978-1-009-60623-3 — Satellite Radar Interferometry
David T. Sandwell , Xiaohua Xu , Jingyi Chen , Robert J. Mellors , Meng Wei , Xiaopeng Tong ,
John B. DeSanto , Qi Ou

1

Introduction

Interferometric synthetic aperture radar (InSAR) is an active remote sensing method that uses repeated radar scans of the Earth's solid surface to measure relative deformation at cm precision over a wide swath. This method has revolutionized our understanding of the earthquake cycle, volcanic eruptions, landslides, glacier flow, ice grounding lines, ground fluid injection/withdrawal, underground nuclear tests, and many other applications that require high spatial resolution measurements of ground deformation (Biggs and Wright, 2020; Joughin et al., 2010; Galloway and Hoffmann, 2007; Wei, 2017). The most recent generation of InSAR satellites has transformed the method from investigating 10s to 100s of synthetic aperture radar (SAR) images to processing 1 000s and 10 000s of images using a wide range of computer facilities. This transformation was largely driven by the Sentinel-1 mission, which was designed for InSAR time series analysis. There are a number of excellent review papers on this topic ranging from InSAR overviews (e.g., Massonnet and Feigl (1998); Bürgmann et al. (2000)) to detailed mathematical descriptions of the measurement and associated errors (Rosen et al. (2000); Simons and Rosen (2007, 2015); Franceschetti and Lanari (2018)).

The focus of this book is to provide the basic physical principles of the method as well as the mathematical model and algorithms used in InSAR processing. These algorithms are implemented in a computer software package called GMTSAR – an open-source (GNU General Public License) InSAR processing system designed for users familiar with generic mapping tools (GMT – (Wessel et al., 2019)). Although the book refers to this particular software, it is also relevant to other modern InSAR processing systems (e.g., ISCE – (Rosen et al., 2012), GAMMA – (Wegnüller et al., 2016), SNAP). This book has been developed over the past 25 years from a Space Geodesy class at Scripps Institution of Oceanography as well as GMTSAR short courses at UNAVCO and EarthScope. Each chapter has a set of problems and solutions to highlight the important concepts. In addition, we will answer 16 basic questions related to InSAR, including:

1. What are the six main categories of satellite remote sensing and why is active remote sensing at microwave wavelengths (2.5 cm–25 cm) best for measuring Earth deformation? (Section 1.2)
2. Why are SAR antennas typically 10 m long in the along-track direction and 1–2 m wide in the cross-track direction? (Section 2.3)
3. What factors control the spatial resolution of a SAR image? (Sections 2.4)
4. What factors limit the width of a SAR swath? (Section 2.5)
5. What are the typical orbital characteristics (e.g., altitude, speed, repeat cycle) of an InSAR satellite and why? (Section 3.2)
6. A C-band SAR operates at a 5 GHz carrier frequency. Why can the full-resolution radar data be recorded at a much lower rate of perhaps 15 MHz? (Section 4.2)
7. How is the raw SAR data focused by a computer to obtain a high spatial resolution image of amplitude and phase? (Chapter 4)
8. What orbit accuracy is needed to achieve an optimally focused SAR image? (Sections 4.5 and 4.6)
9. How is an interferogram constructed from repeated SAR images and what are the main contributions to the phase? (Sections 5.1 and 5.2)
10. How close in space does the repeat orbit need to match the reference orbit to achieve high correlation and good phase recovery? (Section 5.5)
11. How does decorrelation depend on radar wavelength and surface properties? (Section 6.2)
12. What is the best way to detect phase unwrapping errors? (Section 7.5)
13. What are the three main methods for achieving wide swath InSAR? (Chapter 8)
14. What are the main sources of error in most interferograms? (Chapter 9)
15. How can one construct an InSAR time series using existing freely open data sets? (Chapter 10)
16. How can a scalar InSAR time series be integrated with a vector GNSS time series? (Chapter 11)

The processing details associated with the last two questions are addressed in a companion book entitled – *Satellite Radar Interferometry: Application – GMTSAR*. This book mainly covers the theoretical aspects of InSAR. We will begin with a very brief overview of satellite remote sensing to answer the first question.

