ACTIVE ANTENNAS

The present article is an introduction to the topic of active antennas. The first section is a description of the field suitable for reading by almost any undergraduate science major. The next section is an in-depth reexamination of the subject, including equations and some derivations. Its basic idea is to provide the reader with enough tools that he or she can evaluate whether it is an active antenna that he or she might need for a specific application. The final section is a discussion of where active antennas are finding and will find application.

We should mention here that, if one really needs to design active antennas, one will need to go further than this article. The set of references to the primary research literature given in this article is by no means complete, nor is it meant to be. A good way to get started on the current literature on this topic would be a reading of the overview monograph of Navarro and Chang (1). We will not cover active amplifiers in this article. However, this topic is treated in the book edited by York and Popović (2).

AN INTRODUCTION TO ACTIVE ANTENNAS

An antenna is a structure that converts electromagnetic energy propagating in free space into voltage and current in an electrical circuit and/or vice versa. In a transceiver system, the antenna is used both to receive and to transmit free-space waves. At minimum, a transceiver then must consist of a signal source that serves to drive the antenna as well as a receiver circuit that reads out the signal from the antenna. Until recently, practically all antenna systems operating in the microwave frequency regime (operation frequencies greater than 1 billion cycles per second, or 1 GHz) were mostly designed to isolate the antenna from the circuits—that is, to find ways to make system operation independent of the antenna's electrical characteristics. In contradistinction, an active antenna is one in which the antenna actually serves as a circuit element of either the driver or the readout circuit. To understand why this is different from conventional antenna driving or readout will require us to take a bit of a historical trip through the last century or so.

Actually, the first antenna was an active one. Heinrich Hertz, back in 1884 (2a), was the first to demonstrate that one could generate radio waves and that they would propagate from a transmitter to a receiver at the speed of light. The apparatus used is schematically depicted in Fig. 1. The idea of the transmitter is that, by discharging an induction coil (a wire looped about a magnetic core such that the composite device can store significant amounts of magnetic energy) into a spark gap, one can generate a current in the 5 mm diameter wire. The voltage in the spark gap induces a current in the wires, which in turn induces a voltage in the wires, and this voltage in turn induces current, so that the voltage and current propagate along the two pieces of the wire

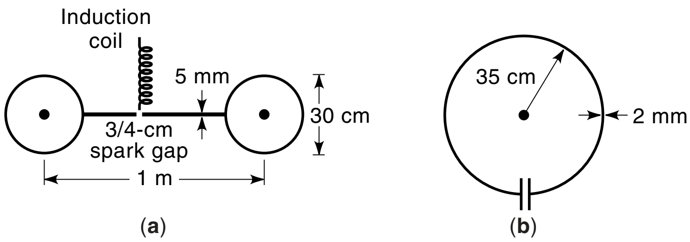


Figure 1. Hertz apparatus for (a) transmitting and (b) receiving radio waves, where the transmitting antenna serves to choose a specific frequency of the spark gap voltage to transmit to the receiving antenna, which also serves to pick out this special frequency from the free-space waveform and turn this electromagnetic disturbance into a voltage across the receiver antenna gap.

to either side of the gap as waves, appearing much like a one-dimensional slice through a water wave propagating away from the point where a pebble has struck the water's surface (the spark gap). A wave will propagate rectilinearly until it encounters an obstruction, at which point it can suffer reflection from or transmission into the barrier that the obstruction presents. There will be reflections then off the metal spheres on the ends of the wire. The spark will generate a broad spectrum of frequencies or wavelengths. The reflections off the two ends, though, will tend to cancel each other except at certain special frequencies. The effect at these wrong frequencies is much like the effect of throwing a handful of pebbles into the pond and noting that, in between the points where the pebbles struck, the waves are much more indistinct than they are far from where the handful struck the surface. The special frequencies are ones which just fit into the region between the spheres. The current needs to be zero at the two ends in order to fit, whereas the voltage needs to be maximum at the ends. The current and voltage waves at the right frequencies may appear as depicted in Fig. 2.

The Hertz transmitter is the archetypical active antenna. The source is the spark gap, which is actually placed in the antenna. The antenna then acts as a filter to pick the right frequency out of a large number of frequencies that could be launched from the gap. The receiver is picked to be of a length to also select this primary frequency.

Hertz-style spark gap transmitters, after further development and popularization by Marconi, were in use for fifty years after Hertz. However, such transmitters exhibit some rather severe drawbacks. The main problem is that the simple resonant dipole antenna (that is, a straight wire antenna with a gap or a feeder cable used to feed in current) is a filter with a poor frequency selection. Namely, if one increases the

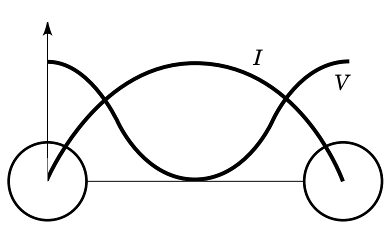


Figure 2. Current and voltage waveforms for the lowest-order (least number of zeros) waveform for the Hertz transmitter of Fig. 1(a). The current must go to zero at the points where the wire ends, whereas the potential will be highest there.

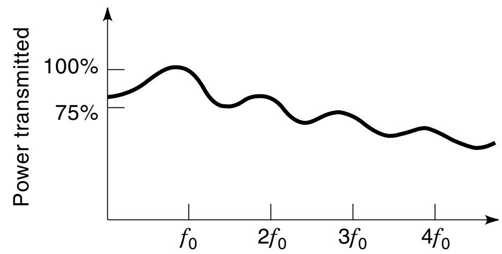


Figure 3. A sketch of what the transmission as a function of frequency might look like for the Hertzian dipole antenna of Figs. 1 and 2.

frequency by 50%, there is 75% as much power transmitted at this frequency as at the first resonance, which is called the *fundamental*. There is a second resonance at twice the frequency of the first resonance, and another at each integer multiple of the fundamental. With increasing frequency, the transmitted power decreases a little and then flattens out around the second resonance, decreases a little, flattens out at the third resonance, etc., as is illustrated in Fig. 3. If the spark discharge is really broadband (that is, if it generates a large number of frequencies where the highest frequency may be many times the lowest), then what is transmitted by the antenna will also be broadband, although with somewhat higher transmission at the fundamental frequency and its harmonics than in between. In the very early days of radio, this was somewhat acceptable, although any information impressed on such a broadband carrier would be rather severely degraded upon reception. However, the demise of the spark gap transmitter was really instigated by the early success of radio, which caused the available frequency bands to begin to fill up rapidly. This band filling led to the formation of the Federal Communications Commission (FCC) in 1934, which was charged with allocation of frequency bands. The allocation by nature led to a ban on spark gap transmitters, which were needlessly wasting bandwidth.

In a later experiment, Hertz noticed that the waves he was generating would tend to have a component that hugged the ground and could therefore travel over the horizon and, in fact, across the Atlantic Ocean, skimming along the surface of the water. Other researchers noticed that the effect became more pronounced at wavelengths longer than the roughly 2 m wavelength that Hertz originally used. (For the frequencies and wavelengths of some important frequency bands, see Table 1.) In order for wave transmission to be useful, however, the transmitted signal needs to carry information. Impressing information on the wave is called *modulating* the carrier. One can modulate the height (amplitude), the frequency, and so on. The discovery of a technique to *amplitude-modulate* the waves coming off an antenna (in 1906) then led to the inception of AM radio in bands with wavelengths greater than 300 m, which corresponds to roughly 1 MHz. AM radio became commercial in 1920. By the 1930s, other researchers noted that waves with frequencies around 10 MHz, corresponding to a wavelength around 30 m, could be quite efficiently propagated over the horizon by bouncing the wave off the ionosphere. This led to the radio bands known as *short-wave*. In 1939, a researcher realized a technique to modulate the frequency of the wave. This realization led in the 1950s to FM radio, which was allocated the band around 100 MHz with

Table 1. A Listing of the Allocated Microwave and Millimeter-Wave Bands as Defined by the Frequency and Wavelength Range Within Each Band

Band Designation	Frequency (GHz)	Wavelength
L	1–2	15–30 cm
S	2–4	7.5–15 cm
C	4–8	3.75–7.5 cm
X	8–12	2.5–3.75 cm
Ku	12–18	1.67–2.5 cm
K	18–26	1.15–1.67 cm
Ka	26–40	0.75–1.15 cm
Q	33–50	6–9 mm
U	40–60	5–7.5 mm
V	50–75	4–6 mm
E	60–80	3.75–5 mm
W	75–110	2.7–4 mm
D	110–170	1.8–2.7 mm
G	140–220	1.4–2.1 mm
Y	220–325	0.9–1.4 mm

a corresponding wavelength around 3 m. However, the FM technique was used first during World War II as a radar modulation technique. Radars today are at frequencies above roughly 1 GHz or wavelengths below 30 cm.

There is a fundamental difference between circuits that operate at frequencies whose corresponding wavelengths are less than the maximum circuit dimension and those that are large compared to the carrier wavelength. The effect is closely related to the concept of impedance. As was mentioned above, in the wire antenna, the voltage and current reinforce each other and thereby travel on the antenna as waves. The same effect takes place in a circuit. At any point along the path (line) in a circuit, one defines the ratio of voltage at one frequency to the current at the same frequency as the impedance at that frequency. For a sinusoidal waveform, if the impedance tends to preserve the phase relationship (where the wave peaks lie, relatively), then we say that the impedance is *resistive.* If the impedance tends to drive the voltage peaks forward with respect to the current peaks, we say that the impedance is *capacitive;* in the opposite case we say that the impedance is *inductive.* In a small circuit (small compared to a wavelength), one generally tries to carefully design passive components—resistors, capacitors, and inductors—so that

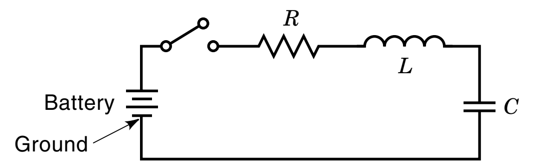


Figure 5. A circuit with lumped elements connected by wire segments.

they exhibit large local impedance, that is, large impedance within their physical dimensions. When the circuit is small, one would like to control the phase and amplitude of the wave at discrete points by using lumped elements and thereby minimizing line effects. The lines (wires) between the components have little or no effect on the electromagnetic disturbances passing through the circuit, then, as the impedances in the wires are small and reasonably independent of their lengths. When the circuit is large, the lines themselves effectively become circuit elements, and they themselves must be carefully designed in order to exhibit the proper impedances. To illustrate, consider the parallel plate capacitor of Fig. 4. The capacitance is maximized by maximizing the permittivity ϵ (a material parameter equal to the ratio of electrial displacement to applied electric field) and area A while minimizing the plate spacing d. However, the fact that the capacitance depends on the plate spacing d is the important point here. Consider the circuit of Fig. 5 as an example. The only ground in the figure is the one on the battery, but the wires connecting the circuit elements together in essence form at each point a capacitor, with a point on the wire that is carrying charge as the upper plate and the ground as the lower. This capacitance changes as a function of position along the wire. For a small enough circuit (relative to the wavelength of the highest frequency carried by the circuit), the effect is not too important, as the wire–ground pair has small capacitance and the position-varying effect is small. For a large circuit, the effect is disastrous, as we shall consider below. The effect is identical to the effect of Fresnel coefficients in optics.

Consider the circuit of Fig. 6. We will now discuss what happens when impedances are not carefully controlled. This leads to the concept of *impedance matching.* Let us first say that the circuit is short (compared to a wavelength). If the load resistor, R_L, is not matched to (that is, is not equal to, or, one could say, *not impedance matched to*) the resistance of the source, R_S, some amount of reflection will occur at R_L, propagate back to R_S, be reflected with a reversal of sign at R_L,

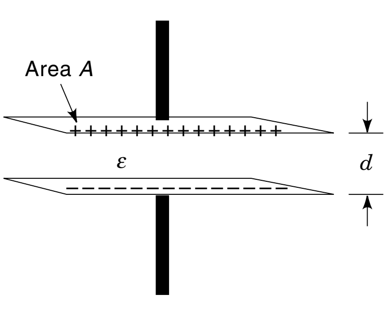


Figure 4. Schematic depiction of a parallel plate capacitor in which the flow of a current will tend to change the upper plate, causing a voltage difference between upper and lower plates. The capacitance is defined as the ratio of the amount of change of the upper plate to the magnitude of the voltage this change induces between the plates.

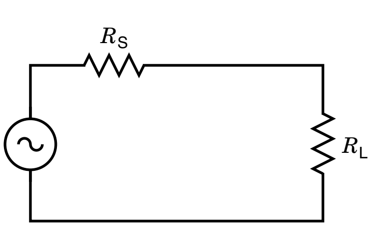


Figure 6. A circuit in which one is trying to supply power from a source with internal resistance R_S to a load with resistance R_L. The power transfer is maximized only when R_S and R_L are equal, in which case half the power supplied by the source is supplied to the load, the other half being dissipated in the source and causing it to heat.

