M

MADDEN-JULIAN OSCILLATION (MJO)

Baijun Tian and Duane Waliser
Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

Synonyms

30–60 day oscillation; 40–50 day oscillation; Intraseasonal oscillation (ISO); Intraseasonal variability (ISV)

Definition

The MJO is a planetary-scale quasiperiodic oscillation of atmospheric wind and convective cloudiness anomalies that moves slowly eastward along the equator mainly over the tropical Indian and Pacific Oceans with a timescale on the order of 30–60 days.

Introduction

In 1971, Roland Madden and Paul Julian stumbled across a 40–50 day oscillation when analyzing the zonal (east–west) wind data from rawinsondes at Kanton Island (3 °S, 172 °W) over the equatorial western Pacific. Until the early 1980s, little attention was paid to this oscillation, which later became known as the MJO. Since the 1982–1983 El Niño event, low-frequency variations in the tropics, both on intra-annual (less than a year) and interannual (more than a year) timescales, have received much more attention, and the number of MJO-related publications grew rapidly. The MJO turned out to be the dominant form of the intraseasonal (30–90 day) variability in the tropical atmosphere and has many important influences on the global weather and climate system.

The MJO is a naturally occurring mode of variability of the tropical ocean–atmosphere system. It is characterized by an eastward propagation of large regions of both enhanced and suppressed tropical convection, cloudiness and rainfall near the equator mainly over the tropical Indian and Pacific Oceans, and associated large-scale atmospheric circulation (wind) anomalies over the whole globe. The anomalous cloudiness or rainfall usually first emerges over the equatorial western Indian Ocean and intensifies and remains evident as it propagates eastward over the warm ocean waters of the equatorial eastern Indian Ocean and western Pacific, the so-called Indo-Pacific warm pool. This pattern of anomalous cloudiness and rainfall then generally weakens and disappears as it moves over the cooler ocean waters of the equatorial eastern Pacific, the so-called equatorial cold tongue. Along with this eastward-propagating pattern of equatorial cloudiness and rainfall anomalies, there also exist eastward moving distinct baroclinic patterns of lower- and upper-level atmospheric circulation anomalies in the tropics and subtropics. The circulation anomalies extend around the globe and are not confined to the eastern hemisphere as opposed to the cloudiness and rainfall anomalies. When the MJO moves eastward, it modulates the background cloud, rainfall, and circulation in the tropical Indian and Pacific Oceans on timescales shorter than a season but longer than a couple of weeks. The length of a typical MJO cycle is approximately 30–60 days but normally 40–50 days. Thus, the MJO is also known as the 30–60 day oscillation, 40–50 day oscillation, intraseasonal oscillation (ISO), or intraseasonal variability (ISV) after its typical timescale. A complete MJO cycle can be divided into two distinct phases according to the intensity of its convective activity and rainfall: convectively active (enhanced) phase or wet phase and convectively inactive (suppressed) phase or dry phase. The wet phase of the MJO cycle is characterized by enhanced tropical convection and large moist convective storms with higher cloud-top heights, more cloud cover, and heavier rainfall (thus more atmospheric latent heating)

E.G. Njoku (ed.), *Encyclopedia of Remote Sensing*, DOI 10.1007/978-0-387-36699-9,

than average. In contrast, the dry phase of the MJO cycle is typified by dry and clear conditions with lower cloud-top heights, less cloud cover and rainfall (thus less atmospheric latent heating) than normal. The MJO appears to be predictable with lead times of 2–3 weeks. This can help bridge the gap in environmental forecast skill between that of weather (lead times up to a few days) and that of seasonal-to-interannual climate predictions (lead times from a few months to a few years).

Satellite remote sensing data have played an important role in the MJO studies during the last three decades because of the high spatial (a few kilometers) and temporal (3 h or daily) resolutions and global coverage of the satellite data especially over the tropical oceans where the rawinsondes are sparse and the global reanalyses have had large uncertainties. These satellite-based studies have significantly advanced our knowledge in the MJO description and mechanism as well as its global impacts. They have also led to considerable improvement in our numerical modeling capability and theoretical understanding of the MJO. This entry briefly reviews the central role of satellite remote sensing data in studying the description, mechanisms, and global impacts of the MJO.

Description

The MJO and its eastward-propagating convective feature are most active during the Northern Hemisphere (boreal) winter season (November–April) when the Indo-Pacific warm pool is centered near the equator. During the Northern Hemisphere (boreal) summer season (May–October), the change in the large-scale wind patterns associated with the Asian summer monsoon results in the large-scale convective disturbances propagating northeastward, from the equatorial Indian Ocean into Southeast Asia. The discussions in this entry mainly focus on the boreal winter MJO events although many aspects of these discussions can be equally applied to the boreal summer MJO events.

The MJO can be detectable in several important atmospheric and oceanic parameters, such as atmospheric cloudiness, atmospheric wind speed and direction, atmospheric temperature, atmospheric moisture, surface pressure, surface rainfall, sea surface temperature (SST), and surface heat and freshwater fluxes. However, the fundamental quantities related to the MJO are large-scale organized convection, cloudiness, rainfall, and tropospheric winds. Thus, our discussion of the MJO description mainly centers on how the remote sensing data help us understand the characteristics of large-scale convective cloudiness, rainfall, and tropospheric winds associated with the MJO.

Outgoing longwave radiation (OLR) and infrared window radiance (or brightness temperature) provide broadband and narrowband measures of the total flux of longwave radiation lost to space at the top of the atmosphere. Deep convective clouds in the tropics have cold cloud tops and therefore have low values of OLR and brightness temperature. Typically, an OLR value of less than 200 W m^{-2} or a brightness temperature value of less than 220 K indicates the presence of deep convection in the tropics. Because of this simple property of OLR and brightness temperature, they have been widely used as a proxy for deep convection over the tropics, where the background longwave radiation from low clouds or surface is much higher. During the wet phase of the MJO, both OLR and window radiance can be significantly reduced due to higher and colder convective cloud tops. On the other hand, during the dry phase of the MJO, both OLR and window radiance can be significantly enhanced because of clear skies and lower and warmer cloud tops. Thus, OLR and infrared window radiance are good indicators of the convective intensity of the MJO. Over the tropical warm ocean waters where MJO convection is active, satellites are the only way to observe convective cloudiness with large spatial coverage. As a result, the satellite remote sensing data have played a central role in studying the general spatial and temporal structure and eastward-propagating features of the MJO convection and cloudiness during the past three decades.

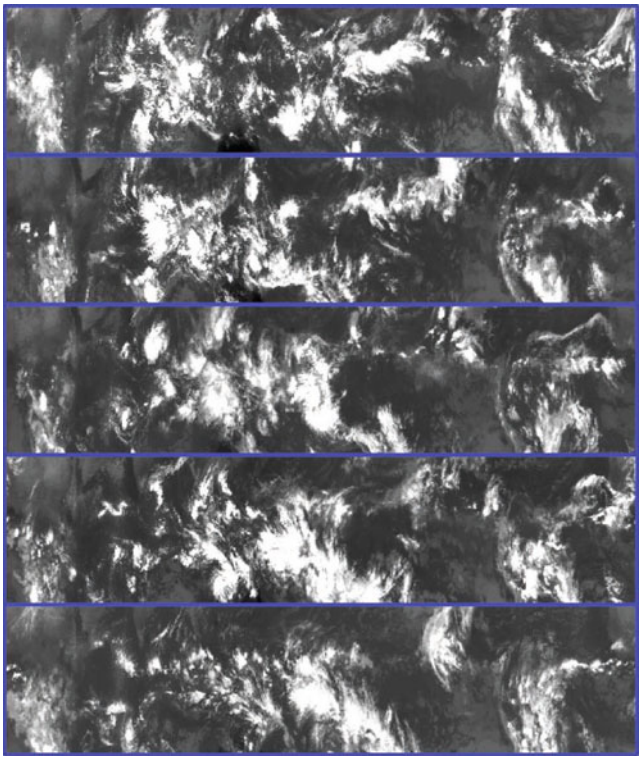
Based on limited rawinsonde and surface station data, Madden and Julian speculated that the MJO is characterized by slowly eastward-propagating, large-scale oscillations in the tropical convective cloudiness over the equatorial Indian Ocean and western Pacific as the result of an eastward movement of large-scale atmospheric circulation cells oriented in the equatorial zonal plane. Evidence of such eastward-propagating clouds in satellite data was first presented by Arnold Gruber in 1974 and Abraham Zangvil in 1975 who both found large-scale eastward-propagating features near 40–50 days at the equator in the cloud brightness data obtained from the Environmental Science Services Administration (ESSA) satellites (ESSA 3 and 5). However, no further evidence was found until the early 1980s when NOAA OLR data and wind analyses from US National Meteorological Center (NMC) became available. These Advanced Very High Resolution Radiometer (AVHRR) OLR data started from the mid-1970s and were mainly from a series of polar-orbiting satellites, such as the scanning radiometer (SR) series and the Television Infrared Observation Satellite–Next Generation (TIROS-N) series. These OLR data have a twice-daily resolution and a good global coverage each day. By the early 1980s, almost 10 years of daily AVHRR OLR data were archived and available to the research community. In the early to mid-1980s, a series of observational papers on the MJO (at the time still referred to as the 40–50 day or 30–60 day oscillation) using the AVHRR OLR and NMC wind analyses appeared. These studies clearly demonstrated the existence of the slowly eastward propagation of the tropical cloudiness at the intraseasonal timescale and documented many detailed and important convective cloudiness and circulation features of the MJO. These papers also helped to bring the MJO to the attention of the scientific community. The NOAA polar orbiter satellites have been operating almost continuously over the

past 30 years. As a result, the NOAA OLR data have a relatively long record (over 30 years) and have been and are still extensively employed for studying the MJO.

Figure 1 shows the infrared brightness for a MJO event from December 7 to 26, 1987, with each panel separated by 5 days. Each map covers the all longitudes (0–360°) between 20 °S and 20 °N. Bright white areas indicate cold high clouds, and dark regions indicate cloud-free or warm low cloud conditions. The slow eastward propagation of cold high clouds associated with the MJO is evident. The cold high clouds first form over the western equatorial Indian Ocean on December 7, 1987 and, over the course of the following ~20 days, then intensify and propagate eastward across the equatorial Indian Ocean and the Maritime Continent to the equatorial western Pacific Ocean.

In addition to the large-scale eastward-propagating pattern of MJO convective cloudiness, there exist many fine-scale structures within the convective cloudiness. The high spatial (100 km) and temporal (daily) resolution AVHRR OLR data from NOAA polar orbiters and much higher spatial (a few kilometers) and temporal (3 h) resolution window-channel infrared data from the geostationary satellites, such as Geostationary Meteorological Satellite (GMS) from Japan, are particularly useful in investigating the fine structure of the MJO convective cloudiness due to their high spatial and temporal resolutions. For example, the OLR data indicated many short-period, synoptic-scale convective systems within the planetary-scale 30–60 day fluctuations. Along the equator, these active convective systems move eastward with a phase speed of 10–15 m s^{-1} and have a horizontal spatial scale of several thousand kilometers and a timescale of less than 10 days. These synoptic-scale, eastward-propagating convective systems within the MJO envelope are referred to as super cloud clusters or more recently convectively coupled Kelvin waves. The GMS infrared window radiance data have revealed that a super cloud cluster consists of many fine-scale cloud clusters. These fine-scale cloud clusters typically propagate westward along the equator with a lifetime of about 1–2 days. Although each cloud cluster moves westward, a super cloud cluster moves eastward due to the successive formation of a new cloud cluster east of the mature-stage cloud cluster. This suggested the existence of a hierarchy of convective activity within the MJO that is still an outstanding avenue of research of today.

In the Tropics, surface rainfall is closely related to convective cloudiness and thus is another key quantity of interest to characterize the MJO. The tropical rainfall can be estimated from the satellite-observed infrared and microwave radiances. The surface rainfall can first be indirectly derived from infrared window radiance and OLR which are very sensitive to cloud-top temperatures that are indirectly tied to surface rainfall. The microwave radiances are very sensitive to the hydrometeors that directly result in surface precipitation and thus can be used more directly to retrieve surface precipitation. The microwave-based rainfall retrievals can be divided into passive and active microwave (radar) retrievals. The passive retrievals can further be divided into the microwave emission-based (sensitive to cloud liquid water) and the microwave scattering-based (sensitive to ice particles and large water drops) rainfall retrievals. Starting from the 1990s, several global rainfall data have been generated from satellite-observed infrared and microwave data and were instrumental in studying the MJO during the last two decades. For example, daily, global oceanic rainfall data retrieved based on microwave emission from the Microwave Sounding Unit (MSU) on the NOAA TIROS-N satellites was first used to study the MJO convective feature in the 1990s. During the late 1990s, the NOAA Climate Prediction Center (CPC) generated a global rainfall data set, referred to CPC Merged Analysis of Precipitation (CMAP), through the merged analysis of precipitation from several sources, such as gauges, satellites, and numerical model outputs. The satellite rainfall estimates for the CMAP includes the window infrared-based rainfall estimate from NOAA geostationary satellites (e.g., Geostationary Operational Environmental Satellites, GOES), the OLR-based rainfall estimate from the NOAA polar-orbiting satellites, the microwave emission-based rainfall estimate from the MSU on the NOAA polar-orbiting



Madden-Julian Oscillation (MJO), Figure 1 Infrared satellite observations for the global tropics (20 °N–20 °S) for (*from the top*) 03 GMT December 7, 12, 17, 22, and 26, 1987. *Bright white areas* indicate high clouds and deep convection, and *dark regions* indicate cloud-free conditions (Based on Global Cloud Imagery (GCI) data courtesy of M. Salby, University of Colorado).

satellites and the Special Sensor Microwave/Imager (SSM/I) on the Defense Meteorological Satellite Program (DMSP) satellites, and the microwave scattering-based rainfall estimate from SSM/I. Similar merged global rainfall data were also produced by the Global Precipitation Climatology Project (GPCP) based on similar inputs. However, some subtle differences exist between the CMAP and GPCP precipitation data sets, such as diurnal cycle adjustment and atoll precipitation adjustments. Both CMAP and GPCP rainfall data with pentad (5 day) resolution have been extensively used to study the convective features of the MJO and identify MJO events. Launched in 1997, the Tropical Rainfall Measurement Mission (TRMM) satellite provided the first spaceborne precipitation radar (PR) (active microwave) to monitor global rainfall from space in addition to the passive TRMM Microwave Imager (TMI) instrument and the visible and infrared scanner (VIRS) instrument. The TRMM PR and TMI data have also been used to study the MJO. However, their spatial and temporal sampling is rather coarse. To alleviate these sampling deficiencies of the TRMM PR and TMI, the TRMM Multisatellite Precipitation Analysis (TMPA) project provides a calibration-based sequential scheme for combining precipitation estimates from multiple satellites, as well as gauge analyses where feasible, at fine scales ($0.25^\circ \times 0.25^\circ$ and 3 hourly). The input satellite data for the TMPA are mainly from two sources: (1) passive microwave-based precipitation from the TMI on TRMM, the SSM/I on DMSP satellites, the Advanced Microwave Scanning Radiometer–Earth Observing System (AMSR-E) on Aqua, and the Advanced Microwave Sounding Unit-B (AMSU-B) on the NOAA polar-orbiting satellites; (2) window infrared-based precipitation data collected by the international constellation of geostationary satellites. This TMPA data set, also known as the TRMM 3B42, has relatively better retrieval accuracy and sampling at fine spatial and temporal scales. It has been used extensively for the recent MJO studies.

In association with the eastward-propagating equatorial convective cloud and rainfall system are strong variations in lower- and upper-level large-scale atmospheric wind fields along the equator and in the subtropics. Unlike the convective cloudiness that is mostly confined over the equatorial Indian and western Pacific Oceans, the large-scale wind anomalies of the MJO extend globally along the equator and into the subtropics. For example, along the equator, low-level zonal winds converge into the convective center, while upper-level zonal winds diverge away from the convective center. These lower- and upper-level zonal winds are interconnected through ascending (upward vertical movement) moist air within the convective center and descending dry air outside the convective center. These large-scale zonal winds propagate eastward together with the convective cloudiness along the equator and can reach into the western hemisphere (eastern Pacific, Atlantic, and Africa). In addition to these zonal winds along the equator are large-scale gyre circulations extending into the subtropics in both the lower and upper troposphere that are tied to the eastward-propagating convective cloudiness and zonal winds along the equator. For example, in the lower troposphere, a subtropical cyclonic couplet (counterclockwise in the Northern Hemisphere and clockwise in the Southern Hemisphere) flanks or lies to the west of the MJO convective region, while a subtropical anticyclonic couplet (clockwise in the Northern Hemisphere and counterclockwise in the Southern Hemisphere) lies to the east of the MJO convective region. On the other hand, in the upper troposphere, a subtropical anticyclonic couplet flanks or lies to the west of the MJO convective region, while a subtropical cyclonic couplet lies to the east of the MJO convective region due to the baroclinic nature of the tropical large-scale wind fields. The near equatorial large-scale zonal wind anomalies are a Kelvin wave response to the MJO convective heating, while the off equatorial meridional wind anomalies are a Rossby wave response to the MJO convective heating. The Kelvin wave is named after Lord Kelvin, who studied water waves along a vertical side boundary under rotation conditions. In this case, the equator, where the vertical component of the Earth's rotation vector changes sign, serves the vertical side boundary. The Rossby wave is due to the latitudinal variation of the vertical component of the Earth's rotation and is named after C. G. Rossby, who was the first to clearly isolate the so-called Rossby wave dynamics (a balance between inertia and rotation). These large-scale patterns of convective cloudiness and wind fields are components of what are collectively referred to as equatorial waves or convectively coupled equatorial waves. It is the upper-level circulation features of these waves that allow the convective signatures of the MJO over the Indo-Pacific warm pool to influence weather "downstream" over the eastern Pacific and Atlantic (e.g., hurricanes and tropical cyclones) as well as the midlatitudes (e.g., precipitation extremes along the US west coasts).

In terms of satellite observations, tropospheric winds are difficult to observe directly except via cloud tracking, such as the cloud-drift winds derived from the Multi-angle Imaging SpectroRadiometer (MISR) and NOAA geostationary satellites. However, the surface winds can be measured directly using the spaceborne radar scatterometers, such as the SeaWinds instrument on NASA's Quick Scatterometer (QuikSCAT) satellite. The importance of the cloud-drift winds and QuikSCAT surface winds for the MJO study has been recognized but still in the early stages of exploration. The overall large-scale dynamic structure of the MJO, especially in the upper levels, is still mainly derived from the global reanalyses or radiosondes at the moment.

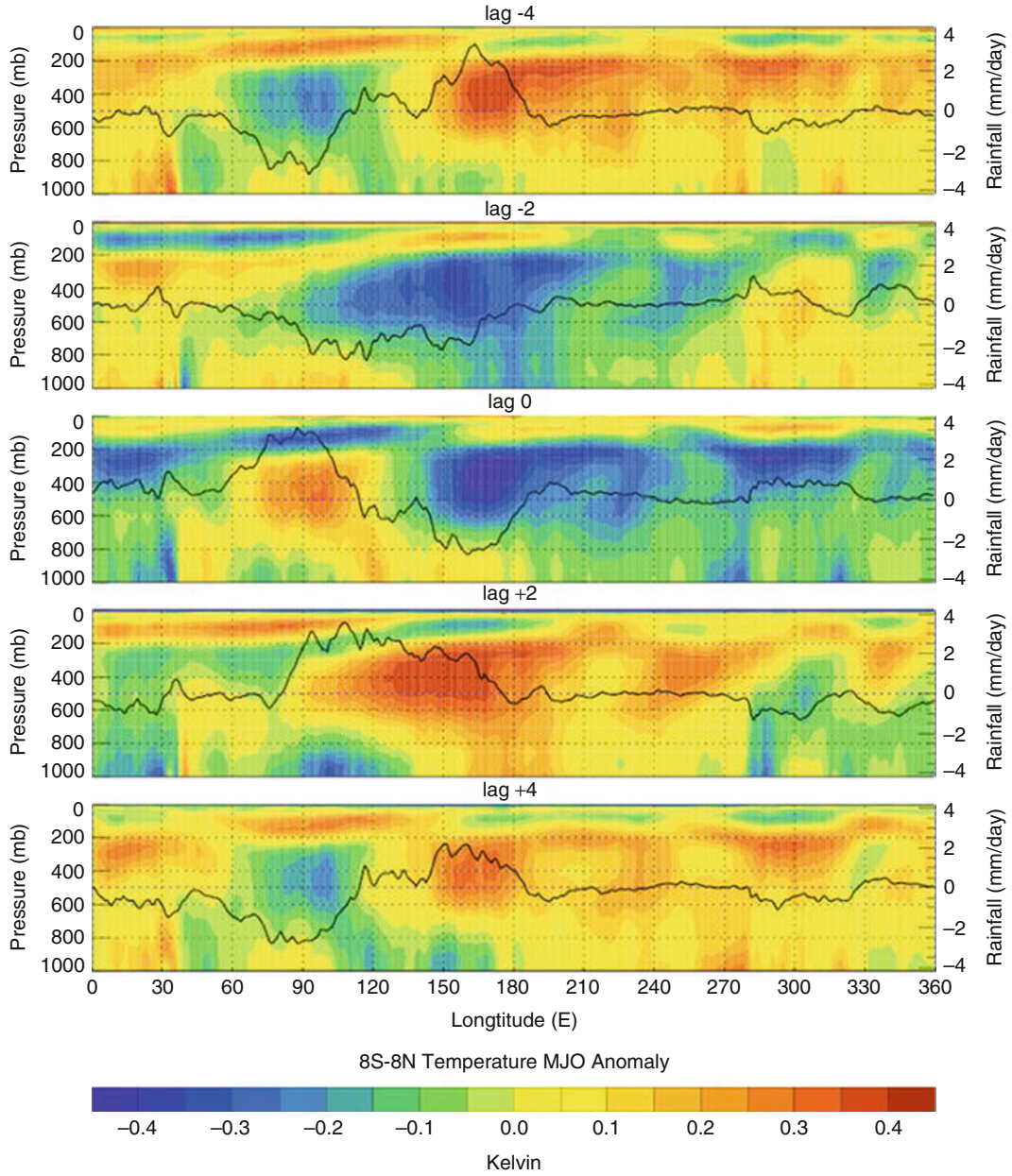
Mechanisms

To understand the mechanisms responsible for the initiation and maintenance of the MJO, it is important to quantify the evolution of the thermodynamic environment and surface conditions associated with the MJO. In particular,

documenting the three-dimensional temperature and moisture structure of the MJO is crucial in advancing our theoretical understanding of the MJO. The availability of satellite-based temperature and moisture soundings makes these studies possible. For example, tropical mean tropospheric temperature, lower-troposphere (surface–300 mb) temperature, and upper-troposphere (500–100 mb) temperature were derived from the MSU channels in the early 1990s. These data were used to study the relationship between MJO convection and temperature anomalies in the 1990s. It was found that when the MJO is amplifying (e.g., over eastern Indian Ocean), convective heating anomalies are positively correlated to temperature anomalies. This implies production of eddy available potential energy (EAPE), which can in turn be used to drive atmospheric motion and sustain the MJO. When the MJO is decaying (e.g., over the eastern Pacific or east of the Date Line), temperature anomalies are nearly in quadrature with convective heating anomalies. As a result, their correlation and production of EAPE are small which is no longer an energy source for the MJO. For water vapor, the TIROS Operational Vertical Sounder (TOVS) provided the water vapor fields at five different levels in the troposphere in the 1990s. These data were used to study the three-dimensional structure and evolution of water vapor over the life cycle of the MJO in the early 2000s. The composite evolution of moisture shows markedly different vertical structures as a function of longitude. There is a clear westward tilt with the height of the moisture maximum associated with the MJO propagating eastward across the Indian Ocean. These disturbances evolve into nearly vertically uniform moist anomalies as they reach the western Pacific. Near-surface (below 850 mb) positive water vapor anomalies were observed to lead the convection anomaly by 5 days over the Indian Ocean and western Pacific. Upper-level positive water vapor anomalies were observed to lag the peak in the convection anomaly by 5–10 days, as the upper troposphere is moistened following intense convection. In the eastern Pacific, the moisture variations then become confined to the lower levels (below 700 mb), with upper-level water vapor nearly out of phase.

While the MSU and TOVS provided useful initial insights into the three-dimensional temperature and moisture structure of the MJO, their vertical resolution was too low to describe the detailed vertical structure, especially near the tropopause and boundary layer. Recently, global atmospheric moisture and temperature profiles with a much higher vertical resolution were produced by the Atmospheric Infrared Sounder (AIRS)/Advanced Microwave Sounding Unit on the NASA Aqua satellite. The AIRS data provide an unprecedented opportunity to document the vertical moist thermodynamic structure and spatial–temporal evolution of the MJO. Figure 2 presents the pressure-longitude cross sections of the temperature anomaly and its relationship to the convective rainfall anomaly for a composite MJO cycle (from −20 day to +20 day, separated by 10 days each). In the Indo-Pacific warm pool, the temperature anomaly exhibits a trimodal vertical structure: a warm anomaly in the free troposphere (800–250 hPa) and a cold anomaly near the tropopause (above 250 hPa) and in the lower troposphere (below 800 hPa) for the wet phase and vice versa for the dry phase. The moisture anomaly also shows markedly different vertical structures as a function of longitude and the strength of the convection anomaly. Most significantly, the AIRS data demonstrate that, over the Indian Ocean and western Pacific, enhanced convection and precipitation is generally preceded in both time and space by a low-level warm and moist anomaly and followed by a low-level cold and dry anomaly. This zonal asymmetry in the low-level moisture and temperature anomaly provides a favorable moist thermodynamic condition for the eastward propagation of the MJO. Furthermore, the comparison between the AIRS observations and the National Center for Environmental Prediction (NCEP)/National Center for Atmospheric Research (NCAR) and NCEP/Department of Energy (DOE) reanalyses revealed the poor representation of the low-level moisture and temperature structure associated with the MJO in these reanalyses particularly over the equatorial Indian and Pacific Oceans, where there are very few conventional data to constrain the reanalyses. Despite their reliance on (imperfect) numerical models, these reanalyses have been widely used as "observations" to validate MJO theories and model simulations.

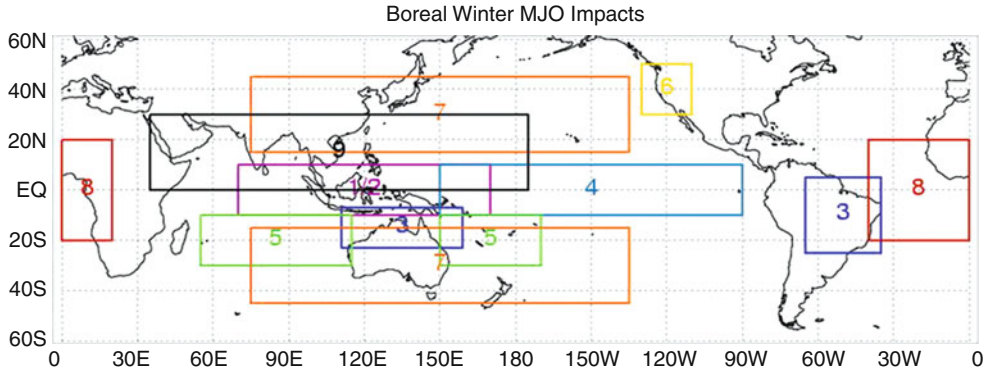
The ocean surface fluxes of heat (solar and infrared radiation, latent and sensible heat), mass (rainfall and water vapor flux), and momentum (wind) are considered important for the initiation and eastward propagation of the MJO by some researchers. Understanding the manner and degree the atmospheric components of the MJO are coupled to ocean surface fluxes is vital to understand the MJO dynamics and establish MJO theories. This coupling includes how the atmospheric components of the MJO influence the ocean surface heat, mass, and momentum fluxes and how the ocean surface provides the needed heat and moisture sources for the MJO convection. Surface heat fluxes have typically been very difficult to measure remotely from space although satellites offer the only viable way to estimate these quantities with regular temporal and spatial samplings over the tropical oceans. Apart from precipitation discussed above, satellite measurements of clouds and water vapor have been used with atmospheric radiative transfer models to provide estimates of solar and infrared radiation fluxes at the ocean surface. Moreover, satellite estimates of ocean surface winds (e.g., QuikSCAT and SSM/I discussed above) have been used in conjunction with water vapor measurements from satellites to construct estimates of latent heat (or evaporative) flux from the ocean. These types of observations, in conjunction with satellite SST retrievals (e.g., Advanced Very High Resolution Radiometer (AVHRR) or TMI), have been used to study how the MJO convection interacts with the ocean surface and explore the degree the ocean and atmosphere are coupled at intraseasonal and other



Madden-Julian Oscillation (MJO), Figure 2 Pressure-longitude cross sections of equatorial mean (averaged from 8 °S to 8 °N) temperature anomaly for a composite MJO cycle based on AIRS data from 2002 to 2005. The color *red* denotes warm anomalies, while the color *blue* indicates cold anomalies. The superimposed *solid black line* denotes rainfall anomaly from TRMM (Reproduced from Tian et al., 2006, their Figure 3, by permission of American Meteorological Society).

timescales. Satellite-based ocean surface wind speed observations (e.g., QuikSCAT and SSM/I discussed above) have been used to derive momentum fluxes (e.g., wind stress) at the ocean surface. These observations have been critical in documenting and understanding the role of the MJO in influencing the development and evolution of El Niño and Southern Oscillation (ENSO) events. ENSO is the most important interannual variability in the coupled tropical atmosphere ocean system with a dominant timescale of 2–7 years and has significant impacts on the global weather and climate.

During the dry phase of the MJO, suppressed convection is associated with decreased cloud cover and increased surface insolation and anomalous surface easterlies. These anomalous surface easterlies act to decrease the surface wind speed because the background surface winds are weak westerlies in the equatorial Indian and western Pacific Oceans, hence decreasing the surface latent heat flux



Madden-Julian Oscillation (MJO), Figure 3 Geographical regions of a number of impacts of the boreal winter MJO on the global climate system. (*1*) Influencing tropical weather, alternative periods of wetter/drier conditions in the tropical Indian Ocean and western Pacific. (*2*) Modulating the diurnal cycle of tropical deep convection and rainfall in the tropical Indian Ocean and western Pacific. (*3*) Modulating the onsets and breaks of the Australian and South American monsoon systems. (*4*) Influencing ENSO cycle over the equatorial central and eastern Pacific. (*5*) Impacting tropical cyclone genesis over the tropical Indian Ocean and western Pacific. (*6*) Influencing the development of heavy rainfall events over US west coast. (*7*) Changing the subtropical total-column ozone over the eastern Hemisphere and Pacific Ocean. (*8*) Affecting the aerosol and air pollution over the equatorial Indian and western Pacific Oceans as well as the tropical Africa and Atlantic Ocean. (*9*) Influencing the ocean surface Chl across the tropical ocean coasts.

(or evaporation). Increased surface shortwave radiation and reduced surface evaporation contribute to the warming of SST for the dry phase. During the subsequent wet phase of the MJO, enhanced convection is associated with increased cloud cover and decreased surface insolation. As a result, the SST warming trend is arrested and a cooling trend is initiated. Subsequently, the continued cooling of the upper ocean is accelerated by increased westerly surface winds leading to enhanced surface evaporation and increased entrainment of cold water from below the thermocline. Then the wet phase is followed by another dry phase when SST warming occurs. Therefore, over the Indian Ocean and western Pacific, the enhanced convection is usually led by a warm SST anomaly to the east due to enhanced insolation and decreased evaporation and followed by a cold SST anomaly to the west due to decreased insolation and enhanced evaporation. When the convective anomaly approaches the Date Line, the surface evaporation anomaly and surface solar radiation anomaly tend to cancel each other. Thus, the SST anomaly is rather small over the eastern Pacific, so does the convective anomaly. This convection–SST phase relationship leads many scientists to believe that the MJO is a coupled mode of the tropical ocean–atmosphere system.

Impacts

During the past three decades, the MJO has been shown to have important influences on various weather and climate phenomena over the globe at many timescales, such as the diurnal cycle, tropical weather, monsoon onsets and breaks, ENSO, tropical hurricanes and cyclones, extreme precipitation events, extratropical and high-latitude circulation, and weather patterns. Given evidence that the MJO is predictable with lead times of 2–3 weeks, the strong modulation of the global climate system by the MJO implies that many other components of the global climate system may be predictable with similar lead times. Some examples of the MJO impacts are listed and described below, and the regions of impacts are illustrated in Figure 3.

First, the MJO significantly impacts the tropical synoptic weather, such as alternative periods of wet and dry conditions, especially over the tropical Indian Ocean and west Pacific, through its influences of tropical rainfall and cloudiness (Figure 3). During the wet phase of the MJO, the tropical atmosphere is very moist and cloudy with heavy rainfall. In contrast, during the dry phase of the MJO, the tropical atmosphere experiences dry and clear conditions with plenty of sunshine.

Second, the MJO can influence the diurnal cycle of tropical deep convection and rainfall through its effect on the background state over the equatorial Indian and western Pacific oceans (Figure 3). The diurnal cycle is enhanced over both land and water during the convectively active phase of the MJO, while it is reduced during the convectively suppressed phase of the MJO. However, the diurnal phase is not significantly affected by the MJO.

Third, the MJO can substantially modulate the intensity of monsoon systems around the globe, such as the Australian and South American monsoons for boreal winter and the Asian and North American monsoons for boreal summer (Figure 3). The wet phase of the MJO can affect both the onset timing and intensity of the monsoon, while the dry phase of the MJO can prematurely end a monsoon and also initiate breaks during already existing monsoons.

Fourth, there is evidence that the MJO influences the ENSO cycle (Figure 3). It was argued that the westerly wind bursts associated with the MJO over the equatorial western Pacific are an important trigger for an El Niño event. The MJO may not cause an El Niño event, but

can contribute to the speed of development, and perhaps the overall intensity of an ENSO cycle.

Fifth, the MJO is known to modulate tropical cyclone activity in the Indian Ocean, Pacific Ocean, Gulf of Mexico, and Atlantic Ocean by providing a large-scale environment that is favorable (unfavorable) for storm development (Figure 3). For example, westerly wind anomalies at the surface in and just behind the area of enhanced convection of the MJO may generate cyclonic (anticyclonic) rotation north (south) of the equator, respectively. At the same time, in the upper levels, anticyclonic (cyclonic) rotation develops along and just behind the area of enhanced convection resulting in a means to reduce vertical wind shear and increase upper-level divergence – both of which are favorable for tropical cyclone development and intensification. The strongest tropical cyclones tend to develop during the wet phase of the MJO. As the MJO progresses eastward, the favored region for tropical cyclone activity also shifts eastward from the Indian Ocean to the Pacific Ocean and eventually to the Atlantic Ocean.

Sixth, boreal winter extreme precipitation events along the US west coast are often connected with the pattern of tropical rainfall and circulation anomalies associated with the MJO (Figure 3). When the heavy tropical rainfall associated with the MJO is concentrated at the Maritime Continent, a strong blocking anticyclone is located in the Gulf of Alaska with a strong polar jet stream around its northern flank. During this time, the US west coast typically experiences a dry spell. When the enhanced tropical rainfall associated with the MJO shifts to the central Pacific and weakens, a deep low pressure system typically forms near the Pacific Northwest coast and can bring up to several days of heavy rain and possible flooding to the Pacific Northwest coast. These events are often referred to as "Pineapple Express" events, so named because a significant amount of deep tropical moisture traverses the Hawaiian Islands on its way toward western North America.

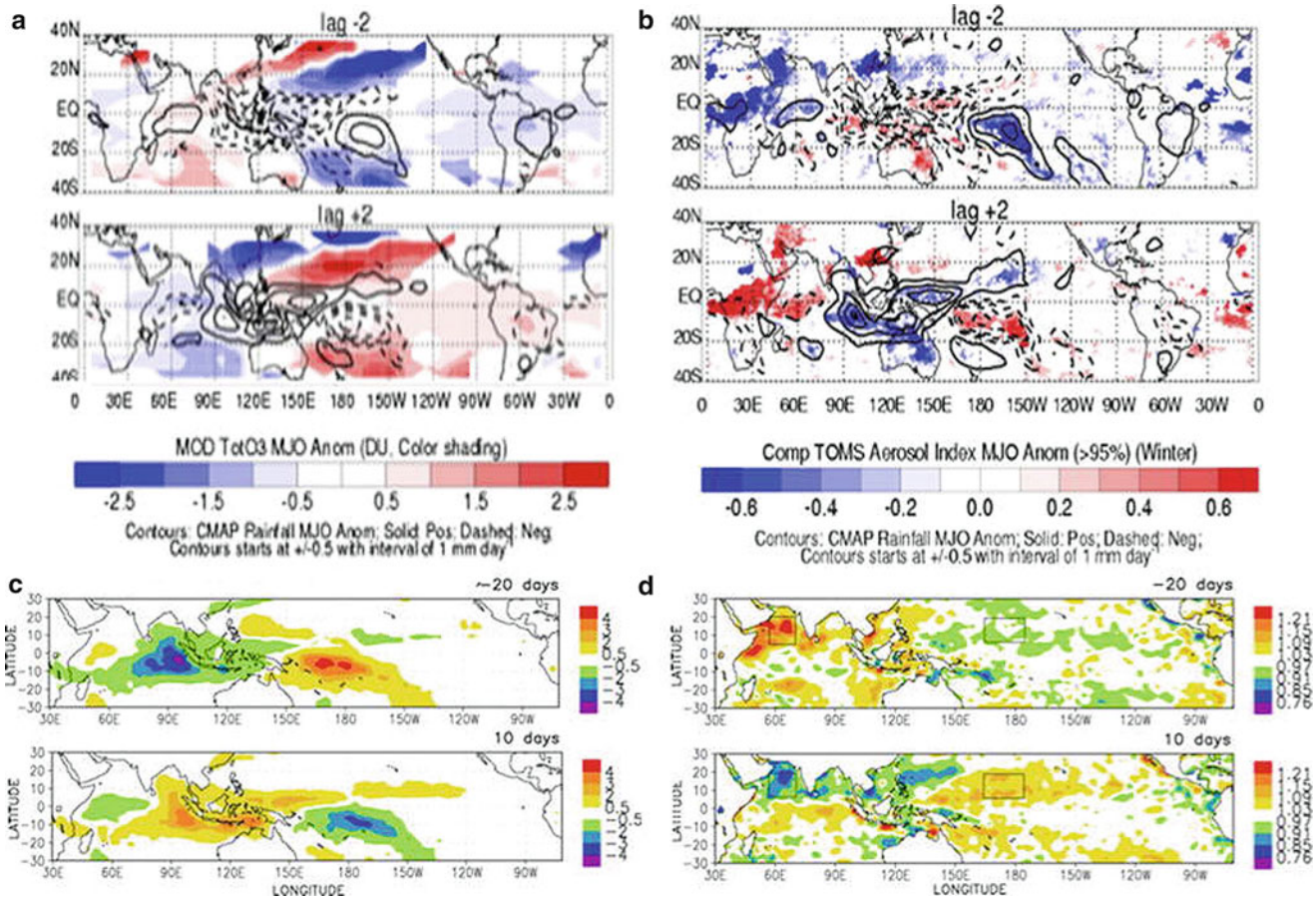
Although most quantities/processes/phenomena of interest, such as hurricanes, monsoons, and extratropical circulation and weather patterns, are not wholly described by satellite data, the satellite-based OLR or rainfall data were usually used to identify the MJO events and often to characterize aspects of the impacts (e.g., rainfall). Some studies totally depend on the available satellite data. For example, in the study of the MJO impact on the diurnal cycle of tropical deep convection, the International Satellite Cloud Climatology Project cloud product was employed to characterize the diurnal cycle, and the TRMM 3B42 precipitation product was used to identify the MJO events.

