E

EARTH RADIATION BUDGET, TOP-OF-ATMOSPHERE RADIATION

Bing Lin
NASA Langley Research Center, MS 420, Hampton, VA, USA

Synonyms

Extraterrestrial radiation; Longwave radiation; Shortwave radiation; Solar radiation; Terrestrial radiation; Thermal radiation

Definition

1. Extraterrestrial radiation: the radiative energy that radiates into the Earth's climate system from the space.
2. Terrestrial radiation: the radiative energy emitted by the Earth's climate system.

Introduction

The radiation at top of the atmosphere (TOA) is the only energy exchange process between the Earth and space for the Earth's climate system. The incoming extraterrestrial radiation, dominantly from the sun, is the driving forcing for the general circulation of the climate system. In response to this solar radiative forcing, the climate system releases heat to the space through terrestrial thermal emission. For the Earth, the incoming solar radiation is generally balanced by the outgoing thermal radiation at TOA at the annual and longer timescales to maintain the climate system in a quasi-equilibrium state. Some anthropogenic processes, such as the continuous increase of the CO_2 concentration within the atmosphere caused by the increased fossil fuel burning, inject extra thermal radiative forcing into the climate system, which breaks the basic radiation balance and forces the climate to change. Since the coupling of solar and thermal radiative energy processes with the Earth's hydrological processes such as precipitation and evaporation builds a fundamental aspect of the global water and energy cycle of the climate system, variations in radiative fields would cause corresponding changes in global precipitation patterns and freshwater distributions and have a strong social-economical impact. The observations of TOA radiation, thus, become a critical part of the remote sensing of the Earth's climate system.

The top-of-the-atmosphere radiation

Satellite direct measurements of TOA radiation (Barkstrom et al., 1989; Wielicki et al., 1996) focus on broadband spectra of the total ($0.2 \sim 100$ μm), shortwave ($0.2 \sim 5$ μm), and longwave ($5 \sim 100$ μm) radiation, which cover more than 99.5 % of the Earth's radiant energy. Absolute calibration and long-term stability are the core elements emphasized by satellite TOA radiation-measuring missions. Both wide and narrow field of view (FOV) techniques have been used in satellite TOA radiation measurements for decades (Barkstrom et al., 1989; Wielicki et al., 1996). Although wide FOV observations can obtain the information of global net radiation, detailed radiative transfer processes for regional and small-scale phenomena such as clouds and aerosols cannot be measured. The narrow FOV measurements, on the other hand, have reasonable spatial resolution in monitoring all large, regional, and small-scale physical processes. But to obtain the full radiant energy of these physical processes, a series of complicated satellite retrievals is required for narrow FOV radiance measurements. Each instantaneous satellite observation of the Earth's radiation field from narrow FOV broadband radiation instruments can only measure the radiance at a particular viewing angle of the targeted scene, and each scene type has its distinct three-dimensional (3-D) radiation features. In order to retrieve the radiative flux of the targeted scene from the radiance

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

measurement at the particular viewing angle, a 3-D angular dependent model (ADM) of radiance field for the scene type has to be used. The ADM is built from the statistics of large amounts of observational multi-angle radiance data. Thus, the spatial resolution of radiation measurements should be both large enough to satisfy the 3-D radiation statistics of individual scene types and fine enough to determine the radiative characteristics of major climate processes. Satellite radiation sensors, such as those from the Earth Radiation Budget Experiment (ERBE) (Barkstrom et al., 1989) and Clouds and the Earth's Radiant Energy System (CERES) (Wielicki et al., 1996), normally have a spatial resolution from 10 to 40 km and generally not only measure the fluxes of total, shortwave, and longwave radiation at TOA but also differentiate the radiation from individual Earth's scene types, such as clouds, aerosols, clear skies, deserts, grasses, forests, and ice sheets. Temporal resolutions of satellite radiation measurements depend on both satellite orbits and scan patterns of radiation sensors. Precession orbits can provide radiative flux measurements of broad solar incident angles for each place and scene type but usually have low repeating sampling rate. Polar orbital sensors, on the other hand, have high repeating rates of measurements for similar weather and climate conditions, but their measurements are only available for limited local times.

Another way in satellite estimations of TOA radiant energy is to observe the radiative properties of all critical agents and components within the atmosphere and at surfaces through satellite narrowband radiation measurements such as those from passive and active visible, infrared, and microwave remote sensing techniques and then to calculate the radiant energy based on these radiative properties using radiative transfer theories. The spatial and temporal resolutions of these radiative flux estimates, thus, depend on their corresponding narrowband satellite measurements, which have a wide range of variations. As the methods of satellite direct broadband radiation measurements, this indirect measurement method has also been used for decades (Rossow and Schiffer, 1999) and is a major part of the satellite remote sensing of global radiation budgets.

Satellite radiative energy measurements track all important physical processes within the Earth's climate system and are currently one of the keys in monitoring and understanding of the climate system. Climate data records from the satellite TOA radiation measurements capture global, regional, and small-scale signatures of the climate variability and change and exhibit distinct features of the tropics and polar regions in response to global warming trends. These TOA radiation measurements, especially those for cloudy and clear skies, also provide strong energy constraints for global climate models (GCMs) and lead advanced GCM cloud parameterizations. With the continuous, long-term radiation observations from CERES and other satellite radiation missions, an improved projection of future climate could be obtained.

Summary

Satellite measurements of the Earth's radiation budget have significant influences on the understanding of aerosols, cloud feedbacks, and atmospheric dynamics as well as other major climate physical processes and are critical for validations and improvements of climate models. Continuous, long-term radiation observations from satellite missions would lead to accurate predictions of the future climate.

Bibliography

Barkstrom, B., et al., 1989. The Earth Radiation Budget Experiment (ERBE) archival and April 1985 results. *Bulletin of the American Meteorological Society*, **70**, 1254–1262.

Rossow, W. B., and Schiffer, R. A., 1999. Advances in understanding clouds from ISCCP. *Bulletin of the American Meteorological Society*, **80**, 2261–2287.

Wielicki, B. A., Barkstrom, B., Harrison, E. F., Lee, R., Smith, G., and Cooper, J., 1996. Clouds and the Earth's Radiant Energy System (CERES): An Earth observing system experiment. *Bulletin of the American Meteorological Society*, **77**, 853–868.

Cross-references

Aerosols
Atmospheric General Circulation Models
Calibration and Validation
Climate Data Records
Climate Monitoring and Prediction
Cloud Properties
Optical/Infrared, Radiative Transfer
Radiation (Natural) Within the Earth's Environment
Surface Radiative Fluxes
Water and Energy Cycles

EARTH SYSTEM MODELS

Andrea Donnellan
Science Division, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

Definition

Earth system. Various parts of the Earth that interact with each other.
Models. A representation, either physically, graphically, or using a computer, of a scientific process or phenomenon developed for providing understanding of the real process or phenomenon.
Earth system model. A representation through graphics, computer software, or physically of interacting parts of the Earth developed to provide improved understanding.

Introduction

The solid Earth includes the core, mantle, lithosphere and crust, and the cryosphere consisting of ice sheets and glaciers. The core is liquid with a solid middle and generates our Earth's magnetic field. The mantle is made of hot rock

that convects, driving plate tectonics in the lithosphere and crust. Earthquakes and volcanoes occur as a result of plate tectonics. The ice masses that make up the cryosphere interact with the solid Earth, oceans, and atmosphere.

Satellite observations are used to probe characteristics of the solid Earth from the outer liquid core to the Earth's surface.

Core

The Earth's core is made up of a solid inner core and a liquid outer core. The outer core is composed of a liquid metal alloy and forms a layer 2,260 km thick. The electrically conducting liquid metal alloy rotates and convects, generating the Earth's magnetic field. The motions within the liquid outer core vary, resulting in both long-term and short-term changes in the magnetic field of the Earth. The Earth's magnetic dipole or north and south poles are generated from the outer core and change slowly, resulting in magnetic reversals that occur on timescales of millions of years. Shorter-term variations are superposed on the long-term variations.

Measurements of the magnetic field provide a way of probing the properties of the flow in the outer core. It is necessary to separate out the magnetic field from the magnetized rocks in the Earth's crust, electrical currents from the ocean, atmosphere, and ionosphere, as well as external fields. Measurements that have been carried out over the last 150 years indicate that the Earth's magnetic field has decayed by 10 %, consistent with an ongoing reversal of the magnetic poles. The bulk of the decay is occurring at the South Atlantic Ocean. Remote sensing can detect magnetic variations on the order of 1 year, which provides information on how the fluids move within the outer core. Remote sensing measurements can also further separate the internal and external processes that generate the Earth's magnetic field. Understanding the motion of the fluids in the outer core will result in a better understanding of the motions of the oceans and atmospheres and their circulation by allowing that contribution to be separated out from the observed total angular momentum budget of the Earth.

Mantle

The mantle is a layer about 2,900 km thick between the core and the crust. The mantle is composed primarily of crystalline silicate rock, which is hot and also convects. Seismic data and gravity data collected by remote sensing indicate that there is a discontinuity in material at about 400–670 km, which marks the boundary between the upper and lower mantle. The seismic waves are deflected by the discontinuity. Scientists predict the gravitational field of the Earth based on what they know about the distribution of densities, and they then compare their prediction to ground- and/or space-based measurements of the gravitational field of the Earth. Gravity anomalies are the differences between the observed and predicted gravity field. The convection of the mantle results in varying mass distributions throughout the mantle. Gravity anomalies have been used to show that the viscosity of the mantle increases dramatically at the boundary between the upper and lower mantle, suggesting that convection occurs in the upper mantle separately from the lower mantle.

The asthenosphere makes up the upper part of the mantle and is about 200 km thick. The rocks in the asthenosphere flow readily and are typically made up of olivine. The lithosphere, which is marked at the bottom by the transition from ductile to brittle olivine, makes up the very top of the mantle and the lower part of the Earth's crust and is made up of rigid rocks. Convection of the mantle drives plate tectonics and molten rock from the mantle, comes to the surface by way of volcanoes, and spreads ridges between the tectonic plates.

Lithosphere and crust

The rigid lithosphere is typically 50–200 km thick under the oceans and is up to 200 km thick under the continents. The Earth's crust is found in the top 30–50 km of the lithosphere over continents. The chemical composition of the rocks changes in the crust, and the base is marked by the Moho discontinuity, determined from seismology.

Plate tectonics

The mantle convection that drives plate tectonics creates convergent zones, where one plate subducts under another plate, transform, and extensional zones. Transform plate boundaries are typically found in the ocean but can also occur on land, such as in California. Extension is typically found at spreading ridges in the ocean. Hot spots, which are upwellings of hot material from deep in the mantle, can also cause spreading in the crust. The Basin and Range Province in the Western United States is an example of spreading induced by a hot spot. Volcanoes with different characteristics can form over hot spots, spreading centers, and subduction zones.

Seismology has been used extensively to understand earthquakes as well as characteristics of the solid Earth and of volcanoes. The advent of remote sensing has allowed for the measurement of surface motions to millimeter precision. This makes it possible to measure plate tectonics, comparing the rates to those determined through other geologic techniques. Measurement of surface deformation yields information about the material properties such as strengths and rigidity of the lithosphere and mantle and provides insight into the complete strain accumulation and release related to earthquakes and volcanoes.

Earthquakes

Earthquakes occur in the brittle part of the crust and mantle. They occur when stress increases, typically from motion due to the tectonic plates, and exceeds a threshold, at which point the crust breaks. The resulting shaking is an earthquake, which can be measured through the use of seismometers. However, the earthquake makes up only a small part of the entire earthquake cycle of strain

accumulation and release. Earthquakes typically occur along faults that have broken many times in the past. They are associated with convergent, transform, and extensional plate tectonics. There are three major types of faults, which are thrust, strike-slip, and normal faults.

Thrust faults and thrust earthquakes occur where the crust is shortening or converging. In these types of earthquakes, one block slides up over the bottom block along a fault oriented diagonally from depth toward the surface. Thrust earthquakes typically occur at subduction zones but can also occur where there is mantle downwelling or plate collision underneath the crust. Subduction zones are found along the margins of the Pacific plate along the coasts of South America, the Pacific Northwest of the United States, Canada, and Alaska, and underneath Japan. A prime example of a collision zone is the Himalayas. The thrust earthquakes in Southern California are associated with a downwelling underneath the Transverse Ranges.

Strike-slip earthquakes result when one side of a fault slips horizontally in either direction on both sides of a fault. Strike-slip faults occur along transform boundaries that offset spreading centers along divergent plate boundaries. The San Andreas Fault in California is a major strike-slip fault and makes up the primary boundary between the Pacific and North American Plates. The actual plate boundary is about 100 km wide, however.

Normal faults occur in regions where extension is occurring. A normal fault is also oriented diagonally from the upper part to the lower part at depth, but in this case, the upper block moves down and away relative to the lower block along the fault. The Basin and Range Province of the Western United States produces many normal faults. As the crust stretches, blocks in the valley, known as grabens, move down. The mountains move upward and are referred to as horsts.

Various techniques can be used to measure the strain associated with the earthquake cycle. Ground-based observations provided measurement of strain. The Global Positioning System (GPS) is now used to measure surface deformation at mm/year accuracy around the globe. The measurements provide time continuous precise measurements at distinct points. Another technique, called interferometric synthetic aperture radar (InSAR), relies on radar waves transmitted from a satellite and reflected back to the satellite from the Earth's surface. When the satellite exactly repeats its track, then differences in the radar return are due to surface motions, as well as error sources. The two observations are coupled together and form an interferogram where the radar waves interfere due to the surface motions. InSAR provides a precise map of deformation of the Earth's surface to/from the satellite.

Faults typically creep at depth, where the crust or mantle is hotter, and are locked at the surface except for when they break in earthquakes. The brittle part of the upper crust behaves elastically and moves in a predictable manner when strained. Observations of surface deformation yield information about the entire earthquake cycle, the cycle of strain accumulation and release, and provide information about fault activity and slip rates.

Recent work using remote sensing has demonstrated that silent slip events that mimic earthquake slip without the shaking are fairly common. Additionally a substantial amount of continued motion, referred to as postseismic response, can occur following earthquakes. The postseismic motions can result from relaxation of the more ductile lower crust or continued slip on the main fault plane. Models that incorporate seismic and crustal deformation are providing a system level understanding of earthquake processes and are improving assessments of earthquake hazards.

Volcanoes

Volcanoes are ruptures at the Earth's surface that allow magma and gases to escape from the Earth's interior. There are three primary types of volcanoes associated with divergent and convergent plate boundaries and hot spots. At spreading centers associated with divergent plate boundaries, the crust is typically thin, and the magma is of low viscosity and basaltic, coming from melted mantle. Subduction zone volcanoes associated with divergent plate boundaries tend to be explosive as the high-silicate-content subducting crust is heated and rises. A rising plume from a mantle hot spot also creates volcanoes. In cases such as Hawaii, chains of volcanoes form as the tectonic plate moves over the relatively fixed hot spot.

Magma moving beneath the surface of the Earth causes the surface of the Earth to deform and change temperature. GPS and interferometric radar (InSAR) are used to measure surface deformation, while short-wavelength infrared data are used to measure thermal characteristics of the surface. These observations are used to infer the location and migration of magma at depth, which are then used to determine the activity of a volcano and likelihood of eruption.

Cryosphere

Ice masses covering the Earth make up the cryosphere. This includes ice sheets, glaciers, and sea ice. The cryosphere has shown tremendous variability in the geologic record. The components of the cryosphere are sensitive to inputs and thermal exchanges from the atmosphere, ocean, and solid Earth. Volcanoes, hot spots, and subsurface topography, and geology at the base can control behavior of ice masses that overlie them. Ocean sea level and temperature affect the margins of ice sheets, which can lead to changes in the interiors. Changes in atmospheric temperature and precipitation can also modify ice sheet behavior. Remote sensing can be used to understand long-term ice sheet and glacier records as well as present-day ice dynamics.

Glacio-isostatic adjustment

Ice sheets depress the crust and mantle that underlie them. As ice sheets thicken, the surface of the rock is depressed downward, flexing the crust beyond the extent of the ice

sheet. When ice sheets thin, the opposite occurs and the crust and mantle rebound. However, the mantle is viscous and as a result has a delayed response for complete rebound. Knowing the mantle viscosity allows inferences to be made about past ice sheet history. Conversely, understanding ice sheet history can provide insight into mantle viscosity.

GPS measurements provide the rates and distribution of present-day glacio-isostatic adjustment. Time-varying gravity measurements from space indicate locations of mass changes, which can be related to glacio-isostatic adjustment. These measurements are used in Earth system models to understand past ice sheet history and crust and mantle properties.

Ice sheet dynamics

Ice sheets are quite variable on short timescales as well as on long timescales. The mass balance of an ice sheet refers to whether the ice sheet is thickening or thinning (losing or gaining mass). Understanding the mass balance of ice sheets requires knowledge of the thickness changes of the ice sheet, but also the flow rates of the ice sheet. Repeat altimetry, such as laser altimetry from LIDAR (Light Detection and Ranging), informs scientists regarding height changes of the ice sheet, which can be translated to thickness changes. Adding in flow measured from GPS or InSAR allows for a complete measurement of the volume changes of the ice sheet. Flow can also be inferred from tracking features in visible images. It is also possible to observe where the changes are occurring and whether those changes are migrating with time. For example, the Greenland ice sheet measurements indicate that the thinning started near the coast and is migrating inland, suggesting that ocean changes are a bigger driver than atmospheric changes. Earth system models take into account the basal characteristics of the ice sheet, flow and thickness parameters, as well as ocean and atmospheric coupling.

Ice shelves floating on the margins of ice sheets can be quite unstable. Radar and other imagery can be used to observe breakup of ice shelves as it occurs. The breakup of ice shelves also needs to be factored into flow models of the adjacent ice sheet.

Sea ice

Changes in sea ice cover provide indicators of how the climate is changing and also impacts shipping lanes. Radar and other imagery can be used to track sea ice flow, coverage characteristics, and floe distributions. Altimetry, such as from LIDAR, can be used to estimate the sea ice freeboard, or the amount of sea ice floating above the surface. Passive microwave remote sensing is used to measure the thermal, dielectric, and geophysical properties of snow on the sea ice. The measured parameters coupled with models show the annual evolution of the Arctic sea ice, atmosphere, and ocean system.

Conclusions

The solid Earth functions as a system, ranging from the core through the mantle to the crust and cryosphere. The geodynamo operating within the core provides the Earth's magnetic field, which varies due to flow within the core. The mantle interacts with and is a key component of plate tectonic, earthquake, and volcanic systems, and earthquakes and volcanoes result from plate tectonics. The cryosphere also interacts with the crust and the mantle depressing the solid Earth surface due to the mass of ice sheets. Tectonic plate motions can in turn affect ice sheets. It is important to model the Earth as a system in order to separate out different effects and to understand the individual components of the system.

Remote sensing offers a way of probing various components of the solid Earth, providing key parameters to Earth system models. The mechanisms that drive various solid Earth processes primarily occur underneath the surface and as such are not directly observable. Furthermore, remote sensing provides a regional view that is more difficult to obtain with in situ measurements.

Acknowledgments

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

Bibliography

Bamber, J. L., and Rivera, A., 2007. A review of remote sensing methods for glacier mass balance determination. *Global and Planetary Change*, **59**(1–4), 138–148.

Braun, A., Kuob, C.-Y., Shum, C. K., Wu, P., van der Wal, W., and Fotopoulos, G., 2008. Glacial isostatic adjustment at the Laurentide ice sheet margin: models and observations in the Great Lakes region. *Journal of Geodynamics*, **46**(3–5), 165–173.

Bürgmann, R., Rosen, P. A., and Fielding, E. J., 2000. Synthetic aperture radar interferometry to measure earth's surface topography and its deformation. *Annual Review of Earth and Planetary Sciences*, **28**, 169–209.

Board on Atmospheric Sciences and Climate, 2008. *Earth Observations from Space: The First 50 Years of Scientific Achievements*.

Delouis, B., Giardini, D., Lundgren, P., and Salichon, J., 2002. Joint inversion of InSAR, GPS, teleseismic, and strong-motion data for the spatial and temporal distribution of earthquake slip: application to the 1999 Izmit Mainshock. *Bulletin of the Seismological Society of America*, **92**(1), 278–299.

Donnellan, A., Parker, J. W., and Peltzer, G., 2002. Combined GPS and InSAR models of postseismic deformation from the Northridge earthquake. *Pure and Applied Geophysics*, **159**, 10.

Glaze, L., Francis, P. W., and Rothery, D. A., 1989. Measuring thermal budgets of active volcanoes by satellite remote sensing. *Nature*, **338**, 144–146.

Hager, B. H., 1984. Subducted slabs and the geoid: constraints on mantle rheology and flow. *Journal of Geophysical Research*, **89**, 6003–6016.

http://www.es.ucsc.edu/~glatz/geodynamo.html

http://www.esa.int/esaLP/ESA3QZJE43D_LPswarm_0.html.

Jeanloz, R., 1990. The nature of the earth's core. *Annual Review of Earth and Planetary Sciences*, **18**, 357–386.

Korsnes, R., 1993. Quantitative analysis of sea ice remote sensing imagery. *International Journal of Remote Sensing*, **14**(2), 295–311.
Langlois, A., and Barber, D. G., 2007. Passive microwave remote sensing of seasonal snow-covered sea ice. *Progress in Physical Geography*, **31**, 539.
Rogers, G., and Dragert, H., 2003. Episodic tremor and slip on the cascadia subduction zone: the chatter of silent slip. *Science*, **300**(5627), 1942–1943.
Segall, P., Desmarais, E. K., Shelly, D., Miklius, A., and Cervelli, P., 2006. Earthquakes triggered by silent slip events on Kilauea volcano, Hawaii. *Nature*, **442**, 71–74.
Solomon, S. C., 2002. *Living on a Restless Planet*. *Solid Earth Science Working Group Report.* National Aeronautics and Space Administration. JPL 400–1040.
Stevenson, D. J., 1981. Models of the earth's core. *Science*, **214**, 4521.
Tralli, D. M., Blom, R. G., Zlotnicki, V., Donnellan, A., and Evans, D. L., 2005. Satellite remote sensing of earthquake, volcano, flood, landslide and coastal inundation hazards. *ISPRS Journal of Photogrammetry and Remote Sensing*, **59**(4), 185–198.
Zebker, H. A., Falk, A., and Jonsson, S., 2000. Remote sensing of volcano surface and internal processes using radar interferometry. *Geophysical Monograph*, **116**, 179–205.

Cross-references

Atmospheric General Circulation Models
Calibration and Validation
Calibration, Microwave Radiometers
Calibration, Optical/Infrared Passive Sensors
Calibration, Scatterometers
Calibration, Synthetic Aperture Radars
Climate Data Records
Climate Monitoring and Prediction
Cryosphere and Polar Region Observing System
Cryosphere, Climate Change Effects
Cryosphere, Climate Change Feedbacks
Cryosphere, Measurements and Applications
Data Access
Data Archival and Distribution
Data Archives and Repositories
Data Assimilation
Data Processing, SAR Sensors
Decision Fusion, Classification of Multisource Data
Emerging Technologies, LIDAR
Emerging Technologies, Radar
Emerging Technologies, Sensor Web
Geodesy
Geomorphology
Geophysical Retrieval, Overview
Geophysical Retrieval, Inverse Problems in Remote Sensing
Global Earth Observation System of Systems (GEOSS)
Global Land Observing System
Ice Sheets and Ice Volume
Magnetic Field
Ocean Modeling and Data Assimilation
Remote Sensing, Physics and Techniques
Sea Ice Albedo
Sea Ice Concentration and Extent
Sea Level Rise
Solid Earth Mass Transport
Subsidence

ELECTROMAGNETIC THEORY AND WAVE PROPAGATION

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

Definition

Electromagnetic theory. Theory based on the Maxwell equations to study the properties and dynamics of electromagnetic fields.
Wave propagation. The process of transporting electromagnetic energy.

Maxwell equations

The Maxwell equations lay the foundation of all classical electromagnetic phenomena. They describe how time-varying electric fields give rise to magnetic fields and vice versa. The Maxwell equations are a set of four equations

$$\nabla \cdot \mathbf{D} = \rho (\text{Coulomb's law})$$
$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} (\text{Faraday's law}) \quad (1)$$
$$\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t} + \mathbf{J}$$
$$(\text{Ampere's law with displacement current addition})$$
$$\nabla \cdot \mathbf{B} = 0 (\text{Absence of free magnetic poles})$$

The six quantities in terms of which the electromagnetic field equations are expressed are as follows:

- $\mathbf{D}$ = Electric flux density or electric displacement (coulombs per square meter)
- $\mathbf{E}$ = Electric intensity (volts per meter)
- $\mathbf{B}$ = Magnetic flux density (Weber's per square meter)
- $\mathbf{H}$ = Magnetic intensity (amperes per meter)
- $\mathbf{J}$ = Electric current density (amperes per square meter)
- ρ = Electric charge density (coulombs per cubic meter)

In vacuum $\mathbf{D} = \varepsilon_0 \mathbf{E}$ and $\mathbf{B} = \mu_0 \mathbf{H}$, where $\varepsilon_0 \approx 8.85 \times 10^{-12}$ farad/m is the permittivity, and $\mu_0 = 4\pi \times 10^{-7}$ Henry/m is the permeability. The continuity equation for charge density and current density can be derived from the above Maxwell equations

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} = 0 \quad (2)$$

When electromagnetic wave propagates in some medium other than vacuum, since the characteristic wavelength is very large compared to the atoms of which the medium is composed, the detailed behavior of the fields over atomic distance can be neglected. What are relevant are the quantities averaged over the atomic scale, including the macroscopic fields and macroscopic

sources. The constitutive relations, which relate **D** and **B** to **E** and **H**, may take forms much more complicated than those in vacuum (see *Media, Electromagnetic Characteristics*).

Boundary conditions

Maxwell equations in the form of Equation 1 concern ordinary points of space in whose neighborhood the physical properties of the medium vary continuously. However, at boundary of two different media, the permittivity, permeability, and conductivity can change sharply. The discontinuous change on a macroscopic scale will lead to corresponding discontinuity in the field vectors. The discontinuity equations are collectively called boundary conditions (Jackson, 1998; Kong, 2005).

The boundary conditions for the normal components of the electric flux density **D** and magnetic flux density **B** on either side of the boundary surface are

$$\begin{aligned}(\mathbf{D_1}-\mathbf{D_2})\cdot\hat{n}&=\sigma\\(\mathbf{B_1}-\mathbf{B_2})\cdot\hat{n}&=0\end{aligned}\tag{3}$$

where the normal $\hat{n}$ is pointing from medium 1 into medium 2, σ is an idealized surface current density that is related to the charge density ρ singular at the interface by $\int_{\Delta V}\rho dv'=\sigma\Delta s$, and ΔV is an infinitesimal pillbox of area Δs and height Δh, with half of its volume in each medium and its top and bottom parallel to the surface. Equation 3 states that at any point on the boundary the normal component of **B** is continuous, while that of **D** is discontinuous with the discontinuity equal to the surface charge density.

The boundary conditions for the tangential components of **E** and **H** on either side of the boundary surface are

$$\begin{aligned}\hat{n}\times(\mathbf{E}_2-\mathbf{E}_1)&=0\\\hat{n}\times(\mathbf{H}_2-\mathbf{H}_1)&=\mathbf{J}_s\end{aligned}\tag{4}$$

where $\mathbf{J}_s$ is an idealized surface current density flowing on the boundary surface which satisfies the condition $\int_{\Delta S}\left(\frac{\partial\mathbf{D}}{\partial t}+\mathbf{J}\right)\cdot\hat{t}ds=\hat{t}\cdot\mathbf{J}_s\Delta l$ and ΔS is an open surface spanning an infinitesimal contour of line Δl parallel to the surface, with half of its area in each medium and its normal $\hat{t}$ tangent to the surface. Equation 4 states that at any point on the boundary the tangential component of **E** is continuous, while that of **H** is discontinuous by an amount of identical magnitude with direction parallel to $\mathbf{J}_s\times\hat{n}$.

Scalar and vector potentials

Potentials are useful auxiliary functions in analysis of an electromagnetic field. This is because coupling between the electric and magnetic fields in the Maxwell equation makes direct solution difficult, except in simple situations. Potentials, however, if introduced properly, can lead to decoupled equations of their own while identically satisfying some of the Maxwell equations. The potentials commonly introduced are the scalar potential ϕ and vector potential **A** (Jackson, 1998; Kong, 2005).

Since the magnetic flux density **B** is always solenoidal, i.e., $\nabla\cdot\mathbf{B}=0$, it can be represented as the curl of the vector potential **A**.

$$\mathbf{B}=\nabla\times\mathbf{A}\tag{5}$$

This representation, together with Faraday's law, implies that the electric intensity **E** can be expressed in terms of the scalar and vector potentials as

$$\mathbf{E}=-\nabla\phi-\frac{\partial\mathbf{A}}{\partial t}\tag{6}$$

Consider the case where the medium is homogeneous and isotopic, for which the constitutive relations are $\mathbf{D}=\varepsilon\mathbf{E}$, $\mathbf{B}=\mu\mathbf{H}$. The arbitrariness of choice of **A** inherent in Equation 5 and of ϕ through Equation 6 allows the choice of a set of potentials $(\mathbf{A},\phi)$ to satisfy the Lorenz condition,

$$\nabla\cdot\mathbf{A}+\mu\varepsilon\frac{\partial\phi}{\partial t}=0\tag{7}$$

Making use of the remaining two Maxwell equations leads to two inhomogeneous wave equations, uncoupled for ϕ and **A** as

$$\begin{aligned}\Delta^2\phi-\mu\varepsilon\frac{\partial^2\phi}{\partial t^2}&=-\frac{\rho}{\varepsilon}\\\Delta^2\mathbf{A}-\mu\varepsilon\frac{\partial^2\mathbf{A}}{\partial t^2}&=-\mu\mathbf{J}\end{aligned}\tag{8}$$

These equations, equivalent to the Maxwell equation in all respect in a homogeneous and isotopic medium, can be conveniently analyzed using the Green's function.

Green's function for the wave equation

The underlying structures are identical for the wave equations (Jackson, 1998)

$$\nabla^2\psi-\mu\varepsilon\frac{\partial^2\psi}{\partial t^2}=-f(\mathbf{r},t)\tag{9}$$

with $f(\mathbf{r},t)$ being a known source distribution. After Fourier transform, $\psi(\mathbf{r},\omega)$ satisfies the inhomogeneous Helmholtz wave equation

$$(\nabla^2+k^2)\psi(\mathbf{r},\omega)=-f(\mathbf{r},\omega).\tag{10}$$

The corresponding Green's function $g(\mathbf{r},\mathbf{r}')$ satisfies the inhomogeneous equation

$$(\nabla^2+k^2)g(\mathbf{r},\mathbf{r}')=-\delta(\mathbf{r}-\mathbf{r}')\tag{11}$$

where $\delta(\mathbf{r})$ is the Dirac delta function. In an unbounded medium, the Green's function is

$$g(\mathbf{r},\mathbf{r}')=\frac{e^{ik|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|}\tag{12}$$

where $k = \omega\sqrt{\mu\varepsilon}$ is the wave number in the medium and $|\mathbf{r} - \mathbf{r}'|$ is the distance between the field point $\mathbf{r}$ and the source point $\mathbf{r}'$. This form of Green's function represents a diverging spherical wave propagating from the origin.

