1 Phased Arrays for High-sensitivity Receiver Applications

Phased arrays date back to the very earliest days of radio. The German physicist Karl Ferdinand Braun constructed a three element, switchable array in 1909 to enhance radio transmission in one direction. Early phased arrays achieved beam steering through applying a progressive phase to each element of a one- or two-dimensional array; the concept may be found in almost every book on antenna theory, e.g. [1]. The contemporary usage extends to include control of both the amplitude and phase (or time-delay) excitations of each radiating element in a multiantenna system [2].

While the analytical tools covered in this book are applicable to phased array antennas for all applications, the concepts and examples in the book are organized around the design and optimization of high-sensitivity radio frequency and microwave receivers. Radio astronomy is an especially challenging application of this technology, and will feature strongly in this book. Although parabolic dishes have dominated antenna technology since the early 1960s, to the point where dishes have become largely synonymous with radio telescopes in the popular imagination,[1] many early discoveries in radio astronomy were made using phased arrays [3]. The same is true for the large dishes (often over 30 m in diameter) used by telecommunication ground stations and for deep space tracking in the same timeframe; but again, phased arrays were far from forgotten, playing an important role in the first Approach and Landing System (ALS) and post-WWII early warning systems.

Parabolic dishes have probably reached the apogee of their design in recent years, and since they are fundamentally large mechanical systems, their cost is dominated by the cost of materials and labour – neither of which is likely to change dramatically in the foreseeable future. In the radio astronomy community, the currently accepted guideline is that the cost of a dish scales as $D^{2.7}$ [4]; since the area only increases as D^2, building ever-larger steerable dishes is clearly not a viable method for increasing sensitivity, which is directly proportional to collecting area. Additionally, steerable dishes in particular involve moving parts, bringing significant maintenance requirements. Phased arrays, on the other hand, are fundamentally electronic systems, whose cost is increasingly dominated by processing. Moore's Law provides the prospect of continuing – and dramatic – reductions in processing costs.

[1] The iconic Arecibo dish has a spherical, not parabolic surface. Arecibo was the largest single aperture radio telescope in the world, until supplanted in 2016 by the Five hundred meter Aperture Spherical Telescope (FAST) in China.

Whilst dishes may continue to dominate some applications, there has been a resurgence of interest in phased arrays for radio astronomy since the 1990s, with a number of systems currently deployed or in an advanced state of design. In radio astronomy, these are generally known as either aperture arrays (AAs) or phased array feeds (PAFs), depending upon whether the array views the sky directly, or is placed at the focal plane of a dish. PAFs blend the most attractive features of both phased arrays and dishes; however, these are very challenging systems to design, and the accurate analysis of PAF-fed dishes has driven the development of state-of-the-art design methods. Overall, technological advances in areas such as signal processing, digital electronics, computing and data storage mean that this new generation of radio astronomy phased arrays is quite unlike its predecessors.

Beyond radio astronomy, highly sensitive receiving arrays find application as satellite communications ground terminals for mobile and airborne platforms, and in passive and active remote sensing applications. The common thread is that the external noise environment viewed by the receiver is primarily the cool microwave sky, which at L band has a brightness temperature of only 4–5 K. When the external noise environment has a low brightness temperature, the incremental gains of improving the receiver noise figure and reducing antenna losses are much more significant than for terrestrial communication and radar applications, for which the external noise environment is closer to the ambient temperature (280 K) or is dominated by interference from other transmitters. Traditional methods for antenna analysis and design are often insufficiently accurate to realize phased array antennas that can compete with reflector-based receivers. The goal of this book is to develop a modern approach to phased array design suitable for high-performance applications.

1.1 Contemporary Design Methods for Phased Arrays

The traditional approach to array design uses the element pattern and array factor approach. As we discuss in Chapter 4, for high performance receiving arrays, the approximate factorization of the array radiation pattern into an element pattern and array factor is not accurate enough to use when designing for stringent performance requirements. Mutual coupling causes element radiation patterns to differ across the array, and this must be taken into account in the design process from the beginning. Whilst we review classical array factor analysis, the focus of this book is on more sophisticated methods based on overlap integrals and network theory. This is particularly important for small to moderate sized arrays, such as the phased array feeds which are currently being deployed to replace single feeds on parabolic dishes.