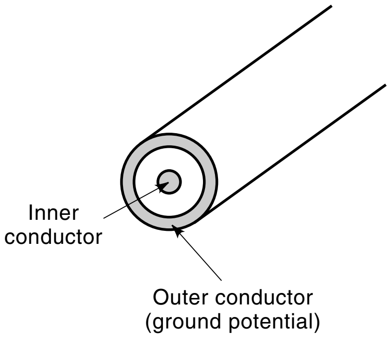


Figure 7. A coaxial cable in which signals are carried on an inner conductor and in which the grounded outer conductor serves to carry the ground plane along with the signal in order to give a constant impedance along the line.

propagate back to R_L, etc. The reflections add up perfectly out of phase (that is, simply subtract from one another) at the source and load, and the amount of power supplied to the load is less than optimal. In this limit of a small circuit, it is as if the load will not allow the source to supply as much power as it is capable of. Let us now say that the line is "well-designed" but long compared to the wavelength used. Then the same argument applies to the reflections, but in this case the source does not know that the load is there until several wave periods have passed (several maxima and minima of the waveform have left the source), so the source supplies all the power it can. The power, though, is not allowed to be fully absorbed by the load, and some of it will rattle around the line until it is radiated or absorbed. As we mentioned above, in a long enough circuit the wire itself becomes a distributed element—that is, one with an impedance of its own. If the distance to the nearest ground is not kept fixed along the line, the inductance and capacitance become dependent on the position. In this case, we have distributed reflections all along the line and the circuit will probably not work at all. This spatial variability of the line impedance is remediable, though, as illustrated by the drawing of a coaxial cable in Fig. 7. The idea is that, if the line brings along its own ground plane in the form of a grounded outer conductor, the characteristic impedance of the line can be kept constant with distance. Such a line, which carries its own ground plane, is called a *transmission line*. The problem becomes the connection of the line to the source and load (i.e., impedance matching).

Before going on to discuss the conventional solution versus the new active antenna solution, perhaps we should summarize a bit. In AM, short-wave, and FM applications, the wavelengths are of order greater than meters. If one considers typical receivers, the whole circuit will generally be small compared to the carrier wavelength. This is also to say that in all of these cases, the antennas will be active in the sense that the antenna presents an impedance to the circuit. (Recall that an active antenna is any antenna in which an active element lies within a wavelength of the antenna and is used as an element to match the antenna impedance to the decoder impedance.) To passively match an antenna to the receiver circuit, one needs pieces of line comparable to a wavelength. However, from here on we shall not be interested in the low-frequency case but rather in the well-above-1-GHz case, as AM, FM, and TV technologies are mature technologies. During World War II, radar was the application that drove the frequencies above 1 GHz (wavelength less than 30 cm). In a radar, one sends out a pulse and, from the returned, scattered wave, tries to infer as much as possible about the target. Target resolution is inversely proportional to wavelength. There has been a constant drive to shorten wavelength. Therefore, as is indicated by Table 1, bands have been allocated out to hundreds of gigahertz. Presently, however, there are a plethora of nonmilitary drivers for pushing to higher-frequency communication systems that are compact and have lower power dissipation. However, the conventional solution, which was developed originally for radars, is really not conducive to compactness nor to the pressures of cost minimization of the commercial market.

A typical conventional transmitter is schematically depicted in Fig. 8. A main concept here is that the transmission lines and matching networks are being used to isolate the oscillator from the amplifier and the amplifier from the antenna, in contrast to the situation in an active antenna. There were a number of reasons why the conventional solution took on the form it did. Among them was the urgency of World War II. Radar was developed rapidly in both Great Britain and the United States in the 1930s and 1940s. Rapid development required numerous researchers working in parallel. When operating frequencies exceeded 1 GHz (corresponding to 30 cm wavelengths), passive matching networks, whose main requirement is that they must consist of lines of lengths comparable to a wavelength, became convenient to construct (in terms of size) for ground-based radar. In this case, then, the oscillators could be optimized independently of the amplifiers, which in turn could be optimized independently of the antennas and the receiver elements. The impedances of the individual pieces didn't matter, as the matching networks could be used to effectively transform the effective impedances looking into an element into something completely different for purposes of matching pieces of the network to each other. There are costs associated with such a solution, though, such as total system size as well as the tolerances that components must satisfy. However, once the technique was in place, the industry standardized on the conventional solution and perfected it to the point where it was hard to challenge. The reemergence of the active solution owes itself to two indepen-

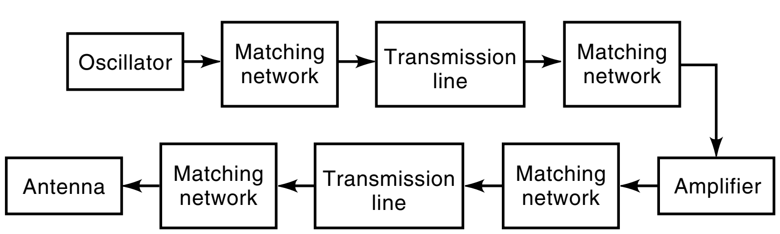


Figure 8. Schematic of a conventional RF microwave transmitter in which each individual element of the transmitter is matched to each other element.

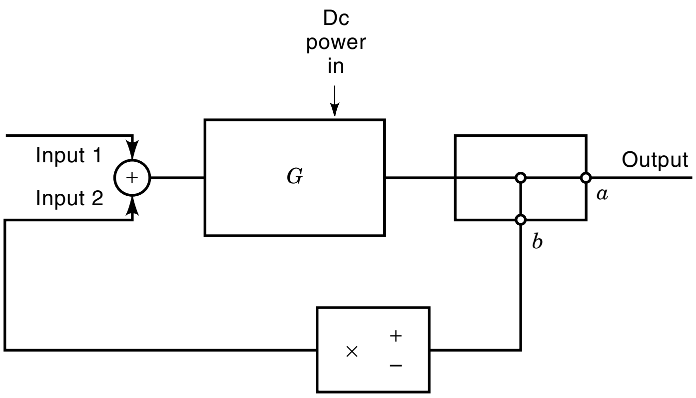


Figure 9. Schematic depiction of a feedback system that can operate as an oscillator when G is greater than 1, the feedback is positive, and there is a delay in feeding back the output to the input.

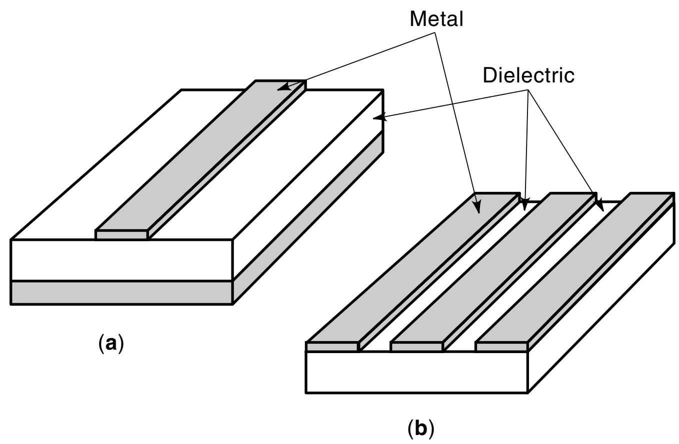


Figure 10. Views of (a) a microstrip and (b) a coplanar waveguide line. In the microstrip, the ground plane is the lower electrode, whereas in the coplanar waveguide the ground plane is placed on the surface of the dielectric substrate.

dent technologies, the emergence of high-frequency solid-state devices and the development of planar circuit and planar antenna technology.

A single frequency of electromagnetic energy must be generated in a so-called *oscillator*—that is, a circuit that converts dc electrical power to ac electromagnetic power at the proper frequency. The basic operation of an oscillator can be described with respect to Fig. 9. What is shown here schematically is an amplifier in which a portion b (<1) of the output is fed back to the input with either a plus or a minus sign. When the feedback is off ($b = 0$), then the signal out will be just G times the input. When the feedback is negative, the output will be less than G times the input. However, in the negative feedback mode, the stability to noise increases, since fluctuations will be damped. That is, if the output fluctuates up, this lowers the effective input, whereas if the output fluctuates down, the output is driven up. The opposite is true in the positive feedback case. In the positive feedback case, if there were no fluctuations, any input would cause the output to increase until all of the dc power in as well as all of the input signal showed up at the output. (This is all of the power that can show up at the output. Such behavior is typical of unstable operation.) This would not be such an interesting case; however, there are always fluctuations of the input, and the positive feedback will cause these to grow. If there is a delay from output to input, then fluctuations with a period corresponding to this delay will be favored, as a rise in the input will show up as a rise in the output one period later, and rapidly all of the dc power in will be converted to power out at this magic frequency.

A real circuit operates a bit more interestingly than our ideal one. In a real circuit, as the fluctuations build up, the gain is affected and some elements absorb power, but the oscillations still take place, although perhaps with a different frequency and amplitude from what one would have predicted from nondynamic measurements.

The transistor was first demonstrated in 1947, with publication in 1948 (3), and the diode followed shortly (4). Although the field effect transistor (FET) was proposed in 1952 (5), it was not until the mid 1960s that the technology had come far enough that it could be demonstrated (6). The FET (and variations thereof) is presently the workhorse microwave three-terminal device. Two-terminal transfer electron devices (TEDs) were used before the FET for microwave applications and are still in use, but tend to have a much lower wall plug efficiency (dc to ac conversion), especially as the amplifying device of an oscillator. Radar systems, however, were already in use in the late 1930s. Essentially all of the microwave sources in radars up until the 1970s operated on principles that required that the source have physical dimensions larger than a wavelength, and perhaps many wavelengths. This fact almost required the conventional solution to be used. Transistors, though, can have active areas with dimensions of micrometers; even packaged hybrid devices can have complete packages of dimensions smaller than a millimeter. The transistor can therefore act as an amplifier with dimensions much smaller than a wavelength and does not, therefore, need to be placed in a conventional (passive) solution design.

The last piece of our story of the new active antenna era involves the development of printed circuit technology, along with slot and patch antennas. The two most common planar "open waveguide" designs are microstrip line and coplanar waveguide (CPW). Depictions of these waveguide lines are given in Fig. 10. The idea behind the microstrip line is to propagate electromagnetic energy along the lines by confining the electric field between the upper signal line and a lower ground plane. As the upper line carries current, a magnetic field encircles the upper line. As power flow takes place in a direction perpendicular to the electric and magnetic fields, the power flow is mostly between the signal line and the ground

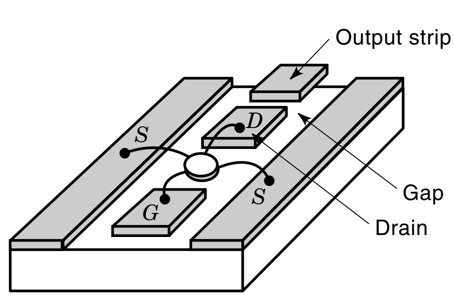


Figure 11. A simple transistor oscillator implemented in CPW technology.

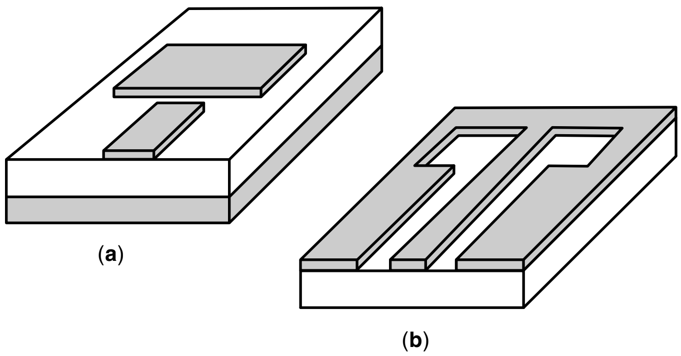


Figure 12. A depiction of (a) a patch antenna in a microstrip line and (b) a slot antenna in a CPW line.

line in the dielectric. On a low-frequency wire (a line whose transverse dimensions are small compared to a wavelength), the voltage and current waveforms reinforce each other. The coupling of the electric and magnetic fields in the microstrip is analogous to the coupling of voltage and current on the Hertz antenna wire, except that the microstrip line can be electrically long in the sense that the distance from the signal line to the ground plane is kept constant so that the impedance can be kept constant, as with the earlier-discussed coaxial cable. Lines that carry along their ground planes are generally referred to as *transmission lines.* Components (i.e. capacitors and inductors) can be built into the line by changing the width, cutting gaps into the upper line, or putting slits in the ground plane. In this sense, we can still describe transmission line circuits by conventional circuit theory if we use a special circuit model for the line itself. The CPW line is quite similar to the microstrip line except that there the ground planes are on top of the dielectric slab. Either of these line types is reasonably easy to fabricate, as one needs only to buy a metal-coated dielectric plate and then pattern the needed shapes by photographically defining the patterns using a technique known as *photolithography,* a process common to all present-day circuit fabrication. These planar structures are quite compatible with transistor technology, as is indicated by the simple transistor oscillator circuit depicted in Fig. 11. The gap in the line on the drain side is there in order to provide the proper feedback for oscillation. In this case, the total oscillator linear dimension can be less than a wavelength.