Recently, a number of studies have documented the MJO impacts on atmospheric composition, air quality, and biogeochemical cycle. This discovery has critically depended on the availability of satellite data. For example, the MJO impact on tropical total-column ozone has been recently characterized using the satellite-observed tropical total-column ozone from the AIRS and Total Ozone Mapping Spectrometer (TOMS). It was found that tropical total ozone intraseasonal variations are large ($\sim \pm 10$ Dobson unit) and comparable to those in the annual and interannual timescales. These intraseasonal total ozone anomalies are mainly evident in the subtropics over the Pacific and eastern hemisphere, with a systematic relationship to the MJO convection and wave dynamics discussed earlier (Figures 3 and 4a). The subtropical negative ozone anomalies typically flank or lie to the west of the equatorial anomalous convection and are collocated with the subtropical upper-troposphere anticyclones generated by the equatorial anomalous convective heating. On the other hand, the subtropical positive ozone anomalies generally lie to the east of the equatorial anomalous convection and are collocated with the subtropical upper-troposphere cyclones generated by the equatorial anomalous convective heating. The subtropical ozone anomalies are anticorrelated with geopotential height anomalies near the tropopause and thus mainly associated with the ozone variability in the stratosphere rather the troposphere.

Another example, the recent availability of multiple, global satellite aerosol products from TOMS, Moderate Resolution Imaging Spectroradiometer (MODIS), and Advanced Very High Resolution Radiometer (AVHRR) has made the investigation of the MJO modulation of aerosols possible. Large aerosol variations are found over the equatorial Indian and western Pacific Oceans where MJO convection is active, as well as the tropical Africa and Atlantic Ocean where MJO convection is weak, but the background aerosol level is high (Figures 3 and 4b). Although significant uncertainties still exist in the satellite aerosol retrievals, the satellite data indicate that the MJO and its associated cloudiness, rainfall, and circulation variability may systematically influence the aerosol variability.

The impacts of the MJO on the carbon monoxide (CO) abundances in the tropical tropopause layer (TTL) were also recently reported based on the Aura Microwave Limb Sounder (MLS) CO data. The effects of the eastward propagation of MJO on CO abundances in the TTL are evident. This indicates that the anomalous deep convection associated with the MJO can inject CO from the lower troposphere up to the TTL.

The availability of satellite-derived ocean surface chlorophyll (Chl) from Sea-viewing Wide Field-of-view Sensor (SeaWiFS) provided a means for the discovery of the MJO impacts on oceanic biology. It was found that the MJO produces systematic and significant variations in ocean surface Chl in a number of regions across the tropical Oceans, including the northern Indian Ocean, a broad expanse of the northwestern tropical Pacific Ocean, and a number of near-coastal areas in the far eastern Pacific Ocean (Figures 3 and 4c, d). This indicates a need to further investigate the MJO modulation of the biogeochemical cycle properties and higher levels of food chains.



Madden-Julian Oscillation (MJO), Figure 4 (**a**) Map of total-column ozone anomaly for a composite MJO cycle based on TOMS/SBUV data from 1980 to 2006. The color *red* denotes high ozone anomalies, while the color *blue* indicates low ozone anomalies. The superimposed *solid black line* denotes rainfall anomaly from CMAP (Reproduced from Tian et al., 2007, by permission of American Geophysical Union). (**b**) As (**a**) but for TOMS aerosol index anomaly (Reproduced from Tian et al., 2008, by permission of American Geophysical Union). (**c**) As (**a**) but for CMAP rainfall anomalies. (**d**) As (**a**) but for SeaWiFS ocean surface Chl anomaly (Reproduced from Waliser et al., 2005, by permission of American Geophysical Union).

Summary

The MJO is a large-scale quasiperiodic oscillation of tropical atmospheric circulation and convection anomalies that moves slowly eastward along the equator mainly over the tropical Indian and Pacific Oceans with a timescale on the order of 30–60 days. The MJO is the dominant form of the intraseasonal variability in the tropical atmosphere and has many important influences on the global weather and climate system. Since the 1970s, the satellite remote sensing data have played a fundamental role in advancing our knowledge in the MJO, particularly in terms of its description, theoretical mechanisms, and global impacts. First, the satellite data provided us the fundamental knowledge of the convective and dynamic features of the MJO. Second, the satellite data presented us the three-dimensional thermodynamic structure and the surface condition (e.g., SST and surface heat flux) evolution associated with the MJO that helped us to better understand the MJO and propose theoretical description. Third, the satellite data offered us the opportunity to discover the global impacts of the MJO that have relevance to societal concerns such as extreme precipitation events, atmospheric composition, air quality, and biological markers in the ocean.

Acknowledgment

This research was carried out at Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA.

Bibliography

Gottschalck, J., Kousky, V., Higgins, W., and L'Heureux, M., 2008. Madden-Julian oscillation. http://www.cpc.noaa.gov/products/precip/CWlink/MJO/mjo.shtml

Hendon, H. H., and Salby, M. L., 1994. The life-cycle of the Madden-Julian oscillation. *Journal of the Atmospheric Sciences*, **51**(15), 2225–2237.

Lau, W. K. M., and Waliser, D. E. (eds.), 2005. *Intraseasonal Variability of the Atmosphere–Ocean Climate System*. Heidelberg: Springer. 474 pp.

Madden, R. A., 2003. Tropical meteorology: intraseasonal oscillation (Madden-Julian oscillation). In Holton, J. R., et al. (eds.), *Encyclopedia of Atmospheric Sciences*. London: Academic, pp. 2334–2338.
Madden, R. A., and Julian, P. R., 1971. Detection of a 40–50 day oscillation in the zonal wind in the tropical pacific. *Journal of the Atmospheric Sciences*, **28**(7), 702–708.
Madden, R. A., and Julian, P. R., 1972. Description of global-scale circulation cells in tropics with a 40–50 day period. *Journal of the Atmospheric Sciences*, **29**(6), 1109–1123.
Madden, R. A., and Julian, P. R., 1994. Observations of the 40–50-day tropical oscillation: a review. *Monthly Weather Review*, **122**(5), 814–837.
Tian, B., Waliser, D. E., Fetzer, E. J., Lambrigtsen, B. H., Yung, Y. L., and Wang, B., 2006. Vertical moist thermodynamic structure and spatial-temporal evolution of the MJO in AIRS observations. *Journal of the Atmospheric Sciences*, **63**(10), 2462–2485, doi:10.1175/JAS3782.1.
Tian, B., Yung, Y. L., Waliser, D. E., Tyranowski, T., Kuai, L., Fetzer, E. J., and Irion, F. W., 2007. Intraseasonal variations of the tropical total ozone and their connection to the Madden-Julian oscillation. *Geophysical Research Letters*, **34**(8), L08704, doi:10.1029/2007GL029451.
Tian, B., Waliser, D. E., Kahn, R. A., Li, Q. B., Yung, Y. L., Tyranowski, T., Geogdzhayev, I. V., Mishchenko, M. I., Torres, O., and Smirnov, A., 2008. Does the Madden-Julian oscillation influence aerosol variability? *Journal of Geophysical Research*, **113**(D12), D12215, doi:10.1029/2007JD009372.
Waliser, D. E., 2006. Intraseasonal variability. In Wang, B. (ed.), *The Asian Monsoon*. New York: Springer/Praxis, pp. 203–257.
Waliser, D. E., Murtugudde, R., Strutton, P., and Li, J. L., 2005. Subseasonal organization of ocean chlorophyll: prospects for prediction based on the Madden-Julian oscillation. *Geophysical Research Letters*, **32**(23), L23602, doi:10.1029/2005GL024300.
Zhang, C. D., 2005. Madden-Julian oscillation. *Reviews of Geophysics*, **43**(2), RG2003, doi:10.1029/2004RG000158.

Cross-references

Aerosols
Radars

MAGNETIC FIELD

Nils Olsen
DTU Space, Technical University of Denmark, Lyngby, Denmark

Definition

Core field. The main part of the geomagnetic field, caused by dynamo action in the Earth's fluid outer core.
Crustal field. The magnetic field contribution caused by (permanent or remnant) magnetized material in the Earth's crust.
Internal sources. Magnetic field contributions caused by electrical currents or magnetized material in the Earth's interior.
External sources. Magnetic field contributions produced by electric currents in the ionosphere or magnetosphere.

Introduction

The Earth has a large and complicated magnetic field, the major part of which is produced by a self-sustaining dynamo operating in the fluid outer core. Magnetic field observations provide one of the few tools for remote sensing the Earth's deep interior, especially regarding the dynamics of the fluid flow at the top of the core. However, what is measured at or near the surface of the Earth is the superposition of the core field and fields caused by magnetized rocks in the Earth's crust, by electric currents flowing in the ionosphere, magnetosphere, and oceans, and by currents induced in the Earth by time-varying external fields. These sources have their specific characteristics in terms of spatial and temporal variations, and their proper separation, based on magnetic measurements, is a major challenge. Such a separation is a prerequisite for remote sensing by means of magnetic field observations.

Data sources

Prior to the satellite era, only near-surface (ground-based, marine, and airborne) magnetic observations were available for magnetic probing the Earth's interior. Presently about 150 geomagnetic observatories monitor the time changes of the field; their data are available through the World Data Center system (e.g., www.ngdc.noaa.gov/wdc, www.wdc.kugi.kyoto-u.ac.jp, www.wdc.bgs.ac.uk) and INTERMAGNET (www.intermagnet.org). However, the global distribution of the observatory network is very uneven, with large uncovered areas especially over the oceans.

While observatories monitor the field change at a given fixed location, regional surveys map the magnetic field at a given epoch (short period fluctuations of external origin have to be removed from the raw data; for this purpose, measurements of the field change at a nearby observatory or a temporary station are used). However, only parts of the globe are presently covered by airborne and marine surveys. The obtained data have been used for instance in the preparation of the World Digital Magnetic Anomaly Map (WDMAM, Korhonen et al. (2007), see also http://projects.gtk.fi/WDMAM/) and dedicated anomaly maps like EMAG2 (Maus et al., 2009).

True global coverage with magnetic field observations is only possible with satellites. Although spaceborne probing began more than 60 years ago with the launch of the *Sputnik 3* satellite in 1958, the data coverage that is necessary for global field modeling was first obtained by the *POGO* satellite series (Cain, 2007) which measured the magnetic field intensity between 1965 and 1972. The first high-precision vector measurements from space were taken by the Magsat satellite (Purucker, 2007) in 1979–1980. More recently, the launch of the satellites Ørsted in February 1999 (Olsen, 2007), CHAMP in July 2000 (Maus, 2007), and SAC-C in November 2000 opened revolutionary new possibilities for probing the Earth from space. In the near future, the *Swarm* satellite constellation mission (Friis-Christensen et al., 2006, 2009), comprising

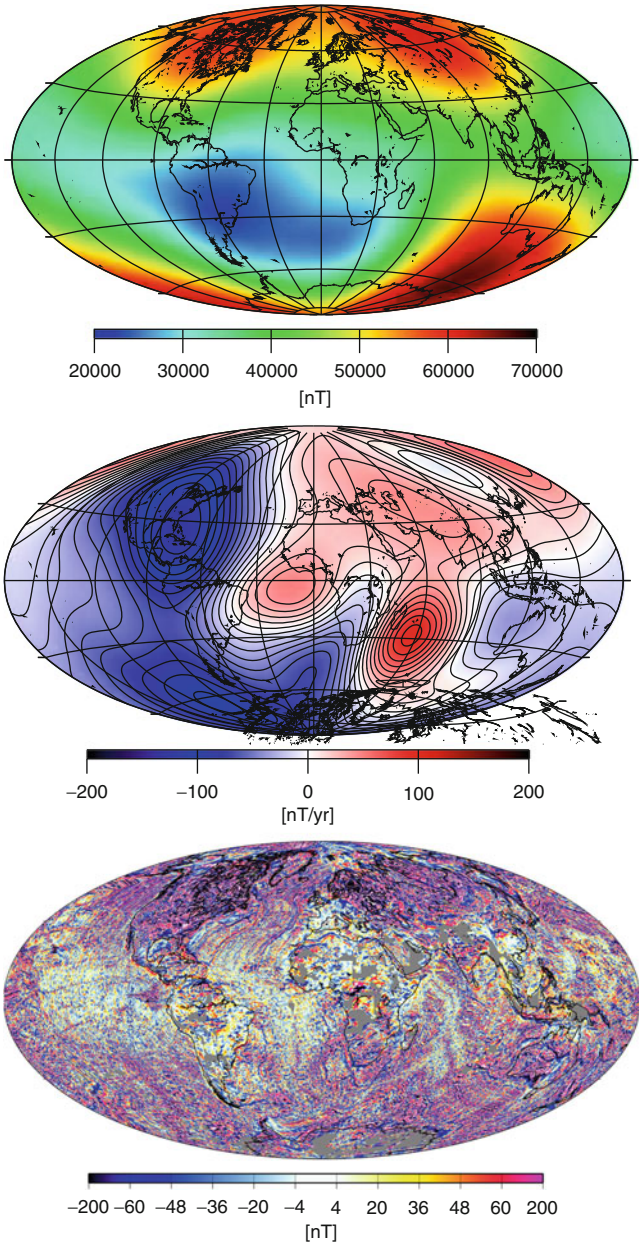
three satellites expected to be launched in late 2013, will provide even better possibilities for magnetic remote sensing the Earth's interior.

Characteristics of the geomagnetic field

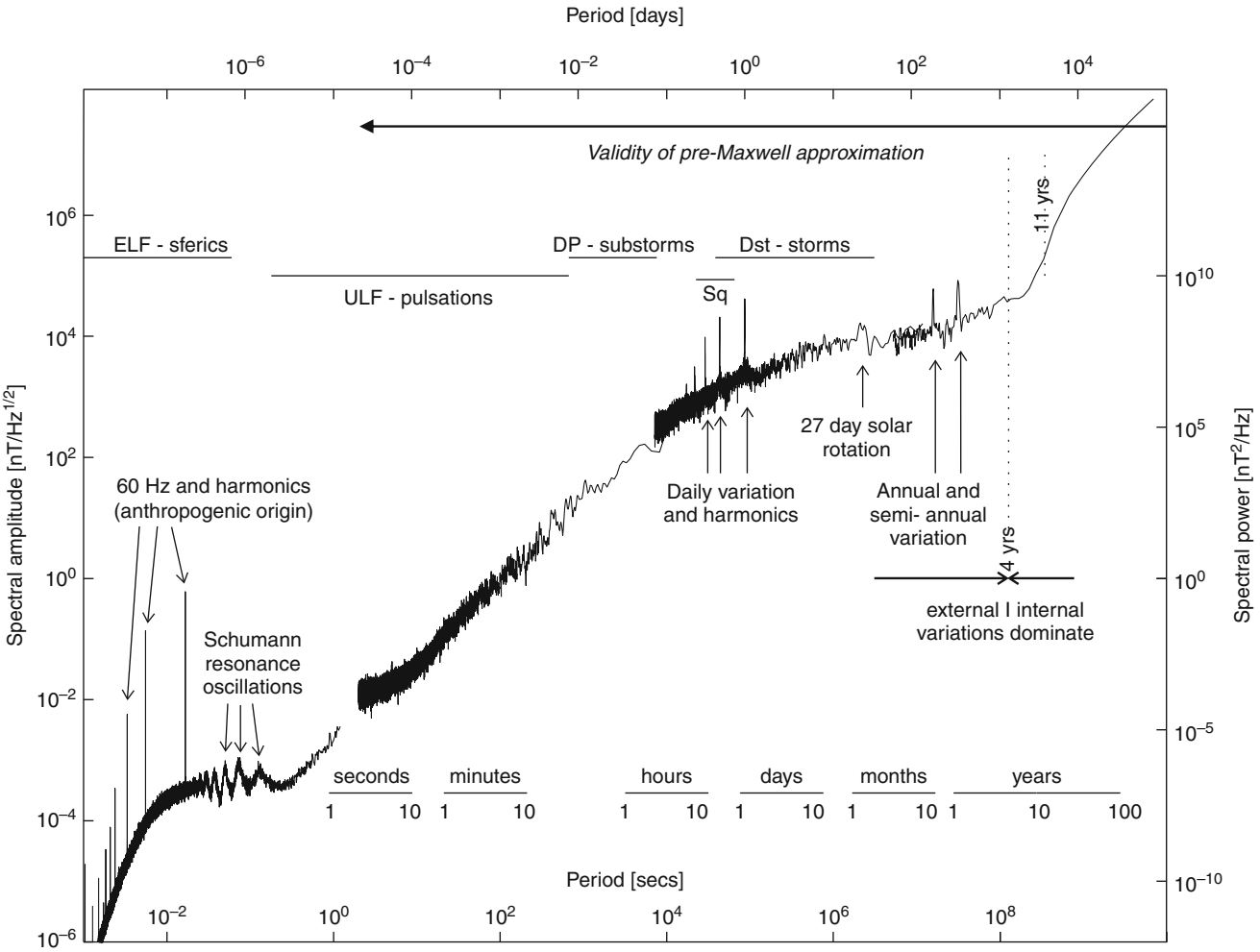
The strength of the magnetic induction **B**, in following for simplicity denoted as "magnetic field," varies at Earth's surface between about 25,000 nT near the equator and about 65,000 nT near the poles (1 nT = 10^{-9} T, with 1 T = 1 tesla = 1 V/s^{-1} m^{-2}). By far the largest part (95 % or more at Earth's surface) is due to dynamo action in the core (e.g., Roberts, 2007); magnetized material in the crust (e.g., Purucker and Whaler, 2007) accounts on average for only a few % of the total field but can locally reach magnitudes of several hundreds or even thousands of nT. Crustal magnetization consists of two parts: *Induced magnetization* is proportional, both in strength and direction, to the ambient field within which the rock is embedded. Were the core field to disappear, induced magnetization would vanish, too. Then, only the second type of magnetization, *remnant magnetization*, remains. As a general rule, remnant magnetization is weak in continental regions (where induced magnetization dominates), while both types of magnetizations are significant in oceanic areas.

The top panel of Figure 1 shows the magnetic field intensity at Earth's surface in 2010.0, and the middle panel shows the yearly change in field intensity (also for epoch 2010.0), based on the International Geomagnetic Reference Field (IGRF) (Finlay et al., 2010). Core and crust contribute to all spatial scales of the field, but the core field dominates for horizontal scales larger than 2,800 km (corresponding to spherical harmonic degrees $n < 14$) while the crustal field dominates for horizontal scales smaller than 2,800 km ($n > 14$). The bottom panel of the figure shows the crustal anomaly field (scales smaller than 2,400 km) close to the surface (4 km above the geoid), based on the EMAG2 model of Maus et al. (2009). The crustal field is static (at least on timescales of centuries or shorter), but the core field undergoes a significant time change, known as secular variation. The temporal change shown in the middle panel of Figure 1 is therefore caused by processes in the Earth's core.

In addition to these internal sources, there are contributions from electric currents in the ionosphere (90–1,000 km altitude) and the magnetosphere (at distance larger than several Earth radii); these are called external sources. They are very dynamic, ranging from a few nT during geomagnetic quiet conditions to several hundreds or even thousands of nT during disturbed times, with especially large amplitudes at polar latitudes. Figure 2 shows a typical spectrum of the magnetic North component for a site at mid-latitudes. Time changes with periods longer than 4 years are dominated by core processes, while those at shorter periods, and especially signals with seasonal and daily periodicity, are caused by ionospheric and magnetospheric sources. See Schmucker (1985) for details on the external field variations ELF, sferics, ULF, pulsations, DP, Dst, and Sq. These short period changes, which are caused by external currents, produce secondary, induced, currents in the Earth's interior, the magnetic field of which adds to that of the primary, external, currents. The observed magnetic time variations are thus a superposition of primary (inducing) and secondary (induced) contributions. Analysis of these time changes



Magnetic Field, Figure 1 Earth's magnetic field intensity (*top*) and its time change (*middle*) in 2010, at Earth's surface. *Bottom*: intensity of magnetic field anomalies at 4 km altitude above the geoid.



Magnetic Field, Figure 2 Amplitude of magnetic field variations (*north component*) at middle latitudes in dependence on period, determined using data from the sites Socorro/USA (1 ms to 1 s), Kakioka/Japan (2 s to 2 h), and Niemegk/Germany (longer than 2 h). Note that the pre-Maxwell approximation Equation 1 is only valid for periods $\gg$1 s.

provides information on the electrical conductivity of the Earth's interior (e.g., Constable, 2007).

Although field changes of internal as well as external origin occur at all timescales, a common practice in separating them relies on their different temporal variations. Over the last 150 years, the axial dipole component of the Earth's magnetic field has decayed by nearly 10 %. This is ten times faster than the natural decay, in case the dynamo was switched off. The current decay rate is characteristic of magnetic reversals, which – as paleomagnetic data have shown – occur on average about once every half million years. Geographically, the recent dipole decay is largely due to changes in the field beneath the South Atlantic Ocean, connected to the growth of the South Atlantic anomaly. The core field and its temporal change (secular variation) directly reflect the fluid flow in the outermost core and provide a unique observational constraint on geodynamo theory. However, only the part of the core field that varies on timescales longer than, say, 1 year is observable at the Earth's surface; shorter fluctuations are heavily attenuated due to the electrical conductivity of the mantle. Hence, variations with timescales longer than a few years are usually attributed to processes in the core, whereas those with periods shorter than 1 year are attributed to external field contributions. Yet interesting features occur at intermediate timescales. A serious limitation regarding the investigation of core processes at timescales of months to years is the effect of geomagnetic variations of external origin, since they contribute significantly on timescales up to that of the 11 year solar cycle.

Basic equations

Magnetic field investigation of core and crust is typically done in the quasi-static approximation, which requires that the timescales in consideration are longer ($\gg$1 s) compared to the time required for light to pass the length scale of interest ($<$a few thousand kilometers). In this so-called

pre-Maxwell approximation, displacement currents can be neglected, and the magnetic field **B** is given by

$$\times \mathbf{B} = \mu_0 \mathbf{J} \quad (1)$$

where $\mu_0 = 4\pi 10^{-7}$ Vs/(Am) is vacuum permeability and current density

$$\mathbf{J} = \mathbf{J}_e + \mathbf{J}_m \quad (2)$$

(expressed in units of A/m^2) is the sum of free-charge current density $\mathbf{J}_e$ and the equivalent current density, $\mathbf{J}_m = \nabla \times \mathbf{M}$, due to material of magnetization **M** (units of A/m). The sources of magnetic fields are therefore electric currents (for instance, in the Earth's core, the ionosphere or the magnetosphere) and/or magnetized material (for instance, in the Earth's crust).

Outside its sources (i.e., in regions with $\mathbf{J} = 0$), the magnetic field, $\mathbf{B} = -\nabla V$, is a Laplacian potential field and can be derived from a scalar magnetic potential *V*. In that case, the magnetic field has similar properties as the gravity field and the same methods for studying both fields might be used (Blakely, 1995). However, the similarity is only true under certain assumptions.

While near-surface magnetic field measurements are compatible with the potential assumption (since electric currents are negligible in the atmosphere and thus the data are obtained in a source-free region), the situation is different for spaceborne observations: Satellites probing the Earth's magnetic field typically fly at altitudes between 300 and 1,000 km in the ionospheric *F*-region, and especially at polar latitudes, the occurrence of local electric currents violates the assumption of a current-free data-sampling region. Satellite magnetic field data therefore cannot be described by a Laplacian potential field alone. Strategies for handling satellite data are discussed for instance in Hulot et al. (2007), Olsen et al. (2010), and Sabaka et al. (2010).

Applications of magnetic remote sensing

There are three main aspects of probing the Earth's interior using magnetic field observations.

Firstly, the dynamics of fluid flow at the top of the core can be determined from observations of the core field and its time change. This is a nonunique problem which requires assumptions on the flow dynamics, for instance to be steady or purely horizontal (i.e., toroidal). In addition to the fundamental nonuniqueness, the fact that only the slowly varying large-scale part of the core field is observable (periods shorter than a few months are heavily attenuated when propagating through the electrically conducting mantle and thus not observable at the surface, and the small-scale part of the core field is masked by the crustal field) puts additional limitations on the fraction of core flow that can be determined. Nevertheless, the magnetic field has proven to provide an extremely useful (and the only) tool to probe core flow. Recent overviews on this subject are given by Holme (2007) and Whaler (2007).

Secondly, information on the structure, composition, and temperature of the crust can be obtained from investigations of the crustal magnetic field. Its small-scale part, mapped by regional near-surface surveys, provides information on shallow geological structures useful to petroleum and mineral exploration, while the long-wavelength part (up to 2,800 km – longer scales are masked by the core field) is a valuable tool to investigate the lower crust and upper mantle. See Langel and Hinze (1998), Purucker and Whaler (2007), and Ravat (2007) for more details on crustal field studies.

Finally, observations of the time-changing part of the magnetic field (of periods between seconds and several days) form the basis for electromagnetic induction studies, providing information on the electrical conductivity of the crust and mantle. The obtained conductivity distribution gives insight into geodynamic processes. This is complementary to seismic remote sensing since conductivity reflects the connectivity of constituents such as fluids, partial melt, and volatiles (all of which may have influence on rheology, mantle convection, and tectonic activity), while seismology ascertains bulk mechanical properties. A description of the techniques for probing the conductivity of the Earth's interior, and examples of obtained results, is given for instance in Constable (2007).

Conclusions

Magnetic field observations provide one of the few tools for remote sensing the Earth's interior, especially regarding the dynamics of core fluid flow and the structure, composition, and temperature of the crust. In addition, studying time variations of the magnetic field allows for determination of the electrical conductivity of the crust and mantle.

Bibliography

Blakely, R. J., 1995. *Potential Theory in Gravity and Magnetic Applications*. Cambridge: Cambridge Press.

Cain, J. C., 2007. POGO (OGO-2, -4 and -6 spacecraft). In Gubbins, D., and Herrero-Bervera, E. (eds.), *Encyclopedia of Geomagnetism and Paleomagnetism*. Heidelberg: Springer.

Constable, S., 2007. Geomagnetic induction studies. In Kono, M. (ed.), *Treatise on Geophysics*. Amsterdam: Elsevier, Vol. 5, pp. 237–276.

Finlay, C. C., Maus, S., Beggan, C. D., Bondar, T. N., Chambodut, A., Chernova, T. A., Chulliat, A., Golovkov, V. P., Hamilton, B., Hamoudi, M., Holme, R., Hulot, G., Kuang, W., Langlais, B., Lesur, V., Lowes, F. J., Luehr, H., Macmillan, S., Mandea, M., McLean, S., Manoj, C., Menvielle, M., Michaelis, I., Olsen, N., Rauberg, J., Rother, M., Sabaka, T. J., Tangborn, A., Toeffner-Clausen, L., Th'ebault, E., Thomson, A. W. P., Wardinski, I., Wei, Z., and Zvereva, T. I., 2010. International geomagnetic reference field: the eleventh generation. *Geophysical Journal International*, **183**, 1216–1230, doi:10.1111/j.1365-246X.2010.04804.x.

Friis-Christensen, E., Lühr, H., and Hulot, G., 2006. Swarm: a constellation to study the earth's magnetic field. *Earth Planets Space*, **58**, 351–358.
Friis-Christensen, E., Lühr, H., Hulot, G., Haagmans, R., and Purucker, M., 2009. Geomagnetic research from space. *EOS, Transactions, American Geophysical Union*, **900**(25), 213–215.
Holme, R., 2007. Large-scale flow in the core. In Olson, P. (ed.), *Treatise on Geophysics*. Amsterdam: Elsevier Science, Vol. 8, pp. 107–130.
Hulot, G., Sabaka, T. J., and Olsen, N., 2007. The present field. In Kono, M. (ed.), *Treatise on Geophysics*. Amsterdam: Elsevier, Vol. 5, pp. 33–75.
Korhonen, J., Fairhead, J., Hamoudi, M., Hemant, K., Lesur, V., Mandea, M., Maus, S., Purucker, M., Ravat, D., Sazonova, T., et al., 2007. *Magnetic Anomaly Map of the World*. Helsinki: Map published by Commission for Geological Map of the World, supported by UNESCO, GTK.
Langel, R. A., and Hinze, W. J., 1998. *The Magnetic Field of the Earth's Lithosphere: The Satellite Perspective*. Cambridge: Cambridge University Press.
Maus, S., 2007. CHAMP magnetic mission. In Gubbins, D., and Herrero-Bervera, E. (eds.), *Encyclopedia of Geomagnetism and Paleomagnetism*. Heidelberg: Springer.
Maus, S., Barckhausen, U., Berkenbosch, H., Bournas, N., Brozena, J., Childers, V., Dostaler, F., Fairhead, J., Finn, C., von Frese, R., et al., 2009. EMAG2: a 2–arc min resolution earth magnetic anomaly grid compiled from satellite, airborne, and marine magnetic measurements. *Geochemistry, Geophysics, Geosystems*, **100**(8), Q08005.
Olsen, N., 2007. Ørsted. In Gubbins, D., and Herrero-Bervera, E. (eds.), *Encyclopedia of Geomagnetism and Paleomagnetism*. Heidelberg: Springer.
Olsen, N., Hulot, G., and Sabaka, T. J., 2010. Sources of the geomagnetic field and the modern data that enable their investigation, Chapter 5. In Freeden, W., Nashed, Z., and Sonar, T. (eds.), *Handbook of Geomathematics*. Heidelberg: Springer, pp. 106–124, doi:10.1007/978-3-642-01546-5_5.
Purucker, M., and Whaler, K., 2007. Crustal magnetism. In Kono, M. (ed.), *Treatise on Geophysics*. Amsterdam: Elsevier, Vol. 5, pp. 195–235.
Purucker, M. E., 2007. Magsat. In Gubbins, D., and Herrero-Bervera, E. (eds.), *Encyclopedia of Geomagnetism and Paleomagnetism*. Heidelberg: Springer.
Ravat, D., 2007. Crustal magnetic field. In Gubbins, D., and Herrero-Bervera, E. (eds.), *Encyclopedia of Geomagnetism and Paleomagnetism*. Heidelberg: Springer.
Roberts, P. H., 2007. Theory of the geodynamo. In *Treatise on Geophysics*. Amsterdam: Elsevier, Vol. 8, pp. 67–106.
Sabaka, T. J., Hulot, G., and Olsen, N., 2010. Mathematical properties relevant to geomagnetic field modelling, Chapter 17. In Freeden, W., Nashed, Z., and Sonar, T. (eds.), *Handbook of Geomathematics*. Heidelberg: Springer, pp. 504–538, doi:10.1007/978-3-642-01546-5_17.
Schmucker, U., 1985. Sources of the geomagnetic field. In *Landolt-Börnstein, New-Series, 5/2b*. Berlin/Heidelberg: Springer, pp. 31–73.
Whaler, K. A., 2007. Core motions. In Gubbins, D., and Herrero-Bervera, E. (eds.), *Encyclopedia of Geomagnetism and Paleomagnetism*. Heidelberg: Springer.

Cross-references

Electromagnetic Theory and Wave Propagation
Fields and Radiation
Geodesy
Observational Systems, Satellite

MEDIA, ELECTROMAGNETIC CHARACTERISTICS

Yang Du
Zhejiang University, Hangzhou, People's Republic of China

Definition

Media, electromagnetic characteristics. Macroscopic permittivity and permeability properties of media.

Introduction

When electromagnetic wave propagates in some medium other than vacuum, since the characteristic wavelength is several orders larger than the atoms of which the medium is composed, the detailed behavior of the fields over atomic distance becomes irrelevant. What do matter are the quantities averaged over the atomic scale, including the macroscopic fields and macroscopic sources. Such treatment implies that the inhomogeneous medium is essentially replaced by a homogeneous one with macroscopic permittivity and permeability (Jackson, 1998).

Macroscopic properties

In macroscopic media, the electric displacement field **D** and magnetic field **H** are related to the electric field **E** and magnetic induction **B** through the macroscopically averaged electric dipole, magnetic dipole, electric quadrupole, and higher-order multipoles. In most materials, only the electric and magnetic polarizations are significant.

In an isotropic medium, the constitutive relations are $\mathbf{D} = \varepsilon\mathbf{E}$, and $\mathbf{B} = \mu\mathbf{H}$, where ε and μ are the permittivity and permeability, respectively. A diamagnetic medium has μ larger than μ_0, since diamagnetic substance, whose atoms or molecules have no angular momentum, creates induced magnetic moments that tend to oppose the applied magnetic field. A paramagnetic medium has μ smaller than μ_0, since paramagnetic substance has a net angular momentum which is aligned parallel to the applied magnetic field.

For anisotropic media, the constitutive relations are $\mathbf{D} = \overline{\overline{\varepsilon}}\mathbf{E}$ and $\mathbf{B} = \overline{\overline{\mu}}\mathbf{H}$, where $\overline{\overline{\varepsilon}}$ and $\overline{\overline{\mu}}$ are the permittivity and permeability tensor, respectively. Crystals are generally characterized by symmetric permittivity tensors, which can be transformed into a diagonal matrix as

$$\overline{\overline{\varepsilon}} = \begin{bmatrix} \varepsilon_x & 0 & 0 \\ 0 & \varepsilon_y & 0 \\ 0 & 0 & \varepsilon_z \end{bmatrix}$$

When $\varepsilon_x = \varepsilon_y = \varepsilon_z$, the crystals are isotropic, such as the cubic crystals. Uniaxial crystals have two of the three parameters equal. Examples are tetragonal, hexagonal, and rhombohedral crystals (Kong, 2005).

For bianisotropic media, the constitutive relations are $\mathbf{D} = \overline{\overline{\varepsilon}}\mathbf{E} + \overline{\overline{\xi}}\mathbf{H}$ and $\mathbf{B} = \overline{\overline{\zeta}}\mathbf{E} + \overline{\overline{\mu}}\mathbf{H}$ (Kong, 2005).

Dispersion

All media are dispersive in that the phase velocity of a wave depends on the frequency. That is, waves with different frequencies travel in a dispersive medium at different speeds. The frequency dependence of ε and μ leads to new effects when an arbitrary wave train containing a range of frequencies travels.

What enter the picture of dispersion are the molecular constitution of matter and the dynamics of molecules. A simple model leads to a dispersion formula which in most cases serves as an adequate representation of the dielectric constant as a function of frequency. The success of this model is truly amazing considering the several assumptions being made, including treating the relative permeability equal to unity, neglecting magnetic force effects, and confining the oscillation to be sufficiently small. The forces exerting upon an electron include an electric force due to the electric field, a restoring force, and a damping force. The resultant formula for the dielectric constant is

$$\varepsilon(\omega) = \varepsilon_0 + \frac{Ne^2}{m} \sum_k \frac{\nu_k}{\omega_k^2 - \omega^2 - i\omega\gamma_k},$$

where $-e$ and m are the charge and mass of an electron, respectively. Here it is assumed that there are N molecules per unit volume and ν_k electrons per molecule with binding frequency ω_k and phenomenological damping constant γ_k (Jackson, 1998).

When real part of $\varepsilon(\omega)$ increases with ω, it is called normal dispersion. Anomalous dispersion refers to the reverse. Since in general the resonant frequencies ω_k are large compared to the damping factors γ_k, in the neighborhood of ω_k, the behavior of $\varepsilon(\omega)$ is expected to be rather violent, with disruption of the normal dispersion and appearance of appreciable imaginary part of $\varepsilon(\omega)$ or resonant absorption.

At frequencies far beyond the highest resonant frequency, the dielectric constant is simply expressed in terms of the plasma frequency ω_p of the medium as $\varepsilon(\omega) \approx \varepsilon_0 - \frac{\omega_p^2}{\omega^2}\varepsilon_0$.

The Kramers-Kronig relations or dispersion relations are simple integral formula relating the real part to imaginary part of the complex permittivity $\varepsilon(\omega)$ or a dispersive process to an absorption process. The validity of these relations is very general due to the fact that very few assumptions are made, among which are the monochromatic components of the displacement $\mathbf{D}(\mathbf{r}, \omega)$ and the electric field $\mathbf{E}(\mathbf{r}, \omega)$ at position $\mathbf{r}$ related by $\mathbf{D}(\mathbf{r}, \omega) = \varepsilon(\omega)\mathbf{E}(\mathbf{r}, \omega)$, as well as a causality requirement on the kernel relating $\mathbf{D}(\mathbf{r}, t)$ and $\mathbf{E}(\mathbf{r}, t)$. The wide generality of the dispersion relations makes them very useful in all areas of physics (Toll, 1956). For instance, in the scattering of nuclear particles, these relations connect real and imaginary parts of the diagonal elements of the scattering matrix (Jost et al., 1950).

Inhomogeneous composite

A macroscopically inhomogeneous medium is a typical representation of many materials. For instance, a snow pack is a composite of ice particles and water; a porous rock is a composite of the rock matrix and salt water if the latter is present. In each example, the medium possesses spatially varying quantities such as permittivity and conductivity.

