-digital system include the antenna, low-noise block (LNB) converter, tuner–demodulator–decoder circuitry and other digital circuitry for demultiplexing, decryption under conditional access control, video and audio decompression, and video and audio output signal generation. For example, in digital receivers for the United States marketplace the final output circuitry recreates an analog composite NTSC or S-video signal for delivery to a standard television set. In a typical digital satellite receiver, a removable device, often in the form of an International Organization for Standardization (ISO) smart card, provides the conditional access control function. (See Section 6.)

3. THEORETICAL MODELS

3.1. Information Theory

Figure 3 shows a theoretical model useful in DTH system design, and the corresponding system elements used to implement the theoretical model. A text such as Ref. 6 describes an “ultimate” system design in which source encoding is used to remove redundancy information in the bitstream representing the source, that is, the television signal, and then channel encoding to protect the encoded source by carefully adding redundancy. Information theory tells us that source codes exist that can drive the number of bits necessary to encode the source toward a theoretical minimum. MPEG2, shown in the lower portion of Fig. 3, provides a practical realization of the information theory by a complex set of transform, run-length, and other source codes. The MPEG algorithm further reduces the information content by selective removal of detail not subjectively important. Channel-coding theory indicates that channel codes exist that can drive the error rate toward zero while not driving the useful throughput toward zero. In 1966 Forney [7] demonstrated a path to realization of this theory by showing that concatenating multiple, simpler channel codes can create a powerful channel code. Figure 3 illustrates a DTH implementation using concatenated convolutional and Reed–Solomon (RS) codes. A bit interleaver is also used to “smooth” burst error sequences entering the RS decoder.

3.2. Layered Model

Figure 4 provides a “layered” or “protocol” model for DTH systems [8]. The layers shown are for the consumer
electronics part of the system, but of course the same layers are necessary within the broadcasting facility equipment. As in layered, communications protocols, it is intended that the design tradeoffs of one layer do not interact with the design tradeoffs of the adjoining layers. For example, the design of MPEG decoder chips is largely independent of the design of the video output circuitry, which may be targeted for either NTSC, phase alternation line (PAL), or sequential couleur avec memoire (SECAM) television receivers. As another example, the MPEG coder to decoder syntax was largely designed without great concern about the specific error characteristics of the channel. However, to improve recovery in the event of channel errors, the MPEG standard does include a Macro Block Slice structure that generally limits error propagation to a portion of a frame.

Each layer is discussed in the following and in Sections 4 and 5. The realization of these protocol layers is discussed in Section 6.

3.2.1. Physical Layer. The physical layer at the bottom of Fig. 4 presents the RF to an intermediate-frequency (IF) LNB converter and the resulting IF interface to the digital receiver itself. An IF frequency starting at 950 MHz is typical but not required [8].

3.2.2. Link Layer. This layer is discussed in detail later in Section 5.

3.2.3. Transport Layer. The transport layer is a multiplexing layer or, for example, the systems layer of the MPEG2 standard. In each format given in Ref. 8, this layer provides common, fixed-length packets for all service types including video, audio, data, or overhead data such as electronic program guide information. Fixed-length packets ease high-speed processing and use of direct memory access.

3.2.4. Conditional Access Layer. This layer provides decoding of specialized conditional access (CA) packets, sometimes called entitlement management messages (EMMs) and entitlement control messages (ECMs) [9]. The EMMs give instructions to the subscriber electronics regarding the authorized entitlements, for example, current subscriptions or pay per view status. The ECMs indirectly provide cryptographic keys for decryption of the individual services. In several systems these packets are passed from the receiver to a smart card with an embedded secure microprocessor. The microprocessor decodes ECMs and returns the corresponding keys. A decryption circuit within the receiver uses the keys and provides decrypted packets for each service to the network services layer. The receiver to microprocessor interface is often similar to the ISO standard [10].

3.2.5. Network Services Layer. This layer delivers the underlying DTH technical services. These services include video plus audio or “television,” standalone audio services, and data delivery services. Separate processes handle the decompression of each service type. For example, video decompression algorithms are quite distinct from those used for audio decompression. Video compression is discussed in greater detail later in Section 4. Other network services include decoding of the electronic program guide and service information syntax.

3.2.6. Presentation Layer. This layer puts the network services in final form for the end user. The layer includes the NTSC or PAL encoders and output circuitry and the on-screen user interface. Although the electronic program guide information delivered by the network is common to all receiver types, each receiver designer may choose a unique user interface concept. For example, for a typical
television program schedule grid, the grid extent (that is, numbers of view channels and time extent) and the color scheme are entirely up to the designer. The presentation layer also receives inputs from the user remote control, which is typically linked to the receiver using infrared or RF.

3.2.7. Customer Services Layer. In most DTH systems the customer provides the final display device such as the television or personal computer. This key assumption bounds the complexity of the satellite receiver and defines the characteristics of its output circuitry. For example, although a typical all-digital DTH system can deliver a three-component television signal, most existing televisions in the United States accept only a composite NTSC input. Since many new sets in the United States also accept a "separate chroma/luma" S-video signal, many satellite receivers in the American marketplace have supplied an S-video output in addition to the composite output. Figure 4 shows the remote control interfacing with both the presentation layer (user interface) and the customer services layer. The latter interface permits control of the display device by the same remote control device—for example, the remote may control the television volume level.

4. COMPRESSION

4.1. Fundamentals

Source coding may be lossless and permit a complete reconstruction by the source decoder, or source coding may be lossy and trade the quality of the reconstructed signal against the bits needed to transmit or store the signal. The nature of the compression algorithms vary with the signal type, its intended audience, and the cost relationship between the value of "saving bits" versus the value of the codec development and production. Reference 11 provides an excellent overview of the television compression state of the art through 1994. The MPEG1 [4] and MPEG2 standards [5] have been broadly deployed in consumer products. The MPEG1 standard is intended for noninterlace video and data rates up to about 1.5 Mbps. The MPEG2 standard accommodates both noninterlace and interlace video, standard definition applications up to about 10 Mbps, and high-definition formats at bit rates up to about 15–50 Mbps. Note that while these standards provide details on the syntax and semantics between the encoder and decoder and are specific to a standard decoder, they say very little about the encoder. They are very abstract and do not dictate the technology of implementation. Both standards utilize two distinct processes in tandem to achieve high compression levels: discrete cosine transform (DCT) coding and motion-compensated interframe prediction. The MPEG2 standard makes more complex algorithms available for motion compensation with interlace video. The discussion that follows provides a very simplified description of MPEG processing.

4.2. Discrete Cosine Transform Coder

Figure 5 illustrates the first major MPEG process, a DCT of pixel element values, the lossy quantization of these values, and then the lossless encoding of the result. Consumer television signals are acquired and displayed as linescan images, but since substantial spatial redundancy exists, the linescan images are first converted to 8 x 8 pixel blocks for MPEG processing. The DCT represents the horizontal and vertical information in the block using cosine functions as the basis vectors. The quantization step ignores the near-zero coefficients and tends to concentrate the energy in the transform domain into the low-frequency components. The "zigzag" readout of the coefficients starts with the DC coefficient and proceeds in zigzag fashion toward the highest frequency vertical and horizontal component. If insufficient bits are available, the higher-frequency coefficients may not be encoded. The next processing steps use tables of run-length and variable-length codes, which, based on experiment, will require the lowest average number of bits to represent the coefficients. The run-length codes use short codes for very likely bit sequences and long codes for less likely sequences. The variable-length codes are created such that no codeword is the prefix of another codeword. The buffer feedback path recognizes that image redundancy varies substantially across the blocks of a given image, but that for most applications the required output bit rate must be constant. As buffer fullness approaches, quantization can be increased, the bit rate reduced and, unfortunately, the quality will be reduced as well. Note that in multichannel systems, the output of a single encoder need not be at a fixed bit rate. In a DTH system, when a buffer strategy is used across all the video channels carried in a single stream, the technique is called statistical multiplexing.

In MPEG the spatial DCT coder described above is supplemented with an interframe predictor to also exploit the temporal redundancy of a given pixel block. Since motion within the image will cause the pixel values to "move" across the frame, the MPEG algorithms also include the technique discussed in the following.

4.3. Motion-Compensated Interframe Prediction

In MPEG, motion compensation determines the translation vector of 16 x 16 pixel blocks of luminance (called macroblocks) across multiple frames. Redundancy reduction is achieved by transmitting the vectors and quantized

![Diagram](image-url)
prediction errors, rather than the blocks, and further efficiencies are achieved by differentially encoding the vectors and also using variable length codes. The vectors are determined by finding the best macroblock match in the previous (and possibly also the future) reference frame. These searches generally are restricted in horizontal and vertical extent and can be very computationally intensive.

The MPEG2 algorithm is more sophisticated than MPEG1 in several areas, particularly motion compensation modes. Both compression schemes permit forward prediction, backward prediction, and interpolated prediction between images. The images may be either video frames or fields. MPEG1 can use only frame-based prediction; however, MPEG2 optionally can use field-based prediction, which allows increased coding efficiency for interlaced video. From video material in which the motion is slow, frame prediction is more efficient and MPEG2 performs similarly to MPEG1. As motion increases, field prediction coding becomes more efficient.

The MPEG2 toolkit is very complex. It is impractical to recreate the entire toolkit in every application. The MPEG2 group has therefore defined a handful of subsets or profiles of the full syntax. Also, within a profile, sets of parameter constraints have been identified as levels, with each higher level including all constraints from the lower levels.

5. TRANSMISSION (OR LINK LAYER)

5.1. Link Equation, Antenna Size, and Coverage

The most fundamental design equation in a DTH satellite system is the communications link equation. Ignoring uplink noise and interference contributions, the downlink carrier power \( C \) to noise power density \( N_0 \) ratio is, in decibels, as follows [12]:

\[
\frac{C}{N_0} = \text{EIRP}_s - \text{BO}_0 - L_d + \frac{G}{T_e} - k - L_r \text{ dB} \cdot \text{Hz} \tag{1}
\]

where the EIRP\(_s\) is the effective power of the satellite with respect to an isotropic radiator, \( \text{BO}_0 \) is the backoff of the satellite transmitter with respect to saturation, \( L_d \) is the free space loss at the carrier frequency, \( \frac{G}{T_e} \) is a receive figure of merit for the DTH subscriber terminal, \( k \) is Boltzmann’s constant \((-228.6 \text{ dB} \cdot \text{W/K} \cdot \text{Hz})\), and \( L_r \) is the link loss due to rain. Figure 6 illustrates the definition of these link parameters. For a typical DTH design, each satellite transmitter handles a single carrier, so carrier intermodulation is not a concern and the transmitter output backoff is nominally zero. As a reference case, assume a downlink frequency of 12 GHz, a path loss of \(-205.8 \text{ dB}\), and a clear weather situation with \( L_r = 0.0 \text{ dB}\). Equation (1) then becomes simplified to

\[
\frac{C}{N_0} = \text{EIRP}_s + \frac{G}{T_e} + 22.8 \text{ dB} \cdot \text{Hz} \tag{2}
\]

Using the parameters of Ref. 13 as an example, the typical edge-of-coverage EIRP is 52.0 dB\( \cdot \)W and the subscriber terminal \( \frac{G}{T_e} \) is 11.3 dB/K for a 45-cm dish. The clear-weather, edge-of-coverage performance is then

\[
\frac{C}{N_0} = 86.1 \text{ dB} \cdot \text{Hz} \tag{3}
\]

The required \( C/N_0 \) is determined by the information bit rate and the required \( E_b/N_0\) energy per information bit over the noise density, for the system’s modulation and coding with an implementation margin. The equation relating the two ratios is

\[
\frac{C}{N_0}_{\text{req}} = \frac{E_b}{N_0}_{\text{req}} + r + R \text{ dB} \cdot \text{Hz} \tag{4}
\]

where \( r \) is the coding rate and \( R \) is the transmission rate in dB \( \cdot \) Hz. The information bit rate is the product of the code
rate and transmission rate, which is expressed as a sum in decibels.

Using \( r = \frac{10}{9} \text{ Mbps} \) and \( R = 40 \text{ Mbps} \) (for an information rate of 30.6 Mbps or 74.9 dB \( \cdot \) Hz), and a threshold \( E_b/N_0 \) value of 7.8 dB \( \cdot \) Hz, then

\[
C \left| \frac{N_0}{N_0 \text{ req}} \right| = 82.7 \text{ dB} \cdot \text{Hz}
\]

The clear-weather, edge-of-coverage performance is then the difference of Eqs. (3) and (5), or 3.4 dB. This is the clear-weather margin for a 45 cm dish for the parameters of Ref. 13. Below a 30 cm diameter (for BSS systems in the Americas), intersystem interference sources due to adjacent satellites cause the simplified analysis used above to become quite inappropriate. Above about 90 cm, the narrow beam of the subscriber antenna may actually be detrimental to satisfactory performance. For narrowbeam subscriber antennas, small satellite stationkeeping errors may cause the line of sight to move outside the subscriber antenna’s main beam.

### 5.2. Propagation Effects at 12 GHz

In the preceding link example, the value \( L_c \) is a link margin against rain and other propagation phenomena. In fact, a system design is typically based on statistical and geometric models to predict the rain degradation along the line of sight. Although a variety of propagation-related impairments can occur, the dominant effects are due to rain and wet snow and result in signal attenuation and impairments can occur, the dominant effects are due to line of sight. Although a variety of propagation-related geometric models to predict the rain degradation along the line of sight, comprehensive models have been developed for DTH system design. Figure 7 shows the rain regions assumed by the International Telecommunication Union (ITU) for BSS planning for the Americas. Figure 8 illustrates the attenuation predicted by the ITU model for Region K of the Americas. The outage value assumed in Fig. 8 (1% of the worst-case month) is a requirement that should be reevaluated by the designer in each new application.

### 5.3. Interference

In addition to rain degradations, DTH designs must consider intrasystem and intersystem interference. Interference into the subscriber dish is a primary concern. Received interference includes cross-polarized, cofrequency, intrasystem interference, interference from other satellites operating at adjacent orbit locations, and emissions from terrestrial users of the same frequency band. The Broadcasting Satellite Service was carefully planned to separate orbital “slot” assignments for satellites with beams with common coverage [15–17], for example, for the United States the primary orbital assignments have 9° of longitude separation. In the Federal Communications Commission’s (FCC’s) Fixed Satellite Service assignments at 11.7–12.2 GHz, the satellites are separated by as little as 2° of longitude; the 2° separation causes dishes of less than 60 cm in diameter, with relatively little spatial isolation, to have generally unacceptable adjacent satellite interference. In the BSS in the United States, the use of 45-cm dishes is a common practice and one service provider has announced plans to use 30-cm dishes.

### 5.4. Satellite Design

Figure 9 illustrates a typical DTH satellite platform. All operational DTH systems have used satellites in geosynchronous orbit (GSO) with microwave frequency translation repeaters. After placement in a GSO, the satellite orbital period is equal (synchronous) with the rotational period of Earth, and the satellite appears to be stationary over a given longitude at the equator. This greatly simplifies the design of the millions of receive terminals that point toward the satellite.

A frequency translating repeater typically receives an uplink carrier via the receive coverage antenna beam,
mixes the signal to the downlink frequency, and then amplifies it for transmission to the transmit coverage beam. This type of translating design is highly reliable and flexible. The receive and transmit coverages need not be identical—for example, for a typical national system the receive beam coverage may be the 48 contiguous states while the transmit beam coverage may consist of all 50 states. (See Satellite Antennas.) The DTH satellite’s total DC and total RF power levels are key attributes since most of the satellite weight and hence cost are involved in generating high power levels. This relationship between weight and cost is largely due to the high cost and relative unreliability of launch vehicles. As one example, the Galaxy Latin America satellite launched at the end of 1997 had an end-of-life solar array power capability of about 8.0 kW. (See Ref. 18.)

The satellite electronics associated with processing and amplifying a single carrier, such as a multiplexed digital television carrier, is called a satellite transponder. The final output stage or transmitter is normally a traveling-wave-tube amplifier (TWTA), a device with very high gain, high efficiency, perhaps in excess of 50%, and wide bandwidth. Although high-power TWTA reliability was a DTH design issue through 1990, dozens of transmitters with power levels above 100 W have operated without apparent incident for more than three years during the 1990s.

5.5. Regulatory Considerations

Direct-to-home systems are typically regulated at the national level. Until relatively recently virtually all nations had a government-owned post, telephone, and telegraph (PTT) or quasigovernment agency that owned and operated all national telecommunication facilities. With the worldwide trend toward commercialization and competition, the government role is tending toward regulation of DTH businesses via RF and business licensing. For example, in Japan, the Ministry of Post and Telecommunications (MPT) has allowed three new commercial DTH businesses to compete with the traditional quasigovernment broadcaster NHK. Each of these businesses has required MPT approval over multiple aspects including the provider (consignor) of each programming channel, the content of the channel, the business viability of each channel, the RF licensing of the uplink site, and the RF licensing of the satellite.

The situation in the United States is somewhat unique in that competing private telecommunication businesses have existed since the 1970s. The FCC licenses and regulates satellite systems primarily via licensing of the satellites themselves. Small receive-only antennas do not require licensing, and the FCC has ordered that, in general, dishes smaller than 1 m cannot be regulated by state or local authorities [19].

Broadcasting from one nation into another commonly requires official landing rights in the distant nation, particularly if the broadcaster intends to collect subscriber fees. Issues of frequency use and coordination between nations are handled by an agency of the United Nations, the International Telecommunication Union [20]. The regulatory agency, that is, the PTT, MPT, or FCC, of each UN nation sends representatives to various ITU working groups to establish mutually agreed international regulations. In general, each agency makes the international
regulations a part of their national regulations; the ITU itself has no powers of enforcement. The ITU has established a number of frequency bands for satellite communications; for DTH applications the bands utilized have been both in the FSS and BSS. The FSS name comes from the fact that, for frequency coordination purposes, the transmitters and receivers are assumed to be at fixed locations, that is, not mobile. The BSS bands also assume "fixed" RF sources but with the added assumption that the primary usage is direct broadcast. In the early 1980s, when the international BSS arrangements were completed, this distinction was quite important since the representatives to the ITU sessions wished to be sure that their nations would someday have the benefit of DTH service. Thus the use of the BSS band (around 12 GHz) has been strictly planned such that every nation existing at the time of the agreement has a specific set of assigned frequencies, polarizations, and satellite antenna coverage. Note that the difference between the FSS and BSS bands is entirely regulatory, not technical—in fact, the BSS band in the United States is the same frequency as the FSS band in Japan and vice versa.

The ITU regulations are based on dividing the world into three regions with specific geographic boundaries. These regions are roughly defined as follows: Region 1 consists of Europe, Africa, and the former Soviet Union; Region 2 consists of the Americas; and Region 3 consists of Asia, excluding the former Soviet Union. Among these different regions the specific regulations can vary substantially, and within a given country the national administration may impose additional regulations. For example, under ITU auspices the BSS frequency bands vary by region, the frequency assignments vary by country, and each country assigns frequencies to a system or company. The downlink plan for all regions is contained in App. 30 of the ITU Radio Regulations [16]. The uplink or "feeder" link plan is contained in Appendix 30A of the ITU Regulations [17]. The 1997 World Radio Conference made changes to certain parameters for Regions 1 and 3 only.

Traditionally, telecommunications standardization has been performed by international groups such as the ITU, International Organization for Standardization (ISO), International Electrotechnical Committee (IEC), and the Joint Technical Committee (JTC1) of the ISO and IEC [21]. As a growing trend, standards are being addressed by regional groups such as the European Telecommunications Institute (ETSI) or the T1 committee in the United States, an organization accredited by the American National Standards Institute. Additionally, specialized, ad hoc groups have been formed to address certain areas of technology, for example, the Asynchronous Transfer Mode (ATM) Forum, the Internet Engineering Task Force (IETF), and the Digital Audio Visual Council (DAVIC).

The ITU has developed a standard for "Digital multi-programme television emissions by satellite" [8], but this is an international recommendation, not a requirement, unless implemented by a national agency as a national standard. This ITU recommendation includes three closely related broadcast formats that can be decoded by the same or similar receiver circuitry. Table 1 compares the characteristics of the three formats. All three utilize QPSK modulation, concatenated convolutional and Reed–Solomon coding, MPEG compression, and fixed-length transport packets.

In Europe the Digital Video Broadcasting (DVB) Project to develop specifications for all aspects of digital television broadcasting was launched in 1993, following a two-year effort. Since producing its first digital satellite standard, the DVB Project has developed specifications, guidelines, and recommendations for the many ancillary parts of digital broadcasting. These have been accepted as standards for DTH systems by more than 200 broadcasters, manufacturers, network operators, and by regulatory bodies in over 30 countries [9].

In Japan, the Association of Radio Industries and Businesses (ARIB) has selected a DVB variant as the standard for digital CS systems [22]. In the United States, the FCC has not required specific requirements for DTH services.

In general, these various standards have dealt with the link, transport, and network services layers, but not the conditional access layer. Certain governments have standards to restrict or specify the encryption method controlled by the conditional access system. For example, Japan’s MPT has specified the encryption algorithm for the new digital CS systems. Recently, DVB, DAVIC, and the Advanced Television Systems Committee (ATSC) in the United States have begun work on conditional access standardization. (See also TELEVISION BROADCAST TRANSMISSION STANDARDS.)