In order to have an active antenna, one needs to have a radiating element—that is, a passive antenna element in the active antenna. There are antenna technologies which are compatible with microstrip and CPW technologies, and the resulting antenna types are illustrated in Fig. 12. The idea behind either of these antenna types is that the patch (slit) is designed to have a transverse length that matches the operating wavelength (as we discussed in conjunction with Hertz dipole antennas). In the case of the patch, the electric field points primarily from the patch to the ground plane, as is illustrated in Fig. 13. The edges of the transverse (to the input line) dimension will then have a field pattern as sketched in Fig. 13(a), and the longitudinal edges will have a field pattern as sketched in Fig. 13(b), with a composite sketch given in Fig. 13(c). The important part of the sketches, however, is really the so-called fringing fields in Fig. 13(a)—

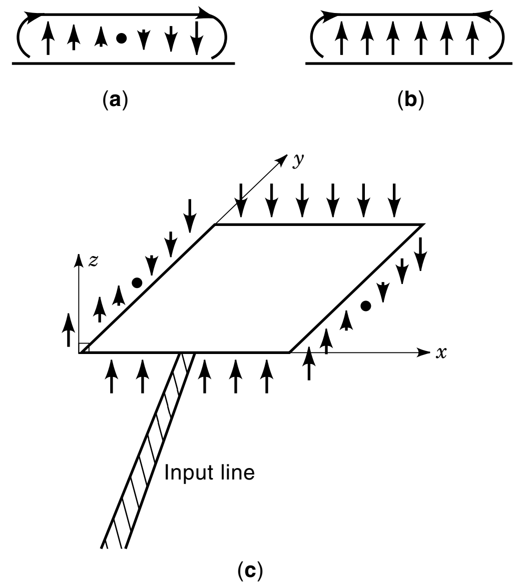


Figure 13. Illustration of the electric field directions along (a) the nonradiating edge and (b) the radiating edge, and (c) a schematic depiction of the edge fields around the patch.

that is, the fields that point neither up nor down but rather across. Beyond the longitudinal edges of the patch are fields, in phase for the two edges, that are normal to the surface. It is these fields (when combined with transverse magnetic fringe fields in the same strips) that give rise to the upward radiation. Similar arguments describe the operation of the slit antenna if one exchanges the electric and magnetic fields in the argument.

We have now introduced all of the pieces necessary to describe the new resurgence in active antenna research. A possible active antenna design could appear as in Fig. 14 (7), where the transistor is actually mounted right into the patch antenna element, and therefore the design can be quite compact. That is, the source plus oscillator plus antenna can all be fitted into less than a wavelength. The design of Fig. 14, which comes from R. Compton's group at Cornell (31,32), will be discussed further in the next section.

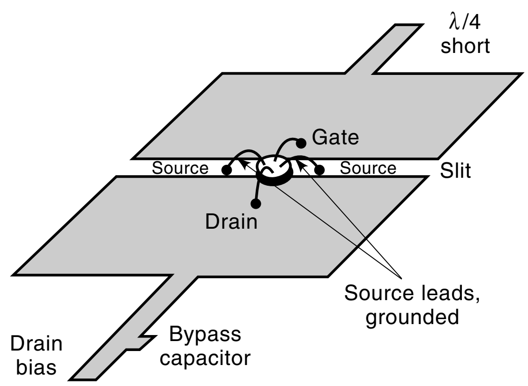


Figure 14. Depiction of the upper surface metallization of a microstrip active patch antenna discussed in Ref. 7. The short on the gate together with the slit between gate and drain provides the proper feedback delay to cause oscillation.

There are a number of advantages to the use of active antennas. One is that an active antenna can be made compact. Compactness in itself is advantageous, as throughout the history of microelectronics, miniaturization has led to lowered costs. There are two more advantages, though, which relate to compactness. One is that the power-handling capabilities of a device go down with increasing frequency. We would therefore like to find ways to combine the power from several devices. One can try to add together outputs from various oscillators in the circuit before feeding them to the elements, but this goes back to the conventional solution. A more advantageous design is to make an array of antennas, with proper spacing relative to the wavelength and antenna sizes, and add the power of the locked oscillators in the array quasi-optically in free space. (In other words, optical radiation tends to radiate into free space, whereas radio frequency in microwave radiation needs to be kept in guiding waveguides until encroachment on radiating elements. *Quasi-optics* uses the principle of the optical interferometer to combine multiple coherent microwave fields in free space.) The locking requires that the oscillators talk to each other so that the phases of all of the array elements stay in a given relation. As will be discussed in more detail in the next section, however, an important problem at present in the active antenna field relates to keeping elements locked yet still being able to modulate the output as well as steer the beam in order to be able to electronically determine on output direction. These issues will be discussed in the next section and taken up in more detail in the last section.

SOME QUANTITATIVE DISCUSSION OF ASPECTS OF ACTIVE ANTENNAS

In order to be able to make calculations on active antennas, it is important to know what level of approximation is necessary in order to obtain results. An interesting point is that, although the operating frequency of active antennas is high, the circuit tends to be small in total extent relative to the operating wavelength, and therefore the primary design tool is circuit theory mixed with transmission line theory. These techniques are approximate, and a most important point in working with high frequencies is to know where a given technique is applicable. Exact treatments of all effects, however, prove to be impossible to carry out analytically. Numerical approaches tend to be hard to interpret unless one has a framework to use. The combined circuit transmission-line framework is the one generally applied. When it begins to break down, one tends to use numerical techniques to bootstrap it back to reality. We will presently try to uncover the basic approximations of transmission line and circuit theory.

Maxwell's equations are the basic defining equations for all electromagnetic phenomena, and they are expressible in MKSA units as (8)

$$\nabla \times \boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t}$$
$$\nabla \times \boldsymbol{H} = \boldsymbol{J} + \frac{\partial \boldsymbol{D}}{\partial t}$$
$$\nabla \cdot \boldsymbol{D} = \rho$$
$$\nabla \cdot \boldsymbol{B} = 0$$

where $\boldsymbol{E}$ is the electric field vector, $\boldsymbol{B}$ is the magnetic induction vector, $\boldsymbol{H}$ is the magnetic field vector, $\boldsymbol{D}$ is the electric displacement vector, $\boldsymbol{J}$ is the current density vector, and ρ is the volume density of charge. An additional important quantity is $\boldsymbol{S}$, the Poynting vector, defined by

$$\boldsymbol{S} = \boldsymbol{E} \times \boldsymbol{H}$$

If one takes the divergence of $\boldsymbol{S}$, one finds

$$\nabla \cdot \boldsymbol{S} = \nabla \cdot (\boldsymbol{E} \times \boldsymbol{H})$$

If one assumes a free-space region,

$$\boldsymbol{D} = \epsilon_0 \boldsymbol{E}$$
$$\boldsymbol{B} = \mu_0 \boldsymbol{H}$$

which is therefore lossless,

$$\boldsymbol{J} = 0$$

and charge-free,

$$\rho = 0$$

(where ϵ_0 is the permittivity of free space and μ_0 is the permeability of free space), one can use vector identities and Maxwell's equations to obtain

$$\nabla \cdot \boldsymbol{S} = -\frac{\epsilon_0}{2}\frac{\partial}{\partial t}(\boldsymbol{E} \cdot \boldsymbol{E}) - \frac{\mu_0}{2}\frac{\partial}{\partial t}(\boldsymbol{H} \cdot \boldsymbol{H})$$

Integrating this equation throughout a volume V and using Gauss's theorem,

$$\int \nabla \cdot \boldsymbol{S}\, dV = \int \boldsymbol{S} \cdot \boldsymbol{dA}$$

where $\boldsymbol{dA}$ is the differential area times the unit normal pointing out of the surface of the volume V, one finds that

$$\int \boldsymbol{S} \cdot \boldsymbol{dA} = -\frac{\partial}{\partial t}W_e - \frac{\partial}{\partial t}W_m$$

where W_e is the electric energy density

$$W_e = \frac{\epsilon_0}{2}\int \boldsymbol{E} \cdot \boldsymbol{E}\, dV$$

and W_m is the magnetic energy density

$$W_m = \frac{\mu_0}{2}\int \boldsymbol{H} \cdot \boldsymbol{H}\, dV$$

The interpretation of the above is that the amount of $\boldsymbol{S}$ flowing out of V is the amount of change of the energy within. One therefore associates energy flow with $\boldsymbol{S} = \boldsymbol{E} \times \boldsymbol{H}$. This is important in describing energy flow in wires as well as transmission lines and waveguides of all types. As was first described by Heaviside (9), the energy flow in a wire occurs not inside the wire but around it. That is, as the wire is highly conductive, there is essentially no field inside it except at the surface, where the outer layer of oscillating charges have no outer shell to cancel their effect. There is therefore a radial electric field emanating from the surface of the wire, which combines with an azimuthal magnetic field that rings the current flow to yield an $\boldsymbol{E} \times \boldsymbol{H}$ surrounding the wire and pointing down its axis.

It was Pocklington in 1897 (10) who made the formal structure of the fields around a wire a bit more explicit and, in the effort, also formed the basis for the approximation upon which most of circuit and transmission line theory rests, the *quasi-static approximation.* A simplified version of his argument is as follows. Assume an x–y–z Cartesian coordinate system where the axis of the wire is the z axis. One then assumes that all of the field quantities $\boldsymbol{f}(x, y, z, t)$ vary as

$$\boldsymbol{f}(x, y, z, t) = \boldsymbol{f}(x, y)\cos(\beta z - \omega t + \phi)$$

If one assumes that the velocity of propagation of the above-defined wave is $c = (\mu_0\epsilon_0)^{-1/2}$, the speed of light, then one can write that

$$\beta = \frac{\omega}{c}$$

The assumption in the above that $\boldsymbol{f}(x, y)$ is independent of z, by substitution of the above into Maxwell's equations, can be shown to be equivalent to the assumption that the transverse field components E_x, E_y, B_x, and B_y all satisfy relations of the form

$$\left|\frac{\partial E_x}{\partial z}\right| \ll \beta|E_x|$$

which is the crux of the quasistatic approximation. With the above approximation, one finds that

$$\nabla_{\rm t} \times \boldsymbol{E}_{\rm t} = \rho$$
$$\nabla_{\rm t} \times \boldsymbol{H}_{\rm t} = \boldsymbol{J}$$

where

$$\nabla_{\rm t} = \hat{\boldsymbol{e}}_x\frac{\partial}{\partial x} + \hat{\boldsymbol{e}}_y\frac{\partial}{\partial y}$$

which is just the transverse, and therefore two-dimensional, gradient operator. These equations are just the electro- and magnetostatic equations for the transverse fields, whereas the propagation equation above shows that these static transverse field configurations are propagated forward as if they corresponded to a plane wave field configuration. If the magnetic field is caused by the current in the wire, it rings the wire, whereas if the electric field is static, it must appear to emanate from charges in the wire and point outward at right angles to the magnetic field. If this is true, then the Poynting vector $\boldsymbol{S}$ will point along the direction of propagation and the theory is self-consistent, if approximate.

If we wish to guide power, then the quasistatic picture must come close to holding, as the Poynting vector is in the right direction for guidance. The more general approximate theory that comes from Pocklington's quasistatic approximation is generally called *transmission line theory.* To derive this theory, first consider the two-wire transmission line of Fig. 15. If we are to have something that we can actually call a transmission line, then we would hope that we can find equiphase fronts of the electromagnetic disturbance propagating in the gap crossing the gap conductor and that we can find lines along which the current flows on the current-carrying conductor. Otherwise (if the equiphases closed on themselves and/or we had eddies in the current), it would be hard to think of the structure as any form of guiding structure. Let us say we form an area in the gap with two walls of the four-sided contour C_1 surrounding this area following equiphases an infinitesimal distance dz from each other. We can then write

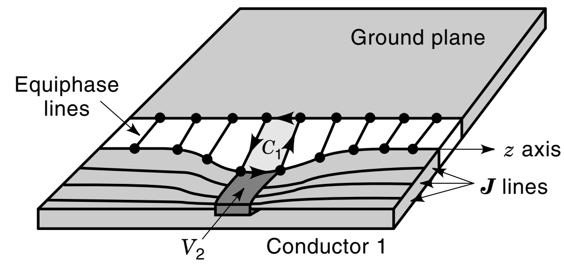


Figure 15. A sketch of a two-conductor transmission line where some equipotentials and some current lines are drawn in, as well as a volume V_1 with outward-pointing normal $d\boldsymbol{A}_1$. There is also an outward-pointing normal $d\boldsymbol{A}_2$ associated with the area bounded by contour C_2.