1.1 Essential Ingredients for Satellite Remote Sensing

Satellite remote sensing of the surface of the Earth relies on four essential ingredients:

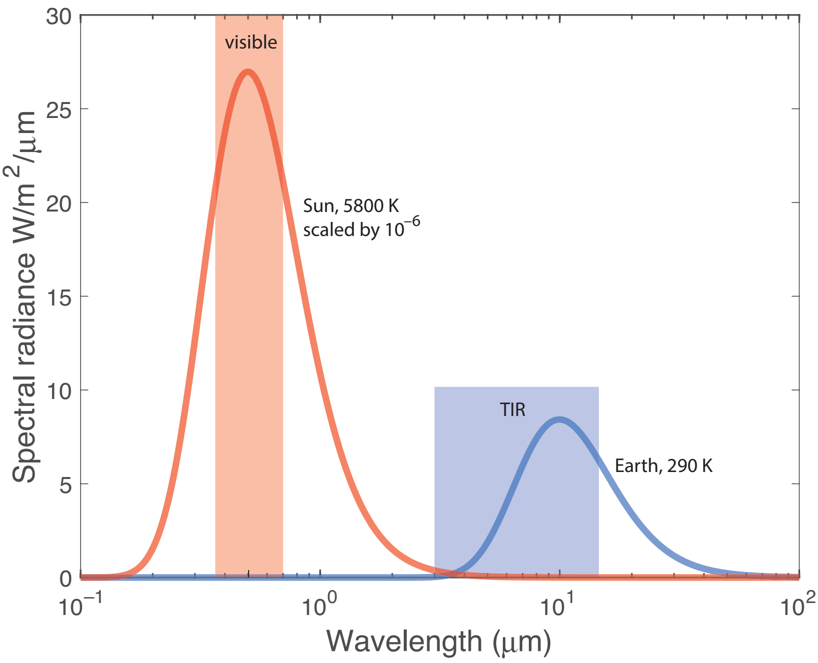

Figure 1.1 Blackbody radiation curves for a temperature of 5 800 K corresponding to the Sun and 290 K roughly corresponding to the Earth. The peak in the solar radiation (scaled by 10^{-6}) is centered on the visible part of the spectrum between 400 and 800 nanometers. The peak in the Earth's thermal radiation occurs in the thermal infrared (TIR) between 3 and 15 micrometers.

First, there must be a source of electromagnetic (EM) radiation. This could be passive radiation (Figure 1.1) from reflected sunlight, which peaks in the visible part of the spectrum (red curve), or passive radiation from the Earth's thermal emissions, which peaks in the thermal infrared (TIR) part of the spectrum (blue curve). An alternate radiation source is for the satellite to illuminate the Earth using an active sensor. While this requires significant power to generate the EM signal, the amplitude and phase of the source can be well controlled so one can measure the round-trip travel time of the pulse.

Second, the EM radiation must be able to pass through the Earth's atmosphere. As shown in Figure 1.2, there are three main windows in the visible (shaded red), thermal infrared (TIR; shaded blue), and microwave region (shaded green).

Third, there needs to be a satellite in space to collect the EM signals, digitize the signals, and transmit the data back to Earth. The satellite must be in orbit around the Earth so the inward force of gravity must be equal to the outward centrifugal force. A variety of possible orbits are discussed in Chapter 3. One can tune the orbit and platform characteristics to vary altitude, speed, revisit time, phase of the orbit