The above Green's function is the scalar Green's function because it relates scalar field to scalar source. If both field and source are vectors, dyadic Green's function must be used instead which satisfies the inhomogeneous equation (Kong, 2005)

$$\nabla \times \nabla \times \overline{\overline{G}}(\mathbf{r}, \mathbf{r}') - k^2 \overline{\overline{G}}(\mathbf{r}, \mathbf{r}') = \overline{\overline{I}}\delta(\mathbf{r} - \mathbf{r}') \tag{13}$$

where $\overline{\overline{I}}$ is the identity dyad. The dyadic Green's function can be expressed in terms of the scalar Green's function as

$$\overline{\overline{G}}(\mathbf{r}, \mathbf{r}') = \left[\overline{\overline{I}} + \frac{1}{k^2}\nabla\nabla\right] g(\mathbf{r}, \mathbf{r}') \tag{14}$$

Poynting's theorem and conservation of energy

Poynting's theorem is a law of conservation of energy for the electromagnetic field. From the Maxwell equations, considering the fields in a finite volume V, the following equation can be readily established (Jackson, 1998):

$$-\int_V \mathbf{J} \cdot \mathbf{E} dv' = \int_V \left[\nabla \cdot (\mathbf{E} \times \mathbf{H}) + \mathbf{E} \cdot \frac{\partial \mathbf{D}}{\partial t} + \mathbf{H} \cdot \frac{\partial \mathbf{B}}{\partial t}\right] \tag{15}$$

Each term of the above equation has physical meaning. The left side of the equation represents the negative of the total work done by the fields on the sources within the volume. Integral of the first term on the right side of the equation represents the energy that flows out of the volume through the boundary surfaces per unit time. Since energy flow is such an important concept, to represent it a vector $\mathbf{S}$ is specifically defined which is called the Poynting vector given by

$$\mathbf{S} = \mathbf{E} \times \mathbf{H}. \tag{16}$$

The last two terms on the right side are collectively associated with the total energy density u (electric and magnetic energy density combined) for a linear medium with negligible dispersion or losses. To be specific, they are the time rate of change of the total energy density, i.e., $\frac{\partial u}{\partial t}$. Integral of the last two terms is then the time rate of change of electromagnetic energy within a certain volume. Equation 15 is a statement of conservation of energy; the diminution of electromagnetic energy stored in V is partly due to electromagnetic energy flows outward across the boundary surfaces and partly due to a conversion of electromagnetic energy into mechanical or heat energy.

For more general case of linear dispersive media, interpretation of the right side of Equation 15 is more involved.

For harmonic electromagnetic fields, it is more convenient to use the complex Poynting vector given by $\mathbf{S} = \frac{1}{2}(\mathbf{E} \times \mathbf{H}^*)$.

Plane wave

Plane waves are the simplest and fundamental electromagnetic waves. All wave functions can be expressed as superpositions of plane waves. In an unbounded uniform isotropic linear media, where $\mathbf{D} = \varepsilon\mathbf{E}$, $\mathbf{B} = \mu\mathbf{H}$, the harmonic plane wave fields are

$$\begin{aligned} \mathbf{E}(\mathbf{r}, t) &= \mathbf{E}e^{i\mathbf{k}\cdot\mathbf{r} - i\omega t} \\ \mathbf{H}(\mathbf{r}, t) &= \mathbf{H}e^{i\mathbf{k}\cdot\mathbf{r} - i\omega t}, \end{aligned} \tag{17}$$

which satisfy the homogeneous Helmholtz wave equation. The Maxwell equations demand that both $\mathbf{E}$ and $\mathbf{H}$ are perpendicular to the vector $\mathbf{k}$. If $\mathbf{k}$ is real, then it can be written as $\mathbf{k} = k\hat{k}$, where $k = \omega\sqrt{\mu\varepsilon}$ is the wave number and $\hat{k}$ is a real unit vector indicating the direction of propagation. If $\mathbf{k}$ is complex, then it is expressed through two real vectors as $\mathbf{k} = \beta + i\alpha$, and the plane wave represented by Equation 17 propagates in the direction of β and attenuates in the direction of α (Harrington, 1961).

If the media are semi-infinite with different properties separated by a plane surface, then at the interface electromagnetic waves experience reflection and refraction. Angle of reflection θ_r equals angle of incidence θ_i, while angle of refraction θ_t is determined by Snell's law, which states that the ratio of sine of angle of refraction to sine of angle of incidence equals the ratio of index of refraction n in the medium of incidence to index of refraction n' in the medium of transmission. The Fresnel reflection coefficient, which specifies the ratio of the reflected electric intensity to the incident electric intensity, depends on how the electric field is directed. If the electric field is perpendicular to plane of incidence, then the Fresnel reflection coefficient is (Jackson, 1998)

$$R_\perp = \frac{\cos\theta_i - \frac{\mu}{\mu'}\frac{n'}{n}\cos\theta_t}{\cos\theta_i + \frac{\mu}{\mu'}\frac{n'}{n}\cos\theta_t} \tag{18}$$

where μ and μ' are the permeability in the medium of incidence and of transmission, respectively. If the electric field is parallel to plane of incidence, then the Fresnel reflection coefficient is

$$R_{//} = \frac{\frac{\mu}{\mu'}\frac{n'}{n}\cos\theta_i - \cos\theta_t}{\frac{\mu}{\mu'}\frac{n'}{n}\cos\theta_i + \cos\theta_t}. \tag{19}$$

There is an important aspect associated with this case; when angle of incidence equals Brewster's angle, there is no reflected wave. Brewster's angle can be simply determined by $\theta_B = \tan^{-1}\left(\frac{n'}{n}\right)$ when $\mu = \mu'$.

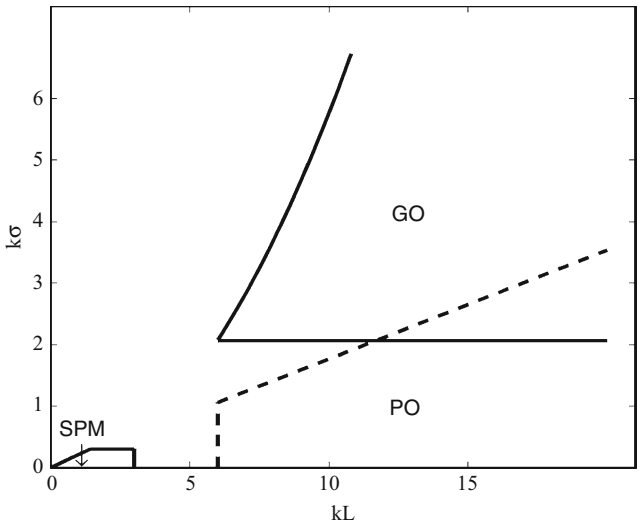
If index of refraction in the medium of incidence is larger than that in the medium of refraction, the phenomenon called total internal reflection can occur when there is no energy flow across the surface.

Electromagnetic wave scattering from randomly rough surfaces

Electromagnetic wave scattering from randomly rough surfaces is fundamental to geophysics, terrestrial, and extraterrestrial remote sensing, and wireless and satellite communications. This field is a rich and active research area, and it has become impossible to cover all recent advances in analytical approaches here. The readers are referred to earlier general survey papers or books on analytical models (Elfouhaily and Guerin, 2004; Fung, 1994; Tsang et al., 1985; Ulaby et al., 1982).

The complex character of electromagnetic wave interaction with rough surface may require different classes of models in different situations. A qualitative sketch of regions of validity for analytical models is shown in Figure 1. For the region where the RMS height is very small compared to the wavelength, the small perturbation method (SPM) is applicable, which is derived rigorously from the extended boundary condition without resorting to any a priori assumption about the field. For the region where surfaces have large surface curvatures, the Kirchhoff approximation (KA) is commonly used. It is a local approximation in that it uses tangent plane approximation by assuming the surface field to be generated as if a tangent plane were at the same point while ignoring the contribution from the surface elsewhere. In the high-frequency regime, by invoking stationary phase method, the scattered amplitude reduces to a probability density function of slopes that are evaluated at specular points. In this case KA is also called the geometrical optics (GO) method. For the regions in between, there are active ongoing researches in seek of so-called unifying methods. In the survey by Elfouhaily and Guerin (Elfouhaily and Guerin, 2004), these unifying methods are classified into generic families, including the Meecham-Lysanov method, the phase perturbation method (PPM), the small slope approximation (SSA), the operator expansion method (OEM), the tilt invariant approximation (TIA), the local weight approximation (LWA), the Wiener-Hermite approach, the unified perturbation expansion (UPE), the full wave approach (FWA), the improved Green's function methods, the volumetric methods, and the integral equation method (IEM). Considering the wide use of the IEM model in the analysis of scattering from terrestrial rough surfaces on the one hand and the presentation of IEM and its related models in (Elfouhaily and Guerin, 2004) being very brief or barely touched on the other hand, these models will be elaborated a bit here to shed more light into their respective strength and limitations so as to help the practitioner make more intelligent choice.

The original IEM model (Fung, 1994) has shown to provide good predictions for forward and backward scattering coefficients. To make the derivation of the IEM model mathematically tractable, several assumptions were made, which include the following (Chen et al., 2003):



Electromagnetic Theory and Wave Propagation,
Figure 1 Qualitative sketch of regions of validity for the rough surface scattering models.

1. Spatial dependence of the local incident angle of the Fresnel reflection coefficient is removed, by either replacing it with the incident angle or the specular angle.
2. For the cross-polarization, the reflection coefficient used to compute the Kirchhoff fields is approximated by $(R_{//} - R_{\perp})/2$, where $R_{//}$ and $R_{\perp}$ are the Fresnel reflection coefficients of transverse magnetic and transverse electric waves, respectively.
3. Edge diffraction terms are excluded.
4. Complementary field coefficients are approximated by simplifying the surface Green's function and its gradient in the phase terms.

Concerns over the assumptions have prompted several modifications and variations of IEM in the literature. Regarding the spectral representation of the Green's function, the simplification was discarded and full form was restored, resulting in a modification to the complementary components. The resulted model is the so-called improved IEM model (I-IEM) (Hsieh and Fung, 1999). Additional restoration of the spectral representation of the gradient of the Green's function in its full form leads to the advanced IEM model (AIEM) (Chen et al., 2003) and the IEM2M model (Alvarez-Perez, 2001).

However, there are some technical subtleties in connection with the restoration of the full Green's function that have not been adequately reflected in these models. For example, in calculating the Kirchhoff-complementary incoherent power, one of the new quantities which need to be evaluated is the following:

$$I_F = < \exp\{-i[k_{sz}z + k_z z' - q|z - z'|]\} >$$

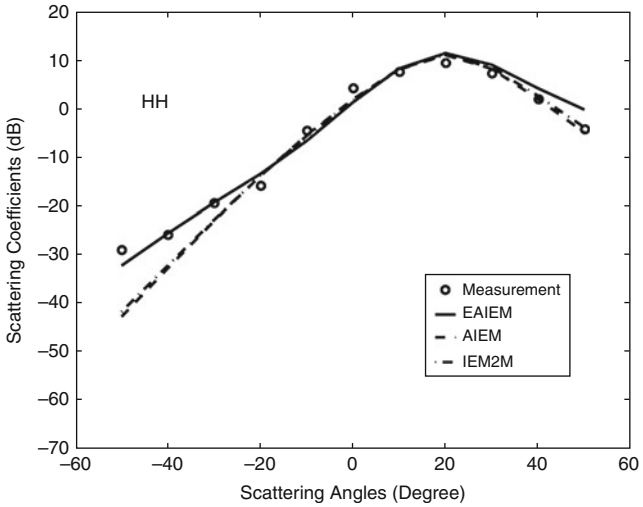
where k_{sz} and k_z are the z-components of the scattered and incident wave vectors, respectively; q is the z-related term

in the spectral representation of the Green's function; z and z' are the heights of two points on the surface, respectively, with the surface being assumed to be a Gaussian process; and the corner brackets denote the expected value. After introducing transformation of variables, one finds that I_F contains a factor of the following form:

$$I_{F2} = \frac{1}{\sqrt{2\pi\sigma\sqrt{1-\rho}}} \times \int_{-\infty}^{+\infty} dy_2 \left\{ \exp\left\{ -i\left[\frac{k_{sz}y_2}{\sqrt{2}} - \frac{k_z y_2}{\sqrt{2}} - \sqrt{2}q|y2| \right] \right\} \times \exp\left[-\frac{{y_2}^2}{2\sigma^2(1-\rho)} \right] \right\},$$

where σ is the RMS height and ρ is the surface correlation function. Due to the presence of the absolute phase term, the result on I_{F2} will contain the error function. However, all the terms related to the error function have been ignored in the above models. After inclusion of the missing terms (Du, 2008), for the cross and the complementary incoherent powers, it is shown that each consists of two terms: one free from and the other related to the error function. Moreover, for the cross incoherent power, the term free from error function appears to be identical to that reported in the literature. So the other term related to error function serves as a correction term for previous calculations. It is not the case for the complementary incoherent power: The term free from error function takes a different form from that in the literature. The reason is that the evaluation of this term is more technically involved, and certain assumptions are required to make the derivation mathematically tractable. The assumptions tacitly made in Chen et al. (2003) become more explicit in Alvarez-Perez (2001). The work in Du (2008) departs from above models in another essential way by making fewer assumptions. As a result, the new model, called the enhanced advanced IEM (EAIEM), is expected to apply to a wider range with a better accuracy. This point is at least partially verified by the comparative performance (see Figure 2).

The weakness of the conventional IEM model can be examined from another angle. Assumption (1) of IEM is an oversimplification of the statistic behavior of the unit normal $\hat{n}$ directed out of the surface at an arbitrary point $\mathbf{r}$ on the surface. Specifically, $\hat{n}$ is effectively restricted to a deterministic quantity, with the direction depending on the angles of incidence and scattering, while giving up all its statistical features. Such simplification may introduce some unwanted effect on the predictive power of the model, since the statistical features of a unit normal can be rich and strong correlations can be observed between unit normal vectors. For example, consider a randomly rough surface described by a Gaussian process with Gaussian power spectrum. If $\hat{n}$ is expressed in terms of the spherical coordinates (θ_n, ϕ_n) as $\hat{n} = \hat{x}\sin\theta_n\cos\phi_n + \hat{y}\sin\theta_n\sin\phi_n + \hat{z}\cos\theta_n$, then the azimuthal angle ϕ_n and the elevation angle θ_n are

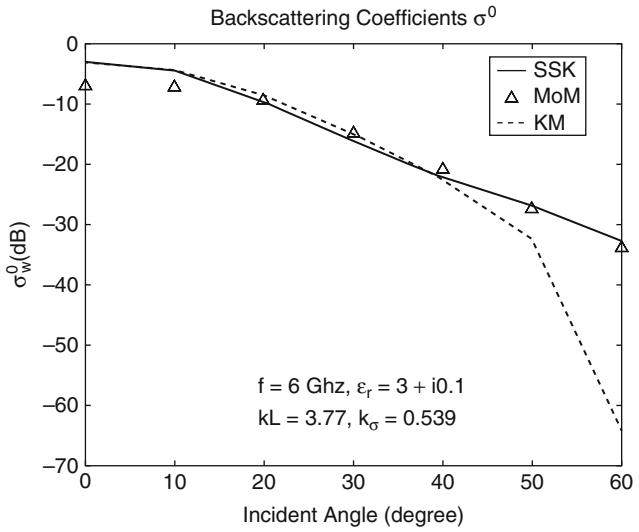


Electromagnetic Theory and Wave Propagation, Figure 2 Comparisons of bistatic scattering models EAIEM, AIEM, and IEM2M against measurement data for a Gaussian surface. The relative permittivity of the surface is $\varepsilon_r = 5.5 + i2.2$, and the frequency is 10 GHz, leading to a normalized correlation length $kL = 12.56$and a normalized surface RMS height $k\sigma = 0.84$. The incidence angle is fixed at $\theta_i = -20°$, while the scattering angle varies from $-50°$ to $50°$.

independent random variables, where ϕ_n is uniformly distributed in the interval $[0, 2\pi]$ and the probability density function of θ_n is

$$p(\theta_n) = \frac{1}{\sigma_s^2}\exp\left(-\frac{\tan^2\theta_n}{2\sigma_s^2}\right)\frac{\tan\theta_n}{\cos^2\theta_n},$$

where σ_s is the RMS slope. The distribution of θ_n, unlike treated in IEM, is independent of the combination of the angles of incidence and scattering. Moreover, for two-unit normal vectors $\hat{n}$ and $\hat{n}'$ at two points on the surface, respectively, if we express them in terms of the directional partial derivatives such that $\hat{n} = \frac{-\hat{x}Z_x - \hat{y}Z_y + \hat{z}}{\sqrt{1+Z_x^2+Z_y^2}}$ and form the four dimensional vector $\mu = \left[Z_x, Z_y, Z'_x, Z'_y\right]^t$, then μ is a multivariate Gaussian random vector with zero mean and covariance matrix C. Inclusion of the statistical behaviors of the surface unit normal vectors in conjunction with the Kirchhoff approximation in the analysis of electromagnetic scattering from a randomly rough surface was presented in Du et al. (2005). The new model was called the slope statistical Kirchhoff (SSK) model. It was shown by numerical simulations that the predictions of SSK are in better agreement with MoM results than those of the conventional Kirchhoff model (KM). A typical comparison is shown in Figure 3. One finds from this figure that besides an overall better performance, SSK appears almost immune to the Brewster angle effect for vertical polarization. This feature is expected since the directions of the unit normal which lead the local angles of incidence to approach the Brewster angle occupy only a small



Electromagnetic Theory and Wave Propagation,
Figure 3 Comparisons among MoM, KM, and SSK simulations at 6 GHz with $\varepsilon_r = 3 + i0.1$, correlation length $L = 3$ cm, RMS height $\sigma = 0.429$ cm.

portion of the directional distribution; contributions from the rest of the distribution become appreciable in this new treatment.

Inclusion of the statistical behaviors of the surface unit normal vectors in combination with the IEM formulism was presented in Du et al. (2007), where the shadowing effect is also included but is not expected to have any appreciable impact on the scattering behaviors predicted by the model for the range of angles of incidence considered, which is limited to be not large (say, under 70°), a restriction inherent to the IEM formulism. The new model was called the statistical IEM (SIEM) model. It was shown by numerical simulations that the predictions of SIEM are in better agreement with MoM results than those of the IEM model.

Electromagnetic wave scattering from vegetation canopy

The potential use of microwave observations to monitor vegetation water content (VWC) and soil moisture is of great importance (e.g., Storvold et al., 2006; Singh and Kathpalia, 2007). Detection of VWC is useful to monitor vegetation stress and important for irrigation management and yield forecasting. Soil moisture is often the limiting factor in transpiration of plants and evaporation from soil surface, which in turn has a significant impact on the energy cycle. Soil moisture is also a key determinant of the global carbon cycle.

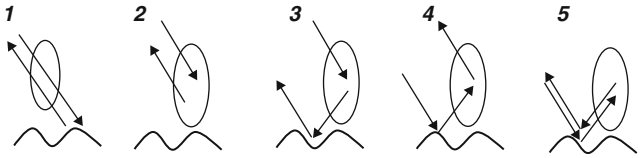
Yet to extract VWC and soil moisture from microwave observations presents a big challenge, which calls for good management of many important issues, among which are the development of a high fidelity scattering model.

In developing such scattering model, a number of factors associated with the vegetation canopy and with the underlying bare soil should be taken into account. Regarding the vegetation canopy, it is important to include the coherent effect caused by the vegetation structure at low frequencies (Zhang et al., 1995), an effect that has been well addressed in a number of recent models. For instance, the branching model due to Yueh et al. (1992) addressed the coherence effects caused by the vegetation structure, where a two-scale branching vegetation structure was used for soybeans, and the scattered fields from constituents were added coherently. A similar treatment in considering coherent effects was proposed in Lin and Sarabandi (1997) for forest canopies. Chiu and Sarabandi also considered the second-order, near-field interaction between constituents in addition to the coherence effect in their scattering model for soybean (Chiu and Sarabandi, 2000).

Yet the roughness effect of the underlying bare soil has not been adequately addressed in the above coherent models. Notarnicola and Posa, in their study of inferring VWC of corn and soybean from C- and L-band SAR images, observed that in the inversion procedure, the introduction of the dependence on roughness improves the estimates (Notarnicola and Posa, 2007). They inferred from such observation that, even for dense vegetation, the contribution from bare soil greatly influences the radar signal. To predict backscattering from the rough surface, the Kirchhoff approximation (KA) was used in Yueh et al. (1992), while a second-order small perturbation model (SPM) and a physical optic (PO) model were incorporated in Chiu and Sarabandi (2000). It is well known that the SPM model and the KA model are applicable for slightly rough surfaces and surfaces with small surface curvatures, respectively (Ulaby et al., 1982; Kong, 2005).

Regarding modeling of the vegetation, the vegetation constituents are represented by simple geometries as in Yueh et al. (1992*)* and Chiu and Sarabandi (2000). Specifically, stems, branches, and pods are modeled as dielectric circular cylinders of finite length. In Chiu and Sarabandi (2000) leaves are represented by elliptical thin dielectric disks, yet it was found through numerical simulation that unless ellipticity ratio is much larger than unity, the final backscatter is insensitive to the ellipticity ratio, so for simplicity leaves can be modeled as circular thin dielectric disks as in Yueh et al. (1992). The orientation distribution of the constituents is described by two angles: the elevation angle β and the azimuth angle γ, for the latter an azimuthal symmetry is assumed.

There are five major scattering mechanisms for a vegetation canopy: (1) direct backscatter from the underlying rough surface, (2) direct backscatter from soybean elements, (3) single ground bounce from scatterer to ground, (4) single ground bounce from ground to scatterer, and (5) double ground bounce. These scattering mechanisms are illustrated in Figure 4. Chiu and Sarabandi considered two additional scattering mechanisms (Chiu and Sarabandi, 2000), namely, second-order



Electromagnetic Theory and Wave Propagation,
Figure 4 Major scattering mechanisms for a vegetation canopy.

scattering interaction among vegetation constituents and scattering interaction between main stem and the rough surface. It was concluded that the latter mechanism is only considered for predicting the cross-polarized scattering at L-band (Chiu and Sarabandi, 1999). For the second-order near-field effect, good comparative illustrations with other scattering mechanisms were provided in Chiu and Sarabandi (2000) at both L- and C-band for two data sets: one from polarimetric measurements conducted using the University of Michigan polarimetric scatterometer systems (POLARSCAT) on a soybean field near Ann Arbor, MI, in August 1995, when the soybean plants were fully grown with significant numbers of pods, and the other from backscatter data collected by AIRSAR during its flight over the Kellogg Biological Station near Kalamazoo, MI, on July 12, 1995, when the soybeans were about a month old. It is observed from the decomposed contributions due to different scattering mechanisms that for not fully grown soybeans (the AIRSAR data set), contribution from rough surface is at least 25 dB higher than that from the second-order interaction for copolarized backscatter at L-band, and similar observation applies to the fully grown soybeans, so the authors concluded that the contribution from the second-order near field is negligible at L-band (Chiu and Sarabandi, 2000). At C-band, although the second-order near-field scattering is significant for fully grown soybeans as suggested by the authors, for not fully grown soybeans, contribution from rough surface is about 10 dB higher for the horizontally polarized backscatter up to 50°, while showing a cross-over with the second-order interaction for the vertically polarized backscatter at around 35°. Such observation indicates that it is more important to improve the predictive accuracy of scattering from rough surface than to include the second-order near-field effect, even at C-band, for copolarized backscatter at small to moderate incident angles. In fact, such statement may be further strengthened if one considers the fact that the estimated ground truth in Chiu and Sarabandi (2000) corresponds to a rather smooth surface, with a RMS height of 0.38 cm and a correlation length of 3.8 cm, which is several times smaller than the ground truth used in Yueh et al. (1992), where the RMS height is around 1.5 cm and the correlation length is about 13 cm. The ground truth of Yueh et al. (1992) seems to agree well with another experiment study on characterization of agricultural soil roughness for radar remote sensing (Davidson et al., 2000; De Roo et al., 2001). This suggests that the backscatter from the rough surface might be much stronger than that of Chiu and Sarabandi (2000) if ground truth comparable to that of Yueh et al. (1992), Davidson et al. (2000), and De Roo et al. (2001) were to be used.

In Du et al. (2008) a scattering model for a soybean canopy was proposed which includes the coherent effect due to the soybean structure and takes advantage of the advanced scattering models for rough surface. Some other issues were also taken care of, such as including curvature effect in studying the ground bounce scattering mechanisms and using array theory with perturbation for characterizing the interplant structure to account for the prevailing agriculture practice of soybean.

Electromagnetic wave scattering from cylinders

Determining the electromagnetic properties of those key constituents such as branches and trunks requires knowledge of the scattering properties of dielectric cylinders (Wang et al., 2005; Henin et al., 2007; Ahmed and Naqvi, 2008). In addition, in studying scattering and absorption of electromagnetic waves from ice needles in clouds, the dielectric cylinder are also used to model those needles (Yeh et al., 1982). Thus, finding an effective method to calculate the electromagnetic scattering by dielectric finite cylinders motivated many authors. Because an exact analytical solution for the scattering from finite cylinders does not exist, several approximations have been proposed (Schiffer and Thielheim, 1979; Karam and Fung, 1988; Stiles and Sarabandi, 1996). Among them is the generalized Rayleigh-Gans (GRG) approximation, which was widely applied in the studies of the vegetation samples (Karam and Fung, 1988). It approximates the induced current in a finite cylinder by assuming infinite length. Therefore, this method is valid for a needle-shaped scatterer with radius much smaller than the wavelength. Thereafter, Stiles and Sarabandi (1996) proposed a more general solution for long and thin dielectric cylinders of arbitrary cross section, but still limited to small cross sections. Nevertheless, it should be noted that the solutions of such approximate methods in general fail to satisfy the reciprocity theorem.

In a more general setting, a semi-analytical method named T-matrix approach, originally introduced by Waterman (Waterman, 1956) and is based on the extended boundary condition method (EBCM), is one of the most powerful and widely used tools for rigorously computing volume electromagnetic scattering based on Maxwell's equations and has been applied to particles of various shapes, such as spheroids, finite cylinders, Chebyshev particles, cubes, and clusters of spheres (Mishchenko and Travis, 1994; Roussel et al., 1996; Wielaard et al., 1997). In applying extended boundary condition to calculate the T-matrix that relates the exciting field and scattered field, the exciting field is assumed to be inside the inscribing sphere and the scattered field outside the circumscribing sphere, respectively. However, for particles with extreme geometries

represented by very large aspect ratios, regular EBCM is reported to suffer from convergence problems (Barber, 1954). Physically, this ill-conditioning procedure stems from the fact that, since the exciting field is assumed to be inside the inscribing sphere, for cases of extreme geometries, the exciting fields will not be accurate representative of surface currents. Nor will the scattered fields.

One approach for overcoming the problem of numerical instability in computing the T-matrix for spheroids with large aspect ratio is the so-called iterative extended boundary condition method (IEBCM) (Iskander et al., 1983). The main feature of this technique is to represent the internal field by several subregion spherical function expansions centered along the major axis of the prolate spheroid. The contiguous subregional expansions are linked to each other by being matched in the overlapping zones. The set of unknown expanded coefficients of internal field can be determined by using the point-matching method (PMM). It has been reported that in some spheroidal cases, the use of IEBCM instead of the regular EBCM allows to more than quadruple the maximum convergent size parameter. However, because the first step in this procedure is to approximate the highly lossy dielectric object with a perfectly conducting object of the same shape for its initial solution, it is restricted by the conductivities of the dielectric particles and the maximum convergent size parameter of EBCM for such perfectly conducting object. Moreover, as pointed in (Kahnert, 2003), PMM is less flexible in terms of applications to different particle shapes due to the fact that, the more the particle's geometry departs from that of a sphere, the more unsuitable the expansions of the fields in spherical vector wave functions. Thus, elongated particles require the use of specially adapted PMM implementations with longer computation time and higher computer-code complexity. Another similar technique using PMM to solve scattering from particles enclosed by smooth surfaces is the general multipole technique (GMT), which represents electromagnetic field vectors by multiple spherical expansions about several expansion origins which are located at appropriate positions in the interior region (Al-Rizzo and Tranquilla, 1995). The GMT has been successfully used for particles with smooth surfaces, such as hemispherically or spherically capped cylinders, yet there are issues when it is used in the scattering computations of finite cylinders with flat ends reported. Recently, null field method with discrete sources (NF-DS) is proposed to deal with the instability of conventional EBCM (Doicu and Wriedt, 1999; Eremina et al., 2004; Wriedt et al., 2008). Its numerical stability is achieved at the expense of considerable increase in computer complexity, and the resolution of this method can be affected by the localization of the sources.

A new iterative technique was proposed for electromagnetic scattering by finite dielectric cylinders with large aspect ratio (Yan et al., 2008). With the understanding that for such cylinders a direct application of the EBCM often leads to numerical instability, the procedure starts by dividing the cylinder into several identical subcylinder to reduce the aspect ratio for each part to which the EBCM can be applied. One of the technical issues that the iterative approach confronted is this: Since any two neighboring subcylinders are touching via the division interface, the conventional multiscatterer equation method is not directly applicable because it requires that the circumscribing spheres of the subcylinder exclude each other (Tsang et al., 1985). Rather, boundary conditions at the division interface need to be satisfied and carefully incorporated into the EBCM formalism. The subtlety lies in the fact that boundary conditions at the division interfaces are point-wise while the EBCM is in an integral form. For such concern, some intermediate variables that have specific meanings are introduced, where the boundary conditions are incorporated. Moreover, since these variables are expressed in terms of surface integrals, the drawbacks of PPM inherent in IEBCM or GMT are avoided. The intercoupling relations of multipole expansions for subcylinder are constructed with the help of translational addition theorems and can be solved iteratively. The impact of translational addition theorem on the convergence property of the resulting linear system is also carefully treated in the iterative procedure.

Summary

The Maxwell equations lay the foundation of all classical electromagnetic phenomena. They describe how time-varying electric fields give rise to magnetic fields and vice versa. At boundary of two different media the permittivity, permeability, and conductivity can change sharply. Boundary conditions are discontinuity equations to specify corresponding discontinuity in the field vectors. Scalar and vector potentials are useful auxiliary functions in analysis of an electromagnetic field, under Lorenz gauge condition the decoupled equations of the potentials identically satisfy the Maxwell equations. Green's function is an important tool in obtaining the solution for general source distributions. Poynting's theorem is a law of conservation of energy for the electromagnetic field. Plane waves are the simplest and fundamental electromagnetic waves. In remote sensing applications, electromagnetic wave scattering from randomly rough surfaces, from vegetation canopies, and from cylinders are important and ongoing research topics.