$$\int \nabla \times \boldsymbol{E} \cdot d\boldsymbol{A}_1 = -\int \frac{\partial \boldsymbol{B}}{\partial t} \cdot d\boldsymbol{A}_1$$

where $d\boldsymbol{A}_1$ corresponds to an upward-pointing normal from the enclosed area. One generally defines the integral as

$$\int \boldsymbol{B} \cdot d\boldsymbol{A}_1 = \phi$$

where ϕ is the magnetic flux. We often further define the flux as the inductance of the structure times the current:

$$\phi = Li$$

The integral with the curl in it can be rewritten by Stokes' theorem as

$$\int \nabla \times \boldsymbol{E} \cdot d\boldsymbol{A}_1 = \oint_{C_1} \boldsymbol{E} \cdot d\boldsymbol{l}$$

where C_1 is the contour enclosing the area. If we define

$$v = \int \boldsymbol{E} \cdot d\boldsymbol{l}$$

on the two equiphase lines of the contour C_1, where v is an ac voltage (this is the main approximation in the above, as it is only strictly true for truly static fields), then, noting that v does not change along two of the boundaries of the contour (because they are the infinitesimal walls on constant-voltage plates) and making the other two connecting lines infinitesimal, we note that the relation between the curl of $\boldsymbol{E}$ and the magnetic field reduces to

$$v(z + dz) - v(z) = \frac{\partial}{\partial t}(Li)$$

where it has been tacitly assumed that geometric deviations from rectilinearity are small enough that one can approximately use Cartesian coordinates, which can be rewritten in the form

$$\frac{\partial v}{\partial z} = l\,\frac{\partial i}{\partial t} \tag{1}$$

where l is an inductance per unit length, which may vary with longitudinal coordinate z if the line has longitudinal variation of geometry. A similar manipulation can be done with the second and third of Maxwell's equations. Taking

$$\nabla \cdot (\nabla \times \boldsymbol{H}) = \nabla \cdot \boldsymbol{J} + \frac{\partial}{\partial t}\nabla \cdot \boldsymbol{D}$$

and noting that the divergence of a curl is zero, substituting for $\nabla \cdot \boldsymbol{D}$, we find

$$\nabla \cdot \boldsymbol{J} + \frac{\partial \rho}{\partial t} = 0$$

which is the equation of charge conservation. Integrating this equation over a volume V_2 that encloses the current-carrying conductor whose walls lie perpendicular to the current lines gives

$$\int \nabla \cdot \boldsymbol{J}\, dV_2 = -\frac{\partial}{\partial t}\int \rho\, dV_2$$

where the total change Q, given by

$$Q = \int \rho\, dV_2$$

is also sometimes defined in terms of capacitance C and voltage v by

$$Q = Cv$$

Noting that

$$\int \nabla \cdot \boldsymbol{J}\, dV_2 = \int \boldsymbol{J} \cdot \boldsymbol{dA}_2$$

where $\boldsymbol{dA}_2$ is the outward-pointing normal to the boundary of the volume V_2 and where one usually defines

$$i = \int \boldsymbol{J} \cdot \boldsymbol{dA}_2$$

and letting the volume V have infinitesimal thickness, one finds that

$$\int \boldsymbol{J} \cdot \boldsymbol{dA}_2 = i(z+dz) - i(z)$$

Putting this together with the above, we find

$$\frac{\partial i}{\partial z} = c\,\frac{\partial v}{\partial t} \tag{2}$$

where c is the capacitance per length of the structure, and where longitudinal variations in line geometry will lead to a longitudinal variation of c. The system of partial differential equations for the voltage and current have a circuit represen-

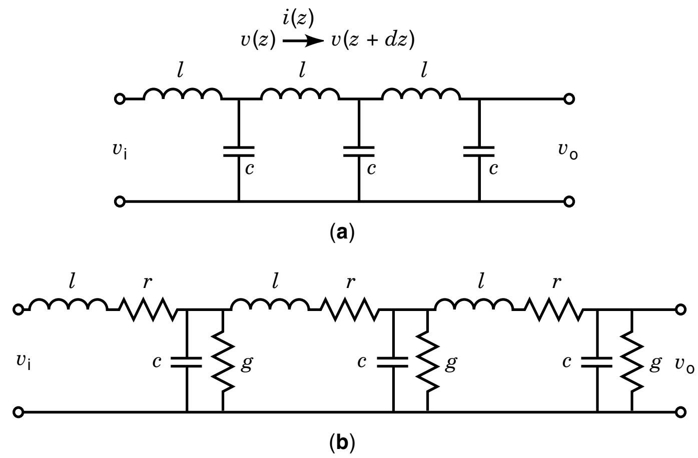


Figure 16. A circuit equivalent for (a) a lossless and (b) a lossy transmission line. The actual stages should be infinitesimally long, and the l's and c's can vary with distance down the line. In reality, one can find closed-form solutions for the waves in nominally constant l and c segments and put them together with boundary conditions.

tation, as is schematically depicted in Fig. 16(a). One can verify this by writing Kirchhoff's laws for the nodes with $v(z + dz)$ and $v(z)$ using the relations

$$v = l\,\frac{\partial i}{\partial t}$$

and

$$i = c\,\frac{\partial v}{\partial t}$$

Figure 16(b) illustrates the circuit equivalent for a lossy (and therefore dispersive) transmission line, where r represents the resistance encountered by the current in the metallization and where g represents any conductance of the substrate material that might allow leakage to ground. A major point of the diagram is that the structure need not be uniform in order to have a transmission line representation, although one may find that irregularities in the structure will lead to longitudinally varying inductances and capacitances.

The solution to the circuit equations will have a wave nature and will exhibit propagation characteristics, which we discussed previously. In a region with constant l and c, one can take a z derivative of Eq. (1) and a t derivative of Eq. (2) and substitute to obtain

$$\frac{\partial^2 v}{\partial z^2} - lc\,\frac{\partial^2 v}{\partial t^2} = 0$$

which is a wave equation with solutions

$$v(z,t) = v_{\mathrm{f}}\cos(\omega t - \beta z + \phi_{\mathrm{f}}) + v_{\mathrm{b}}\cos(\omega t + \beta z + \phi_{\mathrm{b}}) \tag{3}$$

where v_{f} is the amplitude of a forward-going voltage wave, v_{b} is the amplitude of a backward-going voltage wave, and

$$\frac{\omega}{\beta} = \sqrt{lc}$$

Similarly, taking a t derivative of Eq. (1) and a z derivative of Eq. (2) and substituting gives

$$\frac{\partial^2 i}{\partial z^2} - lc\,\frac{\partial^2 i}{\partial t^2} = 0$$

which will have a solution analogous to the one in Eq. (3) above, but with

$$v_{\mathrm{f}} = \sqrt{\frac{l}{c}}\, i_{\mathrm{f}}$$
$$v_{\mathrm{b}} = \sqrt{\frac{l}{c}}\, i_{\mathrm{b}}$$

which indicates that we can make the identification that the line phase velocity v_{p} is given by

$$v_{\mathrm{p}} \triangleq \frac{\omega}{\beta} = \sqrt{lc}$$

and the line impedance Z_0 is given by

$$Z_0 = \sqrt{l/c}$$

Oftentimes, we assume that we can write (the sinusoidal steady-state representation)

$$v(z,t) = \mathrm{Re}[v(z)e^{j\omega t}]$$
$$i(z,t) = \mathrm{Re}[i(z)e^{j\omega t}]$$

so that we can write

$$\frac{\partial v}{\partial z} = -j\omega l i$$
$$\frac{\partial i}{\partial z} = -j\omega c v$$

with solutions

$$v(z) = v_{\mathrm{f}} e^{-j\beta z} + v_{\mathrm{b}} e^{j\beta z}$$
$$i(z) = i_{\mathrm{f}} e^{-j\beta z} - i_{\mathrm{b}} e^{j\beta z}$$

Let us say now that we terminate the line with a lumped impedance Z_l at location l. At the coordinate l, then, the relations

$$Z_l i(l) = v_{\mathrm{f}} e^{-j\beta l} + v_{\mathrm{b}} e^{j\beta l}$$
$$Z_0 i(l) = v_{\mathrm{f}} e^{-j\beta l} - v_{\mathrm{b}} e^{j\beta l}$$

hold, and from them we can find

$$v_{\mathrm{f}} = \tfrac{1}{2}(Z_l + Z_0) i(l) e^{j\beta l}$$
$$v_{\mathrm{b}} = \tfrac{1}{2}(Z_l - Z_0) i(l) e^{-j\beta l}$$

which gives

$$v(z) = \frac{i(l)}{2}[(Z_l + Z_0)e^{j\beta(l-z)} + (Z_l - Z_0)e^{-j\beta(l-z)}]$$
$$i(z) = \frac{i(l)}{2Z_0}[(Z_l + Z_0)e^{j\beta(l-z)} - (Z_l - Z_0)e^{-j\beta(l-z)}]$$

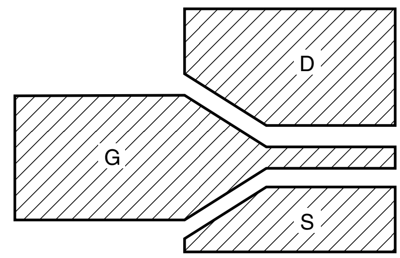


Figure 17. Schematic depiction of a top view of the metallized surface of an FET, where G denotes gate, D drain, and S source.

allowing us to write that

$$Z(z-\ell) = \frac{v(z-l)}{i(z-l)} = Z_0 \frac{Z_l + jZ_0 \tan\beta(z-l)}{Z_0 + jZ_l \tan\beta(z-l)} \qquad (4)$$

This equation allows us to, in essence, move the load from the plane l to any other plane. This transformation can be used to eliminate line segments and thereby use circuits on them directly. However, note that line lengths at least comparable to a wavelength are necessary in order to significantly alter the impedance. At the plane $z = l$, then, we can further note that the ratio of the reflected voltage coefficient v_{b} and the forward-going v_{f}, which is the voltage reflection coefficient, is given by

$$\mathscr{R} = \frac{Z_l - Z_0}{Z_l + Z_0}$$

and has the meaning of a Fresnel coefficient (8). This is the reflection we discussed in the last section, which causes the difference between large and small circuit dimensions.

One could ask what the use was of going at some length into Poynting vectors and transmission lines when the discussion is about active antennas. The answer is that any antenna system, at whatever frequency or of whatever design, is a system for directing power from one place to another. To direct power from one place to another requires constantly keeping the Poynting vector pointed in the right direction. As we can surmise from the transmission line derivation, line irregularities may cause the Poynting vector to wobble (with attendant reflections down the line due to attendant variations in the l and c), but the picture must stay close to correct for power to get from one end of the system to another. For this reason, active antennas, even at very high frequencies (hundreds of gigahertz), can still be discussed in terms of transmission lines, impedances, and circuit equivalents, although ever greater care must be used in applying these concepts at increasingly higher frequencies.

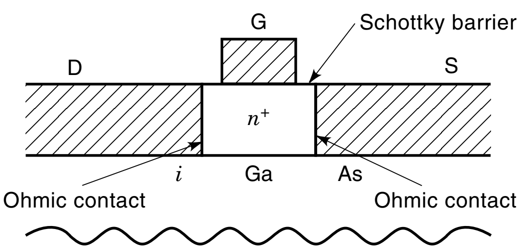
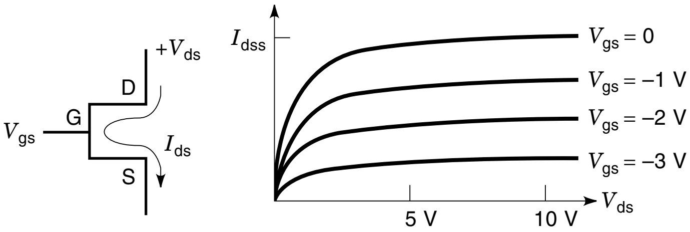


Figure 18. Schematic depiction of the cross section of the active region of a GaAs FET. Specific designs can vary significantly in the field-effect family.



(a) (b)

Figure 19. (a) Circuit element diagram with voltages and currents labeled for (b), where a typical I–V curve is depicted.

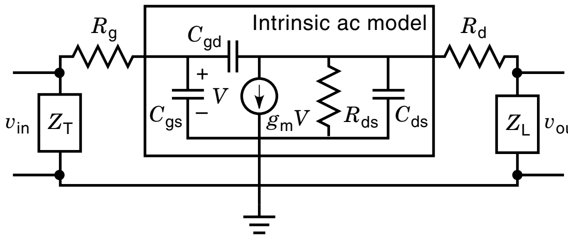


Figure 21. Intrinsic model for a common-source FET with external load and termination impedances and including gate and drain resistive parasitics, where Z_T is the gate termination impedance, R_g is the gate (metallization) resistance, C_{gs} is the gate-to-source capacitance, C_{gd} is the gate-to-drain capacitance, g_m is the channel transconductance, R_{ds} is the channel (drain-to-source) resistance, C_{ds} is the channel capacitance, R_d is the drain (metallization) resistance, and Z_L is the load impedance.