The array factor method certainly has value, and for some receiving array applications considered in this book, such as large, dense, aperture arrays for radio astronomy, it is still a useful method, especially when combined with numerical simulations using infinite array analysis (which do incorporate mutual coupling, albeit in the infinite array environment). For other applications, the array factor method is only useful for rough designs, and the more advanced approach using the overlap integral and network theory approach which will be presented in this book is needed. This approach is fully integrated with numerical methods and optimization tools from the ground up. For the core material, network analysis of phased array antennas, problems are included to guide the reader in understanding and implementing the techniques outlined in the book, making it suitable for use as a textbook in a graduate level course on phased array antennas.

The design of highly sensitive receiving systems requires leveraging modern simulation methods. Key tools ubiquitously applied in the current design of phased arrays are computational electromagnetic (CEM) simulation codes. The ability to rigorously numerically simulate antenna performance is an important enabling technology for this contemporary approach. Both basic numerical techniques and advanced, highly efficient algorithms are surveyed in this book. Array design based on computer simulation is embedded in much of the treatment, and Chapter 8 specifically discusses CEM as it applies to numerically modeling phased arrays.

Regarding numerical simulation methods as used in the design of modern phased array antenna systems, there are two parallel threads in the book. The first is the lossless, resonant, minimum scattering approximation, which allows antenna elements with simple, analytically known radiation patterns to be combined into an array and modeled with mutual coupling effects accounted for with a simple, easily written software code, not requiring sophisticated or expensive software packages. This method represents a natural next step beyond the classical array factor approximation.

The second thread is the use of full-wave, highly accurate and powerful CEM tools to model the antenna array, as described in Chapter 8. CEM simulations can be combined with the overlap integral and network theory formulation to embed the antenna array model into a full system model. The full system model includes the antenna array, a reflector in the case of a phased array feed, receiver electronics, and calibration, beamforming and imaging algorithms in the digital back end or analog beamforming network, and allows the antenna array design and electronics to be optimized to maximize overall system level figures of merit, such as sensitivity, scan range or field of view, bandwidth, and survey efficiency in the case of an astronomical imaging receiver. Figures of merit are dealt with in brief later in this chapter (Sec. 1.5) and in more detail in Chapter 6.

Beamforming is a fundamental function of receiving arrays, and Chapter 10 considers this is detail. Since this implies spatial filtering capability, arrays have some capability to reject or at least mitigate radio frequency interference (RFI). In applications where the source power is fixed, and often very small due to extremely long distances (as encountered in remote sensing and even more acutely in radio astronomy), RFI is an ever-increasing problem. RFI mitigation is also addressed in this chapter.

Most contemporary radio astronomy telescopes comprise large, very sparse, arrays, using interferometry to synthesize a very large aperture and produce high-resolution images of the sky. Each element (or station in radio astronomy parlance) may be a dish, or a beamformed array. Whilst for dishes, interferometry may be considered a mature field, the same is not true for aperture arrays. In particular, the behaviour of the primary beam (or the station beam) in the presence of mutual coupling means that this is a topic of current interest. Again, the availability of modern computer simulation tools plays an important role in this, which we consider in Chapter 11.

A contemporary high-sensitivity receiving array requires attention not just to the antenna elements and the front-end receivers, but also high-speed digital signal processing. Some fields, such as radio astronomy, require very fine frequency channelisation; methods to do this are addressed in Chapter 12, along with a discussion of field programmable gate array (FPGA) and graphics processing unit (GPU) hardware.

In this book, we sometimes use the term receiver to mean the antenna plus the front-end receiving system; this usage is widely encountered in active antenna systems. Some practitioners use the term receptor to denote this combination, but this term has also been used in some radio astronomy projects to denote the complete antenna, receiver, pedestal and control system.

1.2 Phased Arrays in Radio Astronomy

One of the most important motivating applications for advances in the techniques used to model and design phased arrays, both historically and in recent years, has been the field of radio astronomy. In this section, we cover the foundations of radio astronomical observations and survey applications of arrays as astronomical receivers.

1.2.1 Basics of Radio Astronomy

Radio astronomy, the study of electromagnetic radiation at radio frequencies from celestial objects, has a history of over 80 years, since Jansky's 1931 discovery of radio emission from the Milky Way. The discovery of radio emissions from celestial sources was unexpected at the time. Optical astronomy mainly addresses the surfaces of stars, or nearby gas ionized by stars, where the temperatures bring thermal radiation into the visible spectrum. As such, little radiation was expected in the radio spectrum, if stars were indeed the principal source of radiation. Jansky's discovery, and pioneering work by Grote Reber, of course changed this, but it was only in the 1950s and 1960s with the discovery of the 21 cm hydrogen line, quasars, pulsars and the cosmic microwave background, that this work was fully appreciated.