Bibliography

Ahmed, S., and Naqvi, Q. A., 2008. Electromagnetic scattering from a perfect electromagnetic conductor cylinder buried in a dielectric half-space. *Progress In Electromagnetics Research*, **78**, 25.

Al-Rizzo, H. M., and Tranquilla, J. M., 1995. Electromagnetic wave scattering by highly elongated and geometrically composite objects of large size parameters: the generalized multipole technique. *Applied Optics*, **34**, 3502.

Alvarez-Perez, J., 2001. An extension of the IEM/IEMM surface scattering model. *Waves in Random Media*, **11**, 307.
Barber, P. W., 1954. Resonance electromagnetic absorption by nonspherical dielectric objects. *IEEE Transactions on Microwave Theory and Techniques*, **25**, 373.
Chen, K. S., Wu, T., Tsang, L., Li, Q., Shi, J. C., and Fung, A. K., 2003. Emission of rough surfaces calculated by the integral equation method with comparison to three-dimensional moment method simulations. *IEEE Transactions on Geoscience and Remote Sensing*, **41**, 90.
Chiu, T., and Sarabandi, K., 1999. Electromagnetic scattering interaction between a dielectric cylinder and a slightly rough surface. *IEEE Transactions on Antennas and Propagation*, **47**, 902.
Chiu, T., and Sarabandi, K., 2000. Electromagnetic scattering from short branching vegetation. *IEEE Transactions on Geoscience and Remote Sensing*, **38**, 911.
Davidson, M. W. J., Le Toan, T., Mattia, F., Satalino, G., Manninen, T., and Borgeaud, M., 2000. On the characterization of agricultural soil roughness for radar remote sensing studies. *IEEE Transactions on Geoscience and Remote Sensing*, **38**, 630.
De Roo, R. D., Du, Y., Ulaby, F. T., and Dobson, M. C. A., 2001. Semi-empirical backscattering model at L-band and C-band for a soybean canopy with soil moisture inversion. *IEEE Transactions on Geoscience and Remote Sensing*, **39**, 864.
Doicu, A., and Wriedt, T., 1999. Calculation of the T-matrix in the null field method with discrete sources. *Journal of the Optical Society of America*, **16**, 2539.
Du, Y., 2008. A new bistatic model for electromagnetic scattering from randomly rough surfaces. *Waves in Random and Complex Media*, **18**, 109.
Du, Y., Xu, T., Luo, Y. L., and Kong, J. A., 2005. A statistical Kirchhoff model for EM scattering from Gaussian rough surface. *PIERS Proceedings*, **1**(2), 187–191.
Du, Y., Kong, J. A., Yan, W. Z., Wang, Z. Y., and Peng, L., 2007. A statistical integral equation model for shadow-corrected EM scattering from a Gaussian rough surface. *IEEE Transactions on Antennas and Propagation*, **55**, 1843.
Du, Y., Luo, Y. L., Yan, W. Z., and Kong, J. A., 2008. An electromagnetic scattering model for soybean canopy. *Progress In Electromagnetics Research*, **79**, 209.
Elfouhaily, T. M., and Guerin, C.-A., 2004. A critical survey of approximate scattering wave theories from random rough surfaces. *Waves Random Media*, **14**, R1.
Eremina, E., Eremin, Y., and Wriedt, T., 2004. Extension of the discrete sources method to light scattering by highly elongated finite cylinders. *Journal of Modern Optics*, **51**, 423.
Fung, A. K., 1994. *Microwave Scattering and Emission Models and Their Applications*. Norwood, MA: Artech House.
Harrington, R. F., 1961. *Time-Harmonic Electromagnetic Fields*. New York: McGraw-Hill.
Henin, B. H., Elsherbeni, A. Z., and Al Sharkawy, M. H., 2007. Oblique incidence plane wave scattering from an array of circular dielectric cylinders. *Progress in Electromagnetics Research*, **68**, 261.
Hsieh, C. Y., and Fung, A. K., 1999. Application of an extended IEM to multiple surface scattering and backscatter enhancement. *Journal of Electromagnetic Waves and Applications*, **13**, 121.
Iskander, M., Lakhtakia, A., and Durney, C., 1983. A new procedure for improving the solution stability and extending the frequency range of the EBCM. *IEEE Transactions on Antennas and Propagation*, **31**, 317.
Jackson, J. D., 1998. *Classical Electrodynamics*, 3rd edn. New York: Wiley.
Kahnert, F. M., 2003. Numerical methods in electromagnetic scattering theory. *Journal of Quantitative Spectroscopy and Radiative Transfer*, **79–80**, 755.
Karam, M. A., and Fung, A. K., 1988. Electromagnetic wave scattering from some vegetation samples. *IEEE Transactions on Geoscience and Remote Sensing*, **26**, 799.
Kong, J. A., 2005. *Electromagnetic Wave Theory*. Cambridge: EMW Publishing.
Lin, Y. C., and Sarabandi, K., 1997. A Monte Carlo coherent scattering model for forest canopies using fractal generated trees. *IEEE Transactions on Geoscience and Remote Sensing*, **37**, 36.
Mishchenko, M. I., and Travis, L. D., 1994. T-matrix computations of light scattering by large spheroidal particles. *Optics Communications*, **109**, 16.
Notarnicola, C., and Posa, F., 2007. Inferring vegetation water content from C- and L-band SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, **45**, 3165.
Rodriguez, E., 1991. Beyond the Kirchhoff approximation. *Radio Science*, **26**, 121.
Roussel, H., Chew, W. C., Jouvie, F., and Tabbara, W., 1996. Electromagnetic scattering from dielectric and magnetic gratings of fibers: a T-matrix solution. *Journal of Electromagnetic Waves and Applications*, **10**, 109.
Schiffer, R., and Thielheim, K. O., 1979. Light scattering by dielectric needles and disks. *Journal of Applied Physics*, **50**, 2476.
Singh, D., and Kathpalia, A., 2007. An efficient modeling with GA approach to retrieve soil texture, moisture and roughness from ERS-2 SAR data. *Progress In Electromagnetics Research, PIER*, **77**, 121.
Stiles, J. M., and Sarabandi, K., 1996. A scattering model for thin dielectric cylinders of arbitrary crosssection and electrical length. *IEEE Transactions on Antennas and Propagation*, **44**, 260.
Storvold, R., Malnes, E., Larsen, Y., Hogda, K. A., Hamran, S. E., Muller, K., and Langley, K. A., 2006. SAR remote sensing of snow parameters in Norwegian areas – Current status and future perspective. *Journal of Electromagnetic Waves and Applications*, **20**, 1751.
Tsang, L., Kong, J. A., and Shin, R. T., 1985. *Theory of Microwave Remote Sensing*. New York: Wiley.
Ulaby, F. T., Moore, R. K., and Fung, A. K., 1982. *Microwave Remote Sensing: Active and Passive*. Bedham, MA: Artech House.
Wang, L. F., Kong, J. A., Ding, K. H., Le Toan, T., Ribbes, F., and Floury, N., 2005. Electromagnetic scattering model for rice canopy based on Monte Carlo simulation. *Progress In Electromagnetics Research*, **52**, 153.
Waterman, P. C., 1956. Matrix formulation of electromagnetic scattering. *Proceedings of the IEEE*, **53**, 805.
Wielaard, D. J., Mishchenko, M. I., Macke, A., and Carlson, B. E., 1997. Improved *T*-matrix computations for large, nonabsorbing and weakly absorbing nonspherical particles and comparison with geometrical optics approximation. *Applied Optics*, **36**, 4305.
Wriedt, T., Schuh, R., and Doicu, A., 2008. Scattering by aggregated fibres using a multiple scattering T-matrix approach. *Particle and Particle Systems Characterization*, **25**, 74.
Yan, W. Z., Du, Y., Liu, D. W., and Wu, B. I., 2008. EM scattering from a long dielectric circular cylinder. *Progress in Electromagnetics Research, PIER*, **85**, 39.
Yeh, C., Woo, R., Ishimaru, A., and Armstrong, J., 1982. Scattering by single ice needles and plates at 30 GHz. *Radio Science*, **17**, 1503.
Yueh, S. H., Kong, J. A., Jao, J. K., Shin, R. T., and Le Toan, T., 1992. Branching model for vegetation. *IEEE Transactions on Geoscience and Remote Sensing*, **30**, 390.
Zhang, G., Tsang, L., and Chen, Z., 1995. Collective scattering effects of trees generated by stochastic Lindenmayer systems. *Microwave and Optical Technology Letters*, **11**, 107.

Cross-references

Media, Electromagnetic Characteristics

EMERGING APPLICATIONS

William Gail
Global Weather Corporation, Boulder, CO, USA

Synonyms

Emerging societal benefits

Definition

Application. An application of remote sensing is a practical use of information obtained with the technique of remote sensing. Applications are distinct from scientific investigations or engineering development in that their purpose is to consume and apply remote sensing information rather than create and understand it.

Emerging. An emerging application is one in which a new means for using remote sensing information is developed to address a new problem, exploit a new opportunity, or perform an existing function in a new way.

Introduction

Beyond its tremendous value for understanding the Earth, remote sensing also provides a myriad of more practical benefits to society (Group on Earth Observations, 2005). The National Research Council, in its landmark 2007 Decadal Survey of Earth Science and Applications (National Research Council, 2007), underscored the importance of this balance by noting that "attention to securing practical benefits for humankind plays an equal role with the quest to acquire new knowledge about the Earth system." This balance is critical; remote sensing's practical benefits (commonly referred to as *applications*) often arise as a direct result of the cutting-edge knowledge developed by pursuing remote sensing science.

Society's growing need for remote sensing applications is clear. The transformation from an industrial economy to an information economy is driven by knowledge; remote sensing applications are a key element of this progress. Solutions for many of today's critical societal needs, from health to natural disasters to the environment, require sophisticated knowledge about the Earth and its workings. The defining question addressed in this entry is simple: How will remote sensing be used by society – beyond the underlying science – in coming decades?

Classes of applications

To anticipate remote sensing applications that are likely to emerge, it is first helpful to understand today's applications, how they are used, and what motivates their development. While these applications can be categorized in a number of ways, an effective approach is to separate them by public sector, consumer, and business uses. The boundaries between these categories may be fuzzy, but the distinctions between them are meaningful.

The *public* sector, with its obligation for ensuring the security and prosperity of the population it serves, applies remote sensing for a wide variety of purposes. The most prevalent use of remote sensing has long been surveillance for intelligence and military use. In the civilian arena, categories of applications are often described as "societal benefits." The most definitive list of such benefits comes from the Group on Earth Observations (GEO) and includes (a) disasters, (b) health, (c) energy, (d) climate, (e) agriculture, (f) ecosystems, (g) biodiversity, (h) water, and (i) weather. Various international and national governmental bodies (such as the United Nations Environment Programme (UNEP) and Europe's Global Monitoring for Environment and Security (GMES)) exist to develop and use applications, but much of the applications work occurs at the regional and local governmental levels. Typical applications functions within governmental bodies include decision systems, assessments, planning, and monitoring. As with most governmental activities, new public sector applications tend to emerge through deliberate planning processes.

The *consumer* market is a growing user of remotely sensed information. The rapid adoption of consumer mapping in the last several years, both online and through personal navigation devices, provides the most tangible evidence. Remote sensing has become increasingly embedded in consumer applications, often in ways that are not obvious. Remote sensing information is now used in consumer video games, as a framework for sharing images and videos, to plan both local trips and vacation travel, to communicate news events, for exploring real estate, and much more. New applications are largely motivated by the growth of existing markets or the anticipation of new markets. In many cases, remote sensing information creates value not by being an end-use product or service itself but through enabling the purchase of other products and services (online store-locator mapping being one example).

The *business* sector employs remote sensing in two basic ways: as elements of the products and services they sell and to improve the efficiency of their business operations. Imagery from both satellites and aircraft is used to help manage operations – through functions such as geographic information systems (GIS), business intelligence, fleet management, and situational monitoring.

Context of historic trends

For much of its early history, remote sensing was dominated by the surveillance applications of the intelligence community, with an emphasis on high-resolution imaging. This information was often tightly controlled. Starting in the 1970s, the use of remote sensing for regional and global civilian applications expanded as datasets such as those supplied by Landsat and geostationary weather satellites became widely available. In the 1990s, the paradigm of Earth system science led to more sophisticated applications that began to integrate remote sensing observations from multiple sources. Finally, the rapid spread of personal computing and Internet access during the early

2000s made it far easier for businesses and consumers to access remote sensing information, triggering a wave of innovative applications. Overlaid on these four historic periods were the many evolving demands for applications, from the exploding needs of emerging economies to the globalization of commerce.

These historic trends provide a context for the ways in which applications of remote sensing have evolved. Over the last half century, applications have transformed from being a specialists' tool to a public utility, from the activities of individuals to the work of groups, from government-driven to business- and consumer-oriented, from slowly produced to rapidly distributed, from single-discipline to multidisciplinary, and from text-based to highly visual.

Current and future trends

The applications that will arise over the next several decades can be expected to reflect the classic balance between trends in user pull and technology push.

From the perspective of user pull, three applications areas are likely to be particularly important. Environmental security encompasses a number of disciplines for which remote sensing is quite useful. Environmental treaties have proliferated over the last several decades; remote sensing is critical to compliance monitoring and assessment. The growing importance of international climate agreements will only accelerate this need. Many nations are also beginning to recognize the strategic importance of the environment to their national security and economic prosperity, from preservation of water resources to understanding global demographic shifts. One important trend has been to aggregate environmental applications as a set of "ecosystem services" (Millennium Ecosystem Assessment, 2005) – a new approach that leverages market forces by managing Earth's resources as assets (similar to how businesses manage their physical assets). A second important applications area is *business efficiency*. As global markets mature, businesses seek to maintain financial margins by improving operational efficiency. Information technology, and geospatial information in particular, provides particularly high leverage for these businesses. Remote sensing applications can improve asset management, operational responsiveness, energy efficiency, and more. Third, *consumer products and services* using remote sensing are likely to grow dramatically and could substantially alter the nature of remote sensing applications. Typical consumer uses range from navigation to online commerce to gaming. The consumer market is by far the largest component of the global economy, so growth in consumer uses of remote sensing can inject substantial resources into applications development. Many important societal trends cut across all three of these applications areas, from the increasing focus on health issues to the importance of energy.

From the perspective of technology push, a wide variety of trends will shape the future. Ongoing improvements to passive sensors, increased availability of active sensors (radars and LIDARs), and advances in sensor platforms will drive enhanced spatial and temporal resolution. Closely associated are the processing techniques used to transform data into knowledge. These range from algorithms of critical global importance, such as those designed to extract climate information from infrared radiance measurements, to seemingly trivial but high economic-value consumer image-processing algorithms such as red-eye removal. Essential to data processing advances are the ability to integrate (or fuse) data from multiple sources and automation of complex algorithms to facilitate low-cost operation across large datasets. Progress in many of these areas is also greatly enhanced by the development and use of standards, allowing easier interoperability. An often under-recognized area of sensor improvement is human-scale sensors, from professional devices for uses such as surveying to consumer sensors such as digital cameras.

Among the most important trends in supporting technology is the advent of publicly accessible satellite-based global positioning information from the US Global Positioning System (GPS), European Galileo, and Russian GLONASS satellite systems. Computing capability is also critical: Processing power continues to be characterized by Moore's Law, and similar progress metrics apply to information storage capacity and communications bandwidth. Of particular importance, the recent advances in computing and Internet access have seeded a demand for information that extends to the individual consumer. As with the development of television, this "democratization" of information is increasingly funded by advertising, ensuring that the ability to meet demand is not limited by access cost. Information access is also increasingly peer-to-peer rather than from centralized sources, with social networking and information-sharing software being the primary current drivers. The global trend toward universal availability of mobile communications devices, even within the developing world, greatly facilitates this progress. Finally, widespread access to digital photography (in part through inclusion of cameras on mobile telephones) is making people more familiar with the uses of imagery and creating a class of information that can augment and enrich the traditional government-owned sources of remote sensing information.

The rate at which new remote sensing applications emerge based on these trends will also be modulated by larger societal "megatrends." Crisis situations, from military conflict to long-term natural disasters, tend to be strong motivators for rapid advances in remote sensing. The global response to climate change and the international competition for energy may be such drivers; others that are more difficult to anticipate are likely to arise as well. Major changes or opportunities in global economic markets are a second important large trend. The market transformation created by the Internet and the shift of economic growth toward developing nations are important examples.

Characteristics of emerging applications

Emerging remote sensing applications are typically characterized by one or more of the following general attributes:

- *New uses for existing remote sensing capabilities.* Examples include migration of established remote sensing techniques to new markets and new public sector uses arising from newly created programs or agencies.
- *New sensors or techniques which enable applications that were previously technologically, economically, or functionally prohibitive.* In many cases, the applications needs of governments and businesses are well known, but they are difficult or too expensive to address until technology advances make them feasible.
- *New ways of analyzing, processing, or combining remote sensing information.* Useful applications often require extensive algorithmic processing of raw data. Progress is enabled by algorithm development and advances in computing power.
- *New science that enables new uses.* Scientific investigations using remote sensing routinely expand our knowledge of the Earth. This knowledge is used to enhance existing applications and to motivate development of entirely new applications.
- *Technological advances in related areas.* Many new applications emerge as a result of collateral technology advances in computing, communications, and satellite navigation.

These general attributes provide the framework for specific characteristics of emerging applications that can be anticipated over the coming decade and longer:

- *Real-time applications.* The need for remote sensing applications is often time critical, yet access to sensor information and the required processing can take hours or days. Satellite providers have recognized this and are emphasizing reduced revisit times for low-Earth-orbit constellations, radar sensors that can image at night, geostationary platforms to provide persistent imaging, and rapid deployment of aerial platforms (including drones). Advances in automated processing have also been critical to achieving rapid information dissemination.
- *Multidisciplinary integration.* Simple applications tend to focus on a specific need or single information source. The maturing of remote sensing has enabled development of applications that begin to connect multiple information sources. This can involve data from multiple scientific disciplines (e.g., weather and oceans) as well as diverse types of information ranging from science and business.
- *Shared and multiuser applications.* Until recently, most web-based software followed the server–client architecture, in which a centralized server publishes information to multiple users. Increasingly, web-based software employs a peer-to-peer architecture (commonly referred to as Web 2.0 and social networking) that facilitates direct sharing of information between individual users. Similarly, applications have begun to support multiple simultaneous users, enabling shared viewing of dynamic content and collaborative editing of information.
- *Transition from text to visualization.* Visualization is a highly effective means for communicating large amounts of information, particularly when spatial information is involved. Extensive visualization within applications requires systems with large bandwidth and considerable data storage, capabilities that have become widely available within the last few years.
- *Integration across spatial/temporal scales.* Applications have historically been constrained in spatial and temporal scale: limited to a global, regional, or local use and focused on either long-term or short-term. Applications that support linkages between scales are rapidly emerging. A prominent example is medium-range weather forecasting that integrates multidecadal global climatology, regional/seasonal patterns such as El Niño, and localized hourly weather forecasting.
- *Improved data archive and access.* As information volumes have grown, the ability to archive, find, and rapidly access information has become a significant challenge. Considerable resources are being applied to this problem through web-based search, including the areas of metadata, data tagging, indexing, and vertical search. With increasing consumer use, the business model for accessing remote sensing information has evolved; instead of end-users paying for access to the basic data, revenue increasingly comes from use of the applications that consume this data, including "free" applications monetized through advertising revenue.
- *Rapid, widespread communications.* Effective applications quite often require the rapid and widespread dissemination of information, decisions, and alerts. The ongoing revolution in mobile communications, both networks and handsets, has opened many new opportunities for remote sensing applications. Coupled with the Internet, mobile communications enable broad global information sharing and ready access at the point of use.
- *Merging of real and hypothetical information.* The transition from analog to digital information has made possible extensive processing and reprocessing of information. Within this context, it has become easy to combine or merge real and hypothetical data, and this has proven valuable for remote sensing applications. Simple examples are the 3D visual representation of a city modified to include a hypothetical development project and the visualization of hypothetical ecosystem changes brought on by climate change.

Anticipated applications

Based on these trends and characteristics, it is possible to anticipate several important classes of applications for the next decade and beyond. *Environmental services*, both

for government and business use, will almost certainly lead the way. These will encompass enhancements to existing applications, such as weather forecasting and resource management, as well as new applications to support emerging needs in the areas of climate, energy, air quality, and health. Society's increasing sensitivity to human and property losses from natural disasters will motivate improved applications addressing *disaster prevention and response*, with emphasis on communicating centralized information and decisions to the general population. The private sector will increasingly integrate remote sensing and other geospatial information into their operations to improve *business competitiveness and efficiency*. Both businesses and consumers will benefit from the growing availability of *location-based services* and augmented reality, allowing them to get information, make decisions, and perform functions specifically tailored to where they are or where they are going. Associated with this will be significant advances in the capabilities of *personalized mapping*, the ultimate tool for getting remote sensing out to the individual, including both Internet and mobile uses. Finally, *gaming and entertainment* cannot be underestimated due to the considerable financial resources available within this market; innovative uses of remotely sensed information will be developed to enhance the realism of applications in these areas.

It is also possible to anticipate some of the more innovative applications that may emerge in coming decades. High on this list are applications that involve *community-based mapping and remote sensing*. The ability of government and businesses to monitor the Earth at all space and time scales is currently highly constrained. Community-based data gathering (with sensors such as mobile phone cameras), and community-based applications, will be increasingly important. Various technological advances, from the proliferation of mobile phones to the spread of social networking software, make this possible. Linkages between these community-based networks and more formal government/business networks will leverage the capabilities of each. As the world becomes increasingly monitored across all space and time scales, this knowledge will further our ability to create replicas of the real world – 3D *virtual worlds* in which many of the real world's functions can be mirrored. These worlds – built and updated with remote sensing information – provide a virtual context in which applications can be visualized and applied without the constraints encountered in the real world. Perhaps most exciting, they create a platform for "what if" applications, allowing us to implement functionality ("experience a tsunami") and perform actions ("assess the societal consequences of a global crop blight") that cannot be accomplished within the real world for safety, ethics, or other considerations.

Summary

Applications of remote sensing have proven, over the last several decades, to have tremendous benefit to society – increasing our prosperity and making us more secure. The need for such applications is only accelerating as we transform from an industrial society to one that is information-driven. Advances across all areas of remote sensing, as well as the enabling computing and information technologies, ensure that emerging applications of remote sensing will be increasingly important to society's progress.

Bibliography

Group on Earth Observations, 2005. *Global Earth Observations System of Systems GEOSS: 10-Year Implementation Plan Reference Document*. Noordwijk: ESA Publications Division.

Millennium Ecosystem Assessment, 2005. *Living Beyond our Means: Natural Assets and Human Well-Being*. Washington, DC: Island Press.

National Research Council, 2007. *Earth Science and Applications from Space: National Imperatives for the Next Decade and Beyond*. Washington, DC: The National Academies Press.

Cross-references

Cost Benefit Assessment
Emerging Technologies
Environmental Treaties
Policies and Economics

EMERGING TECHNOLOGIES

Jason Hyon
Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

Synonyms

Active sensing; Data Management; Instrument; Passive sensing; Technology trends

Definition

Technologies to enable measuring new science requirements.

Emerging technologies

The objective of the *emerging* technology area is to develop highly capable instruments that are a key element in contributing to an ability to successfully develop and operate high-performance future remote sensing missions. The current state of the art remote sensing sciences tend to demand higher special and spectral resolution from next generation remote sensing capabilities. With a growing emphasis on cost-effective approaches due to high spacecraft and launch vehicle costs, multiband and low mass sensors are of key interests in this area.

There are a number of high payoff technologies within this area that include active remote sensing, massive data processing techniques, and technology push. The primary technology challenges for active remote sensing is

development of systems in ever-broadening envelopes of transmission; large aperture, light-weight, high-efficiency, and high-power transmitters; and reliable systems. These technologies will enable RADAR and LIDAR for a consistent coverage of the globe for CO_2, biomass, cryosphere, water cycle, and earthquake (Richard et al., 2007).

There is a robust and mature capability that is in place for Earth science enterprise to understand surface, ocean, and atmospheric characterizations of Earth. There have been many high resolution instruments to measure various parameters to study Earth such as RADAR, LIDAR, hyperspectral imager, and spectrometer instruments. With use of distributed data architecture, Earth science enterprise distributes data based on middleware architecture and provides a framework for data mining, modeling, simulation, and visualization capability. With significant increase in data volumes, traditional methodology for data processing will not be able to meet challenges of high-throughput modeling and data extraction. These developments should be incorporated to provide a uniform interface to the operational team. Handling large volume of data and information generation to decision support should take advantage of advancement from GIS and industry de facto standards.

In addition to the high-capability technology areas that are discussed, it is important to note the technologies that promise the future. These are the innovative technologies that have the potential to enable critical new science or exploration capabilities. These are promising because they do one or more of the following: stretch the state of the art in some respect, have the potential for high performance, provide a novel operational concept, and attempt to perform a measurement in way never before envisioned. For technology push areas, the technologies required for these instruments focus primarily on detector technology, whether in the detector sensitivity, size of detector arrays, or detection stability and calibration. For some applications, large apertures are also a challenge. In the areas of in situ sensing, challenges associated with in situ sensors focus on development of sensor webs that produce science-lab quality measurements in a field instrument. Furthermore, high-capability instruments tend to produce a lot of data volume and would require high downlink capability; optical communication will play an important role in the future.

Summary

Future science requirement demands higher resolution and temporal coverage. Active sensing techniques and large focal plane array detection will enable those objectives along with communication and information technology infrastructure.

Acknowledgment

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

Bibliography

Richard, A., Berrien, M., II, et al., 2007. *Earth Science and Applications from Space: National Imperatives for the Next Decade and Beyond*. Washington, DC: The National Academies Press.

Cross-references

Data Access
Data Archival and Distribution
Emerging Technologies, LIDAR
Emerging Technologies, Radar
Emerging Technologies, Radiometer
Emerging Technologies, Sensor Web
Microwave Horn Antennas
Reflector Antennas
Thermal Radiation Sensors (Emitted)

EMERGING TECHNOLOGIES, FREE-SPACE OPTICAL COMMUNICATIONS

Hamid Hemmati
Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

Introduction

Future pursuit of the vision for robotic and human space exploration would utilize instruments with ever-increasing capability and require orders of magnitude increase in data return rates from planetary distances. Without resorting to high-power transmitters with hundreds of watts and antenna diameters on the order of 5 m, deep-space science and exploration will rapidly encounter a bandwidth ceiling with radio frequencies of S, X, and Ka bands. In addition, communicating at a given data rate with an outer planet such as Neptune or Pluto is 1,000 times more difficult (when considering link margin degradation) than communicating at nominal Mars distance. Similarly, a link from Mars is approximately 10 billion times more challenging than communicating from geosynchronous orbit (GEO) to the ground (Hemmati, 2006). On space platforms, mass and electrical power are expensive commodities, and thermal gradients and ionizing radiation pose challenges as well (Hemmati, 2009).

Laser communications (lasercom or optical communications) through free-space, utilizing a narrow laser beam transmitted from planetary distances, hold great promise for delivering the increased data rates while imposing reasonable mass and power burdens on a host spacecraft. This technology has the potential to enable deep-space communication between Earth and planets at the farthest reaches of our solar system and beyond. In the coming decade, with comparable mass and power of a radio-frequency communication system (including optics one tenth the diameter of the flight antenna aperture), lasercom systems are expected to deliver at least an order of magnitude higher

data rate than conventional radio-frequency systems. Lasercom could offer these benefits with no additional attitude-control burden on the spacecraft and any spectrum bandwidth allocation or restrictive frequency regulations.

The high-capacity communications offered by lasercom could enable scientists to communicate information from other planets using, for example, streaming HDTV video, synthetic aperture radar, multi-and hyperspectral imagers, precision navigation, and multichannel telecommunications. Additional possibilities are multifunctionality with space-based optical navigations, laser remote sensing, imaging, and altimetry (Hemmati and Lesh, 1998).

Recent highly successful in-space validations of lasercom technology (e.g., GOLD, GeoLITE, SILEX, and TerraSAR-X) have led to the development of operational systems (Wilson et al., 1997; Caplan, 2007; Tolker et al., 2002; Smutny et al., 2009). Although lasercom has been fairly well proven in the Earth's orbit, it will need to undergo a number of successful deep-space demonstrations – with an order of magnitude higher data rate and at a lower cost per bit than conventional systems – before it can be operationally implemented in space. To date, there have been no laser telecommunications with a planetary spacecraft, although there have been demonstrations of aiming a ground-based laser at spacecraft into deep space, and vice versa (Wilson et al., 1993; Smith et al., 2006).

An introduction to significant system design and system engineering aspects of the flight and ground portion of a planetary laser communications will follow.

System design and engineering

Telecommunication at planetary distances involves link ranges that can range up to billions of kilometers. Due to the $1/r^2$ dependence for the space loss factor, a system engineer faces the challenge of designing a system that can compensate for an additional loss of 60 to greater than 100 dB over that of a near-earth satellite-to-ground link. A clear improvement area is to increase the aperture diameter at both ends of the link (Hemmati et al., 2007). However, an increased flight transceiver aperture exasperates the laser beam pointing challenge, since efficient implementation necessitates beam pointing to the order of one tenth of the telescope's beamwidth. Moreover, larger ground apertures collect larger amounts of background light, and even the highest quality telescopes are constrained by atmospheric turbulence-induced spreading of the signal (blur circle) at the telescope's focal plane. As discussed below, before a system can be designed, the link requirements and parameters have to be considered in its entirety.

Constraints and desired characteristics of an optical link

Although optical communication systems offer many potential advantages in deep data return capability, the current implementation also has a number of limitations that can affect the link availability and overall data return strategy. These constraints include limited weather-related availability from a given ground station and limited solar conjunction availability arising from Sun platform geometry. Also, until the spacecraft/mission design can provide sufficient confidence in autonomous systems, radio-frequency (RF) link coverage (albeit with reduced capability) will likely augment the optical link during the attitude-constrained mission phases and critical maneuver.