The next piece of an active antenna that needs to be discussed is the active element. Without too much loss of generality, we will take our device to be a field effect transistor (FET). The FET as such was first described by Shockley in 1952 (5), but the MESFET (metal–semiconductor FET), which is today's workhorse active device for microwave circuitry, was not realized until 1965 (6), when gallium arsenide (GaAs) fabrication techniques became workable albeit only as a laboratory demonstration. [Although we will discuss the MESFET in this section, it should be pointed out that the silicon MOSFET (metal–oxide–semiconductor FET) is the workhorse device of digital electronics and therefore the most common of all electronic devices presently in existence by a very large margin.] A top view of an FET might appear as in Fig. 17. As is shown clearly in the figure, an FET is a three-terminal device with gate, drain, and source regions. A cross section of the active region (that is, where the gate is very narrow) might appear as in Fig. 18. The basic idea is that the saturation-doped n region causes current to flow through the ohmic contacts from drain to source (that is, electrons flow from source to drain), but the current is controlled in magnitude by the electric field generated by the reverse bias voltage applied to the gate electrode. The situation is described in a bit more detail in Fig. 19, where bias voltages are defined and a typical I–V curve for dc operation is given. Typically the bias is supplied by a circuit such as that of Fig. 20. In what follows, we will simply assume that the biases are properly applied and isolated, and we will consider the ac operation. An ac circuit model is given in Fig. 21. If one uses the proper number of circuit values, these models can be quite accurate, but the values do vary from device to device, even when the devices were fabricated at the same time and on the same substrate. Usually, the data sheet with a device, instead of specifying the circuit parameters, will specify the parameters of the device S, which are defined as in Fig. 22 and which can be measured in a straightforward manner by a network analyzer. The S parameters are defined by the equation

$$\begin{pmatrix} V_1^- \\ V_2^- \end{pmatrix} = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \begin{pmatrix} V_1^+ \\ V_2^+ \end{pmatrix} \tag{5}$$

An important parameter of the circuit design is the transfer function of the transistor circuit, which can be defined as the ratio of v_o to v_i as defined in Fig. 21. To simplify further analysis, we will ignore the package parasitics R_g and R_d in comparison with other circuit parameters, and thereby we will carry out further analysis on the circuit depicted in Fig. 23. The circuit can be solved by writing a simultaneous system of equations for the two nodal voltages v_i and v_o. These sinusoidal steady-state equations become

$$v_i = v$$

$$j\omega C_{gd}(v_o - v_i) + g_m v_i + j\omega C_{ds} v_o + \frac{v_o}{R_{ds}} + \frac{v_o}{Z_L} = 0$$

The system can be rewritten in the form

$$v_o\left(j\omega(C_{gd} + C_{ds}) + \frac{1}{R_{ds}} + \frac{1}{Z_L}\right) = v_i(-g_m + j\omega C_{gd})$$

which gives us our transfer function T in the form

$$T = \frac{v_o}{v_i} = \frac{-g_m + j\omega C_{gd}}{j\omega\,(C_{gd} + C_{gs}) + \dfrac{1}{R_{ds}} + \dfrac{1}{Z_L}}$$

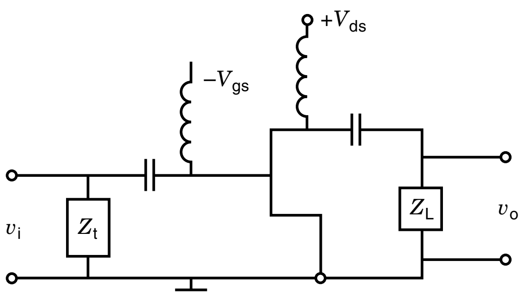


Figure 20. Typical FET circuit including the bias voltages v_{gs} and v_{ds} as well as the ac voltages v_i and v_o, where the conductors represent ac blocks and the capacitors dc blocks.

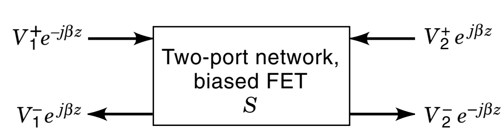


Figure 22. Schematic depiction of an FET as a two-port device that defines the quantities used in the S matrix of Eq. (5).

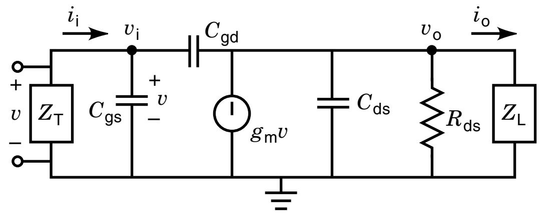


Figure 23. Simplified transistor circuit used for analyzing rather general amplifier and oscillator circuits, where the circuit parameter definitions are as in Fig. 22.

Oftentimes we are interested in open-circuit parameters—for example, the circuit transfer function when Z_L is large compared to other parameters. We often call this parameter G the open-circuit gain. We can write this open-circuit gain in the form

$$G = \left.\frac{v_o}{v_i}\right|_{oc} = \frac{-g_m R_{ds} + j\omega C_{gd} R_{ds}}{j\omega\,(C_{gd} + C_{gs}) R_{ds} + 1}$$

It is useful to look at approximate forms. It is generally true that

$$C_{gd} \ll C_{ds}, C_{gs}$$

and for usual operating frequencies it is also generally true that

$$\frac{1}{\omega C_{ds}} \ll R_{ds}$$

Using both of the above in our equations for T and G, we find

$$T = \frac{-g_m R_{ds}}{1 + \dfrac{R}{Z_L}}$$

$$G = -g_m R_{ds}$$

Clearly, from the above, one sees that the loaded gain will be lower than the unloaded gain, as we would expect. Making only the first of our two above approximations, we can write the above equations as

$$T = \frac{-g_m R_{ds}}{1 + j\omega\tau_{ds} + \dfrac{R_{ds}}{Z_L}}$$

$$G = \frac{-g_m R_{ds}}{1 + j\omega\tau_{ds}}$$

where τ_{ds} is a time constant given by

$$\tau_{ds} = \frac{1}{C_{ds} R_{ds}}$$

We see that, in this limit, the high-frequency gain is damped. Also, an interesting observation is that, at some frequency ω, an inductive load could be used to cancel the damping and obtain a purely real transfer function at that frequency. This effect is the one that allows us to use the transistor in an oscillator.

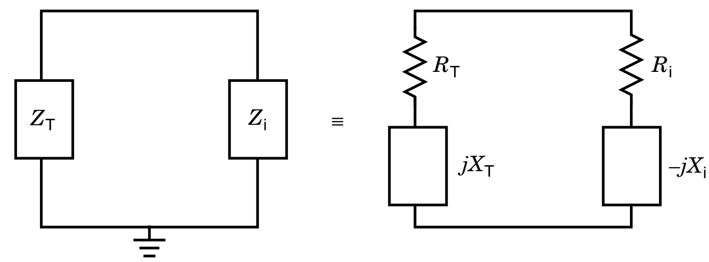


Figure 24. Diagram depicting the transistor and its load as a one-port device that, when matched to its termination so that there is no real or imaginary part to the total circuit impedance, will allow for oscillations.

Let us now consider an oscillator circuit. The basic idea is illustrated in the one-port diagram of Fig. 24. The transistor's gain, together with feedback to the input loop through the capacitor C_{gd}, can give the transistor an effective negative input impedance, which can lead to oscillation if the real and imaginary parts of the total impedance (that is, Z_T in parallel with the Z_i of the transistor plus load) cancel. The idea is much like that illustrated in Fig. 25 for a feedback network. One sees that the output of the feedback network can be expressed as

$$v_o = G(j\omega)[v_i - H(j\omega)v_o]$$

or, on rearranging terms,

$$\frac{v_o}{v_i} = \frac{G(j\omega)}{1 + G(j\omega)H(j\omega)}$$

which clearly will exhibit oscillation—that is, have an output voltage without an applied input voltage—when

$$H(j\omega) = -\frac{1}{G(j\omega)}$$

What we need to do to see if we can achieve oscillation is to investigate the input impedance of our transistor and load seen as a one-port network. Clearly we can write the input current of Fig. 23 as

$$i_i = j\omega C_{gs} v_i + j\omega C_{gd}(v_i - v_o)$$

and then, using the full expression for T to express v_o as a function of v_i, one finds

$$Z_i = \frac{i_i}{v_i} = j\omega C_{gs} + j\omega C_{gd}\left(1 + \frac{g_m - j\omega C_{gd}}{j\omega\,(C_{gd} + C_{ds}) + \dfrac{1}{R_{ds}} + \dfrac{1}{Z_L}}\right)$$

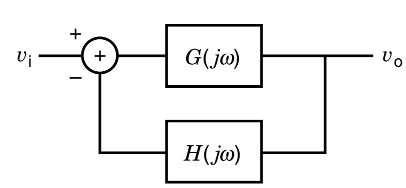


Figure 25. Depiction of a simple feedback network.

which can be somewhat simplified to yield

$$Z_i = j\omega C_{gs} + j\omega C_{gd} \frac{g_m R_{ds} + 1 + j\omega\tau_{ds} + \dfrac{R_{ds}}{Z_L}}{1 + j\omega\tau_{ds} + \dfrac{R_d}{Z_L}}$$

We can again invoke a limit in which $\omega\tau_{ds} \ll 1$ and then write

$$Z_i = j\omega C_{gs} + j\omega C_{gd} \frac{Z_L(1 + g_m R_{ds} + R_{ds})}{R_{ds} + Z_L}$$

Perhaps the most interesting thing about this expression is that if

$$Z_L = j\omega L$$

and

$$g_m R_{ds} \gg 1$$

then clearly

$$R_i < 0$$

Whether or not X_i can be made to match any termination is another question, which we will take up in the next paragraph.

As was mentioned earlier, generally the data sheet one obtains with an FET has plots of the frequency dependence of the S parameters rather than values for the equivalent circuit parameters. Oscillator analysis is therefore usually carried out using a model of the circuit such as that depicted in Fig. 26, where the transistor is represented by its measured S matrix. The S matrix is defined as the matrix of reflection and transmission coefficients. That is to say, with referrence to the figure, S_{11} would be the complex ratio of the field reflected from the device divided by the field incident on the device. S_{21} would be the field transmitted from the device divided by the field incident on the device. S_{12} would be the field incident from the load side of the device divided by the power incident on the device, and S_{22} would be the power reflected from the load side of the device divided by the power incident on the device. For example, if there is only an input from Z_T, then

$$\Gamma_i = S_{11}$$

If there is only an input from Z_L, then

$$\Gamma_o = S_{22}$$

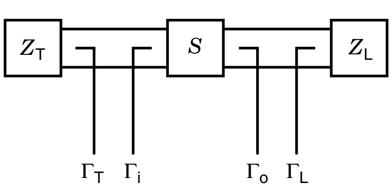


Figure 26. Schematic depiction of an oscillator circuit in which the transistor is represented by its S matrix and calculation is done in terms of reflection coefficients Γ_T looking into the gate termination, Γ_i looking into the gate source port of the transistor, Γ_o looking into its drain source port, and Γ_L looking into the load impedance.

The condition for oscillation in such a system can be expressed in either of the forms

$$\Gamma_i \Gamma_T = 1$$

or

$$\Gamma_o \Gamma_L = 1$$

where the Γ's are defined in the caption of Fig. 26. If both Z_T and Z_L were passive loads—that is, loads consisting of resistance, inductance, and capacitance, then we would have that

$$|\Gamma_T| < 1$$
$$|\Gamma_L| < 1$$

and the conditions for unconditional stability (nonoscillation at any frequency) would be that

$$|\Gamma_i| < 1$$
$$|\Gamma_o| < 1$$

Clearly, we can express Γ_i and Γ_o as series of reflections such that

$$\begin{aligned}\Gamma_i &= S_{11} + S_{12}\Gamma_L S_{21} + S_{12}\Gamma_L S_{22}\Gamma_L S_{21} \\ &\quad + S_{12}\Gamma_L S_{22}\Gamma_L S_{22}\Gamma_L S_{21} + \cdots \\ \Gamma_o &= S_{22} + S_{21}\Gamma_T S_{12} + S_{21}\Gamma_T S_{11}\Gamma_T S_{12} \\ &\quad + S_{21}\Gamma_T S_{11}\Gamma_T S_{11}\Gamma_T S_{12} + \cdots\end{aligned}$$

Using the fact that

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}$$

we can reexpress the Γ's as

$$\Gamma_i = S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L}$$
$$\Gamma_o = S_{22} + \frac{S_{12}S_{21}\Gamma_T}{1 - S_{22}\Gamma_T}$$

If we denote the determinant of the S matrix by

$$\Delta = S_{11}S_{22} - S_{12}S_{21}$$

and define a transistor parameter κ by

$$\kappa = \frac{1 - |S_{11}|^2 - |S_{22}|^2 + |\Delta|^2}{2|S_{12}S_{21}|}$$

then some tedious algebra leads to the result that stability requires

$$\kappa > 1$$
$$\Delta < 1$$

At frequencies where the above are not satisfied, oscillation can occur if the load and termination impedances, Z_L and Z_T respectively, are chosen properly. Oscillator design is discussed in various texts (11–14). Generally, though, oscillator design involves finding instability points and not predicting the dynamics once oscillation is achieved. Here we are discussing only oscillators which are self-damping. External circuits can be used to damp the behavior of an oscillator, but here we are discussing only those that damp themselves independent of an external circuit. The next paragraph will discuss these dynamics.