Astronomical sources

There are two absorption windows in the earth's atmosphere, through which electromagnetic radiation can propagate with minimal absorption. These are clearly visible in Fig. 1.1. The optical window is both well known and relatively narrow, with the ultraviolet end blocked by firstly ozone and then oxygen and nitrogen absorption, and the infrared end by water vapor and carbon dioxide absorption. The blockage at the ultraviolet end is so complete that ultraviolet and X-ray work must be carried out above the atmosphere. There are some windows in infrared wavelengths, permitting ground-based observations from high, dry mountain sites, but generally infrared observations must be done from air- or spaceborne platforms.

For radio, the window extends over some five decades, from approximately 15 MHz ($\lambda \approx$ 20 m) to around 1.5 THz ($\lambda \approx$ 0.2 mm); however, above microwave frequencies, the millimeter regime is characterized by several bands of poor transmittance. At this higher frequency end, the cut-off occurs due to absorption by primarily water vapour and O_2; millimeter wave observatories (such as the Atacama Large Millimeter/submillimeter Array, ALMA) also require high, dry sites (for ALMA, the Atacama desert). The lower frequency end is due to presence of free electrons in the Earth's ionosphere, and is given by the plasma (or critical) frequency

$$f_p = 9\sqrt{N_\mathrm{e}} \tag{1.1}$$

with the plasma frequency f_p in Hz and N_e the electron density in electrons/m^3. The value of N_e varies between day and night time, depending on the electron density; a typical value is around 10^{12} electrons per cubic meter, corresponding to a plasma frequency of 9 MHz. It can be as low as around 4.5 MHz at night, and is typically around 11 MHz in the day, but this depends both on location on Earth and on solar activity. Radio astronomy below the plasma frequency must be done from space, as below this frequency, the wave is totally reflected by the ionosphere.

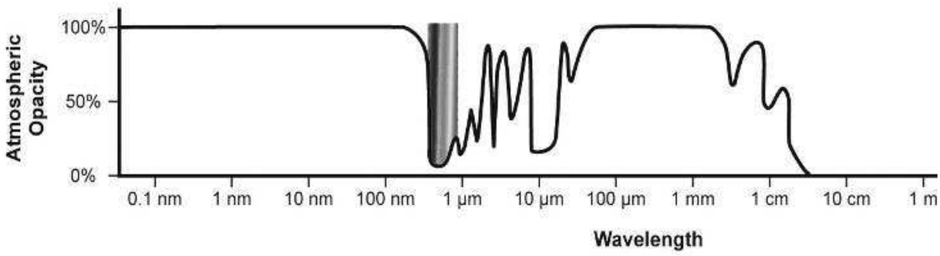

Figure 1.1 A sketch of Earth's atmospheric opacity at various wavelengths of electromagnetic radiation, including visible light. Image credits: NASA, public domain.

Radio astronomical signals are very weak. Received signals are often 30–50 dB or more below the system noise floor. The signals are noise-like, and can be narrowband spectral lines or broadband. The polarization states of the radiated electromagnetic fields vary from random to strongly polarized and are usually characterized by Stokes parameters [5]. This contrasts with manmade transmissions such as those of communication and radar systems, which are fully polarized, coherent emissions, from carefully modulated currents in the transmitting antenna, which are designed to use their allocated radio spectrum as efficiently as possible, and generally carry appreciable signal power. Astronomical signals in the radio band originate mainly in the incoherent emissions from huge clouds of faintly emitting atoms, molecules, or energetic electrons whose intensity is a collective effect arising from their large numbers. The radiated powers, except in a few cases, are added incoherently. Astronomical signals in general are well represented by stochastic or Gaussian noise [6].

As with optical astronomy, radio astronomical observations are generally either continuum (broadband) or spectral line (narrowband). From continuum observations, one obtains the general shape of the spectrum, characterized by the brightness (to be defined shortly) versus frequency. Typical continuum non-thermal sources are bremsstrahlung and synchrotron radiation. Bremsstrahlung, or free-free, radiation is broadband electromagnetic radiation produced by the acceleration of a free electron in the field of an ion [7, p. 133]. Synchrotron radiation arises from electrons with large relativistic energy accelerating through magnetic fields [7, p. 127]. In spectral line observations, the radiation is detected in small frequency intervals, and the shapes of the spectral lines determined. The neutral hydrogen (HI) emission line epitomizes a spectral line. The HI line is due to the hyperfine transition in the ground state of the hydrogen atom; a photon is emitted at 21 cm wavelength (1420.406 MHz) when the atom flips from the parallel to antiparallel spin configurations of the proton and electron [7, p. 175].