Desired characteristics of an optical link include (Hemmati, 2006):

- Support high data rates that are competitive to RF implementation
- Support bidirectional links and spacecraft navigation
- Provide high reliability
- Low mass, power consumption, and volume
- Radiation tolerant
- Not to place demands on the host platform
- Provide extended mission coverage
- Function at small Sun angles

Major design driving issues include:

- Stabilization and pointing of the narrow laser beam transmitted from the remote platform
- Minimization of scattered light, thermally induced telescope deformations, and avoiding telescope damage – considering that transceivers at each end of a link must typically look near the Sun in order not to avoid operational downtime
- Rejection of as much of the collected background light as is feasible (e.g., via optical filtering) prior to data detection

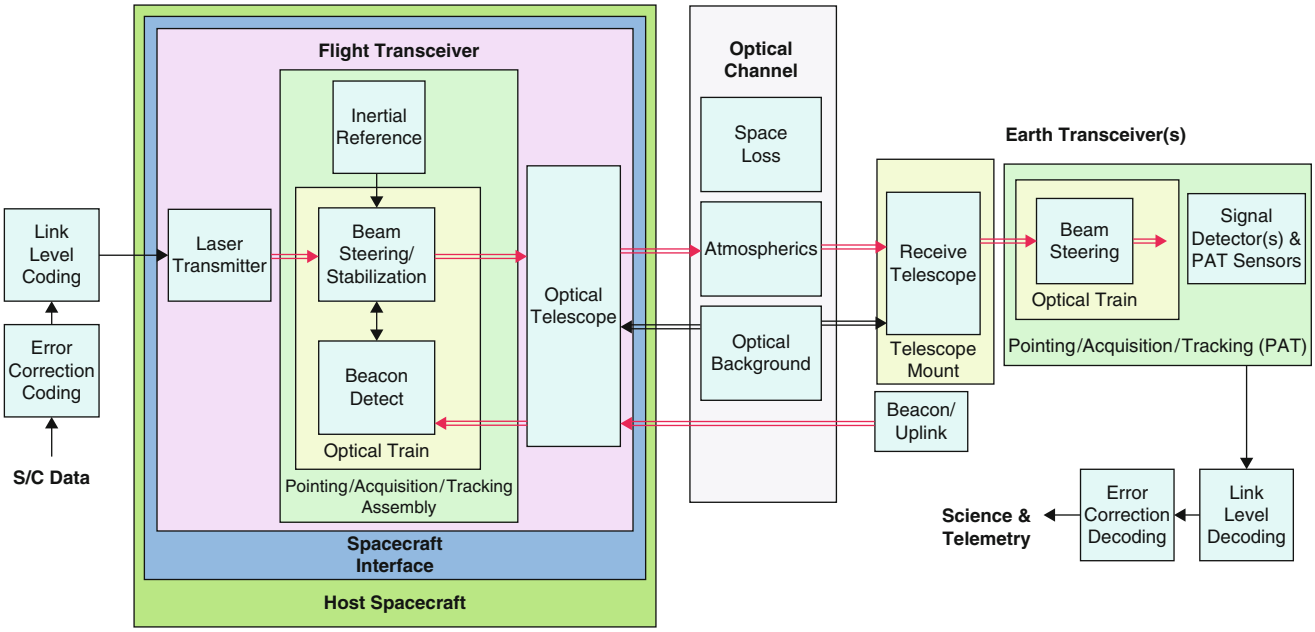
End-to-end link block diagram

A top-level, end-to-end signal flow for an optical communications signal flow is depicted in Figure 1. Mismatch among various blocks or performance degradation in any subsystem block, for example, due to device aging or radiation damage, will result in reduced efficiency of data transmission. The host platform random vibrations, the space platform/transceiver interface, and the space environment (radiation and vacuum) can all influence the flight transceiver's behavior.

Communications link design

The link design control table (DCT) is a listing of link influencing design parameters, such as aperture size, output power, and other requirements placed on the transmit and receive transceivers, and the resulting estimated system at a specific point in time during the mission (Hemmati, 2006). DCT analysis, with the primary purpose of estimating the required signal power to be delivered to the opposite terminal and available margin to maintain the desired link performance for a given worst-case condition, is a critical consideration in the design of a laser communications system.

Factors driving the link analysis are (1) fixed requirements/specifications (e.g., link distance, data rate, and bit error rate) provided by the user, (2) those fixed by laws of physics (i.e., space loss), (3) variables of the hardware



Emerging Technologies, Free-Space Optical Communications, Figure 1 Telecommunications signal flowing from a spacecraft probe to Earth.

design (e.g., aperture sizes and laser powers at opposite ends of the link), and (4) variable hardware qualities driven by implementation approach (e.g., pointing losses and optical losses). By adding the gain factors together (e.g., laser power, photodetector sensitivity, and aperture's area – in units of dBm or dB) and subtracting from the loss factors (e.g., the space loss, mispointing, and imperfection of optics), we are left with the link margin that needs to be positive in order to establish a link.

The mean received signal power at the receiver is related to the transmit power and various link losses, and the transmit and receive antenna gains can be estimated from the following equation:

$$P_R = P_T \eta_T G_T L_{TP} L_{atm} L_s G_R \eta_R$$

Where

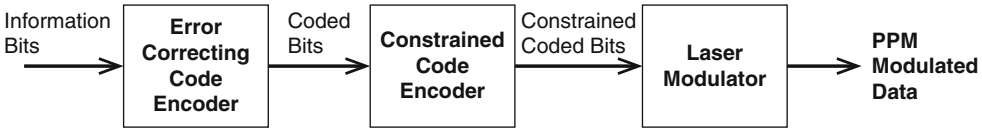
P_R	Signal power incident on the data detector
P_T	Output power of the laser transmitter
η_T	Transmit optics efficiency, accounting for relay optics transmission, wave front quality errors, and beam truncation
G_T	On-axis transmitter gain
L_{TP}	Transmitter pointing loss
L_s	Space loss
L_{atm}	Atmospheric loss comprised of attenuation and turbulence (beam spreading beam jitter losses)
G_R	Receiver gain
η_R	Receiving efficiency accounting for optical transmission and losses associated with mispointing and detector overfilling or truncation (Hemmati, 2006)

Technologies for data rate generation and detection

In this section, trades on data modulation and reception techniques, coding, laser transmitters, single-photon detectors, decoding, demodulation, and data-receive electronics are discussed.

What detection scheme we use is intertwined with what data transmission format we employ, and vice versa. Both direct detection and coherent heterodyne reception are options for data reception. A direct-detection receiver detects the intensity of the received signal and (without background interference) can achieve performance exceeding that of heterodyne detection. Implementation of narrow field-of-view optical receivers and narrow band-pass filters in the receiver chain becomes essential to mitigate the background noise and improve receiver sensitivity. Coherent (heterodyne) receivers first interfere with the downlink optical signal via a local oscillator (laser) prior to photodetecting the combined signal. The signal beam mixing process is mode-selective spatially, and only the signal energy that is mode matched to the local oscillator beam is detected. Therefore, the coherent detection scheme is largely immune to background light noise. However, the coherent receiving process is significantly more complex, particularly for a downlink signal that has traversed the turbulent atmosphere.

Considering the photon-starved regime of operations at ranges typical of planetary distances, our overall goal is for the most efficient detection (measured by the



Emerging Technologies, Free-Space Optical Communications, Figure 2 Example of a coded transmit optical channel, modulated in PPM format.

unit of bits/photon). Coherent receivers utilizing either homodyne techniques or heterodyne techniques with coherent IF electronics have achieved as much as −3 dB bits/photon (for homodyne-detected) signals (Boroson, 2007). The receiver sensitivity for (typically) photon-starved regime planetary communications links is of critical importance. The required receiver sensitivity is on the order of a few photons/bit for expected telemetry downlink communications rates, ranging from 10 to 100 s of megabits/s.

Phase fluctuations (arrival angle variations) constitute the primary influence of turbulent atmosphere on a direct-detection receiver since, by utilizing large (multimeter) diameter ground-based telescopes, the scintillation effect is averaged out to a great extent. In direct detection, the atmospheric turbulence limits the phase coherence across the telescope's optical aperture to a distance of r_0, known as the Fried's atmospheric coherent length (Andrews and Phillips, 2005). The value of r_0 is highly variable, depending on the receiver site and the time of day, and can vary from a few centimeters (daytime – poor seeing conditions) to tens of centimeters (nighttime – excellent seeing conditions). For distances up to r_0, arrival angle/phase remains mostly correlated. Once the daytime adaptive optics technology is sufficiently advanced, a substantial portion of the turbulence effect on the lasercom signal can be mitigated.

Currently, coherent detection approaches require single-spatial-mode reception. However, propagation of the laser beam through the turbulent atmosphere results in distorted (non-flat) wave fronts, leading to an enlarged blurred circle (multimode beam) at the telescope's focal plane. In the absence of an often elaborate adaptive optics system for a large ground-based telescope, the multimode beam is unsuitable for both coherent mixing with local oscillators and for coupling to single-mode fibers. Implementation of adaptive optics during daytime and meeting the unique requirements imposed by lasercom on adaptive optics (for generation of very high Strehl ratios – approaching 1.0) require additional technology development (Wilson et al., 2003). In the absence of an atmosphere (e.g., cross-links between assets above the atmosphere), coherent detection would provide a highly competitive performance.

For a turbulence-constrained system, the simplest detector is one that provides what is known as "direct detection." The most efficient detector of this kind is sensitive to single photons and counts individual photons with high-detection efficiency and extremely low noise (discussed in further detail later) (Farr, 2009). For this detection scheme, the requirement for spatial coherence is relaxed to levels sufficient for focusing a received beam onto the active area of the photodetector, located at the focal point of the optical aperture, and collecting the received photons.

Modulation and coding

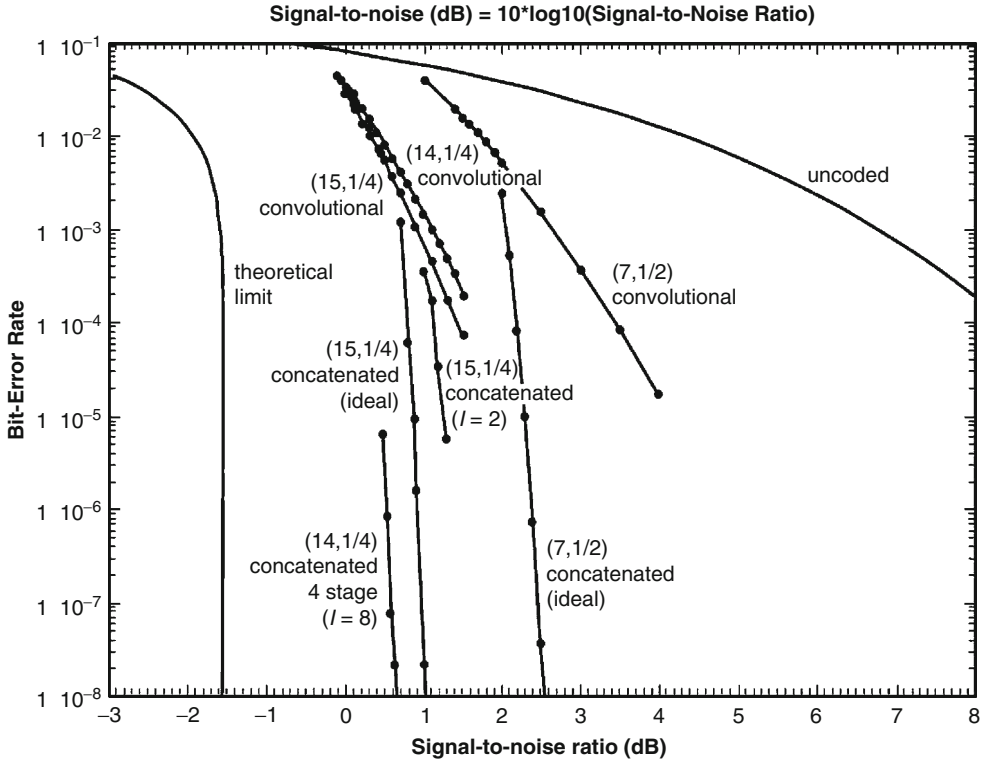
Figure 2 shows an example of signal flow through a coded free-space optical channel. Coding gains exceeding 5 dB can now be expected from the advanced high-efficiency codes, approaching to a fraction of a dB of the Shannon capacity limit (Hamkins and Moision, 2004). The resulting link efficiency gains have the direct effect of relaxing requirements on the transmitted laser power and telescope's aperture diameter, resulting in an overall simplification of the flight transceiver.

One may model a direct-detection free-space optical link as a Poisson point process, with system limitations on bandwidth, average power, and peak power (Moision et al., 2006). Under these constraints, a negligible capacity loss is encountered when confining the modulation to a binary, slotted format (Moision, 2999). The capacity of an optical channel, which represents the highest data rate that the channel can consistently support, depends on multiple factors including the modulation format, the medium through which the beam traverses, the optical preamplifier (if any), and the photodetector. A channel error-correction code (ECC) may be found with negligible output probability of error, when data rate is less than a given channel's capacity. However, the probability of error is enhanced when the data rate exceeds the channel capacity (Wyner, 1988).

Most efficient optical links result when the received signal's peak-to-average power ratio is high (Hamkins and Moision, 2004; Lipes, 1980). M-ary pulse-position modulation (PPM) is a modulation scheme that can create such a condition – resulting in near-capacity performance and negligible loss relative to alternatives. In the PPM format, $\log_2 M$ bits choose the location of a single-pulsed slot in an M-slot frame (Hamkins, 2008). Error-correction codes based on *serially concatenated codes* have been designed for the PPM channel, where the constituent codes are:

(a) Convolutional code serially concatenated with PPM (and preceded by a bit accumulator)
(b) Low-density parity-check (LDPC) code concatenated with PPM (Tan et al., 2008)

These codes – when used in conjunction with photon-counting detectors – achieve efficiencies better than 1 dB photons/bit. Implementation of PPM does result in

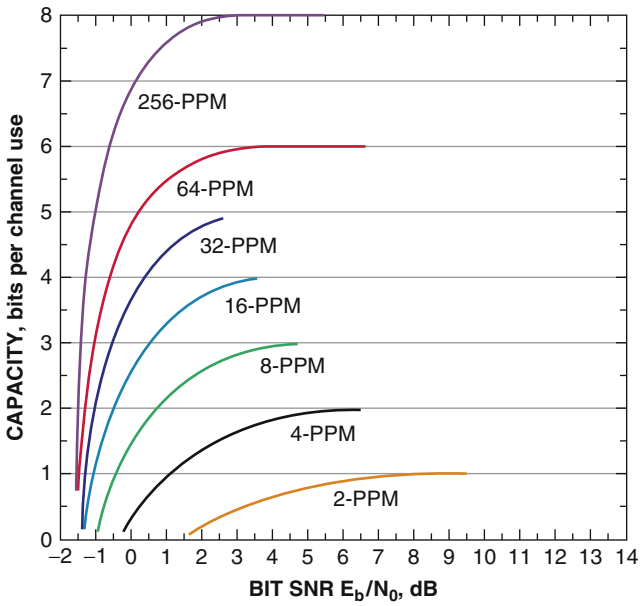


Emerging Technologies, Free-Space Optical Communications, Figure 3 Efficiency of state-of-the-art optical codes compared with channel capacity.

finite losses (due to modulator constraints) and more importantly necessitates high (kW level)-peak-power lasers.

Figure 3 shows the power efficiencies available at various data rates, using current technology of modulation, coding, and reception. On-off keying (OOK) modulation in conjunction with low-noise optically preamplified receivers is a common approach for fiber-optic telecommunications at rates exceeding 40 Gb/s. Utilizing the best available decoder chips, in conjunction with Shannon capacity-achieving codes (e.g., turbo codes) and soft decisions, efficiencies on the order of 6 dB photons per bit (at 1E-9 BER) have been achieved (Kiasaleh, 1998).

Figure 4 shows the efficiency of PPM at differing alphabet size, M. This figure illustrates how one may trade bandwidth for efficiency by using higher and higher M, and how a noisy environment significantly degrades the performance efficiency of a photon-counted PPM-modulated signal (every 3 dB of extra noise requires about 1 dB of extra signal) (Dolinar et al., 2000). The difference between a noiseless and just one noise count/slot can be as much as 10 dB of efficiency, diminishing the significant advantages provided by the photon-counting technique (Moision and Hamkins, 2003).



Emerging Technologies, Free-Space Optical Communications, Figure 4 Efficiency of different PPM orders indicating that the recently developed codes for optical detection are within a fraction of 1 dB of capacity limit. PPM capacity on the additive white Gaussian noise (AWGN) channels is compared with other performance curves.

Laser transmitter

Implementation of high-order PPM requires laser transmitters with a high peak-to-average power ratio.

Emerging Technologies, Free-Space Optical Communications, Table 1 Current status of key single-photon-sensitive detectors

Detector	Wavelength (nm)	Detection efficiency	Dark rate (Hz/mm^2)	Notes
Geiger-mode APD	400–850 900–1,600	40 % 55 %	1E6 2E4	High jitter. Reset time with each detection event
Intensified hybrid photodiodes	900–1,600	40 %	1E6	Photocathode degradation
Superconducting	400 to > 1,600	80 %	1E2	3°K temperature operation
Negative avalanche feedback	450–800 900–1,600	35 % 11 %	3E4 5E8	Intolerant of over bias. Low maturity

Master-oscillator power-amplifier (MOPA) sources based on Yb-doped fiber (operating at 1,060 nm) or Er/Yb-doped fiber (operating near 1,550 nm) satisfy this requirement and are compact enough for spacecraft use (Hemmati, 2009). Modulation extinction ratios on the order of 80 dB have been demonstrated with picosecond lasers (Braun et al., 1995). Also, mode-locked fiber lasers with 1 GHz pulse repetition rate and over 60 dB secondary mode suppression have been achieved (Deng et al., 2004).

Utilization of the laser in deep space also necessitates an overall high (input electrical to optical) efficiency of operation. A wall plug oscillator/amplifier efficiency of greater than 30 % is theoretically achievable (Hemmati et al., 2000). Currently, 1,060 nm lasers are proving to be more efficient than their 1,550 nm counter parts (Wysocki et al., 2006; Spellmeyer et al., 2005). MOPAs modulated at hundreds of megahertz with average output powers exceeding 10 W, peak powers approaching 1 kW, and sub-ns pulses are now available (Dupriez et al., 2006). The performance of current fiber amplifiers is limited by self-phase modulation (SPM) and to a lesser extent by stimulated Brillouin scattering (SBS) nonlinear effects when pulse width is <0.1 ns (Charplyvy and Tkach, 1993).

Laser designers also emphasize suppression of undesired background radiation noise, known as amplified spontaneous emission (ASE), which is generated along with the amplified signal (Bromage et al., 2003). ASE's spectrum spans that of the amplifier's gain profile and maintains the same polarization as the amplified signal. This can cause transmit/receive isolation challenges for the transceiver. Winzer et al. show that for near-Earth ranges of communications, the ASE noise levels may rival the background radiation received from the Sun (Winzer et al., 1999).

Uplink laser(s) for transmitting data to spacecraft and/or as a beacon (to assist with the acquisition and tracking of the earth station by the remote transceiver) typically requires orders of magnitude higher power and orders of magnitude lower pulse repetition frequency (Biswas et al., 2005). Unlike the downlink laser, sub-nanosecond pulse widths will not be required for the uplink laser.

Single-photon-sensitive photodetectors

Efficient detectors at each end of the optical communications link reduce the required laser power at the opposite end of the link. Top-level photodetector requirements, when employing PPM modulation in conjunction with photon-counting receivers, include high single-photon detection efficiency, low output pulse timing jitter, low reset time, large active area simultaneously with high bandwidth, and low dark count rate (Robinson et al., 2005).

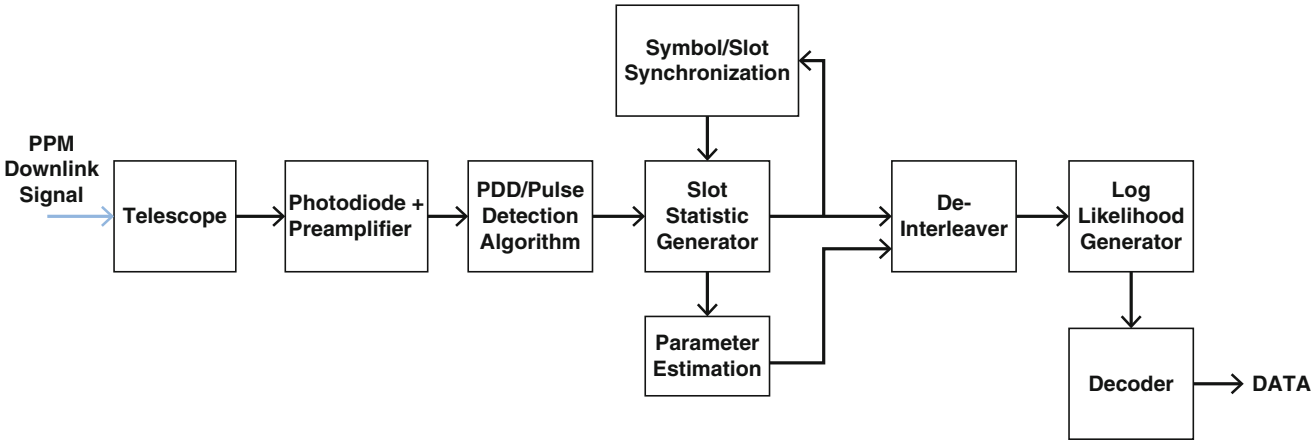
The advent of new generations of ultra-low-noise, high-gain, and high-bandwidth single-photon-sensitive detectors (also known as photon counters) has resulted in nearly 8 dB of link margin gain relative to APDs, enabling optical communication links with sensitivities approaching one photon/bit (Farr, 2009; Stern and Farr, 2007). When photon counters are utilized along with near optimal (channel capacity approaching) codes and high peak-to-average power lasers, nearly 20 dB gain may be realized relative to direct-detection systems operating with on-off keying (OOK) modulation (Boroson, 2007). Minute levels of detector noise also enable implementation of an array of optical receivers each equipped with a photon counter, followed by electrical signal summation (Quirk and Gin, 2006). Table 1 compares current merits of leading photon counters.

The *reset time*, an imperfection that is characteristic of most photon counters, is the period of time for which the detector is inactive upon each signal or noise detection event (Kerman et al., 2006). Depending on the detector type, the reset time period may span from a few nanoseconds to a few milliseconds. This drawback limits the achievable data rate and the total detectable photon flux (detection event/s).

For detection of very faint signals and to avoid detector shutdown due to dark count events, it is advantageous if the mean time between dark count (noise) signals is shorter than the reset time (Kachelmyer and Boroson, 2007). In a photon-starved regime (such as planetary links), daytime background light flux exceeds dark count rate of the detector. The minimum resolvable arrival time for a photodetection event is known as the output pulse *timing jitter* of a given detector (Moision, 2008a). This timing resolution restricts the number of bits accessible from each detection event, setting an upper limit on the achievable data rate (Moision and Farr, 2008).

Receive electronics

Various architectures have been envisioned for data-receive (synchronization, demodulator, decoder) electronics. In one *distributed and scalable* photon-counting



Emerging Technologies, Free-Space Optical Communications, Figure 5 Schematic block diagram of a generic PPM receiver.

decoder architecture, a receiver that independently synchronizes to the received signal follows each detector (Gin et al., 2007). Sets of pulse-position modulation (PPM) slot statistics generated by multiple receivers are combined in each slot accumulator and fed to a decoder for recovering clock and transmitted data bits. Multiple decoders will be required for high (tens of megabytes per second) data rate links. In distributed decoder architecture, each unit that follows the receiver contains a single PPM decoder. A decoder that is unavailable relays the undecided code words onto subsequent decoders. The receivers, slot accumulators, and decoder boards are connected using multi-gigabit serial deserializers. The distributed decoder and slot accumulator boards may share a common hardware platform but differently programmed field-programmable gate arrays. The scalability feature of this architecture has been utilized to develop receiver electronics for planetary-type links utilizing high-order PPM demodulation/decoding and capable of handling hundreds of megabytes per second of data rate throughput (Moision, 2008b). Figure 5 depicts a generic block diagram of a PPM receiver, including synchronization, demodulation, and decoding of the photodetected signal.

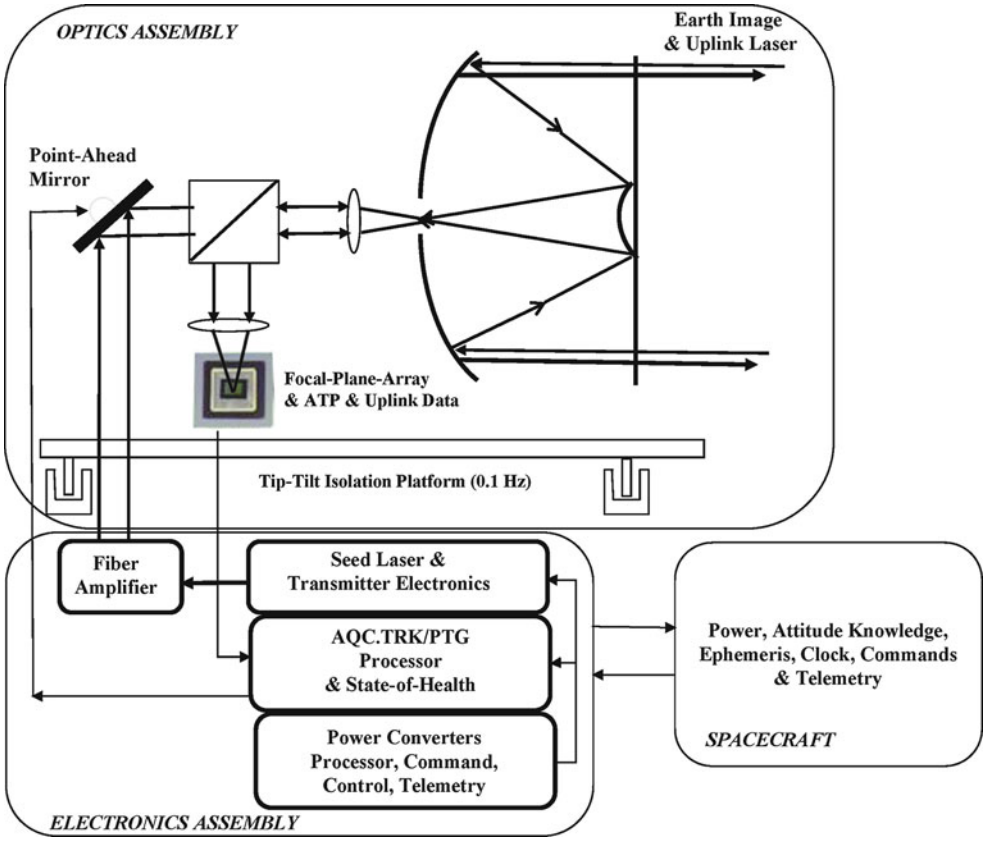
Optics for flight transceiver

Figure 6 shows a generic schematic diagram of the flight optomechanical assembly. This subsystem is comprised of three primary channels: laser beam transmit, data receive, and pointing, acquisition, and tracking. The laser transmitter assembly (typically located remotely to simplify thermal management) is connected to the optics assembly via a single-mode fiber. The data-receive channel may include an avalanche photodiode detector (APD) or a single-photon-sensitive detector. Synchronizer, demodulator, decoder, and amplifiers assisting with clock and data recovery comprise the remaining receiver electronics.

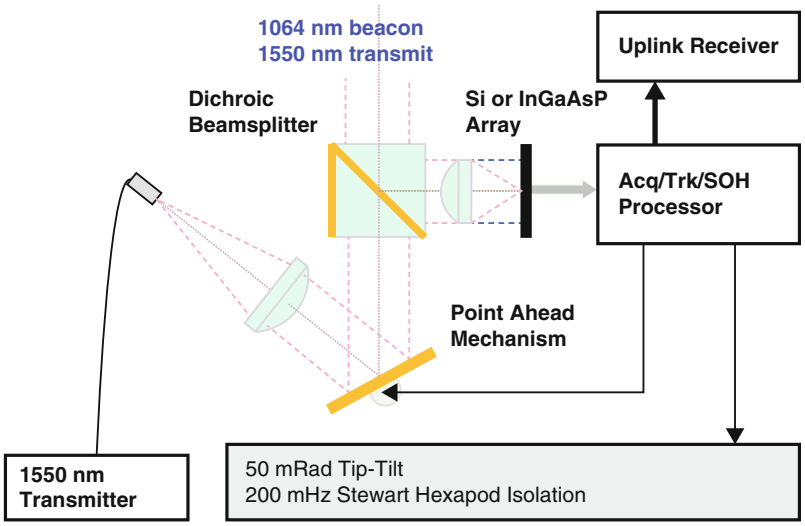
Channeling the uplink signal to the data receiver may be accomplished through free-space or efficient coupling with an optical fiber. Optical amplification of the received signal is also a possibility, as long as the signal level exceeds the amplifier's noise level. The pointing, acquisition, and tracking (PAT) assembly consists of a dichroic mirror to separate the transmit and receive beam, a two-axis fine-pointing mirror whose function is to remove the host platform vibrations, a two-axis point-ahead mirror to compensate for (point ahead or point behind) angle that arises from the cross velocity between the spacecraft and the target (e.g., Earth), and a focal plane array (FPA) consisting of as little as four pixels to hundreds of pixels. The fine-pointing (fast steering) mirror may require an update rate bandwidth of hundreds of hertz to several kilohertz, depending on the nature of the host platform vibration. It is also conceivable for the PAT's pixilated array to also serve as sensor for uplink data detection (Ortiz and Farr, 2009).

An alternative approach involves use of an isolation platform that combines passive and active isolators, as the interface between transceiver and the host platform (Chen et al., 2006). An advantage of this architecture is elimination of the fine-pointing mirror, since the base plate of the transceiver performs that function. Figure 7 illustrates such a configuration schematically.

The requirement for micro-radian level laser beam pointing necessitates optomechanical and thermomechanical integrity under the adverse conditions encountered in space. Maintaining optical isolation between the high-power transmit beam and the faint incoming beam, typically greater than 150 dB, is often a challenge. Even though transmit/receive signal isolation is more of a challenge for optical systems that combine both the transmitter and the (beacon and signal) receive aperture in one assembly, this configuration is preferred due to



Emerging Technologies, Free-Space Optical Communications, Figure 6 Block diagram example for a transceiver's optics assembly.



Emerging Technologies, Free-Space Optical Communications, Figure 7 Schematic of a laser communication transceiver utilizing isolation platform to simplify its architecture.

lower susceptibility for drift between channels (Chen and Lesh, 1994). A solar background, light-rejecting window at the telescope will minimize heat load from sunlight entering the telescope. Stringent requirements on cleanliness of the optics for flight use are an optical system class with cleanliness of 300 or better. This diminishes the scattered light that can find its way to any one of the sensors in the system (Spyak and Wolf, 1992). Redundancy of channels (if required for reliability reasons) may be implemented by polarization coupling for coupling two units and wavelength diversity for coupling additional units. Fiber coupling of all or some of the channels would simplify the optical assembly, while possibly encountering some additional coupling losses.

Pointing, acquisition, and tracking (PAT)

Fine *pointing, acquisition*, *and tracking* (PAT) is conceivably the most critical aspect of any lasercom link. This challenge stems from the narrow beamwidth of a communication beam, requiring the communication system to point its laser beam with sub-micro-radian (μrad) accuracy to a target tens to hundreds of millions of kilometers away. A combination of coarse and fine beam pointing is required to accomplish fine PAT. For planetary missions, a dedicated coarse-pointing mechanism is not required. The coarse-pointing (approximately, 0.1°), typically provided by the spacecraft to the planetary radio-frequency communications systems, is adequate for lasercom transceivers to acquire the Earth-based receiver within its acquisition field of view (Hemmati, 2006, 2009). Fine beam pointing and point ahead/behind is accomplished by way of internal mechanism(s) within the optomechanical assembly.