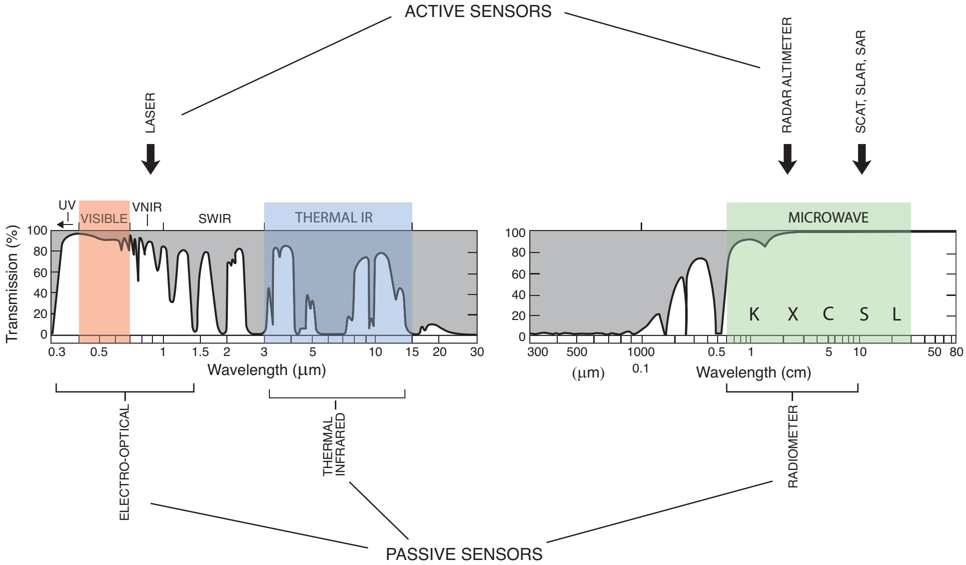

Figure 1.2 Percent transmission of EM waves through the Earth's atmosphere as a function of wavelength. There are three main windows where the atmosphere is mostly transparent. They are the visible (shaded red) centered at 0.5 micrometer, the TIR (shaded blue) centered at 8 micrometers with prominent windows at 3.5 and 10 micrometers, and the microwave (shaded green) at wavelengths between 0.6 cm and 30 cm. Note that EM waves at wavelengths greater than about 30 cm are strongly distorted and reflected by the ionosphere. The microwave region has bands noted by letters. The most common bands for radar interferometry are X, C, S, and L. The K-band is mainly used for radar altimetry and microwave radiometry.

plane with respect to solar illumination or lunar/solar tides, and platform orientation (yaw, pitch, and roll).

Fourth, there must be a computer and software (e.g., GMTSAR) to ingest the downlinked imagery/orbital information and construct maps of surface reflectance and deformation.

1.2 Main Categories of Satellite Remote Sensing

Based on these constraints, there are roughly six possible categories of Earth remote sensing instruments (i.e., three atmospheric windows and two modes – passive or active). These categories include:

The first type of passive sensor is an electro-optical instrument that operates in the visible part of the spectrum and uses cameras to construct imagery. A typical sensor is LANDSAT-7. The satellite was placed into a Sun-synchronous orbit having an inclination of 98.2 degrees and an altitude of 705 km. This seventh

LANDSAT had eight bands comprising six bands in the visible and near infrared with 30 m spatial resolution, a thermal infrared band with 60 m spatial resolution, and a panchromatic band with 15 m spatial resolution. These electro-optical sensors are usually placed into a Sun-synchronous orbit to provide relatively constant Sun illumination for each ascending (or descending) overflight. Very high spatial resolution optical sensors can measure sharp horizontal variations in surface deformation at a precision of about one tenth of a pixel size using cross-correlation methods on repeated images and these data are highly complementary to deformation measurements with InSAR (e.g., Avouac et al. (2006); Milliner et al. (2015)).

The second type of passive sensor is a TIR camera using the atmospheric window between 8 and 15 micrometers. The TIR camera measures thermal emissions from land or ocean surface, which can be used to infer surface temperature (e.g., McClain et al. (1985); Donlon et al. (2012)). These TIR cameras do not need to be in Sun-synchronous orbits and the acquisitions are designed to avoid specular reflections of solar radiation.

The third type of passive sensor is a microwave radiometer. This instrument uses the highly transparent window in the microwave part of the EM spectrum (1 cm–30 cm) to measure thermal emissions from the land (e.g., Reichle et al. (2007)) or ocean (e.g., Chelton and Wentz (2005)). Since they operate at wavelengths much longer than the ~10 micrometer peak in the blackbody radian spectrum from the Earth at 290 K, they must stare at the surface for a relatively long time (~0.1 s) to gather enough independent looks to achieve an accurate measurement of temperature. Moreover, since they operate at this relatively long wavelength λ, a large antenna diameter D is needed to achieve a useful angular resolution θ_a. In Chapter 2, we develop the Fraunhofer diffraction resolution $\sin(\theta_a) = 1.22\,\lambda/D$ to illustrate the pronounced resolution differences between short-wavelength optical systems ($\lambda = \sim 0.5$ micrometers) and longer wavelength microwave systems ($\lambda = \sim 50$ mm).

The fourth sensor is an active laser altimeter that can be used to measure the topography of a wide range of surfaces, including ocean, ice, land, vegetation, and buildings. The atmosphere is sometimes opaque over the visible part of the spectrum because of clouds. The lidar can measure distances very accurately by sending a sharp pulse of light and measuring the two-way travel time of the reflected pulse. Monitoring changes in ocean ice and continent ice thickness are two of the most important applications of this method (Markus et al. (2017)).