6. CONSUMER ELECTRONICS

Consumer electronics equipment for DTH applications has achieved very low cost, high performance, and excellent perceived value by carefully designed very-large-scale integration (VLSI) and mass production. By use of standards, for example, MPEG2, and standard techniques, many VLSI have been used in more than one platform and thus achieved greater economies of scale.

6.1. Outdoor Electronics

The offset fed parabolic reflector continues to be the dominant antenna type due to its simplicity and high gain for a given aperture size. The offset geometry achieves an aperture efficiency greater than 60% by eliminating the "feed blockage" present in a focus fed geometry. Figure 10 shows a 45 cm parabolic dish, digital receiver, and remote control produced by Panasonic in 1997 for the CS market in Japan. Single-polarization, fixed-scan, phased array antennas are also used, but generally not where their size would be greater than 60 cm x 60 cm, when distribution losses become significant.

6.2. Receivers

In the design shown in Fig. 10, the receiver supplies DC power to the outdoor electronics via the coaxial cable delivering the digital signals to the receiver. Additionally, biasing this supply voltage above or below the nominal value implements polarization selection at the feed. Figure 11 gives a reference architecture for a digital
DTH receiver [8]. This common architecture can be applied to any of the three digital formats contained within the reference.

6.2.1. Hardware. In implementing the reference hardware architecture of Fig. 11, the underlying large-scale integrated circuits (LSI) have since 1994 undergone multiple stages of evolutionary development. As an example, Table 2 summarizes the LSI evolution of the RCA-brand receivers produced for the United States marketplace. Each generation has seen greater levels of integration [23] with the fourth generation being a “two-chip” receiver.

Similar levels of LSI integration are also expected with the availability of such chips as the Texas Instruments Series AV 7000 shown in Fig. 12. This chip provides the equivalent of $2.5 \times 10^6$ transistors using 0.35 μm complementary metal oxide semiconductor (CMOS) technology. Advance information [24] indicates that this circuit will incorporate the following:

- 32-bit reduced instruction set computer (RISC) central processing unit (CPU) [40 million instructions per second (mips)]
- Advanced graphics accelerator
- Memory manager
- Transport/decryption (DES)
- MPEG2 video decoder (MPEG1 and MPEG2)
- Audio decoder (MPEG1)
- NTSC/PAL encoder

## Table 1. Summary of ITU Direct-to-Home Formats

<table>
<thead>
<tr>
<th>Function</th>
<th>System A</th>
<th>System B</th>
<th>System C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Randomization for energy dispersal</td>
<td>Yes</td>
<td>Explicit</td>
<td>Yes</td>
</tr>
<tr>
<td>Reed–Solomon outer code</td>
<td>(204, 188, $T = \frac{8}{23}$)</td>
<td>(146, 130, $T = \frac{8}{23}$)</td>
<td>(204, 188, $T = \frac{8}{23}$)</td>
</tr>
<tr>
<td>RS field generator polynomial</td>
<td>$X^8 + X^4 + X^3 + X^2 + 1$</td>
<td>$X^8 + X^4 + X^3 + X^2 + 1$</td>
<td>$X^8 + X^4 + X^3 + X^2 + 1$</td>
</tr>
<tr>
<td>Interleaving</td>
<td>Forney</td>
<td>Ramsey II</td>
<td>Forney</td>
</tr>
<tr>
<td>Inner coding</td>
<td>Convolutional, $K = 7$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basic code</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Generator polynomial</td>
<td>$2$</td>
<td>$\frac{8}{23}$</td>
<td>$\frac{8}{23}$</td>
</tr>
<tr>
<td>Forward error correction (FEC)</td>
<td>$1 2 3 5$ and $\frac{8}{7}$</td>
<td>$1 2 3 4 5 6$ and $\frac{8}{7}$</td>
<td>$1 2 3 4 5 6$ and $\frac{8}{7}$</td>
</tr>
<tr>
<td>Signal modulation</td>
<td>QPSK</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Symbol rate</td>
<td>Variable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transport layer</td>
<td>MPEG2</td>
<td>System B</td>
<td>MPEG2</td>
</tr>
<tr>
<td>Packet size [bytes] [payload]</td>
<td>188 [184]</td>
<td>130 [127]</td>
<td>188 [184]</td>
</tr>
<tr>
<td>Identification ID (bit)</td>
<td>13</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>Statistical multiplexing</td>
<td>Not restricted</td>
<td>Capable</td>
<td>Capable</td>
</tr>
<tr>
<td>Method of synchronization for video and audio</td>
<td></td>
<td>Timestamping (27 MHz reference)</td>
<td></td>
</tr>
<tr>
<td>Video source decoding</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Syntax</td>
<td></td>
<td>MPEG2</td>
<td></td>
</tr>
<tr>
<td>Levels</td>
<td></td>
<td>At least main level</td>
<td></td>
</tr>
<tr>
<td>Profiles</td>
<td></td>
<td>At least main profile</td>
<td></td>
</tr>
<tr>
<td>Audio source decoding</td>
<td>MPEG2 layers I and II</td>
<td>MPEG1 layer II (included in MPEG2)</td>
<td>ATSC A/53 or MPEG2 layers$^b$</td>
</tr>
<tr>
<td>Typical transponder bandwidth (MHz)</td>
<td>Not specified</td>
<td>24 or 27 MHz</td>
<td>24, 27, or 36 MHz</td>
</tr>
<tr>
<td>Selectable conditional access</td>
<td></td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Service information</td>
<td>ETS 300 468$^a$</td>
<td>System B</td>
<td>ATSC A/56 + 'SCTE' DVS/011$^b$</td>
</tr>
<tr>
<td>Electronic program guide</td>
<td>ETS 300 707$^a$</td>
<td>System B</td>
<td>User-selectable</td>
</tr>
<tr>
<td>Teletext</td>
<td>Supported</td>
<td></td>
<td>Not specified</td>
</tr>
<tr>
<td>Subtitling</td>
<td></td>
<td>Supported</td>
<td></td>
</tr>
<tr>
<td>Closed caption</td>
<td>Not specified</td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>Delivered TV standards</td>
<td>Not specified</td>
<td>NTSC and PAL M</td>
<td>NTSC and PAL</td>
</tr>
<tr>
<td>Aspect ratios</td>
<td>4:3 and 16:9 (2.21:1 optionally)</td>
<td>4:3 and 16:9</td>
<td>4:3 and 16:9</td>
</tr>
<tr>
<td>Video resolution formats</td>
<td>Not restricted</td>
<td>MPEG subset</td>
<td>MPEG subset</td>
</tr>
<tr>
<td>Frame rates (frames/s)</td>
<td>Not specified</td>
<td>29.97</td>
<td>25 (PAL) 29.97 (NTSC)</td>
</tr>
<tr>
<td>Compatibility with other</td>
<td>ISO/IEC 13818$^d$</td>
<td>Some processing required</td>
<td>ISO/IEC 13818$^d$</td>
</tr>
<tr>
<td>MPEG2 delivery systems</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^a$European Telecommunications Standards
$^b$Advanced Television Systems Committee (standards)
$^c$Society of Cable Television Engineers
$^d$See Ref. 5.
To complete a typical receiver, the designer will add the following:

- Tuner
- Link integrated chip
- Memory [read-only memory (ROM), random-access memory (RAM), dynamic RAM (SDRAM)]
- Smart card
- Telephone modem
- Other peripherals

The AV 7000 chip also provides an interface to external consumer electronics or computer devices using the IEEE 1394 aerial digital protocol.

6.2.2. Software. Software architectures have been driven by the functionality of the major VLSI. The “transport” chips have provided some degree of logical filtering of the high-speed data streams—and hence not required filtering by the primary processor. The MPEG chips have, in general, relied on the processor to perform group of pictures (GOP) and picture sequence-level processing. The primary processor also typically handles program guide and user interface tasks.

From a subscriber standpoint new product generations have exhibited faster response times and improved graphical interfaces. As a point of reference for 1998 technology, the Texas Instruments AV 7000 chip [24] was designed to provide graphics support for up to eight windows, 8 bit color depth, 16 levels of blending and transparency, overlapping windows, and other features.

More recent products have also included interactive or multimedia software layers, which include, for example, Panasonic’s DVX for DIRECTV JAPAN and Thomson Sun Interactive LLC’s OpenTV. The latter, a forerunner of an emerging interactive services industry, began in 1994 as an alliance of Thomson Multimedia and Sun Microsystems. OpenTV supplies interactive operating systems and services for digital receivers used by pay television services, among other activities [25].

7. MORE RECENT DEVELOPMENTS

7.1. Technology

Technological progress continues in most disciplines important to DTH digital systems. Satellite manufacturers have announced platforms with total DC power levels of at least 15 kW [26]. With a power generation capability 4 times that of DTH satellites launched in 1995, these newer platforms were in orbit before the year 2001.
Modulation and coding show continuing improvements, particularly in turbo coding [27], and advanced modulation formats with higher information content per unit bandwidth (bps/Hz) [28] are under consideration in newer system designs. Compression standards development continued with the MPEG4 standard [29], finally approved in late 1998 for version 1 and late 1999 for version 2. The MPEG4 architecture permits different compression algorithms to be applied to different source material. The proposed inventory of algorithms includes image decomposition into multiple objects and the existing transform-based algorithms such as MPEG2.

7.2. Proposed Systems and Services

New system filings in 1997 at the FCC included a DTH system [30] operating at a downlink frequency of 17.3–17.8 GHz, a band commonly referred to as the Ka band. This system filing proposed to accelerate initial use of this frequency band as a new BSS expansion band, planned by the ITU to first come into operation in 2007. Also in 1997 a group proposed a system called Skybridge [31], which would reuse the existing BSS band but with nongeosynchronous satellites. The system design uses 64 satellites in 1457 km altitude orbits for a variety of telecommunications services. The plan suggests that frequency reuse can be achieved by not broadcasting from a particular Skybridge satellite to a particular region unless, as seen from the user location, the separation angle is sufficiently large between the line of sight to Skybridge and to the geosynchronous satellite arc. According to the Skybridge plan, if the separation angle is large, and certain other conditions are met, the discrimination of the user antenna will lower the Skybridge interference to an acceptable level.

In the area of new service offerings, the Hughes DirecPC service is one example of DTH satellite broadcasting to a personal computer platform. This service, available in the United States, Japan, and Europe, utilizes a small, outdoor dish and, installed in a conventional PC, a digital satellite receiver card. During 1998 in the United States, the DirecPC services included both pull (two-way) and push (one way) Internet access. In early 1998, DirecTV, Inc., demonstrated high-definition DTH broadcasting with delivery of 1280 × 1080 picture elements (pixels) in interlaced signals to a television provided by Thomson Consumer Electronics. DirecTV announced that it would initiate nationwide high-definition (HD) broadcasts before the end of 1998 and coincident with the first terrestrial digital broadcasts [32].

Figure 12. Rapid VLSI progress now permits a single chip to provide all DTH receiver core functions and many secondary functions, such as a microprocessor. (Copyright © 1998, Texas Instruments Incorporated.)

### Table 2. Evolution of LSI in RCA-Brand DSS

<table>
<thead>
<tr>
<th></th>
<th>First Generation</th>
<th>Second Generation</th>
<th>Third Generation</th>
<th>Fourth Generation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major large scale integrated circuits (LSI)</td>
<td>Microprocessor</td>
<td>Microprocessor</td>
<td>ARM IC</td>
<td>DXXa</td>
</tr>
<tr>
<td>Transport/decrypt</td>
<td>Transport/decrypt</td>
<td>Link IC</td>
<td>Link IC</td>
<td>Link IC</td>
</tr>
<tr>
<td>QPSK demodulator</td>
<td>Convolutional decoder</td>
<td>MPEG A/V</td>
<td>MPEG A/V</td>
<td>DXXa</td>
</tr>
<tr>
<td>RS decoder</td>
<td>RS decoder</td>
<td>Video decompression</td>
<td>NTSC encoder</td>
<td>Telephone modem</td>
</tr>
<tr>
<td>Audio decompression</td>
<td></td>
<td></td>
<td>NTSC encoder</td>
<td>Telephone modem</td>
</tr>
<tr>
<td>Video DRAM</td>
<td></td>
<td></td>
<td>Telephone modem</td>
<td>Video DRAM</td>
</tr>
<tr>
<td>Telephone modem</td>
<td></td>
<td></td>
<td>Telephone modem</td>
<td>Video DRAM</td>
</tr>
<tr>
<td>Video DRAM</td>
<td></td>
<td></td>
<td>Telephone modem</td>
<td>Video DRAM</td>
</tr>
</tbody>
</table>

*aSingle chip

Source: Information courtesy of Thomson Consumer Electronics, Inc.

**BIBLIOGRAPHY**

and associated audio: parts 1-3, video, audio and systems MPEG1.
29. ISO/IEC 14496-x, Information technology—coding of audio-visual objects (MPEG4).
31. Skybridge L.L.C., Application for authority to launch and operate a global network of low earth orbit communications satellites providing broadband services in the fixed satellite service, file nos. 48-SAT-P/LA-97, 89-SAT-AMEND-97.
32. F. Biddle, DirecTV unit will offer HDTV programs in fall, Dow Jones & Co. wire service, Jan. 7, 1998.

FURTHER READING

For further reading and current information and general information about the major systems in the United States, please consult the World Wide Websites listed below:

DirecTV http://www.directv.com/
EchoStar http://www.dishnetwork.com/
Primestar http://www.primestar.com/
USSB http://www.ussb.com/

Information about evolving standards can be found at the following:

ATSC http://www.atsc.org/
DAVIC http://www.davic.org/
DVB http://www.dvb.org/

DIRECTION OF ARRIVAL ESTIMATION AND ADAPTIVE PROCESSING USING A CONFORMAL PHASED ARRAY

TAPAN K. SARKAR
Syracuse University
Syracuse, New York
MAGDALENA SALAZAR-PALMA
Madrid Polytechnic University
Madrid, Spain

1. INTRODUCTION

The phased arrays are used to sort out signals in space specifically when there are coherent multipaths of the
signal, as the information about the multipaths does not exist in the temporal domain. A more sophisticated form of phased arrays is an adaptive array which spatially filters out the signal of interest in the midst of clutter, jammer, and interference. The science of adaptive phased arrays, which started in the early 1950s, was implemented primarily through analog processing, as during that period of time the analog correlation processing was the only way. However, with the advent of digital signal processing we are currently using the same algorithms in the digital scenario without taking a critical look at whether such a methodology is really meaningful. For example, when using an analog adaptive algorithm, it is imperative that the adaptive weights can never be greater than unity as that would be equivalent to using an amplifier for the weights. Hence, the antenna array pattern for analog systems has a great physical significance. Therefore, with analog processing the array resolution is essentially limited by the size of the aperture and how closely one can resolve spatially spaced signals are dictated by the Rayleigh resolution criteria. So with the advent of digital signal processing one can go beyond the Rayleigh resolution and resolve signals within the width of the mainbeam of the array. Moreover, the adaptive weights can take any complex value numerically as we are now processing the signals digitally, and the purpose of the weights is equivalent to multiplying the voltages at the antenna feed points by some numbers. In such scenarios, the antenna array pattern looses to have any special significance at all.

The second shortcoming of the current phased-array methodology is that the mathematical description of the adaptive problem is set up as a detection problem rather than as an estimation problem. To illustrate this point, when we are dealing with radars, we are sending out a waveform, and the assumption is that we are trying to detect the same transmitted waveform delayed in time and attenuated, since the radar signal is not getting dispersed as it is propagating through free space. Hence, in such a situation we only want to know whether the transmitted radar signal reflected from the target exists. If the radar return exists then we know how far is the target from the time delay and from the Doppler shift, the velocity of the target. The optimum way to detect the existence of a particular waveshape we transmitted is to use a matched filter whose transfer function is the complex conjugate of the Fourier transform of the transmitted radar waveform. However, in a multi-path-rich mobile communication, detection is not the problem that we want to address, because we know that the signal is there; what we want to do is to estimate its correct amplitude, as that will correct for fading. The current mathematical approach of adaptive filtering is to use a Wiener filter–based model where a pilot signal is transmitted before every transmission and the channel along with the array is calibrated before the actual signal is sent. This leads to serious problems for real-time transmission as the assumption that the environments are identical when the pilot signal was sent and when the actual transmission is arriving may not be the same in a mobile environment. That is why a direct data domain least-squares approach based on a single snapshot of the voltages measured at the feedpoint of the antenna elements at a particular instance of time has been used to solve the estimation problem directly [1].

In this article we are going to extend the single-snapshot-based direct-data-domain approach to deal with a conformal adaptive phased arrays. In the current way of thinking one wants to use essentially an antenna element that has as close to an omnidirectional pattern as possible and then derive the gain by using hundreds and even thousands of them to get any significant gain from the array. Now one has to put a receiver channel at each of these antenna elements and that exponentially increases the cost. So, in order to minimize the cost, one then defines a subaperture and then sums up all the voltages in an analog fashion at the feedpoint of these antennas. Then either (1) a sum-and-difference beam pattern is formed to resolve the target or (2) digital beamforming using these summed up voltages is employed. This is not a sound procedure as it defeats the entire purpose of digital beamforming. It can easily be seen that this sub aperture summation negates the basic fundamental model of any adaptive signal processing algorithm! What is proposed in this section is to use directive elements on a conformal surface to perform adaptive processing. Use of directive elements will significantly reduce the number of antenna elements, and then, if one places these directive elements on a conformal surface to do adaptive processing, a significantly smaller number of antenna elements will be required, without sacrificing the gain. Such a methodology is illustrated here.

2. DOA ESTIMATION USING A CONFORMAL MICROSTRIP PATCH ARRAY ON THE SIDE OF AN AIRCRAFT

Estimation of direction of arrival (DoA) is one of the primary applications of phased arrays. One DoA estimation algorithm that deals with nonuniformly spaced antenna array elements is the MUSIC (mutlisignal classification) algorithm [1,2]. However, the performance of this signal processing technique deteriorates if the mutual coupling between the antenna elements and the effect of near-field scatterers are not accounted for in the analysis. Through this example we illustrate how to merge the electromagnetic analysis with the signal processing. In addition, we use a complex antenna element such as a microstrip patch antenna to illustrate how the antenna effects can be taken into account.

In this example, we estimate the DoA of the various signals impinging on a conformal microstrip patch array on the side of a Fokker aircraft. As seen in Fig. 1, the elements of the 11-element microstrip patch array are not placed uniformly, as the two groups of six and five elements are separated much more widely than those in a uniform interelement spacing, nor do they lie on a flat surface. A detailed view of the microstrip patch antenna element is also shown in Fig. 1. The half-wave rectangular microstrip patch element is fed by a probe and is situated on a high-dielectric-constant substrate so that it resonates at an operating frequency of 100 MHz. The thickness of the dielectric substrate is 0.012 and has an \( e_r = 32 \). There is a strong mutual coupling between the antenna elements, including the wings and fuselage of the
carried out by using a transformation matrix. Point radiators radiating in free space. This is numerically induced in ULVA consisting of omnidirectional isotropic radiators radiating in free space. The real 11-element array is interpolated into a virtual array consisting of 11 uniformly spaced omnidirectional point sources separated by 0.25λ at 100 MHz as shown in Fig. 1. We consider three signals arriving from 139°, 99°, and 60°, and impinging at the microstrip patch array from the side of the aircraft. The azimuthal angle is defined from the nose of the aircraft. The intensities and associated DoAs of the three signals to be simulated for this real case are summarized in Table 1. The simulation is done using the electromagnetic simulation tool WIPL-D as described in Ref. 1, we first transform the 11 voltages available at the terminals of the antennas to that induced in a ULVA as explained in Ref. 1. Then the matrix pencil approach [1,4,5] is applied to the processed voltages to estimate their DoAs and their complex amplitudes. The matrix pencil approach is also a direct data domain approach to estimate the DoAs of the signals and had been shown to have the least Cramer–Rao bound in the estimation of the poles as compared to various other contemporary methods. Since no covariance matrix is formed from the data, the matrix pencil approach can resolve coherent signals and is computationally very efficient as the poles are found in a direct fashion as the solution of a generalized eigenvalue problem. The simulation results from the two-step process is presented in Table 2. First, the electromagnetic analysis is carried out to compensate for the mutual coupling and near-field effects and then the matrix pencil method is applied to estimate the DoA of the three signals and their complex amplitudes. From Table 2 it is seen that not only the DoAs but also the complex amplitudes of the coherent signals have been recovered with engineering accuracy using a conformal phased array consisting of nonuniformly spaced microstrip patch elements.

### Table 1. Complex Amplitudes of all Signals and Their DoAs Incident on Aircraft

<table>
<thead>
<tr>
<th>Signal</th>
<th>Magnitude (V/m)</th>
<th>Phase (°)</th>
<th>DoA (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal 1</td>
<td>1</td>
<td>0</td>
<td>139</td>
</tr>
<tr>
<td>Signal 2</td>
<td>1</td>
<td>0</td>
<td>99</td>
</tr>
<tr>
<td>Signal 3</td>
<td>1</td>
<td>0</td>
<td>60</td>
</tr>
</tbody>
</table>

When operating on [A(φ)], produces numerically a modified manifold [A_ν(φ)], which is due to a ULVA manifold matrix. The elements of this virtual array are isotropic omnidirectional point radiators radiating in free space. Thus, we compensate not only for the lack of nonuniformity but also for the presence of mutual coupling between the elements in the real array in addition to near-field coupling effects between the elements of the array and the aircraft.