Prior to commencing a laser communications link between two remote transceivers, a number of knowledge-acquisition and handshaking steps need to be accomplished simultaneously:

- Knowing that the remote transceivers are within each other's field of view.
- Identifying each transceiver's position (e.g., attitude and locations). If range (and therefore roundtrip travel time) is short enough, the position knowledge may be communicated between transceivers.
- Utilizing the acquired position knowledge to derive the pointing vector to the remote transceiver.
- Deriving time-varying attitude variations and utilizing coarse-pointing and fine-pointing mechanisms to align the optomechanical assembly with the pointing vector.

Severely degraded link margin and system performance (signal fades at the receiver) will result from inaccurate laser beam pointing. The pointing loss allocation in the link control table is commonly divided into bias and jitter mispoint errors. Jitter represents high-frequency disturbances – normally root-sum squared to achieve an estimate of pointing accuracy – while bias represents the low bandwidth disturbances, normally added linearly. The 3 sigma total pointing accuracy is calculated by adding 3 × jitter to bias (Hemmati, 2007).

To account for the statistical pointing-induced fades (PIF) and mispoint angles, a pointing loss is typically allocated to the power link margin. For instance, the diffraction-limited angular beamwidth of a 1,000 nm laser traversed through a 30 cm-diameter telescope is approximately 3.3 μrad. In this example, a 2 dB pointing loss allocation with a 1 % PIF requires a total (3 sigma) pointing accuracy of 2 μrad. A promising new loss mitigation technique utilizes interleaver codes that spread the PIF over many code words of an error-correction code (Barron and Boroson, 2006).

Effective delivery of the spacecraft-emanated signal to a ground station requires the lasercom transceiver to track the receiving station such that the tracking jitter error is less than approximately 10 % of the transmit beamwidth (roughly 0.33 μrad in the above example) (Ortiz and Lee, 2006). For reference, from Mars, Earth angular extend is about 35 μrad. Therefore, in the above example, the spaceborne laser beam needs to be pointed to approximately 1/18th of the Earth's diameter, a pointing requirement orders of magnitude more precise than required for RF system operating over the same range.

Spacecraft platform's angular microvibrations (with frequencies in the 0.1 Hz to hundreds of hertz range) stemming from scheduled and unscheduled guidance and control events (e.g., thruster firings and momentum dumping) constitute additional complexities for the already challenging sub-micro-radian level pointing requirement. Passive isolators and inertially referenced stabilized platforms can effectively mitigate the high (greater than several hundred hertz)-frequency vibrations (McMickell et al., 2007). A reliable beacon signal (e.g., laser beam, Sun-illuminated Earth or celestial references), in conjunction with a dedicated fine beam pointing control subsystem, is required to reject lower frequencies effectively (Hemmati, 2006, 2009).

The total laser beam pointing error for an optical transceiver is primarily comprised of inaccuracies in boresight calibration, the errors in position determination, and residual tracking errors not compensated for by the pointing subsystem control loop. A residual tracking error results from inadequate compensation of the host platform vibrations by the pointing control loop. The higher the bandwidth of the control loop, the better the platform jitter compensation and therefore the lower the residual laser beam pointing error. Data from inertial navigation subsystem (located in close proximity with the transceiver) can reduce the required image centroiding update rates from several kilohertz to hundreds of hertz or lower (Ortiz et al., 2001).

As indicated earlier, mounting the lasercom transceiver on a disturbance-free platform (DFP) effectively isolates the transceiver from the host platform, circumventing the need for fast-tracking capability (Scozzafava et al., 2007). Such architecture will minimize the amount of

line-of-sight jitter coupled into the optics assembly and allow the beacon signal analyzing focal plane array to be integrated over a much longer period of time, as well as enable the use of faint reference beacons as pointing references. With the ability for long integration times, laser beam pointing becomes considerably less sensitive to scintillation-induced fades of the uplink beacon, thereby relaxing the spacecraft pointing requirements. Moreover, a DFP can also provide coarse and fine-pointing adjustments, eliminating the need for additional high-bandwidth line-of-sight control (e.g., fast-steering mirror) within the system (Sannibale et al., 2009).

Optical channel (atmospherics)

The adverse effects on a laser beam traversing through the atmosphere may be put into two categories: (1) *attenuation effects*, caused by *scattering* and *absorption, and* (2) *refractive index fluctuation effects*, including spatial and temporal *scintillation*, *angle-of-arrival fluctuations*, large-scale *beam steering*, and *beam spreading*. The net effect of atmospheric attenuation and turbulence on communication link is signal fades, beam spreading at the telescope focal plane, wave front tip/tilt, background light due to sky radiance, signal scintillation, and other undesirable phenomena (Andrews and Phillips, 2005).

Attenuation reduces available signal power at the receiver and is a function of weather conditions, zenith angle, and the receiver site. Scattering is highly variable and depends on particle size in the atmosphere. If δ is the particle size and λ is laser's wavelength, different scattering cross sections (σ) can result depending on the relative size of δ and $\boldsymbol{\lambda}$. A laser beam propagating through the atmosphere may experience Rayleigh scattering, which occurs when $\delta \ll \lambda$ (gases), Mie scattering when $\delta \sim \lambda$ (aerosols and fog), and geometric scattering when $\delta \gg \lambda$ (Andrews et al., 2001).

Scintillation is caused by refractive index variations in the atmosphere and results from spatial and temporal variations in light intensity caused by small-scale atmospheric turbulence. This effect creates multipath summation of signals at the receiver, resulting in phase differences for each beam with the net effect of variable arrival angles (Hemmati, 2006, 2009). Strong scintillation is difficult to characterize well, affects bit-error-rate performance, its influence increases with communications distance, and is more pronounced for horizontal links and when transmitting narrow beamwidth laser beams from the ground to space. Turbulence effects, scattering, and attenuation are all altitude dependent. At high altitudes, the effect is small for large telescopes, and turbulence along the path results in beam steering. Turbulence above the ground-based receiver results in wave front distortions in the form of a blur at the focal spot on the photodetector located behind the telescope. A significant portion of the atmospheric-induced optical link degradation can be allayed through proper site selection (Wojcik et al., 2005).

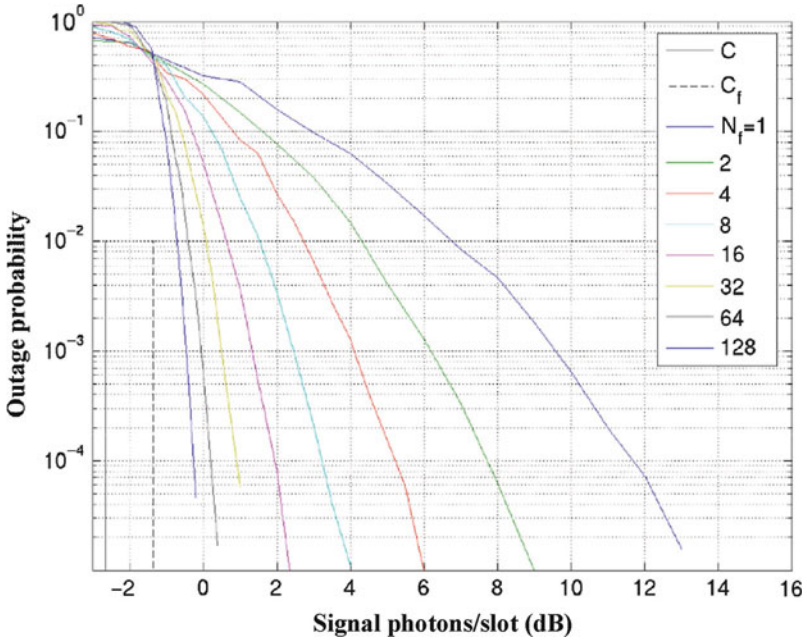
The atmospheric effects, singly or as an aggregate, can lead to dynamic attenuation of the signal strength. This phenomenon is referred to as signal fading. A fade is a transient condition that occurs when signal intensity at the receiver at any given instance is less than what is required to maintain the required link margin for a specified bit error rate (Andrews et al., 2001). Scintillation, spot dancing, and a time-varying speckle pattern at the detector are among major causes of fade events. The received signal fades away when a dark portion of the pattern overlaps the active region of the photodetector. Deep fades typically last 1–100 ms, depending on the strength of the turbulence effect for a particular link, and speed of the airflow across the beam driven by wind (Wojcik et al., 2005). In a high data rate optical link, significant amount of data could be lost during a fade period.

Fade mitigation techniques include use of large aperture diameters (greater than, approximately, 0.3 m), transmission of multiple beams that are mutually incoherent and physically separated (on the order of centimeters, use of long interleaver codes), implementation of adaptive optics technology, use of multimode lasers, and time or frequency diversity (Quaale et al., 2005; Moision and Biswas, 2007). As briefly explained below, each of these techniques suffers from implementation complexities or limitations.

For *downlink signal reception*, the *aperture averaging effect* (which becomes prominent when utilizing multimeter diameter ground telescopes) can mitigate scintillation effects on the beam and therefore decrease the fluctuations in total received power (Yuksel et al., 2005). Aperture averaging occurs when average size of the atmospheric turbulence cells is smaller than the diameter of the optical receiver. For *uplink beam transmission*, multibeam (4–16 beam) uplink can also mitigate some of the atmospheric turbulence effects on the beam (Wilson et al., 2007). For the multibeam fade mitigation to work, the laser beams have to be mutually incoherent and spatially separated enough to traverse uncorrelated atmospheric paths (Cameron et al., 2005).

Figure 8 shows that as much as 13 dB of the atmospheric fade depth can be mitigated via gains from implementation of long interleaver codes (Moision and Biswas, 2007).

Link availability is the fraction of time out of a year over which a laser communications link is operational. Link availability is highly dependent on the time of day and location of the telescope receiver site. Telescope site location options include space, platforms situated above clouds, and ground. Space-based receiving stations provide high link availability due to cloud-cover mitigation but suffer from limited aperture diameter and the lack of ability to upgrade and maintain the system after launch (Hurd et al., 2005; Biswas et al., 2006). High-altitude platforms avoid some of the drawbacks of space-based systems. But due to high winds that are typically encountered at those altitudes, these platforms are often challenged in providing lasercom-suitable



Emerging Technologies, Free-Space Optical Communications, Figure 8 Effect of interleaver in atmospheric fade mitigation is illustrated.

attitude-control levels. In the short term, ground-based receiver networks with atmospheric conditions limiting link availability provide an affordable solution. A recent study indicated that a network of five optical stations located globally will provide link availability of approximately 93 % (Link et al., 2004).

Optics: earth-based aperture

The Earth-based receiver collects the downlink signal photons and directs it to the photodetectors. Clearly, the larger the effective ground transceiver diameter, the greater the margin for the link. The excess link margin may be applied to shift hardware implementation burden away from the spacecraft (and onto the ground), to reduce flight system mass and power requirements, or to increase the amount of data delivered to Earth. The optical aperture may be a single telescope several meters in diameter or an array of distributed telescopes with meter-scale diameters. Each telescope may have a monolithic or a segmented primary mirror. Also, this transceiver may be located on the ground or at high altitudes above clouds, with clear advantages for each case.

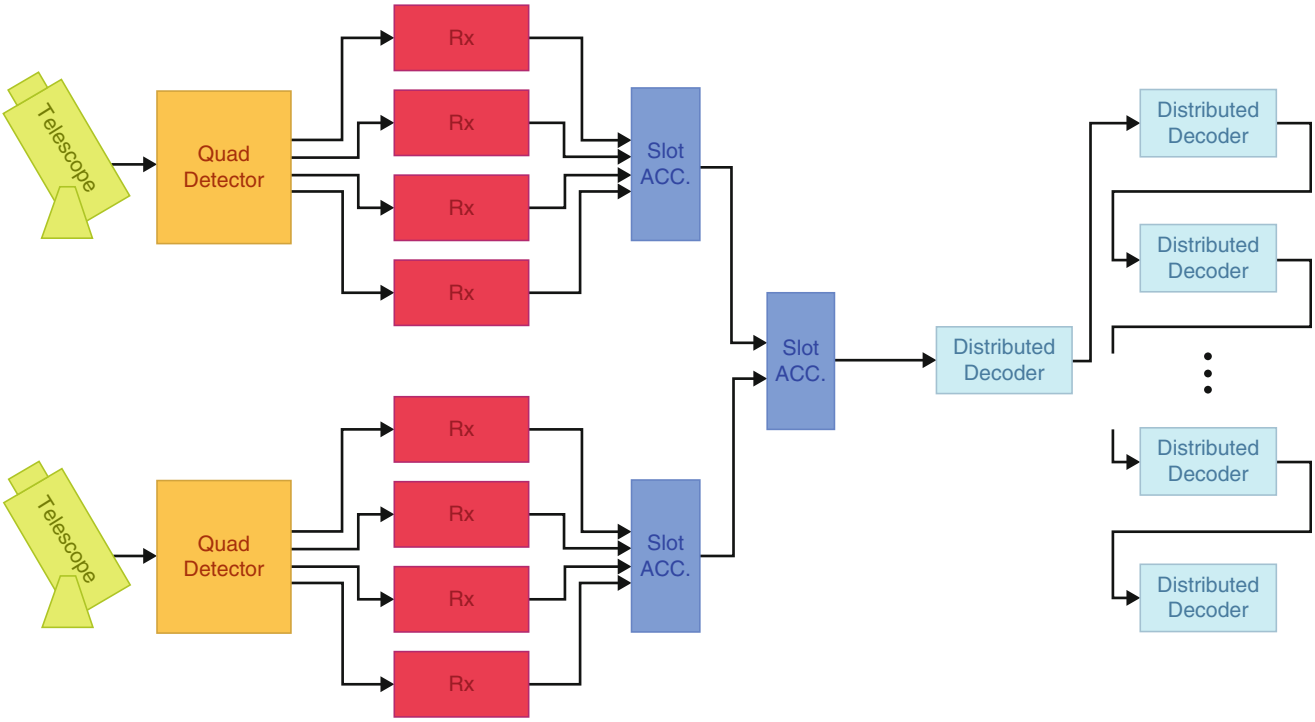
The recent advancements in solid-state, very-low-noise, high-detection-efficiency photon-counting detectors (that operate near room temperature) have given credence to optical receiver architectures that utilize meter-class telescope diameters, each equipped on at least a 2×2 array of photon counters. Features of an array of telescope receivers are redundancy, graceful degradation, affordability scaled by the number of elements in the array, and potential for cost savings of a network of stations. Figure 9 illustrates the schematic of a distributed and scalable ground receiver architecture.

The focal spot diameter of meter-scale diameter telescopes is set by the atmospheric turbulence-induced blurred conditions. The smallest achievable blur circle diameter for a downlink laser beam of wavelength λ that has traversed through an atmosphere with its seeing set by r_0 is set by an angle that is approximately $(r_0/\lambda)^2$ steradians (Roberts, 2005). Large-diameter membrane medium (10s of nm) band-pass optical filters (located at the entrance aperture of the telescope) and small-diameter narrow (sub-nm) band-pass filters, located just before the photodetectors(s), will combine to provide an essential level of background light rejection (Degnan, 1993).

Hybrid laser communications and precision laser ranging

An active laser-ranging transponder and laser communications transceiver have multiple-system-level requirements and subsystems in common. Steps necessary to operate a laser communications system in a transponder mode to infer millimeter-level range resolution are outlined below.

The interest in high-accuracy laser ranging to Mars is motivated by tests of relativistic gravitation, including the tests of the strong equivalence principle, Mars interior science (through exploitation of sensitivity to Mars precession, nutation, polar motion), and planetary science (through improvement of basic dynamical model parameters for the solar system). Centimeter-level resolution of Earth-Moon distance has been demonstrated via laser ranging to passive corner-cube retroreflectors deployed



Emerging Technologies, Free-Space Optical Communications, Figure 9 Scalable optical receiver followed by distributed decoder architecture.

on the lunar surface (Williams et al., 2009). Millimeter-level accuracy of lunar laser-ranging (LLR) data is expected in the near future (Turyshev and Williams, 2007; Murphy et al., 2007). However, due to the high space loss stemming from the $1/R^4$ energy transfer from passive corner cubes, passive laser ranging to planets is not possible at this time. Due to the benefit of $1/R^2$ energy transfer, active laser ranging to transponders located at the Moon or Mars will provide higher signal-to-noise ratios. The development of active laser-ranging techniques would extend the accuracies characteristic of passive laser tracking to interplanetary distances. Picosecond timing resolution time tagging of received and transmitted signal can now enable mm-class resolution ranging between the Earth and Mars.

A pulsed laser transponder system (based on time-of-flight measurements) is an efficient approach, requiring lower laser power than other available schemes and may be used in photon-counting mode without the need for phase referencing. A simple laser transponder might employ an "echo" technique, wherein the laser pulses from the remote unit are produced in response to detections of received pulses (Degnan, 1993). A more robust solution is paired one-way ranging, an asynchronous scheme in which both Earth and Moon/Mars stations transmit free-running pulse trains, time tagging all transmit and receive events (Degnan, 2002). Advantages of the asynchronous scheme include (1) noise immunity, the remote station does not transmit pulses in response to numerous background photon detections; (2) stable laser performance, steady pulse generation results in thermally stable conditions; and (3) multiple Earth stations can range to the remote unit simultaneously without placing conflicting demands on the remote transmitter. With this scheme, clocks at each end of the link provide the baseline for time measurements, though only the Earth clock needs to keep absolute time. A frequency error in the remote clock is separable from velocity effects given multiple bidirectional measurements.

Despite the difference between beam pointing requirements for laser ranging and a lasercom transceiver, system-level requirements may be merged effectively to form a dual-functioning instrument serving both applications. Key differences between the two subsystems and methods to mitigate them are as follows:

1. Extensive transmit/receive isolation is typically required at each transceiver to maximize the link capacities of a bidirectional laser communications system. In a laser-ranging system, each transceiver must direct a portion of the transmitted light to its receiver in order to get an unbiased measurement of the transmit/receive time intervals.
2. Planetary communications links are typically designed for highly asymmetric performance, with a downlink data rate much higher than the uplink. A ranging system requires allocation of equal timing precision on each link.

3. The bandwidth and peak-to-average power ratio are chosen to maximize the data rate in a communications system. In a ranging system, a higher bandwidth always improves the precision at a given power, as does a higher peak-to-average power ratio in the presence of background light. Thus, a ranging system will typically require shorter laser pulses, higher pulse energies at lower repetition rates, and higher detector bandwidths (lower jitter).

Techniques to mitigate each of the above possible conflicts have been identified and are being implemented in the laboratory (Hemmati et al., 2009).

Summary

Technology of free-space laser communications at the component, subsystem, and system levels has advanced enough to successfully validate the lasercom technology at the maximum Mars range with data rates approaching 500 Mb/s – nearly two orders of magnitude beyond the state-of-the-art radio-frequency communications systems. There is opportunity for significant additional link margin improvement. When realized, the additional link margin may be used to reduce mass, power consumption, size, and complexity of the flight transceiver. The enhanced data rate capability will, for example, enable streaming high-quality video from other planets to the Earth. The "light science" enabled by lasercom (akin to radio science) remains a largely unexplored field.

Key remaining engineering challenges include circumventing attenuation due to dense clouds/fog, communicating while the receiver field of view is within a 1° of the Sun, and addressing the inadequacy of ground station infrastructure. While future Earth-based optical receivers may be located on airborne or Earth-orbiting platforms to circumvent atmospheric issues, technology and operational demonstrations for the first decade or so of deep-space laser communications will likely rely on lower risk and less-expensive ground-based receivers. Ground station diversity can adequately mitigate cloud cover. Large-diameter membrane filters located at the entrance aperture of ground-based telescopes, or telescopes designed to circumvent Sun interference, will aid with communications at small Sun angles. The advent of affordable, large-diameter, non-diffraction-limited ground telescopes should make infrastructure development realizable.

Acknowledgments

The author is indebted to the members of the Optical Communications Group and associates at JPL whose work is reported here, particularly W. Farr, A. Biswas, J. Kovalik, G. Ortiz, K. Wilson, S. Piazzolla, W. T. Roberts, M. Wright, K. Birnbaum, B. Moision, J. Hamkins, K. Quirk, J. Gin, S. Townes, F. Pollara, P. Estabrook, D. Antsos, and F. Davarian.

Bibliography

Andrews, L. C., and Phillips, R. L., 2005. *Laser Beam Propagation through Random Media*, 2nd edn. Bellingham: SPIE Press.

Andrews, L. C., Phillips, R. L., and Hopen, C. Y., 2001. *Laser Beam Scintillation with Applications*. Bellingham: SPIE Press.

Barron, R. J., and Boroson, D. M., 2006. Analysis of capacity and probability of outage for free-space optical channels with fading due to pointing and tracking error. *Proceedings of SPIE*, **6105**, 61050B.

Biswas, A., Wright, M. W., Kovalik, J., and Piazzolla, S., 2005. Uplink beacon laser for Mars laser communication demonstration (MLCD). *Proceedings of SPIE*, **5712**, 931–100.

Biswas, A., Khatri, F., and Boroson, D., 2006. Near-Sun free-space optical communications from space. In *IEEE Aerospace Conference* paper# 1611

Boroson, D., 2007. A survey of technology-driven capacity limits for free-space laser communications. *Proceedings of SPIE*, **6709**, 670918.

Braun, A., Rudd, J. V., Cheng, H., Mourou, G., Kopf, D., Jung, I. D., Weingarten, K. J., and Keller, U., 1995. Characterization of short-pulse oscillators by means of a high-dynamic-range autocorrelation measurement. *Optics Letters*, **20**, 1889.

Bromage, J., Winzer, P. J., Nelson, L. E., Mermelstein, M. D., Horn, C., and Headley, C. H., 2003. Amplified spontaneous emission in pulsed-pumped Raman amplifiers. *IEEE Photonics Technology Letters*, **15**, 667–669.

Cameron, A., Dios, F., Rodriguez, A., Rubio, J. A., Reyes, M., and Alonso, A., 2005. Modeling of power fluctuations induced by refractive turbulence in a multiple-beam ground-to-satellite optical uplink. *Proceedings of SPIE*, **5892**, 196–205.

Caplan, D. O., 2007. Laser communication transmitter and receiver design. *Journal of Optical Fiber Communications*, **4**, 225–362.

Charplyvy, A. R., and Tkach, R. W., 1993. What is the actual capacity of single-mode fibers in amplified lightwave systems. *IEEE Photonics Technology Letters*, **5**, 666–668.

Chen, C.-C., and Lesh, J. R., 1994. Overview of the optical communications demonstrator. *Proceedings of SPIE*, **2123**, 85–94.

Chen, C.-C., Hemmati, H., Biswas, A., Ortiz, G., Farr, W., and Pedreiro, N., 2006. Simplified lasercom system architecture using a disturbance-free platform. *Proceedings of SPIE*, **6105**, 610505.

Degnan, J. J., 1993. Millimeter accuracy satellite laser ranging: a review. In Smith, D. E., and Turcotte, D. L. (eds.), *Contributions of Space Geodesy to Geodynamics: Technology, Geodyn. Ser.*, vol. 25, pp. 133–162, AGU, Washington, DC.

Degnan, J. J., 2002. Asynchronous laser transponders for precise interplanetary ranging and time transfer. *Journal of Geodynamics*, **34**, 551.

Deng, Y., Koch, M. W., Lu, F., Wicks, G. W., and Knox, W. H., 2004. Colliding-pulse passive harmonic mode-locking in a femtosecond Yb-doped fiber laser with a semiconductor saturable absorber. *Optics Express*, **12**, 3872.

Dolinar, S., Divsalar, D., Hamkins, H., and Pollara, F., 2000. Capacity of pulse position modulation (PPM) on gaussian and Wenn channels. *The Interplanetary Network Directorate Progress Report 42–142, Jet Propulsion Laboratory, Pasadena, California*. pp. 1–31. http://ipnpr.jpl.nasa.gov/progress-report/.

Dupriez, P., Piper, A., Malinowski, A., Sahu, J. K., Ibsen, M., Thomsen, B. C., Jeong, Y., Hickey, L. M. B., Zervas, M. N., Nilsson, J., and Richardson, D. J., 2006. High average power, high repetition rate, picosecond pulsed fiber master oscillator power amplifier source seeded by a gain-switched Laser diode at 1060 nm. *IEEE Photonics Technology Letters*, **18**, 1013–1015.

Farr, W. H., 2009. Negative avalanche feedback detectors for photon-counting optical communications. *Proceedings of SPIE*, **7199**, 23.

Gin, J. W., Nguyen, D. H., and Quirk, K. J., 2007. High data rate receiver for optical PPM communications. In *NASA Earth Science and Technology Conference Proceedings*, College Park, NSTC-07-0140.

Hamkins, J., 2008. Pulse position modulation. In Bidgoli, H. (ed.), *The Handbook of Computer Networks*. Hoboken: Wiley.

Hamkins, J., and Moision, B., 2004. Selection of modulation and codes for deep space optical communications. *Proceedings of SPIE*, **5338**, 123–130. San Jose, CA (USA).

Hemmati, H. (ed.), 2006. *Deep Space Optical Communications*. Hoboken: Wiley.

Hemmati, H., 2007. Interplanetary laser communications. *Optics & Photonics News*, **18**, 22–27. Washington D. C, (USA).

Hemmati, H. (ed.), 2009. *Near-Earth laser Communications*. Boca Raton: CRC Press. Chapter 10.

Hemmati, H., and Lesh, J. R., 1998. A combined laser communication and imager for micro-spacecraft (ACLAIM). *Proceedings of SPIE*, **3266**, 165. San Jose, CA (USA).

Hemmati, H., Wright, M. W., Biswas, A., and Esproles, C., 2000. High-efficiency pulsed laser transmitters for deep-space communication. *Proceedings of SPIE*, **3932**, 188–195. San Jose, CA (USA).

Hemmati, H., Biswas, A., and Boroson, D. M., 2007. Prospects for improvement of interplanetary laser communication data rates by 30dB. *Proceedings of IEEE*, **95**, 2082–2092. 10.1109/JPROC.2007.905057

Hemmati, H., Birnbaum, K. M., Farr, W. H., Turyshev, S., and Biswas, A., 2009. Combined laser-communications and laser-ranging transponder for Moon and Mars. *Proceedings of SPIE*. **7199**, 71990N-1 to N-12.

Hurd, W. J., MacNeal, B. E., Ortiz, G. G., Cheng, E. S., Moe, R. V., Walker, J. Z., Fairbrother, D., Dennis, M., Eegholm, B., and Kasunic, K. J., 2005. Exo-atmospheric telescopes for deep-space optical communications. In *IEEE Aerospace Conference*, paper # 1557.

Kachelmyer, A. L., and Boroson, D. M., 2007. Efficiency penalty of photon-counting with timing jitter. *Proceedings of SPIE*, **6709**, 670906.

Kerman, A. J., Dauler, E. A., Keicher, W. E., Yang, J. K., Berhhren, K. K., Gol'Tsman, G., and Voronov, B., 2006. Kinetic-inductance-limited reset time of superconducting nanowire photon counters. *Applied Physics Letters*, **88**, 111116–111119.

Kiasaleh, K., 1998. Turbo-coded optical PPM communications. *Journal of Lightwave Technology*, **16**, 18–26.

Link, R., Alliss, R., and Craddock, M. E., 2004. Mitigating the impact of clouds on optical communications. *Proceedings of SPIE*, **5338**, 223.

Lipes, R. G., 1980. Pulse-position-modulation coding as near-optimum utilization of photon counting channel with bandwidth and power constraints. In *DSN Reporting Program*. Vol. 42–56, pp. 108–113.

McMickell, M. B., Kreider, T., Hansen, E., Davis, T., and Gonzalez, M., 2007. Optical payload isolation using the miniature vibration isolation system. *Proceedings of SPIE*, **6527**, 652703.

Moision, B., 2008a. Communication limits due to photon detector jitter. *IEEE Photonics Technology Letters*, **20**, 715–717.

Moision, B., 2008. Universal decoder for variable duty-cycle optical communications., In *JPL's IPN Progress Report*, pp. 42–173.

Moision, B., and Biswas, A., 2007. Memory requirements to mitigate fading loss on an optical channel. In *JPL's IPN Progress Report*, pp. 42–170.

Moision, B., and Farr, W. H., 2008. Communication losses due to photon-detector jitter. *IEEE Photonics Letters*, **20**, 715–717.

Moision, B., and Hamkins, J., 2002. Constrained coding for the deep-space optical channel. *Proceedings of SPIE*, **4635**, 127–137. San Jose, CA (USA).

Moision, B., and Hamkins, J., 2003. Deep-space optical communications downlink budget: modulation & coding. In *IPN Progress Report 42–154*. pp. 1–28.

Moision, B., Hamkins, J., and Cheng, M., 2006. Design of a coded modulation for deep space optical communications. In *Information Theory and Applications*. San Diego: University of California, February 6–10, 2006.

Murphy, T. W., Michelsen, E. L., Orin, A. E., Adelberger, E. G., Hoyle, C. D., Swanson, H. E., Stubbs, C. W., and Battat, J. E., 2007. APOLLO: a new push in lunar laser ranging. In *Proceedings of the International Workshop From Quantum to Cosmos: Fundamental Physics Research in Space*. May 22–24, 2006, Warrenton, International Journal of Modern Physics D, **16** (12a), 2127.

Ortiz, G. G., and Farr, W. H., Charles, J. R., Roberts, W. T., Sannibale, V., Gin, J., Saharaspude, A., and Garkanian, V., 2009. "Canonical deep space optical communications transceiver", Proc. SPIE 7199, Free-Space Laser Communication Technologies XXI, 71990K (February 24,2009); doi:10.1117/12.812410; http://dx.doi.org/10.1117/12.812410.

Ortiz, G. G., and Lee, S., 2006. Chapter 5.3, deep-space acquisition, tracking and pointing. In Hemmati, H. (ed.), *Deep Space Optical Communications*. Hoboken: Wiley.

Ortiz, G. G., Le, S., and Alexander, J. W., 2001. Sub-microradian pointing for deep-space optical telecommunications network. In *19th AIAA International Communications Satellite Systems Conference*. Toulouse.

Quaale, R., Hindman, B., Engberg, B., and Collier, P., 2005. Mitigating environmental effects on free-space laser communications. In *IEEE Aerospace Conference*. Montana.

Quirk, K. J., and Gin, J. W., 2006. Optical PPM combining loss for photon counting receivers. In *Military Communications Conference, IEEE (MILCOM)*. pp. 1–4.

Roberts, W. T., 2005. Optical membrane technology for deep space optical communications filters. In *IEEE Aerospace Conference Proceedings*. Big Sky, 1991–2000.

Robinson, B., Caplan, D. O., Stevens, M. L., Barron, R. J., Dauler, E. A., and Hamilton, S. A., 2005. 1.5-photons.bit photon-counting optical communications using Geiger-mode avalanche photodiodes. In *IEEE LEOS Summer Topical Meetings*.