Astronomical objects of interest may be either effectively point sources, or have extended structure on the sky. An example of the former is a pulsar; an example of the latter is a galaxy close to our own. When the resolution of the telescope is significantly smaller than the source extent, the source is resolved; conversely, for very small sources, they are unresolved.

Sky brightness and source flux density

Measuring the sky brightness, $\mathcal{B}$, as a function of angle, is the basic aim of observations. Sky brightness (also known as surface brightness, or simply brightness, and also often conflated with intensity, which has the same units) is defined as infinitesimal power dW per unit area per solid angle $d\Omega$ per unit frequency, with units $\mathrm{Wm^{-2}Hz^{-1}rad^{-2}}$; see Fig. 1.2.

Brightness is a function of position and frequency. In astronomy, the position of objects is usually described using declination (δ) and right ascension (RA) (or equivalently hour angle), the celestial equivalents of terrestrial latitude and longitude respectively. Later, it will be seen that in radio interferometry, direction cosines provide an appropriate angular reference system, although the imaging software will usually provide the images in the conventional RA–δ system.

The angular integral of the brightness distribution gives the source flux density:

$$\mathcal{S} = \iint \mathcal{B}(\theta, \phi)\, d\Omega \tag{1.2}$$

The SI units are $\mathrm{Wm^{-2}Hz^{-1}}$, but in recognition of the very low fluxes of typical celestial sources, and honouring Karl Jansky's contributions, the Jansky (Jy) is extensively used nowadays. One Jy is $10^{-26}\ \mathrm{Wm^{-2}Hz^{-1}}$, and the flux densities of the most powerful celestial radio sources are typically on the order of tens of Janskys. Brightness is sometimes given in the radio astronomy literature as Jy/beam area or often Jy/beam.[2]

The source flux density is generally used for randomly polarized signals radiated by wideband, continuum sources, and includes power in two orthogonal polarizations. In electromagnetic modeling and antenna analysis, it is common to analyze systems with

[2] One should be aware that this is not a proper spectral brightness as it depends on the synthesized beam solid angle and not just on the radio source. For a more detailed discussion, including the definition of the beam area, see [8, §10.6].

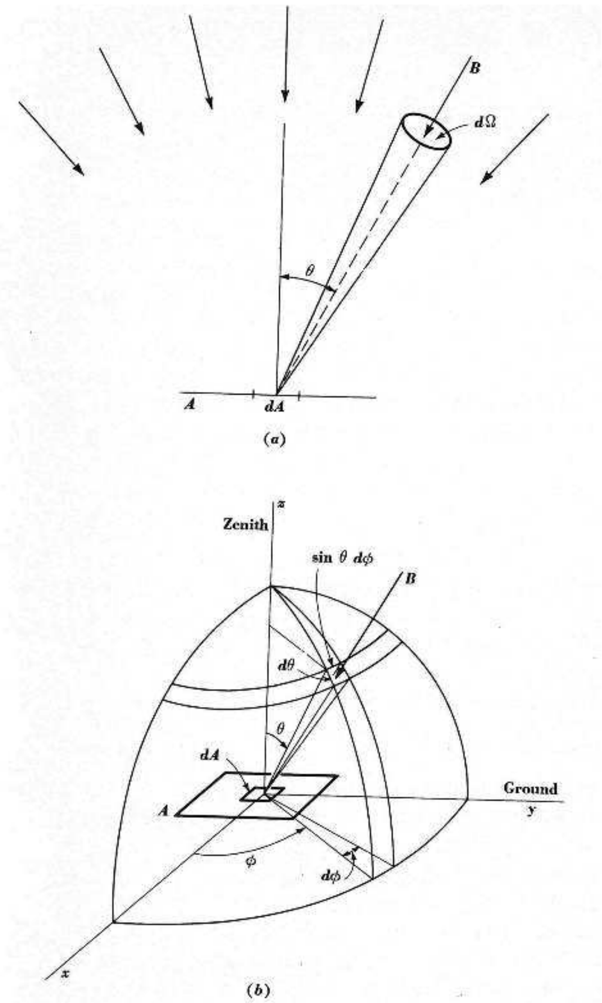

Figure 1.2 Basic geometry for radiation of brightness $\mathcal{B}$ incident on a flat area shown in elevation in (a) and perspective in (b). Reprinted from [5, Fig. 3.1] with the permission of the Estate of John D. Kraus.