If a transistor circuit is designed to be unstable, then as soon as the dc bias is raised to a level where the circuit achieves the set of unstable values, the circuit's output within the range of unstable frequencies rises rapidly and dramatically. The values that we took in the equivalent ac circuit, though, were small-signal parameters. As the circuit output increases, the signal will eventually no longer be small. The major thing that changes in this limit is that the input resistance to the transistor saturates, so that (14)

$$R_{\mathrm{i}} = -R_{\mathrm{i}\phi} + mv^2$$

where the plus sign on the nonlinearity is necessary, for if it were negative the transistor would burn up or else burn up the power supply. Generally, m has to be determined empirically, as nonlinear circuit models have parameters that vary significantly from device to device. For definiteness, let us assume that the Z_{T} is resistive and the Z_{L} is purely inductive. At the oscillation frequency, the internal capacitance of the transistor then should cancel the load inductance, but to consider dynamics we need to put in both C and L, as dynamics take place in the time domain. The dynamic circuit to consider is then as depicted in Fig. 27. The loop equation for this circuit in the time domain is

$$L\frac{\partial i}{\partial t} + (R_{\mathrm{i}} + R_{\mathrm{T}})i + \frac{1}{C}\int i\,dt = 0$$

Recalling the equivalent circuit of Fig. 23 and recalling that

$$C_{\mathrm{gs}} \gg C_{\mathrm{gd}}$$

we see that, approximately at any rate, we should have a relation between v_{i} and i_{i} of the form

$$i_{\mathrm{i}} = C_{\mathrm{gs}}\frac{\partial v_{\mathrm{i}}}{\partial t}$$

Using this i–v relation in the above, we find that

$$\frac{\partial^2 v}{\partial t^2} - \frac{R_{\mathrm{i}} - R_{\mathrm{T}}}{L}\left(1 - \frac{mv^2}{R_{\mathrm{i}} - R_{\mathrm{T}}}\right)\frac{\partial v}{\partial t} + \frac{v}{LC} = 0$$

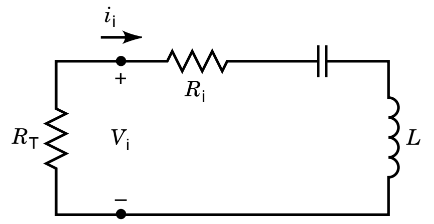


Figure 27. Circuit used to determine the dynamical behavior of a transistor oscillator.

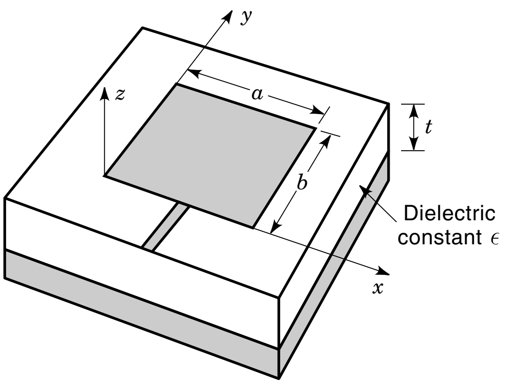


Figure 28. A patch antenna and Cartesian coordinate system.

which we can rewrite in terms of other parameters as

$$\frac{\partial^2 v}{\partial t^2} - \epsilon(1 - \gamma^2 v^2)\frac{\partial v}{\partial t} + \omega_0^2 v = 0$$

which is the form of Van der Pol's equation (15,16), which describes the behavior of essentially any oscillator.

Now that we have discussed planar circuits and dynamical elements that we can put into theory, the time has arrived to discuss planar antenna structures. Perhaps the best way to gain understanding of the operation of a patch antenna is by considering a cavity resonator model of one. A good review of microstrip antennas is given in Carver and Mink (17) and is reprinted in Pozar and Schaubert (18). Let us consider a patch antenna and coordinate system as is illustrated in Fig. 28. The basic idea behind the cavity model is to consider the region between the patch and ground plane as a resonator. To do this, we need to apply some moderately crude approximate boundary conditions. We will assume that there is only a z-directed electric field underneath the patch and that this field achieves maxima on the edges (open-circuit boundary condition). The magnetic field $\boldsymbol{H}$ will be assumed to have both x and y components, and its tangential components on the edges will be zero. (This boundary condition is the one consistent with the open-circuit condition on the electric field and becomes exact as the thickness of the layer approaches zero, as there can be no component of current normal to the edge at the edge, and it is the normal component of the current that generates the transverse $\boldsymbol{H}$ field.) The electric field satisfying the open-circuit condition can be seen to be given by the modes

$$\boldsymbol{e}_{mn} = \hat{\boldsymbol{e}}_z\frac{\chi_{mn}}{\sqrt{\epsilon abt}}\cos k_n x \cos k_m y$$

where

$$k_n = n\pi/a$$
$$k_m = m\pi/b$$
$$\chi_{mn} = \begin{cases} 1, & m = 0 \quad \text{and} \quad n = 0 \\ \sqrt{2}, & m = 0 \quad \text{or} \quad n = 0 \\ 2, & m \neq 0 \quad \text{and} \quad n \neq 0 \end{cases}$$

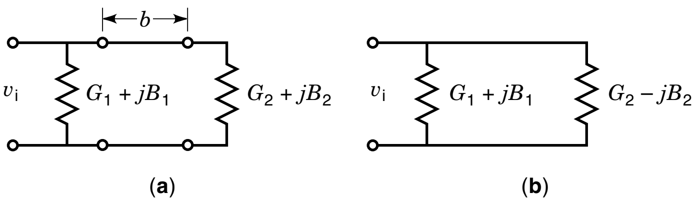


Figure 29. (a) A transmission line model for a patch antenna, and (b) its circuital equivalent as resonance.

The $\boldsymbol{H}$ field corresponding to the $\boldsymbol{E}$ field then will consist of modes

$$\boldsymbol{h}_{mn} = \frac{1}{j\omega\mu}\frac{\xi_{mn}}{\epsilon abt}(\hat{e}_x k_m \cos k_n x \sin k_m y - \hat{e}_y k_n \sin k_n x \cos k_m y)$$

As can be gathered from Fig. 13, the primary radiation mode is the mode with $m = 1$ and $n = 0$.

The basic operation is described by the fact that the boundary conditions are not quite exact. Recall from the earlier argument that accompanied Fig. 13 that the z-directed field gives rise to a fringe field at the edges $y = 0$ and $y = b$ such that there are strips of y-directed electric field around $y \leq 0$ and $y \geq b$. Because the boundary conditions are not quite correct on $\boldsymbol{H}$, there will also be strips of x-directed magnetic fields in these regions. As the Poynting vector is given by $\boldsymbol{E} \times \boldsymbol{H}$, we note that these strips will give rise to a z-directed Poynting vector. Similar arguments can be applied to the edges at $x = 0$ and $x = a$. However, the x-directed field at $x \leq 0$ has a change of sign at the center of the edge and is pointwise oppositely directed to the x-directed electric field at $x = 0$. These fields, therefore, only give rise to very weak radiation, as there is significant cancellation. Analysis of the slot antenna requires only that we interchange the $\boldsymbol{E}$ and $\boldsymbol{H}$ fields.

The picture of the patch antenna as two radiating strips allows us to represent it with a transmission line as well as a circuit model. The original idea is due to Munson (19). The transmission line model is depicted in Fig. 29. The idea is that one feeds onto an edge with an admittance (inverse impedance) $G_1 + jB_1$ and then propagates to a second edge with admittance $G_2 + jB_2$. When the circuit is resonant, then the length of transmission line will simply complex-conjugate the given load [see Eq. (4)], leading to the circuit representation of Fig. 29(b). The slot admittance used by Munson (19) was just that derived for radiation from a slit in a waveguide (20) as

$$G_1 + jB_1 = \frac{\pi a}{\lambda_0 Z_0}(1 - j0.636 \ln k_0 t)$$

where Z_0 is the impedance of free space ($\sqrt{\mu_0/\epsilon_0} = 377\ \Omega$), λ_0 is the free-space wavelength, and k_0 is the free-space propagation vector, and where a and t are defined as in Fig. 28. When the edges are identical (as for a rectangular patch), one can write

$$G_2 + jB_2 = G_1 + jB_1$$

to obtain the input impedance in the form

$$Z_i = \frac{1}{Y_i} = \frac{1}{2G_1}$$

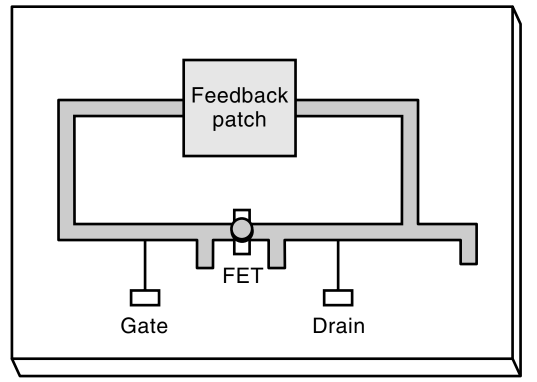


Figure 30. A design of a microstrip active radiating element.

We have now considered all of the pieces, and therefore it is time to consider a couple of actual active antenna designs. Figure 30 depicts one of the early designs from Kai Chang's group at Texas A&M (21). Essentially, the patch here is being used precisely as the feedback element of an amplifier circuit (as was described in connection with Fig. 9). A more compact design is that of Fig. 14 (7). There, the transistor is actually mounted directly into the patch antenna. The slit between the gate and drain yields a capacitive feedback element such that the effective ac circuit equivalent of this antenna may appear as depicted in Fig. 31. The capacitor–inductor pair attached to the gate lead forms what is often referred to as a *tank circuit,* which (if the load were purely real) defines a natural frequency through the relation

$$\omega = \sqrt{\frac{1}{LC}}$$

As was discussed at some length in the last section of this article, a major argument for the use of active antennas is that they are sufficiently compact that they can be arrayed together. Arraying is an important method for free-space power combining, which is necessary because as the frequency increases, the power-handling capability of active devices decreases. However, element size also decreases with increasing frequency so that use of multiple coherently combined elements can allow one to fix the total array size and power more or less independently of frequency, even though the number of active elements to combine increases. In the next paragraph, we shall consider some of the basics of arrays.

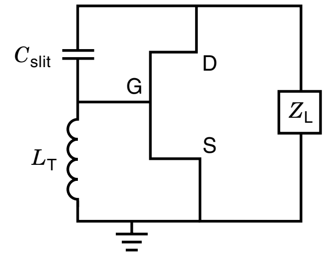


Figure 31. Ac circuit equivalent of the active antenna of Fig. 14.

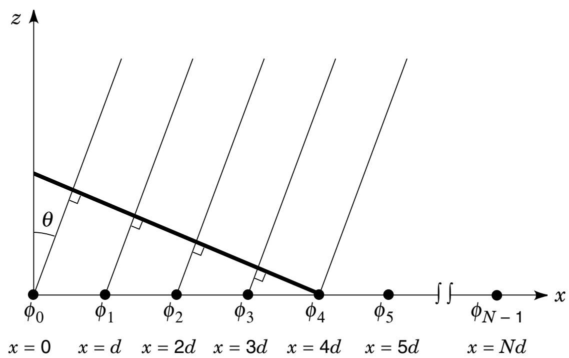


Figure 32. Depiction of a linear array of N identical radiating elements.

Consider a linear array such as is depicted in Fig. 32. Now let us say that the elements are nominally identical apart from phases that are set by the array operator at each of the elements. The complex electric field far from the nth element due to only the nth element is then given by

$$\boldsymbol{E}_n = \boldsymbol{E}_e e^{i\phi_n}$$

where $\boldsymbol{E}_e$ is the electric field of a single element. To find out what is radiated in the direction θ due to the whole array, we need to sum the fields from all of the radiators, giving each radiator the proper phase delay. Each element will get a progressive phase shift $kd \sin\theta$ due to its position (see Fig. 32), where k is the free-space propagation factor, given by

$$k = \frac{2\pi}{\lambda}$$

where λ is the free-space wavelength. With this, we can write for the total field radiated into the direction θ due to all n elements

$$\boldsymbol{E}_t(\theta) = \boldsymbol{E}_e \sum_{n=0}^{N-1} e^{-inkd\sin\theta} e^{i\phi_n}$$

The sum is generally referred to as the *array factor.* The intensity, then, in the θ direction is

$$I_t(\theta) = I_e \left| \sum_{n=0}^{N-1} e^{-iknd\sin\theta} e^{i\phi_n} \right|^2$$

One notes immediately that, if one sets the phases ϕ_n to

$$\phi_n = nkd\sin\theta$$

then the intensity in the θ direction is N^2 times the intensity due to a single element. This is the effect of coherent addition. One gets a power increase of N plus a directivity increase of N. To illustrate, let us consider the broadside case where we take all the ϕ_n to be zero. In this case, we can write the array factor in the form

$$\left| \sum_{n=0}^{N-1} e^{-ind\sin\theta} \right|^2 = \left| \frac{1 - e^{-iNkd\sin\theta}}{1 - e^{-ikd\sin\theta}} \right|^2$$

which in turn can be written as

$$\mathrm{AF} = \frac{\sin^2\left(N\dfrac{kd}{2}\sin\theta\right)}{\sin^2\left(\dfrac{kd}{2}\sin\theta\right)} \tag{6}$$

which is plotted in Fig. 33. Several interesting things can be noted from the expression and plots. For kd less than π, there is only one central lobe in the pattern. Also, the pattern becomes ever more directed with increasing N. This is called the *directivity effect.* If the array has a power-combining efficiency of 100% (which we have built into our equations by ignoring actual couplings, etc.), then the total power radiated can only be N times that of a single element. However, it is radiated into a lobe that is only $1/N$ times as wide as that of a single element.