The fifth sensor is an active radar altimeter that is best for measuring the topography of the ocean surface and changes in sea surface height due to a wide range of oceanographic processes (Fu and Cazenave (2000)). As we will discuss in

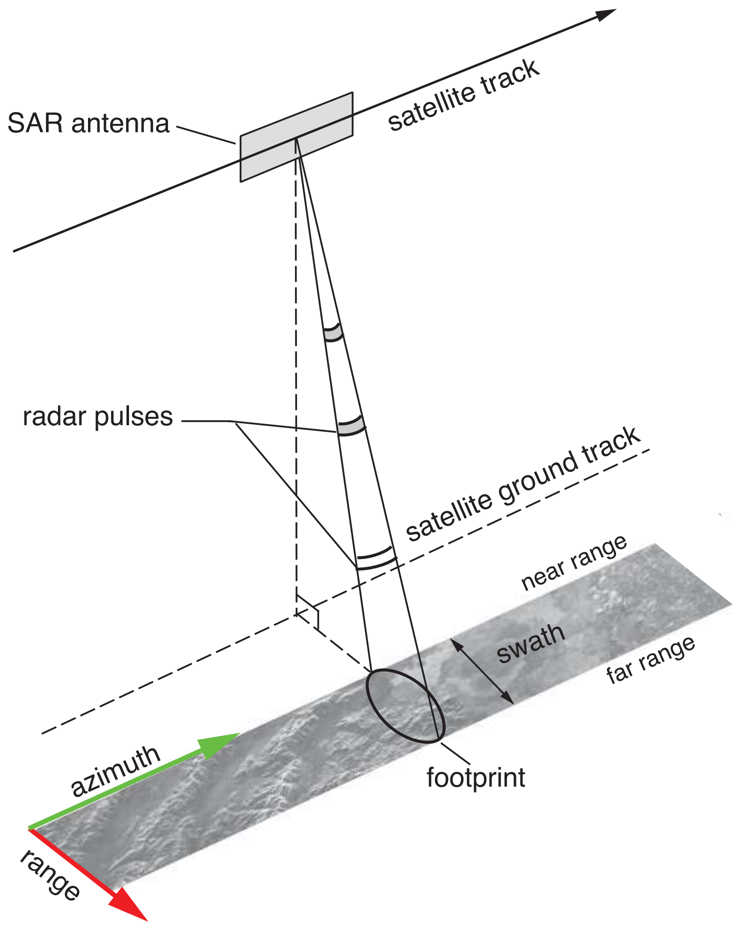

Figure 1.3 Schematic diagram of a SAR satellite in orbit. The SAR antenna has its long axis in the flight direction (also called the *azimuth* direction) and the short axis in the *range* direction. The radar sends pulses to one side of the *ground track* that illuminate the Earth over a large elliptical footprint. The reflected energy returns to the radar where it is recorded as a function of *fast time* in the range direction and *slow time* in the azimuth direction.

Chapter 2, the beam-limited footprint of a radar altimeter is 10s of km in diameter. To reduce the diameter of the footprint, the radar emits a sharp pulse having a spherical wavefront. The radar measures the two-way travel time of the leading edge of the pulse, which corresponds to the closest reflector from the satellite. Over the ocean this is also the *nadir* point as well as the zero-Doppler shift point. In practice, it is difficult to send a very sharp radar pulse so the radar emits a rather long frequency-modulated chirp that reflects from the ocean and returns to the sensor.

A sharp pulse is formed after reception by deconvolving the return chirp with a replica of the outgoing chirp.

The sixth sensor is an active radar that sends pulses of microwave energy to the side of the sub-satellite ground track (e.g., Curlander and McDonough (1991)) as shown in Figure 1.3. Like a radar altimeter, these pulses have broad spherical wavefronts that reflect off a large ellipsoidal *footprint* forming a wide *swath*. The first reflections come from the *near-range* edge of the swath. The pulse sweeps across the swath to the *far-range* edge of the swath.