single frequency or time harmonic signals. When this is done, the electric field intensity is often modeled as a complex phasor (see Sec. 2.1 for further discussion of the phasor notation). The time average Poynting power flux density vector field can be shown to be

$$\overline{S} = \frac{1}{2}\mathrm{Re}\left[\overline{E} \times \overline{H}^*\right] \tag{1.3}$$

where $\overline{E}$ and $\overline{H}$ are the phasor electric and magnetic field intensity vectors. For a time harmonic plane wave, the magnitude of the Poynting flux density is

$$S = \frac{|\overline{E}|^2}{2\eta} \quad (\mathrm{W/m^2}) \tag{1.4}$$

in terms of the electric field intensity $\overline{E}$. For a single frequency wave, the quantity in (1.4) is not a power spectral density and does not include the Hz^{-1} unit.

If $\overline{E}$ represents a randomly polarized wave, then the Poynting flux density contains the power density in two orthogonal polarizations. When analyzing antenna systems, it is common to illuminate the antenna with singly polarized waves, and to consider two orthogonal polarizations separately. If this is done, then the Poynting flux density (1.4) is half of the total power flux density associated with the randomly polarized wave. In many formulas in this book which use the Poynting flux density for a singly polarized wave, such as (2.75), this leads to a factor of two difference between the expression as given and its modification in terms of the source flux density $\mathcal{S}$, which here includes the power in both polarizations for randomly polarized waves. In astronomically focused treatments, the source flux as a power spectral density and including power in both polarizations is favored, whereas in antenna analysis, it is more convenient to use the Poynting flux density associated with a single frequency and a single polarization.

Radiometry, minimum detectable signal level, and resolution

In an astronomical receiver, the output power is generally integrated to achieve a highly stable and accurate estimate of the system noise. The signal of interest is then detectable as a perturbation in frequency, time, or space to the system noise power estimate. Detecting a weak signal as a perturbation between an on-source power measurement and an off-source measurement is referred to as radiometry.

The statistical variation of the integrated total system noise power determines the minimum detectable signal level or radiometer equation [5]

$$\Delta\mathcal{S} = \frac{2k_\text{B}T_\text{sys}}{A_\text{e}\sqrt{B\tau}} \tag{1.5}$$

where $\Delta\mathcal{S}$ is the minimum detectable source flux density in $\text{W/m}^2\text{/Hz}$, k_B is Boltzman's constant ($\approx 1.3806488 \times 10^{-23}\ \text{WHz}^{-1}\text{K}^{-1}$), A_e is receiving area in m^2, B is the system noise equivalent bandwidth set by filters or other signal processing components in the receiver, τ is the integration time in s, and T_sys is the system equivalent noise temperature in K.

To detect weak astronomical signals, the sensitivity A_e/T_sys or G/T_sys, where G is the gain, must be as large as possible. Consequently, development efforts in astronomical instrumentation have emphasized large collecting area, low system temperature, and wide bandwidth. Integration time is a function of the digital back-end, as well as the science case; times in the order of seconds are common and much longer times (minutes to hours) may be required for deep surveys.

Another key parameter is the angular resolution of a radio telescope, which is given approximately by

$$\Delta\theta \approx \frac{\lambda}{D} \tag{1.6}$$

where $\Delta\theta$ is the resolution in radians, λ is the wavelength, and D is the dish diameter (for a single dish) or the longest baseline (in an interferometric array). In short, the inverse of the largest dimension of the telescope, measured in wavelengths, gives the resolution of the instrument, at least to first order. When high resolution is paramount, this motivates widely distributed and very spare aperture synthesis arrays with long baselines.

Thermal noise and blackbody radiation

With the exception of the cosmic microwave background radiation, most radio sources are non-thermal in origin – in other words, they are not the result of blackbody radiation, but of other physical processes. Other unwanted noise sources prevalent in high sensitivity receivers, including ground noise, noise introduced by losses in the antenna, and electronic noise are either thermal in nature or represented as equivalent thermal noise sources.

Blackbody radiation is characterized by the Planck radiation law:

$$\mathcal{B} = \frac{2hf^3}{c^2}\frac{1}{e^{hf/k_{\mathrm{B}}T} - 1} \tag{1.7}$$

Here, f is frequency in Hz, h is Planck's constant ($\approx 6.63 \times 10^{-34}$ J s), T is temperature in K, and c is the velocity of light in m/s. At typical radio frequencies, $hf \ll k_{\mathrm{B}}T$ and the Rayleigh–Jeans approximation follows:

$$\mathcal{B} \approx \frac{2k_{\mathrm{B}}T}{\lambda^2} \tag{1.8}$$

which is the low frequency approximation of classical physics, pre-Planck. This is also used ubiquitously in the modeling of noise in electronic systems, as will be seen later in this book.