Electromagnetic characterization of inhomogeneous composite materials is important in a wide variety of applications, such as in astronomy and atmospheric physics and selective absorbers of solar and infrared radiation.

In the effective medium approximation of Bruggeman and Landauer, a composite of two components is considered. To obtain the effective conductivity of this composite, each particle is imagined to be immersed in a homogeneous effective medium of conductivity σ_e instead of in its actual inhomogeneous medium. The self-consistency condition is invoked such that the average electric field within a particle shall equal the electric field far from the particle. In essence, in the effective-medium approximation, only electron-dipole scattering contributes. The characteristic particle dimension in the composite is required to be small compared to the characteristic wavelength. If a two-component composite has one component dominant in volume fraction, then both approaches work well. However, it is found that EMA produces significant error in far-infrared absorption by a composite composed of dielectric and small metal particles with concentrations below the percolation threshold (Stroud, 1998).

A self-consistency condition is to choose an effective dielectric constant such that for particles embedded in the medium, on the average, their forward scattering amplitude should vanish (Stroud and Pan, 1978). Generalization to continuous size distribution of particles as well as arbitrary number of components is provided in Chylek and Srivastava (1983), where both the contribution of electric and magnetic dipole terms are considered. Special attention is paid to a composite material where small metallic particles are among the components. The form of the size distribution of metallic particles is found to determine the critical volume fraction at which there is a drastic absorption increase.

Metamaterial

Metamaterial is a new class of artificially constructed electromagnetic material that can exhibit electromagnetic properties difficult or impossible to find in conventional materials (Pendry, 2004; Smith et al., 2004). For example, the material property of negative index of refraction has been realized at both GHz and optical frequencies using metamaterial concepts.

Recent advances in construction of metamaterial with independent and arbitrary varying capability of

permittivity and permeability values pave the way for totally new electromagnetic phenomena. For instance, the electric displacement **D**, the magnetic field intensity **B**, and the Poynting vector **S**, by controlling the material electromagnetic properties, can be directed at will, made focused, or avoid objects. In particular is the design of cloaking of objects from electromagnetic fields (Pendry et al., 2006).

Summary

In macroscopic media, the electric displacement field and magnetic induction can be related to the electric intensity and magnetic intensity in ways quite different from that in vacuum. With increasing complexity, the medium can be characterized as isotropic, anisotropic, and bianisotropic. All media are dispersive in that the phase velocity of a wave depends on the frequency. The real and imaginary parts of the frequency-dependent complex permittivity are related by the Kramers-Kronig relations, which are very useful in all areas of physics. In characterizing the electromagnetic properties of an inhomogeneous medium, the effective-medium approximation is useful within its region of validity. Beyond that, the method base on the self-consistency condition can be employed. In terms of artificial materials, metamaterial is a new class of artificially constructed electromagnetic material that can exhibit electromagnetic properties difficult or impossible to find in conventional materials, which leads to totally new electromagnetic phenomena.

Bibliography

Chylek, P., and Srivastava, V., 1983. Dielectric constant of a composite inhomogeneous medium. *Physical Review B*, **27**, 5098.

Jackson, J. D., 1998. *Classical Electrodynamics*, 3rd edn. New York: Wiley.

Jost, R., Luttinger, J. M., and Slotnick, M., 1950. Distribution of recoil nucleus in pair production by photons. *Physical Review*, **80**, 189.

Kong, J. A., 2005. *Electromagnetic Wave Theory*. Cambridge: EMW.

Pendry, J. B., 2004. A chiral route to negative refraction. *Science*, **306**, 1353.

Pendry, J. B., Schurig, D., and Smith, D. R., 2006. Controlling electromagnetic fields. *Science*, **312**, 1780.

Smith, D. R., Pendry, J. B., and Wiltshire, M. C. K., 2004. Metamaterials and negative refractive index. *Science*, **305**, 788.

Stroud, D., 1998. The effective medium approximation: some recent developments. *Superlattices and Microstructures*, **23**, 567.

Stroud, D., and Pan, F. P., 1978. Self-consistent approach to electromagnetic wave propagation in composite media: application to model granular metals. *Physical Review B*, **17**, 1602.

Toll, J. S., 1956. Causality and the dispersion relation: logical foundations. *Physical Review*, **104**, 1760.

Cross-references

Electromagnetic Theory and Wave Propagation

MICROWAVE DIELECTRIC PROPERTIES OF MATERIALS

Martti Hallikainen
Aalto University, Espoo, Finland

Synonyms

Dielectric constant; Permittivity

Definitions

Permittivity characterizes the electrical properties of materials: The real part gives the contrast with respect to vacuum and the imaginary part gives the electromagnetic loss of the material. Loss tangent is defined as the imaginary part divided by the real part of permittivity. Total electromagnetic loss in a medium consists of absorption loss (electromagnetic power transformed into other forms of energy, such as heat) and scattering loss (energy is caused to travel in directions other than that of incident radiation). Penetration depth provides an approximate value to the maximum depth of the medium that contributes to the backscattering coefficient and brightness temperature. Optical depth between two points in a medium provides a measure of how transparent the medium is to electromagnetic power passing through it. Both the real and imaginary parts of the permittivity of water are very large compared to other natural media; hence, the dielectric properties of water tend to dominate those of materials containing water.

Basic electromagnetic quantities for describing materials

Several basic electromagnetic quantities are defined in order to characterize the microwave interaction with matter. Two electrical quantities, electric flux density and electric field intensity, constitute a fundamental pair of electromagnetic fields, and they are related via

$$\vec{D} = \varepsilon\varepsilon_0 \vec{E} \tag{1}$$

where $\vec{D}$ is the electric flux density (displacement) vector, $\vec{E}$ is the electric field vector, ε is the relative (compared to vacuum) permittivity, and $\varepsilon_0 = 8.854 \times 10^{-12}$ As/V/m is the vacuum permittivity. The permittivity of free space (air) is equal to that of vacuum with a high degree of accuracy.

Relative permittivity and loss tangent

The relative permittivity is a complex number and is denoted by

$$\varepsilon = \varepsilon' - j\varepsilon'' \tag{2}$$

where $j = \sqrt{-1}$. It characterizes the electrical properties of materials: The real part (ε') gives the contrast with respect to free space ($\varepsilon'_{air} = 1$), and the imaginary part (ε'') gives the electromagnetic loss of the material. The real

and imaginary parts of ε are often referred to as the dielectric constant and the dielectric loss factor, respectively. The loss tangent, defined as

$$\tan(\delta) = \frac{\varepsilon''}{\varepsilon'} \tag{3}$$

may also be used to characterize the loss of materials.

Propagation, absorption, and phase constant

For an electromagnetic plane wave traveling in the z-direction, the intensity of the electric field at point z can be expressed as

$$E(z) = E_0\, e^{-\gamma z} \tag{4}$$

where E_0 is the field intensity at $z = 0$. The complex propagation constant of the medium is denoted by γ and is expressed as

$$\gamma = \alpha + j\beta \tag{5}$$

where α is the absorption constant and β is the phase constant. The absorption constant describes transformation of wave energy into other forms of energy, such as heat. The phase constant is equal to the wave number $k = 2\pi/\lambda$, where λ is wavelength, in a lossless medium. α and β are related to the complex permittivity by

$$\alpha = k_0 \left|\mathrm{Im}\left\{\sqrt{\varepsilon}\right\}\right| \tag{6}$$

$$\beta = k_0 \mathrm{Re}\left\{\sqrt{\varepsilon}\right\} \tag{7}$$

where k_0 is the wave number in free space.

Power absorption coefficient

The power absorption coefficient κ_a is defined as

$$\kappa_a = 2\alpha \tag{8}$$

and is often expressed in dB/m through the relation

$$\kappa_a(\mathrm{dB/m}) = 8.686\,\alpha\,(\mathrm{Np/m}) \tag{9}$$

Extinction and scattering coefficient

The total electromagnetic loss in a medium consists of absorption loss (electromagnetic power transformed into other forms of energy, such as heat) and scattering loss (energy is caused to travel in directions other than that of the incident radiation). Scattering loss is caused by particles of different ε embedded in a host medium. The extinction coefficient (total loss) is thus

$$\kappa_e = \kappa_a + \kappa_s \tag{10}$$

where κ_s denotes the scattering loss.

Penetration depth

Part of a wave incident upon the surface of a medium from the air in the direction z is transmitted across the boundary into the medium. The penetration depth δ_p is defined as the depth z at which (Ulaby et al., 1986)

$$\frac{P(z = \delta_p)}{P(z = 0+)} = \frac{1}{e} \tag{11}$$

where $P(\delta_p)$ is the transmitted power at depth δ_p and $P(0+)$ is the transmitted power just beneath the surface. If scattering in the medium is ignored, $\kappa_e \equiv \kappa_a = 2\alpha$ and if, additionally, α is constant as a function of z,

$$\delta_p = \frac{1}{\kappa_a} = \frac{\sqrt{\varepsilon'}}{k_0 \varepsilon''}; \quad \varepsilon'' << \varepsilon' \tag{12}$$

The penetration depth provides an approximate value to the maximum depth of the medium that contributes to the backscattering coefficient and brightness temperature.

Optical depth

The optical depth between two points r_1 and r_2 in a medium provides a measure of how transparent the medium is to electromagnetic power passing through it. It is defined as

$$\tau(r_1, r_2) = \int_{r_1}^{r_2} \kappa_e dr \tag{13}$$

The optical depth is often used in atmospheric remote sensing.

Dielectric properties of natural media

The first book on the dielectric properties of materials was published by Debye, who presented his theory on polar liquids in 1929. Von Hippel published, in 1954, his book on theoretical and experimental aspects of dielectric media. The two-volume analysis of electric polarization was published by Böttcher (1973) and Böttcher et al. (1978). Followed by progress in microwave remote sensing in the 1970s, both theoretical and experimental studies of the dielectric properties of natural media increased rapidly. Electromagnetic mixing theories were analyzed by Sihvola (1999).

Electromagnetically, pure water is homogeneous matter, because its characteristics are not a function of the position within water. It is also isotropic, meaning that its characteristics are the same in all directions. However, most natural media are either nonhomogeneous or anisotropic; they may be mixtures of various substances, possibly with certain geometry. For example, soils consist of soil solids particles, air voids, and water. Seawater contains a variety of dissolved salts, making its dielectric behavior different from that of pure water. An example of anisotropic media is sea ice, which consists of ice, air, and mostly vertical brine inclusions.

In microwave remote sensing, interaction of electromagnetic waves with matter is modeled using the macroscopic dielectric properties of targets. Due to the

complex structure of many natural media, their dielectric characteristics cannot be derived theoretically with a sufficiently high degree of accuracy. Hence, the dielectric models for most media are semiempirical, relying on both theoretical studies and measurements made under laboratory conditions. In the following, the dielectric properties of natural media essential in microwave remote sensing are discussed. Various mixing models for computing the bulk dielectric properties of natural media are presented.

Most natural media contain water. Since the complex permittivity of water is substantially larger than that of other natural media at frequencies relevant to microwave radiometer and radar remote sensing, the dielectric properties of water tend to dominate those of other media including soils, vegetation, snow, and ice. The volumetric water content in soils, especially clay, may be up to 50 %, in vegetation parts up to 70 %, and in snow over 10 %. The permittivity of most natural dry materials is small ($\varepsilon' < 5$ and $\varepsilon'' < 0.1$). The dielectric properties of pure water, freshwater, and seawater are discussed first in this presentation, followed by those of other natural media important for microwave remote sensing.

Dielectric properties of pure water, freshwater, and seawater

Pure Water. For polar liquids like water, the complex permittivity can be derived by assuming that a polar molecule interacts with an electromagnetic field by rotating in a viscous medium. This results in the so-called Debye equation for water (Debye, 1929):

$$\varepsilon_w = \varepsilon_{w\infty} + \frac{\varepsilon_{w0} - \varepsilon_{w\infty}}{1 + j2\pi f \tau_w} \tag{14}$$

where ε_{w0} is relative permittivity of water in a static electric field, $\varepsilon_{w\infty}$ is relative permittivity of water at an "infinitely high" frequency, τ_w is relaxation time of water, and f is frequency. ε'_w and ε''_w depend, in addition to frequency, on temperature. The real and imaginary part of permittivity can be solved from (Equation 14):

$$\varepsilon'_w = \varepsilon_{w\infty} + \frac{\varepsilon_{w0} - \varepsilon_{w\infty}}{1 + (2\pi f \tau_w)^2} \tag{15a}$$

$$\varepsilon'_w = \frac{2\pi f \tau_w (\varepsilon_{w0} - \varepsilon_{w\infty})}{1 + (2\pi f \tau w)^2} \tag{15b}$$

Numerical expressions for the terms of (Equation 14) as a function of temperature and salinity are available in Klein and Swift (1977); these equations have been regularly used for computing the complex permittivity of seawater. The frequency at which the maximum dielectric loss occurs is called the relaxation frequency; it is given by

$$f_0 = \frac{1}{2\pi\tau_w} \tag{16}$$

It has been recently concluded that the model based on the Debye relaxation process (Equation 14) is accurate to within less than 1 % up to 30 GHz at temperatures around 25 °C, but differences between the extrapolated value and independently measured permittivity values are about 10 % at 90 GHz (Ellison, 2006).

There are two approaches to increase accuracy at frequencies beyond 30 GHz and at low temperatures, the double Debye model of Liebe et al. (1991) and the extended double Debye model of Ellison (2006). The double Debye model is basically the same as that given in (Equation 14), but it includes two relaxation processes instead of one (Liebe et al., 1991):

$$\varepsilon_w = \frac{\varepsilon_{w01} - \varepsilon_{w\infty 1}}{1 + j\left(\frac{f}{f_{01}}\right)} + \frac{\varepsilon_{w\infty 1} - \varepsilon_{w\infty 2}}{1 + j\left(\frac{f}{f_{02}}\right)} + \varepsilon_{w\infty 2} \tag{17}$$

where ε_{w01} is the static permittivity, f_{01} and f_{02} are the first and second relaxation frequency, respectively, and $\varepsilon_{w\infty 1}$ and $\varepsilon_{w\infty 2}$ are the first and second high-frequency permittivity, respectively, and

$$\theta = 1 - \frac{300}{T(K)} \tag{18}$$

$$\varepsilon_{w01} = 77.66 - 103.3\theta \tag{19}$$

$$\varepsilon_{w\infty 1} = 0.0671\varepsilon_{w01} \tag{20}$$

$$f_{01} = 20.20 + 146.4\theta + 316\theta^2 \tag{21}$$

$$\varepsilon_{w\infty 2} = 3.52 + 7.52\theta \tag{22}$$

$$f_{02} = 39.8 f_{01} \tag{23}$$

where T(K) is the temperature in Kelvins.

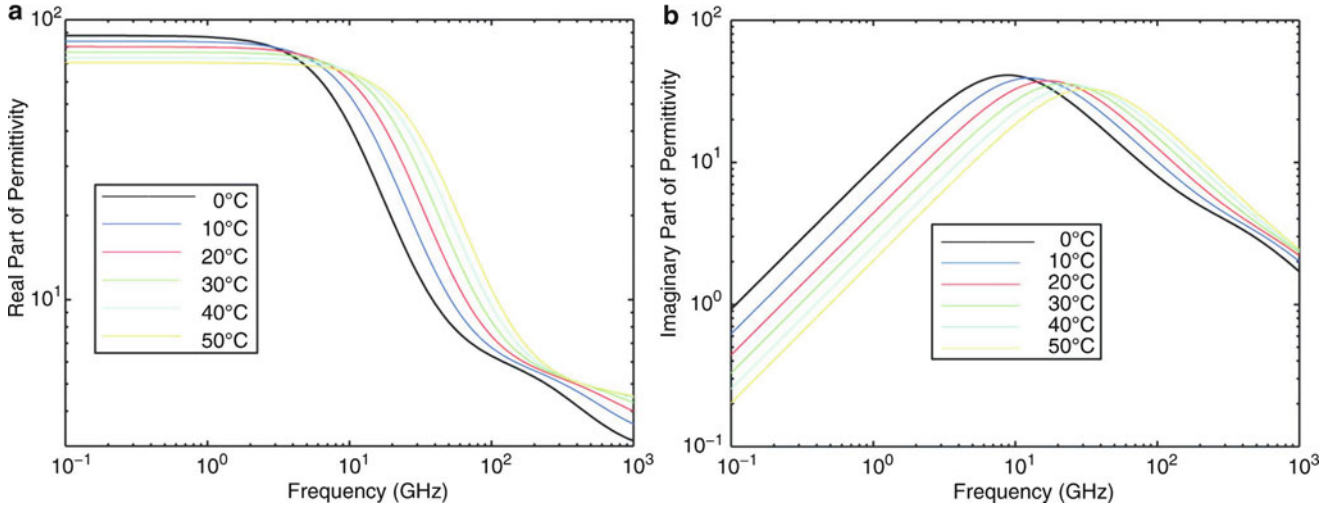
Based on the double Debye model in (Equations 17–23), Figure 1 shows the complex permittivity of pure water over the frequency range of 100 MHz–1,000 GHz for temperatures between 10 °C and 50 °C.

Freshwater. There are relatively few dissolved ions in freshwater, and hence, ε' for freshwater is practically the same as that for pure water. However, the loss term must be corrected by adding a conductivity term:

$$\varepsilon''_{iw} = \varepsilon''_w + \frac{\sigma}{2\pi f \varepsilon_0} \tag{24}$$

where subscript *i* refers to freshwater (impure water) and σ is the conductivity. Due to the frequency term in the denominator of (Equation 24), the effect of conductivity increases with decreasing frequency.

Pure Water and Seawater. Seawater contains dissolved salts, including NaCl, $MgCl_2$, Na_2SO_4, $CaCl_2$, and KCl, and the average salinity in the oceans is 32.5 psu (Anderson, 1960). Dissolved salts in seawater increase the ionic conductivity. Consequently, the dielectric loss factor of seawater is substantially higher than that of pure water at low microwave frequencies.



Microwave Dielectric Properties of Materials, Figure 1 Complex dielectric permittivity of pure water according to Liebe et al. (1991).

For this reason, L-band is optimal for retrieval of ocean salinity from radiometer data.

Recently, the dielectric behavior of water has been approximated using an extended double Debye model up to 25 THz for pure water and up to 105 GHz for seawater (Ellison, 2006). His results are based on several data sets and indicate that there are three relaxation processes in the microwave region and two resonances in the far-infrared region. At 25 °C, the relaxation frequencies are 18.56 GHz, 167.83 GHz, and 1.944 THz; the two resonance frequencies are 4.03 and 14.48 THz. Freshwater and seawater contain dissolved ions and gases; hence, their dielectric behavior is different from that of pure water. Using the extended double Debye model, the complex permittivity is then expressed (Ellison, 2006) as

$$\varepsilon_w = \varepsilon_{w\infty} + \frac{\varepsilon_{w0} - \varepsilon_{w1}}{1 + j2\pi f \tau_{w1}} + \frac{\varepsilon_{w1} - \varepsilon_{w\infty}}{1 + j2\pi f \tau_{w2}} - j\frac{\sigma_w}{2\pi f \varepsilon_0} \tag{25}$$

The parameters ε_{w0}, ε_{w1}, $\varepsilon_{w\infty}$, τ_{w1}, τ_{w2}, and σ_w depend on temperature and salinity of water. Their highly complex numerical expressions are given in (Ellison, 2006, pp. 445–454):

$$\varepsilon_s(T,S) = 87.85306 \exp\left(-0.00456992T - a_1 S - a_2 s^2 - a_3 ST\right) \tag{26}$$

$$\varepsilon_1(T,S) = a_4 \exp(-a_5 T - a_6 S - a_7 ST) \tag{27}$$

$$\tau_1(T,S) = (a_8 + a_9 S)\exp\left(\frac{a_{10}}{T + a_{11}}\right) \tag{28}$$

$$\tau_2(T,S) = (a_{12} + a_{13} S)\exp\left(\frac{a_{14}}{T + a_{15}}\right) \tag{29}$$

$$\varepsilon_\infty(T,S) = a_{16} + a_{17} T + a_{18} S \tag{30}$$

$$\sigma(T,S) = \sigma(T,35)\cdot P(S)\cdot Q(T,S) \tag{31}$$

where

$$\sigma(T,35) = 2.903602 + 8.607\cdot 10^{-2} T + 4.738817\cdot 10^{-4} T \tag{32}$$

$$- 2.991\cdot 10^{-6} T^3 + 4.3041\cdot 10^{-9} T^4 \tag{33}$$

$$P(S) = S\frac{37.5109 + 5.45216\ S + 0.014409\, S^2}{1,004.75 + 182.283\ S + S^2} \tag{34}$$

and

$$Q(T,S) = 1 + \frac{\alpha_0 (T - 15)}{T + \alpha_1} \tag{35}$$

where

$$\alpha_0 = \frac{6.9431 + 3.2841 S - 0.099486\, S^2}{84.85 + 69.024\ S + S^2} \tag{36}$$

$$\alpha_1 = 49.843 - 0.2276\, S + 0.00198\, S^2 \tag{37}$$

Equations 25–37 with Table 1 are stated to represent the permittivity of pure water to within 1 % over the frequency range 0–20 GHz, to within 3 % over the frequency range 30–100 GHz, and to within 5 % over the frequency range 100–1,000 GHz (Ellison, 2006). The permittivity of seawater is represented to within 3 % over the frequency

Microwave Dielectric Properties of Materials, Table 1 Coefficients for Equations 26–30

$a_1 = 0.46606917$ E (−2)	$a_7 = 0.34414691$ E (−4)	$a_{13} = 0.38957681$ E (−6)
$a_2 = -0.26087876$ E (−4)	$a_8 = 0.17667420$ E (−3)	$a_{14} = 0.30742330$ E (3)
$a_3 = -0.63926782$ E (−5)	$a_9 = -0.20491560$ E (−6)	$a_{15} = 0.12634992$ E (3)
$a_4 = 0.63000075$ E (1)	$a_{10} = 0.58366888$ E (3)	$a_{16} = 0.37245044$ E (1)
$a_5 = 0.26242021$ E (−2)	$a_{11} = 0.12634992$ E (3)	$a_{17} = 0.92609781$ E (−2)
$a_6 = -0.42984155$ E (−2)	$a_{12} = 0.69227972$ E (−4)	$a_{18} = -0.26093754$ E (−1)

range 3–105 GHz within temperatures of 0–30 °C and salinities of 0–40 psu. Below 3 GHz, the results of Klein and Swift (1977) can be used.

Dielectric properties of snow

Dry Snow. Dry snow consists of ice crystals and air voids. Any water in the snow medium collects at points of contact between the ice grains. The metamorphism caused by melting and freezing changes the microstructure of snow. The ice grains become rounded during the melting process and some of the smaller grains disappear completely. Snow that has undergone several melt–freeze cycles tends to form multiple clusters. In general, the density of snow slowly increases with time due to metamorphism and melt–freeze cycles. The density of dry snow varies from 0.1 g/cm^3 (newly fallen snow) to 0.5 g/cm^3 (refrozen snow).

Since dry snow is a dielectric mixture of ice and air, its complex permittivity is governed by the dielectric properties of ice, snow density, and ice-particle shape. Since the real part of the permittivity of ice is $\varepsilon'_i = 3.17$ at frequencies between 10 MHz and 1,000 GHz (Mätzler and Wegmüller, 1987) and practically independent of temperature, the dielectric constant of dry snow, ε'_{ds} is only a function of density. For the most extensive data set used in dielectric measurements of snow, the result between 3 and 37 GHz is (Hallikainen et al., 1986)

$$\varepsilon'_{ds} = 1 + 1.9\,\rho_{ds}; \quad \rho_{ds} \le 0.5 \text{ g/cm}^3 \tag{38}$$

where ρ_{ds} is the density of dry snow in g/cm^3. From (Equation 38), $\varepsilon'_{ds} = 1.57$ for a density of 0.3 g/cm^3. A piecewise linear relationship between ε'_{ds} and ρ_{ds} has been reported based on measurements at 1 GHz and related modeling (Mätzler, 1996).

Experimental data for the loss factor of dry snow are limited to frequencies below 13 GHz (Cumming, 1952; Tiuri et al., 1984). The best fit to the experimental data of Tiuri et al. was provided by the following expression:

$$\varepsilon''_{ds} = 1.59 \cdot 10^6 (0.52\rho_{ds} + 0.62\rho_{ds}^2) \cdot \left(f^{-1} + 1.23 \cdot 10^{-14}\sqrt{f}\right) \exp(0.036T) \tag{39}$$

where T is the temperature in °C and f is the frequency in Hz. An alternative equation has been provided by Ulaby et al., (1986) using a two-phase dielectric mixing formula for spherical inclusions. Both approaches indicate that the dielectric loss factor of dry snow has a minimum value between 1 and 5 GHz and then increases with increasing frequency. It is of the order of 10^{-4}–10^{-3} at all frequencies below 100 GHz; hence, dry snow is a loss–loss medium.

Due to its granular structure, dry snow is a strongly scattering medium at high microwave frequencies. Based on theoretical calculations using the Mie theory (assuming independent scattering), scattering dominates over absorption in dry snow (ice grain size 1 mm) at frequencies above 15 GHz (Ulaby et al., 1986). Scattering loss increases rapidly with increasing frequency; hence, the extinction coefficient at and above 35 GHz consists mainly of scattering loss. Based on an extensive data set of snow samples ranging from newly fallen to refrozen snow, with the ice grain diameter ranging from 0.2 to 1.6 mm and snow density from 0.17 to 0.39 g/cm^3, the following empirical expressions have been developed to relate the extinction coefficient of dry snow, denoted κ_{eds}, to the observed snow particle diameter d_s (Hallikainen et al., 1987):

$$\kappa_{eds} = 1.5 + 7.4\,d_s^{2.3} \quad \text{dB/m at 18 } GHz \tag{40}$$

$$\kappa_{eds} = 30\,d_s^{2.1} \text{ dB/m at 35 GHz} \tag{41}$$

$$\kappa_{eds} = 180\,d_s^{2.0} \quad \text{dB/m at 60 GHz} \tag{42}$$

$$\kappa_{ds} = 300\,d_s^{1.9} \text{ dB/m at 90 GHz} \tag{43}$$

The particle diameter (observed by photography) is in millimeters. Equations 40–43 hold for particle sizes below 1.6 mm. Equations 40–42 can be combined into a single equation of the form

$$k_{eds} = 0.0018\,f^{2.8} d_s^{2.0} \text{ dB/m} \quad 18 \text{ GHz} < f < 60 \text{ GHz} \tag{44}$$

where f is in GHz and d_s is in mm. Due to the exponent value of 2.8 in (Equation 44), the extinction coefficient increases rapidly with increasing frequency; at 35 GHz it is about 20 dB/m for snow with a grain size of 1 mm, whereas at 90 GHz it is well over 100 dB/m. This is why radiometers operating at 35 GHz are able to measure snow water equivalents less than 100 mm and those operating at 90 GHz "see" only the snowpack surface. The strong fluctuation theory (Stogryn, 1986) provides results that agree reasonably well with the empirical extinction coefficient values in the 18–60 GHz range for all realistic grain sizes and also at 90 GHz for grain sizes smaller than 0.9 mm.

Wet Snow. Wet snow is a mixture of ice crystals, liquid water, and air. The geometry and porosity of wet snow depend on its liquid water content. It has been concluded that snow has two distinct regimes of liquid saturation (Colbeck, 1982). In the lower range (pendular regime), air is continuous throughout the pore space, and liquid water occurs in the form of isolated inclusions. In the higher range of liquid saturation (funicular regime), liquid water is continuous throughout the pore space, and air occurs as distinct bubbles trapped by narrow constrictions in the pores. There is a sharp transition between the two regimes.

Electromagnetically, wet snow is a three-component dielectric mixture consisting of ice particles, air, and liquid water. Both water and ice exhibit Debye-type relaxation spectra. The relaxation frequency (frequency at which the maximum dielectric loss occurs) of ice is in the kilohertz range, whereas that for water at 0 °C is 9 GHz. The complex permittivity of ice and water depends on frequency and temperature. Consequently, the permittivity of wet snow is a function of frequency, temperature, volumetric water content, snow density, ice-particle shape, and the shape of water inclusions.

Since the permittivity of water is substantially higher than that of ice and air, the dielectric behavior of wet snow is governed by the volume fraction of water. Basically, ε'_{ws} and ε''_{ws} are compressed versions of those for water at 0 °C. The dielectric behavior of wet snow was measured by Hallikainen et al. (1986) at nine frequencies between 3 and 18 GHz and, additionally, at 37 GHz. They tested four dielectric models: Debye-like model, modified Debye-like model, two-phase Polder–Van Santen model (Polder and Van Santen 1946), and three-phase Polder–Van Santen model. The two-phase Polder–Van Santen mixing formula is applied by assuming wet snow to consist of dry snow as the host material with water inclusions embedded in it. The shape of the water inclusions was included in the modeling study. Tiuri et al. (1984) used a resonator in the 0.5–1 GHz range and also the 4 GHz data set of Hallikainen et al. (1982) and developed a numerical model for the permittivity of wet snow. Mätzler et al. (1984) made resonator measurements at 1 GHz and applied their results for spectral modeling of wet snow. See Hallikainen et al. (1986) for a review of dielectric wet snow studies.

The samples of Hallikainen et al. (1986) had densities ranging from 0.09 to 0.42 g/cm^3 and liquid water contents ranging from 0 % to 12.3 % by volume. The snow particle size varied between 0.5 and 1.5 mm and the sample temperatures ranged from −5 °C to 0 °C. The two-phase Polder–Van Santen model for wet snow is

$$\varepsilon_{ws} = \varepsilon_{ds} + \frac{m_v \varepsilon_{ws}}{3}(\varepsilon_w - \varepsilon_{ds}) \cdot \sum_{j=1}^{3} \left[\varepsilon_{ws} + (\varepsilon_w - \varepsilon_{ws}) A_{wj}\right]^{-1} \tag{45}$$

where m_v is the volumetric liquid water content and water droplets are randomly distributed and randomly oriented ellipsoids with depolarization factors A_{w1}, A_{w2}, and A_{w3}. The general Debye-like model is expressed as

$$\varepsilon'_{ws} = A + \frac{B m_v^x}{1 + (f/f_0)^2} \tag{46a}$$

$$\varepsilon''_{ws} = \frac{C(f/f_0) m_v^x}{1 + (f/f_0)^2} \tag{46b}$$

where f_0 is the relaxation frequency (9.07 GHz) and A, B, C, and x are constants determined by fitting the model to the measured data in the 3–37 GHz range. The obtained values are given in Hallikainen et al. (1986). In the simplified Debye-like model, A, B, and C depend only on dry snow density and volumetric water content, whereas in the modified Debye-like model, they depend, additionally, on frequency. The Debye-like equations do not take into account the geometry of the wet snow medium (shape of water inclusions, funicular and pendular regimes).

Liquid water inclusions in snow are approximately needle shaped in the pendular regime (low values of water content), but they become approximately disk shaped in the funicular regime (high values of water content) (Colbeck, 1982). The best fit of the developed two-phase Polder–Van Santen mixing model to experimental data was obtained by assuming that their shapes depend on snow water content and, additionally, that the water inclusions are nonsymmetrical in shape (Hallikainen et al., 1986). Transition from the pendular regime to the funicular regime was observed to take place around snow volumetric water content $m_v = 3$ %.

Both the modified Debye-like model and the two-phase Polder–Van Santen model (with variable, nonsymmetrical shape factors) provide good accuracy (Hallikainen et al., 1986). The simplified Debye-like model works well only at frequencies below 15 GHz, but the modified Debye-like model works well also at higher frequencies. The two-phase Polder–Van Santen model with wetness-dependent shape factors provides a slightly better overall fit to the experimental observations of wet snow. The modified Debye-like model is easier to use and is likely to provide adequate accuracy for most applications. The three-phase Polder–Van Santen model provided accuracies comparable to those of the two-phase Polder–Van Santen model.

The real part of the dielectric constant of wet snow varies from the values for dry snow up to 3.0 at 3 GHz and to 1.9 at 37 GHz for a liquid water content of 12 % and, generally, decreases with increasing frequency. Its dielectric loss factor increases with frequency until it reaches its maximum value of 0.9 at 9 GHz for a wetness of 12 %. At frequencies above 9 GHz, it decreases monotonically. Even a volumetric water content of 2 % results in absorption of tens of decibels per meter, depending on frequency. The penetration depth for wet snow is generally less than 20 cm at frequencies above 5 GHz, suggesting that microwave radar and radiometer measurements can provide information only on the top of the snowpack.

Dielectric properties of ice

Pure and Freshwater Ice. Freshwater ice is free from salt, but usually it includes air bubbles, impurities, and cracks. Pure ice is an idealization of freshwater ice (frozen distilled water); it is not discussed here. The dielectric properties of freshwater ice at microwave frequencies are determined by two physical processes, namely, the high-frequency tail of a relaxation spectrum (relaxation frequency in the kilohertz range) and the low-frequency tail of far-infrared absorption bands (Evans, 1965; Mätzler and Wegmüller, 1987). These processes result in a practically constant value for the dielectric constant ε'_i at microwave frequencies and in a minimum of the dielectric loss factor ε''_i centered around 1–5 GHz.

Based on experimental data, the relative dielectric constant of pure and impure ice may be assigned the constant value of

$$\varepsilon'_i = 3.17 \tag{47}$$

for frequencies between 10 MHz and 100 GHz (Mätzler and Wegmüller, 1987; Cumming, 1952).

For the imaginary part of the permittivity, results of various investigators show substantial scatter. The scatter is caused, in addition to measurement inaccuracies (partly due to small values of ε''_i), by varying amounts of impurities in the ice samples. Reviews of results are available, for example, in Ulaby et al. (1986). A numerical equation for ε''_i was developed by Mätzler and Wegmüller (1987):

$$\varepsilon''_i = \frac{A}{f} + Bf^C \tag{48}$$

where f is the frequency in GHz. The values of constants A, B, and C were determined for pure and impure ice at temperatures −5 °C and −15 °C. The values for impure ice at −5 °C are $A = 0.0026$, $B = 0.00023$, and $C = 0.87$. For pure ice, they are $A = 0.0006$, $B = 0.000065$, and $C = 1.07$. For −15 °C, they are for impure ice $A = 0.0013$, $B = 0.00012$, and $C = 1.0$. For pure ice, they are $A = 0.00035$, $B = 0.000036$, and $C = 1.2$. Based on (Equation 41), ε''_i for impure ice is between 0.001 and 0.01 in the 1–100 GHz range and slightly smaller for pure ice. Thus, freshwater ice is a low-loss medium, and its power absorption coefficient at 1 GHz (temperature −5 °C) is about 0.1 dB/m and penetration depth is 100 m. At 10 GHz, the corresponding numbers are about 1 dB/m and 10 m. This means that contributions to the brightness temperature and the backscattering coefficient are not limited to the topmost ice layers.

Sea Ice. Sea ice consists of freshwater ice, liquid brine, and air. In addition to the dielectric properties of these constituents, several additional parameters influence its complex permittivity, including the volume fraction of each constituent and geometry (shape, size, and orientation) of brine pockets with respect to the propagation direction of the electromagnetic wave. Most of these parameters depend on ice temperature. The salinity of liquid brine in sea ice is governed by its temperature, and it is over 200 ‰ by weight at −20 °C and decreases to about 35 ‰ at −2 °C (Assur, 1960). The relative volume of liquid brine is directly proportional to the salinity; for sea ice with a salinity of 1 ‰, it is about 3 ‰ at −20 °C, 24 ‰ at −2 °C, and 103 ‰ at −0.5 °C (Frankenstein and Garner, 1967). Due to the high volumetric content and salinity, the complex permittivity of brine is substantially higher than that of seawater (Stogryn and Desargeant, 1985). Hence, brine characteristics dominate the permittivity of sea ice.

Experimental investigations of the microwave dielectric behavior of sea ice have been carried out at frequencies below 40 GHz. Reviews of these investigations are given in Vant (1976) and Hallikainen and Winebrenner (1992). The total number of permittivity measurements for arctic ice is small. Additionally, most of the investigations are limited to first-year ice. Frequency-wise, extensive sets of experimental Arctic sea ice data were acquired by Vant (1976), Vant et al. (1978) (0.1–7.5 GHz; 13 samples), and Sackinger and Byrd (1972) (26–40 GHz; sample number not reported). Sample-wise, the most extensive studies are those of Bogorodskii and Khokhlov (1975) (10 GHz; 130 samples) and Hallikainen (1983) (0.6 and 0.9 GHz; 373 samples). References Vant (1976) and Vant et al. (1978) report the same results, but the former includes detailed results for each ice sample. Hoekstra and Cappillino (1971) (0.1–24 GHz) measured ice samples of various salinities, made from seawater flash frozen in a sample holder; their results have very likely been influenced by sample preparation technique.

In most studies, temperatures below −5 °C were used, although rapid changes in the permittivity occur at temperatures close to 0 °C due to increasing relative volume of brine. Hallikainen (1983) measured low-salinity ($S < 2$ ‰) ice samples at temperatures up to −0.1 °C. In order to avoid brine drainage, each ice sample was measured immediately after removing it from the ice field and using only the original temperature. The relative brine volume of low-salinity sea ice near 0 °C may be higher than that of high-salinity sea ice at lower temperatures. The relative brine volume of sea ice of salinity 1 ‰ is 103 ‰ at −0.5 °C, whereas the corresponding number for sea ice of salinity 4 ‰ is 40 ‰ at −5 °C.

A limited amount of data is available on the dielectric properties of multiyear sea ice (Vant, 1976, two samples; Vant et al., 1978, two samples). Due to lower salinity, the complex permittivity of multiyear ice is lower than that of first-year ice.

Only Vant et al. (1978) (NaCl ice; orientation from vertical 0, 30, 45, 60, and 90°), Sackinger and Byrd (1972) (Arctic first-year ice; 0 and 90°), and Hallikainen (1983) (low-salinity first-year ice; 0°, 30°, and 90°) have investigated the effect of sample orientation with respect to the propagation direction of the electromagnetic wave. Sackinger and Byrd (1972) observed that the dielectric loss factor at −7 °C may be up to 300 % higher for 90° (horizontal orientation) than for 0° (vertical); it decreases

with decreasing temperature and is about 150 % higher at −21.5 °C. Hence, the dielectric loss of sea ice increases with increasing incidence angle. Based on their data, Vant et al. (1978) developed a model to account for sample orientation. The permittivity also depends on ice type; in general, the dielectric loss is higher for frazil than for columnar sea ice with the electromagnetic wave propagating in the vertical direction.