Sannibale, V., Ortiz, G. G., Farr, W. H., 2009. A sub-Hertz vibration isolation platform (LFVIP) for a deep space optical communication transceiver. *Proceedings of SPIE*. **7199**, 7199–17.

Scozzafava, J. J., Boroson, D. M., Burnside, J. W., Glynn, M. L., DeVoe, C. E., DeFranzo, C. M., and Doyle, K. B., 2007. Design of a very small inertially stabilized optical space terminal. *Proceedings of SPIE*, **6709**, 670905. San Jose, CA (USA).

Smith, D. E., Zuber, M. T., Sun, X., Neumann, G. A., Cavanaugh, J. F., McGarry, J. F., and Zagwodzki, T. W., 2006. Two-way laser link over interplanetary distance. *Science*, **311**(5757), 53.

Smutny, B., Kaempfner, H., Muehlnikel, G., Sterr, U., Wandernoth, B., Heine, F., Hildebrand, U., Dallma, D., Reinhardt, M., Freier, A., Lange, R., Boehmer, K., Feldhaus, T., Mueller, J., Weichert, A., Greulich, P., Seel, S., Meyer, R., and Czichy, R., 2009. 5.6 Gbps optical intersatellite communication link. *Proceedings of SPIE*, **7199**, 719906. San Jose, CA (USA).

Spellmeyer, N. W., Caplan, D. O., and Stevens, M. L., 2005. Design for a 5-watt PPM transmitter for the mars laser communications demonstration. In *Digest of the LEOS Summer Topical Meetings*. Paper TuA4.1, pp. 51–52.

Spyak, P. R., and Wolf, W. L., 1992. Scatter from particulate-contaminated mirrors. Part 1: theory and experiment for polystyrene spheres and $\lambda = 0.632$ μm. *Optical Engineering*, **31**, 1746.

Stern, J. A., and Farr, W. H., 2007. Fabrication and characterization of superconducting NbN nanowire single photon detector. *IEEE Transactions on Applied Superconductivity*, **17**, 306–309.

Tan, Y., Guo, J., Ai, Y., Liu, W., and Fei, Y., 2008. A coded modulation scheme for deep-space optical communications. *IEEE Photonics Technical Letters*, **20**, 372–374.
Tolker Nielsen, T., and Oppenhaeuser, G., 2002. In orbit test result of an operational optical intersatellite link between ARTEMIS and SPOT4, SILEX. *Proceedings of SPIE*, **4635**, 1. San Jose, CA (USA).
Turyshev, S. G., and Williams, J. G., 2007. Space-based tests of gravity with laser ranging. *International Journal of Modern Physics*, **D16**, 2165–2179.
Williams, J. G., Turyshev, S. G., and Boggs, D. H., 2009. Lunar laser ranging tests of the equivalence principle with the earth and moon. *International Journal of Modern Physics*, **D18**, 1129–1175.
Wilson, K. E., Lesh, J. R., and Yan, T.-Y., 1993. GOPEX: a laser uplink to the Galileo spacecraft on its way to Jupiter. *Proceedings of SPIE*, **1866**, 138. San Jose, CA (USA).
Wilson, K., Lesh, J. R., Araki, K., and Arimoto, Y., 1997. Overview of ground-to-orbit lasercom demonstration. *Proceedings of SPIE*, **2990**, 23–30.
Wilson, K., Troy, M., Srinivasan, M., Platt, B., Vilnrotter, V., Wright, M., Garkanian, V., Hemmati, H., 2003. Daytime adaptive optics for deep space optical communications. In *Conference Proceedings International Space Conference of Pacific-basin Societies (ISCOPS)*, Tokyo.
Wilson, K., Kovalik, J., Biswas, A., and Roberts, W. T., 2007. Development laser beam transmission strategies for ground-to-space optical communications. *Proceedings of SPIE*, **6551**, 65510B. San Jose, CA (USA).
Winzer, P. J., Kalmar, A., and Leeb, W. R., 1999. Role of amplified spontaneous emission in optical free-space communication links with optical amplification – impact on isolation and data transmission; utilization for pointing acquisition and tracking. *Proceedings of SPIE*, **3615**, 34–141. San Jose, CA (USA).
Wojcik, G. S., Szymczak, H. L., Alliss, R. J., Link, R. P., Craddock, M. E., and Mason, M. L., 2005. Deep-space to ground laser communications in a cloudy world. *Proceedings of SPIE*, **5892**, 589203. San Jose, CA (USA).
Wyner, A. D., 1988. Capacity and error exponent for the direct detection photon channel – part I. *IEEE Transactions on Information Theory*, **34**, 1449–1461.
Wysocki, P., Wood, T., Grant, A., Holcomb, D., Chang, K., Santo, M., Braun, L., and Johnson, G., 2006. High reliability 49 dB gain, 13 W PM fiber amplifier at 1550 nm with 30 dB PER and record efficiency. In *Optical Fiber Communication Conference*. Long Beach, paper PDP17.
Yuksel, H., Milner, S., and Davis, C. C., 2005. Aperture averaging for optimizing receiver design and system performance on free-space optical communication links. *Journal of Optical Networking*, **4**, 462–475.

EMERGING TECHNOLOGIES, LIDAR

David M. Tratt
The Aerospace Corporation, Los Angeles, CA, USA

Synonyms

Ladar; Laser radar; Optical radar

Definitions

Lidar. LIght Detection And Ranging.
Ladar. LAser Detection And Ranging.

Introduction

The term *lidar* (i.e., the optical analogue of radar) embraces a broad collection of active optical probing techniques for remotely determining the physical and chemical characteristics of diffuse or condensed matter targets, both natural and otherwise. Although lidar methods have been practiced since before the dawn of the laser era, their classification here as an emerging technology reflects the fact that they have only in relatively recent years begun to see systematic exploitation from space-based vantage points (most especially with the advent of robust, high-performance laser systems).

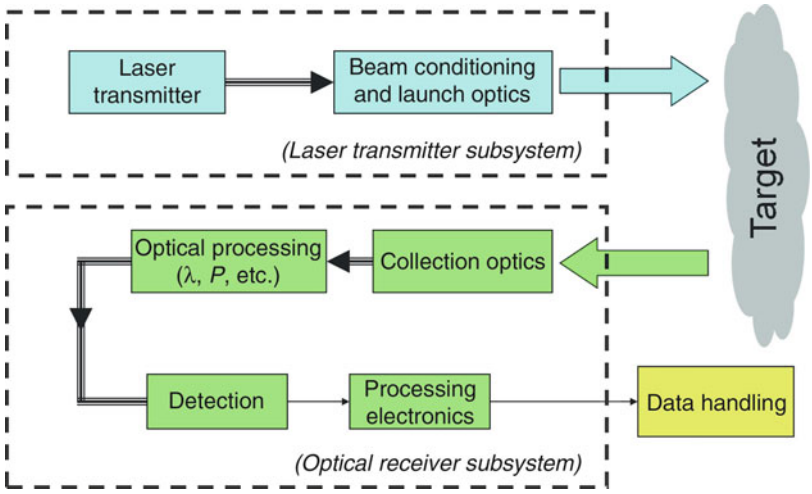
Some remarks on terminology are in order as a guide to those unfamiliar with this field. Although the terms *lidar* and *ladar* are essentially interchangeable (except for exceedingly rare instances in which an incoherent source is used and which would therefore exclude consideration of *ladar*), different application communities have preferentially adopted one over the other. Hence, the atmospheric science community uses the form *lidar*, while those working in areas that involve remote characterization of hard targets have adopted *laser radar* or *ladar* (McManamon, 2012). It is also possible to find instances where *ladar* and *lidar* appear in upper case, reflecting their origin as acronyms. This is becoming less common now that these terms (like *radar* before them) have entered general parlance, but the hybrid form LIDAR is nevertheless commonplace albeit used exclusively by the airborne laser terrain mapping community.

A number of textbooks are available that treat lidar systems and methodology from a variety of standpoints. Measures (1992) provides a thorough grounding in the basic theory and practice of laser remote sensing, while Fujii and Fukuchi (2005), Weitkamp and Walther (2005), and Kovalev and Eichinger (2004) collectively offer a comprehensive overview of the wide array of measurement applications that have benefitted from lidar. Grant et al. (1997) provide an anthology of seminal lidar papers from the earliest days, when arc lamps were the brightest sources available, up to the mid-1990s.

Lidar basics

A canonical lidar system comprises the set of subsystems depicted in Figure 1. Depending on the intended application, the laser transmitter may require stringent control of wavelength, spectral purity, polarization, pulse duration, or some combination of all of these properties. By corollary, the receiver system may also be required to analyze one or all of these same properties of the return radiation.

Following the detection stage, processing electronics convert the signals to a range-gated (i.e., time-of-flight registered) data stream which is then logged for further processing to extract the desired observable quantity. This is accomplished through an understanding of the often complex interaction phenomenology of the laser radiation with the target and frequently leads to a suboptimally constrained retrieval process which must be augmented



Emerging Technologies, Lidar, Figure 1 Functional schematic of a generic lidar system.

by ancillary measurements, modeled parameters, or a combination of both. The optical power $P(R)$ returned to the lidar detection subsystem by a scattering volume located at range R is given by the standard basic lidar equation:

$$P(R) = \left[E\eta O(R)\frac{cA}{2R^2}\right][T(R)T'(R)\beta(\lambda, R)], \qquad (1)$$

where E is the pulse energy emitted into the atmosphere (outward transmission T, return transmission T'), A is the receiver collecting area, and β is the backscatter from the range element of interest at the specified emission wavelength, λ. η is the overall optical efficiency of the receiver subsystem and $O(R)$ is the range-dependent transmit/receive overlap function. (In general, the optical efficiency represents the optical transmission. However, for a *coherent* detection lidar system, this parameter will also include a component due to heterodyne mixing efficiency (Zhao et al., 1990).) The distinction made between outgoing and return atmospheric transmission recognizes that the incident radiation in some cases will be inelastically scattered, as explained below. Terms grouped together in the first set of square parentheses in Equation 1 are lidar system-dependent parameters, while those contained in the second set of square parentheses characterize the atmospheric response. This expression also assumes single scattering. Many lidar systems are subject to multiple scattering to a greater or lesser extent, and this will complicate the data inversion procedure. Eloranta (1998) provides a discussion of these issues and techniques for their resolution.

While pulsed operation is implicitly assumed in Equation 1, other forms of range gating have used modulation of continuous wave (cw) laser sources, such as pseudorandom noise encoding (Takeuchi et al., 1986), in conjunction with synchronous detection of the return to extract the time-dependent signal behavior.

Most, if not all, lidars are operated in the backscattering or quasi-backscattering configuration in which the transmitter and receiver are collocated and form part of an integral instrument. When the transmit and receive optical axes are coaxial, the lidar is said to be *monostatic*, whereas if they are offset, the lidar is described as *bistatic*. Note that this nomenclature differs from that applied in the radar field in that a bistatic radar system comprises transmitter and receiver elements that are geographically separated, whereas a monostatic radar implies collocated transmitter and receiver systems that may or may not be coaxial. Multistatic lidar configurations comprising two or more receiver stations have been considered for certain specialized applications (e.g., Olofson et al., 2008), but thus far, none have been implemented beyond the experimental stage.

An additional distinction that is frequently used to classify lidar systems is the mode of photon detection employed. In a *direct* (*incoherent*) *detection* lidar, the received photons are converted directly to electrical impulses by the detector with the loss of all phase information, and this is the detection method found in the overwhelming majority of lidar systems. In a *heterodyne* (*coherent*) *detection* lidar, the received photons are mixed with a reference oscillator signal and detection is performed on the intermediate frequency. Such a configuration is technologically more complex than a direct detection system, and also makes considerably more stringent demands on laser frequency stability and beam quality, but preserves phase knowledge and is essential if the ultimate goal is to directly retrieve motion-induced Doppler signatures, such as for atmospheric wind measurement (Huffaker and Hardesty, 1996) or object tracking (Osche and Young, 1996). A few coherent detection lidars have been described that are non-Doppler capable (e.g., Grant et al., 1987; Menzies and Tratt, 1994; Gibert et al., 2006). In these cases the additional system

complexity is tolerated because of the much greater noise rejection that is afforded by coherent detection, which can be as much as an order of magnitude above that of a comparable direct detection system (Menzies, 1976).

Inelastic backscatter lidar

Inelastic backscatter lidars make use of wavelength conversion properties of the target to infer details of its physical and chemical makeup. The chief downshifting processes that are commonly employed in the lidar field are Raman and fluorescence scattering, both of which require excitation in the UV-visible spectral range.

Although the weak cross sections for Raman scattering tend to limit its usefulness for long-range remote sensing purposes, this disadvantage is offset by the high specificity that it affords (because the observed wavelength shift is diagnostic of the molecule being probed). Hence, many users tolerate the need for a high-power transmitter, large collection aperture, and long accumulation intervals in order to gain the benefit of unambiguous measurements, which is clearly of central importance when such systems are used to validate satellite data products (Wessel et al., 2000; Tratt et al., 2005). Raman lidar has met with significant success in the ground-based remote sensing of atmospheric constituents and structure (Turner and Whiteman, 2006), atmospheric dynamics (Koch et al., 1991), and hazardous substances (Sharma et al., 2005) and has also seen limited airborne application (Heaps and Burris, 1996).

Fluorescence lidars are generally used for the investigation of chemical and biological targets. For example, laser-induced fluorescence has been successfully used in the remote assessment of plant health (Saito et al., 2000), airborne monitoring of phytoplankton populations and ocean productivity (Hoge et al., 2005), and the diagnosis of biodeterioration in building fabrics (Weibring et al., 2001). These techniques rely on measurement of the 690 nm chlorophyll-α fluorescence signature (sometimes in association with other features) and an understanding of the way it is modified by environmental disturbances. Fluorescence lidar has also been proposed for monitoring of marine oil spills (Karpicz et al., 2006) and the identification of airborne chemical agents or pathogens (Buteau et al., 2007).

So-called resonance fluorescence techniques, as the name implies, resonantly excite a target atomic species in order to directly probe the atmospheric properties at the residence altitude of the species in question, which are typically metal ions entrained in the middle and upper atmosphere (80–100 km altitude) that are derived from, and continually replenished by, meteoritic deposition. Hence, there is a long heritage of using sodium lidars to track atmospheric dynamics in these difficult-to-access regions (Gardner, 1989) as well as to retrieve upper atmospheric temperatures through measurement of the Doppler-broadened atomic lineshape, which can then be related directly to the ambient temperature. Other atomic metal species that have been studied for similar purposes are potassium (Papen et al., 1995) and iron (Kane and Gardner, 1993). More recently, iron Boltzmann lidars have been used to infer atmospheric temperature by measuring the spectral content of radiation emitted by resonantly excited mesospheric iron layer atoms (Gelbwachs, 1994; Raizada and Tepley, 2002). An extensive treatment of this field has been given by Chu and Papen (2005).

An intriguing new class of inelastic backscatter lidar is the so-called femtosecond white-light lidar. This employs an ultrashort-pulse transmitter operating in the near-IR at around 800 nm whose high intensity gives rise to nonlinear self-phase modulation that results in broadband upconverted continuum emission across the UV-IR spectral range (Kasparian et al., 2005). The critical factor in such systems is that the broadband continuum is emitted predominantly in the backward direction and is thus a particularly strong and effective backscattering mechanism (Yu et al., 2001). Applications for the femtosecond lidar center on investigations requiring multiple wavelength channels to constrain multivariate retrievals, such as high-resolution atmospheric absorption spectrum measurement and aerosol composition and microphysics determination (Kasparian et al., 2003).

Elastic backscatter lidar

The majority of lidar systems in use can be broadly classed as elastic backscatter systems. Elastic backscattering processes conserve the centroid wavelength and consist of Rayleigh (or molecular) scattering from air molecules and particle scattering from suspended particulate material (aerosols and clouds). (The term Mie scattering is frequently used to implicitly denote particle scattering; however, the Mie descriptor strictly speaking only applies to scattering from *spherical* particles, which in general account for only a small percentage of the total atmospheric aerosol population.) Rayleigh scattering magnitude is inversely proportional to the fourth power of wavelength, which in practice limits the usefulness of Rayleigh lidar approaches to wavelengths not exceeding the visible spectrum. Rayleigh lidar systems tend to be used for profiling fundamental atmospheric properties such as temperature and density (Chanin and Hauchecorne, 1984).

By far the commonest application of lidar is concerned with cloud and aerosol (i.e., particulate) detection and characterization. In its simplest form, this can entail measurements at a single wavelength to infer the broad characteristics of scattering structure in atmospherically entrained dust and cloud features. However, in the absence of ancillary data, such minimally constrained measurements can only acquire estimates of attenuated backscatter. To properly retrieve aerosol optical properties and physical parameters requires a multiwavelength lidar approach and calls for backscatter coefficients at a minimum of three wavelengths and aerosol extinction coefficients for at least two wavelengths (e.g., Müller et al., 1998; Althausen et al.,, 2000). Aerosol extinction

has typically been obtained from collocated, contemporaneous column optical depth measurements (using sunphotometers) or from auxiliary Raman lidar channels, which measure the Raman backscatter from a well-mixed atmospheric constituent (usually nitrogen) and observe its departure from expected values to ascribe the aerosol extinction (Ansmann et al., 1992). However, in recent years there has been increasing interest in the high spectral resolution lidar (HSRL) technique because of its ability to separate and directly measure the aerosol and molecular scattering components with a single, self-contained multiwavelength instrument that avoids recourse to ancillary measurements (Eloranta, 2005; Hair et al., 2008).

In addition to the wealth of ground-based and airborne cloud/aerosol lidar systems that are today operated around the world, some functioning as components of autonomous, long-term monitoring networks, three multiwavelength systems have flown in Earth orbit, one on the Space Shuttle (Winker et al., 1996) and two on dedicated free-flying platforms (Spinhirne et al., 2005; Winker et al., 2007). There has even been a cloud/aerosol lidar operated on the surface of Mars (Whiteway et al., 2008).

Polarimetric lidar

The shape of cloud/aerosol particles can also modify the polarization state of scattered radiation. Polarization lidars exploit this phenomenon by separately measuring the unaltered (p) and depolarized (s) components of the return radiation, for which a range-dependent *backscatter depolarization ratio*, $\delta(R)$, can then be defined:

$$\delta(R) = \frac{P_s(R)}{P_p(R)}. \tag{2}$$

(Note, however, that Gimmestad (2008) has offered a recasting of the traditional polarization lidar formulation that is more descriptive of the microphysical processes involved and which may in time supersede this notation.)

The magnitude of the depolarization ratio can be used to deduce the degree of nonsphericity of the particles. The relative polarimetric properties of various aerosol and cirrus particles show sufficient diversity that polarization lidar can provide for some level of discrimination between the different classes of such particles (Sassen, 1991; Sassen, 2007). However, care must be taken when inferring scattering or attenuation coefficients from lidar measurements of cirrus, as the preferential horizontal orientation of cirrus platelets has been shown to dramatically enhance lidar returns through the addition of a specular reflection component to the backscatter signal (Platt et al., 1978). In order to separate the intrinsic cirrus scattering properties from this effect, it is therefore advisable to offset the lidar viewing angle by $\sim 3^\circ$ from the vertical.

Differential absorption lidar

The differential absorption lidar (DIAL) evolved as a means for identifying and quantifying gaseous atmospheric constituents. The technique entails transmitting two different wavelengths, one tuned to a characteristic spectral feature of the target species and the other offset by a small but sufficient amount that it is unaffected by the target gas and so acts as an atmospheric transmissivity reference. The usual practice is to contrive the on-and off-resonant wavelengths to be as narrowly separated as possible (in order to eliminate dispersive effects from other atmospheric constituents that could confuse the measurement), so that the following close approximation for the number density $N(R)$ of absorbing molecules can be applied:

$$N(R) = \frac{1}{2|\sigma|} \frac{d \ln\left[P_{on}(R)/P_{off}(R)\right]}{dR}, \tag{3}$$

in which σ is the differential absorption cross section between the on-and off-resonant wavelengths.

A large number of DIAL systems are in operation worldwide to measure a wide variety of atmospheric trace gases from the ground and from the air. Because of their climatic impacts, water vapor and ozone profiling are two common DIAL measurement objectives (Browell et al., 2004), while a plethora of systems operating from the UV to the thermal-IR are in use to track the transport and dispersion of plumes containing common pollutants such as sulfur dioxide (Weibring et al., 1998), ammonia (Zhao et al., 2002), hydrocarbons (Murdock et al., 2008), and many other compounds.

The ongoing debate over the role of carbon dioxide in climate change has driven a resurgence in interest concerning DIAL techniques for monitoring this atmospheric constituent. The requirement is for mixing ratio measurement globally throughout the lowermost levels of the troposphere with a precision of $\sim$0.5 %, which is considerably more challenging than the typical $\sim$20 % DIAL precision specification. Conventional DIAL is one of two approaches that are under evaluation for this application (Koch et al., 2004; Gibert et al., 2006), the other being integrated path differential absorption (IPDA), which entails cw or quasi-cw illumination of the Earth's surface at on-and off-resonant probe wavelengths (Menzies and Tratt, 2003). The retrieved differential absorption return can then be transformed into a path-averaged column content (Abshire et al., 2010; Spiers et al., 2011). Detailed trade studies will ascertain the optimum wavelength combination and detection configuration for acquiring high-precision gas mixing ratios by this technique.

Doppler lidar

Doppler lidars are used to measure target motion, either for atmospheric wind measurement (Baker et al., 1995) or object tracking (Osche and Young, 1996), and can be divided into two distinct classes. *Coherent detection Doppler lidars* directly measure the Doppler shift imposed on the optical return by the atmospheric target by mixing the return signal with a well-defined frequency reference

laser, termed a *local oscillator*. The heterodyne mixing efficiency in such systems is a sensitive function of frequency purity, wave front quality, and optical aberration factors, any one of which can be the source of significant performance degradation if not adequately managed (Zhao et al., 1990; Chambers, 1997). For hard target coherent Doppler lidar systems, where backscatter is high, these considerations are less of an issue, but for atmospheric applications, where performance is contingent on aerosol and cloud backscatter, one is effectively driven to operating wavelengths in the IR spectral region.

By contrast, *direct detection Doppler lidars* operate in the "photon bucket" mode and thus do not mandate the wave front quality requirements of coherent systems. In a direct detection system, the Doppler shift is inferred indirectly by measuring the differential transmission of the return signal through a high-resolution filter mechanism whose dispersion characteristics must be known to high accuracy and very well controlled. Freed from the IR wavelength constraint to which coherent systems are subject, direct detection Doppler lidars are thus designed to take advantage of the omnipresence of molecules in the atmosphere by operating in Rayleigh backscatter mode (usually in the eye-safe UV). Direct detection systems are less photon efficient than their coherent counterparts, and their retrieval accuracy can be compromised by the individual and collective stability of optical components within the receiver subsystem, but are preferred for operation against atmospheres that have little or no aerosol content. The combining of both Doppler lidar approaches in a *hybrid* instrument has been suggested as a means to make best use of the advantages of both types of system (Emmitt, 2004).

There is a significant body of experience with both surface-based (Huffaker and Hardesty, 1996; Grund et al., 2001; Henderson and Hannon, 2005) and airborne (Bilbro et al., 1986; Rothermel et al., 1998; Werner et al., 2001; Hannon et al., 1999) coherent Doppler wind lidars. In addition, there have been several ground-based direct detection Doppler wind lidars (von Zahn et al., 2000; Gentry et al., 2000; Irgang et al., 2002) and also a few airborne systems (Dehring et al., 2006; Durand et al., 2006). All require some means of scanning the transceiver in order to obtain the multiple perspective line-of-sight retrievals that are necessary to resolve 2D or 3D wind vector fields.

The European Space Agency (ESA) plans to deploy the first space-based Doppler lidar. The Atmospheric Dynamics Mission (ADM-Aeolus) will carry a direct detection Doppler wind lidar and an integral high spectral resolution lidar (HSRL) with two receiver channels for analysis of aerosol/cloud backscatter and molecular backscatter, respectively (Stoffelen et al., 2005). This instrument will address only a single off-nadir line-of-sight perspective but will nevertheless represent the first on-orbit demonstration of Doppler wind lidar techniques.

It is worth noting here that the first lidar wind measurements from space were made by a non-Doppler elastic backscatter system using optical scatterometry of the ocean surface (Menzies et al., 1998). These measurements were necessarily indirect, scalar in character, and restricted to sea surfaces only, but more recent measurements indicate that this approach can provide significant new information on ocean winds and associated air–sea interaction processes (Hu et al., 2008).

Laser ranging and geodetic imaging

Laser ranging to hard targets was among the earliest of lidar applications and has seen by far the most commercial exploitation. Apart from the proliferation of laser-augmented rangefinding binoculars and surveying instruments, the technique is now in high demand for airborne high-density terrain mapping (Slatton et al., 2007) and shallow-water hydrographic surveying (Banic and Cunningham, 1998), for which a significant commercial services capability has arisen. Laser altimetry was also the first lidar application to be implemented in space (Sjogren and Wollenhaupt, 1973; Bufton, 1989; Garvin et al., 1998), and global-scale geodetic mapping by laser altimetry can in many respects be regarded as commonplace, having now been demonstrated with moderate horizontal resolution (10–100 m) on Earth (Schutz et al., 2005), Mars (Smith et al., 2001), Mercury (Zuber et al., 2008), the Moon (Smith et al., 1997; Smith et al., 2010), and the irregular asteroids Eros (Cheng et al., 2002) and Itokawa (Mukai et al., 2007). Most of these were first-return time-of-flight instruments, the exception being the Geoscience Laser Altimeter System (GLAS), which was designed to transcribe the entire return waveform (Abshire et al., 2005). Full waveform recovery makes it possible for laser altimetry to resolve volumetric vegetation structure and sub-canopy topography by taking advantage of the optical porosity of vegetation cover (Blair et al., 1999; Lefsky and McHale, 2008).

Some measurement scenarios call for a level of areal density of individual ranging measurements that pushes horizontal resolution toward the meter range. For these applications, limited pulse energy and the need for rapid scanning mean that "photon starvation" becomes a significant issue (Abrams and Tratt, 2005). To address this concern, single-photon counting detector arrays with integrated readout circuitry have been developed to acquire 3D imagery of laser flood-illuminated scenes in what is often termed a *flash lidar* scheme (Albota et al., 2002). In addition, work is currently in process to extend active image formation techniques from the established synthetic aperture radar domain into the optical spectral region (Beck et al., 2005). This technique faces nontrivial technical challenges but is nevertheless being pursued for tactical applications (Ricklin and Tomlinson, 2005) and has also been proposed for planetary remote sensing (Karr, 2003).

High-density laser ranging is also used in space-to-space scenarios where the uncompromising demands of autonomous rendezvous and proximity operations rule out less precise instrumental approaches (Nimelman et al., 2006).

Summary

Active optical remote sensing (lidar) techniques have provided demonstrable enhancements in spatiotemporal resolution and reduced dependence on the diurnal cycle in comparison to other remote sensing methods used for Earth and planetary science applications. While the breadth and diversity of lidar methodology attests to its fundamental versatility, there are certain measurements that are enabled uniquely by active optical sensors (e.g., aerosol and cloud vertical structure, global-scale tropospheric wind measurement, 3D vegetation structure, high-resolution sub-canopy topography mapping). The intrinsic benefits of lidar now drive a steadily increasing number of fielded systems on the ground, in the air, and in space. While not exhaustive, this entry describes the basic characteristics and associated applications of the commonest types of lidar likely to be encountered and also provides an introduction to some promising emerging concepts that have yet to fully mature into accepted field-capable systems.

Bibliography

Abrams, M. C., and Tratt, D. M., 2005. Progress in laser sources for lidar applications: laser sources for 3D imaging remote sensing. *Proceedings of SPIE*, **5653**, 241, doi:10.1117/12.580323.

Abshire, J. B., Sun, X., Riris, H., Sirota, J. M., McGarry, J. F., Palm, S., Yi, D., and Liiva, P., 2005. Geoscience laser altimeter system (GLAS) on the ICESat mission: on-orbit measurement performance. *Geophysical Research Letters*, **32**, L21S02, doi:10.1029/2005GL024028.

Abshire, J. B., Riris, H., Allan, G. R., Weaver, C. J., Mao, J., Sun, X., Hasselbrack, W. E., Kawa, S. R., and Biraud, S., 2010. Pulsed airborne lidar measurements of atmospheric CO_2 column absorption. *Tellus*, **62B**, 770, doi:10.1111/j.1600-0889.2010.00502.x.

Albota, M. A., Heinrichs, R. M., Kocher, D. G., Fouche, D. G., Player, B. E., O'Brien, M. E., Aull, B. F., Zayhowski, J. J., Mooney, J., Willard, B. C., and Carlson, R. R., 2002. Three-dimensional imaging laser radar with a photon-counting avalanche photodiode array and microchip laser. *Applied Optics*, **41**, 7671, doi:10.1364/AO.41.007671.

Althausen, D., Müller, D., Ansmann, A., Wandinger, U., Hube, H., Clauder, E., and Zörner, S., 2000. Scanning 6-wavelength 11-channel aerosol lidar. *Journal of Atmospheric and Oceanic Technology*, **17**, 1469, doi:10.1175/1520-0426(2000)017<1469:SWCAL>2.0.CO;2.

Ansmann, A., Wandinger, U., Riebesell, M., Weitkamp, C., and Michaelis, W., 1992. Independent measurement of extinction and backscatter profiles in cirrus clouds by using a combined Raman elastic-backscatter lidar. *Applied Optics*, **31**, 7113, doi:10.1364/AO.31.007113.

Baker, W. E., Emmitt, G. D., Robertson, F., Atlas, R. M., Molinari, J. E., Bowdle, D. A., Paegle, J., Hardesty, R. M., Menzies, R. T., Krishnamurti, T. N., Brown, R. A., Post, M. J., Anderson, J. R., Lorenc, A. C., and McElroy, J., 1995. Lidar-measured winds from space: a key component for weather and climate prediction. *Bulletin of the American Meteorological Society*, **76**, 869, doi:10.1175/1520-0477(1995)076<0869:LMWFSA>2.0.CO;2.

Banic, J. R., and Cunningham, A. G., 1998. Airborne laser bathymetry: a tool for the next millennium. *EEZ Technology*, **3**, 75. http://www.jalbtcx.org/downloads/Publications/32Banic_Cunningham_98.pdf.

Beck, S. M., Buck, J. R., Buell, W. F., Dickinson, R. P., Kozlowski, D. A., Marechal, N. J., and Wright, T. J., 2005. Synthetic-aperture imaging laser radar: laboratory demonstration and signal processing. *Applied Optics*, **44**, 7621, doi:10.1364/AO.44.007621.

Bilbro, J. W., DiMarzio, C. A., Fitzjarrald, D. E., Johnson, S. C., and Jones, W. D., 1986. Airborne Doppler lidar measurements. *Applied Optics*, **25**, 3952, doi:10.1364/AO.25.003952.