After we account for the strong electromagnetic coupling effects between the elements of the microstrip patch array and the aircraft at 100 MHz, such as the effects of nonuniformity in the spacing between the elements on a conformal nonplanar surface and the effects of mutual coupling between the elements of the patch array and their interactions with the wings and fuselage by using the approach described in Chap. 6 of Ref. 1, we first transform the 11 voltages available at the terminals of the antennas to that induced in a ULVA as explained in Ref. 1. Then the matrix pencil approach [1,4,5] is applied to the processed voltages to estimate their DoAs and their complex amplitudes. The matrix pencil approach is also a direct data domain approach to estimate the DoAs of the signals and had been shown to have the least Cramer–Rao bound in the estimation of the poles as compared to various other contemporary methods. Since no covariance matrix is formed from the data, the matrix pencil approach can resolve coherent signals and is computationally very efficient as the poles are found in a direct fashion as the solution of a generalized eigenvalue problem. The simulation results from the two-step process is presented in Table 2. First, the electromagnetic analysis is carried out to compensate for the mutual coupling and near-field effects and then the matrix pencil method is applied to estimate the DoA of the three signals and their complex amplitudes. From Table 2 it is seen that not only the DoAs but also the complex amplitudes of the coherent signals have been recovered with engineering accuracy using a conformal phased array consisting of nonuniformly spaced microstrip patch elements.

### Table 2. Estimated DoAs and Strengths of the Three Signals Incident on Phased Array

<table>
<thead>
<tr>
<th>Signal</th>
<th>Magnitude (V/m)</th>
<th>Phase (°)</th>
<th>DoA (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal 1</td>
<td>1.04</td>
<td>−0.05</td>
<td>139.71</td>
</tr>
<tr>
<td>Signal 2</td>
<td>0.91</td>
<td>0.06</td>
<td>101.86</td>
</tr>
<tr>
<td>Signal 3</td>
<td>1.08</td>
<td>−0.01</td>
<td>61.75</td>
</tr>
</tbody>
</table>

In this section, we illustrate how to employ the same interpolation technique of Chap. 6 in Ref. 1 to carry out adaptive processing using a single snapshot of the data using a nonuniformly spaced array in the presence of mutual coupling and near-field scatterers. In all the
examples, we are basically interested in three different kinds of antenna configurations. The three different arrangements we have used are as follows:

1. A nonuniformly spaced linear array (NLA) (as shown in Fig. 2)
2. A semicircular array (SCA) (as shown in Fig. 3)
3. A spatially sinusoidally modulated array (SSA) (as shown in Fig. 4)

In all the antenna configurations, as shown in Figs. 2–4, the antenna elements of the real array are considered to be half-wavelength-long thin-wire dipoles. Each element of the array is identically point loaded by 50 $\Omega$ at the center. The dipoles are all $z$-directed and hence are parallel to each other. They are of length $L = \lambda/2$ and radius $r = \lambda/200$.

The details of the antenna array elements are presented in Table 3.

As described in Ref. 1, all the real arrays operating in the three different configurations mentioned above are interpolated into a similar ULVA consisting of 17 uniformly spaced omnidirectional point sources separated by a distance $d/\lambda$. Typically $d$ is chosen to be close to $\lambda/2$. By choosing the reference point at the center of the real array in configuration 2 for the SCA and configuration 3 for the sinusoidally modulated array, the steering vectors associated with the virtual array is given by the following equation:

For configuration 1 (NLA = a nonuniformly spaced linear array), we choose the first element in the array as a reference point. Then the steering vectors for the ULVA are given by

$$\begin{eqnarray*}
[A_v(\phi)] &=& \begin{bmatrix}
1,
[e^{i(\pi/2z\Delta)/\lambda]} \cos \phi, 
[e^{i(3\pi/2z\Delta)/\lambda]} \cos \phi, & \ldots,
\end{bmatrix}^T 
\end{eqnarray*}$$

Here the number $(2k + 1)$ of elements of the virtual array is considered to be odd and equal to 17, and $\lambda$ is the wavelength of the signal located in the far-field region of the array, and the distance $\Delta$ between the elements in the virtual array is 0.4775$\lambda$. Here we want to correct for all the electromagnetic effects in the azimuth sector of $\left[\phi_q, \phi_{qq}\right] = [30, 150^\circ]$. The incremental angle $\phi$ for the computation of the transformation matrix is chosen to be $1^\circ$. The sector chosen here is of width $120^\circ$ symmetrically located around the broadside. Then, a set of voltages induced at the antenna elements are measured and computed for far-field sources located at each of the angles $\phi_q, \phi_q + \phi, \phi_q + 2\phi, \ldots, \phi_{qq}$. The measured/computed vector $A(\phi)$ is then distorted from the ideal steering vector defined by (1), due to the presence of mutual coupling between the elements of

Table 3. Physical Sizes for Dipole Antenna Elements in all Three Antenna Arrays

| Number of elements in the three arrays | 24 |
| Length of $z$-directed wires | $\lambda/2$ |
| Radius of wires | $\lambda/200$ |
| Loading at center of element | 50 $\Omega$ |

Figure 2. Geometry of a nonuniformly spaced linear array (NLA) and a ULVA (shown by dots) electromagnetically representing the NLA.

Figure 3. Geometry of a semicircular array (SCA) and a ULVA (shown by dots) electromagnetically representing the SCA.

Figure 4. Geometry of a sinusoidally spaced array (SSA) and a ULVA (shown by dots) electromagnetically representing the SSA.
the real array. The actual steering vectors having all the undesired electromagnetic effects are computed using the electromagnetic analysis code WIPL-D [3]. We compute the transformation matrix to compensate for the effects of non-uniformity in spacing and the presence of mutual coupling between the elements of the real array. The induced voltages at the elements of the antenna array result in a compensated set of input voltages in which the nonuniformity in spacing and mutual coupling effects are eliminated from the actual measured/computed voltages. We then apply the direct-data-domain least-squares approach described in Ref. 1 to estimate the complex amplitude of the signal of interest (SoI) given its DoA. Next we illustrate the performance of these three complex arrays through numerical electromagnetic simulations.

3.1. Constant Signal

First, we consider a case where we have two jammers arriving at the array along with the SoI. The interference-to-signal ratio for the jammers impinging on the array from an angle of 60° is varied from 0 to 54 dB. The other jammer arrives at the array at an angle of 70° with respect to the x axis. The SoI is incident at an angle of 80° at a frequency of 300 MHz. The various electrical parameters of the signals are summarized in Table 4. The received signal-to-thermal-noise ratio (SNR) at the antenna elements is set at 20 dB in this example. For this input, the signal strength is estimated while rejecting jammers. This estimation is applied to all the three antenna configurations.

The output signal-to-interference-plus-noise ratio (SINR) is an indicator of the accuracy of our estimate. It is defined as

\[
\text{SINR}_{\text{out}} = 20 \log \left| \frac{z}{z_{\text{est}}} \right| \quad (2)
\]

where \(z\) is the amplitude of the desired signal and \(z_{\text{est}}\) is the estimate of the amplitude of the reconstructed signal.

The results are shown in Figs. 5 and 6 for all the three configurations. The thermal noise is ignored in Fig. 5, while the signal-to-noise ratio is set at 20 dB in Fig. 6. The noise is additive and is modeled as a Gaussian random variable. The x axis of the graph corresponds to the second jammer-to-SoI power level (as defined in Table 4), while the y axis corresponds to the output signal-to-interference-plus-noise ratio, as defined in (2). It can be observed in these figures that the interpolation technique using measured/computed steering vector shows proper compensation for all the electromagnetic effects of the real arrays to estimate the magnitude of the SoI based on the direct-data-domain approach.

Table 4. Complex Amplitudes of Signals and Their DoAs

<table>
<thead>
<tr>
<th></th>
<th>Magnitude (V/m)</th>
<th>Phase</th>
<th>DoA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal</td>
<td>1.0</td>
<td>0.0°</td>
<td>80°</td>
</tr>
<tr>
<td>Jammer 1</td>
<td>1.0</td>
<td>0.0°</td>
<td>70°</td>
</tr>
<tr>
<td>Jammer 2</td>
<td>1.0–500.0</td>
<td>0.0°</td>
<td>60°</td>
</tr>
</tbody>
</table>

Figure 5. Output SINR as a function of interference-to-SoI ratio in decibels (dB) without noise.

Figure 6. Output SINR as a function of interference-to-SoI ratio in dB with additive thermal noise.
and a spatially sinusoidally modulated array in estimating the DoA when the interferer is in the mainlobe is better than that for a nonuniform linearly spaced array.

3.3. Effects of Blockage Produced by Near-Field Scatterers

Next, we consider the effects of a near-field scatterer located near each array configuration. The near-field scatterer is located within a distance that is twice the radius of the array and is located along the direction of 60° in such a way that this near-field scatterer interferes with the direct line of sight of the SoI. The length and width of this square cylindrical near-field scatterer are 1.26 wavelengths, and the height of the scatter is 3.28 wavelengths. The desired signal and the jammers are as summarized in Table 5. The results are shown in Fig. 8 for the three different array configurations. It can be observed in these figures that the output signal-to-interference-plus-noise ratio is much lower than where there was no blockage of the SoI, due to a near-field scatterer. However, one can still obtain a proper estimate for the complex amplitude of the SoI after compensating for the various unwanted effects even when the jammer is in the mainlobe and operating in the presence of a large near-field scatterer.

4. DOA ESTIMATION USING A PHASED ARRAY LOCATED ON A CONFORMAL HEMISPHERICAL SURFACE

In this section, we describe how to carry out DoA estimation using directive antenna elements over a conformal surface like a hemisphere. We consider the use of three directive antenna elements, such as a short-circuited dual-patch antenna, a dielectric resonator antenna, and a horn antenna as elements in a conformal array. The surface selected for the implementation is a hemisphere. The procedure is to carry out the analysis to obtain the steering vectors over a conical scan subtended by the azimuth angles $[\varphi_q, \varphi_{q+1}]$ and from elevation angles spanning $[\theta_q, \theta_{q+1}]$. This is illustrated by Fig. 9. In the examples to follow, we transform the real array where the antenna elements are located over a hemispherical surface into a two-dimensional uniform planar linear virtual array (2D ULVA) lying in the x–y plane containing the projection of the hemisphere, as shown in Fig. 10. The 2D ULVA have been formulated in the form of a cross, a L-shaped, or a two-dimensional grid configuration. Here, in this methodology, the one-dimensional transformation matrix, for azimuth angles is extended to two dimensions to handle both elevation and azimuth angles [1].

### Table 5. Complex Amplitudes of Various Signals and Their DoAs

<table>
<thead>
<tr>
<th></th>
<th>Magnitude (V/m)</th>
<th>Phase</th>
<th>DoA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal</td>
<td>1.0</td>
<td>0.0°</td>
<td>100°</td>
</tr>
<tr>
<td>Jammer</td>
<td>1.0</td>
<td>0.0°</td>
<td>101–110°</td>
</tr>
</tbody>
</table>

### Figure 7. Output SINR as a function of separation between SoI and jammer.

### Figure 8. Output SINR as a function of separation between SoI and jammer with blockage.

### Figure 9. Parameters of the sector in which the signals of interest lie.
this 2D transformation matrix, we map the voltages that are induced at the feedpoint of these real antenna elements operating in the presence of mutual coupling and other near-field scatterers to a planar 2D ULVA.

4.1. Short-Circuited Dual-Patch Antenna on a Hemispherical Surface

The first structure to be considered in this technique is the application of a short-circuited dual-patch antenna (SDPA) to form an array over a hemispherical surface. An antenna element representing a SDPA is shown in Fig. 11, and the specific dimensions are given in Fig. 12 \cite{6}. A SDPA consists of two layers, in which the upper patch, which is of trapezoidal shape, is connected along one edge to the ground by a vertical metal wall and the other edge is connected to the lower patch by another vertical wall. These novel features produce a significant reduction in the resonant frequency of the antenna. The numerical results obtained by using a dynamic electromagnetic simulator such as WIPL-D \cite{3} are very similar to the experimental results described in Ref. 6. The actual geometry of the SDPA is shown in Fig. 12, and the input impedance for this geometry as a function of frequency is given in Fig. 13. For this structure the resonant frequency occurs at about 2.41 GHz, as shown in Fig. 13. Finally, the radiation pattern of the SDPA along the E plane is shown in Fig. 14, illustrating that the element pattern has some directivity and the frontlobe is larger than the backlobe. These results are close to the measured ones given in Ref. 6. The SDPA is considered to be fabricated over a finite square ground plane of size \(G\) centimeters on each side. We now consider a 48-element array consisting of an SDPA that is distributed over a hemisphere of radius 1.4\(\lambda\). The elements are placed in a star configuration as seen in Fig. 15. We now transform the actual voltages that are induced at the feedpoint of these antenna elements located on the surface of a hemisphere to a 16-element planar 2D ULVA located on the \(x\)-\(y\) plane with \(z = 0\). The 2D ULVA consists of omnidirectional point antennas radiating in free space as shown in Fig. 16. The spacing between the virtual elements is 0.5\(\lambda\) at 2.41 GHz, and all the omnidirectional point sources are placed as a 4 \(	imes\) 4 array located within the base of the hemisphere as seen in Fig. 16.

For all the examples that we will present, the azimuth scan angle \(\phi\) varies from \(\phi_q\) to \(\phi_{qq}\) with an angular step of 1° (i.e., \(\phi = 1\)). The step size for the elevation angle is also

\[\phi = 1\]
1° (i.e., \( \theta = 1° \)). This information is now used to calculate the transformation matrix \([\mathcal{A}]\).

We next define a transformation error to check the accuracy of this interpolation in both \( \varphi \) and \( \theta \). The error is defined by

\[
\text{Interpolation error} = \sqrt{\frac{1}{IJ} \sum_{i=1}^{I} \sum_{j=1}^{J} \left| [A_i(\varphi_q; \Theta_q)]_{ij} - \mathcal{A}_i[\mathcal{A}(\varphi_q; \Theta_q)]_{ij} \right|^2}
\]

The interpolation error and the condition number of the transformation matrix for different values of the sectors \( q \) are given in Table 6. The condition number is defined as the ratio of the largest singular value to the smallest singular value of the transformation matrix. It is an indicator of how sensitive is the solution to changes of the physical dimensions of the antenna array. As observed from Table 6, the error in the transformation to a ULVA remains more or less constant until \( 0° \leq \theta \leq 50° \). The azimuth scan is the entire 360°. We do not present results beyond \( \theta \leq 50° \) because all the SDPA elements on the hemisphere do not see the signal, and hence the transformation matrix \([\mathcal{A}]\) becomes undefined. In that case we restrict the azimuth angle scan, as will be seen later on in this section. Therefore this procedure works if the interpolation error is reasonably small. This can be determined apriori for the structures of interest.

![Figure 14. Radiation pattern of the short-circuited dual-patch antenna; normalized; E plane.](image)

![Figure 15. Short-circuited dual-patch antenna array.](image)

![Figure 16. Location of the 16-element virtual array when the scanned sector is defined by the azimuth angle \((\varphi_q, \varphi_q) = (30°, 150°)\) and elevation angle \((\theta_q, \theta_q) = (0°, 40°)\).](image)

<table>
<thead>
<tr>
<th>Elevation Scan</th>
<th>Interpolation Error (%)</th>
<th>Condition Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>0° (\rightarrow) 10°</td>
<td>0.2</td>
<td>(2.7 \times 10^8)</td>
</tr>
<tr>
<td>0° (\rightarrow) 20°</td>
<td>0.36</td>
<td>(5 \times 10^5)</td>
</tr>
<tr>
<td>0° (\rightarrow) 30°</td>
<td>2.4</td>
<td>(8 \times 10^5)</td>
</tr>
<tr>
<td>0° (\rightarrow) 40°</td>
<td>5.9</td>
<td>17,191</td>
</tr>
<tr>
<td>0° (\rightarrow) 50°</td>
<td>10.2</td>
<td>15,368</td>
</tr>
</tbody>
</table>
Here the real elements are short-circuited dual-patch antennas placed over the hemispherical array in a star configuration. The estimated DoAs of all the three signals given by the matrix pencil method using the voltages induced in the 2D ULVA are given in Table 7. The estimated amplitudes of all three signals are also shown in Table 7.

4.2. Dielectric Resonator Antennas on a Hemispherical Surface

The next example deals with an antenna array consisting of rectangular dielectric resonator (RDRs) antennas as antenna elements located on a hemispherical surface. The reason that the dielectric resonators are chosen as antenna elements is because they have a high radiation efficiency, typically greater than 98%.

An antenna element representing a RDR is shown in Fig. 18, where the specific dimensions of the RDR are as given in Fig. 19 [7]. The coaxial probe extends into a dielectric of relative permittivity of 8.9. The antenna element is situated over a square finite ground plane of dimension \( G = 6.0 \text{ cm} \) along each side. The input impedance as a function of frequency is shown in Fig. 20 and was obtained using WIPL-D [3]. The results are similar to those in Ref. 7. The actual resonant frequency of this RDR occurs at 3.07 GHz, as in Fig. 20. Finally, the radiation pattern along the \( E \) plane is shown in Fig. 21.

A total of 48 RDR elements are placed on a hemisphere of radius 2.1\( l \), as shown in Fig. 22, in order to create an RDR hemispherical array (RDR HA). We now transform the voltages induced in these real antenna elements to those that will be induced by the same incident fields if they were to illuminate the ULVA. The 18 antenna elements that are used in this transformation are represented by the black squares in Fig. 23. The sector over which the transformation is carried out is covered by the azimuth angle of \((\phi_q, \phi_{qq}) = (30^\circ, 150^\circ)\), and the elevation scan angle varies from 10\(^\circ\) to 50\(^\circ\). The actual voltages induced in the 18 real elements are mapped to the voltages induced that will be induced in the 16-element 2D ULVA, as shown in Fig. 16. For the ULVA the elements are point radiators radiating in free space. These virtual elements are separated by 0.5\( l \). The results of the accuracy and stability of this transformation are given in Table 8. To carry out an estimation of DoA for three signals impinging on this 48-RDR-element array, we transform the voltages induced in the real 18-element array (as shown by the black squares in Fig. 23) to a 16-element ULVA. The goal here is to estimate the directions of arrival and the amplitudes of three signals that are incident on the array from \((\varphi, \theta)_1 = (50^\circ, 5^\circ)\), \((\varphi, \theta)_2 = (80^\circ, 20^\circ)\), and \((\varphi, \theta)_3 = (140^\circ, 30^\circ)\). The amplitude of all the signals is 1 V with a zero phase angle. The scan to generate the transformation matrix is carried

---

**Table 7. Estimation of DoA of Three Signals: \((\varphi, \theta)_1 = (50^\circ, 5^\circ), (\varphi, \theta)_2 = (80^\circ, 20^\circ), (\varphi, \theta)_3 = (140^\circ, 30^\circ)\)**

<table>
<thead>
<tr>
<th>( \varphi )</th>
<th>( \theta )</th>
<th>Estimation of ( \varphi )</th>
<th>Estimation of ( \theta )</th>
<th>Estimation of Amplitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>50(^\circ)</td>
<td>5(^\circ)</td>
<td>51.91(^\circ)</td>
<td>5.83(^\circ)</td>
<td>1.03</td>
</tr>
<tr>
<td>80(^\circ)</td>
<td>20(^\circ)</td>
<td>78.80(^\circ)</td>
<td>18.84(^\circ)</td>
<td>0.87</td>
</tr>
<tr>
<td>140(^\circ)</td>
<td>30(^\circ)</td>
<td>141.38(^\circ)</td>
<td>30.51(^\circ)</td>
<td>0.93</td>
</tr>
</tbody>
</table>
out over the azimuth angles of \((\varphi,\theta) = (30^\circ,150^\circ)\) and the elevation angles of \((\theta_q,\theta_{qq}) = (0^\circ,40^\circ)\).

The estimated DoAs of all the three signals given by the matrix pencil method using the voltages induced in the 2D ULVA are given in Table 9. The estimated amplitudes of all three signals are also shown in Table 9.

### 4.3. An Array of Horn Antennas on a Hemispherical Surface

Finally, we consider an array of horn antennas located on a hemispherical surface. The probe-fed horn antenna is shown in Fig. 24, and its dimensions are shown in Fig. 25. The input impedance at the probe feed is shown in Fig. 26, as a function of frequency demonstrating that the resonant frequency of the horn occurs at 2.41 GHz. At that frequency the element radiation pattern is given by Fig. 27.

A nonplanar conformal array on a hemispherical surface is formed by using 48 horn antennas placed on a
hemispherical surface of radius 2.4\text{m} as shown in Fig. 28. The properties of the transformation matrix when we transform the voltages induced in the actual array consisting of 18 elements (see Fig. 28) to the voltages that will be induced in a 16-element planar ULVA (see Fig. 16) are given in Table 10. The sector scan in azimuth and in elevation is limited by \((\varphi_q,\varphi_{qq}) = (30^\circ, 150^\circ)\) and \((\theta_q,\theta_{qq}) = (0^\circ, 40^\circ)\), respectively. The interpolation error and the condition number of the transformation matrix \([3_q]\) for different sectors \(q\) are shown in Table 10.