Comparisons between various data sets have shown that there is substantial scatter for values of both the real part and imaginary part of the permittivity for mostly used frequencies like 1 GHz, 5 GHz, and 10 GHz (Hallikainen and Winebrenner, 1992). Efforts to model the dielectric behavior of sea ice are scarce. Vant et al. (1978) used two approaches based on the relative brine volume and a semiempirical approach to account for the orientation. Stogryn (1987) employed strong fluctuation theory to compute the tensor dielectric constant for sea ice. Hallikainen and Winebrenner (1992) generated numerical equations, based on available data sets, to relate the dielectric properties of sea ice to the relative brine volume V_b at three frequencies – 1 GHz, 4 GHz, and 10 GHz. Their expressions are given below in a slightly modified form, due to revisiting the data sets.

$$\varepsilon'_{si} = 3.12 + 0.009\,V_b; \quad 1\,\text{GHz} \tag{49a}$$

$$\varepsilon''_{si} = 0.04 + 0.005\,V_b; \quad 1\,\text{GHz} \tag{49b}$$

$$\varepsilon'_{si} = 3.05 + 0.007\,V_b; \quad 4\,\text{GHz} \tag{50a}$$

$$\varepsilon''_{si} = 0.02 + 0.0033\,V_b; \quad 4\,\text{GHz} \tag{50b}$$

$$\varepsilon'_{si} = 3.04 + 0.004V_b; \quad 10\,\text{GHz} \tag{51a}$$

$$\varepsilon''_{si} = 0.01 + 0.007V_b; \quad 10\,\text{GHz} \tag{51b}$$

where the relative brine volume $V_b \leq 70$ ‰ and is given by (Frankenstein and Garner, 1967):

$$V_b = S\left(-\frac{52.56}{T} - 2.28\right) \quad -0.5^\circ C \geq T \geq -2.06^\circ C \tag{52a}$$

$$V_b = S\left(-\frac{45.917}{T} + 0.930\right) \quad -2.06^\circ C > T \geq -8.2^\circ C \tag{52b}$$

$$V_b = S\left(-\frac{43.795}{T} + 1.189\right) \quad -8.2^\circ C > T \geq -22.9^\circ C \tag{52c}$$

Equations 49a, b–51a, b do not explicitly take into account the effect of ice density upon the real part of the permittivity or direction of wave propagation in the ice medium. The correlation coefficient for Equations 49a, b–51a, b is, in general, between 0.7 and 0.8.

In general, the absorption coefficient κ_a increases with increasing frequency and temperature. The loss of multiyear ice is considerably lower than that of first-year ice, mainly due to much lower salinity. The values for first-year low-salinity sea ice show that the temperature dependence of the absorption coefficient is high at temperatures near 0 °C. At 10 GHz, the differences between the results for columnar, frazil, and multiyear ice are substantial.

In the 1–10 GHz range, the penetration depth is between 100 and 5 cm for first-year ice and 500 and 30 cm for multiyear ice. Sensors operating at X-band provide information on sea ice mainly from the topmost 5–80 cm, depending on ice type, salinity, and temperature, whereas the corresponding numbers for L-band sensors are 40–500 cm.

Dielectric properties of soils

Soils consist of bulk soil, air, and water. A soil's textural composition is usually given in terms of weight percentages of sand, silt, and clay. Sand includes particles with diameters in the range between 0.05 and 2.0 mm, silt includes particles with diameters in the 0.002–0.05 mm range, and clay includes particles with diameters smaller than 0.002 mm. The water contained in the soil is usually divided into bound water and free water, although the transition between these is not sharp. Molecules of bound water are contained in the first few molecular layers surrounding the soil particles; these are tightly held by the soil particles due to the matric and osmotic forces (Baver et al., 1977). The amount of bound water is proportional to the total surface area of the soil particles, which depends on the soil particle size distribution and mineralogy: The smaller the particles, the more there can be bound water. Hence, clay has more bound water than sand. Water molecules located further away from the soil particle surface can move within the soil medium and are referred to as free water.

From the electromagnetic point of view, soil is a heterogeneous medium consisting of bulk soil, air, bound water, and free water. Its dielectric properties as a function of frequency depend on the soil bulk density (compaction), soil composition (particle size distribution and mineralogy), the volume fractions of bound and free water, the salinity of the soil solution, and temperature (Dobson et al., 1985). Wang and Schmugge (1980) showed that various dielectric mixing formulas describing soil as a two-component system (soil with free water) fail to describe the complex permittivity of wet soil realistically at 1.4 and 5 GHz. They were the first to explicitly include both free water and bound water in their empirical mixing formula by examining two moisture regions: (a) water contents less than the maximum bound water fraction and (b) water contents higher than the bound water fraction. They included two free parameters in the model and, by optimizing their values, were able to explain the behavior of the complex permittivity, especially its real part. Wang (1980) modeled the soil–water system with

a Debye-like relaxation over a finite band of relaxation frequencies and, by adjusting two free parameters, adequately predicted the behavior of data over the 0.3–1.4 GHz range. The two free parameters were the width of the activation energy of the soil solution and the mean relaxation frequency of the soil–water mixture at a given frequency and for a given soil and water content.

Hallikainen et al. (1985) performed dielectric measurements of five soil types at 12 frequencies between 1.4 and 18 GHz. Soil moisture values from 0 % to the highest moisture contents that can be supported by that soil type without drainage taking place were used; additionally, data for frozen soils were acquired as well. They developed for each frequency polynomial expressions, separately for real and imaginary parts of the permittivity, dependent on volumetric moisture content and the percentage of sand and clay contained in the soil. Dobson et al. (1985) developed two dielectric models for wet soil based on the data set of Hallikainen et al. (1985), a theoretical model and a semiempirical model. The theoretical model accounts explicitly for the presence of bound water adjacent to hydrophilic soil particle surfaces and employs a four-component dielectric mixing model that describes the soil–water system consisting of dry soil solids as a host medium with randomly distributed and randomly oriented disc-shaped inclusions of bound water, bulk water, and air. The bulk water component characteristics in the model depend on frequency, temperature, and salinity. Based on comparisons with data, the theoretical model was determined to be an appropriate formulation, and it yields values that describe well the observed effects of frequency and soil type. However, its accuracy is limited by the uncertainty concerning the dielectric properties of bound water.

The driving force in the development of the semiempirical model of Dobson et al. (1985) was to develop a user-friendly, frequency-dependent model that is based on readily measured soil characteristics including volumetric moisture and weight fractions of sand and clay. The final expressions for the real and imaginary parts of the permittivity are

$$\varepsilon'_{ws} = \left[1 + \frac{\rho_b}{\rho_s}(\varepsilon_s^{\alpha} - 1) + m_v^{\beta'}\varepsilon_{fw}'^{\alpha} - m_v\right]^{\frac{1}{\alpha}} \tag{53a}$$

$$\varepsilon''_{ws} = \left[m_v^{\beta''}\varepsilon_{fw}''^{\alpha}\right]^{\frac{1}{\alpha}} \tag{53b}$$

where α is a constant shape factor, ρ_b is soil bulk density, $\rho_s = 2.66$ g/cm^3 is soil specific density, ε_s is the soil solid permittivity, m_v is the volumetric moisture, β' and β'' are the soil-texture-dependent coefficients for computing the real and imaginary part, respectively, and ε_{fw} is the permittivity of free water. The optimum value for the shape factor was determined to be $\alpha = 0.65$. The relative permittivity of soil solids is

$$\varepsilon_s = (1.01 + 0.44\rho_s)^2 - 0.062 \tag{54}$$

and the soil-texture-dependent coefficients are

$$\beta_{\varepsilon'} = 0.01(127.48 - 0.519S - 0.152C) \tag{55a}$$

$$\beta_{\varepsilon''} = 0.01(1.33797 - 0.603S - 0.166C) \tag{55b}$$

where S and C are the percentage of sand and clay, respectively. The dielectric properties of free water are computed using an expression similar to that for saline water:

$$\varepsilon_w = \varepsilon_{w\infty} + \frac{\varepsilon_{w0} - \varepsilon_{w\infty}}{1 + j2\pi f\tau_w} - j\frac{\sigma_{eff}}{2\pi f\varepsilon_0}\cdot\frac{\rho_s - \rho_b}{\rho_s m_v} \tag{56}$$

where the effective conductivity is

$$\sigma_{eff} = -1.645 + 1.939\rho_b - 0.02013S + 0.01594C \tag{57a}$$

The above expressions were determined using the whole soil dielectric data set between 1.4 and 18 GHz. For the frequency range of 0.3–1.3 GHz, another expression for the effective conductivity was developed (Peplinski et al., 1995):

$$\sigma_{eff} = 0.0467 + 0.2204\rho_b - 0.4111S + 0.66144C \tag{57b}$$

For a reader interested in the dielectric properties of soils at a certain frequency between 1.4 GHz and 18 GHz, the polynomial expressions in Hallikainen et al. (1985) are useful. Mironov et al. (2004) developed a generalized refractive index mixing dielectric model for moist soils incorporating the dielectric characteristics of bound water.

The above studies show that the complex permittivity of wet soil is a compressed version of that for slightly saline water. In the 1.4–18 GHz range, the real part of the permittivity increases with increasing volumetric moisture and decreasing frequency, whereas the imaginary part increases with increasing moisture and increasing frequency. At any given moisture content and at all frequencies, the real part is roughly proportional to sand content and inversely proportional to clay content; however, this effect decreases with increasing frequency. The effect of soil texture on the imaginary part varies with frequency. At 1.4 GHz, the real part increases with increasing clay content for moisture levels above 0.2 cm^3/cm^3, whereas at 4–6 GHz, it is nearly independent of soil texture at all moisture levels and at frequencies of 8 GHz and above it decreases with increasing clay content. This behavior is obviously due to the ionic conductivity (strongest at low frequencies) and the relation of volume fraction of bound water to soil specific surface. At moisture levels close to saturation, ε may reach values up to 23-j3 for sand ($m_v = 0.35$) and 33-j9 for clay ($m_v = 0.50$) at 1.4 GHz. At 18 GHz, the corresponding values are 13-j8 and 18-j12, respectively. The penetration depth for wet soil decreases drastically with increasing moisture content and frequency;

consequently, retrieval of soil moisture from microwave radiometer and radar data works best at low frequencies. Even at 1.4 GHz, information only on surface moisture is obtained.

The permittivity of soils decreases drastically at temperatures below 0 °C (Hallikainen et al., 1985). Between −11 °C and −24 °C, both the real and imaginary parts of permittivity still depend on temperature, demonstrating that not all water is frozen. Recently, Mironov et al., (2010) conducted dielectric measurements on organic rich permafrost soil from 1 to 16 GHz and from −30 °C to +25 °C. They also extended the previously developed dielectric soil model of Mironov et al. (2004) to cover frozen soil.

Dielectric properties of vegetation

The characteristics of plants and trees vary substantially, including the size, density, moisture content, and geometry. Hence, the physical characteristics of vegetation material are not discussed in this presentation. For a detailed discussion on timber characteristics, the reader is referred to Dinwoodie (2000). Dielectric properties of wood are discussed in detail in Torgovnikov (1993); they have been primarily examined for industrial purposes, especially for microwave drying of timber at 2.45 GHz.

The dielectric properties of plants including trunks, stalks, and leaves for remote sensing have been measured and modeled by El-Rayes and Ulaby (1987), and Ulaby and El-Rayes (1987) in the 0.2–20 GHz range, and monitored in situ at L-band by McDonald et al. (1999). Mätzler (1994) measured and modeled various leaves at frequencies 1–94 GHz. Ulaby et al. (1987) measured the propagation constant for vegetation canopies with vertical stalks between 1.6 and 10.2 GHz. Dielectric spectroscopy of Scots pine at frequencies up to 1 GHz has been done by Tomppo et al. (2009).

Ulaby and El-Rayes (1987) modeled the complex permittivity of vegetation as a simple additive mixture of plant material, free water, and bound water due to their observation that bound water has a substantially lower relaxation frequency than free water. At 22 °C, their Debye–Cole dual-dispersion model for the permittivity of vegetation is

$$\varepsilon_v = \varepsilon_r + v_{fw}\left[4.9 + \frac{75.0}{1 + j\cdot 0.556f} - j\frac{18\sigma}{f}\right] + v_{bw}\left[2.9 + \frac{55.0}{1 + (j\cdot 0.556f)^{0.5}}\right] \quad (58)$$

where ε_r is the permittivity of bulk vegetation material, f is the frequency in GHz, σ is the ionic conductivity of the free-water solution in S/m, and v_{fw} and v_{bw} are the relative fractions of free and bound water, respectively. The volumetric moisture content consists of free water and bound water:

$$M_u = v_{fw} + v_{bw} = \frac{M_g\rho}{1 - M_g(1-\rho)} \quad (59)$$

where M_g denotes the gravimetric moisture content and ρ is the bulk density. If M_g and ρ are known, the following expressions can be used for obtaining higher accuracy:

$$\varepsilon_r = 1.7 + 3.2M_v + 6.5M_v^2 \quad (60)$$

$$v_{fw} = M_v(0.82M_v + 0.166) \quad (61)$$

$$v_{bw} = \frac{31.4M_v^2}{1 + 59.5M_v^2} \quad (62)$$

The Debye–Cole dual-dispersion model has been determined to provide an estimate of ε'_v for leaves, stalks, branches, and trunks with an accuracy of 5 %. If the salinity S is known in order to provide an estimate for the conductivity (S/m) via

$$\sigma = 0.16S - 0.0013S^2 \quad (63)$$

similar accuracies apply to ε''_v.

Summary

Dielectric properties of natural media at microwave frequencies have been reviewed, and the effects of various physical parameters to the behavior of their permittivity have been discussed. Since most natural media are heterogeneous, theoretical derivation of their dielectric properties is not straightforward. Measurements provide useful information for model development and evaluation, but only within the range of the covered physical characteristics and employed frequencies. The permittivity of pure and saline water is the key to modeling the behavior of most natural media; gaps still occur for its values at near-freezing temperatures, especially at high microwave frequencies. More experimental data are needed to confirm the behavior of wet snow at frequencies above 18 GHz. Measurements of the permittivity of sea ice for various salinities and ice types, especially for high brine volumes and at temperatures above −5 °C, would substantially benefit modeling of the behavior of sea ice. Permittivity measurements of various types of vegetation are needed to test presently available semiempirical models and further develop them.

Bibliography

Anderson, D. L., 1960. The physical constants of sea ice. *Research*, **13**, 310–318.

Assur, A., 1960. *Composition of Sea Ice and Its Tensile Strength*. Wilmette: U.S. Army Snow, Ice and Permafrost Research Establishment.

Baver, L. D., Gardner, W. H., and Gardner, W. R., 1977. *Soil Physics*. New York: Wiley.

Bogorodskii, V. V., and Khokhlov, G. P., 1975. Electrical properties of ice in the ice edge zone of the Bering Sea at 10 GHz frequency. In *Proceedings of the Final Symposium on the Results of the Joint Soviet-American Expedition*. Leningrad: Gidrometeoizdat, pp. 219–233 (in Russian).

Böttcher, C. J. F., 1973. *Theory of Electrical Polarisation*, 2nd edn. Amsterdam: Elsevier, Vol. 1.

Böttcher, C. J. F., Van Belle, O. C., Bordewijk, P., and Rip, A., 1978. *Theory of Electrical Polarisation*, 2nd edn. Amsterdam: Elsevier, Vol. 2.

Colbeck, S. C., 1982. The geometry and permittivity of snow at high frequencies. *Journal of Applied Physics*, **53**, 4495–4500.

Cumming, W., 1952. The dielectric properties of ice and snow at 3.2 cm. *Journal of Applied Physics*, **23**, 768–773.

Debye, P., 1929. *Polar Molecules*. New York: Dover.

Dinwoodie, J. M., 2000. *Timber, Its Nature and Behaviour*, 2nd edn. New York: Taylor & Francis.

Dobson, M. C., Ulaby, F. T., Hallikainen, M. T., and El-Rayes, M., 1985. Microwave dielectric behavior of wet soil – part II: dielectric mixing models. *IEEE Transactions on Geoscience and Remote Sensing*, **23**, 35–46.

Ellison, W., 2006. Microwave dielectric properties of water. In Mätzler, C., Rosenkranz, P. W., Battaglia, A., and Wigneron, J. P. (eds.), *Thermal Microwave Radiation: Applications for Remote Sensing*. Stevenage: Institute of Engineering and Technology. IET Electromagnetic Waves Series, Vol. 52, pp. 431–455.

El-Rayes, M. A., and Ulaby, F. T., 1987. Microwave dielectric spectrum of vegetation – part I: experimental observations. *IEEE Transactions on Geoscience and Remote Sensing*, **25**, 541–549.

Evans, S., 1965. Dielectric properties of ice and snow – a review. *Journal of Glaciology*, **5**, 773–792.

Frankenstein, G., and Garner, R., 1967. Equation for determining the brine volume of sea ice from −0.5 °C to −22.9 °C. *Journal of Glaciology*, **6**, 943–944.

Hallikainen, M. T., 1983. A new low-salinity sea-ice model for UHF radiometry. *International Journal of Remote Sensing*, **4**, 655–681.

Hallikainen, M. T., and Winebrenner, D., 1992. The physical basis for sea ice remote sensing. In Carsey, F. (ed.), *Microwave Remote Sensing of Sea Ice*. American Geophysical Union, Geophysical Monograph 68, Chap. 3, Washington, DC, pp. 29–46.

Hallikainen, M. T., Ulaby, F. T., and Abdel-Razik, M., 1982. Measurements of the dielectric properties of snow in the 4–18 GHz frequency range. In *Proceedings of the 12th European Microwave Conference*. Helsinki, pp. 151–156.

Hallikainen, M. T., Ulaby, F. T., Dobson, M. C., and El-Rayes, M., 1985. Microwave dielectric behavior of wet soil – part I: empirical models and experimental observations. *IEEE Transactions on Geoscience and Remote Sensing*, **23**, 25–34.

Hallikainen, M. T., Ulaby, F. T., and Abdelrazik, M., 1986. Dielectric properties of snow in the 3 to 37 GHz range. *IEEE Transactions on Antennas and Propagation*, **34**, 1329–1340.

Hallikainen, M. T., Ulaby, F. T., and Van Deventer, T. E., 1987. Extinction behavior of dry snow in the 18 to 90 GHz range. *IEEE Transactions on Geoscience and Remote Sensing*, **25**, 737–745.

Hoekstra, P., and Cappillino, P., 1971. Dielectric properties of sea and sodium chloride ice at UHF and microwave frequencies. *Journal of Geophysical Research*, **76**, 4922–4931.

Klein, L. A., and Swift, C. T., 1977. An improved model for the dielectric constant of seawater at microwave frequencies. *IEEE Transactions on Antennas and Propagation*, **25**, 104–11.

Liebe, H. J., Hufford, G. A., and Manabe, T., 1991. A model for the complex permittivity of water at frequencies below 1 THz. *International Journal of Infrared and Millimeter Waves*, **12**, 659–675.

Mätzler, C., 1994. Microwave (1–100 GHz) dielectric model of leaves. *IEEE Transactions on Geoscience and Remote Sensing*, **32**, 947–949.

Mätzler, C., 1996. Microwave permittivity of snow. *IEEE Transactions on Geoscience and Remote Sensing*, **34**, 573–581.

Mätzler, C., and Wegmüller, U., 1987. Dielectric properties of freshwater ice at microwave frequencies. *Journal of Physics D: Applied Physics*, **20**, 1623–1630. Errata, 1988. 21:1660.

Mätzler, C., Aebischer, H., and Schanda, E., 1984. Microwave dielectric properties of wet snow. *IEEE Journal of Oceanic Engineering*, **9**, 366–371.

McDonald, K. C., Zimmermann, R., Way, J. B., and Chun, W., 1999. Automated instrumentation for continuous monitoring of the dielectric properties of woody vegetation: system design, implementation, and selected in situ measurements. *IEEE Transactions on Geoscience and Remote Sensing*, **437**, 1880–1894.

Mironov, V. L., De Roo, R. D., and Savin, I. V., 2010. Temperature-dependable microwave dielectric model for an Arctic soil. *IEEE Transactions on Geoscience and Remote Sensing*, **48**, 2544–2556.

Mironov, V. L., Dobson, M. C., Kaupp, V. H., Komarov, S. A., and Kleshchenko, V. N., 2004. Generalized refractive mixing dielectric model for moist soils. *IEEE Transactions on Geoscience and Remote Sensing*, **42**, 773–785.

Peplinski, N. R., Ulaby, F. T., and Dobson, M. C., 1995. Dielectric properties of soils in the 0.3–1.3 GHz range. *IEEE Transactions on Geoscience and Remote Sensing*, **33**, 803–807. Correction: 33: 1340.

Polder, D., and Van Santen, J. H., 1946. The effective permeability of mixtures of solids. *Physica*, **12**, 257–271.

Sackinger, W. M., and Byrd, R. C., 1972. *Reflection of Mm waves from snow and sea ice*, IAEA Report 7203, Institute of Arctic Environmental Engineering, University of Alaska, Fairbanks.

Sihvola, A., 1999. *Electromagnetic Mixing Formulas and Applications*. London: Institution of Electrical Engineers. IEE Electromagnetic Wave Series, Vol. 47.

Stogryn, A., 1986. A study of the microwave brightness temperature of snow from the point of view of strong fluctuation theory. *IEEE Transactions on Geoscience and Remote Sensing*, **24**, 220–231.

Stogryn, A., 1987. An analysis of the tensor dielectric constant of sea ice at microwave frequencies. *IEEE Transactions on Geoscience and Remote Sensing*, **25**, 147–158.

Stogryn, A., and Desargeant, G. J., 1985. The dielectric properties of brine in sea ice at microwave frequencies. *IEEE Transactions on Antennas and Propagation*, **33**, 523–532.

Tiuri, M., Sihvola, A., Nyfors, E., and Hallikainen, M. T., 1984. The complex dielectric constant of snow at microwave frequencies. *IEEE Journal of Oceanic Engineering*, **9**, 377–382.

Tomppo, L., Tiitta, M., Laakso, T., Harju, A., Venäläinen, M., and Lappalainen, R., 2009. Dielectric spectroscopy of scots pine. *Wood Science and Technology*, **43**, 653–667.

Torgovnikov, G. I., 1993. *Dielectric Properties of Wood and Wood-Based Materials*. Berlin: Springer.

Ulaby, F. T., and El-Rayes, M. A., 1987. Microwave dielectric spectrum of vegetation – part II: dual-dispersion model. *IEEE Transactions on Geoscience and Remote Sensing*, **25**, 550–557.

Ulaby, F. T., Moore, R. K., and Fung, A. K., 1986. *Microwave Remote Sensing – Active and Passive, III: From Theory to Applications*. Dedham: Artech House.

Ulaby, F. T., Tavakoli, A., and Senior, T. B. A., 1987. Microwave propagation constant for a vegetation canopy with vertical stalks. *IEEE Transactions on Geoscience and Remote Sensing*, **25**, 714–725.

Vant, M. R., 1976. *A Combined Empirical and Theoretical Study of the Dielectric Properties of Sea Ice over the Frequency Range 100 MHz to 40 GHz*. Technical Report, Carleton University, Ontario, Ottawa.

Vant, M. R., Ramseier, R. O., and Makios, V., 1978. The complex-dielectric constant of sea ice at frequencies in the range 0.1–40 GHz. *Journal of Applied Physics*, **49**, 1264–1280.

Von Hippel, A., 1954. *Dielectrics and Waves*. Boston: MIT Press.

Wang, J. R., 1980. The dielectric properties of soil-water mixtures at microwave frequencies. *Radio Science*, **15**, 977–985.

Wang, J. R., and Schmugge, T. J., 1980. An empirical model for the complex dielectric permittivity of soils as a function of water content. *IEEE Transactions on Geoscience and Remote Sensing*, **18**, 288–295.

MICROWAVE HORN ANTENNAS

Yahya Rahmat-Samii
Department of Electrical Engineering, University of California at Los Angeles, Los Angeles, CA, USA

Definition

Horn Antenna. An antenna with metallic flare walls which provides a transition between waves propagating in a waveguide and electromagnetic waves directivity radiated into space.

Microwave horn antennas

Horns are one of the simplest and most widely used microwave antennas. Interests in horn antennas date back to the turn of the nineteenth century and then considerably revived during World War II. Microwave horn antennas are essentially a device to make a transition from waves propagating in a waveguide into electromagnetic signals transmitting in another open medium such as free space. They can have metallic walls, dielectric material, or the combination of both, and occur in various shapes and sizes to fulfill many practical applications, such as communication systems, remote sensing, radio frequency heating, reference sources for other antenna testing and evaluation. Other than being a stand-alone directive power transmitter/receiver, horns are also used as feeds for other antennas such as reflectors, lenses, and compound antennas. Their widespread application is due to their simple, solid geometry and excellent performance when beam directivity is required (Love, 1976; Balanis, 1988, 1996; Olver et al., 1994; Bird and Love, 2007).

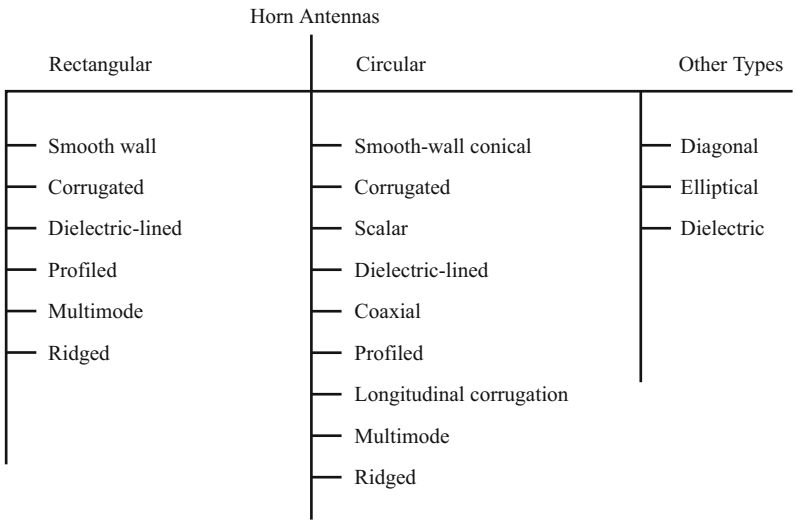
Figure 1 summarizes the primary categories of horns, and Figure 2 shows some representative examples of well-known horn designs. Rectangular pyramidal horns are probably the most commonly used class of horns. One of the primary applications of a pyramidal horn is the standard gain antenna since its gain may be calculated very accurately by knowing their dimensions. The beamwidths in the two principal planes can also be independently controlled by varying the rectangular-aperture dimensions. Some other forms of rectangular horns with nonlinear profiles are also used to increase aperture efficiency and lower side lobes compared with a linear flare. Circular geometries are also in widespread use. The axial symmetry of conical horns allows them to generate any polarization of the dominant mode which makes them well suited for circular polarization. The beamwidths, however, are usually unequal in the two principal planes. In order to overcome this problem, the dual-mode and corrugated (hybrid-mode) conical horns have been developed. Higher-order modes are excited so that the aperture field of the horn is modified in such a way as to produce radiation patterns with axial symmetry and very low cross-polarization.

Horn specifications

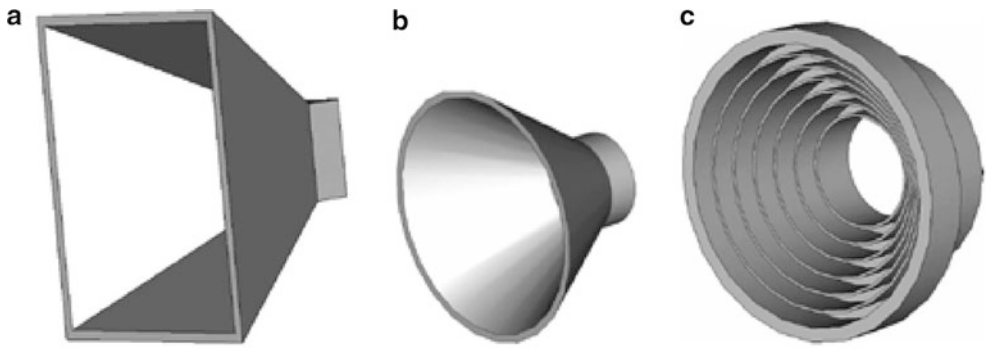
Ample types of horns can provide various radiation performances, and for every specific application, it is necessary for an antenna designer to consider some critical specifications which dictate the choosing of an appropriate type of horn antennas.

Bandwidth

The frequency band over which the system is to operate is usually specified as a range of frequencies (for single-band applications) where the antenna must satisfy a required return loss (or VSWR) and provide a radiation pattern adequate for the application. Consider an antenna required to operate from f_{min} to f_{max}. The bandwidth, in general, can be calculated as



Microwave Horn Antennas, Figure 1 A summary of primary types of horn antennas.



Microwave Horn Antennas, Figure 2 Representative types of horn antennas: (**a**) pyramidal horn, (**b**) circular horn, and (**c**) corrugated horn.

$$\text{Bandwidth } (\%) = \frac{2(f_{\max} - f_{\min})}{(f_{\max} + f_{\min})} \times 100.$$

Based on this definition, a bandwidth below 20 % is generally considered narrowband whereas a bandwidth over 20 % is considered wideband. Often, there is a specific frequency determined in specifications as the design frequency at which the design provides optimum performances. For single-band applications, a practical choice is a design frequency of $f_c = \sqrt{f_{\min} \cdot f_{\max}}$ for narrowband operations and $f_c = 1.2\, f_{\min}$ for wideband operations.

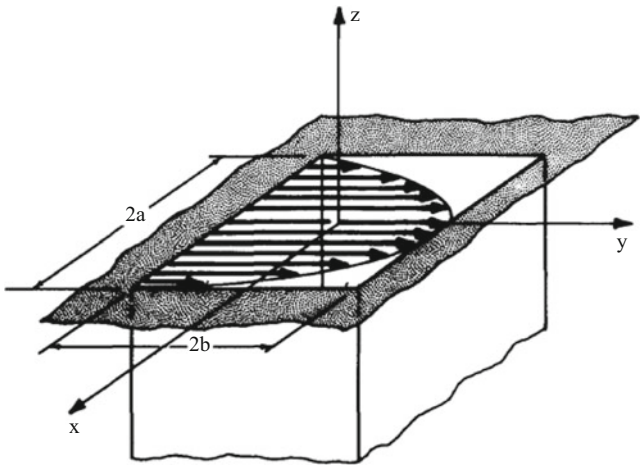
Gain and aperture efficiency

The antenna gain takes into account both the directional capability and efficiency of the antenna. Among many different types of gains, two definitions are usually used to describe horn antennas (Balanis, 1996). The most commonly accepted definition of gain only considers the accepted power and removes the power reflected from the input. In this definition, an ideal horn without losses is assumed and the gain equals directivity as the efficiency is 100 %. Another definition, called absolute gain, retains the reflected power from the input, as this is an intrinsic property of the antenna. For horn antennas, the results of these two definitions are usually very similar since the reflection coefficient is small. The difference is a factor of $1/(1 - |\Gamma_{in}|^2)$, where Γ_{in} is the reflection coefficient at the input.

Another useful measure for comparing the performances of horn antennas is the aperture efficiency. It is defined as the ratio of the antenna gain in the direction (θ, ϕ) and the gain of the same antenna aperture with ideal uniform illumination given by

$$\eta_{aperture} = \frac{\lambda^2}{4\pi A} G(\theta, \varphi),$$

where A is the total area of the aperture, λ is the wavelength at the operating frequency, and $G(\theta, \phi)$ is the gain function. In many applications it is assumed that the aperture efficiency is in the boresight direction of the antenna.

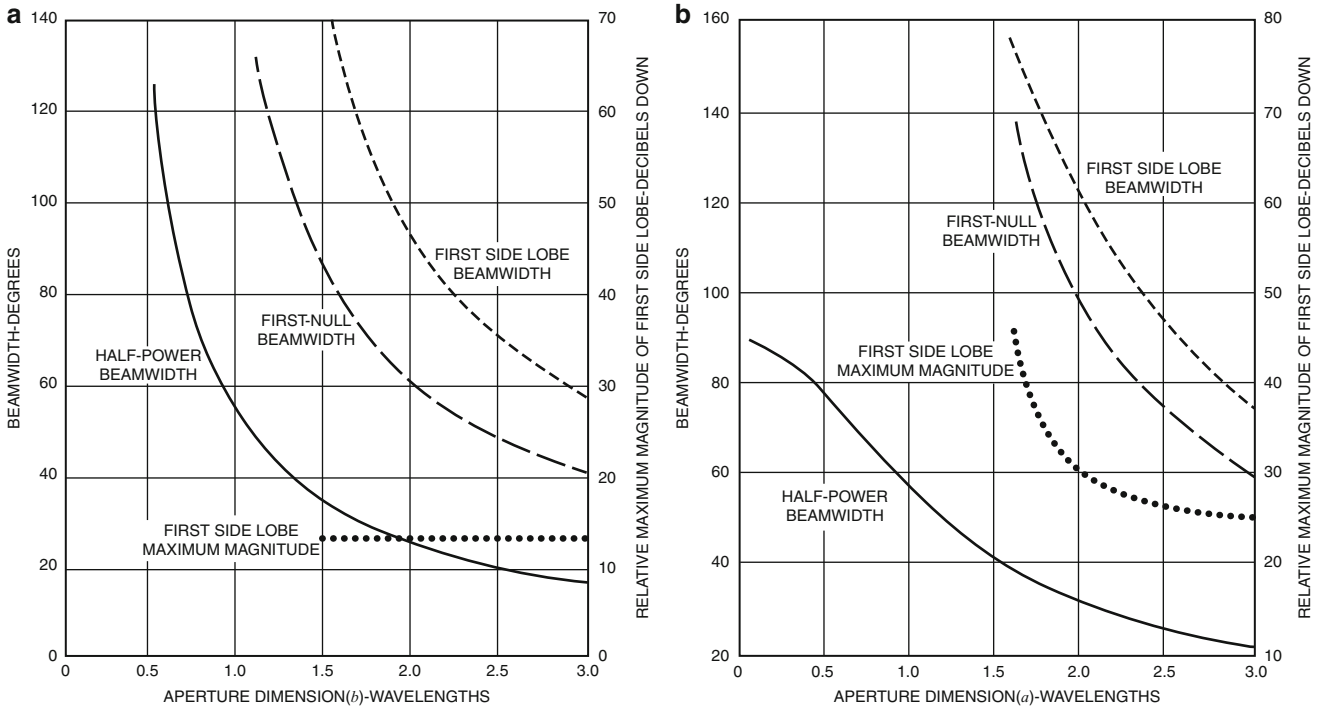


Microwave Horn Antennas, Figure 3 A TE_{10}-mode open-ended rectangular waveguide feed with dimensions 2a and 2b.

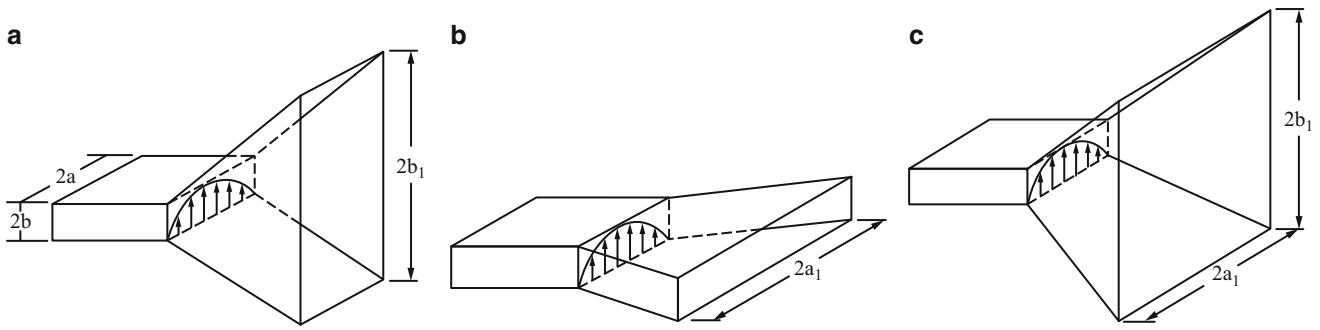
It has to be noted that what is usually defined is the gain of the whole antenna system, not just the gain of the aperture antenna in isolation. Therefore, all other losses such as mismatch and ohmic losses have to be taken into consideration. A well-designed aperture antenna fed by a poorly designed feed system results in poor performance.

Polarization

The radiated far-field of a horn antenna can be decomposed into two components: The one in the direction of the desired polarization is called co-polarization, and the orthogonal component is called cross-polarization. Polarization performance measurement can determine, in general, how much power is radiated in the undesired polarization which cannot be received by receiving antennas. The most common measurement is the peak cross-polarization level for linear polarization and the axial ratio for circular polarization. The measurement of the cross-polarization can be achieved in different ways. For linearly polarized horns, the co-polar pattern is measured by aligning the test antenna with the



Microwave Horn Antennas, Figure 4 Different beamwidths and first side lobe levels for TE_{10}-mode open-ended rectangular waveguide on ground plane: (**a**) E-plane and (**b**) H-plane.



Microwave Horn Antennas, Figure 5 Typical rectangular horn antennas: (**a**) E-plane sectoral horn, (**b**) H-plane sectoral horn, and (**c**) pyramidal horn.

polarization of the distant source antenna, and the cross-polar pattern is obtained by rotating the source antenna by 90° and repeating the radiation pattern measurement.