The data collected by this side-looking radar have very poor spatial resolution related to the large size of the footprint. Chapters 2–4, we will describe how these data are focused first in the range direction and then in the azimuth direction using a synthetic aperture method (SAR). This is a good place to note that the raw SAR data come as a 2-D array of numbers. The rows of the file correspond to the *range* direction marked by the red arrow in Figure 1.3. The columns in the file correspond to the along-track direction of the satellite, which is also called *azimuth* marked by the green arrow in Figure 1.3. The data are equally spaced in time in both dimensions. The time spacing in the range direction is 1/*range sampling rate*. If the range cells are $\Delta\rho = 5$ meters, then the range sampling rate $\Delta t = c/(2 * \Delta\rho) = 30$ MHz, where c is the speed of light. This dimension is also called *fast time* because of the high sampling rate. The sampling rate in the azimuth direction, also called the *pulse repetition frequency* (PRF) is much lower at ~1 500 Hz; this dimension is called *slow time*. In Chapter 2, we will discuss how these sampling rates are related to the resolution of the SAR image as well as the velocity of the spacecraft and the length of the antenna.

2

Principles of Synthetic Aperture Radar

2.1 Introduction

Synthetic aperture radar (SAR) satellites collect swaths of side-looking echoes at a sufficiently high-range resolution and along-track sampling rate to form high-resolution imagery (see Figure 1.3). These notes are derived from excellent books by Curlander and McDonough (1991), Elachi and Van Zyl (2021), and Elachi (1988). As discussed in this chapter, the *range resolution* of the raw radar data is determined by the *pulse length* and the *incidence angle*. For real aperture radar, the along-track or *azimuth* resolution of the outgoing microwave pulse is diffraction limited to an angle corresponding to the wavelength of the radar (e.g., 0.05 m) divided by the length of the aperture (e.g., 10 m). When this beam pattern is projected onto the surface of the Earth at a range of say 850 km, it illuminates 8 000 m in the along-track dimension so the raw radar data are horribly out of focus in azimuth. Using the synthetic aperture method, the image can be focused on a point reflector on the ground by coherently summing thousands of consecutive echoes, thus creating a synthetic aperture perhaps 5 000-m long. Proper focus is achieved by summing the complex numbers along a constant range (Chapter 4). The focused image contains both *amplitude* (backscatter) and *phase* (range) information for each resolution cell.

2.2 Fraunhofer Diffraction

To understand why a synthetic aperture is needed for microwave imaging from orbital altitude, one must understand the concepts of diffraction and resolution. Consider the projection pattern of coherent radiation after it passes through an *aperture*. First we will consider a 1-D aperture and then go on to a 2-D rectangular aperture to simulate a rectangular SAR antenna. The 2-D case provides the shape and dimension of the footprint of the radar. Although we will develop the resolution characteristics of apertures as transmitters of radiation, the resolution

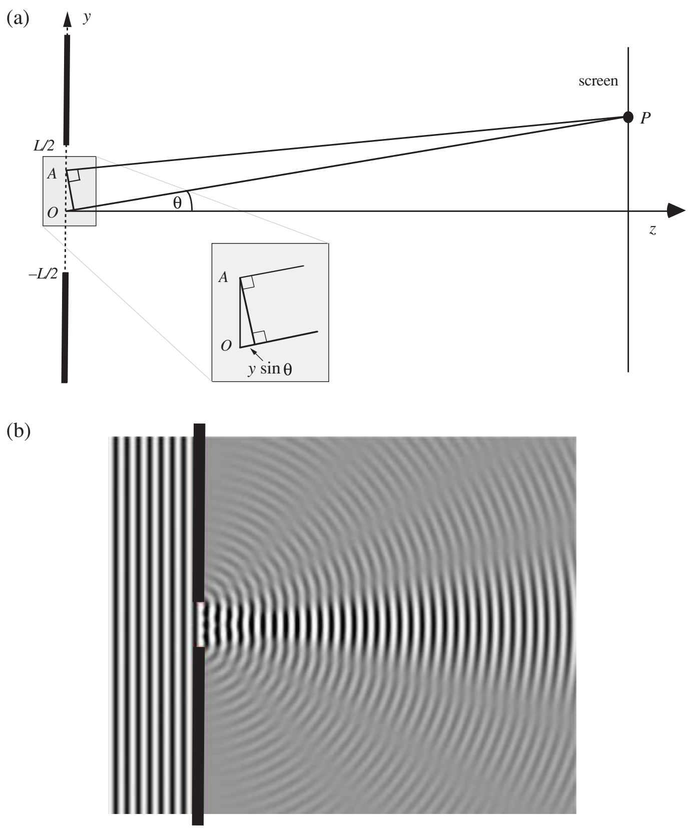

Figure 2.1 (a) Geometric diagram for the projection of coherent waves on a screen that is far from the aperture of length L. (b) Image of coherent waves passing through a slit (aperture) in a wall.

characteristics are exactly the same when the aperture is used to receive radiation. These notes were developed from Rees (2001) and Bracewell (1978).