To carry out DoA estimation of the three signals \((\varphi, \theta)_1 = (50^\circ, 5^\circ)\), \((\varphi, \theta)_2 = (80^\circ, 20^\circ)\), and \((\varphi, \theta)_3 = (140^\circ, 30^\circ)\) impinging on the hemispherical horn array, we transform the voltages that are induced on the shaded 18-horn antenna elements as shown in Fig. 17 to the 2D ULVA of Fig. 16. The amplitude of all the signals is 1 V.

The estimated DoAs of all three signals given by the matrix pencil method using the voltages induced in the 2D ULVA are given in Table 11. The estimated amplitudes of all three signals are also shown in Table 11. The estimates using a single snapshot of the data voltages are quite accurate. It is important to note that the error in the transformation matrix increases as we increase the azimuth angle \(\vartheta\) toward 90\(^\circ\) because the increase in degrees of freedom is not commensurate with the number of antenna elements. The error can be controlled either by adding additional arrays or by reducing the scan angle for the group of active elements.

It is seen that it is possible to carry out a DoA estimation by using directive antenna elements located on a hemispherical surface. The 2D matrix pencil method is applied to the outputs from the 2D ULVA to estimate the DoAs of the various signals illustrates that the results can be obtained with engineering accuracy. The use of directive elements is made possible through this methodology, which increases the efficiency of the system, as we now have a larger signal-to-noise ratio to operate in. In addition, such a
system significantly reduces the cost of a phased-array system, as there are fewer antenna elements and hence fewer receiver channels and therefore fewer number of A/Ds are required to carry out digital beamforming.

5. CONCLUSION

This article has presented a preprocessing technique that transforms a nonuniformly spaced array operating in the presence of mutual coupling between the elements of the array and near-field scatterers into a virtual array of omnidirectional isotropic point elements operating in free space that is suitable to the application of a direct data domain algorithm. Through such a transformation formulated using an interpolation technique, we have shown that one can not only compensate for the effects of mutual coupling in a nonuniformly spaced array but also eliminate the effects of strong near-field scatterers. Since the transformed output voltages are those of a uniformly spaced linear array consisting of omnidirectional point radiators, a conventional adaptive algorithm can easily be applied to extract the SoI in the presence of jammers. Finally, it is shown how to use directive antenna elements located on a conformal hemispherical surface is used to perform DoA estimation using a single snapshot of the data.

BIBLIOGRAPHY


DIRECTIONAL COUPLERS

The directional coupler (DC) is the most useful four-port microwave device. Substantially, it consists of a four-port device, where a wave entering any port excites two other ports with prescribed amplitudes, whereas the remaining port is isolated. This property makes the DC an essential component when dealing with microwave devices, as it allows to distinguish reflected waves from incident ones. After a definition of the ideal DC, the typical parameters and the electrical characteristics of actual devices are illustrated, then, the main applications are considered and, finally, some common realizations of DC are shown, particularly focusing the attention on different coupling mechanisms and technologies.

1. DEFINITION

An ideal directional coupler is a linear, reciprocal, and lossless four-port device, often indicated as in Fig. 1, with the following properties [1]:

1. The ports are matched.
2. Each port is only coupled to other two ports, the fourth one being isolated.

![Figure 28. An array of conformal horn antenna aligned on a hemispherical surface.](image-url)
tracting one from the other, one obtains the equation

\[ \begin{bmatrix} 0 & s_{12} & s_{13} & 0 \\ s_{12} & 0 & 0 & s_{24} \\ s_{13} & 0 & 0 & s_{24} \\ 0 & s_{24} & s_{34} & 0 \end{bmatrix} \]

Therefore, the scattering matrix of an ideal directional coupler takes the form

\[ S = \begin{bmatrix} 0 & s_{12} & s_{13} & s_{14} \\ s_{12} & 0 & s_{23} & s_{24} \\ s_{13} & s_{23} & 0 & s_{24} \\ s_{14} & s_{24} & s_{34} & 0 \end{bmatrix} \]

apart from an insignificant permutation of the port indices. Thus, a wave incident at port 1, not being reflected at all, splits between ports 2 and 3, while port 4 is isolated. Conversely, a wave incident at port 2 is coupled to ports 1 and 4, with port 3 isolated under that excitation (hence the name directional coupler). The fundamental parameter characterizing the ideal directional coupler is the so-called coupling \( C \), defined as the reciprocal of the magnitude of the transmission coefficient between ports 1 and 3 (i.e., \( C = -20 \log |s_{13}| \)).

It is also shown that any linear, reciprocal, lossless, and matched four-port must be a directional coupler. Its scattering matrix \( S \) is in fact

\[ S = \begin{bmatrix} 0 & s_{12} & s_{13} & s_{14} \\ s_{12} & 0 & s_{23} & s_{24} \\ s_{13} & s_{23} & 0 & s_{24} \\ s_{14} & s_{24} & s_{34} & 0 \end{bmatrix} \]

Since losslessness implies the unitarity of the scattering matrix, \( SS^T = I \), where \( I \) is the unit matrix, the following equations must be satisfied:

\[ |s_{12}|^2 + |s_{13}|^2 + |s_{14}|^2 = 1 \]

(3)

\[ |s_{12}|^2 + |s_{23}|^2 + |s_{24}|^2 = 1 \]

(4)

\[ |s_{13}|^2 + |s_{23}|^2 + |s_{34}|^2 = 1 \]

(5)

\[ |s_{14}|^2 + |s_{24}|^2 + |s_{34}|^2 = 1 \]

(6)

\[ s_{13}s_{23}^* + s_{14}s_{24}^* = 0 \]

(7)

\[ s_{12}s_{23}^* + s_{14}s_{24}^* = 0 \]

(8)

\[ s_{12}s_{24}^* + s_{13}s_{34}^* = 0 \]

(9)

By multiplying Eq. (7) by \( s_{12} \) and Eq. (8) by \( s_{13} \) and subtracting one from the other, one obtains the equation

\[ s_{14}(s_{13}s_{24}^* - s_{13}s_{34}^*) = 0 \]

(10)

whose solutions are

\[ s_{14} = 0 \]

(11)

\[ s_{12}s_{24}^* = s_{13}s_{34}^* \]

(12)

Let us consider the first solution. From Eq. (7) it also follows that \( s_{23} = 0 \) [otherwise we would have \( s_{12} = s_{13} = 0 \), in contradiction with Eq. (3)]. Moreover, by subtracting Eq. (4) from Eq. (3) and Eq. (5) from Eq. (3), one obtains

\[ |s_{13}| = |s_{24}| = \alpha \quad \text{and} \quad |s_{12}| = |s_{34}| = \beta. \]

By setting \( \phi_{ij} = \angle S_{ij} \), from Eq. (9) one obtains the relationship linking the phases:

\[ (\phi_{12} - \phi_{13}) + (\phi_{34} - \phi_{24}) = \pi \]

(13)

Therefore, the scattering matrix of the device takes the form

\[ S = \begin{bmatrix} 0 & \beta e^{i\phi_{12}} & \alpha e^{i\phi_{13}} & 0 \\ \beta e^{i\phi_{12}} & 0 & 0 & \alpha e^{i\phi_{23}} \\ \alpha e^{i\phi_{13}} & 0 & 0 & \beta e^{i\phi_{34}} \\ 0 & \alpha e^{i\phi_{23}} & \beta e^{i\phi_{34}} & 0 \end{bmatrix} \]

(14)

Note that when \( \phi_{12} - \phi_{13} = \phi_{34} - \phi_{24} \), as often occurs in practical cases when the coupler is symmetric \( (s_{13} = s_{24} \) and \( s_{12} = s_{34} \), we obtain

\[ \phi_{12} - \phi_{13} = \pi/2 \]

(15)

Therefore, the outputs from ports 2 and 3 are in quadrature. In this case, it is always possible to choose the reference planes in such a way that the scattering matrix takes the form

\[ S = \begin{bmatrix} 0 & j\beta & \alpha & 0 \\ j\beta & 0 & 0 & \alpha \\ \alpha & 0 & 0 & j\beta \\ 0 & \alpha & j\beta & 0 \end{bmatrix} \]

(16)

When we consider Eq. (12), we note immediately that the solutions \( s_{12} = 0 \) and \( s_{34} = 0 \), or \( s_{13} = 0 \) and \( s_{24} = 0 \), respectively, are the same as that obtained previously, after exchanging the coupled ports with the isolated ones. For instance, under excitation of port 1, the isolated port becomes port 2 while port 3 and port 4 are coupled. Therefore, the first case examined can be taken as typical.

An alternative picture representing a directional coupler commonly used in measurement benches is shown in Fig. 2. The figure emphasizes that port 4 of the DC is terminated on a matched load. Actually, in many applications, port 4 is not used at all. Nevertheless, this port must be loaded in order to suppress the secondary line signals due to its mismatch. Moreover, the sketch suggests an intuitive idea of the realization of a directional coupler: two parallel transmission lines electromagnetically coupled to each other. A portion of the wave traveling from port 1 toward port 2 couples to port 3. According to common
parlance, the branch connecting ports 1 and 2 is called the “main arm,” while the one linking ports 3 and 4 is called the “secondary arm.”

2. REAL DIRECTIONAL COUPLERS

Of course, it is impossible to obtain both perfect match at the four ports and full isolation of the uncoupled ports [2,3]. Therefore, in order to characterize an actual coupler, it is necessary to define, in addition to the coupling, the directivity

\[
D = 20 \log \frac{|s_{13}|}{|s_{14}|} \tag{17}
\]

which represents the ratio between the power flowing from 1 to 3 and that from 1 to 4. In the ideal case the directivity is infinite. Sometimes it is preferred to use another parameter, the isolation \(I\), defined by

\[
I = 20 \log \frac{1}{|s_{14}|} = C + D \tag{18}
\]

Additionally, actual devices are typically characterized by the following quantities:

- **Frequency range**—operational band of the coupler. Typically, commercial waveguide couplers operate over the whole waveguide band. Over that band the following parameters are defined:
- **The nominal coupling \(C\)**—typical values are 3, 6, 10, 20, 30, and 40 dB.
- **Coupling sensitivity/deviation**—the maximum deviation \(C\) with respect to its nominal value.
- **Minimum directivity \(D\)**—typically between 30 and 50 dB.
- **Insertion loss**—the maximum insertion loss on the main path (ports 1–2).
- **Primary-arm VSWR**—the VSWR on the main path.
- **Secondary-arm VSWR**—that of the secondary path.
- **Power-handling capability (cw)/peak**—the maximum continuous or peak power that can be carried by the coupler.
- **Connectors**—indicate the connectors by which the coupler is fed.

3. MEASUREMENT OF THE COUPLER PARAMETERS

Since a directional coupler is a linear four-port device, its characteristic parameters \(C\) and \(D\) are commonly deduced by measuring the scattering parameters with the help of a network analyzer. This is very easy in coaxial and rectangular waveguides, where analyzer flanges are the same as those of the device under test. In microstrip and in planar circuits, some attention must be paid to the feed transitions. Such transitions could noticeably alter the response of the coupler. Figure 3 shows the typical measured parameters of a commercial 10-dB directional coupler in a rectangular waveguide.

4. APPLICATIONS

The main feature of the coupler is its ability to detect whether a wave traveling along the main branch is propagating from 1 to 2 or in the opposite direction. That makes the coupler an essential component in telecommunication and measurement systems. A few of the most important applications are discussed next.

4.1. Reflectometer

From the detection of the propagation direction of a wave follows the possibility of measuring the reflection of an unknown device. This is achieved by arranging the measurement bench as depicted in Fig. 4.

In the ideal case [i.e., when the scattering matrix of the directional coupler is given by Eq. (1), and assuming port 4 to be perfectly matched], the signal detected at port 3, \(b_3\), is proportional to \(\rho_1\), the reflectivity of the device under test (DUT) loading port 1, while port 2 is fed by a microwave generator (4). In fact

\[
b_3 = s_{13}s_{12}\rho_1 \tag{19}\]
The measurement requires knowledge of the term $s_{13}s_{12}$, which can be easily measured by substituting the unknown load $r_1$ with a known one, typically a totally reflecting load, as a short circuit ($r_k = -1$). Under such a condition, the signal delivered at port 3 is

$$b_{3k} = -s_{13}s_{12} = -s_{13}s_{12} \quad (20)$$

Once this step, commonly referred to as the calibration, has been completed, one can find the reflection $\rho_1$ as

$$\rho_1 = -b_3/b_{3k} \quad (21)$$

### 4.2. Monitoring, Feedback, and Power Measurements

In the configuration of Fig. 5 the coupler is used for the purpose of monitoring the output level of a given source. The signal at port 3 can also be used as a feedback for the source itself—for instance, when a leveled output power is required (see Feedback Amplifiers).

The use of coupler with large coupling is very useful when dealing with measurements of high power signals, which could destroy power meters. The typical bench for such measurements is shown in Fig. 6.

### 4.3. Couplers as Part of More Complex Systems

The coupler often represents an essential part of more complex devices, such as the mixer or the four-way circulator, shown in Figs. 7 and 8, respectively. In the first case the coupler provides a strong separation between the local oscillator (LO), while in the second case two 3 dB couplers, connected as shown in Fig. 8, permit one to obtain (in the ideal case) the following scattering matrix:

$$S = \begin{bmatrix} 0 & 0 & 0 & 0 & e^{j\phi_{14}} \\ 0 & 0 & e^{j\phi_{41}} & 0 & 0 \\ 0 & e^{j\phi_{14}} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (22)$$

When port 1 is fed, the power equally splits between ports 2 and 3, while port 4 is isolated. Then, thanks to the phase shifter, the signal traveling from port 3 to port 3' is shifted by 180° with respect to the signal directly arriving at port 2'. The output of the combination of the two signals 180° out of phase is different from zero only at port 4'. Therefore, $s_{14} = 0$, $s_{14} \neq 0$ and $s_{11} = 0$. Analogously, when port 4 is fed, ports 1 and 4' are isolated, while port 1' is coupled. It is easy to repeat the same reasoning when ports 1' and 4' are fed and to recover the scattering matrix of Eq. (22).

### 5. COUPLING MECHANISMS

The two lengths of transmission line forming the coupler can be coupled to each other by different mechanisms. Schematically, a first distinction can be made between lumped and distributed coupling [5].

Considering that at microwave frequencies purely lumped coupling does not exist, since propagation effects occur, we will consider the coupling as lumped if the region where the coupling physically takes place is much shorter
than the wavelength. Lumped coupling can be achieved in both waveguide and planar technology.

In the first case, a further distinction can be made as to whether the coupling is directive. Coupling may be made directive by shaping the coupling region in such a way that on the secondary guide it produces two waves that are in phase opposition in one direction. The double aperture invented by Saad and Riblet \[6\] (Fig. 9a) provides a clear example of such a mechanism. In fact, in the secondary guide the thin vertical slot produces two waves 180° out of phase, propagating in opposite direction, while the thin horizontal slot produces two waves in phase. It is therefore possible to choose the dimensions and the positions of the two slots so that the amplitudes of the waves that are separately excited are almost the same. Thus, in one direction the two waves sum in phase, while in the opposite direction they cancel. Note that the bandwidth of this coupler is wide enough since the aforementioned mechanism does not take advantage of resonance effects. The same mechanism can be exploited by collapsing the two thin apertures into an elliptical one and adjusting eccentricity

Figure 9. Riblet and Saad directional aperture. In the second arm, the magnetic current \(M_v\) excites two waves whose amplitudes are 180° out of phase and propagating in opposite directions. On the contrary, the waves excited by the magnetic current \(M_h\) are in phase. Hence, it is possible to adjust the dimensions of the apertures in such a way that the two scattered waves cancel in one direction and sum in the opposite over a broad band.

Figure 8. A four-way circulator is supposed to feed one port at a time. For instance port 1: In that case port 4 is uncoupled and the signal is split equally among the remaining ports of the first junction. At the second junction, the two signals combine 180° out of phase because of the phase shifter inserted in the lower pattern. Therefore, due to the symmetry, they can couple only to port 4', while port 1' is isolated. The same reasoning applies when port 4' is being fed. In that case, however, the phase shifter does not operate as the signal flows counterclockwise in the lower pattern. Therefore, the signals arriving at the first coupler are in phase and combine only to port 1. Analogous reasoning holds for ports 1' and 4.

Therefore, due to the symmetry, they can couple only to port 4. Analogous reasoning applies when port 4 is being fed. In that case, however, the phase shifter does not operate as the signal flows counterclockwise in the lower pattern. Therefore, the signals arriving at the first coupler are in phase and combine only to port 1. Analogous reasoning holds for ports 1' and 4.

and the position according to the preceding criteria \[7,8\], as shown in Fig. 9b.

Of course, coupler performance can be strongly improved by cascading several properly spaced apertures. To understand the principle of operation, let us consider two lengths of rectangular waveguides positioned side by side and coupled via two apertures drilled in the broad wall, as illustrated in Fig. 10. We want to couple them in such a way that a fraction of the power traveling from port 1 to port 2 is delivered to port 3 while port 4 is isolated. On the contrary, a wave traveling from 2 to 1 must couple only to port 4. If \(C_l\) and \(D_k\) denote the fraction of signal delivered to ports 3 and 4, respectively, when a wave travels from port 1 to 2, then the amplitude of the wave at port 3 is given by

\[
A_3 \approx |(C_1 + C_2)|
\]  

while at port 4 we have

\[
A_4 \approx |(D_1 + D_2 e^{-j2\beta d})|
\]  

Therefore, if \(D_1 \approx D_2\), \(C_1 \approx C_2\), and \(2\beta d = \pi\) (that is, the apertures are equal and spaced by \(\lambda_g/4\)), \(A_4 = 0\). Hence, the structure shows, at least at one frequency, the characteristics we are looking for. It is apparent that the design of a directional coupler providing given performance over almost the whole waveguide band is quantitatively much more complicated and requires the use of an appropriately dimensioned array of apertures, as schematically sketched in Fig. 11.

Under the hypothesis that the power coupled by a single aperture to the secondary waveguide is small, let us assume \(C_n\) and \(D_n\) to be the coupling coefficients of the \(n\)th aperture in the forward and reverse directions. Let us also suppose to separate the apertures by a distance \(d\). Hence, the whole coupling into port 3 (\(B_3\)), computed in correspondence of the last aperture, is

\[
B_3 = A e^{-j\beta d} \sum_{n=0}^{N} C_n e^{-j\beta (N-n)d}
\]
Chebyshev polynomial of order $N$, $T_N$:

$$F = \left| \sum_{n=0}^{N} r_n^3 e^{-j2\beta d n} \right|^2$$

$$= K |T_N(\sec \theta_m \cos \theta)| \text{ where } \theta = \beta d$$

The midband frequency is $\theta = \pi/2$ and corresponds to $d = l_0/4$, and $\theta_m$ is the value of $\beta d$ at the band edges. The positive constant $K$ is chosen to obtain the desired coupling at the midband frequency:

$$C = -20 \log K |T_T||T_N(\sec \theta_m)|$$

When $\theta = 0$, $F = |\sum_{n=0}^{N} r_n^3| = |T_N(\sec \theta_m)|$. Therefore, the directivity is given by the formulas

$$D = 20 \left[ \log \frac{T_T}{T_b} + \log \left| \frac{T_N(\sec \theta_m)}{T_N(\sec \theta_m \cos \theta)} \right| \right] \theta = \frac{\pi}{2}$$

$$D = 20 \left[ \log \frac{T_T}{T_b} + \log |T_N(\sec \theta_m)| \right] \theta = \frac{\pi}{2}$$

Although $T_T/T_b$ depends on frequency and the characteristic is, in principle, different from the Chebyshev one, nevertheless such a shift is almost negligible, except very close to the midband frequency. Therefore, in the wideband case that contribution is negligible in the band where $\beta d = \theta_m$. Correspondingly

$$D = D_m = 20 \log |T_N(\sec \theta_m)|$$

where $D_m$ is the minimum directivity in the passband, due to the array factor. Hence, once the coupler specifications are set in terms of midband frequency $f_0$, bandwidth $\Delta f$, coupling, and directivity, the distance $d$ separating two adjacent apertures is given by

$$d = \frac{\pi}{2f_0}$$

It is immediate to compute $\cos \theta_m \approx d(\beta f_0 \pm \Delta f)$ and, from Eq. (35), the degree $N$ of the Chebyshev polynomial yielding the specifications on directivity. Because of the non-linearity of $\beta(f)$, $\cos \theta_m$ is only approximately calculated.

The constant $K$ is obtained from Eq. (32):

$$K = 10^{-\langle C/20 \rangle |T_T||T_N(\sec \theta_m)|}$$

Once the coefficients $T_T$ are properly computed, either by means of an electromagnetic analysis or by Bethe’s more accurate closed formulas, one has only to equate the coefficients of the array factor to those of $T_N$ and to determine the aperture radii. The preceding theory could be further improved with the help of a more accurate analysis of the coupling mechanisms, as the one proposed by Levy [10, 11]. At present, however, thanks to the availability of efficient and accurate field theory based computer-aided design (CAD), it seems to be more convenient to improve the design by performing an optimization directly on the
electromagnetic model of the actual physical structure. The resulting design is exact, and device tuning is unnecessary [12] (see ELECTROMAGNETIC FIELD MEASUREMENT) [13].