Despite most radio-sources being of non-thermal origin, equivalent temperatures of radio sources are frequently given using

$$T_{\mathrm{eq}} = \frac{\mathcal{B}\lambda^2}{2k_{\mathrm{B}}} = \frac{\mathcal{B}c^2}{2k_{\mathrm{B}}f^2} \tag{1.9}$$

This is often useful in system calculations, since system noise is conveniently described using equivalent noise temperatures.

Science goals in radio astronomy

Instrument development in radio astronomy is fundamentally driven by science. Current topics of intense interest include probing the cosmic dawn (by imaging neutral hydrogen); pursuing the "Cradle of Life" (including the successful search for exoplanets, and the to-date unsuccessful Search for Extra-Terrestrial Intelligence, SETI); testing Einstein's theory of gravity; studying how galaxies evolve and looking for dark matter; and understanding cosmic magnetism. These are the amongst the top science drivers for the international Square Kilometre Array (SKA) project. Optimal specifications for the telescope differ between science cases, and some compromises are required. Furthermore, a number of the most notable discoveries in radio astronomy have been serendipitous, including of course Jansky's original observations of the radio sky back in 1931, but also more recently the accidental discovery of the cosmic microwave background radiation by Penzias and Wilson in 1964 and Hewish and Bell Burnell's discovery of pulsars in 1967.[3] As such, no matter how compelling a particular science case may be at a particular point in time, flexible instrumentation is a key requirement in contemporary radio astronomy.

For radio astronomy observatories operating in the radio and microwave bands, the development of increasingly sensitive instruments and an ever-growing amount of RFI have driven radio astronomers to remote sites, starting with the VLA which was built in New Mexico, USA and other instruments that are located far from populated areas. The SKA, which will be mentioned on many occasions in this book, is currently being built on two sites, one in the Western Australian outback, and the other in the Karoo semi-desert in South Africa.

To read more about radio astronomy, Kraus' textbook remains a classic [5], [9]. Although his descriptions of instrumentation have obviously been overtaken by events, his treatments of the engineering electromagnetic basics of radio astronomy are still widely referenced.[4] The chapters on polarization and wave propagation are particularly insightful, presenting the development of material (such as Stokes parameters) often referenced only in a very terse summary form in more modern texts. Kraus is also meticulous with his use of units, unlike some later texts. Burke and Graham-Smith's book is widely referenced, current, and provides a good overview of the field [7], as does the text by Wilson, Rohlfs and Hüttermeister [10]. On pulsars, Lorimer and Kramer's book is probably the standard reference [11]. On interferometry, Thompson, Moran, and Swenson is without question the canonical reference [12]; as the present book was going to print, a third edition was published in open access format [8]. Usually only the third edition is cited here unless there is a different section reference in the earlier edition, or for copyright purposes. A concise history of the origins of radio astronomy may be found in [7, Appendix 3]. A more extensive history of its early days is presented in [13]. For a wider perspective on astronomy in general, Kutner's text provides a comprehensive introduction [14].

1.2.2 Types of Astronomical Arrays

Phased arrays are not new to radio astronomy, having been used since the early days of the field. An interferometric array was used by Ryle in 1946 to increase the effective aperture size and achieve a narrower effective beamwidth as compared to a single antenna. Interferometers have matured technologically into large-scale synthesis imaging arrays for high-resolution observations. Synthesis arrays include the Very Large Array (VLA) and the Very Long Baseline Interferometer (VLBI). These instruments typically consist of multiple medium-sized parabolic reflector or dish antennas located many wavelengths apart. Since the element spacing is much larger than the wavelength, these arrays are highly sparse. Correlation products of received signals by each array

[3] The former saw Penzias and Wilson share the Nobel Prize in Physics in 1978; however, Bell Burnell was excluded from the 1974 prize, awarded jointly to Ryle and Hewish for their work in radio astronomy.
[4] Unusually, the 2nd edition was self-published; it can be readily identified as it was ring-bound. Although long out of print, both editions may be found in specialist second-hand bookshops. The 1st edition is still well worth acquiring, as it contains the core material on engineering electromagnetics.