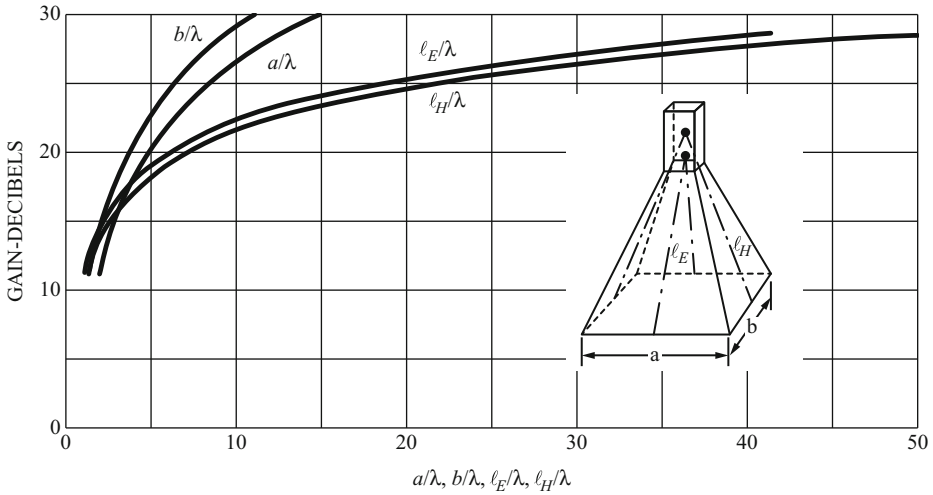
Functioning environment

Consideration of the working environment in which the antenna is likely to operate is important in design process as well. In many cases this too will dictate the choice of materials and the mechanical aspects of the system. One consideration is the potential effects of surrounding objects on the horn's performance. A horn antenna may work perfectly well in isolation but give poor performance when surrounded by nearby scatterers. A typical effect is caused by conductive surrounding objects which are fitted afterward in forms of blockage or radome effect. Other considerations are any special needs for horn antennas such as pressurization, high-power handling, low passive intermodulation level, robustness, and low-loss requirements.

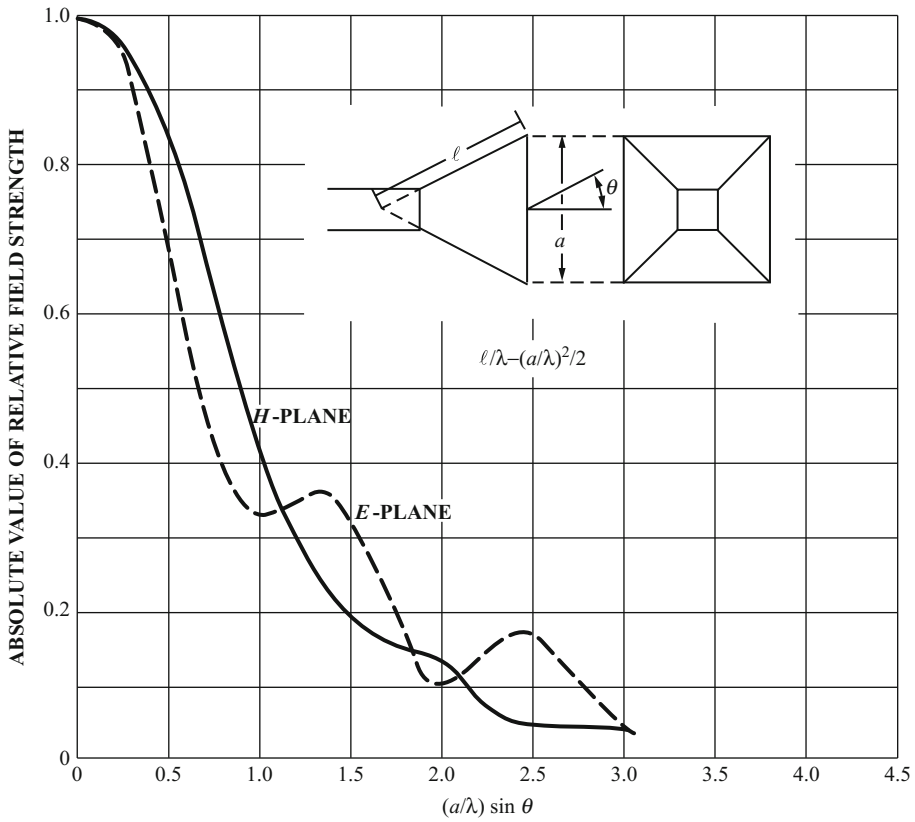
Conventional horn antennas

Rectangular horn

Rectangular horns are constructed by gradually flaring a rectangular waveguide to provide a smooth transition from the input to free space. If the flare angle is small enough so that the higher-order modes can be



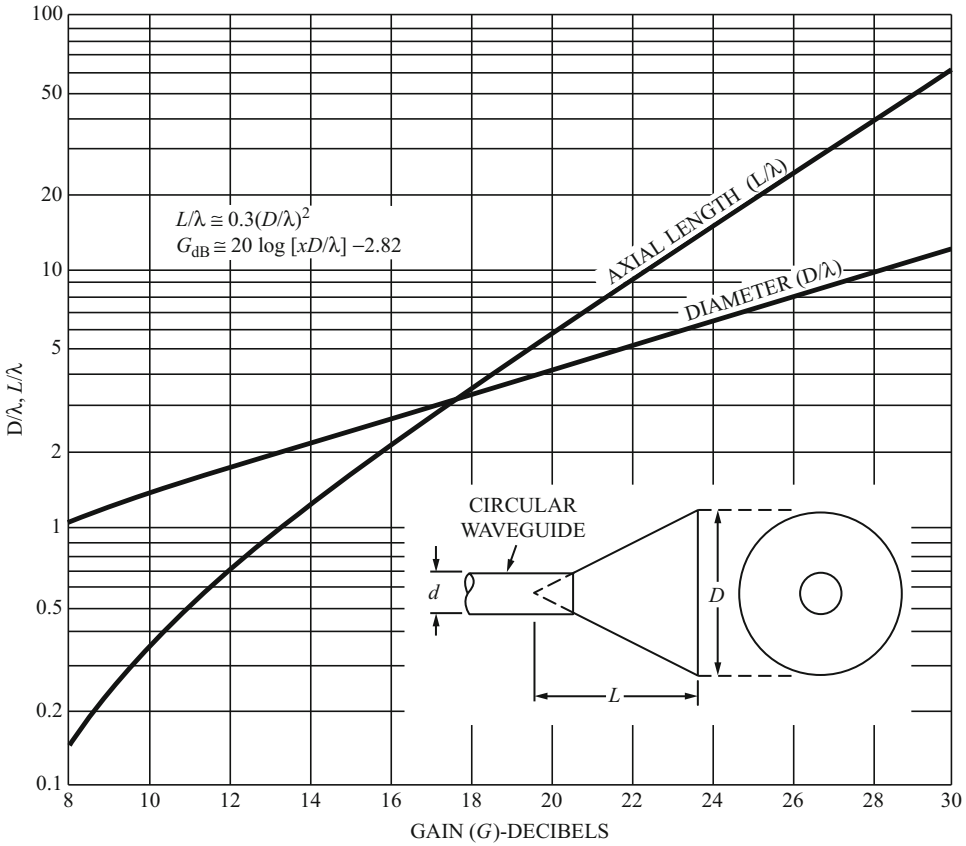
Microwave Horn Antennas, Figure 6 Gain characteristics of pyramidal horns of optimum-gain design.



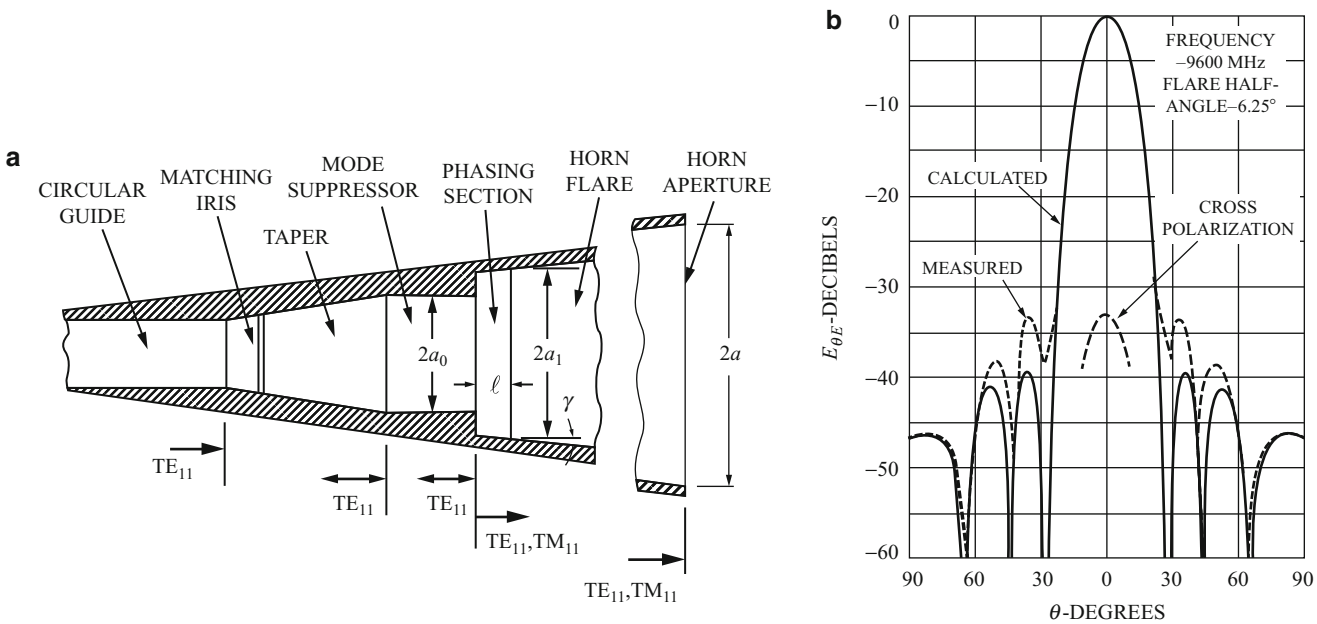
Microwave Horn Antennas, Figure 7 Universal patterns of pyramidal horns of optimum-gain design.

neglected, the transverse electric field of the dominant mode of the waveguide is a good first approximation to the aperture field at the horn mouth. For reference, some important characteristics of the open-ended rectangular waveguide (Figure 3) are shown in Figure 4.

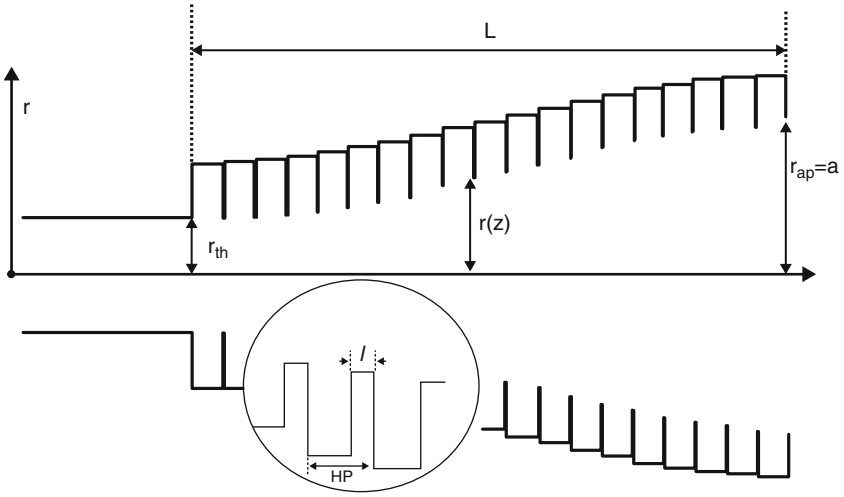
Depending on to what plane its opening tapers rectangular horns may be classified as E-plane sectoral, H-plane sectoral, or pyramidal horns (Figure 5). The type, direction, and amount of taper can have a significant effect on the overall performance of the horn. The radiation characteristics can be determined using the aperture field



Microwave Horn Antennas, Figure 8 Gain characteristics of conical horns of optimum-gain design.



Microwave Horn Antennas, Figure 9 Dual-mode conical Potter horn: (**a**) horn structure and (**b**) E-plane pattern.



Microwave Horn Antennas, Figure 10 Profiled conical corrugated horn.

method or GTD construction (Jull, 1973; Huang et al., 1983). The latter is also capable of predicting the near field and back-radiated field. Considerable amounts of design data are available on rectangular horns, and the reader is referred to the just-mentioned references (Balanis, 1988, 1996; Bird and Love, 2007).

For pyramidal horns, the geometrical parameters may be chosen to achieve the so-called optimum-gain design (Balanis, 1988). Figure 6 shows the relationship between the parameters to design an optimum-gain horn for a specified gain (the overall efficiency of these horns is typically about 50 %). Figure 7 is the plot of the E- and H-plane patterns of such horns.

Conical horns

The geometry of a circular horn is shown in Figure 8. In contrast to pyramidal horns which are typically fed by rectangular waveguides, the circular horn is usually fed by circular waveguides. The aperture field of these horns can be constructed in a fashion similar to that of pyramidal horns by simply multiplying the aperture field of the circular waveguide by a quadratic phase term (Balanis, 1988). The resulting integral for the computation of the radiated field can then be evaluated numerically. Figure 8 gives the proper horn dimensions for constructing an optimum-gain horn for a specified gain. These horns typically possess more symmetric E- and H-plane patterns than do their pyramidal horn counterparts. They can also be used more effectively to create circularly polarized fields.

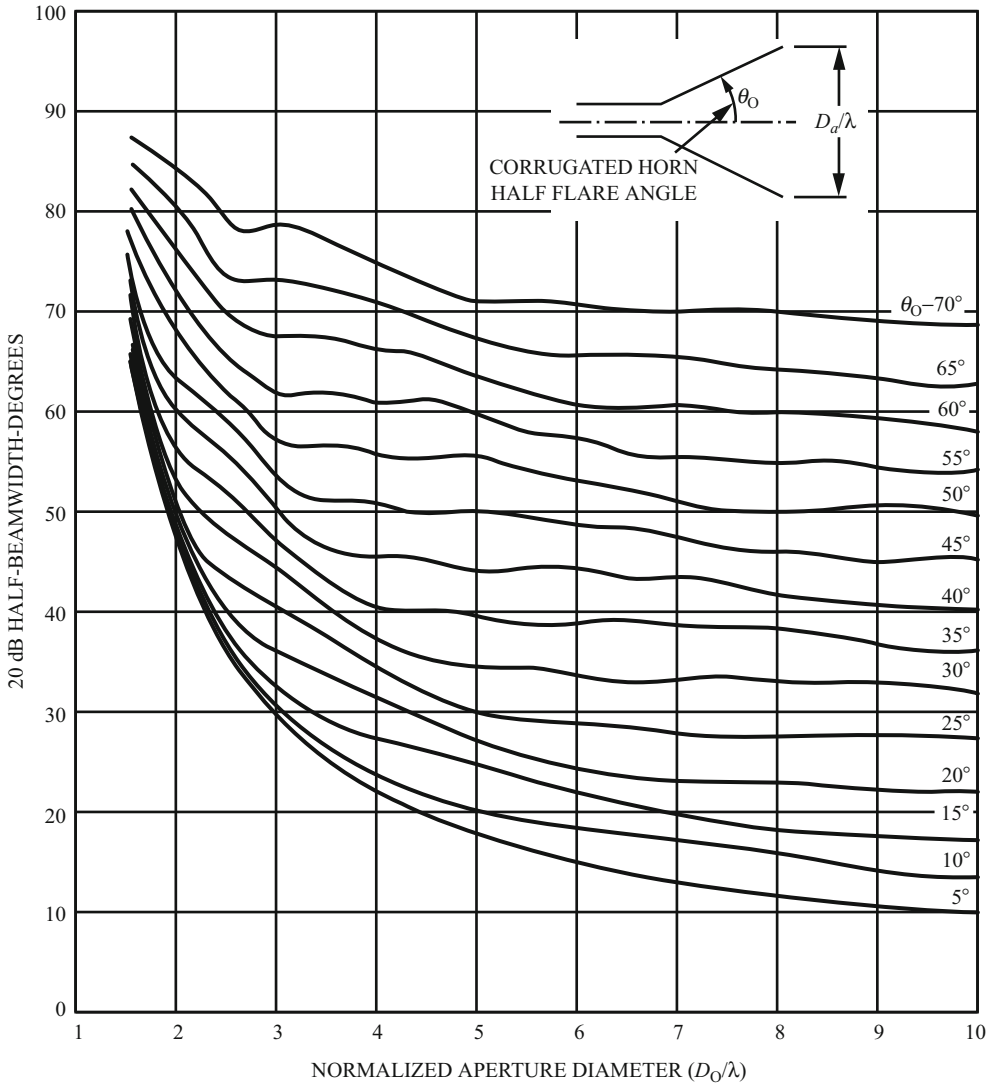
Multimode, corrugated, matched horns, and others

The requirements for designing more advanced remote sensing systems have resulted in the generation of a new class of feeds. These feeds in particular are used in large reflectors for radio astronomy and remote sensing reflectors. Among them one may refer to multimode (Potter, 1963; Thomas, 1970; Bhattacharyya and Goyette, 2004), corrugated (hybrid mode) (Clarricoats and Olver, 1984; James, 1984; Thomas et al., 1986; Olver et al., 1994), and matched horns (Rudge and Adatia, 1975; Bahadori and Rahmat-Samii, 2006).

Multimode horns were invented to equalize the pattern asymmetry of single-mode horns. For instance, in the Potter horn (Potter, 1963) the TM_{11} mode is generated along with the dominant TE_{11} mode of a circular horn. Although this new TM_{11} mode does not have any appreciable effect on the H-plane radiation pattern, when it is properly phased and combined with the TE_{11} mode, it can effectively alter the E-plane aperture distribution, which results in a symmetric radiation pattern. All these favorable features, however, cannot be properly realized until the feed aperture diameter exceeds about 1.3 λ. For example, Figure 9 shows the radiation pattern of a dual-mode Potter horn as reported in Potter (1963). A partial conversion of the TE_{11} mode energy to a TM_{11} mode happens in the flared section of the horn, while the straight section enforces the condition that both modes have the proper phase relation at the aperture which can be maintained over a bandwidth of less than 10 %. There are also available other types of multimode horns, which result from a combination of modes such as TE_{10}, TE_{12}, and TM_{12} in a square-aperture pyramidal horn.

Corrugated horns have become the main choice of antenna feeds in recent advanced communications, radar, and remote sensing applications where demanding performance is required. First introduced in the 1960s (Minnett and Thomas, 1966; Simmons and Kay, 1966), they can provide superior radiation performances such as symmetric patterns and low cross-polarization levels compared with conical horns.

Corrugated horns are capable of creating similar boundary conditions at all polarizations which result in similar tapers in the aperture field distribution in all planes. Due to these boundary conditions, symmetric radiation



Microwave Horn Antennas, Figure 11 20 dB half beamwidths of corrugated horns.

patterns can be obtained at levels as low as −25 dB in both the E- and H-planes. A corrugated horn can be realized by grooving the E-plane of a pyramidal horn or the entire wall of a circular horn with, typically, six or more slots (corrugations) per wavelength (Figure 10). For circular corrugated horns, Figure 11 shows the plots of the pattern widths at 20 dB level as a function of the opening angle.

The existence of the corrugations, especially near the waveguide-horn junction, affects the VSWR of the horn. The usual practice is to begin the corrugations at a small distance from the junction. These horns are also classified as hybrid-mode horns because they support modes in which both longitudinal E- and H-field components are present. In a circular corrugated horn, a natural mixture of TE_{11} and TM_{11} results in the generation of a hybrid-mode HE_{11}. In contrast to dual-mode horns there is no need of a mode converter, and therefore, the hybrid-mode horns have typically wider bandwidths. In particular, one version of these hybrid horns (scalar horns), which has a large flare angle with a relatively short horn, has radiation characteristics with little dependence on frequency. There now exists advanced computational and optimization methods (Sinton et al., 2002) to tailor corrugated horn antennas for various applications including profiled corrugated horns with shorter lengths than the standard straight corrugated horns.

There are other horn feeds which are properly tailored to specifically overcome some undesirable characteristics of reflectors. For example, matched horn feeds (Rudge and Adatia, 1975; Bahadori and Rahmat-Samii, 2006) are used to significantly reduce the generation of unwanted cross-polarized fields in an offset parabolic reflector illuminated with a tilted horn feed. The basic concept behind these designs is to match the horn

aperture field distribution with the receiving focal plane distribution of the reflector.

There also exists a class of dielectric horns with some unique properties, as described in Oh et al. (1970), Baldwin and McInnes (1973), and Jha et al. (1975), and other configurations (Love, 1962; Vokurka, 1979).

Summary

Microwave horn antennas are widely used as standard gain antennas for antenna testing or evaluation or feeds for other antennas due to their simple and solid geometries and excellent performances. Among many different types of horns, rectangular pyramidal horns and conical horns, multimode and corrugated horns are the most popular and practical configurations, and a selection of a specific type is made based on the design specifications such as bandwidth, gain and aperture efficiency, polarization, and some other unique radiation characteristics. With the advances in computational and optimization methods, more sophisticated designs such as profiled corrugated horns and matched horns are developed for demanding applications in radars, satellite communications, remote sensing, deep-space telemetry, and radio astronomy.

Bibliography

Bahadori, K., and Rahmat-Samii, Y., 2006. "A Tri-Mode Horn Feed for Gravitationally Balanced Back-to-Back Reflector Antennas", *IEEE Antennas and Propagation Society Symposium,* pp. 4397–4400, Albuquerque, NM.

Bahadori, K., and Rahmat-Samii, Y., 2009. "Tri-Mode Horn Feeds Revisited: Cross-pol Reduction in Compact Offset Reflector Antennas," *IEEE Trans. Antennas Propag,* vol. 57, no. 9, pp. 2771–2775.

Balanis, C. A., 1988. Horn antennas. In Lo, Y. T., and Lee, S. W. (eds.), *Antenna Handbook*. New York: Van Nostrand Reinhold.

Balanis, C. A., 1996. *Antenna Theory: Analysis and Design*, 2nd edn. New York: Wiley.

Baldwin, R., and McInnes, P. A., 1973. Radiation patterns of dielectric loaded rectangular horns. *IEEE Transactions on Antennas and Propagation*, **21**, 375–376.

Bhattacharyya, A., and Goyette, G., 2004. A novel horn radiator with high aperture efficiency and low cross-polarization and applications in arrays and multibeam reflector antennas. *IEEE Transactions on Antennas and Propagation*, **52**, 2850–2859.

Bird, T. S., and Love, A. W., 2007. Horn antennas. In Volakis, J. (ed.), *Antenna Engineering Handbook*, 4th edn. New York: McGraw-Hill.

Clarricoats, P. J. B., and Olver, A. D., 1984. *Corrugated Horns and Microwave Antennas*. London: Peregrinus.

Huang, J., Rahmat-Samii, Y., and Woo, K., 1983. A GTD study of pyramidal horns for offset reflector antenna applications. *IEEE Transactions on Antennas and Propagation*, **31**, 305–309.

James, G. L., 1984. Design of wide-band compact corrugated horns. *IEEE Transactions on Antennas and Propagation*, **32**, 1134–1138.

Jha, R. K., Misra, D. K., and Jha, L., 1975. Comparative study of dielectric and metallic horn antennas. In *IEEE Antennas and Propagation Society International Symposium*, pp. 16–18.

Jull, E. V., 1973. Errors in the predicted gain of pyramidal horns. *IEEE Transactions on Antennas and Propagation*, **21**, 25–31.

Love, A. W., 1962. The diagonal horn antenna. *Microwave Journal*, **5**, 117–122.

Love, A. W. (ed.), 1976. *Electromagnetic Horn Antennas*. New York: IEEE Press.

Minnett, H. C., and Thomas, B. M., 1966. A method of synthesizing radiation patterns with axial symmetry. *IEEE Transactions on Antennas and Propagation*, **14**, 654–656.

Oh, L. L., Peng, S. Y., and Lunden, C. D., 1970. Effects of dielectrics on the radiation patterns of an electromagnetic horn. *IEEE Transactions on Antennas and Propagation*, **18**, 553–556.

Olver, A. D., Clarricoats, P. J. B., Kishk, A. A., and Shafai, L., 1994. *Microwave Horns and Feeds*. New York: IEEE Press.

Potter, P. D., 1963. A new horn antenna with suppressed sidelobes and equal beamwidths. *Microwave Journal*, **6**, 71–78.

Rudge, A. W., and Adatia, N. A., 1975. New class of primary-feed antennas for use with offset-parabolic-reflector antennas. *Electronics Letters*, **11**, 597–599.

Simmons, A. J., and Kay, A. F., 1966. The scalar feed – A high performance feed for large paraboloid reflectors. In *IEE Conference Publications*, Vol. 21, pp. 213–217.

Sinton, S., Robinson, J., and Rahmat-Samii, Y., 2002. Standard and micro genetic algorithm optimization of profiled corrugated horn antennas. *Microwave and Optical Technology Letters*, **35**, 449–453.

Thomas, B. M. A., 1970. Prime-focus one- and two-hybrid-mode feeds. *Electronics Letters*, **6**, 460–461.

Thomas, B. M. A., James, G. L., and Greene, K. J., 1986. Design of wide-band corrugated conical horns for Cassegrain antennas. *IEEE Transactions on Antennas and Propagation*, **34**, 750–757.

Vokurka, V. J., 1979. Elliptical corrugated horn for broadcasting satellite antennas. *Electronics Letters*, **15**, 652–654.

Cross-references

Electromagnetic Theory and Wave Propagation
Emerging Technologies, Radar
Microwave Radiometers
Microwave Radiometers, Conventional
Radars

MICROWAVE RADIOMETERS

Niels Skou
National Space Institute, Technical University of Denmark, Lyngby, Denmark

Synonyms

Passive microwave radiometer (PMR)

Definition

The microwave radiometer is a calibrated receiver that measures properties of the natural emission from the environment as picked up by an antenna system.

Introduction: what is radiometry about?

All bodies at a temperature above the absolute zero (0 K = −273 °C) radiate power, according to Planck's law. At microwave frequencies the Rayleigh-Jeans approximation holds, and the radiation is proportional to physical temperature. Actually, this is only true for the

so-called blackbodies, which are perfect emitters. Natural bodies radiate less, and we introduce the term emissivity (ε) which is a number between 0 and 1 describing how well the body radiates relative to a blackbody. Within radiometry the radiated power is expressed as the so-called brightness temperature, T_B, so that $T_B = \varepsilon \cdot T_{phys}$. The brightness temperature of a blackbody is thus equal to its physical temperature, while all natural bodies will have brightness temperatures lower than that. The brightness temperature of a natural body can also be understood as that physical temperature a blackbody would have to have in order to emit the power in question. Typical values related to planet Earth range from almost 0 K (looking up toward free space) to about 300 K.

The radiated energy can be picked up by an antenna in order to be measured by a radiometer. Unfortunately antennas are not perfect: a perfect antenna would have a sharply defined main lobe so that when this is pointed toward the target, only this contributes to the received power. But the main lobe is not with sharp cutoff, so some power is received from the surroundings of the target. Furthermore, antennas have side lobes far from the main lobe meaning that some power is also received from other directions. The process can be expressed by the following equation:

$$T_A = \frac{1}{4\pi} \cdot \int \int_{4\pi} T_B(\theta, \phi) \cdot G(\theta, \phi) d\Omega. \qquad (1)$$

This convolution integral, where $G(\theta, \phi)$ is the antenna radiation pattern, shows how the received power, denoted as the so-called antenna temperature T_A, is a gain-weighted summation of the brightness temperatures $T_B(\theta, \phi)$ from each direction (see Ulaby et al., 1981).

Sensitivity

As stated above, the task of the radiometer is to measure the power picked up by the antenna. So in fact the radiometer is basically a calibrated microwave receiver. The basic and simplest radiometer, the total power radiometer, is illustrated in Figure 1a.

The received power is amplified with a gain G, a certain bandwidth B around the center frequency is selected by the filter, the microwave signal is detected by a square law detector, and since we are handling noise-like signals, a certain integration time τ is required. Finally, it is indicated that we cannot make a receiver without introducing an additional noise T_N.

The output will represent a power measure expressed as $P = k \cdot B \cdot G \cdot (T_A + T_N)$ where k is Boltzmann's constant: $1.38 \cdot 10^{-23}$ J/K. As already stated the signals are noise-like, so the output fluctuates, but the fluctuations are smoothed by the integration. The level of fluctuations, the standard deviation of the output to be specific, is expressed as the radiometer sensitivity, ΔT, and it is calculated from Eq. 2:

$$\Delta T = \frac{T_A + T_N}{\sqrt{B.\tau}}. \qquad (2)$$

Stability and accuracy

A fundamental problem with this simple and direct implementation of the radiometer is that the output, $P = k \cdot B \cdot G \cdot (T_A + T_N)$, is dependent on stability of the gain G and the receiver noise temperature T_N. Hence, stability may be problematic, and frequent calibration is required, or a more developed radiometer principle must be employed (Skou and LeVine *cpoo*).

If k, B, G, and T_N are not only constants but known constants, we have no absolute accuracy problems: a given T_A results in a given P that can be calculated. Such knowledge of the constants is rarely available, leading to the necessity for calibration. This illustrates the fundamental difference between stability and accuracy: stability is a highly appreciated virtue of an instrument, but a stable instrument need not be accurate. The steps toward accuracy include the calibration process.

In the following we will describe a slightly different aspect of absolute accuracy, which stresses the care that must be exercised when designing or working with radiometers. Consider losses in a signal path – it could be the waveguide connecting the antenna with the radiometer input or a passive component in the radiometer front end. Let ℓ denote the fractional loss (or the absorption coefficient) and T_0 the physical temperature. T_1 is the input temperature and the output temperature is then given by $T_2 = T_1 (1 - \ell) + \ell \cdot T_0$.

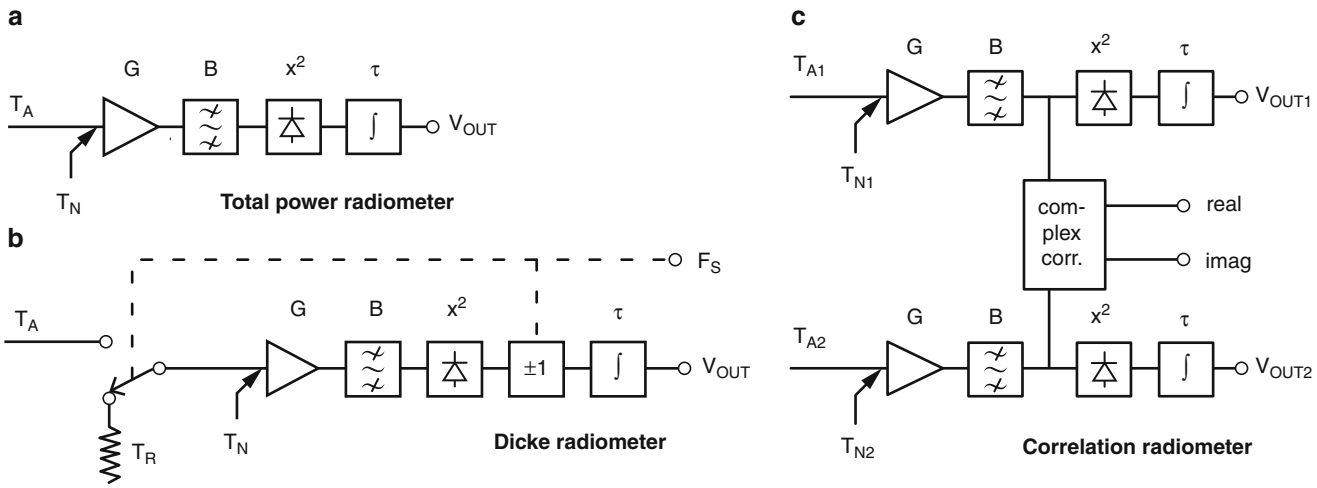
The difference between output and input is $T_D = T_2 - T_1 = \ell (T_0 - T_1)$. If T_1 is 100 K and T_0 is 300 K, a loss as small as 0.01 dB ($\ell = 0.0023$) results in a difference, T_D, of 0.5 K. Bearing in mind that the losses of a real signal path are much greater than 0.01 dB, the physical temperature of the path must be measured and used for correction of the measured brightness temperature. The corresponding losses must be known to an accuracy of better than 0.01 dB and must remain stable within the same limits.

Consider a mismatch, for example, at the input of a radiometer; with a reflection coefficient ρ, we similarly find $T_D = \rho (T_{RAD} - T_1)$ where T_{RAD} is the microwave temperature as seen from the point of reflection into the radiometer. T_{RAD} is typically 300 K and if T_1 again is assumed to be 100 K, a reflection coefficient of −26 dB will give an error (T_D) of 0.5 K. Care must be exercised to obtain reflection coefficients better than −26 dB.

Radiometer types

The conventional radiometers are illustrated to the left in Figure 1, while the more advanced types are illustrated by the correlation radiometer to the right.

The total power radiometer has already been discussed. It is the simplest and basic radiometer that has the optimum sensitivity but may have problems with stability.



Microwave Radiometers, Figure 1 Two conventional radiometer types: (**a**) total power and (**b**) Dicke and (**c**) the correlation radiometer.

The classical way of alleviating stability problems is to employ switching as done in the Dicke radiometer (Figure 1b). The idea here is that the radiometer in fact measures the difference between the unknown antenna temperature and an internal, known reference. The Dicke radiometer has proven very good behavior and has been widely used. An enhanced version of the Dicke radiometer is the noise injection radiometer (NIR) in which a null-balancing technique is employed resulting in total independence of gains and receiver noise temperatures. A price has to be paid for the enhanced stability of Dicke type switching radiometers: the sensitivity is degraded by a factor of 2 compared with that of the total power radiometer.

More advanced radiometers are very often based on the correlation radiometer shown in Figure 1c. It basically consists of two total power radiometers, but in addition to the usual detected outputs, the signals before detection in the two channels are cross correlated to produce two additional outputs: the real and imaginary part of the correlation. The correlation radiometer is the building block of interferometric, synthetic aperture radiometer systems and often also of polarimetric radiometer systems.

Radiometer antennas

An important part of any radiometer system is its antenna. The purpose of the antenna is to collect the emitted energy from a target and present it to the radiometer input. Thus, the antenna radiation pattern determines the spatial resolution of the radiometer system. Important antenna parameters are main beam efficiency and side lobe level. A good beam efficiency and low side lobes ensure that the radiometer system primarily measures what the antenna points toward and not so much the surroundings. Of paramount importance for a radiometer antenna is low loss following the previous accuracy discussion. Hence, the most popular antenna types are the microwave horn and reflector antennas.

Horns are bulky, low-gain devices and as such normally not used as primary spaceborne antennas, but they play an important role in ground-based and airborne systems and as feeds for reflector antennas. Especially Potter horns and dual-mode horns are favored due to excellent radiation patterns.

Reflector antennas – typically based on offset parabolic reflectors – have been used extensively in spaceborne systems. They feature low loss and good radiation patterns, and, very importantly, they can serve a wide range of frequencies (e.g., 5–90 GHz) through the use of clusters of feed horns.

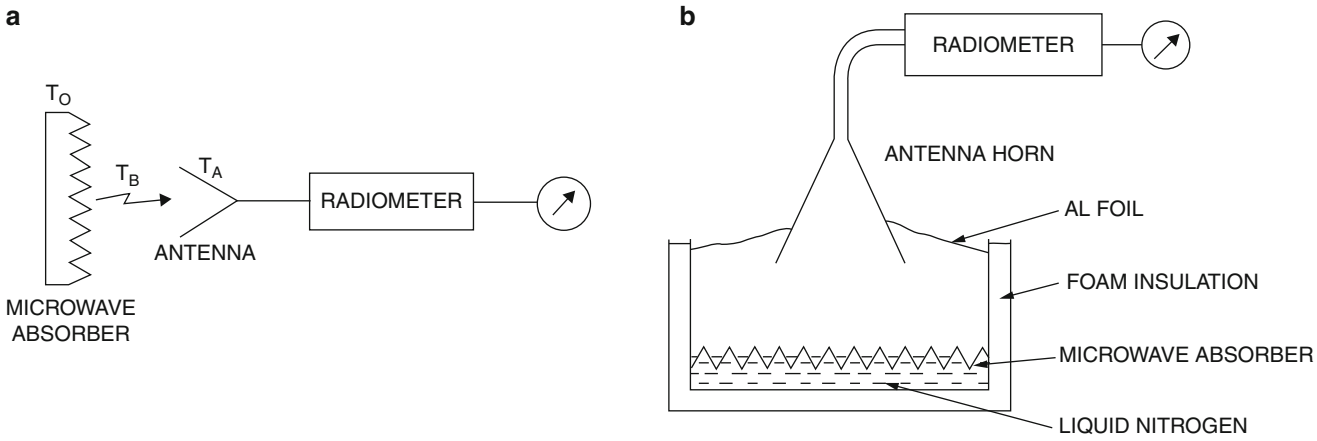
For the antennas discussed here, where the electrical aperture is roughly equal to the physical aperture, it is very simple to estimate the width of the main beam. The 3 dB beamwidth is slightly larger than the reciprocal of the aperture measured in wavelengths of the operating frequency:

$$\theta = 1.4 \cdot \frac{\lambda}{D}. \tag{3}$$

The factor 1.4 reflects the need for high beam efficiency. A 1 m reflector antenna used at 30 GHz (1 cm wavelength) will thus have a 0.014 (= 0.8°) beamwidth. Looking straight down from a satellite in an 800 km orbit, the footprint on Earth will be 11.2 km. Spaceborne radiometer systems are not high-resolution devices like optical scanners or synthetic aperture radars, but they do supply unique data for geophysical interpretation with frequent coverage of the globe.

Calibration

The purpose of calibration is to establish the relation between the input brightness temperature and the



Microwave Radiometers, Figure 2 Antenna target calibrator, principle, and practical layout.

radiometer output – usually digital numbers following analog to digital conversion of the radiometer output voltage. Assuming that the radiometer transfer is linear, two calibration points are required, typically one hot around 300 K and one cold in the 0–77 K range. The most obvious solution is a cooled microwave termination. A well-matched load will generate a noise temperature equal to its physical temperature – and the physical temperature is easily measured. If the load is left at room temperature, it provides a hot calibration point, and if it is submerged into liquid nitrogen, it will be a cold load at around 77 K. However, the latter is not quite as easy as it sounds: heat flow into the load must be prevented, and the loss in the input transmission line must be corrected for.