In planar technology, lumped coupling is often obtained by physically connecting two lines. In the preceding configuration, coupling due to electromagnetic induction can be considered negligible with respect to direct coupling. The main couplers employing such a coupling mechanism are the branchline coupler and the hybrid ring coupler. Both permit one to obtain large coupling values easily, since the lines are electrically connected. In particular, the electromagnetic analysis of the first one (Fig. 12) is difficult to carry out rigorously; a circuit analysis, however, is simple if one takes advantage of the four-fold symmetry. In fact, it is easy to study the equivalent circuit under four independent excitations: \((1,1,1,1), (1,1,1,1), (1,1,1,1), (1,1,1,1)\). The corresponding reflections at the port 1, \(\Gamma_a, \Gamma_b, \Gamma_c, \Gamma_d\) are given by [5]

\[
\Gamma_a = \frac{1 - jY_{t}t_{t} - jY_{b}t_{b}}{1 + jY_{t}t_{t} + jY_{b}t_{b}}
\]

\[
\Gamma_b = \frac{t_{t} + jY_{t}t_{t} - jY_{b}t_{b}}{t_{t} - jY_{t}t_{t} + jY_{b}t_{b}}
\]

\[
\Gamma_c = \frac{t_{b} - jY_{b}t_{b} + jY_{b}t_{b}}{t_{b} + jY_{b}t_{b} - jY_{b}t_{b}}
\]

\[
\Gamma_d = \frac{t_{t} + jY_{t}t_{t} + jY_{b}t_{b}}{t_{t} - jY_{t}t_{t} - jY_{b}t_{b}}
\]

where \(t_{t} = \tan \beta d_{t}/2\), \(t_{b} = \tan \beta d_{b}/2\), and \(Y_{t}\) and \(Y_{b}\) are the normalized characteristic admittances of the throughline and the branchlines, respectively. The characteristic admittance of the input line is normalized to 1.

It is immediate to combine Eqs. (38)-(41), thus finding the scattering parameters of the branchline coupler:

\[
S_{11} = \frac{1}{4} (\Gamma_a + \Gamma_b + \Gamma_c + \Gamma_d)
\]

\[
S_{12} = \frac{1}{4} (\Gamma_a - \Gamma_b + \Gamma_c - \Gamma_d)
\]

\[
S_{13} = \frac{1}{4} (\Gamma_a - \Gamma_b - \Gamma_c + \Gamma_d)
\]

\[
S_{14} = \frac{1}{4} (\Gamma_a + \Gamma_b - \Gamma_c - \Gamma_d)
\]

**Figure 12.** The branchline coupler.

If \(t_{t} = t_{b} = 1\) (i.e., the electrical lengths of the throughline and the branchline are the same) and also \(Y_{t}^{2} - Y_{b}^{2} = 1\), the four ports are matched, \(S_{14} = 0\), and

\[
S_{21} = -\frac{Y_{b}}{Y_{t}}
\]

\[
S_{21} = -\frac{1}{Y_{t}}
\]

\(S_{31}\) and \(S_{21}\) are in quadrature, which is predictable because of the symmetry of the coupler. The coupling \(C = 20 \log 1/|S_{31}|\) depends on the ratio between the two characteristic impedances. When \(Y_{t}/Y_{b} = \sqrt{2}\), \(C = 3\) dB. In that case a coupler, also having the outputs in quadrature, is commonly called “hybrid.” The preceding characteristics hold exactly only in a narrow interval around the working frequency. However, it is possible to enlarge considerably the bandwidth of the coupler by cascading sections [14,15].

The principal of operation of the hybrid ring (Fig. 13) is similar to that of the branch coupler. When port 1 is fed, the signal splits equally into two signals traveling clockwise and counterclockwise that recombine at the remaining ports. At ports 2 and 3 the signals sum in phase, while at port 4 they are out of phase. The same reasoning applies to the remaining ports.

**Figure 13.** Ring coupler. A wave entering port 1 is split into two waves that recombine in phase at ports 2 and 4. On the other hand, port 3 is isolated, since the waves propagating clockwise and counterclockwise respectively are 180° out of phase. The same reasoning applies to the remaining ports.

In planar structures (e.g., microstrip, stripline, finline, which are the most common), as well as in TEM lines (e.g., coaxial cables), the coupling is often obtained by placing the two lines parallel and close to each other over a certain length, in such as way that a portion of the field wave traveling along the first guide couples by electromagnetic induction into the second one in contraflow, as schematically sketched in Fig. 14 [16–18]. The coupling depends mainly on the gap separating the two strips: The wider the gap, the weaker the coupling.
Hence, the scattering matrix takes the form

\[ S = \begin{bmatrix} s_{11} & s_{12} & s_{13} & s_{14} \\ s_{12} & s_{11} & s_{14} & s_{13} \\ s_{12} & s_{14} & s_{11} & s_{13} \\ s_{14} & s_{13} & s_{12} & s_{11} \end{bmatrix} \]

The scattering parameters can be calculated by considering two independent sets of excitations. The first is \( a_1 = 1, a_2 = 0, a_3 = 0, a_4 = 1 \) and corresponds to a magnetic wall on the symmetry plane \( x = 0 \). The second is \( a_1 = 1, a_2 = 0, a_3 = 0, a_4 = -1 \) and corresponds to an electric wall on the same symmetry plane. The reflected waves at ports 1 and 2 are given by

a) under the first excitation

\[ b_1^e = s_{11} + s_{13} \]  
\[ b_2^e = s_{12} + s_{14} \]

b) under the second excitation

\[ b_1^o = s_{11} - s_{13} \]  
\[ b_2^o = s_{12} - s_{14} \]

It is immediate to calculate the parameters of the four-port:

\[ s_{11} = \frac{b_1^e + b_1^o}{2} \]  
\[ s_{12} = \frac{b_2^e + b_2^o}{2} \]  
\[ s_{13} = \frac{b_1^e - b_1^o}{2} \]  
\[ s_{14} = \frac{b_2^e - b_2^o}{2} \]

The preceding two situations can be modeled by two transmission lines having normalized characteristic impedances \( Z_0^e \) and \( Z_0^o \) and, correspondingly, electrical lengths \( \theta^e \) and \( \theta^o \). Although the physical lengths of the lines are the same, their electrical lengths are different in the non-TEM case, and the propagation constants are different in the two cases. Let us suppose both lines are fed by transmission lines with unit characteristic impedance. The transmission matrices are given by

\[ T = \begin{bmatrix} \cos \theta^o \sin \theta^e \\ \sin \theta^o \cos \theta^e \end{bmatrix} \]

As observed previously, losslessness implies that the condition under which the device is a directional coupler is \( s_{11} = 0 = b_1^e + b_1^o \). Such a condition is satisfied when

\[ \frac{\left( Z_0^e - \frac{1}{Z_0^o} \right) j \sin \theta^o}{2 \cos \theta^o + \left( Z_0^e + \frac{1}{Z_0^o} \right) j \sin \theta^e} = \frac{\left( Z_0^o - \frac{1}{Z_0^e} \right) j \sin \theta^e}{2 \cos \theta^e + \left( Z_0^o + \frac{1}{Z_0^e} \right) j \sin \theta^o} \]

An immediate solution is obtained when \( \theta^e = \theta^o = \theta \), as occurs when the strips are embedded in a homogeneous medium, and \( Z_0^e = 1/Z_0^o \). The more immediate solution is therefore to place over the strip a dielectric layer that has the same permittivity as the substrate. This particular case pertains to TEM couplers. In such a case, the coupling \( C \) is given by the formula

\[ C = 20 \log \left( 1 - \frac{c^2 \cos^2 \theta}{c \sin \theta} \right) \]

where \( c = Z_0^e - Z_0^o/Z_0^e + Z_0^o \). Since in microstrip technology \( c_{\text{max}} \approx \frac{1}{2} \), the maximum coupling achievable by the ordinary photolithographic technique is about 6 dB and occurs when \( \theta = \pi/2 \). Moreover, its bandwidth is rather narrow and the directivity moderate. Much better performances are obtained by the Lange interdigitated coupler, shown in Fig. 15 [19].

This configuration permits one to achieve 3 dB coupling easily, as well as an octave bandwidth and a good...
directivity. This coupler is, however, difficult to realize and the bond wires are critical at the higher frequencies. Moreover, the higher the frequency, the more difficult it is to equalize the phase velocities of the even and odd modes. Nevertheless, its good electrical characteristics and compactness make the Lange coupler suitable for applications up to 30 GHz with standard technology, but the gallium arsenide (GaAs) monolithic version is used up to 100 GHz. A comprehensive and accurate analysis of the interdigitized coupler is reported in Ref. 20.

The compensation of the different phase velocities over a wide band is one of the more difficult tasks in microstrip couplers. At present, designers often adopt one of the following strategies:

1. Placing two capacitances across the lines at the input and the output
2. Using nonuniform planar transmission lines
3. Shaping the two strips in a serpentine form or in a shark teeth form (wiggly coupler) as shown in Fig. 16
4. Combining the preceding techniques (for instance, by shaping as shark teeth the strips of a nonuniform coupler)

Similar to waveguide couplers, bandwidth and directivity can be much improved by cascading many coupled line sections, as indicated in Fig. 17 [21–23].

A detailed description of planar directional couplers can be found in Ref. 24. Planar couplers are also affected by high conductor losses. In this regard, the use of superconductor technology is attractive, though unfortunately not mature enough for large-scale production. Microstrip couplers do not achieve performances comparable to the ones realized in waveguide; therefore, they are hardly employed in measurement benches, where high directivity and loss losses are required. However, their use in civil telecommunication systems is widespread.

BIBLIOGRAPHY

Radio antennas that are directive have many advantages over stationary antennas. They can observe in one direction while ignoring much larger signals coming in from the sides. In addition, as transmitting antennas, they can direct a small amount of radiation directly to the proper receiver. Conventional directive antennas use large mirrors for short waves, such as are used in radar. For longer-wave antennas, they use reflector and director elements aligned with the antenna. The basic problem with such arrays is that they are mechanically steered, which is slow and difficult to do with large antennas, especially in the presence of wind, ice, or snow.

One way of having a large steerable antenna with no moving parts is to use a phased array. A phased array is based on Huygen’s principle, which states that a wavefront can be determined at a point in time by constructing a surface tangent to a collection of secondary waves. Thus, if one has a large number of small antennas located on a line, and if the antennas are excited in phase, the resultant wavefront is normal to the line. If each small antenna is excited with a small phase shift relative to the next antenna, the wave propagates at an angle to the line. Conversely, if the array is used for reception, the small antenna signals, combined with the individual phase shifts, allow the array to scan in azimuth. The principle can be improved by using a two-dimensional array located on a plane to scan in altitude as well as azimuth. In fact, by properly adjusting the phase shifts, such an antenna can be located on a curved surface, such as the nose or wing of an aircraft. The basic principle is shown in Fig. 1. Since the subject of directive antennas is so large, Professor Herb Neff, UTK Emeritus (an antenna specialist), recommends using Refs. 1 and 2 for reference.

Newer antennas being developed include the “agile mirror” [3,4] designed by Dr. Wallace M. Manheimer of the US Naval Research Laboratory. In this concept, a sheet of ionized air or other gas forms a reflecting surface. Since the sheet is not a mechanical body, it can be tilted and re-formed in a very short span of time. Thus, the direction of transmission or reception can be varied extremely rapidly. A magnetic field is used to help direct the sheet electron beam that forms the plasma. A second advantage of such an antenna is that it vanishes when the electrical discharge is terminated. This is a great advantage for stealth technology, because a mechanical antenna presents a large scattering cross section to radar signals near the antenna resonant frequency. One disadvantage of the “agile mirror” is that it must be formed in a gas at reduced pressure.

A second new type of antenna is the “stealth antenna,” which is being developed by the Patriot Scientific Corporation in San Diego, California. The idea, originally developed by Dr. Igor Alexeff at the University of Tennessee [5], uses glow discharge tubes to comprise elements of an antenna. When the tubes are energized, the antenna is a complete conducting structure. When the tubes are deenergized, the antenna becomes either a large number of separate, nonresonant conducting components or just a nonconducting structure of glass tubing. By selectively energizing various tubes, the antenna also can be directed.

**BIBLIOGRAPHY**

DISTRIBUTED AMPLIFIERS

T. KORYU ISHII
Marquette University
Milwaukee, Wisconsin

1. DEFINITION AND STRUCTURE

The objective of this article is to present various aspects of distributed amplifiers. Distributed amplifiers are by definition, electronic amplifiers consisting of distributed circuit parameters. However, in practice, amplifier systems that consist of a number of discrete amplifiers associated with distributed parameter circuits are often termed distributed amplifiers; this latter amplifier is actually a pseudodistributed amplifier.

In practice, the distributed parameter circuit often takes the form of a transmission line. The circuit parameters, the inductance, the capacitance, and the resistance are distributed throughout the transmission line. If the transmission line is a conventional passive transmission line, the electrical output power of the transmission line is either equal to or less than the electrical input power depending on the power loss of the transmission line.

If the transmission line is active, then the output power is greater than the input. In this case, the transmission line is considered as an amplifier; this is actually a distributed amplifier.

For example, an ordinary optical fiber cable is a passive transmission line for lightwaves. The output light of the optical cable is always less than the input light because of the cable loss. But an erbium-doped optical fiber cable is different. The lightwave output of the cable is greater than the lightwave input. The input lightwaves (which are electromagnetic waves) are amplified. The erbium-doped optical fiber cable is an active transmission line and is one form of distributed amplifier. A schematic diagram of a generic distributed amplifier is shown in Fig. 1a. In this distributed amplifier, the transmission line is continuously loaded by the continuously distributed power-pumping active substrate.

In a pseudodistributed amplifier, a number of discrete amplifiers are periodically loaded as shown in Fig. 1b. The input power is amplified by these discrete amplifiers; therefore the output of the transmission line is greater than the input power.

The objective of the distributed amplifiers is to obtain a high-frequency bandwidth with high-gain amplification. The operating frequency ranges are in RF, microwaves, and lightwaves. Depending on the operating frequency range, the amplifier configurations are markedly different. The transmission line can be a two-wire line, a coaxial (“coax”) line, a waveguide, a microstrip line, a coplanar waveguide, or an optical fiber cable.

The term “distributed amplifier” contrasts against “discrete amplifier” or “lumped amplifier.” A lumped amplifier is represented in a block diagram as shown in Fig. 2. In a lumped or discrete amplifier, where point A is the input and point B is the output, the geometric distance between these two points is negligibly small in comparison with the operating wavelength. A distributed amplifier can also be represented by a block diagram as shown in Fig. 2, but the geometric distance between the actual point A and the actual point B is comparable to the operating wavelength.

1.1. Continuous Active-Diode Distributed Amplifiers

Activated tunnel diodes, Gunn diodes, and varactor diodes are considered as active diodes. When tunnel diodes and Gunn diodes are properly biased, these diodes exhibit negative resistance. Ordinarily a resistance is positive. A positive resistance consumes electrical energy. A negative resistance generates electrical energy. Therefore, if the amount of negative resistance is adjusted by material composition, configuration, and the bias current and if the circuit impedance of the transmission line is properly adjusted, then the active-diode-loaded transmission line can amplify propagating electromagnetic waves on the transmission line. One possible biasing method is illustrated in Fig. 3a. The transmission line is most likely a microstrip line or a coplanar coupled waveguide. The microstrip line is DC-biased through a RF choke. If the active substrate is a tunnel diode of long degenerate pn junction, the properly forward-biased pn junction exhibits negative resistance by the tunnel effect [1]. If the active substrate is a long Gunn diode of properly doped n-type GaAs, the substrate exhibits negative resistance by the carrier momentum transfer effect [2].

Figure 1. Generic configuration of distributed amplifiers. Signals to be amplified are fed at the left terminal. The signals are amplified during propagation on the line. The amplified signals exit from the right. (a) Distributed amplifier. The amplifier consists of a continuous active transmission line. (b) Pseudodistributed amplifier. Lumped amplifiers are periodically loaded on a passive transmission line.

Figure 2. A block diagram representation of a discrete amplifier, a lumped amplifier, or a distributed amplifier. Generic symbols of a generic amplifier: G represents the gain of the amplifier (which can be the voltage, current, or power gain); A is the input, and B is the output terminal.
The negative resistance can also be created by a properly biased and pumped long varactor diode junction. The varactor diode is a reverse-biased pn-junction diode. This is a variable-capacitance diode, and the junction capacitance is varied depending on the bias voltage across the diode. The junction capacitance in the case of Fig. 3b is controlled by the DC bias and the pump oscillator voltage launched on the microstripline transmission line. Some varactor diodes work without DC bias. The pump oscillator frequency $f_p$ is approximately twice of the signal frequency $f_s$ for the best results. When the pump oscillator frequency and phase are properly adjusted, the energy of the pump oscillator transfers to the signal through the variable-junction capacitance and the signal waves are amplified as the waves propagate on the microstripline. The amplifier that functions by use of a junction capacitance is termed a varactor parametric amplifier [1].

1.2. Periodically Loaded Active-Diode Distributed Amplifiers

A schematic diagram of a periodically loaded active-diode microstripline distributed amplifier is shown in Fig. 4a. The active diodes are either discrete tunnel diodes or Gunn diodes. The periodicity $L$ is usually less than a quarter-wavelength to avoid resonance. When the periodicity is made equal to either a quarter-wavelength or a half-wavelength, the amplifier will be at resonance. In such cases, the frequency bandwidth becomes narrow and it may also become unstable and oscillate; therefore the resonance should be avoided. One objective of the distributed amplifier is to obtain a wide frequency bandwidth; thus it is safe to keep the periodicity $L$ less than a quarter-wavelength. The diodes must be DC-biased properly at the middle of the negative-resistance region.

A schematic diagram of a periodically loaded variable-capacitance diode (varactor diode) parametric distributed amplifier is shown in Fig. 4b. As seen from the diagram, varactor diodes are reverse-biased by the DC bias supply and are pumped by the pump oscillator. The pump frequency $f_p$ is approximately twice of the signal frequency $f_s$ to be amplified. The pump wave on the line must be synchronized with the signal wave. The synchronization is accomplished using a variable phase shifter as shown in the pump oscillator circuit. The pump oscillator power is transferred into the signal through the varactor, and the signal wave is amplified [1,2]. The varactor diodes are pumped so that when and where the signal waves crest the junction, the capacitance becomes minimum. This phasing makes the microwave signal voltage amplified. The transmission line can be a microstripline as shown or...
a coplanar coupled waveguide. An example of a coplanar coupled waveguide distributed amplifier is sketched in Fig. 5. As seen from this figure, fabrication of a coplanar coupled waveguide amplifier is easier than fabrication of a microstripline amplifier.

### 1.3. Continuous Transistor Distributed Amplifiers

A schematic diagram of a continuous transistor distributed amplifier is shown in Fig. 6. This is a FET (field-effect transistor) of long configuration. The length of lines must be greater than a wavelength of the operating carrier signals. The microwave input signals are fed into the coplanar coupled waveguide that consists of a gate strip and a source strip. As the input microwaves propagate along this input gate–source coplanar coupled waveguide, the amplified signal waves appear on the drain–gate coplanar coupled waveguide. Then the amplified microwaves exit at the end of the drain–gate coplanar coupled waveguide. The long transistor must be properly DC-biased as shown in Fig. 6.

### 1.4. Periodically Loaded Transistor Distributed Amplifiers

A schematic diagram of a periodically loaded transistor distributed amplifier is shown in Fig. 7. To qualify as a distributed amplifier, the length of the coplanar coupled waveguide must be longer than a wavelength of operating microwaves. If the length is very short, this is a simple

---

**Figure 4.** Periodically loaded active-diode distributed amplifiers. (a) Periodically loaded active-diode microstripline distributed amplifier. A microstripline is periodically loaded by active diodes with periodicity $L$. A properly biased active diode is capable of amplifying electromagnetic signals. (b) Periodically loaded active capacitive parametric distributed amplifier. A microstripline is loaded periodically by properly biased and pumped varactor diodes with periodicity $L$. Such a varactor diode acts as a lumped amplifier.

**Figure 5.** An example of coplanar coupled waveguide distributed amplifier. This is an example of a case in which the transmission line is a coplanar coupled waveguide. Lumped diodes are mounted on it periodically.

**Figure 6.** A schematic diagram of a continuous transistor distributed amplifier. This is a case of extremely long gate field-effect transistor. The length of the gate can be several wavelengths longer than the operating wavelength. As the input signals propagate on the gate–source line, the amplified output signals travel on the drain–gate line. The amplified signals exit at the right.
1.5. Thermionic Traveling-Wave Distributed Amplifiers

A thermionic traveling-wave amplifier is a vacuum tube [2]. Electrons are emitted from an electron gun into a vacuum as shown in Fig. 8. The emitted electrons are pulled by the anode, which is a helical transmission line focused by the longitudinally applied DC magnetic flux \( B \). The electron beam is shot through the helix line and hits an endplate termed the electron collector. The electron collector collects ("spent") electrons. The helix transmission line is a distributed parameter transmission line. The pitch angle of the helix transmission line, the diameter of the helix, and the electron acceleration voltage are adjusted in such a way that the speed of the electron beam is equal to the axial propagation speed of microwaves on the helix transmission line. Then the kinetic energy of the electron beam is transferred to the microwave energy on the helical line through the distributed capacitance between the electron beam and the helical line. As the microwaves on the line and the electrons in the beam travel together, the microwaves are amplified and exit from the output of the tube as shown in Fig. 8 [2]. The helical transmission line is replaced by a meander line or an interdigital transmission line (2).