Blair, J. B., Rabine, D. L., and Hofton, M. A., 1999. The laser vegetation imaging sensor: a medium-altitude, digitization-only, airborne laser altimeter for mapping vegetation and topography. *ISPRS Journal of Photogrammetry and Remote Sensing*, **54**, 115, doi:10.1016/S0924-2716(99)00002-7.

Browell, E. V., Grant, W. B., and Ismail, S., 2004. Environmental measurements: laser detection of atmospheric trace gases. In Guenther, R. D., Steel, D. G., and Bayvel, L. (eds.), *Encyclopedia of Modern Optics*. Amsterdam: Elsevier, pp. 403–416.

Bufton, J. L., 1989. Laser altimetry measurements from aircraft and spacecraft. *Proceedings of the IEEE*, **77**, 463, doi:10.1109/5.24131.

Buteau, S., Simard, J.-R., Lahaie, P., Roy, G., Mathieu, P., Déry, B., Ho, J., and McFee, J., 2007. Bioaerosol standoff monitoring using intensified range-gated laser-induced fluorescence spectroscopy. In Kim, Y. J., and Platt, U. (eds.), *Advanced Environmental Monitoring*. Dordrecht: Springer, pp. 203–216, doi:10.1007/978-1-4020-6364-0_16.

Chambers, D., 1997. Modeling of heterodyne efficiency for coherent laser radar in the presence of aberrations. *Optics Express*, **1**, 60, doi:10.1364/OE.01.000060.

Chanin, M. L., and Hauchecorne, A., 1984. Lidar studies of temperature and density using rayleigh scattering. In *International Council of Scientific Unions Middle Atmosphere Handbook*, Vol. 13, Urbana, Illinois, USA: Scientific Committee on Solar Terrestrial Physics, International Council of Scientific Unions, p. 87.

Cheng, A. F., Barnouin-Jha, O., Prockter, L., Zuber, M. T., Neumann, G., Smith, D. E., Garvin, J., Robinson, M., Veverka, J., and Thomas, P., 2002. Small-scale topography of 433 Eros from laser altimetry and imaging. *Icarus*, **155**, 51, doi:10.1006/icar.2001.6750.

Chu, X., and Papen, G. C., 2005. Resonance fluorescence lidar for measurements of the middle and upper atmosphere. In Fujii, T., and Fukuchi, T. (eds.), *Laser Remote Sensing*. Boca Raton: Taylor & Francis, pp. 179–432, doi:10.1201/9781420030754.ch5.

Dehring, M. T., Tchoryk, P., and Wang, J., 2006. High altitude balloon-based wind LIDAR demonstration: from near space to space. *Proceedings of SPIE*, **6220**, 62200P, doi:10.1117/12.669262.

Durand, Y., Chinal, E., Endemann, M., Meynart, R., Reitebuch, O., and Treichel, R., 2006. ALADIN airborne demonstrator: a Doppler wind lidar to prepare ESA's ADM-Aeolus explorer mission. *Proceedings of SPIE*, **6296**, 62961D, doi:10.1117/12.680958.

Eloranta, E. W., 1998. Practical model for the calculation of multiply scattered lidar returns. *Applied Optics*, **37**, 2464, doi:10.1364/AO.37.002464.

Eloranta, E. W., 2005. High spectral resolution lidar. In Weitkamp, C., and Walther, H. (eds.), *Lidar: Range-Resolved Optical Remote Sensing of the Atmosphere*. Berlin: Springer, pp. 143–163, doi:10.1007/0-387-25101-4_5.

Emmitt, G. D., 2004. Combining direct and coherent detection for Doppler wind lidar. *Proceedings of SPIE*, **5575**, 31, doi:10.1117/12.576539.

Fujii, T., and Fukuchi, T. (eds.), 2005. *Laser Remote Sensing*. Boca Raton: Taylor & Francis. http://www.crcnetbase.com/isbn/9781420030754.

Gardner, C. S., 1989. Sodium resonance fluorescence lidar applications in atmospheric science and astronomy. *Proceedings of the IEEE*, **77**, 408, doi:10.1109/5.24127.

Garvin, J., Bufton, J., Blair, J. B., Harding, D., Luthcke, S., Frawley, J., and Rowlands, D., 1998. Observations of the Earth's topography from the shuttle laser altimeter (SLA): laser-pulse echo-recovery measurements of terrestrial surfaces. *Physics and Chemistry of the Earth*, **23**, 1053, doi:10.1016/S0079-1946(98)00145-1.

Gelbwachs, J. A., 1994. Iron Boltzmann factor lidar: proposed new remote sensing technique for mesospheric temperature. *Applied Optics*, **33**, 7151, doi:10.1364/AO.33.007151.

Gentry, B. M., Chen, H., and Li, S. X., 2000. Wind measurements with 355-nm molecular Doppler lidar. *Optics Letters*, **25**, 1231, doi:10.1364/OL.25.001231.

Gibert, F., Flamant, P. H., Bruneau, D., and Loth, C., 2006. Two-micrometer heterodyne differential absorption lidar measurements of the atmospheric CO_2 mixing ratio in the boundary layer. *Applied Optics*, **45**, 4448, doi:10.1364/AO.45.004448.

Gimmestad, G. G., 2008. Reexamination of depolarization in lidar measurements. *Applied Optics*, **47**, 3795, doi:10.1364/AO.47.003795.

Grant, W. B., Margolis, J. S., Brothers, A. M., and Tratt, D. M., 1987. CO_2 DIAL measurements of water vapor. *Applied Optics*, **26**, 3033, doi:10.1364/AO.26.003033.

Grant, W. B., Browell, E. V., Menzies, R. T., Sassen, K., She, C.-Y., and Thompson, B. J. (eds.), 1997. *Selected Papers on Laser Applications in Remote Sensing*. Bellingham: SPIE.

Grund, C. J., Banta, R. M., George, J. L., Howell, J. N., Post, M. J., Richter, R. A., and Weickmann, A. M., 2001. High-resolution Doppler lidar for boundary layer and cloud research. *Journal of Atmospheric and Oceanic Technology*, **18**, 376, doi:10.1175/1520-0426(2001)018<0376:HRDLFB>2.0.CO;2.

Hair, J. W., Hostetler, C. A., Cook, A. L., Harper, D. B., Ferrare, R. A., Mack, T. L., Welch, W., Izquierdo, L. R., and Hovis, F. E., 2008. Airborne high spectral resolution lidar for profiling aerosol optical properties. *Applied Optics*, **47**, 6734, doi:10.1364/AO.47.006734.

Hannon, S. M., Bagley, H. R., and Bogue, R. K., 1999. Airborne Doppler lidar turbulence detection: ACLAIM flight test results. *Proceedings of SPIE*, **3707**, 234, doi:10.1117/12.351378.

Heaps, W. S., and Burris, J., 1996. Airborne Raman lidar. *Applied Optics*, **35**, 7128, doi:10.1364/AO.35.007128.

Henderson, S. W., and Hannon, S. M., 2005. Advanced coherent lidar system for wind measurements. *Proceedings of SPIE*, **5887**, 58870I, doi:10.1117/12.620318.

Hoge, F. E., Lyon, P. E., Wright, C. W., Swift, R. N., and Yungel, J. K., 2005. Chlorophyll biomass in the global oceans: airborne lidar retrieval using fluorescence of both chlorophyll and chromophoric dissolved organic matter. *Applied Optics*, **44**, 2857, doi:10.1364/AO.44.002857.

Hu, Y., Stamnes, K., Vaughan, M., Pelon, J., Weimer, C., Wu, D., Cisewski, M., Sun, W., Yang, P., Lin, B., Omar, A., Flittner, D., Hostetler, C., Trepte, C., Winker, D., Gibson, G., and Santa-Maria, M., 2008. Sea surface wind speed estimation from space-based lidar measurements. *Atmospheric Chemistry and Physics*, **8**, 3593, doi:10.5194/acp-8-3593-2008.

Huffaker, R. M., and Hardesty, R. M., 1996. Remote sensing of atmospheric wind velocities using solid-state and CO_2 coherent laser systems. *Proceedings of the IEEE*, **84**, 181, doi:10.1109/5.482228.

Huffaker, R. M., Jelalian, A., and Thomson, J. A. L., 1970. Laser-Doppler system for detection of aircraft trailing vortices. *Proceedings of the IEEE*, **58**, 322, doi:10.1109/PROC.1970.7636.

Irgang, T. D., Hays, P. B., and Skinner, W. R., 2002. Two-channel direct-detection Doppler lidar employing a charge-coupled device as a detector. *Applied Optics*, **41**, 1145, doi:10.1364/AO.41.001145.

Kane, T. J., and Gardner, C. S., 1993. Structure and seasonal variability of the nighttime mesospheric Fe layer at midlatitudes. *Journal of Geophysical Research*, **98**, 16875, doi:10.1029/93JD01225.

Karpicz, R., Dementjev, A., Kuprionis, Z., Pakalnis, S., Westphal, R., Reuter, R., and Gulbinas, V., 2006. Oil spill fluorosensing lidar for inclined onshore or shipboard operation. *Applied Optics*, **45**, 6620, doi:10.1364/AO.45.006620.

Karr, T. J., 2003. Synthetic aperture ladar for planetary sensing. *Proceedings of SPIE*, **5151**, 44, doi:10.1117/12.505723.

Kasparian, J., Rodriguez, M., Méjean, G., Yu, J., Salmon, E., Wille, H., Bourayou, R., Frey, S., André, Y. B., Mysyrowicz, A., Sauerbrey, R., Wolf, J.-P., and Wöste, L., 2003. White light filaments for atmospheric analysis. *Science*, **301**, 61, doi:10.1126/science.1085020.

Kasparian, J., Bourayou, R., Frey, S., Luderer, J. C., Méjean, G., Rodriguez, M., Salmon, E., Wille, H., Yu, J., Wolf, J.-P., and Wöste, L., 2005. Femtosecond white-light lidar. In Fujii, T., and Fukuchi, T. (eds.), *Laser Remote Sensing*. Boca Raton: Taylor & Francis, pp. 37–62, doi:10.1201/9781420030754.ch2.

Koch, S. E., Dorian, P. B., Ferrare, R., Melfi, S. H., Skillman, W. C., and Whiteman, D., 1991. Structure of an internal bore and dissipating gravity current as revealed by Raman lidar. *Monthly Weather Review*, **119**, 857, doi:10.1175/1520-0493(1991)119<0857:SOAIBA>2.0.CO;2.

Koch, G. J., Barnes, B. W., Petros, M., Beyon, J. Y., Amzajerdian, F., Yu, J., Davis, R. E., Ismail, S., Vay, S., Kavaya, M. J., and Singh, U. N., 2004. Coherent differential absorption lidar measurements of CO_2. *Applied Optics*, **43**, 5092, doi:10.1364/AO.43.005092.

Kovalev, V. A., and Eichinger, W. E., 2004. *Elastic Lidar: Theory, Practice, and Analysis Methods*. Hoboken: Wiley, doi:10.1002/0471643173.

Lefsky, M., and McHale, M. R., 2008. Volume estimates of trees with complex architecture from terrestrial laser scanning. *Journal of Applied Remote Sensing*, **2**, 023521, doi:10.1117/1.2939008.

McManamon, P. F., 2012. Review of ladar: a historic, yet emerging, sensor technology with rich phenomenology. *Optical Engineering*, **51**, 060901, doi:10.1117/1.OE.51.6.060901.

Measures, R. M., 1992. *Laser Remote Sensing: Fundamentals and Applications*. New York: Krieger.

Menzies, R. T., 1976. Laser heterodyne detection techniques. In Hinkley, E. D. (ed.), *Laser Monitoring of the Atmosphere*. Berlin/Heidelberg: Springer, pp. 297–353, doi:10.1007/3-540-07743-X_21.

Menzies, R. T., and Tratt, D. M., 1994. Airborne CO_2 coherent lidar for measurements of atmospheric aerosol and cloud backscatter. *Applied Optics*, **33**, 5698, doi:10.1364/AO.33.005698.

Menzies, R. T., and Tratt, D. M., 2003. Differential laser absorption spectrometry for global profiling of tropospheric carbon dioxide: selection of optimum sounding frequencies for high-precision measurements. *Applied Optics*, **42**, 6569, doi:10.1364/AO.42.006569.

Menzies, R. T., Tratt, D. M., and Hunt, W. H., 1998. Lidar In-space technology experiment measurements of sea surface directional reflectance and the link to surface wind speed. *Applied Optics*, **37**, 5550, doi:10.1364/AO.37.005550.

Mukai, T., Abe, S., Hirata, N., Nakamura, R., Barnouin-Jha, O., Cheng, A., Mizuno, T., Hiraoka, K., Honda, T., and Demura, H., 2007. An overview of the LIDAR observations of asteroid 25143 Itokawa. *Advances in Space Research*, **40**, 187, doi:10.1016/j.asr.2007.04.075.

Müller, D., Wandinger, U., Althausen, D., Mattis, I., and Ansmann, A., 1998. Retrieval of physical particle properties from lidar

observations of extinction and backscatter at multiple wavelengths. *Applied Optics*, **37**, 2260, doi:10.1364/AO.37.002260.

Murdock, D. G., Stearns, S. V., Lines, R. T., Lenz, D., Brown, D. M., and Philbrick, C. R., 2008. Applications of real-world gas detection: airborne natural gas emission lidar (ANGEL) system. *Journal of Applied Remote Sensing*, **2**, 023518, doi:10.1117/1.2937078.

Nimelman, M., Tripp, J., Allen, A., Hiemstra, D. M., and McDonald, S. A., 2006. Spaceborne scanning lidar system (SSLS) upgrade path. *Proceedings of SPIE*, **6201**, 62011V, doi:10.1117/12.666187.

Olofson, K. F. G., Witt, G., and Pettersson, J. B. C., 2008. Bistatic lidar measurements of clouds in the nordic arctic region. *Applied Optics*, **47**, 4777, doi:10.1364/AO.47.004777.

Osche, G. R., and Young, D. S., 1996. Imaging laser radar in the near and far infrared. *Proceedings of the IEEE*, **84**, 103, doi:10.1109/5.482225.

Papen, G. C., Gardner, C. S., and Pfenninger, W. M., 1995. Analysis of a potassium lidar system for upper-atmospheric wind-temperature measurements. *Applied Optics*, **34**, 6950, doi:10.1364/AO.34.006950.

Platt, C. M. R., Abshire, N. L., and McNice, G. T., 1978. Some microphysical properties of an ice cloud from lidar observations of horizontally oriented crystals. *Journal of Applied Meteorology*, **17**, 1220, doi:10.1175/1520-0450(1978)017<1220:SMPOAI>2.0.CO;2.

Raizada, S., and Tepley, C. A., 2002. Iron Boltzmann lidar temperature and density observations from Arecibo – an initial comparison with other techniques. *Geophysical Research Letters*, **29**, 1560, doi:10.1029/2001GL014535.

Ricklin, J. C., and Tomlinson, P. G., 2005. Active imaging at DARPA. *Proceedings of SPIE*, **5895**, 589505, doi:10.1117/12.622572.

Rothermel, J., Cutten, D. R., Hardesty, R. M., Menzies, R. T., Howell, J. N., Johnson, S. C., Tratt, D. M., Olivier, L. D., and Banta, R. M., 1998. The multi-center airborne coherent atmospheric wind sensor. *Bulletin of the American Meteorological Society*, **79**, 581, doi:10.1175/1520-0477(1998)079<0581:TMCACA>2.0.CO;2.

Saito, Y., Saito, R., Kawahara, T. D., Nomura, A., and Takeda, S., 2000. Development and performance characteristics of laser-induced fluorescence imaging lidar for forestry applications. *Forest Ecology and Management*, **128**, 129, doi:10.1016/S0378-1127(99)00280-7.

Sassen, K., 1991. The polarization lidar technique for cloud research: a review and current assessment. *Bulletin of the American Meteorological Society*, **72**, 1848, doi:10.1175/1520-0477(1991)072<1848:TPLTFC>2.0.CO;2.

Sassen, K., 2007. Identifying atmospheric aerosols with polarization lidar. In Kim, Y. J., and Platt, U. (eds.), *Advanced Environmental Monitoring*. Berlin: Springer, pp. 136–142, doi:10.1007/978-1-4020-6364-0_10.

Schutz, B. E., Zwally, H. J., Shuman, C. A., Hancock, D., and DiMarzio, J. P., 2005. Overview of the ICESat mission. *Geophysical Research Letters*, **32**, L21S01, doi:10.1029/2005GL024009.

Sharma, S. K., Misra, A. K., and Sharma, B., 2005. Portable remote Raman system for monitoring hydrocarbon, gas hydrates and explosives in the environment. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, **61**, 2404, doi:10.1016/j.saa.2005.02.020.

Sjogren, W. L., and Wollenhaupt, W. R., 1973. Lunar shape via the apollo laser altimeter. *Science*, **179**, 275, doi:10.1126/science.179.4070.275.

Slatton, K. C., Carter, W. E., Shrestha, R. L., and Dietrich, W., 2007. Airborne laser swath mapping: achieving the resolution and accuracy required for geosurficial research. *Geophysical Research Letters*, **34**, L23S10, doi: 10.1029/2007GL031939.

Smith, D. E., Zuber, M. T., Neumann, G. A., and Lemoine, F. G., 1997. Topography of the Moon from the clementine lidar. *Journal of Geophysical Research*, **102**, 1591, doi:10.1029/96JE02940.

Smith, D. E., Zuber, M. T., Frey, H. V., Garvin, J. B., Head, J. W., Muhleman, D. O., Pettengill, G. H., Phillips, R. J., Solomon, S. C., Zwally, H. J., Banerdt, W. B., Duxbury, T. C., Golombek, M. P., Lemoine, F. G., Neumann, G. A., Rowlands, D. D., Aharonson, O., Ford, P. G., Ivanov, A. B., Johnson, C. L., McGovern, P. J., Abshire, J. B., Afzal, R. S., and Sun, X. L., 2001. Mars orbiter laser altimeter: experiment summary after the first year of global mapping of Mars. *Journal of Geophysical Research*, **106**, 23689, doi:10.1029/2000JE001364.

Smith, D. E., Zuber, M. T., Neumann, G. A., Lemoine, F. G., Mazarico, E., Torrence, M. H., McGarry, J. F., Rowlands, D. D., Head III, J. W., Duxbury, T. H., Aharonson, O., Lucey, P. G., Robinson, M. S., Barnouin, O.S., Cavanaugh, J. F., Sun, X., Liiva, P., Mao, D.-D., Smith, J. C., and Bartels, A. E., 2010. Initial observations from the lunar orbiter laser altimeter (LOLA). *Geophysical Research Letters*, **37**, L18204, doi:10.1029/2010GL043751.

Spiers, G. D., Menzies, R. T., Jacob, J., Christensen, L. E., Phillips, M. W., Choi, Y., and Browell, E. V., 2011. Atmospheric CO_2 measurements with a 2 μm airborne laser absorption spectrometer employing coherent detection. *Applied Optics*, **50**, 2098, doi:10.1364/AO.50.002098.

Spinhirne, J. D., Palm, S. P., Hart, W. D., Hlavka, D. L., and Welton, E. J., 2005. Cloud and aerosol measurements from GLAS: overview and initial results. *Geophysical Research Letters*, **32**, L22S03, doi:10.1029/2005GL023507.

Stoffelen, A., Pailleux, J., Källénc, E., Vaughan, J. M., Isaksen, L., Flamant, P., Wergen, W., Andersson, E., Schyberg, H., Culoma, A., Meynart, R., Endemann, M., and Ingmann, P., 2005. The atmospheric dynamics mission for global wind field measurement. *Bulletin of the American Meteorological Society*, **86**, 73, doi:10.1175/BAMS-86-1-73.

Takeuchi, N., Baba, H., Sakurai, K., and Ueno, T., 1986. Diode-laser random-modulation cw lidar. *Applied Optics*, **25**, 63, doi:10.1364/AO.25.000063.

Tratt, D. M., Whiteman, D. N., Demoz, B. B., Farley, R. W., and Wessel, J. E., 2005. Active Raman sounding of the Earth's water vapor field. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, **61**, 2335, doi:10.1016/J.SAA.2005.02.032.

Turner, D. D., and Whiteman, D. N., 2006. Remote Raman spectroscopy. Profiling water vapor and aerosols in the troposphere using Raman lidars. In Chalmers, J. M., and Griffiths, P. R. (eds.), *Handbook of Vibrational Spectroscopy*. Chichester: Wiley, doi:10.1002/0470027320.s6803.

Von Zahn, U., von Cossart, G., Fiedler, J., Fricke, K. H., Nelke, G., Baumgarten, G., Rees, D., Hauchecorne, A., and Adolfsen, K., 2000. The ALOMAR Rayleigh/Mie/Raman lidar: objectives, configuration, and performance. *Annales Geophysicae*, **18**, 815, doi:10.1007/s00585-000-0815-2.

Weibring, P., Edner, H., Svanberg, S., Cecchi, G., Pantani, L., Ferrara, R., and Caltabiano, T., 1998. Monitoring of volcanic sulphur dioxide emissions using differential absorption lidar (DIAL), differential optical absorption spectroscopy (DOAS) and correlation spectroscopy (COSPEC). *Applied Physics B*, **67**, 419, doi:10.1007/s003400050525.

Weibring, P., Johansson, T., Edner, H., Svanberg, S., Sundnér, B., Raimondi, V., Cecchi, G., and Pantini, L., 2001. Fluorescence lidar imaging of historical monuments. *Applied Optics*, **40**, 6111, doi:10.1364/AO.40.006111.

Weitkamp, C., and Walther, H. (eds.), 2005. *Lidar: Range-Resolved Optical Remote Sensing of the Atmosphere*. Berlin: Springer, doi:10.1007/b106786.

Werner, C., Flamant, P. H., Reitebuch, O., Köpp, F., Streicher, J., Rahm, S., Nagel, E., Klier, M., and Herrmann, H., 2001. Wind infrared Doppler lidar instrument. *Optical Engineering*, **40**, 115, doi:10.1117/1.1335530.
Wessel, J., Beck, S. M., Chan, Y. C., Farley, R. W., and Gelbwachs, J. A., 2000. Raman lidar calibration for the DMSP SSM/T-2 microwave water vapor sensor. *IEEE Transactions on Geoscience and Remote Sensing*, **38**, 141, doi:10.1109/36.823908.
Whiteway, J., Daly, M., Carswell, A., Duck, T., Dickinson, C., Komguem, L., and Cook, C., 2008. Lidar on the phoenix mission to Mars. *Journal of Geophysical Research*, **113**, E00A08, doi:10.1029/2007JE003002.
Winker, D. M., Couch, R. H., and McCormick, M. P., 1996. An overview of LITE: NASA's Lidar in-space technology experiment. *Proceedings of the IEEE*, **84**, 164, doi:10.1109/5.482227.
Winker, D. M., Hunt, W. H., and McGill, M. J., 2007. Initial performance assessment of CALIOP. *Geophysical Research Letters*, **34**, L19803, doi:10.1029/2007GL030135.
Yu, J., Mondelain, D., Ange, G., Volk, R., Niedermeier, S., Wolf, J. P., Kasparian, J., and Sauerbrey, R., 2001. Backward supercontinuum emission from a filament generated by ultrashort laser pulses in air. *Optics Letters*, **26**, 533, doi:10.1364/OL.26.000533.
Zhao, Y., Post, M. J., and Hardesty, R. M., 1990. Receiving efficiency of monostatic pulsed coherent lidars. 1: theory. *Applied Optics*, **29**, 4111, doi:10.1364/AO.29.004111.
Zhao, Y., Brewer, W. A., Eberhard, W. L., and Alvarez, R. J., 2002. Lidar measurement of ammonia concentrations and fluxes in a plume from a point source. *Journal of Atmospheric and Oceanic Technology*, **19**, 1928, doi:10.1175/1520-0426(2002)019<1928:LMOACA>2.0.CO;2.
Zuber, M. T., Smith, D. E., Solomon, S. C., Phillips, R. J., Peale, S. J., Head, J. W., Hauck, S. A., McNutt, R. L., Oberst, J., Neumann, G. A., Lemoine, F. G., Sun, X., Barnouin-Jha, O., and Harmon, J. K., 2008. Laser altimeter observations from MESSENGER's first Mercury flyby. *Science*, **321**, 77, doi:10.1126/science.1159086.

Cross-references

Cryosphere, Measurements and Applications
Ocean, Measurements and Applications
Optical/Infrared, Atmospheric Absorption/Transmission, and Media Spectral Properties
Optical/Infrared, Radiative Transfer
Optical/Infrared, Scattering by Aerosols and Hydrometeors
Radiation, Multiple Scattering
Radiation, Volume Scattering

EMERGING TECHNOLOGIES, RADAR

Alina Moussessian
Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

Definition

To respond to the wide range of goals in Earth and Planetary sciences, radar systems must operate with a wide range of functionality. Parameters include multipolarization transmit and receive signals, multiband operation, adaptable resolution, scanning for increased swath, and very high measurement accuracy. In addition onboard data processing techniques for data reduction and transmission and handling of large data sets are increasingly important as more complex, higher bandwidth radars are needed to meet mission requirements. Improved tools and support for warehousing and data mining are required to document the large number of missions and retain long duration data records required to perform key analysis, modeling, and monitoring.

Lowering mass and volume of radar systems is especially important for spaceborne missions both to lower spacecraft and launch vehicle costs and to enable combining multiple instruments such as radar/lidar on the same platform. Continued technology advancements focusing on risk and cost reduction are also critical.

Below are some of the technologies for future radar systems:

RF technologies

High-efficiency RF semiconductor devices are critical for future high-power systems. The use of wide bandgap semiconductors technologies such as GaN technology enables high-power, high-efficiency power amplifiers. For the most part, this technology can replace other devices, such as Si and GaAs, in power amplifier components.

In addition to semiconductor technology, some applications will still require the use of extended interaction klystron (EIK) and/or traveling-wave tubes (TWT). Further technology advancement in phase stable, higher peak power, higher efficiency, and longer life devices is needed.

Device technologies such as SiGe allow integration of RF and digital components onto the same chip enabling the concept of "radar on a chip." This is most beneficial for phased array applications where we need hundreds or thousands of these transmit/receive (TR) modules for electronic beam scanning.

In addition to high-efficiency, high-power RF transmitters, phase stable electronics both for receive and transmit are critical for future interferometry applications.

Quasi-optical techniques at millimeter-wave frequencies must be improved to support the front-end optics/electronics for millimeter-wave radars measuring crucial environmental parameters, such as clouds, aerosols, particulates, and other pollution.

Antenna technologies

In most cases the largest part of a radar is its antenna. Developing new materials and technologies for reducing antenna mass density both for passive and active antennas is critical. For reflector-type antennas, mesh technology is currently used at lower frequencies. Future radars will need larger mesh antennas at high frequencies (ka-band above). For these radars the development of multi-frequency shared feeds will allow sharing of resources for multiple measurements.

For very large active-phased arrays (size of 900 m^2 or greater), membrane technology is one way of reducing

mass density. Technology challenges in this area include integrating electronics with the array, manufacturing of these arrays, reliability, and the need for calibration to account for the antenna shape change.

MEMS technologies

MEMS resonators have the potential to incorporate low-loss front-end analog filtering into active antennas. This technology is becoming more important as the RF spectrum becomes more crowded, and digital processing is moving up the RF chain. As MEMS technology matures, other MEMS technologies such as MEMS switches might be usable in radar front-end electronics.

Digital technologies

Analog-to-digital converters (A/D) transform continuous analog radar signals into discrete digital signals. Extensive use of digital signal processing of radar data demands higher sampling rate A/D converters with larger dynamic range. Direct sampling of higher frequency signals, such as L and S band, requires faster A/D converters as well as higher dynamic ranges.

Due to further advancements in digital electronics, radars are using digital techniques in areas that RF technologies were previously used. One of these areas is the use of digital filtering, for example, for RFI mitigation. This allows the filtering of unwanted signals as the radar moves in space encountering different in-band interference. Another area where digital technology is replacing RF is digitizing higher frequency signals without having to downconvert the signal. For example, with current technology an L band (1.2 GHz) direct digital receiver is possible.

Another area which will dominate future radar applications is adaptive onboard science data processing. This along with development of a direct downlink capability provides near real-time data products tailored to users' needs.

Information system technology

We will need information technology advances for collecting, handling, and managing very large amounts of data and information in space and on the ground. These advanced information systems need to support new observations and information products; increase the accessibility and utility of science data; and reduce the risk, cost, size, and development time for future missions. These technologies are not unique to radar systems but apply to all sensors. However, radar systems typically collect large data which can take advantage of these processes. Technologies in this area include incorporating autonomy and intelligence within the sensing process and allowing rapid response to needed measurements, data acquisition on demand (in response to both science and society), sensor-to-sensor coordination, and interoperability. These technologies will increase the effectiveness of observing instruments or missions.

We also need technologies for combining observations of multiple Earth science variables into numerical models, data mining for information characteristics or content, dynamically acquiring and combining data from multiple data sources and distributed processing.

Constellations

Improved orbital control and tracking, along with satellite-to-satellite communication, is required to continue trends.

Acknowledgment

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

Bibliography

Kuhn, W., Mojarradi, M., and Moussessian, A., 2005. A resonant Switch for LNA protection in Watt-level CMOS Transceivers. *IEEE Transactions on Microwave Theory and Techniques*, **53**, 2819–2825.

Moussessian, A., Del Castillo, L., Huang, J., Sadowy, G., Hoffman, J., Smith, P., Hatake, T., Derksen, C., Lopez, B., and Caro, E., 2005. An active membrane phased array radar. In *IEEE MTT-S International Microwave Symposium*, Long Beach, CA.

Moussessian, A., Mojarradi, M., Johnson, T., Davis, J., Grigorian, E., Hoffman, J. P., Kuhn, W., and Caro, E., 2004. A completely integrated L- and S-band radar on a chip. Provisional Patent Application, NASA Case No. NPO 40869.

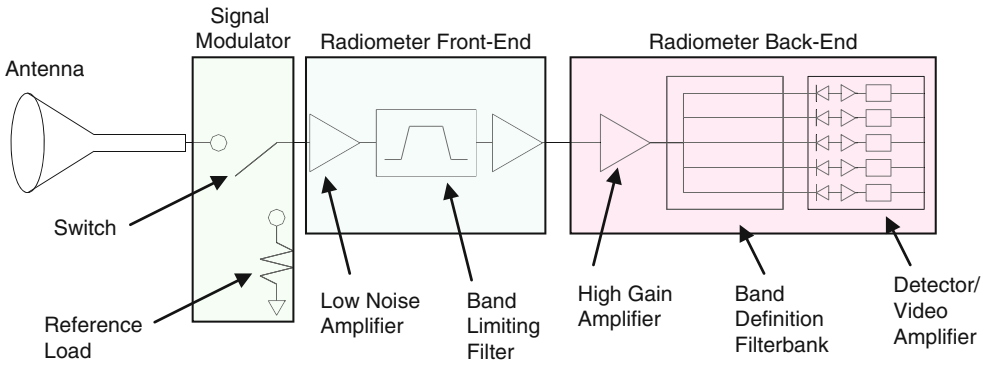
Reuss, R. H., Chalamala, B. R., Moussessian, A., et al., 2005. Macroelectronics: perspectives on technology and applications. *Proceedings of IEEE*, **93**, 1239–1256.