If we are to realize array gain, however, we need to be certain that the array elements are identical in frequency and have fixed phase relations in time. This can only take place if the elements are locked together. The idea of locking is probably best understood in relation to the Van der Pol equation (16), with an injected term, such that

$$\frac{\partial^2 v}{\partial t^2} - \frac{R_{i\phi} - R_T}{L}\left(1 - \frac{mv^2}{R_{i\phi} - R_T}\right)\frac{\partial v}{\partial t} + \omega_0^2 v = A\cos\omega_i t$$

where $R_{i\phi}$ is the input resistance of the transistor circuit as seen looking into the gate source port and R_T is the external termination resistor placed between the gate and common source. In the absence of the locking term, one can see that oscillation will take place with a primary frequency (and some harmonics) at angular frequency ω_0 with amplitude $\sqrt{R_{i0} - R_T/m}$ such that

$$v(t) \approx \sqrt{\frac{R_{i0} - R_T}{m}} \cos\omega_0 t$$

Without being too quantitative, one can say that, if ω_i is close enough to ω_0 and A is large enough, the oscillation will lock

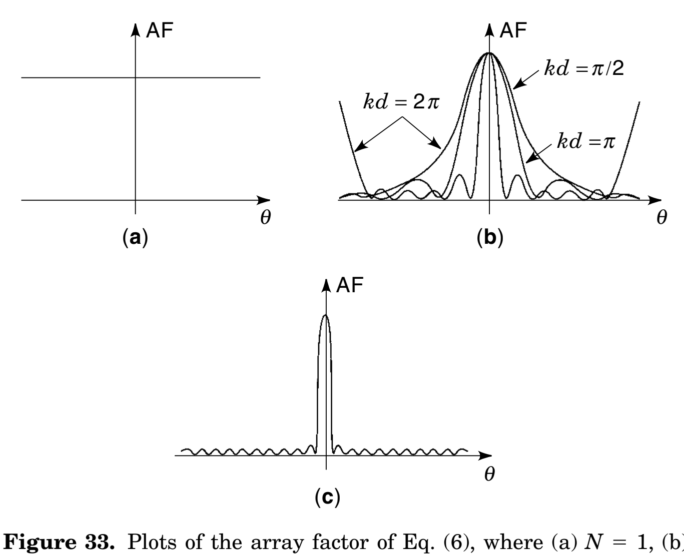


Figure 33. Plots of the array factor of Eq. (6), where (a) $N = 1$, (b) $N = 5$ and $kd = \pi/2$, π, and 2π, and (c) $N = 10$ and $kd = \pi$.

to ω_i in frequency and phase. If ω_i is not quite close enough and A not quite big enough (how big A needs to be is a function of how close ω_i is), then the oscillation frequency ω_0 will be shifted so that

$$v(t) = A_0 \cos[(\omega_0 + \Delta\omega)t + \phi]$$

where $\Delta\omega$ and ϕ are functions of ω_i and A. These ideas are discussed in a number of places including Refs. 1, 15, 16, 22, 23, and 24. In order for our array to operate in a coherent mode, the elements must be truly locked. This locking can occur through mutual coupling or through the injection of an external signal to each of the elements.

Ideally, we would like to be able to steer the locked beam. A number of techniques for doing this are presently under investigation. Much of the thinking stems from the work Stephan (25–28) and Vaughan and Compton (28a). One of the ideas brought out in these works was that, if the array were mutually locked and one were to try to inject one of the elements with a given phase, all of the elements would lock to that phase. However, if one were to inject two elements at the locked frequency but with different phases, then the other elements would have to adjust themselves to these phases. In particular, if one had a locked linear array and one were to inject the two end elements with phases differing by ϕ, then the other elements would share the phase shift equally so that there would be a linear phase taper of magnitude ϕ uniformly distributed along the array.

A different technique was developed by York (29,30), based on work he began when working with Compton (31,32). In this technique, instead of injecting the end elements with the locked frequency and different phase, one injects with wrong frequencies. If the amplitudes of these injected frequencies are set to values that are not strong enough to lock the elements to this wrong frequency, then the elements will retain their locked frequencies but will undergo phase shifts from the injected signal. If the elements of the array are locked due to mutual feedback, trying to inject either end of the array with wrong frequencies will then tend to give the elements a linear taper—that is, one in which the phase varies linearly with distance down the array—with much the same result as in the technique of Stephan. This will just linearly steer the main lobe of the array off broadside and to a new direction. Such linear scanning is what is needed for many commercial applications such as tracking or transmitting with minimum power to a given location.

Another technique, which again uses locking-type ideas, is that of changing the biases on each of the array's active devices (33–35). Changing the bias of a transistor will alter the ω_0 at which the active antenna wants to oscillate. For an element locked to another frequency, then, changing the bias will just change the phase. In this way one can individually set the phase on each element. There are still a couple of problems with this approach (as with all the others so far, which is why this area is still one of active research). One is that addressing each bias line represents a great increase in the complexity that we were trying to minimize by using an active antenna. The other is that the maximum phase shift obtainable with this technique is $\pm\pi$ from one end of the array to the other (a limitation that is shared by the phase-shifts-at-the-ends technique). In many phased-array applications, of which electronic warfare is a typical one, one wants to have true time delay, which means that one would like to have as much as a π phase shift between adjacent elements. I do not think that the frequency-shifting technique can achieve this either. Work, however, continues in this exciting area.

APPLICATIONS OF AND PROSPECTS FOR ACTIVE ANTENNAS

Perhaps the earliest application of the active antenna concept (following that of Hertz) was aimed at solving the small-antenna problem. As we recall, an antenna can be modeled (roughly) by a series RLC network with the R representing the radiation resistance. The input impedance of such a combination is given by

$$Z_i = \frac{1 - \omega^2/\omega_0^2 + j\omega RC}{j\omega C}$$

and so we see that, when the operation frequency ω is well below the resonant frequency

$$\omega_0 = \frac{1}{\sqrt{LC}}$$

and the reciprocal of the RC time constant

$$\tau = RC$$

then the antenna appears as a capacitor and radiates quite inefficiently. The problem of reception is similar. Apparently, already in 1928 Westinghouse had a mobile antenna receiver that used a pentode as an inductive loading element in order to boost the amount of low-frequency radiation that could be converted to circuit current. In 1974, two works discussed transistor-based solutions to the short-aerial problem (36,37). In Ref. 37, the load circuit appeared as in Fig. 34. The idea was to generate an inductive load whose impedance varied with frequency, unlike a regular inductor, but so as to increase the antenna bandwidth. The circuit's operation is not intuitively obvious. I think that it is possible that most AM, short-wave, and FM receivers employ some short-antenna solution whether or not the actual circuit designers were aware that they were employing active antenna techniques.

Another set of applications where active devices are essentially used as loading elements is in the greater-than-100-

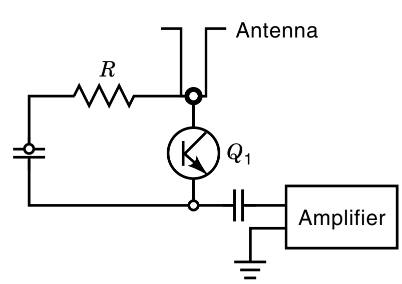


Figure 34. A circuit taken from Ref. 37 in which a transistor circuit is used to load a short antenna. Analysis shows that, in the frequency regime of interest, the loading circuit appears, when looking toward the antenna from the amplifier terminals, to cancel the strongly capacitive load of the short antenna.

GHz regime. Reviews of progress in this regime are given in Refs. 1 and 38. To date, most work at frequencies greater than 100 GHz has involved radio-astronomical receivers. A problem at such frequencies is a lack of components, including circuit elements so basic as waveguides. Microstrip guides already start having extra-mode problems at Ku band. Coplanar waveguides can go higher, although to date, rectangular metallic waveguides are the preferred guiding structures past about 60 GHz. In W band (normally narrowband, about 94 GHz—see Table 1), there are components, as around 94 GHz there is an atmospheric window of low propagation loss. However, waveguide tolerances, which must be a small percentage of the wavelength, are already severe in W band, where the wavelength is roughly 3 mm. Higher frequencies have to be handled in free space or, as one says, quasi-optically. Receivers must therefore by nature be downconverting in this >100 GHz regime. Indeed, these types of solutions are the ones being demonstrated by the group at Michigan (38), where receivers will contain multipliers and downconverting mixers right in the antenna elements in order that CPW can be used to carry the downconverted signals to the processing electronics. Millimeter-wave–terahertz radio astronomy seems to be a prime niche for quasioptical active antenna solutions.

The first applications of active antennas where solid-state components were used as gain elements were primarily for power boosting (39–44). Power combining (see reviews in Refs. 45 and 46) can be hard to achieve. There is a theorem that grew out of the early days of radiometry and radiative transfer (in the 1800s), known variously as the brightness theorem, the Lagrange invariant, or (later) the second law of thermodynamics. (See, for example, Ref. 8, Chap. 5.) The theorem essentially states that one cannot increase the brightness of a source by passive means. This theorem practically means that, if one tries to combine two nominally identical sources by taking their outputs, launching them into waveguides, and then bringing the two waveguides together in a Y junction into a single waveguide, the power in the output guide, if the output guide is no larger than either of the input guides, can be no greater than that of either of the nominally identical sources. This seems to preclude any form of power combining. There is a bit of a trick here, though. At the time the brightness theorem was first formulated, there were no coherent radiation sources. If one takes the output of a coherent radiation source, splits it in two, and adds it back together in phase, then the brightness, which was halved, can be restored. If two sources are locked, they are essentially one source. (As P. A. M. Dirac said, a photon only interferes with itself. Indeed, the quantum mechanical meaning of locking is that the locked sources are sharing a wave function.) Therefore, locked sources can be coherently added if they are properly phased. We will take this up again in a following paragraph.

An alternative to power combining that obviates the need for locking and precise phase control is amplification of the signal from a single source at each element. By 1960, solid-state technology had come far enough that antennas integrated with diodes and transistors could be demonstrated. The technology was to remain a laboratory curiosity until the 1980s, when further improvements in microwave devices were to render it more practical. Recent research, however, has been more concentrated on the coherent power combining of self-oscillator elements. This is not to say that the element-mounted amplifier may not still be of practical use. The main research issue at present, though, is the limited power available from a single active element at millimeter-wave frequencies.

Another application area is that of proximity detection (47). The idea is that an oscillator in an antenna element can be very sensitive to its nearby (several wavelengths) environment. As was discussed previously, variation in distances to ground planes changes impedances. The proximity of any metal object will, to some extent, cause the oscillator to be aware of another ground plane in parallel with the one in the circuit. This will change the impedance that the oscillator sees and thereby steer the oscillator frequency. The active antenna of Ref. 47 operated as a self-oscillating mixer. That is, the active element used the antenna as a load, whereas the antenna also used a diode mixer between itself and a low-frequency external circuit. The antenna acted as both a transmitting and a receiving antenna. If there were something moving near the antenna, the signal reflected off the object and rereceived might well be at a different frequency than the shifting oscillator frequency. These two frequencies would then beat in the mixer, be downconverted, and show up as a low-frequency beat note in the external circuit. If such a composite device were to be used in a controlled environment, one could calibrate the output to determine what is occurring. Navarro and Chang (1, p. 130) mention such applications as automatic door openers and burglar alarms. The original paper (47) seemed to have a different application in mind, as the term *Doppler sensor* was in the title. If one were to carefully control the immediate environment of the self-oscillating mixer, then reflections off more distant objects that were received by the antenna would beat with the stable frequency of the oscillator. The resulting beat note of the signals would then be the Doppler shift of the outgoing signal upon reflection off the surface of the moving object, and from it one could determine the normal component of the object's velocity. It is my understanding that some low-cost radars operate on such a principle. As with other applications, though, the active antenna principle, if only due to size constraints, becomes even more appealing at millimeter-wave frequencies, and at such frequencies power constraints favor use of arrays.

An older antenna field that seems to be going through an active renaissance is that of retroreflection. A retroreflector is a device that, when illuminated from any arbitrary direction, will return a signal directly back to the source. Clearly, retroreflectors are useful for return calibration as well as for various tracking purposes. An archetypical passive retroreflector is a corner cube. Another form of passive reflector is a Van Atta array (48). Such an array uses wires to interconnect the array elements so that the phase progression of the incident signal is conjugated and thereby returned in the direction of the source. As was pointed out by Friis already in the 1930s, though, phase conjugation is carried out in any mixer in which the local oscillator frequency exceeds the signal frequency (49). (A *phase conjugate* signal is one that takes on negative values at each phase point on the incoming wave.) This principle was already being exploited in 1963 for implementing retroreflection (50). This work did not catch on, perhaps for technical reasons. A review in 1994 (51) and designs for such arrays were demonstrated and presented at the 1995 International Microwave Symposium (52,53). Although both demonstrations used transistors and patch-type elements,

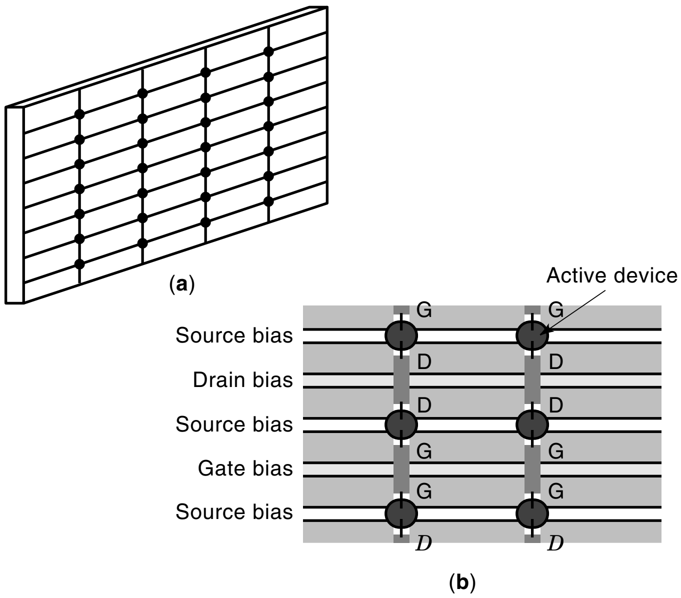


Figure 35. Schematic depiction of (a) the active surface of a grid oscillator and (b) a breakout of an internal region of the grid showing the active device placement relative to the bias lines.

both also employed circulators for isolation and therefore were not actually active array demonstrations. It would seem that retroreflection should motivate an active self-oscillating mixer solution, which will perhaps appear in the future.