1.6. Fiberoptic Distributed Amplifiers

A schematic diagram of a fiber optic distributed amplifier is shown in Fig. 9. The main part of this amplifier is a section of erbium-doped optical glass fiber cable [3]. As seen from this figure, if a lightwave of proper wavelength is pumped into the fiber cable through a directional coupler from a pump laser, and the lightwave signal to be amplified is fed into the input of the fiber cable thorough an isolator, then, while the signal lightwave is propagating in the fiber cable, the signal lightwave is intensified by the emission of radiation from the erbium atoms that are pumped by the lightwave that is propagating in the fiber cable from the pump laser. The pump lightwave travels with the signal lightwave and pumps energy into the signal lightwaves through the stimulated emission of radiation from erbium atoms. The particular optical fiber is considered to be a distributed parameter transmission-line amplifier for propagating optical electromagnetic wave signals.

![Figure 7](image-url)  A schematic diagram of a thermionic traveling-wave distributed amplifier. Discrete transistors are periodically loaded on a coplanar coupled waveguide. The input signals fed on the gate–source coplanar waveguide are amplified as the signals propagate on the line. The amplified output appears on the drain–gate coplanar coupled waveguide. The output propagates on the line and exit from the right.

![Figure 8](image-url)  A schematic diagram of a thermionic traveling-wave distributed amplifier. The pitch of the helical transmission line is adjusted so that the axial speed of microwave propagation on the line is almost equal to the speed of electrons in the electron beam. Under this condition, The kinetic energy of electrons is transferred into the traveling microwaves on the line, and the propagating microwaves are amplified.

![Figure 9](image-url)  A schematic diagram of a fiber cable lightwave distributed amplifier. The active part is a long section of erbium-doped fiber cable. The erbium atoms are pumped by a light from the pump laser at left. The input lightwave signals are amplified by the stimulated emission of radiation from the pumped erbium atoms in the active fiber cable.
**2. GENERAL GOVERNING EQUATIONS**

### 2.1. Gain

A generic configuration of a distributed amplifier transmission circuit is shown in Fig. 10. In this diagram, $R$ is the series resistance per meter of the distributed amplifier, $L$ is the series inductance per meter of the distributed amplifier, $-G$ is the negative conductance per meter of the distributed amplifier, and $C$ is the capacitance per meter of the distributed amplifier. The amplification constant of this amplifier is [1]

$$\alpha = \frac{\omega (CR - LG)}{\sqrt{2}} \left\{ \frac{\omega^2 (CR - LG)^2}{(\omega^2 LC + RG) + (RG + \omega^2 LC)^2} \right\}^{1/2}$$

where $G$ is the magnitude of the negative conductance parameter. In a distributed amplifier, if the propagating power increase per meter is $\Delta P$ (W/m) and the propagating voltage increase parameter is $\Delta V$ (V/m), the magnitude of the negative conductance per meter is $G = 2\Delta P/\Delta V^2$ (S/m).

The phase constant of this amplifier is [1]

$$\beta = \frac{\left\{ \omega^2 (CR - LG) + \sqrt{\omega^2 (CR - LG)^2 + \omega^2 (CR - LG)^2} \right\}^{1/2}}{2}$$

If the length of the active region of the amplifier is $l$ meters long, then the voltage gain of the amplifier is

$$G(f) = z(f)l = \frac{(2\pi f)(C(f)R(f) - L(f)G(f))}{\sqrt{2} \left\{ ((2\pi f)^2 L(F)C(f) + R(f)G(f)) + \sqrt{(2\pi f)^2 (C(f)R(f) - L(f)G(f))^2 + (R(f)G(f) + (2\pi f)^2 L(f)C(f))^2} \right\}^{1/2}}$$

At the edge of the frequency bandwidth at a frequency $f'$

$$G(f') = \frac{1}{\sqrt{2}} G(f_0)$$

where $f_0$ is the center frequency of the amplifier. Usually Eq. (6) is at least the second-order equation of $f$. One root is $f_H'$, which is greater than $f_0$ and the other is $f_L'$, which is less than $f_0$. Then the frequency bandwidth is

$$\Delta f = f_H' - f_L'$$

### 2.2. Frequency Bandwidth

In a generic distributed amplifier, the circuit parameters $R$, $L$, $C$, and $G$ are functions of operating frequency $f$. Therefore, the gain of the amplifier is

$$\Delta l = \beta l \quad \text{[radians (rad)]}$$

### 2.3. Sensitivity

According to the IEEE Standard Dictionary [4], sensitivity is defined as “the minimum input signal required to produce a specified output signal having a specified signal to noise ratio.” This means that

$$\frac{P_s A}{KTBAF} = \frac{S_0}{N_0}$$

where $A$ is the power gain of an amplifier, $K$ is the Boltzmann constant (1.38054 x 10^{-23} J/K), $T$ is the absolute temperature of the input circuit to the amplifier, $B$ is the overall frequency bandwidth of the amplifier, $P_s$ is the input signal power, $F$ is the noise figure of the amplifier, and $S_0/N_0$ is the signal-to-noise power ratio of the amplifier at
the output. Then

\[ P_s = KTBF \frac{S_0}{N_0} \quad (9) \]

As “a specified signal to noise ratio,” often

\[ \frac{S_0}{N_0} = 1 \quad (10) \]

is used for the definition of the sensitivity of the amplifier. Then, the sensitivity is

\[ P_s|_{S_0/N_0 = 1} = KTBF \quad (11) \]

For a distributed amplifier, the value of \( B \) is obtained using Eq. (7). The value of the noise figure \( F \) can be obtained from the next section. Then the sensitivity is

\[ P_s|_{S_0/N_0 = 1} = \frac{N_0}{A} \quad (12) \]

where \( A \) is given by

\[ A = e^{2\alpha l} \quad (13) \]

where \( \alpha l \) is given by Eq. (5).

### 2.4. Noise Figure

The noise figure \( F \) of an amplifier is given by [1]

\[ F = \frac{N_0}{KTBA} \quad (14) \]

where \( N_0 \) is the noise output of the amplifier (W). For a distributed amplifier, both the frequency bandwidth \( B \) and the power amplification \( A \) are given by Eqs. (7) and (5), respectively.

### 2.5. Dynamic Range

A range of input signal level in which the gain of the amplifier is constant is termed the **dynamic range** of the amplifier. Usually, the gain of an amplifier is less at an extremely small input signal level or at a large input signal level.

In the semiconductor distributed amplifiers, thermionic distributed amplifiers or even in fiber optic distributed amplifiers, the values of \( L, C, R, \) and \( G \) are inherent functions of operating signal levels \( V_s \). (Therefore, in Eq. (1)

\[
\zeta(V_s) = \frac{\omega(C(V_s)R(V_s) - L(V_s)G(V_s))}{\sqrt{2} \left( \omega^2L(V_s)C(V_s) + R(V_s)G(V_s) \right) + \frac{\omega^2(C(V_s)R(V_s) - L(V_s)G(V_s))^2}{(R(V_s)G(V_s) + \omega^2L(V_s)C(V_s))^2} \right)^{1/2}}
\]

(15)

If the gain constant in the linear region of the distributed amplifier is \( z_0 \), then the power gain of the amplifier is

\[ A_0 = e^{2\alpha l} \quad (16) \]

where \( l \) is the length of the active region of the distributed amplifier. In a large-signal level \( V_s \), the gain will be compressed by saturation and

\[ A(V_s) = e^{2(z_0 - \zeta(V_s))l} \quad (17) \]

where \( \zeta(V_s) \) is as given in Eq. (15).

If the gain compression of \(- n \) dB is chosen, then

\[ -n \ dB = 10 \log_{10} \frac{A(V_s)}{A_0} \quad (18) \]

and

\[ n \ dB = 10 \log_{10} e^{2(z_0 - \zeta(V_s))l} \quad (19) \]

or

\[ n \ dB = 8.686(z_0 - \zeta(V_s))l \quad (20) \]

In practice, \( n = 1 \) is often chosen, and the value of the input voltage for \( n = 1 \) is termed the input signal voltage at 1 dB compression point. The 1 dB compression point input signal voltage is then

\[ \zeta(V_s) = z_0 - \frac{1}{8.686l} \quad (21) \]

### 2.6. Stability

As seen from Eq. (1), a generic distributed amplifier is inherently stable. A controlling parameter in Eq. (1) is the magnitude of the negative conductance per meter \( G \). Equation (1) does not show any singularity due to the size of \( G \) within the range of practical operation.

### 3. PERIODICALLY LOADED ACTIVE-DIOIDE DISTRIBUTED AMPLIFIERS

#### 3.1. Periodically Loaded Tunnel Diode Distributed Amplifiers

In a periodically loaded tunnel diode disturbed amplifier, a number of discrete tunnel diodes are periodically loaded on a RF transmission line as shown in Fig. 4a. A generic volt-
ampere curve of a tunnel diode is shown in Fig. 11. This is a plot of the diode current and the voltage across the diode. When the diode is biased in a negative-conductance region, the amount of the negative conductance is given by

$$ G = \frac{\partial I}{\partial V} < 0 $$  \hspace{1cm} (22)$$

An equivalent circuit of a discrete tunnel diode is shown in Fig. 12, where $L$ is the lead inductance, $R_s$ is the spreading resistance, $C_J$ is the junction capacitance, $C_p$ is the package capacitance, and $-G$ is the negative conductance of the tunnel junction created by the tunnel effect. With the help of additional impedance matching components, it is possible to tune out the inductances and capacitances, and under a matched and tuned condition, the tunnel diode can be represented by a negative conductance of magnitude $G_D$.

The RF power gain due to a discrete negative conductance $G_D$ that is matched to a characteristic impedance of the transmission line $Z_0$ is [3]

$$ A = \frac{1}{1 - G_D Z_0} $$  \hspace{1cm} (23)$$

At any diode in Fig. 4a, half of the amplified power goes back to the input and only another half of the power amplified continues to travel toward the output. So, the actual power gain of traveling waves toward the output is

$$ A^+ = \frac{1}{2(1 - G_D Z_0)} $$  \hspace{1cm} (24)$$

If $N$ diodes are used in a distributed amplifier as shown in Fig. 4a, after matching and tuning, the total power gain of the amplifier is

$$ A_T = (A^+)^N = \frac{1}{2^N(1 - G_D Z_0)^N} $$  \hspace{1cm} (25)$$

For the impedance matching and tuning, in addition to attaching the impedance-matching circuit components to the diode mount, the adjustment of periodicity together with the diode biasing must be done properly.

### 3.2. Periodically Loaded Gunn Diode Distributed Amplifiers

A voltampere characteristic of a generic Gunn diode is in a similar shape as shown in Fig. 11, except that the negative conductance is smaller than that of a tunnel diode. The negative conductance of Gunn diode is created by the transfer of the electronic momentum between a high-electric-field domain and a low-field domain in the bulk of a semiconductor diode. The equivalent circuit of a Gunn diode is similar to the circuit shown in Fig. 12. Therefore the principle of periodically loaded Gunn diode distributed amplifiers is similar to the principle of periodically loaded tunnel diode distributed amplifiers. Then the power gain equation of a Gunn diode distributed amplifier consists of $N$ Gunn diodes in the negative conductance $G_D$ with the characteristics impedance $Z_0$ is

$$ A_T = \frac{1}{2^N(1 - G_D Z_0)^N} $$  \hspace{1cm} (26)$$

### 3.3. Periodically Loaded Varactor Diode Distributed Parametric Amplifiers

When discrete variable capacitance diodes (varactor diodes) are periodically mounted on a transmission line as shown in Fig. 4b, RF voltage is amplified by the parametric effect of the junction capacitance. The voltage gain across a single parametric capacitance diode is given by [5]

$$ A = \frac{v_p \sqrt{Q_t}}{4(V_0 + V_{r0})} + \sqrt{Q_s} $$  \hspace{1cm} (27)$$

**Figure 11.** Generic voltampere (V-A) curve of a tunnel diode. Note that the V-A curve does not follow Ohm’s Law. Note also the negative differential conductance at the midvoltage region. This is a plot of the diode current versus the diode bias voltage.
where \( v_p \) is the pump voltage across the junction capacitance of the varactor diode, \( V_0 \) is the magnitude of the contact potential barrier of the diode, \( V_{ro} \) is the DC reverse bias voltage, and \( Q_i \) and \( Q_s \) are the quality factors of the idler circuit and the signal circuit per diode, respectively.

In a parametric amplifier, the idle frequency in Fig. 4b is usually

\[
f_i = f_p - f_s
\]

and

\[
f_p \approx 2f_s
\]

for a high gain [5]. So

\[
Q_i \approx Q_s
\]

Applying the same concept expressed in Eqs. (25) and (26), the total voltage gain of \( N \)-diode distributed parametric amplifier is, after matching and tuning

\[
A_{VT} = \left\{ \frac{v_p \sqrt{Q_i}}{4(V_0 + V_{ro})} + \sqrt{Q_s} \right\}^N
\]

4. PERIODICALLY LOADED TRANSISTOR DISTRIBUTED AMPLIFIERS

Similar concepts of Eqs. (25), (26), and (31) are applicable to a periodically loaded discrete-transistor amplifier. The transistors can be either junction transistors or field-effect transistors. If the \( s \) parameter of the discrete transistor from the gate (or base) to the drain (or collector) is \( S_{21} \), then, after impedance matching and tuning, the voltage gain of \( N \)-transistor distributed amplifier is given by

\[
A_{VT} = S_{21}^N
\]

5. THERMIONIC DISTRIBUTED AMPLIFIERS

Thermionic distributed amplifiers are vacuum tubes referred to as traveling-wave tubes (TWTs). A schematic diagram of a generic TWT is shown in Fig. 8, and the principle of the TWT distributed amplifier was briefly explained earlier in this article. While microwaves travel along the helical transmission line with the axial propagation speed approximately equal to the speed of electron beam, the kinetic energy of the electron beam is gradually transferred to the propagating microwaves in the transmission circuit and hence the microwaves are amplified. The propagation constant of a TWT is given by [1,2]

\[
\hat{\gamma} = \beta_e \left[ -\frac{\sqrt{3}}{2} \frac{C + j(1 + \frac{C}{2})}{C} \right] (m^{-1})
\]

where \( \beta_e \) is the phase constant associated with the DC electron beam and

\[
\beta_e = \frac{\omega}{u_0}
\]

where \( \omega \) is the operating angular frequency and \( u_0 \) is the speed of the electrons in the beam. \( C \) in Eq. (33) is termed the gain parameter [1,2] and is given by

\[
C^3 = \frac{Z_0 I_a}{4V_a}
\]

where \( Z_0 \) is the characteristic impedance of the helical line, \( I_a \) is the DC electron-beam current, and \( V_a \) is the acceleration anode voltage of the traveling-wave tube.

If the length of the active interaction region on the helical transmission line is \( l \) meters long, then the voltage gain of this TWT is [2]

\[
A = e^{\alpha(l/2)}\beta_e l
\]

6. QUANTUM ELECTRONIC DISTRIBUTED AMPLIFIERS

A quantum electronic distributed amplifier can be a continuous configuration as shown in Fig. 9. If the signal to be amplified is a lightwave, then this distributed amplifier is a traveling-wave laser. If the signal to be amplified is a microwave, then this distributed amplifier is a traveling-wave maser. For a traveling-wave maser, instead of the optical fiber cable, a microwave transmission line that is continuously loaded with maser material such as a ruby or a rutile crystal, and instead of the pump laser, a microwave local pump oscillator is used.

At any rate, the gain constant of a traveling-wave maser or laser distributed amplifier is given by [6]

\[
x = \frac{\omega}{2Q_m v_g}
\]

where \( \omega \) is the angular frequency of the signal to be amplified, \( Q_m \) is the quality factor/meter of the active cable, and \( v_g \) is the group velocity of the signal in the cable. The quality factor \( Q_m \) is given by

\[
Q_m = \frac{\omega W_{s0}}{\Delta P}
\]

where \( W_{s0} \) is the electromagnetic energy of the signal stored per meter of the cable and \( \Delta P \) is the signal power loss per meter of the cable.

The voltage gain of this continuously loaded distributed laser or maser amplifier is

\[
A = e^{\alpha(l/2Q_m v_g)}
\]

where \( l \) is the length of active part of the cable.

A quantum electronic distributed amplifier can be a periodical loading configuration as shown in Fig. 13. A
microwave transmission line of a periodical structure is continuously loaded with an activated maser crystal and placed in a rectangular microwave waveguide. The pump power from a pump oscillator is fed to the rectangular waveguide to activate the maser crystal. The pumped-up maser crystal emits radiation when stimulated by the input microwave signals.

In a manner similar to the case of continuously loaded quantum electronic distributed amplifier, the voltage gain of the periodically loaded quantum electronic distributed amplifier is given by

$$A = e^{(\omega/dQ_{mp})}$$  \hspace{1cm} (40)$$

where $Q_{mp}$ is the quality factor within the periodicity of the periodical structure. Then

$$Q_{mp} = \frac{\omega W_{sp}}{\Delta P}$$  \hspace{1cm} (41)$$

where $W_{sp}$ is the signal energy stored within the periodicity of the structure of the transmission line and $\Delta P$ is the signal power loss within the periodicity.

7. EXAMPLES OF DISTRIBUTED AMPLIFIERS

7.1. RF Distributed Amplifiers

In practice, at RF $<300$ MHz, a distributed amplifier can be built using discrete components or surface-mountable discrete components. An example of such distributed amplifiers is shown schematically in Fig. 14 [3,6,20,23,24]

Discrete FETs are sequentially excited through the gate delay line or the gate artificial transmission line consisting of $C_g, \frac{1}{2} L_g, L_{gg},$ and $R_g$. These are discrete components. $C_g$ is a DC blocking input coupling capacitor, $L_g$ is an inductor to produce the desired phase delay between stages of the FET amplifiers, and $R_g$ is the matched terminating resistor for the artificial transmission line. The idea is to generate RF traveling waves on the gate artificial transmission line. $L_{gg}$ is a stray inductance of the gate lead. In most cases $L_{gg}$ is negligibly small at most RF frequencies. $R_s$ and $C_s$ are the source bias resistor and bypass capacitor, respectively. $L_{dd}$ is the stray inductance of the drain of the FET. By making the drain lead as short as possible, it is possible to make $L_{dd}$ negligibly small at RF frequencies.

The drain delay line or the drain artificial transmission line is formed by $R_d, L_d,$ and $C_d$. $R_d$ is an impedance-matched resistor to the drain artificial transmission line,
and $L_d$ is the phase-delaying inductor between stages. The value of $L_d$ must be determined so that the phase of waves on the drain artificial transmission line synchronize with the phase of waves on the gate artificial transmission line. $C_d$ is a DC blocking RF coupling capacitor to the output load. The transistors are biased through a RF choke coil and a decoupling capacitor.

### 7.2. Microwave Distributed Amplifiers

Various types of distributed amplifiers in microwave frequency have been built in the past [7–9]. In microwave frequencies, the distributed amplifiers take forms of monolithically developed integrated circuits as shown in Fig. 15 as an example. As is the case in Fig. 14, the microwave input signals to be amplified are fed to the gate microstripline with impedance-matched terminating resistance $R_g$ through a DC blocking coupling capacitor $C_{cg}$. The gate of each FET that is properly biased is sequentially excited. Amplified microwave drain current propagates along the drain microstripline toward the output and is coupled out to the output circuit through a DC blocking coupling capacitor $C_{dd}$. The drain microstripline is terminated with an impedance-matched resistor $R_d$. The drain microstripline is biased through a RF choke coil and a bypass capacitor with $V_{dd}$.

Most microwave monolithic integrated-circuit (MMIC) distributed amplifiers have extremely wide frequency bands even though the total gain is not so high. They are extremely compact. For example, Kimura and Imai [11] monolithically integrated a seven stage distributed amplifier on a $1.5 \times 2.5$-mm IC substrate and reported a flat gain of 9 dB over the frequency range 0–55 GHz with a 6 dB noise figure. Further examples can be found in the literature [11–19,25–27].

### 7.3. Lightwave Distributed Amplifiers

Actual configuration of a lightwave distributed amplifier is shown in Fig. 9. These amplifiers are actually deployed to boost lightwave signals for a long-haul lightwave signal transmission such as transoceanic lightwave cables. For example [3], the lightwave input signal to be amplified has a wavelength of 1500 nm. The pump laser is a 980-nm solid-state laser diode that feeds the pump power through a directional coupler to the main cable. The directional coupler is a pair of lightwave waveguides placed in proximity to each other so that the lightwaves can couple with one waveguide to another. The end of the lightwave guide for the pump laser that is the primary waveguide of the directional coupler is reflectionlessly terminated using a lightwave-absorbing component. The pump laser light is fed into the main lightwave waveguide, which is an erbium-doped optical fiber. The pump laser light excites or “pumps up” the atoms of erbium to prepare for emission of radiation at 1500 nm. When these pumped-up erbium atoms receive stimulating radiation at 1500 nm by the input lightwaves, they emit radiation at the same wavelength 1500 nm. This is a laser amplifier. The emission of radiation continues as the input lightwave travels in the erbium-doped optical fiber. The emitted wave travels together with the stimulating lightwave. The amplified lightwave exits from the output connector. The lightwave gain of 15 dB is reported for several-meters-long erbium-doped plastic optical fiber cable.