An alternative solution is a cooled target viewed by a suitable antenna connected to the radiometer (see Figure 2). A microwave absorber (normally used to cover the inside walls of radio anechoic chambers) will emit a brightness temperature equal to its physical temperature T_0. Under ideal conditions, the antenna will sense nothing but the brightness temperature from the absorber and $T_A = T_B = T_0$. Figure 2 also shows a practical layout of this concept. The radiometer is connected to an antenna horn through a short waveguide (very low losses!). The horn views a microwave absorber soaked with liquid nitrogen. The absorber and the liquid nitrogen are contained in an insulated metal bucket, and the excess opening of the bucket is covered by aluminum foil. In this way the antenna is only able to pick up energy from the absorber, which is cooled to 77 K by the nitrogen. There is generally no problem with losses and heat flow in the antenna and the waveguide, liquid nitrogen is readily available, and the setup is cheap and simple. Overall this is a very useful radiometer calibrator – and before liquid nitrogen is applied it provides also the hot calibration point.

Finally, it shall be noted that pointing an antenna toward free space will provide a cold calibration point of only a few K. The problem is that the atmosphere will give a significant contribution to the temperature, and it must be carefully accounted for. However, this is not the case for satellite-borne systems, and here such a sky view is often used for the cold calibration point.

Going to the more advanced correlation radiometers, also shown in Figure 1, calibration is somewhat more complicated. First, a basic calibration of the two channels is carried out as just described, but in addition to this, the calibration of the correlation channels requires generation of a pair of known signals with a known amount of correlation between them. This issue is outside the scope of the present text.

Conclusion

Microwave radiometers are sensitive receivers requiring special attention to stability and accuracy. Several types have been developed over the years each with its special virtue. The antenna is an important part of the radiometer system determining what is actually measured by the radiometer. Any radiometer must be carefully calibrated – not only after its development but also a frequent check of calibration must be carried out at proper intervals when the radiometer is in use. More information can be found in (Skou and LeVine, 2006).

Bibliography

Skou, N., and LeVine, D., 2006. *Microwave Radiometer Systems, Design and Analysis*. Boston: Artech House.

Ulaby, F. T., Moore, R. K., and Fung, A. K., 1981. *Microwave Remote Sensing*. Norwood: Artech House, Vol. 1.

Cross-references

Calibration, Microwave Radiometers
Microwave Horn Antennas
Microwave Radiometers
Microwave Radiometers, Conventional
Microwave Radiometers, Correlation
Microwave Radiometers, Interferometers
Microwave Radiometers, Polarimeters

MICROWAVE RADIOMETERS, CONVENTIONAL

Niels Skou
National Space Institute, Technical University of Denmark, Lyngby, Denmark

Synonyms

Passive microwave radiometer (PMR)

Definition

The conventional microwave radiometer is a calibrated receiver that measures the power from the environment as picked up by an antenna system.

Introduction

The emission from the environment is received by the antenna. This can typically be horizontally or vertically polarized. The conventional radiometer measures the power corresponding to one of these polarizations, and it is basically a single channel receiver. If both polarizations must be measured, two independent radiometer receivers are employed. In-depth treatment of conventional radiometers is found in Ulaby et al. (1981) and in Skou and LeVine (2006).

Total power radiometer

The simplest and most direct implementation of a radiometer is found in the total power radiometer (TPR) as illustrated in Figure 1a. The gain in the radiometer has been symbolized by an amplifier with a gain G and the frequency selectivity by a filter with a bandwidth B (centered around some given frequency). We cannot make a receiver without introducing an additional noise T_N. The microwave power has to be detected to find some measure of its mean. In the present case, it is very attractive to use a square-law detector: then the output *voltage* will be proportional to the input *power* and hence the input temperature. Finally, we indicate where the integration takes place: the signal from the detector is smoothed by the integrator to reduce fluctuations in the output, and the longer is the integration time, the more smoothing.

The output can be expressed as

$$V_{OUT} = c \cdot (T_A + T_N) \cdot G \tag{1}$$

where c is a constant. V_{OUT} is totally dependent on T_N and G. These can for some applications not be regarded as stable enough to satisfy reasonable requirements to stability and accuracy. In other cases, however, the total power radiometer is very useful, namely, where frequent calibration, for example, once every few seconds, is possible.

The sensitivity (standard deviation of the output signal) of the total power radiometer is given by

$$\Delta T_{TPR} = \frac{T_A + T_N}{\sqrt{B \cdot \tau}} \tag{2}$$

Dicke radiometer

In 1946, R. H. Dicke found a way of alleviating the stability problems in radiometers (Dicke, 1946). By using the radiometer not to measure directly the antenna temperature T_A but rather the difference between this and some known reference temperature T_R, the sensitivity of the measurement to gain and noise temperature instabilities is greatly reduced (see Figure 1b).

The input of the radiometer is rapidly switched between the antenna temperature and the reference temperature. The switch frequency F_S could be 1,000 Hz. The output of the square-law detector is multiplied by +1 or −1, depending on the position of the Dicke switch, before integration. The input to the integrator is then $V_1 = c \cdot (T_A + T_N) \cdot G$ in one half period of F_S and $V_2 = -c \cdot (T_R + T_N) \cdot G$ in the second half period.

Provided that the switch frequency F_S is so rapid that T_A, T_N, and G can be regarded as constants over the period and that the period is much shorter than the integration time, the output of the radiometer is found as

$$V_{OUT} = c \cdot (T_A - T_R) \cdot G \tag{3}$$

It is seen that T_N has been eliminated, while G is still present, although with less weight. Now G multiplies the difference between T_A and T_R, where T_R is reasonably chosen to be in the same range as T_A, while in the total power case, G multiplied the sum of T_A and the rather large T_N.

The Dicke principle has proven to be very useful, and Dicke radiometers (DR) have been used extensively over the years.

A price has to be paid, however, for the better immunity to instabilities. Since only half of the measurement time is spent on the antenna signal (the other half is spent on the reference temperature), the sensitivity is degraded by a factor of 2 compared with the total power radiometer:

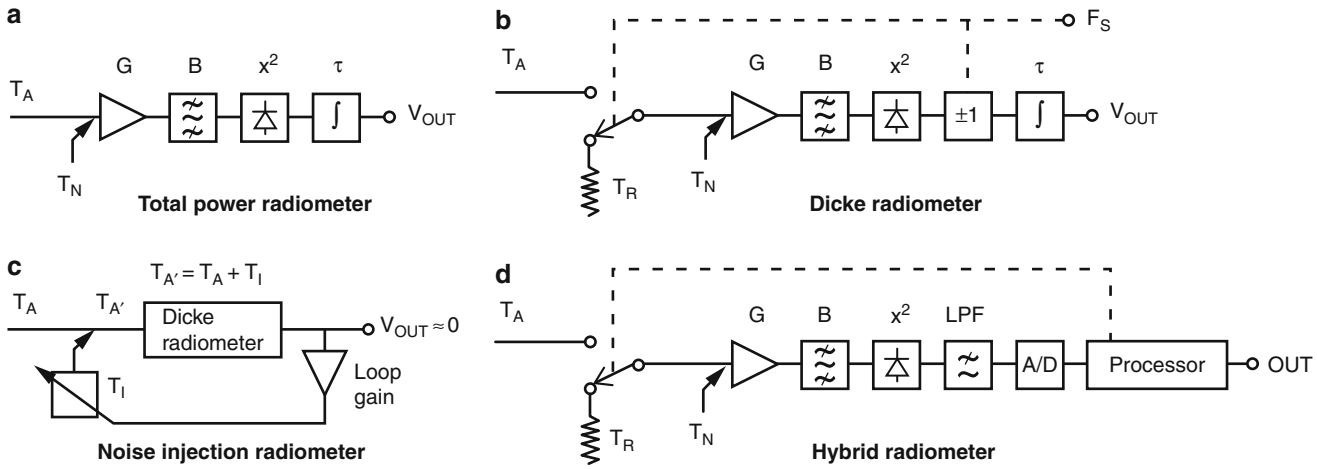
$$\Delta T_{DR} = 2 \cdot \frac{T_A + T_N}{\sqrt{B \cdot \tau}} \tag{4}$$

Noise injection radiometer

The noise injection radiometer (NIR) represents the final step toward stability; that is, the output is independent of gain and noise temperature fluctuations (Goggins, 1967; Hardy et al., 1974).

From Equation 3 it is seen that the output from a Dicke radiometer is zero (independent of G and T_N) if the reference temperature and the antenna temperature are equal. The noise injection radiometer is a specialization of a Dicke radiometer in which this condition is continuously fulfilled by a servo loop.

In almost any case encountered in Earth remote sensing, the antenna temperature is below some 300 K (emissivities between 0 and 1 are multiplied by the physical temperature). The reference temperature in a Dicke radiometer is conveniently equal to the physical



Microwave Radiometers, Conventional, Figure 1 (**a**) Total power radiometer, (**b**) Dicke radiometer, (**c**) noise injection radiometer, and (**d**) hybrid radiometer.

temperature in the microwave front end, that is, 300–320 K. In Figure 1c we show how the output T_I of a variable noise generator is added to the antenna signal T_A, so that the resultant input (T_A') to the Dicke radiometer is equal to the reference temperature (T_R) and a zero output results. A servo loop adjusts T_I to maintain the zero output condition or rather the near-zero output condition: the loop gain can be made large but not infinite.

From Equation 3 we have

$$V_{OUT} = c \cdot (T_A' - T_R) \cdot G = 0$$

and as $T_A' = T_A + T_I$, we find $T_A = T_R - T_I$. T_R is a known constant, and knowledge of T_I is required to find T_A. The accuracies of the Dicke radiometer part of the NIR and of the loop gain are (given large loop gain) completely insignificant for the accuracy with which we determine T_A. This is solely dependent on the accuracy of T_I. Accurate and stable noise sources with variable output can be made and are used for "injecting" the required signal T_I into the input line, so that T_I and T_A are added.

The sensitivity of the noise injection radiometer is very close to that of the Dicke radiometer:

$$\Delta T_{NIR} = 2 \cdot \frac{T_R + T_N}{\sqrt{B \cdot \tau}} \tag{5}$$

Hybrid radiometer

The radiometers as described hitherto in this entry are the classical receiver types, and their implementation is indicated in the classical way using, for example, analog integration after detection and analog subtraction of antenna and reference signals (in the Dicke radiometer). And indeed many radiometers are still implemented this way. But with the advent of analog to digital converters and digital processing, other implementation forms are possible and often used. This is illustrated in Figure 1d.

As soon as possible following detection, the signal is A to D converted – only a low-pass filter is indicated to condition the signal bandwidth to the sampling frequency of the converter. The signal from the converter is led to some kind of digital processor, typically a PC or an FPGA (field-programmable gate array), where suitable data handling takes place. This can typically be digital integration to the required integration time τ, as well as subtraction of the antenna signal and the reference signal. Since these processes are under computer control, flexibility becomes a key word, and the distinction between total power and Dicke radiometer vanishes to some extent: if the processor operates the input switch rapidly and regularly, we can regard it as a Dicke case, while if the measurement situation is such that the antenna signal can be measured (with interruptions when the switch points to the reference signal) without loss of data, then we have a total power case with frequent calibration.

A classical Dicke radiometer spends half of its time measuring the well-known reference temperature and thus only half of its time doing its real job, namely, measure the unknown antenna temperature. Thus over the years researchers have considered a better duty cycle for the antenna measurements (in turn potentially leading to improved sensitivity). However, having analog subtraction of the signals, the 50 % duty cycle is instrumental to having simple and stable circuitry. But with the subtraction done digitally this is no longer a limitation, and optimized duty cycles can be found on a case-by-case basis. A word of warning should, however, be stated: the naive notion that spending more time on the unknown antenna signal will lead to improved sensitivity is not necessarily true! When more time is spent on the antenna, less time is left for the reference and this in turn leads to an increased standard deviation for that measurement. Thus, the final standard deviation after subtraction might not decrease.

However, there are possible improvement schemes: since the reference temperature and the receiver noise temperature are assumed relatively stable while the antenna temperature may change rapidly, averaging over several reference temperature measurements will reduce the standard deviation of this measurement and thus allow a non-50 % duty cycle to be employed. Initial instrument fluctuations and averaging times must of course be considered carefully. See further discussion in Tanner et al. (2003) and Racette and Lang (2005).

Implementation

The radiometer is merely a very sensitive microwave receiver, and like any receiver, it can be implemented as a direct receiver or, by use of a mixer and a local oscillator, as a superheterodyne receiver. In the direct receiver, all amplification takes place at the input RF frequency, and all selectivity is determined by filters in the same frequency range. In the superheterodyne receiver most of the amplification takes place at the much lower intermediate frequency (IF), and selectivity is determined by a combination of filters at RF and IF levels.

Microwave FET amplifiers with excellent noise figures and square-law detectors covering the frequencies well beyond 40 GHz are available. Hence, the general trend is that conventional radiometers be implemented as direct receivers at frequencies below some 40 GHz, while higher-frequency implementations seem to favor the superheterodyne receiver – due to unavailability of amplifiers or very expensive amplifiers.

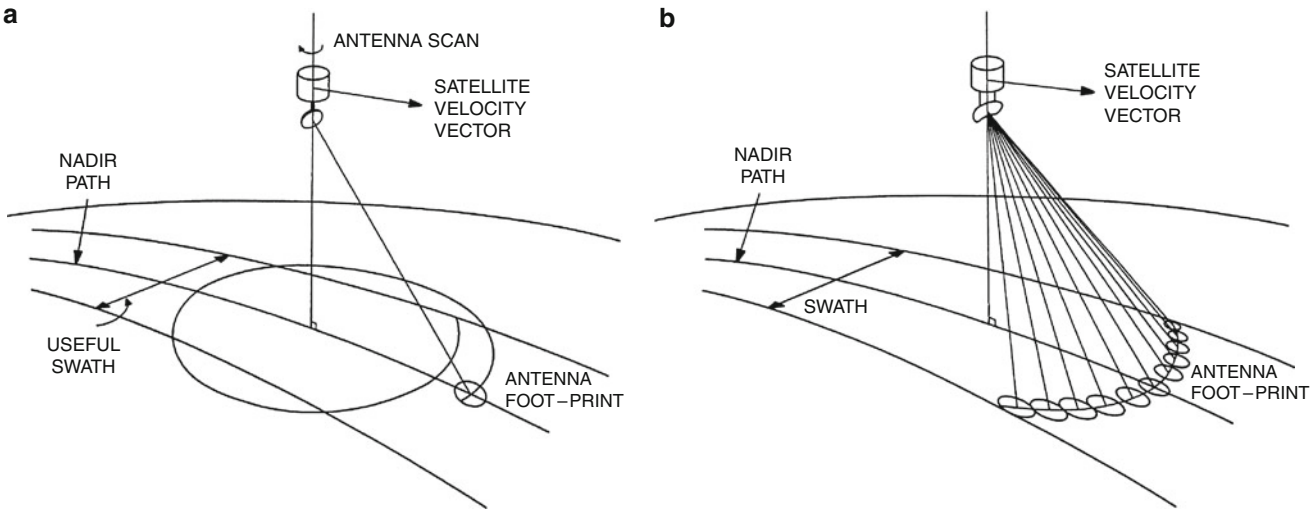
Another very important issue when implementing microwave radiometers is the concept of thermal stabilization. The best and simplest way to achieve a stable radiometer is to carefully temperature-stabilize all important components of the receiver – first and foremost the microwave front end.

Imaging considerations

A microwave radiometer is often mounted on an aircraft or a satellite in order to cover large areas.

If a radiometer with its antenna is mounted on a satellite and the antenna points straight down, the forward movement of the satellite will facilitate measurements on the ground along a straight line (the nadir path of the satellite). Coverage of the entire Earth by such "profiles" will require an enormous number of orbits! A dramatic increase in mapping efficiency results from scanning the antenna (see Figure 2a). The antenna rotates about a vertical axis, and the footprint will cover a wide swath on the Earth dependent on satellite altitude and incidence angle. Other scanning methods are possible, but the rotating scan is attractive due to constant incidence angle on the ground and the lack of accelerations associated with reciprocating scans.

In a scanning system it is obvious that there is only a limited time for the radiometer to carry out its measurement before the footprint moves to another position within the swath. The so-called dwell time per footprint is short. The users in general want small footprints (or high spatial resolution, to put it differently). As technology evolves, high-resolution systems become possible, but a small footprint results in rapid rotation (mechanical problems) and in a very small dwell time per footprint, hence in a short integration time, which, through our radiometer sensitivity formula, directly translate into poor sensitivity! The solution to this fundamental problem is offered by the so-called push-broom concept illustrated in Figure 2b.



Microwave Radiometers, Conventional, Figure 2 (**a**) Conical scan and (**b**) push broom.

In the push-broom radiometer system a multiple-beam antenna covers the swath while the satellite moves forward. A host of radiometer receivers are connected to an equal number of antenna feeds, producing individual beams to sense the Earth simultaneously.

The obvious advantages of the push-broom system compared to a scanning system are no moving antenna to cause problems in the satellite design and much larger dwell time per footprint, hence better sensitivity.

For the scanning radiometer systems, the requirement for receiver sensitivity is severe. At the same time, frequent calibration is easily achieved: once per scan, while the antenna is anyway looking away from the swath, the receiver is calibrated. Hence, the total power radiometer is an obvious candidate for such systems, due to its optimal sensitivity and since potential instabilities are taken care of by frequent calibration.

For the push-broom system, requirements for receiver sensitivity are greatly relaxed due to the much larger dwell time per footprint as compared to the scanner situation. At the same time, frequent calibration is not attractive, as all receivers are always busy sensing the Earth. Hence, the push-broom situation favors a trading of sensitivity for stability, and in conclusion, a Dicke type of switching radiometer is preferred – maybe the NIR.

Conclusion

Microwave radiometers are sensitive receivers requiring special attention to stability and accuracy. Several types have been developed over the years, and in general stability comes at a price: degraded sensitivity. So, when designing a satellite-borne imaging system, an important trade-off between sensitivity, stability, and imaging properties must be carried out.

Bibliography

Dicke, R. H., 1946. The measurement of thermal radiation at microwave frequencies. *The Review of Scientific Instruments*, **17**, 268–279.

Goggins, W. B., 1967. A microwave feedback radiometer. *IEEE-AES*, **3**, 83–90.

Hardy, W. N., Gray, K. W., and Love, A. W., 1974. An S-band radiometer design with high absolute precision. *IEEE-MTT*, **22**, 382–390.

Racette, P. E., and Lang, R. H., 2005. Radiometer design analysis based upon measurement uncertainty. *Radio Science*, **40**, 1–22.

Skou, N., and LeVine, D., 2006. *Microwave Radiometer Systems, Design and Analysis*. Norwood: Artech House.

Tanner, A. B., Wilson, W. J., and Pellerano, F. A., 2003. Development of a high-stability L-band radiometer for ocean salinity measurements. In *IEEE, Proceedings of IGARSS'03*, pp. 1238–1240.

Ulaby, F. T., Moore, R. K., and Fung, A. K., 1981. *Microwave Remote Sensing*. Norwood: Artech House, Vol. 1.

Cross-references

Calibration, Microwave Radiometers
Microwave Horn Antennas
Microwave Radiometers
Microwave Radiometers, Correlation
Microwave Radiometers, Interferometers
Microwave Radiometers, Polarimeters
Reflector Antennas

MICROWAVE RADIOMETERS, CORRELATION

Christopher Ruf
Department of Atmospheric, Oceanic and Space Sciences, University of Michigan, Ann Arbor, MI, USA

Definition and overview

A microwave radiometer measures statistical properties of the electrical field incident on its antenna. A total power radiometer, for example, measures the variance of the field strength, which is proportional to the brightness temperature of the source of the electric field. A correlation radiometer measures the statistical covariance between two incident electric fields. The most common types of correlation radiometers are autocorrelation spectrometers (Ruf and Swift, 1988), spatial interferometers (Kerr et al., 2000), and coherent detection polarimeters (Piepmeier and Gasiewski, 2001). Autocorrelation spectrometers measure the correlation between one electric field and a time-delayed version of itself. Spatial interferometers measure the correlation between the electric field at two locations. Polarimeters measure the correlation between two polarization components of an electric field. Since variance is defined as the statistical correlation of a signal with itself, a total power radiometer can be thought of as a special case of all three correlation radiometers – a spectrometer with zero time delay, an interferometer with no spatial separation, or a polarimeter with a common polarization component.

Autocorrelation spectrometer

The autocorrelation of a time-varying electric field is given by

$$R(\tau) = \langle E^*(t,\vec{r})E(t-\tau,\vec{r})\rangle \quad (1)$$

where $E(t, \vec{r})$ is the electric field at time t and position $\vec{r}$, * denotes complex conjugation, τ is the time delay between the two versions of the electric field being correlated, and $\langle\bullet\rangle$ denotes a statistical expectation operator. In practice, the expectation operation is approximated by a time average because, in most cases, second-order statistics of the electric field associated with microwave thermal emission can be considered to be stationary over short time intervals. The autocorrelation is typically sampled over a suitable range of time delays, τ. The Fourier transform of the autocorrelation with respect to the time delay is its power spectrum. For this reason, the autocorrelation of a radiometric signal is often measured in order to determine the spectral dependence of its brightness temperature. The Fourier transform that recovers the power

spectrum from the autocorrelation is performed in software as part of data post-processing. Autocorrelators are used to determine the Tb spectrum when the spectrum is to be resolved with sufficiently fine resolution that a bank of narrow band filters is impractical.

Spatial interferometer

The correlation between the electric field at two positions is given by

$$V(\vec{s}) = \langle A_1\{E(\vec{r}, \Omega)\}^* A_2\{E(\vec{r} - \vec{s}, \Omega)\}\rangle \quad (2)$$

where $E(\vec{r}, \Omega)$ is the electric field at position $\vec{r}$ arriving from angular direction Ω, $A_n\{\bullet\}$ denotes the angular reception sensitivity of antenna n ($n = 1, 2$) to the incident electric field, and $\vec{s}$ is the separation between the two versions of the electric field being correlated. This correlation statistic is referred to as the visibility of the brightness temperature distribution. The visibility is sampled over a suitable range of separations between antenna pairs. If the two antennas have overlapping angular reception sensitivities, then the Fourier transform of the visibility with respect to the separation, $\vec{s}$, is related to the angular dependence of the power density of the incident electric field received by the antennas. The power density is, in turn, proportional to the brightness temperature of the source of the electric field. Spatial interferometers (also called Fourier synthesis imagers) are used to determine the angular dependence of the brightness temperature, Tb (Ω). The Fourier transform required to convert measured visibilities to a Tb image is performed in software as part of the data post-processing. Interferometers are used to image the Tb when its angular variation is to be resolved with sufficiently fine resolution that a large antenna, capable of comparable spatial resolution, is impractical.

Coherent detection polarimeter

The correlation between two linearly polarized components of the electric field is given by

$$C_{pg} = \left\langle E_p(t, \vec{r})^* E_q(t, \vec{r})\right\rangle \quad (3)$$

where the subscripts p and q denote the two polarization components being correlated. A common polarization component was assumed in Equations 1 and 2 for the two versions of the electric field being correlated. If p and q are the same, then C_{pq} is proportional to that polarization component of the brightness temperature of the source of the electric field. If p and q are orthogonal, then the real and imaginary components of C_{pq} are proportional to the third and fourth Stokes parameters in brightness temperature. This type of correlation radiometer differs from the other two in that the correlation measurement itself, and not a Fourier transform of it, is typically the measurement of fundamental interest.

Conclusions

Correlation microwave radiometers measure the statistical covariance between different components (in time, space, or polarization) of the electric field radiated by a source of thermal emission. Certain properties of the brightness temperature associated with that source (its spectrum, its angular distribution, or its polarization state) can be determined from the correlation measurements. Correlation radiometers are typically used when determination of those properties would be otherwise difficult or impossible.

Bibliography

Kerr, Y., Font, J., Waldteufel, P., and Berger, M., 2000. The soil moisture ocean salinity mission (SMOS). *Earth Observation Quarterly*, **66**, 18.

Piepmeier, J. R., and Gasiewski, A. J., 2001. Digital correlation microwave polarimetry: analysis and demonstration. *IEEE Transactions on Geoscience and Remote Sensing*, **39**(11), 2392.

Ruf, C. S., and Swift, C. T., 1988. Atmospheric profiling of water vapor density with a 205–235 GHz autocorrelation radiometer. *Journal of Atmospheric and Oceanic Technology*, **5**(4), 539–546.

Cross-references

Microwave Radiometers, Interferometers
Microwave Radiometers, Polarimeters

MICROWAVE RADIOMETERS, INTERFEROMETERS

Manuel Martin-Neira
European Space Agency (ESA-ESTEC), Noordwijk, The Netherlands

Synonyms

Aperture synthesis radiometers; Synthetic, interferometric radiometers

Definition

Interferometry. Technique of using the mean product of two random signals to infer some characteristics of the source that generated them.

Aperture synthesis. Application of interferometry to an array of antennas to form an image of the signal source equivalent to one that would be observed by an antenna the size of the whole array.

Introduction

Microwave interferometric radiometers are a particular class of microwave radiometers, which work on the same principles as radio telescopes (Krauss, 1996; Thompson et al., 1988): interferometry and aperture synthesis. In turn, the development of the first aperture synthesis radiometers for space applications has led to a fundamental review of the very principles of radio astronomy

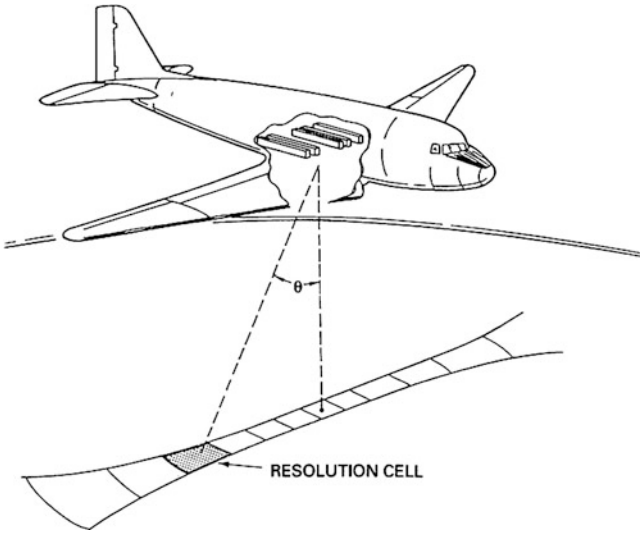
exemplified in the *Corbella equation*. Instead of being pointed to the sky, these radiometers observed the Earth surface from ground-based, airborne, and spaceborne platforms. However, interferometric radiometers present very different features by comparison with radio telescopes, namely, the small size of the antenna elements and the short spacing at which they are clustered together, both of the order of the wavelength. Aperture synthesis radiometers have been developed to achieve fine spatial resolution in those applications where a single scanning antenna is mechanically too complex to realize. Since an interferometric radiometer does not need to scan to make an image, this has been the preferred solution in those cases. The original driving scientific applications have been the mapping of soil moisture and ocean salinity from space (Swift, 1993) and, more recently, atmospheric observations from geostationary orbit.

Historical development of aperture synthesis in remote sensing

While the origin of radio astronomy dates back to 1950 (Ryle et al., 1950), its application to Earth observation was only considered in the late 1970s by the University of Berne (Schanda, 1979) and the beginning of the 1980s by engineers at NASA Goddard Space Flight Center in collaboration with the University of Massachusetts at Amherst (LeVine and Good 1983; Ruf et al., 1988; Tanner, 1990; Swift et al., 1991). The objective behind was to map the Earth's soil moisture and ocean salinity, two important geophysical parameters never measured before at global scale.

The first interferometric radiometer that was built had a synthetic beam in only one dimension, using the real aperture antenna pattern in the other. This was NASA's ESTAR (electronically steered thinned array radiometer), an aircraft demonstrator of such a hybrid instrument; see Figure 1 (LeVine et al., 1992; Swift, 1993). Subsequent developments followed elsewhere with different variations as that using the motion of the platform to save in the required number of receivers (Komiyama, 1990), this being equivalent to the use of Earth rotation in radio astronomy.

Aperture synthesis in two dimensions was developed in Europe during the 1990s. The Technical University of Denmark constructed a laboratory demonstrator (Laursen et al., 1994; Skou, 2004), and the European Space Agency (ESA) started the research of an L-band spaceborne MIRAS (Microwave Imaging Radiometer with Aperture Synthesis) (Martín-Neira, 1993; Goutoule et al., 1994; Bayle et al., 2002). ESA's study involved French scientists at the Centre d'Etudes Spatiales de la Biosphère (CESBIO) (Kerr et al., 2000) and benefited with the participation of radio astronomers from the Observatoire du Midi Pyrenees (Lannes and Anterrieu 1994; Anterrieu et al., 2002). The Polytechnic University of Catalonia (UPC) in Barcelona played an important role in defining the requirements and calibration strategy for MIRAS



Microwave Radiometers, Interferometers, Figure 1 NASA's electronically steered thinned array radiometer: the first one-dimensional airborne radiometer (Courtesy David LeVine, NASA).

Microwave Radiometers, Interferometers, Figure 2 Helsinki University of Technology's two-dimensional airborne interferometer (HUT-2D) (Courtesy Martti Hallikainen, TKK).

(Camps, 1996; Torre et al., 1996). Even more crucial was UPC's research on the completed ESA's MIRAS demonstrator, which led to the *Corbella equation* (Corbella et al., 2004), a fundamental correction to the formulation used by radio astronomers. The Helsinki University of Technology embarked in the manufacturing of an airborne two-dimensional microwave interferometer, the HUT-2D (see Figure 2) (Rautiainen et al., 1999), the first airborne two-dimensional aperture synthesis radiometer that provided good quality images of the Earth surface (Kainulainen et al., 2007). The calibration strategy of SMOS was first tested on HUT-2D (Colliander et al., 2007).

Microwave Radiometers, Interferometers, Figure 3 ESA's soil moisture and ocean salinity (SMOS) mission (Courtesy Yann Kerr, CESBIO).

In 1999, the SMOS (soil moisture and ocean salinity) mission was selected by ESA as second Earth Explorer Opportunity Mission, carrying MIRAS as only payload (McMullan et al., 2008). SMOS was launched 2 November 2009 being the first aperture synthesis radiometer flown in space; see Figure 3. It has successfully demonstrated the technique paving the way for its application in other areas. In fact, aperture synthesis has been proposed from geostationary orbit (Ruf, 1990), and higher frequency interferometers are now considered viable for Earth observation satellites flying in low Earth orbit.

In the 2000s, several other ground-based and airborne microwave interferometric radiometers have been developed by different groups around the world, as NASA Goddard's ESTAR-2D, JPL's and ESA's geostationary sounder demonstrators (Christensen et al., 2007), the C- and X-band one-dimensional interferometers of the Chinese Center of Space Science and Application Research's (Wu et al., 2005; Yan et al., 2005), or ESA's Airborne MIRAS in Europe.

Basic principles in remote sensing aperture synthesis

A microwave aperture synthesis radiometer consists of a collection of N antennas arranged in an adequate geometry for optimum imaging. Every antenna is connected to a microwave receiver, which filters out all frequencies of the incoming radiation except for a narrow band, in such a way that the output signal is quasi-monochromatic, with slowly varying but random amplitude and phase. The receiver output signal is then a random process with *Gaussian statistics* and is best represented by its complex *analytic* signal.

Each of the $N \times (N-1)/2$ possible pairs that can be formed with all the antennas of the collection is called a *baseline*. The two elements of a baseline receive the radiation from a distant target with some delay with respect to each other depending on the angle of arrival of the signal relative to the baseline. The average product of the two receiver output signals of every baseline, named *correlation* or visibility, is then computed. The visibility presents maxima and minima (null value), depending on whether the corresponding spatial delay is or not an integer number of the central wavelength. This means that some directions yield peak visibility values, while others do not contribute to it at all, what is equivalent to a spatial filtering. Every baseline acts then as a spatial filter: the longer the baseline, the shorter the spatial wavelength of the filter. Baselines oriented in different directions provide spatial filtering along those same directions. Thus, the set of all visibilities provides the spatial frequency content of the scene inside a frequency domain limited by the physical extent and location of the antenna elements. The image can then be recovered by a Fourier Transform of the visibilities. This is the formulation of the Van Cittert–Zernike theorem. Nonetheless, as explained later, this theorem would not suffice to describe the operation of an aperture synthesis radiometer, which obeys the Corbella equation instead.

Spatial resolution and windowing

By the properties of the Fourier Transform, the angular resolution of the interferometric radiometer is determined by the extension of the spatial frequency domain. Since a given baseline can be taken in one direction as well as in its opposite direction, the spatial frequency support has a size twice the maximum baseline length, this fixing the angular resolution.

The angular resolution of an aperture synthesis radiometer can now be compared with that of a real aperture instrument. An antenna of the same physical size as the synthetic radiometer would produce a far-field distribution equal to the Fourier Transform of the aperture illumination, following standard antenna theory. The square of the far field gives the radiated power that, when inverse Fourier Transformed, yields the spatial convolution of the original field illumination with itself. For a uniformly illuminated square aperture, the self-spatial convolution becomes a square pyramid with a side twice the physical length of the aperture. This is equivalent to applying a pyramidal weighting function to the visibilities collected by the synthetic radiometer.

In summary, the same spatial resolution of a real aperture with a given illumination can be achieved by an interferometric radiometer of the same physical size and a proper weighting of its visibility function. This is the principle of *aperture synthesis* or *aperture thinning*.

Real aperture radiometers never use uniform illumination as the side lobes become too high. Instead strong tapering is applied to increase the amount of energy collected through the main beam. Similarly, the visibility

samples of an aperture synthesis radiometer are multiplied by weighting functions, or *windows*, that achieve a better beam efficiency than when using a pyramidal weighting. As an example, SMOS weighs the visibility function with a *Blackman window.*

Field of view

The signal received by each element of an interferometer is affected by its antenna pattern. As the gain of the element decreases substantially outside its main beam, the radiometer becomes quite insensitive away from boresight. Imaging is therefore only feasible within the main beam of the element, which defines the field of view of the interferometric radiometer. Small antenna elements are preferred because of their wide beam, which avoids the need for mechanical scanning, a major advantage of interferometric radiometers. This is an important difference with respect to radio astronomy, where mechanically scanned large dish antennas are used. The field of view of an aperture synthesis radiometer is hence very wide and constrained by the element pattern, but not only, as explained below.

Spatial frequency sampling and aliases

The element pattern of a microwave interferometer is chosen to fit the size of the area to be imaged, or *swath*, in each particular application. For instance, an interferometer flying in a low Earth orbit at 800 km altitude will map the complete Earth in less than 3 days if its field of view is 60° wide (900 km on ground), while from geostationary orbit 18° suffices to cover the Earth disk. In matching the element pattern to the area of interest, the physical size of the element becomes defined, which in turn determines the pitch or minimum *spacing* possible between a pair of adjacent antenna elements. The resulting element spacing is usually larger than the maximum sampling period of the spatial frequency domain that guarantees the absence of grating lobes, and *aliases* appear in the image surrounding the alias-free area. The alias borders, like the element pattern, limit the field of view in aperture synthesis radiometry.

The extension of the alias zones depends on the physical arrangement of the antenna elements. For any given spacing, the alias is minimized when the antennas are placed on a hexagonal grid with a pitch the same as the spacing. Examples of such hexagonal geometries are the triangle, the hexagon, the Y (three lines at 120°), and, in general, any snowflake configuration, which lead to visibility domains with the same shapes (triangle, hexagon, a six-point star). An alternative geometry is the rectangular one, like a square, a rectangle, a cross, a U, or a T, all leading to a square or rectangular coverage in the spatial frequency domain. The maximum distances between elements (normalized to the wavelength) before aliases appear are 0.577 and 0.5 for the hexagonal and rectangular geometries, respectively, which shows the benefit of the former.

Radiometric sensitivity

The sensitivity of an aperture synthesis radiometer is comparable to that of a real aperture radiometer. As an example, consider a U-shape interferometer with N elements per arm. On the one hand, the sensitivity is proportional to the collecting area, given by the area of each antenna element a times the number of them $3N$, that is, $3Na$. On the other, the sensitivity is proportional to the square root of the integration time, or, what is equivalent, proportional to the square root of the number of resolution cells inside the field of view. In alias-free conditions, this number equals $2N$. Finally, the sensitivity is also proportional to the volume of the weighting window, normalized by the area of its base, which for a pyramidal window gives 1/3. The final sensitivity of the interferometer is thus determined by $2N^2a$. A uniformly illuminated real aperture radiometer of collecting area A has 4 resolution cells in its field of view, and its sensitivity is thus proportional to $2A$. If both instruments have the same physical size, $A = N^2a$, the sensitivity of the real aperture radiometer becomes $2N^2a$, the same as that of the aperture synthesis radiometer.

The Corbella equation

The Van Cittert–Zernike theorem, basis of radio astronomy, is not compatible with the Bosma theorem (Wedge and Rutledge 1991). This became apparent for the first time within ESA's MIRAS Demonstrator Pilot Project in 2002. A two-dimensional aperture synthesis radiometer was placed inside an anechoic chamber, and all measured visibilities were zero, as expected from the Bosma theorem. This theorem states that for any passive network (as an array of antennas inside an anechoic chamber) in thermal equilibrium with its terminations (as the isolators following the antennas), the cross-correlation of the outcoming noise waves must be null. But according to the Van Cittert–Zernike theorem, the measured correlations should have been equal to the Fourier Transform of the element antenna pattern. The theoretical foundations of aperture synthesis were revised and a new equation established, compatible with the Bosma theorem, which correctly explains aperture synthesis, and by extension, radio astronomy: the *Corbella equation*. According to this equation, the visibility function is proportional to the difference or *contrast* in physical temperature between the target and the instrument. High correlations are expected when looking to the cold sky, and very small correlations when inside an anechoic chamber at similar physical temperature as the instrument. The Corbella equation has been verified from ground, airborne and space.

The flat target transformation

According to the Corbella equation, any uncertainty in the knowledge of the antenna pattern is amplified by the instrument–target temperature contrast. Thus, it is desirable to minimize such contrast. This is possible if the *flat target response* (FTR) of the interferometer is acquired

first. The FTR is the visibility function measured when imaging an unpolarized uniform brightness temperature distribution, which is stable over time. In practice, the cold sky near the galactic poles of the Milky Way is a good approximation of such a target, and the FTR can be measured by pointing the aperture synthesis radiometer to the galactic poles. The FTR is like a mask formed by ripples due to system imperfections, mainly antenna errors. When this mask is scaled down to the temperature contrast between the instrument and an Earth scene, and is subtracted from the measured visibilities, the instrument errors are cancelled to a large extent. Mathematically, the Corbella equation is transformed by this linear combination with the result that the instrument physical temperature is replaced by the average temperature of the scene. The final temperature contrast is then reduced to that between each pixel of the scene and the average of all the pixels. The transformation of the Corbella equation using the FTR is known as the *flat target transformation* (FTT) and is an essential step in aperture synthesis imaging (Martín-Neira et al., 2008).