As was mentioned earlier in this article, a quite important application area for active antennas is free-space power combining. As was pointed out then, a number of groups are working on developing compact elements such as those of Fig. 14 (7) and Fig. 30 (21). As was also previously mentioned, in order to do coherent power combining, the elements must be locked. In designs where the elements are spatially packed tightly enough, proximity can lead to strong enough nearest-neighbor coupling so that the array will lock to a common frequency and phase. Closeness of elements is also desirable in that arrays with less than $\lambda/2$ spacing will have no sidelobes sapping power from the central array beam. In designs that do not self-lock, one can inject a locking signal either on bias lines or spatially from a horn to try to lock to all elements simultaneously. Of course, the ultimate application would be for a high-bandwidth, steerable, low-cost transceiver.

Another method of carrying out power combining is to use the so-called *grid oscillator* (54,55). The actual structure of a grid appears in Fig. 35. The operating principle of the grid is quite a bit different from that of the arrays of weakly coupled individual elements. Note that there is no ground plane at all on the back, and there is no ground plane either, per se, on the front side. Direct optical measurements of the potentials on the various lines of the grid (56), however, show that the source bias lines act somewhat like ac grounds. In this sense, either a drain bias line together with the two closest source biases, or a gate bias line together with the two horizontally adjacent bias lines, appears somewhat like CPW. The CPW lines, however, are periodically loaded ones with periodic active elements alternated with structures that appear like slot antennas. The radiating edges of the slots are, for the drain bias lines, the vertical ac connection lines between drain and drain or, for the gate bias CPW, the horizontal ac gate-to-gate connection lines. Indeed, the grid is known to lock strongly between the rows and more weakly between columns. As adjacent row elements are sharing a patch radiator, this behavior should be expected.

In a sense, this strong locking behavior of the grid is both an advantage and a disadvantage. It is advantageous that the grid is compact (element spacing can be $\leq \lambda/6$) and further that it is easy to get the rows to lock to each other. However, the compactness is also a disadvantage in that it is quite hard to get any more functionality on the grid. Much effort has been made in this area to generate functionality by stacking various grid-based active surfaces such as amplifying surfaces, varactor surfaces for frequency shifting and modulation, doubling surfaces, etc. A problem with stacking is, of course, diffraction as well as alignment. Alignment tolerance adds to complexity. Diffraction tends to ease alignment tolerance, but in an inelegant manner. A 100-transistor array with $\lambda/6$ spacing will have an extent of roughly 1.5λ per side. As the diffraction angle is something like the wavelength divided by the array diameter, the diffraction angle for such an array is a good fraction of a radian. One can say that grids are quasi-optical, but in optics one generally doesn't use apertures much smaller than a millimeter (center optical wavelength of micrometers), for which the diffraction angle would be roughly a thousandth of a radian. As far as pure combining efficiency goes, grids are probably the optimal solution. However, more functionality may well be hard to obtain with this solution.

As we have mentioned, there are a number of techniques for steering being investigated. There seems to be less work on modulation, and I do not know of any simultaneous steering of modulated beams to date. Although the field of active antennas began with the field of radio frequency, it still seems to be in its infancy. However, as I hope this article has brought across, there is a significant amount of work ongoing, and the field of active antennas will grow in the future.

BIBLIOGRAPHY

1. J. A. Navarro and K. Chang, *Integrated Active Antennas and Spatial Power Combining,* New York: Wiley, 1995.
2. R. A. York and Z. B. Popović (eds.), *Active and Quasi-Optical Arrays for Solid-State Power Combining,* New York: Wiley, 1997.

2a. H. Hertz, *Electric Waves,* New York: Macmillan, 1983. (This is a book of reprints of Hertz' work in the 1890s.)

3. J. Bardeen and W. Bratain, The transistor: A semiconductor triode, *Phys. Rev.,* **74**: 435, 1948.
4. W. Shockley, The theory of *p–n* junctions in semiconductors and *p–n* junction transistors, *Bell Syst. Tech. J.,* **28**: 435, 1949.
5. W. Shockley, A unipolar field-effect transistor, *Proc. IEEE,* **40**: 1365–1376, 1952.
6. C. A. Mead, Schottky-barrier gate field-effect transistor, *Proc. IEEE,* **54**: 307–308, 1966.
7. R. A. York, R. D. Martinez, and R. C. Compton, Active patch antenna element for array applications, *Electron. Lett.,* **26**: 494–495, March 1990.
8. A. R. Mickelson, *Physical Optics,* New York: Van Nostrand Reinholt, 1992, chap. 2.
9. B. J. Hunt, *The Maxwellians,* Ithaca, NY: Cornell University Press, 1991, chap. 3.

10. H. C. Pocklington, Electrical oscillations in wires, *Proc. Cambridge Phil. Soc.,* 324–333, 1897.
11. D. M. Pozar, *Microwave Engineering,* Reading, MA: Addison-Wesley, 1990.
12. P. E. Gray and C. L. Searle, *Electronic Principles, Physics, Models and Circuits,* New York: Wiley, 1967.
13. R. E. Collin, *Foundations for Microwave Engineering,* 2nd ed., New York: McGraw-Hill, 1992.
14. K. Chang, *Microwave Solid-State Circuits and Applications,* New York: Wiley, 1994.
15. K. Y. Chen et al., Analysis of an experimental technique for determining Van der Pol parameters of a transistor oscillator, *IEEE Trans. Microw. Theory Tech.,* **46**: 914–922, 1998.
16. B. Van der Pol, Forced oscillations in a circuit with a nonlinear resistance, *Phil. Mag.,* **3**: 65–80, 1927.
17. K. R. Carver and J. W. Mink, Microstrip antenna technology, *IEEE Trans. Antennas Propag.,* **AP-29**: 2–24, 1981.
18. D. M. Pozar and D. H. Schauber (eds.), *Microstrip Antennas,* Piscataway, NJ: IEEE Press, 1995.
19. R. E. Munson, Conformal microstrip antennas and microstrip phased arrays, *IEEE Trans. Antennas Propag.,* **AP-22**: 74–78, 1974.
20. R. F. Harrington, *Time Harmonic Electromagnetic Fields,* New York: McGraw-Hill, 1961, p. 276.
21. K. Chang, K. A. Hammer, and G. K. Gopalakrishnan, Active radiating element using FET source integrated with microstrip patch antenna, *Electron. Lett.,* **24**: 1347–1348, 1988.
22. R. Adler, A study of locking phenomena in oscillators, *Proc. IRE,* **34**: 351–357, 1946.
23. R. Adler, A study of locking phenomena in oscillators, *Proc. IEEE,* **61**: 1380–1385, 1973. This is a reprint of (22).
24. K. Kurokawa, Injection locking of microwave solid-state oscillators, *Proc. IEEE,* **61**: 1386–1410, 1973.
25. K. D. Stephan, Inter-injection-locked oscillators for power combining and phased arrays, *IEEE Trans. Microw. Theory Tech.,* **34**: 1017–1025, 1986.
26. K. D. Stephan and W. A. Morgan, Analysis of inter-injection-locked oscillators for integrated phased arrays, *IEEE Trans. Antennas Propag.,* **35**: 771–781, 1987.
27. K. D. Stephan and S. L. Young, Mode stability of radiation-coupled inter-injection-locked oscillators for integrated phased arrays, *IEEE Trans. Microw. Theory Tech.,* **36**: 921–924, 1988.
28. W. A. Morgan and K. D. Stephan, An x-band experimental model of a millimeter-wave inter-injection-locked phase array system, *IEEE Trans. Antennas Propag.,* **36**: 1641–1645, 1988.
28a. M. J. Vaughan and R. C. Compton, 28 GHz omnidirectional quasioptical transmitter array, *IEEE Trans. Microw. Theory Tech.,* **MTT-43**: 2507–2509, 1995.
29. R. A. York, Nonlinear analysis of phase relationships in quasioptical oscillator arrays, *IEEE Trans. Microw. Theory Tech.,* **41**: 1799–1809, 1993.
30. P. Liao and R. A. York, A new phase-shifterless beam scanning technique using arrays of coupled oscillators, *IEEE Trans. Microw. Theory Tech.,* **41**: 1810–1815, 1993.
31. R. A. York and R. C. Compton, Quasi-optical power combining using mutual synchronized oscillator arrays, *IEEE Trans. Microw. Theory Tech.,* **39**: 1000–1009, 1991.
32. R. A. York and R. C. Compton, Coupled-oscillator arrays for millimeter-wave power-combining and mode-locking, *IEEE MTT-S Int. Microw. Symp. Digest,* 1992, pp. 429–432.
33. P. S. Hall and P. M. Haskins, Microstrip active patch array with beam scanning, *Electron. Lett.,* **28**: 2056–2057, 1992.
34. P. S. Hall et al., Phase control in injection locked microstrip active antennas, *IEEE MTT-S Int. Microw. Symp. Digest,* 1994, pp. 1227–1230.
35. A. Zarrang, P. S. Hall, and M. Cryan, Active antenna phase control using subharmonic locking, *Electron. Lett.,* **31**: 842–843, 1995.
36. T. S. M. Maclean and P. A. Ransdale, Short active aerials for transmission, *Int. J. Electron.,* **36**: 261–269, 1974.
37. P. K. Rangole and S. S. Midha, Short antenna with active inductace, *Electron. Lett.,* **10**: 462–463, 1974.
38. G. M. Rebeiz, Millimeter-wave and terahertz integrated circuit antennas, *Proc. IEEE,* **80**: 1748–1770, 1996.
39. A. D. Frost, Parametric amplifier antennas, *Proc. IRE,* **48**: 1163–1164, 1960.
40. J. R. Copeland and W. J. Robertson, Antenna-verters and antennafiers, *Electronics,* pp. 68–71, 1961.
41. M. E. Pedinoff, The negative conductance slot amplifier, *IRE Trans. Microw. Theory Tech.,* **9**: 557–566, 1961.
42. W. J. Robertson, J. R. Copeland, and R. G. Verstraete, Antennafier arrays, *IEEE Trans. Antennas Propag.,* **2**: 227–233, 1964.
43. K. Fujimoto, Active antennas: Tunnel-diode-loaded dipole, *Proc. IEEE,* **53**: 174, 1964.
44. H. H. Meinke, Tunnel diodes integrated with microwave antenna systems, *Radio Electron. Eng.,* **31**: 76–80, 1966.
45. K. J. Russell, Microwave power combining techniques, *IEEE Trans. Microw. Theory Tech.,* **27**: 472–478, 1979.
46. K. Chang and C. Sun, Millimeter-wave power-combining techniques, *IEEE Trans. Microw. Theory Tech.,* **31**: 91–107, 1983.
47. B. M. Armstrong et al., Use of microstrip impedance-measurement technique in the design of BARITT piplex Doppler sensor, *IEEE Trans. Microw. Theory Tech.,* **28**: 1437–1442, 1980.
48. E. D. Sharp and M. A. Diab, Van Atta reflector array, *IRE Trans. Antennas Propag.,* **8**: 436–438, 1960.
49. H. T. Friis and C. Feldman, A multiple-unit steerable antenna for short-wave reception, *Bell Syst. Tech. J.,* **16**: 337–419, 1937.
50. C. Y. Pon, Retrodirective array using the heterodyne technique, *IEEE Trans. Antennas Propag.,* **12**: 176–180, 1964.
51. B. S. Hewitt, The evolution of radar technology into commercial systems, *IEEE MTT-S Int. Microw. Symp. Digest,* 1994, pp. 1271–1274.
52. C. W. Poblanz and T. Itoh, A conformal retrodirective array for radar applications using a heterodyne phase scattering element, *IEEE MTT-S Int. Microw. Symp. Digest,* 1995, pp. 905–908.
53. Y. Chang, D. C. Scott, and H. R. Fetterman, Microwave phase conjugation using antenna coupled nonlinear optically pumped surface, *IEEE MTT-S Int. Microw. Symp. Digest,* 1995, pp. 1303–1306.
54. Z. B. Popovic, M. Kim, and D. B. Rutledge, Grid oscillators, *Int. J. Infrared Millimeter Waves,* **9**: 647–654, 1988.
55. Z. B. Popovic et al., A 100-MESFET planar grid oscillator, *IEEE Trans. Microw. Theory Tech.,* **39**: 193–200, 1991.
56. K. Y. Chen et al., Noninvasive experimental determination of charge and current distributions on an active surface, *IEEE Trans. Microw. Theory Tech.,* **44**: 1000–1009, 1996.

Alan R. Mickelson
University of Colorado

ACTIVE ANTENNAS. See Antenna arrays.
ACTIVE ARRAYS. See Quasi-optical circuits.
ACTIVE FILTERING. See Power system harmonics.