The amplifier cable can be a praseodymium-doped fluoride fiberglass cable for a wavelength of 1300 nm [3]. A gain of 40 dB for several-meters-long cable has been reported [3].

### 8. CONTINUOUS DISTRIBUTED AMPLIFIERS

#### 8.1. Continuous Active-Diode Distributed Amplifiers

In Fig. 5, by removing all discrete diodes and using an intrinsic semiconductor substrate instead of the dielectric substrate, it is possible to monolithically develop a continuous tunnel diode junction or Gunn effect diode contact between the two strips of metallization. Then, if the conducting strips form a coplanar waveguide that is properly biased at the negative resistance of the diode, the microwave waves fed into one end of the coplanar waveguide will be amplified by the distributed negative resistance as it travels along the coplanar waveguide, and the amplified microwave will exit from the other end of the coplanar waveguide.

**Figure 15.** A schematic diagram of a microwave monolithic distributed amplifier. Both the gate and drain lines are microstriplines. The gate line feeds the FET sequentially. On the drain microstripline, the amplified signals are sequentially combined and propagate out at right.
8.2. Continuous Transistor Distributed Amplifiers

A conceptual diagram of a continuous transistor distributed amplifier is shown in Fig. 6. The length of the microstrips must exceed several wavelengths of the transmission-line wavelength. The transmission-line wavelength on the coplanar waveguide is smaller than the free-space wavelength and is inversely proportional to the square root of the effective relative permittivity of the substrate in the gap between the conducting strips. For a semiconductor substrate, it is not uncommon for the effective relative permittivity to be 10 or higher.

8.3. Continuous Parametric Varactor Diode Distributed Amplifier

A conceptual schematic diagram of a continuous parametric varactor diode distributed amplifier is shown in Fig. 3b. Here, the concept is the same and as that for an MMIC amplifier, but an alternate method and a more convenient approach to MMIC technology is shown. Instead of using the microstripline as shown in Fig. 3b, the configuration is changed to a coplanar waveguide as shown in Fig. 5. A long junction varactor diode is monolithically developed flatly between the gap of long parallel metallization strips of a coplanar waveguide.

8.4. Continuous Ferrimagnetic Distributed Amplifiers

The space between the long gap of metallization strips of either the microstripline as shown in Fig. 3b or the coplanar waveguide as shown in Fig. 5 can be filled with a magnetized ferrimagnetic material or a ferrite. This constitutes a ferrimagnetic continuous distributed amplifier [1,10]. The nonlinear magnetism of a ferrite renders the system as a variable-inductance parametric amplifier when both the pump oscillator power and the signal power are launched into a same transmission line; then the pump oscillator power gradually transfers into the signals through the distributed nonlinear inductance of the ferrites as both signals and the pump oscillator power travel together along the ferrite-loaded transmission line.

9. TRENDS OF DISTRIBUTED AMPLIFIER TECHNOLOGY

One of the most common concepts of a distributed amplifier consists in a number of amplifier units distributed along a transmission line [6,7,16,20–22]. An alternate approach to obtain the wide frequency bandwidth and higher gain is to use the concept of matrix amplifiers [27]. A generic concept of a matrix amplifier is illustrated in Fig. 16. As seen from this figure, a matrix amplifier consists of several tiers of distributed amplifiers. Each tier is connected by bleeder lines. The bleeder lines connect the
output points of each (amplification) stage of a distributed amplifier and the input points of each stages of the next tier distributed amplifier. Thus, when the signal travels down the first tier of the distributed amplifier, each stage of the amplifiers of the second tier are successively excited. This process is repeated until the output signal reaches the output point of the final stage of the amplifier unit in the final tier. In a matrix amplifier, a part of the signal travels in a longitudinal direction along the direction of the transmission line of each distributed amplifier. The remainder of the signal travels in a transverse direction along the amplifier tiers through bleeder lines.

In a matrix amplifier, the amplifiers in each tier need not be identical in type. This feature gives design flexibility for the required performance of an amplifier.

Another, unorthodox, type of distributed amplifier is the transversely distributed amplifier. In a transversely distributed amplifier a number of amplifier units are distributed on a screen that is oriented in transversely to the direction of the incoming microwave propagation. Incoming microwave signals are amplified by the amplifier units distributed throughout the screen and reradiated from the other side of the screen. This structure may be considered simply as an array of amplifiers rather than a distributed amplifier. But the spacing among amplifier units is comparable in distance to the operating wavelengths for proper phasing. Therefore, the transversely distributed amplifiers are distributed amplifiers of a special kind.

A schematic diagram of a generic transversely distributed amplifier is shown in Fig. 17. Incoming microwave or RF signals are amplified by a properly distributed amplifier on a screen and reradiated to space in the other side of the screen in the required direction [28–30]. Usually the transversely distributed amplifier is placed at a midpoint in a microwave transmission beam between the transmitter and the receiver for the microwave signal booster [28], with microwave signal booster and microwave beam steering [29], or in accordance with microwave distribution analysis within the beam [30]. Usually, the direction of the output reradiation is the same as or collinear with that of the incoming radiation. But the direction of the output reradiation can be at any desired direction by proper phasing of the distributed amplifier units.

Research and development trends for the traditional distributed amplifier are pushing the already extremely broad frequency band even wider [24], in a search for a new design approach to obtain optimal amplifier performance [26,31,32], and for ultra-high-speed digital signal-processing amplification [33]. For example, an ultrabroadband distributed amplifier designed to cover 100 MHz–20 GHz has been reported by Virdee and Birdee [24]. In MMIC technology, beside the traditional silicon, gallium nitride [26] and indium phosphide [32] technologies are developing and producing good results. In the digital signal processing, a distributed amplifier capable of handling 40 Gbps has been reported by Shigematsu et al. [33].

Today, the technology of conventional distributed amplifiers is a mature one. Commercial products of distributed amplifiers are available on the market [34,35].

10. CONCLUSION

Distributed amplifiers are electrical transmission lines with periodically or continuously loaded amplifiers. The main feature of the distributed amplifier is the wide frequency bandwidth. Wide-bandwidth amplifiers have a large channel capacity and are also capable of handling extremely short or fast pulses. Distributed amplifiers are useful for fast digital data transmission systems of gigabit rates. Distributed amplifiers can be made compact by the use of the monolithic integrated-circuit technology.

BIBLIOGRAPHY

DUAL- AND MULTI-FREQUENCY MICROSTRIP ANTENNAS

1. INTRODUCTION

There are many applications in wireless communications that involve two or more distinct frequencies. Because of the attractive features of the microstrip patch antenna, such as planar profile, ruggedness, and low cost, there has been considerable interest in the development of these antennas to meet the dual- or multifrequency specifications. It is sometimes possible that a broadband microstrip antenna can cover the frequencies of interest. However, the disadvantage of using a broadband antenna is that it also receives nondesired frequencies unless some kind of filtering network is introduced to reject such frequencies. On the other hand, the advantage of a dual- or multifrequency design is that it focuses only on the frequencies of interest and is thus more desirable. It is the purpose of this article to present several such designs. Dual-frequency designs will be emphasized since they are much more developed than multifrequency designs, which have been introduced by the present authors using the proximity effects. It is sometimes possible that a broadband microstrip antenna can cover the frequencies of interest. However, the disadvantage of using a broadband antenna is that it also receives nondesired frequencies unless some kind of filtering network is introduced to reject such frequencies. On the other hand, the advantage of a dual- or multifrequency design is that it focuses only on the frequencies of interest and is thus more desirable. It is the purpose of this article to present several such designs. Dual-frequency designs will be emphasized since they are much more developed than multifrequency designs, which have become a topic of current interest.

In Section 2, dual-frequency single and stacked patches are reviewed. Section 3 presents dual-frequency designs obtained by loading the patch with either a reactive load or with slots. Section 4 describes a more recent design introduced by the present authors using the proximity effects.
coupling of a U-shaped patch with unequal arms and a wideband E-shaped patch. Dual-frequency wideband patches using a L-probe feed, also introduced by the present authors, are presented in Section 5. Section 6 discusses two methods of designing multi-frequency patch antennas. The article ends with some concluding remarks.

2. DUAL-FREQUENCY SINGLE AND STACKED PATCHES

2.1. Single-Element Dual-Frequency Microstrip Antennas

It is possible for a single-element microstrip antenna to operate at many frequencies corresponding to the various resonant modes pertaining to the structure. However, for most applications, it is required that the radiation pattern, the polarization, and the impedance be similar, if not identical, in all the frequency bands of operation. This immediately rules out many modes. Furthermore, for a given geometry, all the resonant frequencies are related in fixed ratios.

If for a particular patch shape two modes can be found that produce similar radiation patterns with the same polarization, dual frequency is possible with a single patch.

For the rectangular patch, the two modes TM01 and TM03 satisfy this requirement. However, their resonant frequencies are related by a ratio of approximately 3; the exact value is dependent on the edge effect. It has been

Table 1. Resonant Frequencies of the TM01 and TM03 Modes against Short-circuiting Pins Used

<table>
<thead>
<tr>
<th>Number of Pins</th>
<th>Pin Position</th>
<th>f01 (MHz)</th>
<th>f03 (MHz)</th>
<th>f03/f01</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>—</td>
<td>613</td>
<td>1861</td>
<td>3.04</td>
</tr>
<tr>
<td>1</td>
<td>(1)</td>
<td>664</td>
<td>1874</td>
<td>2.82</td>
</tr>
<tr>
<td>2</td>
<td>(1)(2)</td>
<td>706</td>
<td>1865</td>
<td>2.64</td>
</tr>
<tr>
<td>3</td>
<td>(1)(2)(3)</td>
<td>792</td>
<td>1865</td>
<td>2.36</td>
</tr>
<tr>
<td>4</td>
<td>(1)(2)(3)(6)</td>
<td>813</td>
<td>1865</td>
<td>2.29</td>
</tr>
<tr>
<td>5</td>
<td>(1)(2)(3)(5)(6)</td>
<td>846</td>
<td>1865</td>
<td>2.20</td>
</tr>
<tr>
<td>6</td>
<td>(1) to (6)</td>
<td>891</td>
<td>1865</td>
<td>2.09</td>
</tr>
</tbody>
</table>

Source: From Zhong and Lo [1], reproduced with permission from IEE.
shown that placement of short-circuiting pins on the nodal lines of the TM₀₃ mode field has little effect on this mode but a strong effect on the TM₀₁ mode. Zhong and Lo [1] demonstrated that this offers a way of altering not only the separation of the two frequency bands but also the input impedance of the TM₀₁ mode. The geometry of the rectangular patch in their experiment is shown in Fig. 1. It is made of 0.062-in. copper-cladded Rexolite 2200 with six short-circuiting pin positions. The effects of successively adding more and more pins (each approximately 0.05 cm in diameter) at the positions indicated in Fig. 1 are shown in Table 1. It is seen that the ratio of the two operating frequencies \( f₀₃/f₀₁ \) can be varied from approximately 3 to 2. Since all these pins are located on the TM₀₃ mode null lines, \( f₀₃ \) remains constant at about 1865 MHz while \( f₀₁ \) is varied from 613 to 891 MHz. In order for the impedances of the two bands to be close to 50 \( \Omega \) at resonance, it is necessary to attach a short capacitive stub of 0.6 \( \times \) 2.1 cm. With the stub added, the bandwidth with reference to a standing-wave ratio (SWR) of 3:1 is about 2% for the lowband and almost 8% for the highband. Typical low- and highband patterns in both E and H planes are shown in Fig. 2. It is seen that, while the two modes radiate strongest in the broadside, the directivities of the two modes are quite different.

The rectangular patch is not the only geometry capable of providing dual-frequency operation. It will be shown in Section 6 that the TM₀₁, TM₂₀, and TM₂₁ modes of the equiangular patch are all broadside modes with similar polarizations in the broadside direction (refer to Fig. 31 for a detailed radiation pattern). By choosing the location of the feed properly, the impedances of these modes do not vary greatly. Section 6 discusses the utilization of the equiangular patch for dual- or triple-frequency operations.

2.2. Dual-Frequency Stacked Patches

By using a stacked geometry consisting of two patches on two substrate layers, dual-frequency operation is obtained. The first experiment, reported by Long and Walton [2], utilized two stacked circular patches. The geometry is shown in Fig. 3.

The disks were phototetched on separate substrates and aligned so that their centers were along the same line. The sizes of the two disks and their spacings were varied and the resultant behavior of the antenna characteristics measured. The antenna was fed by means of a coaxial line. The center conductor passes through a clearance hole in the lower disk and is connected electrically to the upper disk. If one considers the two regions under the patch as two resonant cavities, it is clear that the system behaves as a pair of coupled cavities. Since the fringing fields are...
Figure 6. Radiation pattern of stacked circular patches antenna: (a) \( f = 2.83 \text{ GHz} \); (b) \( f = 3.1 \text{ GHz} \); (c) \( f = 2.9 \text{ GHz} \). (From Long and Walton [2], © 1979 IEEE.)

Figure 7. Tunable dual-frequency stacked microstrip antenna utilizing the airgap concept. (From Dahele and Lee [3], © 1982 IEEE.)

Figure 8. Geometry of the stacked annular ring antenna. (From Dahele et al. [4], © 1987 IEEE.)

Figure 9. Dual-frequency rectangular patch antenna with monolithic reactive loading. (From Davidson et al. [6], reproduced with permission from IEE.)

Figure 10. Impedance of edge-loaded, 4 \times 6\text{ cm}-patch antenna with \( L = 4.0 \text{ cm}, W = 0.33 \text{ cm}, \varepsilon_r = 2.17, t = 0.079 \text{ cm}; \) coaxially fed near the edge end at the center of the 6-cm side. (From Davidson et al. [6], reproduced with permission from IEE.)
different for the upper and lower cavities, two resonant frequencies are expected even if the diameters of the two disks are the same. Long and Walton [2] showed that,

![Figure 11](image1.png)

**Figure 11.** Geometry incorporating an insert dimension $S$ and a gap spacing $G$. (from Davidson et al. [6], reproduced with permission from IEE.)

<table>
<thead>
<tr>
<th>$W$ (cm)</th>
<th>$L$ (cm)</th>
<th>$G$ (cm)</th>
<th>$S$ (cm)</th>
<th>$f_L$ (GHz)</th>
<th>$f_U$ (GHz)</th>
<th>$f_U/f_L$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.33</td>
<td>4.0</td>
<td>1.0</td>
<td>1.5</td>
<td>2.356</td>
<td>2.494</td>
<td>1.054</td>
</tr>
<tr>
<td>0.33</td>
<td>4.0</td>
<td>0</td>
<td>0</td>
<td>2.275</td>
<td>2.666</td>
<td>1.172</td>
</tr>
<tr>
<td>0.33</td>
<td>8.4</td>
<td>0</td>
<td>0</td>
<td>2.339</td>
<td>2.628</td>
<td>1.124</td>
</tr>
<tr>
<td>0.33</td>
<td>4.0</td>
<td>0.7</td>
<td>1.5</td>
<td>2.437</td>
<td>2.494</td>
<td>1.023</td>
</tr>
<tr>
<td>0.33</td>
<td>4.0</td>
<td>0.3</td>
<td>1.5</td>
<td>2.471</td>
<td>2.514</td>
<td>1.017</td>
</tr>
</tbody>
</table>

*Source: From Davidson et al. [6], reproduced with permission from IEE.*

![Figure 12](image2.png)

**Figure 12.** Geometry of probe feed SPA. (From Maci et al. [7], reproduced with permission from IEE.)

![Figure 13](image3.png)

**Figure 13.** Current distributions on slotted patch ($W=40$ mm; $L=30$ mm; $d=1$ mm; $L_s=28$ mm; $W_s=36$ mm): (a) TM$_{10}$ mode; (b) TM$_{30}$ mode. (From Maci et al. [7], reproduced with permission from IEE.)

different for the upper and lower cavities, two resonant frequencies are expected even if the diameters of the two disks are the same. Long and Walton [2] showed that,

![Figure 14](image4.png)

**Figure 14.** Return loss of SPA without tuning resonators fed by 50-Ω coaxial probe ($W=15.5$ mm; $L=11.5$; $t=0.5$ mm; $w=1$ mm; $d=1$ mm; $W_p=5.5$ mm; $\varepsilon_r=2.2$; $t=0.8$ mm) (— experimental data; —— MoM analysis). (From Maci et al. [7], reproduced with permission from IEE.)
unlike the single-element dual-frequency rectangular patch operated in the TM$_{01}$ and TM$_{03}$ modes, in which the frequency ratio is approximately 3 when no short-circuiting pins are used, this design can produce resonant frequencies that are considerably closer. A typical frequency ratio is 1.2; the exact value is determined by the relative diameters of the upper and lower disks, the thicknesses of the substrate layers, and the relative permittivities of the substrates.

Figure 4 shows the real and imaginary parts of the input impedance for $2a_2 = 3.78$ cm, $t_1 = t_2 = 0.075$ cm, and three values of $2a_1$. The relative permittivity of both layers is 2.47. The resonant frequencies as a function of the upper disk diameter are shown in Fig. 5. Also shown is the resonant frequency of the lowest mode for a single disk of diameter $2a$ and substrate thickness $t = 0.075$ cm, taking into account the fringing field through the effective diameter. It is seen that the lower resonant frequency is relatively constant, remaining near the value of a single disk with $2a = 3.78$ cm and $d = 0.075$ cm. The upper resonance, on the other hand, is highly dependent on the size of the upper disk. Radiation patterns were also given in Ref. 2, which showed that they were similar to the radiation pattern of the lowest mode of the single circular patch. The radiation pattern is shown in Fig. 6.

Dahele and Lee [3] have applied the airgap concept to the design of dual-frequency stacked disks. The geometry is shown in Fig. 7, in which an airgap between the two substrates is introduced. By increasing the width of the airgap, the upper resonant frequency increases, accompanied by a broadening of the bandwidth of the lower resonance.

Dahele et al. [4] also studied a structure consisting of two stacked annular ring patches as shown in Fig. 8. This structure was also found to exhibit dual-frequency behavior. As in the case of circular disks, an upper airgap was found to be a convenient method of altering the separation between the frequency bands. A theoretical analysis of the stacked circular disks and stacked annular rings using Hankel transforms was given by Fan and Lee [5]. The predicted resonant frequencies showed excellent agreement with the measurements in Refs. 2 and 4.

In the two designs discussed in this section, the stacked geometry is suitable when the frequency ratio is in the range of 1.1–1.2, while the single element incorporating short-circuiting pins yields frequency ratios in the range of 2–3. In the next section, two additional designs are presented: patch with reactive loading and patch with slots. The former also yields similar frequency separation as the stacked patches, while loading a rectangular patch with slots can reduce the frequency ratio from about 3 to as low as 1.57.
3. DUAL-FREQUENCY LOADED PATCHES

3.1. Patch with Reactive Loading

A dual-frequency microstrip antenna can be obtained by loading it with a reactive load, such as a short-circuited coaxial line or a short-circuited microstrip line. The latter preserves the low-profile characteristics of the microstrip patch antenna. Such a structure, shown in Fig. 9, was demonstrated experimentally by Davidson et al. [6]. In the experiment, a rectangular patch of dimensions $6 \times 4$ cm etched on a substrate with $\varepsilon_r = 2.17$ and thickness $\frac{e}{c} = 0.1$.

![Figure 17](image_url)

**Figure 17.** Measured return loss for different short-circuited plane widths (--- $s/a = 1$; · · · $s/a = 0.5$; · · · · $s/a = 0.25$; · · · · · $s/a = 0.1$). (From Euo et al. [9], reproduced with permission from IEE.)

![Figure 18](image_url)

**Figure 18.** Measured field patterns: (a) $s/a = 1$, $f_{10} = 2.355$ GHz; (b) $s/a = 1$, $f_{30} = 3.785$ GHz; (c) $s/a = 0.1$, $f_{10} = 2.355$ GHz; (d) $s/a = 0.1$, $f_{30} = 3.785$ GHz. (--- $E$ copolarization; --- · · · · $E$ cross-polarization; --- $H$ copolarization; · · · · $H$ cross-polarization). (From Guo et al. [9], reproduced with permission from IEE.)
0.079 cm. It is coaxially fed near the edge and at the center of the 6-cm side. For a linelength of \( L = 4.0 \) cm and linewidth \( w = 0.33 \) cm, the measured real and imaginary parts of the input impedance are as shown in Fig. 10. Good pattern characteristics were observed at each of the resonant frequencies, 2.275 and 2.666 GHz, respectively.

The separation of the resonances can be varied by (1) changing the length of the microstrip line and (2) introducing an inset dimension \( S \) with an accompanied gap

---

**Figure 19.** Dual-band patch antenna at 900 and 1800 MHz with \( h = 11 \) mm. (From Guo et al. [10], © 2002 IEEE.)

**Figure 20.** Measured SWR of the antenna in Fig. 19 (From [10], © 2002 IEEE.)

---

**Figure 21.** Radiation patterns of the antenna in Fig. 19, 10 dB/div: (a) \( E \) plane at 900 MHz; (b) \( H \) plane at 900 MHz; (c) \( E \) plane at 1800 MHz; (d) \( H \) plane at 1800 MHz (From Guo et al. [10], © 2002 IEEE.)
spacing $G$ between the line and the patch, as shown in Fig. 11. The results for the resonant frequencies are shown in Table 2. It is seen that the frequency ratio varies between 1.017 and 1.172.