Polarimetry in aperture synthesis

An aperture synthesis radiometer is inherently a polarimetric type of instrument because of its very wide field of view. Horizontal *H* and vertical *V* polarizations on the target are transformed into *v* and *h* polarizations on the element antenna according to a different rotation matrix depending on the direction. Since the *H* and *V* fields appear mixed together in the *v* and *h* fields in the instrument polarization frame, a nonzero *vh* cross brightness temperature is measured even from a target with no correlation between *H* and *V*. Reversely, to recover the horizontal and vertical brightness temperatures of the target, it is necessary to measure the *vh* cross brightness temperatures. When these are not available, a singularity along the image diagonals appears, and pixels in those areas have to be discarded. They can be recovered though when, due to the platform motion, they move to directions outside the singularities.

Calibration

An aperture synthesis radiometer is a complex instrument and requires several calibration steps. The first one is the flat target transformation, explained above, for which a cold sky view to acquire the flat target response must have been performed before hand. The second step is the correction of any comparator offset in the digital circuits and quadrature error of the demodulator. Then comes the phase correction, achieved by injecting correlated noise to all receivers internally, and the amplitude calibration, based on two standard temperatures, as the cold sky and an internal matched load at a well-monitored physical temperature. Finally, an uncorrelated load is used to remove any correlation offsets due to residual internal interference. In addition, to form an image, the antenna patterns must have been characterized a priori, and the *fringe-washing function* has to be determined during operation. The fringe-washing function takes into account the decorrelation of the signal with time delay across the array and is estimated from early, punctual, and late correlations.

Summary and conclusions

Aperture synthesis radiometry has developed starting in the early 1980s mainly driven by the need to map soil moisture and ocean salinity from space. During this development, the theoretical fundamentals had to be revised, and a new formulation, the Corbella equation, was found, which describes correctly the operation of this type of instruments, and, in fact, of radio telescopes as well. A consequence of the new formulation is the flat target transformation, a powerful method to calibrate out antenna errors in aperture synthesis.

The spatial resolution and sensitivity of an aperture synthesis radiometer are comparable to those of a real aperture radiometer. Field of view, collecting area, and integration time take different values in each case, but balance each other, yielding the same net result at the end.

ESA's SMOS mission carrying the first aperture synthesis radiometer into space was launched 2 November 2009. Scientists have been able to assess the ultimate benefits of this new remote sensing technique. Following SMOS success, a next generation of interferometric radiometers might follow not only from low Earth orbit but also from geostationary orbit, already under study.

Bibliography

Anterrieu, E., Waldteufel, P., and Lannes, A., 2002. Apodization functions for 2-D hexagonally sampled synthetic aperture imaging radiometers. *IEEE Transaction in Geoscience and Remote Sensing*, **40**(12), 2531–2542.

Bayle, F., Wigneron, J.-P., Kerr, Y. H., Waldteufel, P., Anterrieu, E., Orlhac, J.-C., Chanzy, A., Marloie, O., Bernardini, M., Sobjaerg, S., Calvet, J.-C., Goutoule, J.-M., and Skou, N., 2002. Two-dimensional synthetic aperture images over a land surface scene. *IEEE Transaction in Geoscience and Remote Sensing*, **40**(3), 710–714.

Camps, A., 1996. *Application of Interferometric Radiometry to Earth Observation*. PhD thesis, Barcelona, Polytechnic University of Catalonia.

Christensen, J., Carlstrom, A., Ekstrom, H., Emrich, A., Embretsen, J., De Maagt, P., and Colliander, A., 2007. GAS: the geostationary atmospheric sounder. In *IGARSS-2007 Proceedings*, Barcelona, pp. 223–226.

Colliander, A., Lemmetyinen, J., Uusitalo, J., Suomela, J., Veijola, K., Kontu, A., Kemppainen, S., Pihlflyckt, J., Rautiainen, K., Hallikainen, M., and Lahtinen, J., 2007. Ground calibration of SMOS: NIR and CAS. In *IGARSS Proceedings*, pp. 3631–3634.

Corbella, I., Duffo, N., Vall-llossera, M., Camps, A., and Torres, F., 2004. The visibility function in interferometric aperture synthesis radiometry. *IEEE Transactions on Geoscience and Remote Sensing*, **42**(8), 1677–1682.

Goutoule, J. M., Kraft, U., and Martin-Neira, M., 1994. MIRAS: preliminary concept of a two-dimensional L-band aperture synthesis radiometer. In *MicroRad'94 Proceedings*, Rome.

Kainulainen, J., Rautiainen, K., Tauriainen, S., Auer, T., Kettunen, J., and Hallikainen, M., 2007. First 2D interferometric

radiometer imaging of the Earth from an aircraft. *IEEE Transactions on Geoscience and Remote Sensing Letters*, **4**(2), 241–245.
Kerr, Y. H., Wigneron, J. P., Ferrazzoli, P., and Waldteufel, P., 2000. Soil moisture and vegetation biomass retrievals using L-band, dual polarised and multi angular radiometric data in preparation of the SMOS mission. In *IGARSS-2000 Proceedings*, Hawaii, Vol. 3, pp. 1244–1246.
Komiyama, K., 1990. *Super-synthesis radiometer (SSR) for the remote sensing of the Earth*. Technical Report. Ibaraki: Electrotechnical Laboratory.
Krauss, J. D., 1996. *Radio Astronomy*. New York: McGraw-Hill.
Lannes, A., and Anterrieu, E., 1994. Image reconstruction methods for remote sensing by aperture synthesis. In *IGARSS-94 Proceedings*, Pasadena, Vol. 4, pp. 2228–2230.
Laursen, B., and Skou, N., 1994. A spaceborne synthetic aperture radiometer simulated by the TUD demonstrator model. In *IGARSS-94 Proceedings*, Pasadena, Vol. 3, pp. 1314–1316.
LeVine, D. M., and Good, J. C., 1983. *Aperture synthesis for microwave radiometers in space*. NASA Technical Memorandum 85033. Greenbelt, MD: Goddard Space Flight Center.
LeVine, D. M., Griffis, A., Swift, C. T., and Jackson, T. J., 1992. ESTAR: a synthetic aperture microwave radiometer for measuring soil moisture. In *IGARSS-92 Proceedings*, Houston, TX, Vol. II, pp. 1755–1757.
Martín-Neira, M., 1993. MIRAS: a two-dimensional passive aperture synthesis radiometer. In *Presentation at PIERS'93*. Pasadena, CA: JPL.
Martín-Neira, M., Suess, M., Kainulainen, J., and Martín-Porqueras, F., 2008. The flat target transformation. *IEEE Transactions on Geoscience and Remote Sensing*, **46**(3), 613–620.
McMullan, K., Brown, M., Martin-Neira, M., Rits, W., Ekholm, S., Marti, J., and Lemanczyk, J. 2008. SMOS: the payload. *IEEE TGARS*, **46**, 594–605.
Rautiainen, K., Valmu, H., Jukkala, P., Moren, G., and Hallikainen, M., 1999. Four-element prototype of the HUT interferometric radiometer. In *IGARSS-99 Proceedings*, Hamburg, Vol. 1, pp. 234–236.
Ruf, C. S., 1990. Antenna performance for a synthetic aperture microwave radiometer in geosynchronous Earth orbit. In *IGARSS-90 Proceedings*, pp. 1589–1592.
Ruf, C. S., Swift, C. T., Tanner, A. B., and LeVine, D. M., 1988. Interferometric synthetic aperture microwave radiometry for the remote sensing of the Earth. *IEEE Transactions on Geoscience and Remote Sensing*, **26**(5), 597–611.
Ryle, M., Smith, F. G., and Elsmore, B., 1950. *A Preliminary Survey of the Radio Stars in the Northern Hemisphere*. London: Royal Astronomical Society.
Schanda, E., 1979. Multiple wavelength aperture synthesis for passive sensing of the Earth's surface. In *International Symposium Digest*. Seattle, WA: IEEE Antennas and Propagation Society, p. 762.
Skou, N., 2004. Spaceborne L-band radiometers: pushbroom or synthetic aperture? In *IGARSS-04 Proceedings*, Vol. 2, pp. 1264–1267.
Swift, C. T., 1993. *ESTAR – The Electronically Scanned Thinned Array Radiometer for Remote Sensing Measurement of Soil Moisture and Ocean Salinity*. Goddard Space Flight Center, Washington DC: NASA, p. 4523.
Swift, C. T., Le Vine, D. M., and Ruf, C. S., 1991. Aperture synthesis concept in microwave remote sensing of the Earth. *IEEE Transactions on Microwave Theory and Techniques*, **39**, 1931–1935.
Tanner, A. B., 1990. *Aperture Synthesis for Passive Microwave Remote Sensing: The Electronically Scanned Thinned Array Radiometer*. PhD thesis, Amherst, University of Massachusetts.
Thompson, A. R., Moran, J. M., and Swenson, G. W., 1988. *Interferometry and Synthesis in Radio Astronomy*. New York: Wiley.
Torre, F., Camps, A., Bara, J., Corbella, I., and Ferrero, R., 1996. On-board phase and modulus calibration of large aperture synthesis radiometers: study applied to MIRAS. *IEEE Transactions on Geoscience and Remote Sensing*, **34**(4), 1000–1009.
Wedge, S. W., and Rutledge, D. B., 1991. Noise waves and passive linear multiports. *IEEE Microwave and Guided Wave Letters*, **1**(5), 117–119.
Wu, J., Liu, H., Yan, J., Ban, S., Dong, X., and Jiang, J., 2005. Research activity on synthetic aperture radiometry in CSSAR/CAS. In *PIERS 2005 Proceedings*, Hangzhou.
Yan, J., Wu, J., Liu, H., Dong, X., and Jiang, J., 2005. Design and implementation of digital correlator for CAS synthetic aperture radiometer. In *PIERS 2005 Proceedings*, Hangzhou.

Cross-references

Microwave Radiometers
Microwave Radiometers, Conventional
Microwave Radiometers, Correlation
Microwave Radiometers, Polarimeters
Observational Systems, Satellite
Sea Surface Salinity
Soil Moisture

MICROWAVE RADIOMETERS, POLARIMETERS

David Kunkee
The Aerospace Corporation, Los Angeles, CA, USA

Synonyms

Microwave polarimeter; Polarimetric radiometer

Definition

Microwave Polarimetric Radiometer. Microwave radiometer capable of measuring at least three of the four modified Stokes' parameters of the incident field.
Polarimetry. Measurement of the polarization characteristics of incident electromagnetic radiation.

Introduction

Polarimetric microwave radiometers are capable of measuring at least three Stokes' parameters of the incident microwave radiation or radiometric scene. Traditional dual-polarized radiometers measure vertically and horizontally polarized brightness temperatures or, equivalently, the first two Stokes' parameters of the input scene. In order to completely characterize the spatially averaged input scene, measurements of third and fourth Stokes' parameters are required (Jackson, 1975), and these measurements can be performed utilizing coherent (direct) or incoherent techniques. Direct measurements are performed by polarization-correlating radiometers to measure the third and/or fourth Stokes' parameters. Indirect methods include a combination of linearly polarized brightness temperatures typically oriented at 45° and −45° with respect to the principle directions in order to derive the third Stokes' parameter and by combing left-and right-hand circular polarized brightness temperatures to derive the fourth

Stokes' parameter. Several airborne microwave polarimeters have been part of airborne experiments conducted between 1993 and 2003 in order to investigate application of microwave polarimetry to retrieval of sea surface wind direction (Yueh et al., 1995). The satellite-based wind direction system (WindSat), developed by the US Naval Research Laboratory (NRL) and launched in January 2003, performed the first space-based polarimetric measurements (Gaiser et al., 2004).

Background

Airborne and space-based polarimetric radiometers measure the intensity of the Earth's naturally occurring upwelling thermal emission. This arbitrarily polarized, non-or partially coherent thermal radiation is completely characterized by the four parameter-modified Stokes' vector:

$$\overline{I} = \begin{pmatrix} I_v \\ I_h \\ U \\ V \end{pmatrix} \frac{W}{(m^2 \cdot sr \cdot Hz)}$$

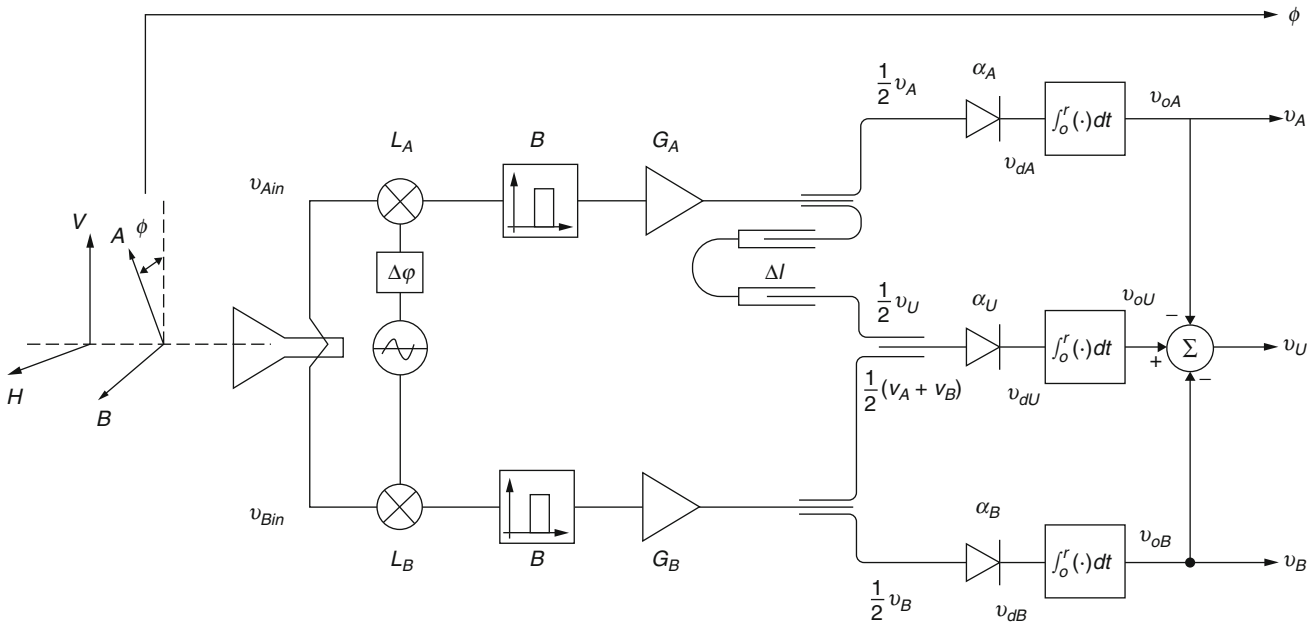
where I_v and I_h are the spectral intensities of the vertically and horizontally polarized field components and U and V are the in-phase and quadrature correlations, respectively, between the orthogonally polarized field components $\hat{v}$ and $\hat{h}$. The spectral intensity of this emission can be described by the brightness temperature vector, $\overline{T}_B$:

$$\overline{T}_B \equiv \frac{\lambda^2}{k}\overline{I} = \begin{pmatrix} T_v \\ T_h \\ T_U \\ T_V \end{pmatrix} = \begin{pmatrix} \langle |E_v|^2 \rangle \\ \langle |E_h|^2 \rangle \\ 2\mathrm{Re}\{\langle E_v E_h^* \rangle\} \\ 2\mathrm{Im}\{\langle E_v E_h^* \rangle\} \end{pmatrix}$$

where E_v and E_h represent the vertically and horizontally polarized electric fields, respectively, $\langle \cdot \rangle$ represents a time-averaged quantity. Microwave polarimeters measure at least three of the four parameters. There are two general methods in passive microwave remote sensing in order to measure the third and fourth Stokes' parameter: (1) cross-correlation between two orthogonally polarized measurements (polarization-correlating radiometer) and (2) measurements of polarized brightness temperature measurements using multiple polarizations. For example, using linearly polarized brightness temperatures at orientations of $\pm 45^\circ$ with respect to the orientation of the vertical polarization vector, the third Stokes' parameter can be found simply by differencing the two measurements: $T_U = T_{45} - T_{-45}$. Similar approaches can be used to measure T_V. One of the first reports on radiometer designs for sea surface microwave emission polarimetry was provided by Dzura (1992).

Polarization-correlating radiometer (direct method)

An example of an analog polarization-correlating radiometer is given below. Figure 1 shows the block diagram of a three-channel polarization-correlating radiometer.



Microwave Radiometers, Polarimeters, Figure 1 Block diagram of a three-channel polarization-correlating radiometer (Gasiewski and Kunkee, 1993).

The measurement sensitivity for v_U (and therefore T_U) is proportional to the geometric mean of the characteristics of v_A and v_B; hence, if channels A and B are identical, channel U will have the same sensitivity (ΔT_{RMS}) as channels A and B. Phase adjustments are made by adjusting the relative oscillator phase ($\Delta\phi$) and/or relative line length (Δl) before the combiner. The phase can be adjusted to provide response to in-phase (U) or phase quadrature (V) components; however, in order to measure V (fourth Stokes' parameter), a single-side band configuration must be used in order to avoid cancelation of the signal from the contributions of the high and low side bands.

The post-detection analog signal summation in the scheme shown in Figure 1 is typically performed by using digital electronic circuitry (Piepmeier and Gasiewski, 2001).

Polarization combining radiometer (indirect method)

An alternate method is to combine brightness temperature measurements at multiple polarizations to derive U or V indirectly. This method is also called the incoherent method and is typically applied through measurements of linearly polarized brightness temperatures offset ±45° from the original basis to derive the third Stokes' parameter T_U and using left- and right-hand-polarized brightness temperatures to derive the fourth Stokes' parameter, T_V. Using a unitary rotational transform, the relationship between linearly polarized brightness temperatures rotated φ degrees is expressed by

$$\overline{T'} = \begin{pmatrix} T_{v'} \\ T_{h'} \\ T_{U'} \\ T_{V'} \end{pmatrix} = \begin{bmatrix} \cos^2\varphi & \sin^2\varphi & 0.5\sin^2\varphi & 0 \\ \sin^2\varphi & \cos^2\varphi & -0.5\sin^2\varphi & 0 \\ -\sin^2 2\varphi & \sin^2 2\varphi & \cos 2\varphi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} T_v \\ T_h \\ T_U \\ T_V \end{pmatrix} = \overline{\overline{U}}(\varphi)\overline{T}$$

where $\overline{T'}$ is the measurement made with the rotated polarization basis. Therefore, by using $T_{\pm 45} = T_{u'}(\varphi = \pm 45^\circ)$, the third Stokes' parameter, $T_U = T_{45} - T_{-45}$. A popular way of deriving brightness temperature measurements at ± 45° is through the use of polarization combining networks (Yueh et al., 1995). Note that in the above equation, the fourth Stokes' parameter, T_V, is rotationally invariant; however, it can be shown that $T_V = T_R - T_L$ where T_R and T_L are the right- and left-hand circularly polarized brightness temperatures of the incident wave, respectively (Tsang et al., 1985).

Ocean surface wind measurements

One of the primary applications of microwave polarimetry is to improve ocean surface wind (OSW) retrievals performed with dual-polarized, T_v and T_h, brightness temperatures at 18 and 37 GHz. Dual-polarized radiometric measurements from space-based radiometers are currently and routinely used to accurately observe ocean surface wind speed to better than 1 m/s accuracy, Wentz (1997). The relationship between wind-roughened ocean surfaces and upwelling brightness temperature is well understood and was first explored by Hollinger (1971). Ocean surface wind speed measurements have been performed from operational microwave radiometers since the first special sensor microwave/imager (SSM/I) radiometer was launched and began operating in 1987 (Goodberlet et al., 1990: see also "*Ocean, Measurements and Applications*").

The first measurements showing a variation of brightness temperature as a function of the relative angle between the surface wind direction and the polarization plane were reported by Bespalova et al. (1979) as part of an airborne radiometry experiment conducted by the Space Research Institute at Moscow. A study by Wentz (1992) showed that there was a systematic bias in wind speed retrievals from SSM/I compared to ocean buoys that was dependent on the angle of observation with respect to the surface wind direction. The directional dependence was used to demonstrate that wind direction could also be retrieved using microwave brightness temperatures. However, the presence of multiple solutions, or directional ambiguities, limits the utility of wind direction derived using dual-polarized microwave brightness temperatures.

Building upon theoretical modeling and laboratory measurements of polarized microwave emission from wave-covered surfaces (Yueh et al., 1994; Gasiewski and Kunkee, 1994), the utility of the third Stokes' parameter for ocean surface wind direction retrieval was demonstrated through airborne experiments that measured polarimetric brightness temperatures over the ocean (Yueh et al., 1995). In order to demonstrate wind direction retrievals from space, the WindSat microwave polarimetric radiometer was developed by the US Naval Research Laboratory and launched in 2003 (Gaiser et al., 2004). Wind vector (speed and direction) retrievals using WindSat data (Bettenhausen et al., 2006) are now assimilated into numerical weather forecasting (NWP) models and used to improve forecasts (Candy et al., 2009).

Other applications

Measurements of the third and fourth Stokes' parameters by WindSat over the Greenland ice sheet have shown responses related to the asymmetrical features of polar ice sheets and potentially the microphysical structure of snow crystals (Li et al., 2008). Over desert, polarimetric measurements are also sensitive to the asymmetrical features of the surface such as the structure of sand dunes (Narvekar et al., 2007). Investigations have also shown the WindSat polarimetric measurements to be highly sensitive to RFI signals, and in some cases, showing responses before contamination is apparent in the vertically or horizontally polarized channels (Ellingson and Johnson, 2006; see also "*RFI*").

Summary

Microwave polarimetry involves the measurement of the third and fourth Stokes' parameter of the input radiometric "scene." Traditional dual-polarized radiometers measure only the first two parameters of the Stokes' vector and miss some characteristics of the brightness temperature scene. Microwave polarimeters can be designed for direct (coherent) polarimetric measurements or indirect (incoherent) measurements that are derived through post-detection combination of polarized measurements. Motivation to develop airborne and space-based microwave polarimeters is provided by improvements to passive ocean surface wind vector measurements. The WindSat radiometer was the first space-based microwave polarimeter, providing global measurements of all four Stokes' parameters. These measurements are used to retrieve ocean surface wind vectors and improve weather forecast quality of NWP models. Other potential applications of microwave polarimetry are under development and include snow and ice characterization, land monitoring, and RFI detection and characterization.

Bibliography

Bespalova, E. A., Veselov, V. M., Glotov, A. A., et al., 1979. Sea ripple anisotropy estimates from variations in polarized thermal emission of the sea. *Doklady Akademii Nauk SSSR*, **246**(6), 1482–1485.

Bettenhausen, M. H., Smith, C. K., Bevilacqua, R. M., Wang, N., and Gaiser, P. W., 2006. A nonlinear optimization algorithm for WindSat wind vector retrievals. *IEEE Transactions on Geoscience and Remote Sensing*, **44**(3), 597–609.

Candy, B., English, S. J., and Keogh, S., 2009. A comparison of the impact of quickscat and windsat wind vector products on met office analyses and forecasts. *IEEE Transactions on Geoscience and Remote Sensing*, **47**(6), 1632–1640.

Dzura, M. S., Etkin, V. S., Khrupin, A. S., Pospelov M. N., and Raev M. D., 1995. Radiometers-polarimeters: principles of design and applications for sea surface microwave emission anisotropy. In *IEEE International Geoscience and Remote Sensing Symposium, 1995*. IGARSS'95, Vol. 2, pp. 1432–1434.

Ellingson, S. W., and Johnson, J. T., 2006. A polarimetric survey of radio frequency interference in C-and X-bands in the continental United States using windsat radiometry. *IEEE Transaction on Microwave Theory and Techniques*, **44**(3), 540–548.

Gaiser, P. W., et al., 2004. WindSat spaceborne polarimetric radiometer: sensor description and early orbit performance. *IEEE Transaction on Microwave Theory and Techniques*, **42**(11), 2347–2361.

Gasiewski, A. J., and Kunkee, D. B., 1993. Calibration and applications of polarization correlating radiometers. *IEEE Transaction on Microwave Theory and Techniques*, **41**(5), 767–773.

Gasiewski, A. J., and Kunkee, D. B., 1994. Polarized microwave emission from water waves. *Radio Science*, **29**(6), 1449–1466.

Goodberlet, M. A., Swift, C. T., and Wilkerson, J. C., 1990. Ocean surface wind speed measurements of the special sensor microwave/imager (SSM/I). *IEEE Transactions on Geoscience and Remote Sensing*, **28**(5), 823–828.

Hollinger, J. P., 1971. Passive microwave measurements of sea surface roughness. *IEEE Transactions on Geoscience Electronics*, **GE-9**(3), 165–169.

Jackson, J. D., 1975. *Classical Electrodynamics*. New York: Wiley, p. 848.

Li, L., Gaiser, P. W., Albert, M. R., Long, D. G., and Twarog, E. M., 2008. WindSat passive polarimetric signatures of the greenland ice sheet. *IEEE Transactions on Geoscience and Remote Sensing*, **46**(9), 2622–2631.

Narvekar, P. S., Jackson, T. J., Bindlish, R., Li, L., Heygster, G., and Gaiser, P. W., 2007. Observations of land surface passive polarimetry with the windsat instrument. *IEEE Transactions on Geoscience and Remote Sensing*, **45**(7), 2019–2028.

Piepmeier, J. R., and Gasiewski, A. J., 2001. Digital correlation microwave polarimetry: analysis and demonstration. *IEEE Transactions on Geoscience and Remote Sensing*, **39**(11), 2392–2410.

Tsang, L., Kong, J. A., and Shin, R. T., 1985. *Theory of Microwave Remote Sensing*. New York: Wiley, p. 613.

Wentz, F. J., 1992. Measurement of oceanic wind vector using satellite microwave radiometers. *IEEE Transactions on Geoscience and Remote Sensing*, **30**(5), 960–972.

Wentz, F. J., 1997. A well-calibrated ocean algorithm for special sensor microwave/imager. *Journal of Geophysical Research*, **102**(C4), 8703–8718.

Yueh, S. H., Nghiem, S. V., Kwok, R., Wilson, W. J., Li, F. K., Johnson, J. T., and Kong, J. A., 1994. Polarimetric thermal emission from periodic water surfaces. *Radio Science*, **29**, 87–96.

Yueh, S. H., Wilson, W. J., Li, F. K., Nghiem, S. V., and Ricketts, W. B., 1995. Polarimetric measurements of sea surface brightness temperatures using an aircraft K-band radiometer. *IEEE Transactions on Geoscience and Remote Sensing*, **33**(1), 85–92.

MICROWAVE SUBSURFACE PROPAGATION AND SCATTERING

Alexander Yarovoy
Delft University of Technology, Delft, The Netherlands

Definitions

Subsurface. Natural materials (soils, rock, snow, ice) below air–ground interface.

Ground-penetrating radar. Radar for subsurface sensing. It locates, images, and characterizes changes in electrical and magnetic properties of subsurface materials.

Introduction

The first description of microwaves use for subsurface sensing is attributed to a German patent by Leimbach and Löwy from 1910 (Daniels, 2004). In this patent, propagation of microwaves between pairs of vertically buried dipole antennas has been used to detect any subsurface objects with higher conductivity than the surrounding medium. Only monochromatic electromagnetic waves have been considered in this patent. The first use of electromagnetic pulses with a broad spectrum to determine the structure of buried objects is attributed to Hülsenbeck (Hülsenbeck et al., 1926). It was noted that any dielectric variation, not necessarily involving conductivity, would also produce reflections. The first ever experiments with subsurface microwave propagation to sound the depth of a glacier were performed in Austria in 1929 (Stern, 1929, 1930). These experiments were largely forgotten

until the late 1950s when US Air Force radars were seeing through ice as planes tried to land in Greenland but misread the altitude and crashed into the ice. This accident together with lunar soil investigations within Apollo program (Simmons et al., 1972) triggered research on microwave penetration into subsurface, including not only ice sounding but also mapping soil properties and the water table. This fundamental research and simultaneous development of ultra-wideband microwave technology resulted, in the early 1970s, in wide use of microwave subsurface propagation and scattering for characterization and mapping of subsurface. Since then, the range of applications has been expanding steadily.

Microwave Subsurface Propagation and Scattering, Table 1 Typical values of relative dielectric permittivity of some dry soils and rocks at microwave frequencies (Daniels, 2004)

Material	Relative dielectric permittivity
Sand	2.4–6
Loam	4–10
Clay	2–6
Granite	5
Limestone	7
Sandstone	2–5
Silt	3–10

Dielectric properties of soil, rock, ice, and snow

Propagation velocity and attenuation of microwaves is controlled by the electrical and magnetic properties of subsurface. Regarding the subsurface, three different types of natural earth materials should be distinguished: rocks, soils, and ice/packed snow. At microwave frequencies, electrical properties (which in most cases are more important than the magnetic properties) are dominantly controlled by density and by the chemistry, fine structure (composition of liquid/gas/solid components), and content of water. Below we concentrate mainly on rock and soil properties.

Following Heimovaara et al. (1994), soils can be considered as a four-component system: irregularly shaped solid particles, air, free water, and bound water. The bound water refers to the first few (up to 10) molecular layers of water near solid surfaces that are rotationally hindered by surface forces. The frequency-dependent complex dielectric permittivity of soils can be described then with a four-component complex dielectric mixing model based on the volumetric mixing of the refractive indices of the soil components. It should be noted that free water and bound water have different dielectric permittivities and thus different impact of the electrical properties of the material. Similarly to this approach, Wobschall (1977) has considered rocks as a three-phase system: irregularly shaped particles, air-or water-filled voids (pores), and crevices. In both cases, presence of moisture strongly influences electrical properties of material.

Solid particles of soils and rocks have different physical properties depending on their material. In the most simple and frequently used approach (Wang and Schmugge, 1980; Dobson, et al., 1985), the soil solid particles are considered to be a mixture of sand particles of diameter $d > 0.005$ cm; silt, $0.0002\ \text{cm} < d < 0.005$ cm; and clay, $d < 0.0002$ cm, weight content of which is expressed in percentage of total weight of soil. Dielectric permittivities of dry sand, clay, and silt are slightly different (see Table 1). More important is that clay, silt, and sand have different abilities to bind the water. It is shown in Wang and Schmugge (1980) that the quantity of bound water in soil depends on the volume of clay fraction in it and the quantity of bound water increases with the volume of clay. This is explained by a large specific area of clay surface compared to other soil fractions. Sand and silt particles are also covered with bound water films; however, the amount of bound water on sand and silt particles is less than 0.1 % (Bojarskii et al., 2002). Further discussions on dielectric properties of soils and rocks can be found in Hoekstra and Delaney (1974), Hipp (1974), De Loor (1983), and Hallikainen et al. (1985).

Providing a deep insight in the interaction of microwaves with rocks and soils, the abovementioned approaches require detailed knowledge of many physical parameters of soils, which are often not known during the field work. Thus, a large number of approximate equations have been proposed to compute dielectric permittivity of soil as a mixture of different components (Shutko and Reutov, 1982; Dobson et al., 1985; Sen et al., 1981). Among them, the so-called Bruggeman–Hanai–Sen (BHS) mixing model (Sen et al., 1981) is widely used nowadays:

$$p = \frac{(\varepsilon_m - \varepsilon_r)\left(\frac{\varepsilon_w}{\varepsilon_r}\right)^C}{\varepsilon_m - \varepsilon_w}$$

where ε_r is bulk soil relative dielectric permittivity, ε_m is the dielectric permittivity of dry soil, ε_w is the dielectric permittivity of water, p is fractional porosity (volume of voids/total volume), and C is a shape factor (1/3 for spherical grains). When the water content is large, wet rock or soil can be considered simply as dry solid/water mixture and simply empirically stated Topp equation (Topp et al., 1980) is frequently used:

$$\varepsilon_r = 3.03 + 9.3 \cdot \theta + 146.0 \cdot \theta^2 - 76.3 \cdot \theta^3$$

where ε_r is bulk soil relative dielectric permittivity and θ is volumetric water content. Increase of the water content in general is responsible for increase of relative dielectric permittivity of the material.

In order to use different dielectric mixture formulas, knowledge of dielectric permittivity of water and dry soils is required. The complex-valued dielectric permittivity of bulk water at microwave frequencies can be described by

the Debye theory and the Cole–Cole model (Cole and Cole, 1941; Heimovaara, 1994) is frequently used:

$$\varepsilon(\omega) = \varepsilon' - j\varepsilon'' = \varepsilon_\infty + \frac{\varepsilon_s - \varepsilon_\infty}{1 + j(w\tau)^{1-\beta}} - j\frac{\sigma_{dc}}{w\varepsilon_0}$$

where ε_s is the low-frequency (static) dielectric permittivity of water (which is of about 88 at a temperature of 0 °C), ε_∞ is the high (infinite)-frequency dielectric permittivity and equals 4.25 for the 0 °C, τ is the relaxation time of water and $\beta = 0.0125$ and is a factor that accounts for the possible spread in relaxation frequencies, σ_{dc} is the DC conductivity at a given temperature, and ε_0 is the permittivity of free space, which is $8.854 \cdot 10^{-12}$ F/m. The relaxation time of water molecules, which equals $7.7 \cdot 10^{-12}$ s at 27 °C, drastically increases due to surface forces, and in monomolecular layers of water covering, the soil particles can reach the value of $5.0 \cdot 10^{-10}$ s at 27 °C (Bojarskii et al., 2002). As a result, the bound water exhibits relaxation at frequencies below 300 MHz and has, at microwave frequencies, considerably lower values for the dielectric permittivity than free water. The permittivity of bound water is temperature dependent, resulting in a substantial increase in bulk dielectric permittivity with temperature at microwave frequencies for soils with high surface area (Or and Wraith, 1999).

In an idealized representation, the water relative permittivity remains constant at high (roughly above 10 GHz) and low (below 100 MHz) frequencies. In a transition region over a frequency band, the real part of complex dielectric permittivity considerably decreases, while imaginary part exhibits a maximum.

Solid parts of most soils and rocks have, when dry, a relative permittivity in the range 2–9 (see Table 1) (Daniels, 2004). Very often, it is assumed that the relative permittivity of dry soils and rock is frequency independent at microwave frequencies.

Total losses in natural rocks and soils are mainly due to conductivity and polarization losses. According to Olhoeft (see, e.g., Olhoeft, 1987), water plays the dominant role in rock and soil conductivity through ionic charge transport through water-filled pore spaces in rocks and soils. Furthermore, conductivity might be due to presence of metals and other good conductors. Conductivity results in dissipation of microwave energy into heat.

Similarly to soils, dielectric properties of snow and seawater ice depend on texture and volumetric amount of water and ice crystal (Bogorodskii et al., 1983; Boyarskii and Tikhonov, 1994; Kovacs et al., 1995). Relaxation frequencies of water molecule in ice crystals lie in kHz region. Thus at microwave frequencies, pure ice has almost constant with frequency dielectric permittivity of about 3.17 and very low losses. Dry snow can be considered as a mixture of ice crystals and air. Depending on density of snow, its relative dielectric permittivity varies from 1.2 to 2. Dielectric permittivity of wet snow depends largely on amount of water in it.

Seawater ice has different properties than freshwater ice, because seawater ice consists of pure ice, air, brine, and possibly solid salts. Furthermore, the sea ice has a multilayer structure, which depends on periodic processes of growth, deformation, and melting. Depending on its age and internal structure, several types of sea ice are distinguished. Relative dielectric permittivity of ice may vary from 3 till 18–30. Losses of coherent energy of microwaves in ice are determined by ice conductivity and noncoherent scattering on spatial heterogeneities.

Attenuation and dispersion of microwaves in subsurface

To simplify the mathematical description of microwave propagation in natural subsurface, a complex-valued apparent dielectric permittivity is used:

$$\varepsilon = \varepsilon' - j\varepsilon''$$

where ε' is a bulk dielectric permittivity of the subsurface and ε'' stands for losses associated with both conductivity and polarization (dipolar) losses:

$$\varepsilon'' = \frac{\sigma}{j\omega} + \varepsilon''_{relaxation}$$

The polarization losses are caused mainly by dielectric relaxation of water and are due to the polarity of the two hydrogen atoms of the water molecule.

The plane electromagnetic wave propagating in such medium can be presented as a function of distance z and time t:

$$E(z,t) = E_0 \cdot e^{-\alpha z} \cdot e^{j(\omega t - \beta z)}$$

where E stands for the electric field component of the field, α is the attenuation factor, and β is the phase constant. Both parameters are determined by the complex-valued apparent dielectric permittivity of the medium:

$$\alpha = \omega\left[\frac{\mu\varepsilon'}{2}\left(\sqrt{1 + \left(\frac{\varepsilon''}{\varepsilon'}\right)^2} - 1\right)\right]^{\frac{1}{2}}$$

$$\beta = \omega\left[\frac{\mu\varepsilon'}{2}\left(\sqrt{1 + \left(\frac{\varepsilon''}{\varepsilon'}\right)^2} + 1\right)\right]^{\frac{1}{2}}$$

The distance d at which the electric field is attenuated in e times is called the skin depth and is equaled to $1/\alpha$. The skin depth is widely used in practice to estimate penetration depth of microwaves into subsurface or thickness of the top subsurface layer which determines microwave emission from the subsurface. It can be seen that the attenuation factor α by ignoring frequency dependence of complex-valued apparent dielectric permittivity is linearly related to frequency. Thus, the skin depth in the very first approximation also linearly decreases with frequency.

Attenuation of microwaves in dry subsurface at 100 MHz varies from typically less than 1 dB/m for snow and sand up to 10 dB/m for some clay, limestone, and sandstone (Daniels, 2004). The attenuation factor slowly increases with frequency in the range from 100 MHz till 1 GHz, while at the frequencies above 1 GHz, it sharply increases due to dielectric relaxation of water. At, for example, 3 GHz, microwave attenuation in subsurface varies typically between several dozens and several hundreds dB/m.