We simulate *coherent radiation* by point sources of radiation distributed along the aperture between $-L/2$ and $L/2$ (Figure 2.1). For simplicity we will assume that all the sources have the same amplitude a, wavelength λ, and phase. Given these sources of radiation, we solve for the illumination pattern on the screen as a function of θ. We will assume that the screen is far enough from the aperture so that the rays AP and OP are parallel. Later we will determine how far away the screen needs to be in order for this approximation to hold. Under these conditions, the ray AP is slightly shorter than the ray OP by an amount $-y \sin\theta$. This corresponds to a phase shift of $\frac{-2\pi}{\lambda} y \sin\theta$. The amplitude of the illumination at point P is the integral over all of the sources along the aperture multiplied by their complex phase value

$$P(\theta) = \int_{-L/2}^{L/2} a(y) e^{-i2\pi yk \sin\theta} \, \mathrm{d}y, \tag{2.1}$$

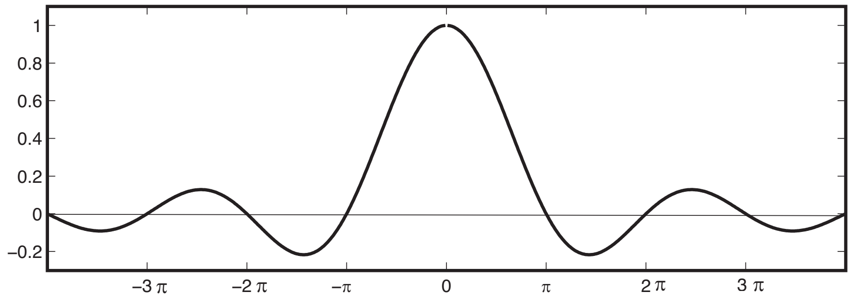

Figure 2.2 Sinc function illumination pattern for the aperture shown in Figure 2.1.

where the wavenumber $k = 1/\lambda$. This is called the Fraunhofer diffraction integral. The illumination across the aperture is uniform in both amplitude and phase, so we set $a(y) = 1$. Now, let $s = 2\pi k \sin\theta$, so the integral is easy to evaluate:

$$P(s) = \int_{-L/2}^{L/2} e^{-isy}\, \mathrm{d}y = \frac{e^{-isL/2} - e^{isL/2}}{-is} = \frac{2}{s} \sin\left(sL/2\right) = L \operatorname{sinc}\left(sL/2\right). \tag{2.2}$$

Replacing s with $2\pi \sin\theta/\lambda$, we arrive at the final result:

$$P\left(\theta\right) = L \operatorname{sinc}\left(\frac{L\pi \sin\theta}{\lambda}\right). \tag{2.3}$$

The illumination pattern on the screen is shown in Figure 2.2.

The first zero crossing, or *angular resolution* θ_a, of the sinc function occurs when the argument is π, so

$$\sin\theta_a = \frac{\lambda}{L}. \tag{2.4}$$

You should remember this formula. Note for small angles $\theta_a \cong \frac{\lambda}{L}$ and $\tan\theta_a \cong \sin\theta_a$. One could modify the screen illumination by changing the strength of the illumination across the aperture. For example, a theoretical Gaussian aperture would produce a Gaussian illumination function on the screen. This would eliminate the *sidelobes* associated with the sinc function, but it would also broaden the projection pattern. In addition, one could vary the phase along the aperture to shift the point of maximum illumination away from $\theta = 0$. Such a phased array aperture is used in some radar systems to continuously illuminate a feature as the satellite passed over it. This is called *spotlight mode* SAR, and it is a favorite technique for military reconnaissance.

One can perform the same type of analysis with any 2-D aperture; many analytic examples are given in Figure 12.4 from Bracewell (1978). For a circular aperture, the angular resolution is given by

$$\sin\theta_a = 1.22\frac{\lambda}{L}. \tag{2.5}$$