3.2. Patch with Slots

A dual-band rectangular patch antenna with a frequency ratio of $<2$ can be obtained by loading the patch with two narrow slots etched inside the patch, which are close to and parallel to the radiating edges [7]. Such a slotted patch antenna (SPA) is shown in Fig. 12.

![Image of the basic L-probe dual-band antenna](image)

**Figure 22.** Geometry of the basic L-probe dual-band antenna. (From Luk et al. [15], reproduced with permission from IEE.)

![Image of the measured return loss S11](image)

**Figure 23.** Measured return loss $S_{11}$. (From Luk et al. [15], reproduced with permission from IEE.)

3.2. Patch with Slots

A dual-band rectangular patch antenna with a frequency ratio of $<2$ can be obtained by loading the patch with two narrow slots etched inside the patch, which are close to and parallel to the radiating edges [7]. Such a slotted patch antenna (SPA) is shown in Fig. 12.

![Image of the measured return loss S11](image)

**Figure 24.** Radiation pattern: (a) $E$ plane at 953.5 MHz; (b) $H$ plane at 953.5 MHz; (c) $E$ plane at 1.7855 GHz; (d) $H$ plane at 1.7855 GHz (--- copolarization; ..., copolarization; .... copolarization without slots—for (c) only). (From Luk et al. [15], reproduced with permission from IEE.) (This figure is available in full color at http://www.mrw.interscience.wiley.com/erfme.)
The reduction in frequency separation can be explained as arising from the perturbation of the TM$_{10}$ and TM$_{30}$ modes, as shown in Figs. 13a and 13b. These narrow slots are placed close to the radiating edges, where the current is nearly minimum for the TM$_{10}$ mode. Consequently, the current distribution for this mode is only slightly perturbed. Hence, its resonant frequency is only slightly different from that of the rectangular patch without a slot. For the TM$_{30}$ mode, slots are located where the current of the unperturbed TM$_{30}$ mode is significant, leading to

**Figure 25.** Geometry of L-probe dual-band antenna with DC ground.

**Figure 26.** Measured standing-wave ratio and gain of the dual-band antenna. (This figure is available in full color at http://www.mrw.interscience.wiley.com/erfme.)
strong modification of the current distribution. As a result, the resonance frequency decreases, due to the increase in the current pathlength introduced by the slot.

Maci et al. [7] performed numerical analysis based on the method of moments (MoM) and fabricated several prototypes at C and X bands. Design criteria were obtained from the MoM analysis. Figure 14 shows the amplitude of the reflection coefficient for a patch with the dimensions shown in the caption. Two resonances were obtained at \( f_{10} = 5.515 \text{ GHz} \) and \( f_{30} = 10.446 \text{ GHz} \). The frequency ratio was 1.89. The experimental curve (continuous line) compares well with the results obtained by MoM analysis (dashed line). The central resonance of lower amplitude that appears in the Fig. 14 is associated with the TM\(_{20}\) mode that is excited in the structure owing to the asymmetric feed. The measured radiation patterns of this antenna in the E and H planes are shown in Fig. 15. The dashed and continuous lines correspond to the lower and the upper frequencies, respectively. A gain of 6.5 dB was found for the lower frequency and 6.8 dB for the upper frequency. The E-plane radiation pattern at the upper frequency is broader than that for the lower frequency.

Maci et al. also showed that, by introducing resonant tuning stubs, the frequency ratio can be further reduced to about 1.57 [7].

### 3.3. Slot-Loaded Short-Circuited Patch Antenna

It is well known that the size of a rectangular patch can be reduced by short-circuiting along the vertical central axis,

### Table 3. Summary of Radiation Characteristics

<table>
<thead>
<tr>
<th></th>
<th>Vertical Plane</th>
<th></th>
<th>Horizontal Plane</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency (GHz)</td>
<td>3 dB Beamwidth</td>
<td>Cross-polarization (dB)</td>
<td>Backlobe (dB)</td>
<td>3 dB Beamwidth</td>
</tr>
<tr>
<td>Lower band</td>
<td>0.9</td>
<td>56.3°</td>
<td>−13</td>
<td>−12</td>
</tr>
<tr>
<td>Upper band</td>
<td>1.8</td>
<td>40.4°</td>
<td>−12</td>
<td>−15</td>
</tr>
</tbody>
</table>
which corresponds to the zero-potential plane for the TM$_{10}$ and TM$_{30}$ modes, and by using only one-half of the patch. The size of the antenna can be further reduced by decreasing the short-circuiting plane width [8]. Guo et al. [9] showed that this technique can be applied to reduce the size of the dual-frequency slot-loaded patch antenna.

The geometry of this antenna is shown in Fig. 16. A single slot with dimensions $L_s \times W_s$ is cut in a rectangular patch with dimensions $a \times b$ with a short-circuited plane with width $s$ placed at its other side. The patch is separated from the ground plane by a foam substrate of thickness $h$. The slot is placed a small distance $d_s$ from the radiating edge of the patch. The parameters of the slot were selected using the design criteria given in Ref. 7. It is found that, by controlling the short-circuiting plane width, both the TM$_{10}$ and TM$_{30}$ modes are strongly perturbed.

Many designs of the proposed antenna with various short-circuiting plane widths were fabricated and measured by Guo et al. [9]. Figure 17 shows typical results of the measured return loss for the cases with $s/a = 1$, 0.5, 0.25, and 0.1 for the dimensions shown in the caption. It is seen that the perturbed TM$_{10}$ and TM$_{30}$ modes are excited with good impedance matching. However, when $s/a$ is $< 0.1$, no feedpoint can be found for exciting the two frequencies with good matching. This indicates that there are limitations to the present dual-band design. It is seen that the frequency ratio $f_{30}/f_{10}$ varies in the range $\approx 1.6 - 2.2$. For the case $s/a = 0.1$, the frequency $f_{10}$ occurred at 1.535 GHz, which is about 0.29 times the frequency $f_{10}$ (5.35 GHz) for a regular half-wavelength patch of the same patch size. The partial shorting plane therefore significantly reduces the resonant length of the patch.

Typical measured far-field radiation patterns at the two operating frequencies for the case with $s/a = 1$ and $s/a = 0.1$ are shown in Fig. 18. For the experiment, the shorted patch was mounted on a large ground plane (a circular disk with diameter $= 3d_0$) to reduce diffraction off the edges. Similar radiation patterns and polarizations for the two operating frequencies $f_{10}$ and $f_{30}$ are observed. It can be seen that for $s/a = 1$, there is a mainbeam squint of about 32° and 43° in the $E$ plane for $f_{10}$ and $f_{30}$, respectively. For $s/a = 0.1$, the corresponding mainbeam squint is about 30° and 6°. For $s/a = 1$, the measured gain in the maximum direction is 2.2 and 2.75 dBi for $f_{10}$ and $f_{30}$, respectively, while for $s/a = 0.1$, the corresponding gain is 0.4 and 1.9 dBi. It should be noted that although the $H$-plane cross-polarization level is quite high, it may not be a disadvantage in some applications, such as indoor mobile communications.

4. DUAL-FREQUENCY COUPLED PATCHES

In this section, we present a dual-frequency design that makes use of the concept of a low-frequency (outer) patch coupled to a high-frequency (inner) patch [10]. The low-frequency element is a U-shaped patch with unequal arms, while the high-frequency element is a E-shaped patch [11], which is a variation of the wideband U-slot patch [12]. The geometry is shown in Fig. 19. The inner patch is fed by a coaxial probe, and the outer patch is excited through proximity coupling to the inner patch.

It is noted that the two arms of the U patch are not equal. This introduces two closely spaced resonances at the lower frequency, making the bandwidth of the low-band wider than the case when the arms are of equal length. Moreover, a short-circuiting wall is introduced in the U patch to reduce the size of the patch. An antenna of this type was designed to operate at both 900 and 1800 MHz, using FR4 substrate material with relative permittivity equal to 4.4 [10]. It has the following parameters: $W = 40 \text{ mm}$, $W_1 = 5 \text{ mm}$, $W_2 = 5 \text{ mm}$, $W_3 = 16 \text{ mm}$, $L_1 = 45 \text{ mm}$, $L_2 = 50 \text{ mm}$, $L_3 = 42.5 \text{ mm}$, $L_4 = 42.5 \text{ mm}$, $a = 4 \text{ mm}$, $b = 2 \text{ mm}$, $c = 44 \text{ mm}$, $d_1 = 2 \text{ mm}$, $d_2 = 2 \text{ mm}$, $e_1 = 3.5 \text{ mm}$, $e_2 = 3.5 \text{ mm}$, $s_1 = 1.5 \text{ mm}$, $s_2 = 1.5 \text{ mm}$, $f = 41 \text{ mm}$, $g = 10 \text{ mm}$, and $h = 11 \text{ mm}$.

The measured SWR of the dual-band antenna is shown in Fig. 20. The bandwidth of the lower-frequency element is 13.3% with a frequency range of 807–922 MHz, and that of the upper band is 20% with a frequency range of 1625–1892 MHz when SWR = 2. The substrate thickness corresponds to 0.033 free-space wavelength at 900 MHz, while the length $L_2$ and the width $W$ are 0.15$d_0$ and 0.12$d_0$, respectively. Figure 21 shows the copolarization and cross-polarization $E$- and $H$-plane radiation patterns at 900 and 1800 MHz, respectively. At 900 MHz, the 3 dB beamwidth in the $E$ plane is 109° and in the $H$ plane, 127°. At 1800 MHz, the 3 dB beamwidth in the $E$ plane is 72° and

Figure 28. Geometry of equilateral triangular patch antenna. (From Lee et al. [16], © 1988 IEEE.)
in the $H$ plane, 73°. The beamwidth in the lower band is larger than that in the upper band. The measured gain of the antenna was $2 \pm 1.5$ dBi and $4.5 \pm 1.8$ dBi at the lower and upper bands, respectively. The measured results confirm that the lower band is due to the short-circuited U-shaped patch with two unequal arms and the upper one is due to the inner coaxially fed E-shaped patch. Note that the cross-polarization was significantly higher than with the single-patch antenna. However, for indoor mobile communications, this would lead to better transmission capabilities in a multipath environment.

5. DUAL-FREQUENCY WIDEBAND L-PROBE PATCH

5.1. Basic L-Probe Dual-Band Antenna

The L-probe patch [13,14] is a wideband patch antenna. This technique can be further extended to dual-band operation [15]. For ease of operation and to reduce cost, the antenna should have only one input for both frequency bands. In this section, the design of a dual-band patch antenna excited by two L probes through one feed is described.

The geometry of a dual-band patch antenna [15] consisting of two L probes is shown in Fig. 22. The antenna operates at 900 MHz ($\lambda_1$, lower-band operation) and 1.8 GHz ($\lambda_2$, upper-band operation). The probes are combined together to form a feed structure. The lower-band patch has length $l_1 = 102$ mm (0.312$\lambda_1$) and width, $w_1 = 110$ mm (0.336$\lambda_1$) and is 45.4 mm (0.139$\lambda_1$) above the ground. Two slots with width $= 2$ mm and length $= 90$ mm are etched $l_1$ away from the radiation edge of the lower-band patch. These two slots are used to suppress the excitation of the TM $20$ mode that would influence the upper-band radiation pattern. The upper-band patch with sides $l_2 = 37$ mm (0.227$\lambda_1$) is 21.5 mm (0.128$\lambda_1$) above the ground. Because of the coupling effect of the lower-band patch, the resonant length of this patch is less than that of an isolated patch operated at the same frequency.

The measured return loss $S_{11}$ is as shown in Fig. 23. The impedance bandwidth ($S_{11} < -10$) of 20.8 and 17.9% was found for the lower and upper bands, respectively. It is wide enough to cover GSM 900 and 1800 cellular phone systems. The maximum gain of 8.4 dBi was found in the upper band. The measured radiation patterns are shown in Fig. 24. The 3 dB beamwidths of lower and upper bands are 71° and 75° in the $H$ plane and are 56° and 60°, in the $E$ plane. The cross-polarization is $-10$ and $-13$ dB in the lower and upper bands, respectively.
The L-probe dual-band patch antenna can be modified to have a DC grounded feature, which protects the antenna against the lightning hazard. This is accomplished by adding a short-circuit stub in the feeding network of the antenna. Such a design is presented in this section.

The geometry of the antenna, which operates at both 900 MHz (\(f_1\)) (lower band) and 1800 MHz (\(f_2\)) (upper band), is shown in Fig. 25. The antenna consists of two rectangular patches of different sizes. The thickness of each aluminum patch is 1 mm. The lower-band patch (patch L) with dimensions 100 mm (0.33\(\lambda_1\)) \times 15 mm (0.35\(\lambda_1\)) is supported 45.37 mm above the ground by plastic pins. The height and length of the larger probe are 29.265 and 57.8 mm, respectively, and the antenna is located 55.55 mm (about \(1/4\lambda_1\)) away from the short-circuiting pin. In order to suppress the TM\(_{20}\) mode of the larger patch, two slots with sides 80 \times 2 mm are etched on the patch. The upper-band patch (patch U) with dimensions 63.5 mm (0.381\(\lambda_2\)) \times 54 mm (0.324\(\lambda_2\)) is located 22.9 mm (0.137\(\lambda_2\)) above the ground. The height and length of the smaller probe are 15.22 and 32.44 mm, respectively, and the probe is located at 141.38 mm (about 0.75\(\lambda_2\)) away from the short-circuiting pin. The thickness and dielectric constant of the microwave substrate are, respectively 1.5 mm and 2.65. The characteristic impedance transmission line is 50\(\Omega\). The size of the ground plane is 200 \times 275 mm.

Figure 26 shows the measured standing-wave ratio (SWR) and gain of the dual-band patch antenna. For SWR < 2, the impedance bandwidth of 27.5% was found for lower-band operation and 14.1% for the upper-band operation. The results show the antenna is wide enough to cover the GSM 900 and 1800 mobile phone systems. An average gain of 8 dBi was found for the lower band and 7.5 dBi for the upper band. The measured radiation patterns of the antenna are shown in Fig. 27. The characteristics of the vertical and horizontal planes are summarized in Table 3.
6. MULTIFREQUENCY PATCH ANTENNAS

6.1. Single-Element Triple-Frequency Triangular Patch Antennas

As mentioned in Section 2.1, it is possible for a single-element patch antenna to operate at several frequency bands corresponding to the various resonant modes pertaining to the patch structure. For dual-band operation, we can simply select the rectangular patch, which can be operated at the TM$_{01}$ and TM$_{03}$ broadside modes with a single feed. For triband operation, other patch shapes have to be considered.

From the cavity model analysis of equilateral triangular patch antennas [16], it appears that we can excite the three broadside modes, namely, TM$_{10}$, TM$_{20}$, and TM$_{21}$.
modes, with a single feed. In particular, if we select the length of each edge of an equilateral triangular patch to be 10 cm, as shown in Fig. 28, we can ensure that the antenna matches at the three modes by choosing the feed position \( g = 50 \text{ mm} \) (Fig. 29) from an edge of the triangle.

Using the IE3D\(^2\) simulation tool, we have analyzed the performance of the triangular patch antenna. The gain, SWR, input impedance, and radiation pattern are shown in Figs. 30 and 31. It can be observed that the antenna does resonate at the TM\(_{10}\), TM\(_{20}\), and TM\(_{21}\) modes with resonant frequencies of 1299, 2599, and 3499 MHz, respectively. The three modes basically have broadside radiation patterns.

### 6.2. Multifrequency Stub-Loaded Patch Antennas

As described in Section 3.1, a patch antenna can be operated in dual frequency if the patch is loaded with a single stub. The technique can be generalized for designing patch antennas with triple-band operation. As demonstrated by Daniel and Kumar [17], it is possible to design a square patch antenna with the triband characteristic by loading the two radiating edges with two open-circuited stubs of unequal lengths, as shown in Fig. 32.

To confirm the results presented in Ref. 17, we have employed the IE3D software to simulate the performance of this antenna. The simulated SWRs of the antenna with and without the two stubs are shown in Fig. 33. It can be observed that without the stubs, only one resonant mode is found within the frequency range 1.2–2 GHz. With the addition of the two stubs, three modes are generated. The antenna can now be operated around 1.252, 1.540, and 1.935 GHz. This antenna has a frequency ratio of 1:1.23:1.55, which indicates that the three operating bands are more closely packed than those of triangular patch antenna with multimode operation. The radiation patterns of the antenna are shown in Fig. 34. It can be observed that the three modes radiate strongly in the broadside direction.

\(^{1}\)IE3D is a tradename for the MoM full-wave electromagnetic simulator manufactured by Zeland Software.

### 6.3. Multifrequency Microstrip Antenna Consisting Of Parallel Microstrip Dipoles

Although the two designs presented in Sections 6.1 and 6.2 can be operated at three frequency bands, the gain varies substantially between different bands. In particular, a very low gain is observed at the middle frequency. Also, in these techniques, it is difficult to generate more than three broadside modes with a single feed. Alternatively, multifrequency operation can be realized by using electromagnetically coupled parallel microstrip dipoles of different lengths excited by the aperture-coupling technique [18]. As shown in Fig. 35, six dipoles and the feedline are centered on the coupling aperture. To achieve a uniform excitation of the dipoles, the longest pair of dipoles is located closer to the center of the aperture and the shortest closer to the edges of the aperture. With appropriate selection of dimensions [18], a triple-band patch antenna operated at \( f_1 = 5.3 \text{ GHz} \), \( f_2 = 6.28 \text{ GHz} \), and \( f_3 = 7.19 \text{ GHz} \) was successfully designed. The bandwidth at bands 1, 2, and 3 are respectively 4%, 4.5%, and 2.1%. This is an attractive feature as the difference between the bandwidths at the different frequency bands is small. Moreover, the antenna has similar broadside radiation patterns at the three bands.

### 6.4. Multifrequency Microstrip Antenna Using Multiple Stacked Elements

Another technique has been proposed for the design of a multifrequency patch antenna with similar bandwidth and gain at different frequency bands. As shown in Fig. 36, the antenna consists of a driven patch and four parasitic patches placed underneath the driven patch [19].

The driven patch, which has dimensions of 51 \( \times \) 51 mm, is fed by capacitive coupling in order to mitigate the mismatching due to the probe inductance. It is etched on a thin dielectric substrate of thickness 0.8 mm and dielectric constant 3.38. The parasitic patches are supported by foam layers and are slightly larger than the driven patch. Their dimensions are adjusted for optimum
performance based on simulation. The total thickness of the antenna is 15.6 mm. Measured results of the antenna are listed in Table 4.

It was found that the antenna could be operated at five frequency bands with gains varying from 7 to 9 dB, and bandwidth varying from 1.5% to 5.5%. The operating bands are closely packed in this design (small frequency ratios). Moreover, the radiation patterns at different frequencies are similar and stable. The radiation pattern is shown in Fig. 37.

7. CONCLUDING REMARKS

Many present-day wireless applications demand aesthetically pleasing devices operating at dual- or triple-frequency bands. Examples include tri-band GSM mobile phones, integrated GPS/GSM receivers, and dual-band wireless local-area networks. Because of its low-profile characteristic, multiband microstrip antennas are becoming the most popular choice as an embedded or conformal antenna in modern wireless devices or systems.

Figure 34. Radiation patterns of square patch antenna. (This figure is available in full color at http://www.mrw.interscience.wiley.com/erfme.)
In this article, major techniques available in the literature for designing dual-band and multiband microstrip antennas have been reviewed, with emphasis on the principles of operation and design guidelines. It has been demonstrated that if a large frequency ratio is required, the multimode technique is preferable, whereas if a small frequency ratio is required, the multiple-resonator techniques, including the stacked patch or coplanar patch geometry, are good choices.

**BIBLIOGRAPHY**


**Table 4. Bandwidth and Gain Experimental Data**

<table>
<thead>
<tr>
<th>Band</th>
<th>Frequency (GHz)</th>
<th>Bandwidth (%)</th>
<th>Gain (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.634</td>
<td>5.5</td>
<td>9.2</td>
</tr>
<tr>
<td>2</td>
<td>1.819</td>
<td>2.0</td>
<td>8.4</td>
</tr>
<tr>
<td>3</td>
<td>1.954</td>
<td>2.7</td>
<td>7.6</td>
</tr>
<tr>
<td>4</td>
<td>2.092</td>
<td>1.5</td>
<td>6.9</td>
</tr>
<tr>
<td>5</td>
<td>2.289</td>
<td>5.3</td>
<td>7.3</td>
</tr>
</tbody>
</table>

**Figure 35.** Geometry of microstrip antenna with parallel microstrip dipoles. (From Croq and Pozar [18], © 1992 IEEE.)

**Figure 36.** Geometry of microstrip antenna with stacked elements. (From Anguera et al. [19], © 2003 IEEE.)

**Figure 37.** Radiation pattern of microstrip antenna with stacked elements (From Anguera et al. [19], © 2003 IEEE.)