The phase constant β determines the phase velocity of microwave propagation in subsurface:

$$v = \frac{\omega}{\beta} = \left[\frac{\mu\varepsilon'}{2}\left(\sqrt{1+\left(\frac{\varepsilon''}{\varepsilon'}\right)^2}+1\right)\right]^{-\frac{1}{2}}$$

In the case of small losses and far away from relaxation frequencies, the microwave propagates with the phase velocity:

$$u = \frac{c}{\sqrt{\varepsilon_r}} = \frac{1}{\sqrt{\mu_0\varepsilon_0\varepsilon_r}}$$

which is frequency independent. In the general case, finite conductivity of subsurface and the dielectric relaxation of water are responsible for dependence of phase velocity on frequency. This dependence is called dispersion. For microwaves, the dispersion is typically observed at the frequencies above 1 GHz.

By propagation of wideband microwave pulses in subsurface, the dispersion and attenuation cause not only decrease of the pulse magnitude but also distortion of the pulse waveform.

Subsurface scattering

Subsurface is heterogeneous not only on microscopic but also on macroscopic scale. On top of regular geological structure created by multiple strata of different soils/rocks, there is a finer irregular structure created by local spatial inhomogeneities. These are created by spatial variations of moisture, all kind of inclusions (e.g., stones or man-made objects), and animal burrows. While propagating in subsurface, microwaves scatter on interfaces between different strata as well as on local inhomogeneities. Fields scattered from large-scale heterogeneities are used in subsurface remote sensing for geological structure reconstruction and minerals prospecting (Fung, 1994; Daniels, 2004). It has been shown by Yarovoy et al. (2000) that the presence of a low-loss (e.g., sand) layer above rock at certain angles and polarizations of the incident field will enhance microwave backscattering from either internal interface (rock surface) or from air–ground interface.

Field scattered on small-scale (local) inhomogeneities is called subsurface clutter. At medium and/or high microwave frequencies, typical size of many abovementioned local inhomogeneities becomes comparable with the wavelength of electromagnetic field. As a consequence, strength of clutter at these frequencies might become comparable to useful (searched) reflections, resulting in masking these reflections. It is widely believed that at frequencies above 1 GHz, the subsurface clutter becomes the major limiting factor for subsurface sensing.

Subsurface sensing

Subsurface propagation and scattering of microwaves makes a basis for subsurface sensing. Nowadays, three major approaches are used: ground-penetrating radar (GPR) and airborne and spaceborne remote sensing.

GPR

GPR is a surface-based active radar system (Daniels, 2004). In order to achieve reasonable penetration depth, frequencies below 1 GHz are typically used in GPR. At a frequency of 1 GHz, penetration depth varies from about 1–2 m (for dry sandy soil or dry rock) to about 10 cm in water-saturated clay. At a frequency of 100 MHz, penetration depth might reach several 100 m (Cook, 1975). Due to its short-range operation, GPR has to radiate very short (of an order of a few ns) microwave pulses into subsurface, which requires ultra-wide operational bandwidth of the system. Commercial GPR systems have typically octave bandwidth (ratio between the highest and lowest operational frequency equals 2), while some experimental systems have this ratio above 10.

While ultra-wide operational bandwidth is one key feature of GPR, another one is ground-coupled antennas. The latter is needed to avoid, as much as possible, reflection of microwaves from air–ground interface and maximize transmitted into the ground microwave power.

GPR nowadays are widely used for a wide scope of applications starting from geological prospecting and ending with landmine detection and classification.

Airborne sensing (CARABAS, radiometers)

Airborne subsurface sensing has been performed both by active (radars) and passive (radiometers) systems. First active systems have been used for polar ice profiling and mapping of central areas of Kalimantan. Furthermore, CARABAS (airborne radar developed by FOA in Sweden and operating at the frequency band 25–85 MHz) has demonstrated capabilities to image buried pipelines in desert conditions (Hellsten et al., 1996), while FOLPEN (radar developed by SRI International and operating at the frequency band 200–400 MHz) has produced images of buried metal-cased antitank landmines in the Yuma Desert (Grosch et al., 1995).

While performance of active radar systems is limited by maximal allowed transmitted power and field breakdown in the antennas, passive systems are free from these limitations. Airborne microwave radiometers have been successfully used in arid regions for search of subsurface

water, water leakage from irrigation channels, and other purposes. Achieving probably similar to active systems penetration depth, passive systems however cannot provide 3D subsurface images as well as comparable to active systems cross-range resolution.

Spaceborne sensing

Similar to airborne systems, satellite-based systems use both radar and radiometer for subsurface sensing. In particular, SIR-A and SIR-C results demonstrated that the L-band radar has penetration capabilities up to 2 m in rigid areas, revealing details of buried rock structures (Elachi et al., 1984).

Conclusion

Having almost 100 years of history, subsurface remote sensing became recently a fast-growing area of science and engineering with steadily expanding range of applications. This subject lies in a border area of three disciplines: microwave interaction with subsurface materials at microscopic level is studied by geophysics; microwave propagation and scattering in lossy, dispersive, and heterogeneous media is studied by electromagnetics; and extraction of information from microwaves radiated or emitted from the subsurface is studied by remote sensing (including GPR). The main research focus in this area remains on interaction of microwaves with heterogeneities of soils and rocks at microscale in order to determine effective dielectric permittivity of matter and development of statistical models for microwave field scattered from subsurface heterogeneities at macroscale.

Bibliography

Bogorodskii, V., Bentli, C., and Gudmandsen, P., 1983. *Russian Radio Glaciology.* Leningrad: Gidrometeoizdat. In Russian.

Bojarskii, D. A., Tikhonov, V. V., and Komarova, N. Y., 2002. Model of dielectric constant of bound water in soil for applications of microwave remote sensing. *Progress in electromagnetic Research*, **35**, 251–269.

Boyarskii, D. A., and Tikhonov, V. V., 1994. Microwave effective permittivity model of media of dielectric particles and applications to dry and wet snow. In *Proceedings of Geoscience and Remote Sensing Symposium.* Vol. 4, p. 2065.

Cole, K. S., and Cole, H. R., 1941. Dispersion and absorption in dielectrics – I. Alternative current characteristics. *Journal of Chemical Physics*, **9**, 341.

Cook, J., 1975. Radar transparencies of mine and tunnel rocks. *Geophysics*, **40**, 865.

Daniels, D. J. (ed.), 2004. *Ground-Penetrating Radar*, 2nd edn. London: The Institution of Electrical Engineers.

De Loor, G. P., 1983. The dielectric properties of wet materials. *IEEE Transactions on Geoscience and Remote Sensing*, **21**, 364.

Dobson, M. C., Ulaby, F. T., Hallikainen, M. T., and El-Rayes, M. A., 1985. Microwave dielectric behavior of soil – part II: dielectric mixing models. *IEEE Transactions on Geoscience and Remote Sensing*, **23**, 35.

Elachi, C. H., Roth, L. E., and Schaber, G. G., 1984. Spaceborne radar subsurface imaging in hyperarid regions. *IEEE Transactions on Geoscience and Remote Sensing*, **22**, 383.

Fung, A. K., 1994. *Microwave Scattering and Emission Models and Their Applications*. Norwood: Artech House.

Grosch, T. O., Lee, C. F., Adams, E. M., Tran, C., Koening, F., Tom, K., and Vickers, R. S., 1995. Detection of surface and buried mines with an UHF airborne SAR. *Proceedings of SPIE*, **2496**, 110.

Hallikainen, M. T., Ulaby, F. T., Dobson, M. C., El-Rayes, M. A., and Wu, L. K., 1985. Microwave dielectric behaviour of wet soil – part I. Empirical models and experimental observations. *IEEE Transactions on Geoscience and Remote Sensing*, **15**, 25.

Heimovaara, T. J., 1994. Frequency-domain analysis of time-domain reflectometry waveforms 1. Measurement of the complex dielectric permittivity. *Water Resources Research*, **30**, 189.

Heimovaara, T. J., Bouten, W., and Verstraten, J. M., 1994. Frequency domain analysis of time domain reflectometry waveforms 2. A four-component complex dielectric mixing model for soils. *Water Resources Research*, **30**, 201.

Hellsten, H., Ulander, L. M., Gustavsson, A., and Larsson, B., 1996. Development of VHF CARABAS II SAR. *Proceedings of SPIE*, **2747**, 48.

Hipp, J. E., 1974. Soil electromagnetic parameters as functions of frequency, soil density and soil moisture. *Proceedings of the IEEE*, **62**, 98.

Hoekstra, P., and Delaney, A., 1974. Dielectric properties of soils at UHF and microwave frequencies. *Journal of Geophysical Research*, **79**, 1699.

Hülsenbeck, R., et al., 1926. German patent No. 489434.

Kovacs, A., Gow, A. J., and Morey, R. M., 1995. The in-situ dielectric constant of polar firn revisited. *Cold Regions Science and Technology*, **23**, 245.

Olhoeft, G. R., 1987. Electrical properties from 10–3 to 10 + 9 Hz – physics and chemistry. In Banavar, J. R., Koplik, J., and Winkler, K. W. (eds.), *Physics and Chemistry of Porous Media II*. New York: American Institute of Physics.

Or, D., and Wraith, J. M., 1999. Temperature effects on soil bulk dielectric permittivity measured by time-domain reflectometry: a physical model. *Water Resources Research*, **35**, 371.

Sen, P. N., Scala, C., and Cohen, M. H., 1981. A self-similar model for sedimentary rocks with application to the dielectric constant of fused glass beads. *Geophysics*, **46**, 781.

Shutko, A. M., and Reutov, E. M., 1982. Mixture formulas applied in estimation of dielectric and radiative characteristics of soil and grounds at microwave frequencies. *IEEE Transactions on Geoscience and Remote Sensing*, **20**, 29.

Simmons, G., Strangway, D. W., Bannister, L., Baker, R., Cubley, D., La Torraca, G., and Watts, R., 1972. The surface electrical properties experiment. In Kopal, Z., and Strangway, D. W. (eds.), *Lunar Geophysics*. Dordrecht: Reidel, p. 258.

Stern, W., 1929. Versuch einer elektrodynamischen Dickenmessung von Gletschereis. *Gerlands Beitrage zur Geophysik*, **23**, 292.

Stern, W., 1930. Uber Grundlagen, Methodik und bisherige Ergebnisse elektrodynamischer Dickenmessung von Gletschereis. *Zeitschrift Gletscherkunde*, **15**, 24.

Topp, G. C., Davis, J. L., and Annan, A. P., 1980. Electromagnetic determination of soil water content: measurement in coaxial transmission lines. *Water Resources Research*, **16**, 574.

Wang, J. R., and Schmugge, T. J., 1980. An empirical model for the complex dielectric permittivity of soils as a function of water content. *IEEE Transactions on Geoscience and Remote Sensing*, **18**, 288.

Wobschall, D., 1977. A theory of the complex dielectric permittivity of soil containing water: the semidisperse model. *IEEE Transactions on Geoscience and Remote Sensing*, **15**, 49.

Yarovoy, A. G., de Jongh, R. V., and Ligthart, L. P., 2000. Scattering properties of a statistically rough interface inside a multilayered medium. *Radio Science*, **35**, 455.

MICROWAVE SURFACE SCATTERING AND EMISSION

David R. Lyzenga
College of Engineering, Naval Architecture and Marine Engineering, University of Michigan, Ann Arbor, MI, USA

Definition

Surface scattering. The process by which microwave radiation incident upon a solid or liquid surface is wholly or partially redirected away from that surface.

Microwave emission. The process by which microwave radiation originates from a solid or liquid surface due to the rotational or vibrational motions of molecules near the surface.

Introduction

This topic deals with the processes through which a solid or liquid surface influences the electromagnetic field (at microwave frequencies) in the vicinity of that surface and the manner in which these processes are determined by the physical and chemical properties of the surface. The processes of surface scattering and emission, though distinctly different, are related by *Kirchhoff's law* and can both be described by means of a single function, the *bidirectional reflectance*. The bidirectional reflectance function is defined in the following section, and the dependence of this function on the physical and chemical properties of the surface is discussed in the subsequent section. For active microwave (radar and scatterometry) purposes, the scattering properties of surfaces are more commonly described in terms of the normalized radar cross section, or cross section per unit area, which is related to the bidirectional reflectance as discussed near the end of this section.

Bidirectional reflectance

The bidirectional reflectance of a given surface, denoted here by the symbol $\rho(\mu,\phi;\mu',\phi')$, may be loosely described as the proportion of the radiance incident upon the surface from the direction (μ',ϕ') which is scattered or reflected from the surface into the direction (μ,ϕ). Here μ refers to the cosine of the angle between the direction of the reflected radiation and the surface normal, and ϕ refers to the azimuthal angle of the reflected radiation about the surface normal. The symbols μ' and ϕ' similarly refer to the direction from which the incident radiation arrives at the surface. If we denote the surface normal direction by the unit vector $\boldsymbol{n}$ and the direction of propagation of the reflected and incident radiation by the unit vectors $\boldsymbol{k}$ and $\boldsymbol{k}'$, respectively, then $\mu = \boldsymbol{k}\cdot\boldsymbol{n}$ and $\mu' = -\boldsymbol{k}'\cdot\boldsymbol{n}$. For a specular surface, all of the incident radiation is reflected into the direction $\mu = \mu'$ and $\phi = \phi' + \pi$, whereas for a diffusely reflecting surface, the radiation incident from a given direction is scattered into a range of angles.

The bidirectional reflectance may be defined more rigorously, albeit implicitly, by the equation

$$L(\mu,\phi) = \int_0^1 \int_0^{2\pi} \rho(\mu,\phi;\mu',\phi') L(\mu',\phi') \mu' d\mu' d\phi' \tag{1}$$

where $L(\mu,\phi)$ is the reflected radiance and $L(\mu',\phi')$ is the incident radiance (Nicodemus, 1965). For a system in thermal equilibrium, the incident radiance is isotropic and is equal to $B_\nu(T_s) = \frac{2}{\lambda^2}\frac{h\nu}{e^{h\nu/kT_S}-1}$ where $\nu = c/\lambda$ is the frequency, h is Planck's constant, k is Boltzmann's constant, and T_S is the surface temperature. In this case, the sum of the reflected and emitted radiation is also equal to $B_\nu(T_S)$. The emitted radiance in the direction (μ,ϕ) is then equal to $\varepsilon(\mu,\phi)\, B_\nu(T_S)$, where

$$\varepsilon(\mu,\phi) = 1 - \int_0^1 \int_0^{2\pi} \rho(\mu,\phi;\mu',\phi') \mu' d\mu' d\phi' \tag{2}$$

is the surface emissivity. This relationship is known as Kirchhoff's law for unpolarized radiation. At microwave frequencies, $h\nu << kT_S$ for temperatures typical of the Earth's surface, and $B_\nu(T_S) \approx 2kT_S/\lambda^2$ (Rayleigh-Jeans law). Thus, the emitted radiance is proportional to the product of the surface temperature and the emissivity and is often converted into units of temperature (K) and referred to as the *brightness temperature*, this being defined as the temperature of a blackbody that emits the same radiance as the actual surface, i.e., $T_B = \varepsilon\, T_S$.

Thus far, we have only considered the case of unpolarized radiation. However, the quantities defined above can be extended to the case of polarized radiation as well. There are various definitions of the state of polarization, but for microwave radiation, the most common definition is in terms of the modified *Stokes parameters* T_h, T_ν, U, and V (see entry on "*Radiation, Polarization, and Coherence*"). A complete description of the reflectivity would require 16 components, corresponding to each possible combination of incident and reflected polarization states. However, for most remote-sensing applications, we are primarily interested in the reflection of atmospheric (and cosmic) radiation which is virtually unpolarized, and we can combine these into four components corresponding to the reflection of unpolarized radiation into each of the four polarization states. For notational purposes, we indicate that the bidirectional reflectance is polarization dependent by using the bold-faced symbol $\boldsymbol{\rho}(\mu,\phi;\mu',\phi')$, to be understood as a vector with components corresponding to the four modified Stokes parameters. Since the blackbody radiation is unpolarized, the blackbody emissivity vector can be written as $\boldsymbol{\varepsilon}_b = [1,1,0,0]$, and Kirchhoff's law for polarized radiation becomes

$$\boldsymbol{\varepsilon}(\mu,\phi) = \boldsymbol{\varepsilon}_b - \int_0^1 \int_0^{2\pi} \boldsymbol{\rho}(\mu,\phi;\mu',\phi') d\mu' d\phi'. \tag{3}$$

Just above the surface of the Earth, the upwelling radiation field includes contributions from both the downwelling (atmospheric and cosmic) radiation that is reflected from the surface and the radiation that is emitted from the surface. The reflected contribution can be written in terms of the brightness temperature as

$$\boldsymbol{T}_r(\mu,\phi) = \int_0^1 \int_0^{2\pi} \boldsymbol{\rho}(\mu,\phi;\mu',\phi')T_d(\mu',\phi')d\mu' d\phi' \tag{4}$$

where $T_d(\mu',\phi')$ represents the (unpolarized) downwelling radiation at the surface, and the surface-emitted component can be written as

$$\begin{aligned}\boldsymbol{T}_e(\mu,\phi) &= T_S\boldsymbol{\varepsilon}(\mu,\phi)\\ &= T_S\boldsymbol{\varepsilon}_b - T_S\int_0^1\int_0^{2\pi}\boldsymbol{\rho}(\mu,\phi;\mu',\phi')d\mu' d\phi'.\end{aligned} \tag{5}$$

Combining these expressions, the total upwelling brightness temperature just above the surface can be written as

$$\begin{aligned}\boldsymbol{T}(\mu,\phi) =& \boldsymbol{T}_e + \boldsymbol{T}_r\\ =& T_S\boldsymbol{\varepsilon}_b + \int_0^1\int_0^{2\pi}\boldsymbol{\rho}(\mu,\phi;\mu',\phi')\\ & [T_d(\mu',\phi') - T_S]d\mu' d\phi'.\end{aligned} \tag{6}$$

Measured at spacecraft altitudes, the observed brightness temperature would of course be reduced by attenuation in the atmosphere and augmented by the upwelling radiation emitted from the atmosphere.

Dependence of bidirectional reflectance on surface properties

The interest in making measurements of the microwave radiation field above the surface of the Earth is to obtain information on geophysical or biological processes occurring at the surface or in the atmosphere or on the physical or chemical state of the surface and/or atmosphere. The types of information and the methods used to extract this information are discussed in other entries (see, e.g., *Land-Atmosphere Interactions, Evapotranspiration*; *Ocean-Atmosphere Water Flux and Evaporation*; *Ocean, Measurements and Applications*; *Cryosphere, Measurements and Applications*). In this section, we discuss some of the relationships that underlie these applications.

As discussed in the previous section, the effects of the surface on the radiation field above the surface can be described in principle by the surface temperature and the bidirectional reflectance vector $\rho(\mu,\phi;\mu',\phi')$. This quantity can be defined for a specific surface but more often refers to a statistical ensemble of surfaces with similar properties. It could in principle be measured, but as can be imagined, this is a difficult and time-consuming process since it involves many different combinations of incidence and reflection angles and must be repeated for many surfaces to obtain a stable statistical mean. Thus it is desirable to model the bidirectional reflectivity when possible, although measurements are necessary for model validation and can sometimes be used directly if some assumptions are made about the directional dependence.

Models have been developed, even for complex surfaces such as vegetation canopies, as discussed in the *Measurements and Applications* entries referenced above. A review of these models is beyond the scope of this entry, but in general the reflectivity depends on the surface roughness, or the geometrical shapes of the elements comprising the surface or canopy, as well as the *dielectric constant* of the component materials. Since the dielectric constant is influenced by such factors as the water content of soils and vegetation, the salinity and phase (liquid or ice) of the ocean surface, and possibly other chemical constituents, the radiation field potentially contains information about these factors. In addition, the roughness of the ocean surface is a function of the wind speed and direction (as well as, possibly, other factors such as the presence of surface slicks, converging or diverging surface currents, and atmospheric stability conditions). As a result, radiometric measurements from aircraft and satellite platforms have been shown to be very useful for measuring surface winds over the ocean on regional and global scales and for observing some of the other features mentioned above on smaller scales. Polarimetric measurements such as those provided by WINDRAD (Yueh et al., 1999) and WindSat (Gaiser et al., 2004) are particularly useful for separating atmospheric from surface effects and for inferring the wind direction as well as wind speed.

For the ocean surface, the bidirectional reflectance is most commonly modeled using a two-scale model (Wentz, 1975; Yueh, 1997) although other models such as the small-slope approximation (Irisov, 1997; Johnson and Zhang, 1999) have also been used. Such modeling also requires a representation of the ocean wave spectrum and its dependence on the wind speed. The spectral model of Durden and Vesecky (1985) has been widely used for this purpose, but some modifications have also been suggested (Lyzenga, 2006) in order to improve comparisons with satellite data. The shortwave (centimeter-scale) portion of the ocean wave spectrum has a strong influence on the microwave surface reflectivity, and there are large uncertainties in this portion of the spectrum due to the difficulty of measuring the surface on these scales. The effects of wave breaking and surface foam on the reflectivity are also incompletely understood, although considerable work has been done in this area.

In the radar literature, the scattering properties of objects are most often described in terms of the radar cross section σ, which is defined as 4π times the reflected power per unit solid angle divided by the power density (power per unit area) incident on the object. It can be defined for any scattering angle but is most commonly used for

backscatter, i.e., for the case in which the source of radiation and the receiver are collocated. This definition applies to discrete objects but was later extended to the case of scattering surfaces by dividing the reflected power by the surface area illuminated or resolved by the radar. This dimensionless quantity is denoted by the symbol σ^o and referred to as the normalized radar cross section, or the cross section per unit area.

The relationship between the radar cross section and the bidirectional reflectance can be found by considering the radiation scattered from a unit surface area. For the case of a distant source, the incident radiance can be written as

$$L(\mu', \phi') = E_o \delta(\mu' - \mu_o)\delta(\phi' - \phi_o) \quad (7)$$

where E_o is the incident power per unit area normal to the direction of the incident radiation (as in the definition of the radar cross section). The scattered radiance for this case is given by Equation 1 as

$$L(\mu, \phi) = \mu_o E_o \rho(\mu, \phi; \mu_o, \phi_o). \quad (8)$$

Following from the definition of the radiance, the reflected power per unit solid angle is equal to $\mu\, L(\mu,\phi)$. Applying the definition of the radar cross section, we then have

$$\sigma^\circ = 4\pi \frac{\mu L(\mu, \phi)}{E_o} = 4\pi\mu\mu_o\rho(\mu, \phi; \mu_o, \phi_o) \quad (9)$$

in agreement with Tomiyasu (1988). For the backscatter case, of course, $\mu = \mu_o$.

Conclusion

The effect of the Earth's surface on the naturally occurring microwave radiation field above the surface can be described in terms of the surface temperature and the bidirectional reflectance of the surface. The bidirectional reflectance is a function of the geometry and composition of the surface. For land surfaces, the reflectance is strongly dependent on the water content of the soil and/or vegetation. For water surfaces, the reflectance depends on the salinity, the phase (liquid or solid), and the surface roughness. For the liquid ocean, the surface roughness is primarily dependent on the wind speed but may also be influenced by surface slicks, currents, and atmospheric stability effects.

Bibliography

Durden, S. P., and Vesecky, J. F., 1985. A physical radar cross-section model for a wind-driven sea with swell. *IEEE Journal of Oceanic Engineering*, **10**, 445–451.

Gaiser, P. W., St Germaine, K. M., Twarog, E. M., Poe, G. A., Purdy, W., Richardson, D., Grossman, W., Jones, W. L., Spencer, D., Golba, G., Cleveland, J., Choy, L., Bevilacqua, R. M., and Chang, P. S., 2004. The WindSat spaceborne polarimetric microwave radiometer**:** sensor description and early orbit performance. *IEEE Transactions on Geoscience and Remote Sensing*, **42**, 2347–2361.

Irisov, V. G., 1997. Small-slope expansion for thermal and reflected radiation from a rough surface. *Waves in Random Media*, **7**, 1–10.

Johnson, J. T., and Zhang, M., 1999. Theoretical study of the small slope approximation for ocean polarimetric thermal emission. *IEEE Transactions on Geoscience and Remote Sensing*, **37**, 2305–2316.

Lyzenga, D. R., 2006. Comparison of WindSat brightness temperatures with two-scale model predictions. *IEEE Transactions on Geoscience and Remote Sensing*, **44**, 549–559.

Nicodemus, F. E., 1965. Directional reflectance and emissivity of an opaque surface. *Applied Optics*, **4**, 767–773.

Tomiyasu, K., 1988. Relationship between and measurement of differential scattering coefficient (σ^o) and bidirectional reflectance distribution function (BDRF). *IEEE Transactions on Geoscience and Remote Sensing*, **26**, 660–665.

Wentz, F. J., 1975. A two-scale model for foam-free sea microwave brightness temperatures. *Journal of Geophysical Research*, **80**, 3441–3446.

Yueh, S. H., 1997. Modeling of wind direction signals in polarimetric sea surface brightness temperatures. *IEEE Transactions on Geoscience and Remote Sensing*, **35**, 1400–1418.

Yueh, S. H., Wilson, W. J., Dinardo, S. J., and Li, F. K., 1999. Polarimetric microwave brightness signatures of ocean wind directions. *IEEE Transactions on Geoscience and Remote Sensing*, **37**, 949–959.

Cross-references

Cryosphere, Measurements and Applications
Land-Atmosphere Interactions, Evapotranspiration
Ocean-Atmosphere Water Flux and Evaporation
Ocean, Measurements and Applications
Ocean Measurements and Applications, Ocean Color
Radiation, Polarization, and Coherence

MISSION COSTS OF EARTH-OBSERVING SATELLITES

Randall Friedl and Stacey Boland
Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

Synonyms

Earth-observing mission costs; Spaceborne mission costs

Definition

Earth-observing satellite mission. An instrumented spacecraft deployed in space that is designed to observe select geophysical parameters of the Earth for purposes of environmental science or monitoring. A satellite mission requires a launch vehicle, observing instruments, and a spacecraft bus which houses the instruments as well as systems for guidance, navigation, control, communications, command and data, power, thermal control, propulsion, and structures.

Satellite mission cost. The total cost of implementing a satellite mission. The cost includes all of the component

costs as well as costs for design, system engineering, operating and data analysis software, and mission operations. The cost does not typically include scientific analysis of the data collected by the mission.

Introduction

Observing the Earth from the vantage point of space has transformed our understanding of the Earth as an integrated system of physical and biological components. The number of Earth-observing satellites has steadily increased with time, with more than 100 satellite missions currently operating. The desire for new satellite missions remains high, especially since many of the geophysical parameters of interest to scientists (e.g., 44 essential climate variables as defined by the Global Climate Observing System Program) are still not observed from space and other key parameters need to be sustained over many satellite lifetimes to enable investigations of long-term climate trends. The need for new missions is tempered by the relatively high cost, currently ranging from $100 M to $3 B, to deploy and operate a satellite mission. Understanding the factors that determine the costs of satellite missions is critical to developing and sustaining an effective system of Earth-observing satellites.

Cost factors

Transporting instrumented science payloads to space requires powerful launch vehicles. The relatively small number (i.e., ~30 total, ~12 USA) of available launch vehicles is divided into four classes, as defined by the US Federal Aviation Administration Office of Commercial Space Transportation, that reflect lift capability as follows: small ($\leq$5,000 lb to Low Earth Orbit or LEO), medium (5,001–12,000 lb to LEO), intermediate (12,001–25,000 lb to LEO), and heavy ($\geq$25,000 lb to LEO). Examples of launch vehicles across the four classes are Pegasus (small), Delta II (medium), Atlas IIAS (intermediate), and Atlas V (heavy). While the cost of a launch vehicle increases with each step-up in class, the per pound payload cost is roughly the same across classes when the launch vehicle lift capability is fully utilized. There is thus a substantial cost incentive to either design a payload to fit into the smallest possible launch vehicle or to combine payloads to take advantage of the full capability of a larger-class vehicle. Most of the past and current Earth science missions range between 1,000 and 10,000 lb (instruments plus spacecraft) and were launched on small- and medium-class launch vehicles.

While launch vehicle costs can be directly correlated to performance requirements in terms of a "price per pound" metric, this is not the case for estimating the cost of the satellite instruments. Science instrument payloads typically represent the most unique and challenging components of Earth-observing missions. Payload designs are generally mission unique, resulting from a careful optimization of science requirements against physical and programmatic constraints (e.g., cost, schedule, risk). Instrument costs can thus be driven by a variety of factors. Increasingly stringent requirements on temporal and spatial resolution, for example, might necessitate larger signal collection optics and/or antennas as well as highly sensitive detectors and precision signal processing electronics, driving payload mass, volume, and component costs higher. Attempts to reduce payload mass and volume might necessitate incorporation of new, yet expensive, lightweight technologies.

Spacecraft costs, like launch vehicle costs, are generally performance driven. Instrument acodation requirements, particularly mass, power, data rate, and configuration constraints, greatly dictate requirements for the platform, or spacecraft bus, that houses the instrument. While many standardized platforms are now available, most missions have the need for some degree of spacecraft customization (e.g., tighter thermal control, improved pointing accuracy or knowledge, larger onboard data storage capacity, additional power allocations). The cost of customization is to a large extent dependent on whether new technology development is required.

Mission operations costs are driven, to a large extent, by mission data volume and latency requirements which can drive ground network, data processing, and archival costs. The spacecraft's onboard data storage capacity (i.e., solid-state recorder size) is also a determining factor in the required frequency of data downlinks and/or number of ground stations.

Cost estimation methods

Three methodologies are generally used to estimate mission costs, namely, analogy, parametric, and "grass roots."

An analogy-based estimate represents the simplest approach to cost estimation. It relies on direct comparisons to analogous ongoing or previous missions. Provided there is little deviation from the chosen analog, rapid and accurate mission cost estimates can be obtained using this method. However, since very few missions share identical subsystems or components, analogy-based estimates are not widely applicable without including cost adjustments based on analysis of differences between the target mission and the analog. Such adjustments typically employ parametric analyses in order to enable quantitative extrapolations between systems with varying degrees of differences. Analogy-based estimates best apply to mission lines which involve launching a series of similar (or near-identical) satellites (e.g., the NOAA POES series of weather satellites) and those which incorporate incremental performance improvements between generations (e.g., the NASA TOPEX-Poseidon, Jason, OSTM-sustained research satellite series).

Parametric cost estimates are based on cost driver relationships derived from analyses of historical data. The primary cost drivers used for parametric studies typically include mass, power, and subsystem complexity. For example, experience shows that mission costs increase roughly linearly with mass within a given launch vehicle

class, allowing derivation of the "price per pound" metric cited above. Once developed, parametric relationships are particularly useful for performing rapid trade studies during the initial scoping of mission concepts, particularly for concepts without direct historical analogy.

Grassroots cost estimates involve detailed consideration of each element of a mission or component of an instrument. For each element or component, a cost estimate is generated and basis of estimate is documented, describing the technique used to derive the estimate. For a true grassroots estimate to be performed, the mission or instrument design must be specified to a fairly high degree to enable rigorous quantitative cost estimates based on material and labor costs allocated over a defined implementation schedule. Consequently, grassroots estimates find their greatest utility during later stages of mission development. During concept development, a "quasi-grassroots" mechanism can be employed in which the grassroots cost estimating structure is preserved, including documentation of a basis of estimate, while leveraging analogy or parametric cost estimates for the individual components for which detailed requirements, designs, and/or schedules are not well known.

Cost estimation accuracy

The ability to conceive, plan, and implement an effective and affordable multi-satellite observing system rests in large part on understanding the costs of individual missions. In particular, decisions regarding mission priorities and sequencing are intimately tied to considerations of cost to benefit.

Because there are relatively small numbers of satellite missions and most of them are partially, or wholly, unique in design, the ability to accurately estimate cost is limited in the beginning phases of a mission. Analysis of historical mission cost data reveals a tendency for substantial increases (10–50 %) in estimates for lower-cost Earth missions when going from initial to later mission phases. These increases have been traced primarily to initial underestimations of instrument and spacecraft components. In many cases, the cost increases reflect more stringent measurement requirements stemming from refinement of the mission science goals, rather than inherent inaccuracies in the estimation methodology.

Acknowledgment

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the NASA.

Bibliography

Bearden, D., 2005. Perspectives on NASA mission cost and schedule performance trends. In *NASA Goddard Flight Center Symposium*.

Committee on Earth Observation Satellites mission database, available at http://www.ceos.org/

Emmons, D. L., Bitten, R. E., and Freaner, C. W., 2006. Using historical NASA cost and Schedule growth to set future program and project reserve guidelines, IEEEAC paper #1545, Version 3.

FAA Commercial Launch Vehicle website. http://www.faa.gov/about/office_org/headquarters_offices/ast/launch_data/.

Futron Corporation, 2002. Space transportation costs: trends in price per pound to orbit 1990–2000. www.futron.com.

Global climate observing system program's essential climate variables. Available from http://www.wmo.int/pages/prog/gcos/index.php?name=essentialvariables.

Larson, W. J., and Wertz, J. R., 1999. *Space Mission Analysis and Design*. El Segundo: Microcosm, Vol. 3.

NASA, 2008. Cost estimation handbook. Available from http://ceh.nasa.gov.

Wertz, J. R., and Larson, W. J., 2007. *Reducing Space Mission Cost*. New York: Springer.

Cross-references

Global Climate Observing System
Mission Operations, Science Applications/Requirements
Observational Systems, Satellite

MISSION OPERATIONS, SCIENCE APPLICATIONS/ REQUIREMENTS

David L. Glackin
Los Angeles, CA, USA

Synonyms

Environmental data record (EDR) requirements

Definition

Science applications. Environmental phenomena to be measured in the Earth's atmosphere, oceans, land surface, solid Earth, ice cover, and near-Earth space environment.
Science requirements. A numerical specification of the phenomena that must be measured and a corresponding set of attributes including spatial resolution, spatial reporting interval, spatial coverage, measurement uncertainty (accuracy/precision), long-term measurement stability, mapping uncertainty, reporting frequency, and timeliness of data delivery.

Introduction

The science applications are described thoroughly elsewhere in this volume (see *Atmospheric General Circulation Models*; *Ocean, Measurements and Applications*; *Cryosphere, Measurements and Applications*; *Land-Atmosphere Interactions, Evapotranspiration*; *Trace Gases, Troposphere - Detection from Space*). The thrust of this entry is the science requirements and how they are motivated, established, codified, analyzed, and iterated.

David L. Glackin: deceased.

Motivation for and establishment of requirements

In an ideal world, science requirements should be motivated by a problem in need of a remote sensing instrument (or set of instruments), not vice versa. Reality is sometimes different, but environmental remote sensing missions, especially large ones, typically have a large user constituency behind them that provides the rationale for the mission. Mission requirements may be driven by the need for new types of data, improvements in the quality of old types of data, or continuity of data for studies of long-term trends.

Ideally, the requirements flow should be from the top down. The science requirements should drive the instrument concepts. Those concepts in turn should drive the instrument designs, which should in turn drive the platform (satellite, aircraft, etc.) concept and design. Examples of important aspects of the platform design that the science requirements drive are pointing accuracy and stability and the choice of satellite orbit (see *Observational Systems, Satellite*).

As an adjunct to the science requirements, the postlaunch data reduction, analysis, and research activities should be considered up front. Too often, the tendency is to focus on building the instruments and the platform to meet the requirements and to delay consideration of the analysis and research. The latter approach can have large systems impacts, and it is best to consider these issues up front while setting requirements. For example, the calibration/validation phase of a mission is a major postlaunch activity and is the one in which the instrument calibration is established and in which the quality of the science data products is validated (see *Calibration and Validation*). If the science requirements lead to data products that are exceedingly difficult to validate, unexpected amounts of postlaunch resources may be required (see *Mission Costs of Earth-Observing Satellites*).

Science requirements are usually established by user working groups, which are organized by scientific discipline. It is important to involve the instrument systems engineers early in this process. They understand the hardware in an end-to-end systems sense and can help the user communities to understand the cost and risk implications of their requirements. Often, scientific "desires" must be tempered with technological reality. Sometimes, it may be possible to achieve half of what is desired at one-tenth the cost. Science requirements are typically iterated between these two groups of people until a realistic set of requirements is established.

Science requirements are often couched in terms of "thresholds" and "objectives." The thresholds are the minimum set of requirements that must be met. The objectives represent the performance that the users would like to meet, if meeting them does not inordinately drive the hardware design. Remote sensing missions may in the end deliver data products that fall between these two benchmarks.

Codification of requirements

The science requirements must be established carefully. Clarity is of the utmost importance. If the requirements are not clear up front, much time will be wasted later in clarifying the requirements to the instrument and platform vendors and then in rewriting the requirements. Definitions must be clear and comprehensive. Units must be consistent and understandable. Requirements must be consistent with basic physics. There should be no conflicting requirements. Very significant is the establishment of requirements for measurement quality. It is possible to generate major confusion over items as seemingly simple as measurement accuracy, precision, and uncertainty. The latter can be a major stumbling block in the process of requirements establishment and should be treated thoughtfully.

Requirements analysis and iteration

Science requirements should be thoroughly analyzed after they are initially established, for clarity, consistency, comprehensiveness, lack of conflict, and impact on the end-to-end mission. This is the point at which the design engineers, the hardware vendors, and the operators should engage with the system engineers to ensure that the set of requirements is sound. At this stage, iterations of the requirements are usually made, in consultation with the user community.

It is paramount that experienced people think through these issues up front. Otherwise, problems can crop up during the course of the program that can be very wasteful of resources.

It is particularly important to define the "driving requirements" (Wertz and Larson, 1991). These are the ones that dictate some of the major hardware characteristics.

"Requirements creep" is almost always a problem in any large remote sensing mission. Once the user communities start to see what the system can do, they usually think of more things that it might do. One solution for that is to define these items as "Pre-Planned Product Improvements" (P3I) that might be addressed later, given sufficient resources.

Conclusion

An intelligent process of establishing the science requirements for a mission can make the difference between mission success and failure or between delivering the hardware on time and under budget and delivering it late and far over budget. Getting the science requirements right at the beginning and thinking things through thoroughly up front is key to mission success.

Bibliography

Wertz, J. R., and Larson, W. J., 1991. *Space Mission Analysis and Design*. Dordrecht: Kluwer.

Cross-references

Calibration and Validation
Cryosphere, Measurements and Applications
Ocean, Measurements and Applications