EMERGING TECHNOLOGIES, RADIOMETER

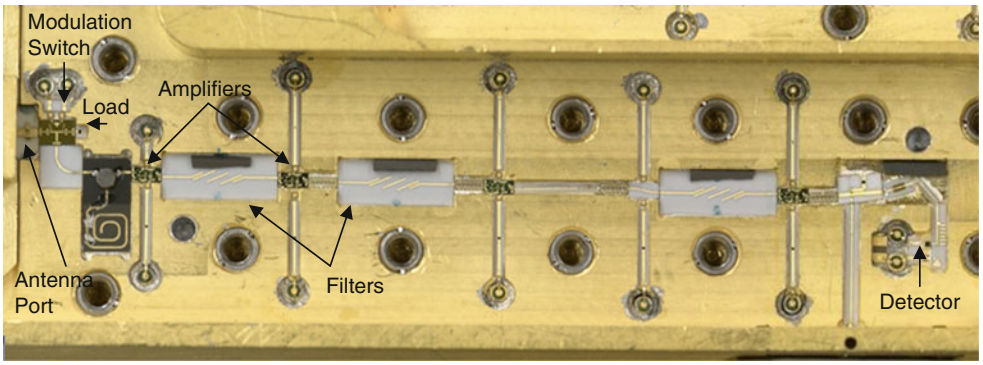
Todd Gaier
Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

Definition and introduction

Radiometric remote sensing involves the detection of emitted radiation and or reflected ambient radiation from an object. Such radiation typically originates as blackbody radiation from a remotely sensed object, undergoing partial reflection or absorption by intervening media. Detection is generally categorized into two types, *coherent detection* or *incoherent detection* (often referred to as direct detection). Coherent detection preserves the phase of the incoming wave, while incoherent detection does not. Examples of incoherent detectors include bolometers, thermopiles, CCD and APS sensors, as well as photoconductors and power detection diodes. Most passive remote sensing applications do not require phase preservation, although the advantages of phase-preserving detection make these systems the preferred choice for environmental applications in the frequency range



Emerging Technologies, Radiometer, Figure 1 Schematic of a basic radiometer suitable for remote sensing.



Emerging Technologies, Radiometer, Figure 2 Photograph of the Jason-II advanced microwave radiometer, 34 GHz channel, which shows all of the major components of a remote sensing radiometer (Courtesy D. Dawson).

1 GHz–1 THz. It is important to note that direct detectors do not have to impart additional noise to the incoming radiation, although practically speaking, cryogenic cooling of direct detectors is required to eliminate this added noise. When the observed object is very cold and the detected signals are very small (as in applications in radio astronomy), direct detection provides the lowest possible noise. The cooling requirement and the fact that most objects observed in remote sensing applications are already warm, resulting in a large inherent noise term, make these detectors less desirable than coherent detectors. For the rest of this entry, the discussion will be limited to coherent detection.

The basic coherent detection radiometer system can be divided into a front end and a back end. The front end amplifies the incoming signal (a voltage waveform) to a level where additional signal processing can occur without consideration of additional noise. The front end usually performs this function with the lowest possible additive noise available, but coherent systems always add some noise. The back end processes the amplified waveform extracting the desired information. Since most electronic components used in these systems exhibit inherent long-term instability, some sort of regular calibration or signal modulation must be applied to the radiometer. A schematic of a basic radiometer is shown in Figure 1. A photograph of an integrated radiometer is shown in Figure 2.

Radiometer front ends

The choice of front-end technologies depends upon the application and the frequency of operation. Below 100 GHz low-noise amplifiers are readily available. Most modern amplifiers use solid-state semiconductor, three-terminal devices (transistors) as the active low-noise element. The development of these transistors has a remarkable history which is driven largely by demands in the computer and telecommunications industries. Common transistors can be classified into two types: bipolar transistors and field-effect transistors (FETs). One can think of a bipolar transistor as a current-controlled current source (i.e., the input current between one terminal and the second controls the current between the second and third terminal). The action of the current control can be used to make a small signal larger, referred to as *signal gain*. A FET is a voltage-controlled current source (i.e., the *voltage* between the first and second

terminals determines the *current* between the second and third). At frequencies below a few GHz, bipolar transistors offer excellent noise and are extremely stable over time. These devices are typically made of doped silicon or sometimes (gallium arsenide) GaAs. Recently, the computer industry has pushed the practical frequency of operation of bipolar transistors well beyond 10 GHz. This is largely due to the development of SiGe transistors. SiGe transistors provide competitive gain and noise up to 10 GHz (Weinreb et al., 2007), at costs of pennies per transistor. This technology is also important for receiver back ends as will be discussed later.

The other primary front-end amplifier technology is a type of FET called a high-electron-mobility transistor (HEMT). HEMTs are devices in which the semiconductor crystal structure is grown one atomic layer at a time. This confines the current carriers (either electrons or "holes") to a two-dimensional surface, relatively free of scattering, providing a combination of high speed (with resulting high-frequency gain) and low noise (Duh et al., 1988).

A second key development in this field is the ability to create very small *gate* structures. The gate is the terminal of the FET which controls the current. In order to obtain very high switching speeds, the time for a carrier to transit the gate must be short. Electron-beam lithography allowed the creation of gate structures well under 1 μ in size (0.1 μ is readily achievable today). HEMTs using GaAs semiconductors have been used to build low-noise amplifiers for decades at frequencies up to 100 GHz. In the 1980s, HEMT technologies were employed in monolithic microwave integrated circuits (MMICs) in which several transistors were integrated onto single chips functioning amplifiers. During the 1990s, experiments were performed with different materials systems. Indium phosphide (InP) emerged as a materials system with very high gain, low noise, as well as low operating power. InP transistors and MMICs are now regularly used in systems up to 100 GHz with *noise temperatures* below 300 K (Pospieszalski, 2005; Lai et al., 2005).

Recent developments in this field have shown that it is still improving. Techniques have become available to reduce gate sizes as small as 35 nm. InP MMIC amplifiers have now been demonstrated at frequencies as high as 350 GHz (Pukala et al., 2008). These very high operational frequencies had been impossible for devices with gain. For remote sensing applications above 100 GHz, it has been typically required to utilize *heterodyne detection*. In a heterodyne receiver, the incoming signal is fed into a nonlinear device (usually a diode) where it is *mixed* with a sinusoidal voltage at almost the same frequency (local oscillator or LO frequency). A waveform results which has an amplitude proportional to the incoming signal but at a frequency which is the difference between the signal and LO frequencies. The heterodyne architecture is extremely useful if the observed object has high-spectral-resolution features, such as high-altitude atmospheric molecular line emission and absorption (Waters et al., 2006).

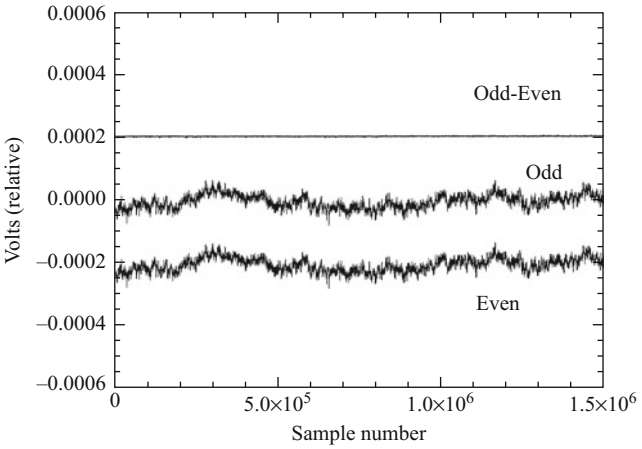
Diode mixers are the dominant radiometer front end between 100 GHz and 1 THz. Most of these diodes are Schottky diodes, formed when a metal and a semiconductor are brought into contact. Schottky diode mixers have been built up to several THz and have been used extensively in astronomy and atmospheric chemistry. The earliest mixers were built by manually pressing a metal whisker onto a piece of doped semiconductor. While these devices were used at very high frequencies, there were obvious reliability issues. Semiconductor and lithographic processes, similar to those used for transistors, are now used to make planar diode structures, also operating at THz frequencies. These planar devices can even be bought commercially. To obtain the lowest possible noise, mixers can be fabricated from superconducting devices called superconductor-insulator-superconductor or SIS mixers. While these devices have noise limited only by quantum mechanics, they require cryogenic cooling, making them impractical for many remote sensing applications. SIS mixers are widely used in radio astronomy (Pan et al., 1989).

Local oscillator technologies are a key part of heterodyne receiver front ends. Detection of signals at several hundred GHz requires oscillators working at these frequencies. For a long time, the primary device used was the Gunn-effect diode. Gunn diodes have a negative resistance, that is, as the voltage is decreased, the current increases. Such a device is inherently unstable. When coupled to a resonant cavity, this instability leads to oscillation, with the frequency of oscillation dependent upon the size of the cavity. Gunn oscillators capable of generating tens of megawatts have been built up to 160 GHz. To obtain even higher frequencies, diode frequency multipliers are employed.

Advances in power amplifier technology have enabled power generation at ever-increasing frequencies. GaAs power amplifiers are capable of generating hundreds of mW at frequencies above 100 GHz. These amplifiers have largely displaced Gunn oscillators as the primary source of LO power in mm-wave radiometers.

Calibration and modulation

The inherent gain instability of high-frequency circuits creates the need to stabilize the radiometer. The most common form of signal stabilization is the Dicke-switched radiometer (Dicke, 1946). The Dicke radiometer has as a first element, a modulation switch which rapidly exchanges the input signal with a stable reference load. The output of the radiometer is then a synchronous signal representing the remote sensed object and the reference load. Differencing of these signals on short timescales provides a calibration on a timescale shorter than the receiver gain can vary. A graphical depiction of this is shown in Figure 3. Many variations of modulated radiometers exist, including noise-adding radiometers (in which a noise source is added synchronously to provide gain calibration) and correlation radiometers (which allow for simultaneous



Emerging Technologies, Radiometer,
Figure 3 Pseudocorrelation Radiometer output. A time series (roughly 1 h) of data are shown with the radiometer alternately viewing a "remote sensed target" (*even*) and an internal reference load (*odd*). Each of the data sets displays long-term noise (due to gain instability). The differenced data are also shown, with considerably lower resulting noise and greatly improved stability. This demonstrates the power of signal modulation in a radiometer in removing time-correlated excess noise (Graphic courtesy M. Seiffert).

detection of two input signals through a common signal path). For each of these systems, signal modulators are required.

In order to be effective, switches used for Dicke modulation are usually in front of the first gain element. Historically, switches have been mechanical, ferrite, or semiconductor diode switches. Because of the high speeds and reliability required, mechanical switching is now limited to radiometers which can have the entire optical and receive system quickly scanned across the observed field of view. Ferrite switches use the time asymmetric properties of magnetic fields to switch signals between two ports of a three-port switch. The magnetic field of a circulating switch is typically provided by an electromagnet. Reversing the current in the electromagnet reverses the magnetic field which switches the microwave signal at a common port. In recent years, diode switches have been used for signal modulation. These switches take advantage of the conducting versus nonconducting states of a diode depending upon their *bias* or DC current. PIN (*p*-intrinsic-*n* doped semiconductor) diodes are particularly useful as switches as they are small, low loss (at low frequencies), support fast switch rates, and are repeatable and reliable with no moving parts.

PIN switches can also be used to make phase modulators. Phase modulators are useful in providing radiometer stability in a correlation (or pseudocorrelation) radiometer. In a correlation radiometer, couplers are used to provide a common signal path for incoming signals and the stable reference. While they do share a signal path, these signals must be separated prior to detection, which creates residual instability. Phase modulation stabilizes this residual by switching the remote sensed and reference signals prior to detection. The advantage of such a system is that the modulation element need not be located prior to the first gain stage, which in turn means that signal loss is not a primary concern. PIN diodes tend to have high loss at high frequencies. For this reason, high-frequency radiometers often use correlation techniques for stabilization (Jarosik et al., 2003).

The noise-adding radiometer is also used to stabilize a signal. While it seems counterintuitive to add noise to a receive system, it is an effective means of providing necessary gain stabilization. In this system, a noise source is coupled into the receiver input (with low efficiency) and is rapidly modulated. The noise source is typically a noise diode (similar to a Zener diode), which is operated alternately in the off state and in a state near avalanche. In avalanche, the diode carriers undergo ballistic collisions with the lattice to produce excess noise. This white noise looks like a blackbody of high temperature (typically thousands of degrees kelvin). By coupling in a small portion of the signal, a stable offset of tens of kelvin can be generated, which, while small compared to the receiver noise, is large enough to provide long-term gain stabilization. Because of the statistics of the noisy signal, gain adding alone is typically insufficient to provide stability and is typically used with some other form of signal modulation.

Radiometer back ends

The back end receives the amplified input signals and performs necessary processing. In the simplest form, the back end is a band definition filter and a power detector. Such a system measures the *radiometric temperature* in a band. Power detectors for remote sensing systems are almost always either Schottky barrier diodes or tunnel diodes. The diode rectifies the incoming signal with a DC voltage response that, to first order, is proportional to the square of the voltage or the power. In order to rectify such high-frequency signals, it is necessary for the diode to have a fast response. Because Schottky diodes can be fabricated with small features, they are typically used for detection applications at frequencies above 10 GHz. Limitations of the diodes include post-detection instability and significant temperature dependence. For this reason, tunnel diodes are the preferred choice below 10 GHz. There have been recent developments in high-frequency tunnel diodes (Moyer et al., 2008), although little data exist for their application in remote sensing. Detector diodes typically output signals at the 1 μV level in order to maintain linearity. The noise fluctuations about this signal are typically three to five *orders of magnitude smaller* than this, requiring low-noise *video* amplification (the term video is typically used to denote the post-detection signal).

Modern back-end processors take advantage of emerging digital technologies. Some of these back-end processors include *spectrometers*. A spectrometer takes the amplified input signal (or mixed signal) and processes the electromagnetic spectrum to determine the power in discrete frequency bands. As digital signal processing speeds improve, more technologies are becoming available for remote sensing applications. Digitizers with speeds faster than 2 GS/s are now commercially available. Processors also exist to handle these data rates. Field-programmable gate arrays (FPGAs) have been used to process these high-speed signals both in time and frequency domain. Spectrometers have been demonstrated using Fourier transform and polyphase filter bank technologies (Parsons et al., 2006).

The same technology can also be used to purge data of radiofrequency interference (RFI). RFI is becoming a greater problem for remote sensing applications in the frequency range 1–15 GHz as commercial communications technologies become widespread. Even with protected bands, contamination of the electromagnetic spectrum requires active monitoring and vetoing by current remote sensing experiments. Fortunately, the statistics of RFI signals differ from thermal signals, allowing for high-speed processing to flag and remove contaminated data (Ruf et al., 2006).

The final type of back-end digital processor is the cross correlator. The correlators multiply signals from one receiver with another. These signal products can be used in aperture synthesis interferometry. Aperture synthesis radiometry has become more widely used in recent years as large apertures without moving parts are becoming required to meet remote sensing requirements for soil moisture and ocean salinity measurements as well as continuous observations of atmospheric (SMOS, GeoSTAR).

Acknowledgment

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

Bibliography

Dicke, R. H., 1946. The measurement of thermal radiation at microwave frequencies. *The Review of Scientific Instruments*, **17**, 268–275.

Duh, K. H. G., Pospieszalski, M. W., Kopp, W. F., Ho, P., Jabra, A. A., Chao, P.-C., Smith, P. M., Lester, L. F., Ballingall, J. M., and Weinreb, S., 1988. Ultra-low-noise cryogenic high-electron-mobility transistors. *IEEE Transactions on Electron Devices*, **35**(3), 249–256.

Jarosik, N., et al., 2003. Design, implementation and testing of the MAP radiometers. *The Astrophysical Journal Supplement Series*, **145**, 413.

Lai, R., Grundbacher, R., Sawdai, D., Uyeda, J., Biedenbender, M., Barsky, M., Gutierrez-Aitken, A., Cavus, A., Chin, P., Liu, P. H., Bhorania, R., Streit, D., and Oki, A., 2005. Production InP MMICs for low cost, high performance applications. In *International Conference on Indium Phosphide and Related Materials*, May 8–12, 2005, pp. 598–602.

Moyer, H. P., Schulman, J. N., Lynch, J. J., Schaffner, J. H., Sokolich, M., Royter, Y., Bowen, R. L., McGuire, C. F., Hu, M., and Schmitz, A., 2008. W-band Sb-diode detector MMICs for passive millimeter wave imaging. *IEEE Microwave and Wireless Components Letters*, **18**(10), 686–688.

Pan, S.-K., Kerr, A. R., Feldman, M. J., Kleinsasser, A. W., Stasiak, J. W., Sandstrom, R. L., and Gallagher, W. J., 1989. An 85–116 GHz SIS receiver using inductively shunted edge junctions. *IEEE Transactions on Microwave Theory and Techniques*, **37**(3), 580–592.

Parsons, A., Backer, D., Werthimer, D., and Wright, M., 2006. A new approach to radio astronomy signal processing: packet switched, FPGA-based, upgradeable, modular hardware and reusable, platform-independent signal processing libraries. In *Proceedings USNC-URSI*, Boulder, CO, January 2006.

Pospieszalski, M. W., 2005. Extremely low-noise amplification with cryogenic FETs and HFETs: 1970–2004. *IEEE Microwave Magazine*, **6**(3), 62–75.

Pukala, D., Samoska, L., Gaier, T., Fung, A., Mei, X. B., Yoshida, W., Lee, J., Uyeda, J., Liu, P. H., Deal, W. R., Radisic, V., and Lai, R., 2008. Submillimeter-wave InP MMIC amplifiers from 300–345 GHz. *IEEE Microwave and Wireless Components Letters*, **18**(1), 61–63.

Ruf, C. S., Gross, S. M., and Misra, S., 2006. RFI detection and mitigation for microwave radiometry with an agile digital detector. *IEEE Transactions on Geoscience and Remote Sensing*, **44**(3), 694–706.

Waters, J. W., et al., 2006. The earth observing system microwave limb sounder (EOS MLS) on the aura satellite. *IEEE Transactions on Geoscience and Remote Sensing*, **44**(5), 1075–1092.

Weinreb, S., Bardin, J. C., and Mani, H., 2007. Design of cryogenic SiGe low-noise amplifiers. *IEEE Transactions on Microwave Theory and Techniques*, **55**(11), 2306–2312.

EMERGING TECHNOLOGIES, SENSOR WEB

Mahta Moghaddam[1], Agnelo Silva[1] and Mingyan Liu[2]
[1]Electrical Engineering – Electrophysics, University of Southern California, Los Angeles, CA, USA
[2]Electrical and Computer Engineering, University of Michigan, Ann Arbor, MI, USA

Synonyms

Environmental sensor network; ESN; Wireless sensor network; WSN

Definition

Sensor webs are networks of sensors that can operate in a coordinated fashion, for example, through centralized or decentralized control centers. The controller can command the sensor nodes to modify their measurement schedule or configuration in response to environmental factors or to achieve certain measurement goals. Sensor webs are envisioned as a means of providing spatially and temporally adaptive in situ measurements for validation of remote sensing observations.

Introduction

Satellite remote sensing often results in data and retrieved geophysical products with resolutions that are significantly coarser than the scale of variations of the phenomena they represent. As an example of a geophysical variable, soil moisture fields retrieved from satellite observations using microwave instruments have resolutions on the order of kilometers if not coarser. However, the soil moisture fields themselves have spatial dynamics at the scales of several meters (e.g., because of topographic and land-cover variations), hundreds of meters (e.g., because of land-cover and soil texture variations), and kilometers (e.g., because of precipitation). Therefore, the coarse-resolution retrievals at the satellite pixel scale may not accurately represent the true mean of the soil moisture field.

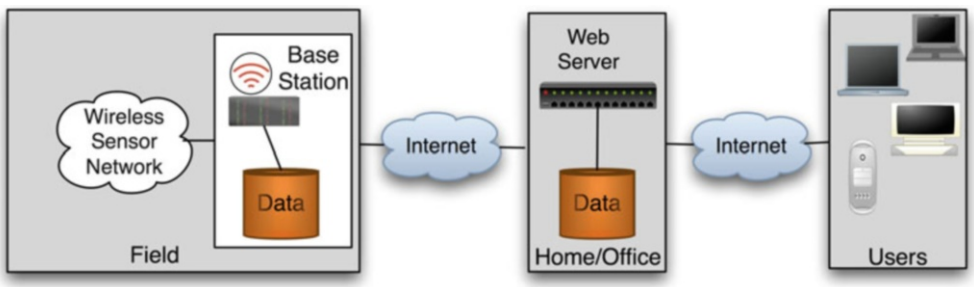
The validation of satellite retrievals is therefore a challenging task. It requires the use of in situ sensor networks, whose node placement has to be such that the proper spatial statistics are represented and such that the in situ estimate of the mean soil moisture within the coarse-resolution pixel is close to the true mean. Furthermore, the measurement schedule of the sensor nodes has to be such that the temporal variations of soil moisture are properly captured. At the same time, it has to be dynamic so that, if necessary, only the minimum number of measurements is taken so that energy usage is optimized by the network.

A large number of sensor nodes may be necessary to sufficiently represent the spatial variations of the coarse-resolution field. For example, for the NASA Soil Moisture Active and Passive (SMAP) satellite mission, scheduled for launch in 2014, the primary soil moisture product will have a resolution of 10 km (Entekhabi et al., 2010). Depending on the topographic, vegetation, and soil texture heterogeneity present in various pixels, tens or hundreds of sensors might be needed for proper retrieval validation. However, it is not feasible to deploy such large number of sensor nodes with conventional data loggers and manual data collection schemes. Instead, wireless network concepts have to be used that allow near real-time data upload. There are many challenges with such large-scale outdoor networks, including energy management, lifetime, environmental robustness, network capacity, and costs. Such challenges are illustrated by the following discussion involving the design of the sensor web for the SMAP mission.

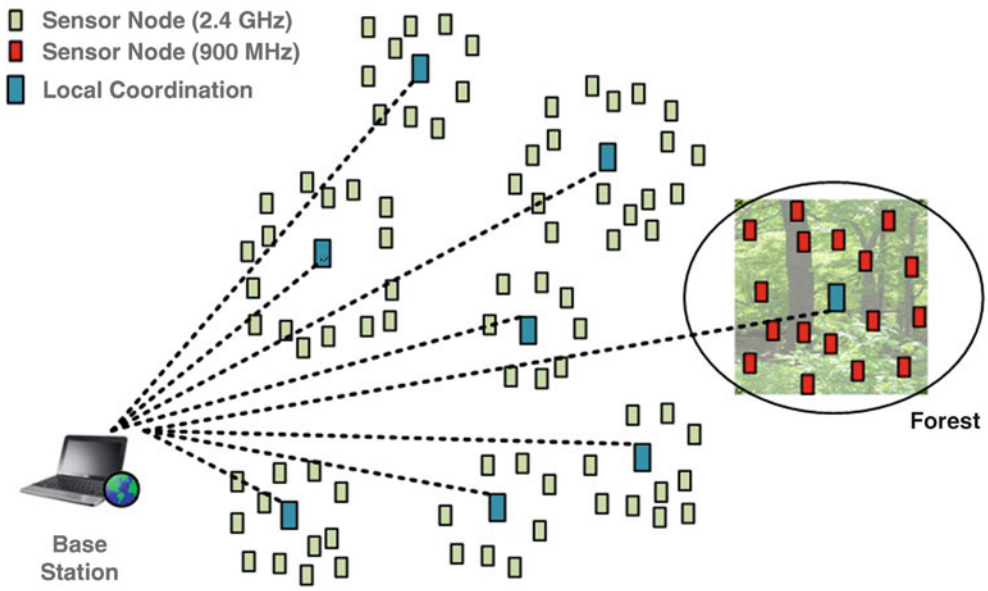
Global architecture of the sensor web

The global architecture of the in situ sensor network is shown in Figure 1. The system consists of a field element and a home/office element. A wireless sensor network is deployed over a target field, along with a base station that performs data collection and sensing control, and a database collocated with the base station for local data storage. At each sensing site, several sensors (for example, soil moisture sensors) are deployed at different depths and connected to the sensor node (i.e., a ground wireless module). The base station receives data wirelessly from each sensor node; it can also control their measurement schedules on demand. It also periodically (every half an hour) uploads the collected sensor data through a long-range link, such as a 3G connection, to a database server located in the home base.

Initially, this network was built using the ZigBee (IEEE 802.15.4 plus higher layer specifications) standard (Moghaddam et al., 2010). The first deployment of this network was in a field near Canton, Oklahoma. ZigBee was chosen as it allowed the formation of a multi-hop network and due to its mature technology that could significantly shorten the development and production cycle. The disadvantage was that a router node under the ZigBee specification could not be put to sleep mode, thus consuming significantly more energy and requiring large batteries and large solar panels. It also was proven to be rather unstable for outdoor field environment due to poor router-base station connection, causing end devices (or nodes) to switch parent–child association. Therefore, the architecture was modified to adopt a two-tiered hierarchy: the lower layer consists of a local coordinator (LC) node and multiple sensor nodes or end devices (ED) associated with the LC node. The upper layer consists of LC node(s) and a base station. In contrast to existing network routers, the LC node can sleep and its energy requirements are significantly reduced. The LC node may be equipped with two radio interfaces, allowing it to communicate within the two layers using different radio technologies, effectively making the two layers logically separated. The lower layer uses the IEEE 802.15.4 standard but not



Emerging Technologies, Sensor Web, Figure 1 Global architecture of the wireless sensor network system.



Emerging Technologies, Sensor Web, Figure 2 A large wireless sensor network may have multiple local coordinators servicing a variety of different landscape types. Choice of local transceivers should be transparent to the network for extensibility.

the ZigBee suite. The advantages of this design include (1) flexibility in developing an open protocol on top of 802.15.4 for the lower layer and multiple candidate solutions for the upper layer; (2) the logical separation between the two layers, making sleep scheduling of the ED nodes much easier to control; (3) a different radio solution for the upper layer, allowing the system to span over much longer distances; and (4) ease of scaling up of system architecture. Using this design, the network can be scaled up (Figure 2) to contain multiple local coordinators that can be deployed in multiple landscape types and to cover a span of several kilometers.

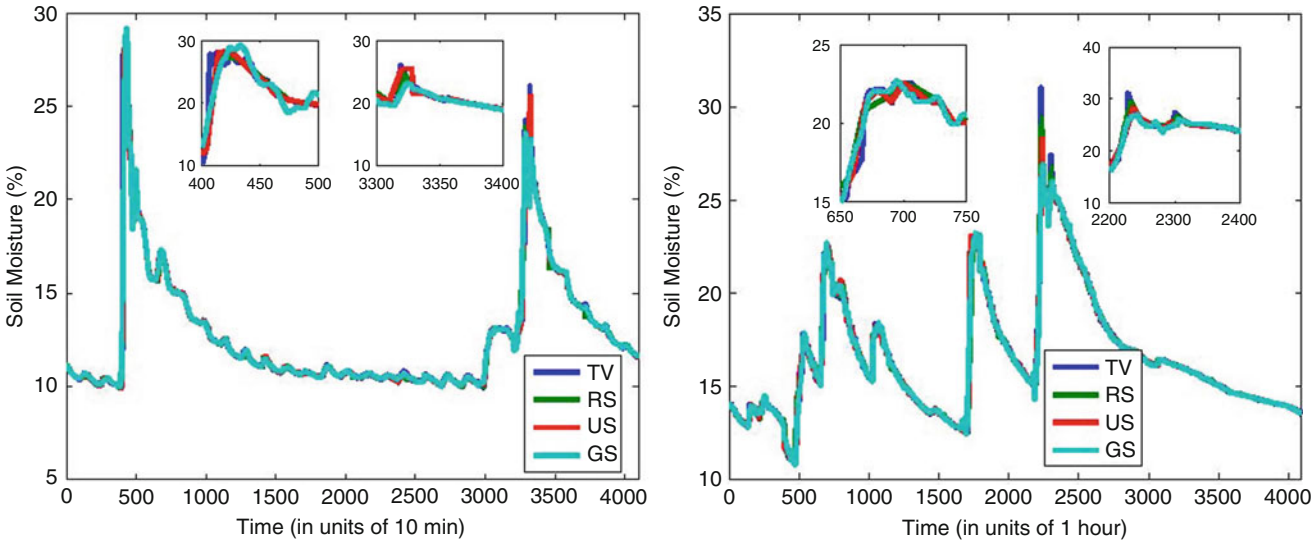
Sensor placement, scheduling, and field mean estimation

The true mean of geophysical variables such as soil moisture fields is a function of time and of the state of the soil surface. Its determination ideally would require a very fine sampling of the remote sensing satellite footprint, both spatially and temporally. This, however, is cost prohibitive; manually installing these sensors is expensive, and their battery power does not allow us to continuously sample, as we need them to last a reasonably long period of time (months or even years). These considerations pose severe limitations on how many sensors can be made available and how frequent they can be used/activated. The overall objective is thus to place and activate sensors such that the field mean may be estimated to a desired accuracy subject to budgetary constraints, e.g., the total number of sensors available, the total amount of available energy at each sensor, and bandwidth (Wu and Liu, 2012a).

There are two elements to the above problem: one is the determination of the best set of locations within the sensing field to place a limited number of sensors (sensor related cost constraint) and the other is the optimal dynamic operation of these sensors (when and which to activate) once they are placed (energy constraint).

These two elements are coupled. For instance, if energy of operation is a bigger concern than placement costs, then one can choose to place more sensors to compensate for a desired, reduced sampling rate. The reverse holds as well. In addition, activation and sampling decisions can influence where sensors should be placed and vice versa. But jointly considering and optimizing both elements leads to a problem whose complexity is prohibitive both analytically and computationally.

For a given placement plan, the sensor measurement scheduling problem becomes one that aims to minimize the estimation error (for the reconstructed soil moisture process using measured samples) subject to a certain energy/sampling rate constraint, or to minimize the sampling rate subject to an estimation accuracy criterion, or to minimize certain weighted sum of both. The estimation can be done both in a closed-loop fashion and an open-loop fashion. Under a closed-loop approach, temporal and spatial statistics of the soil moisture evolution is first learned using training data (either real or simulated). This knowledge, which describes what is likely to happen given what has already happened, together with samples already collected, i.e., what we know has happened, can help predict the future and thus make judicious decisions on the best time to take the next measurements. This is essentially the idea behind a partially observed Markov decision process (POMDP) formulation of this problem (Shuman et al., 2010).



Emerging Technologies, Sensor Web, Figure 3 Recovery accuracy of compressed sensing approach under three types of measurement schedules (US: uniform/periodic sampling; RS: random sampling; GS: Gaussian sampling; TV: true value).

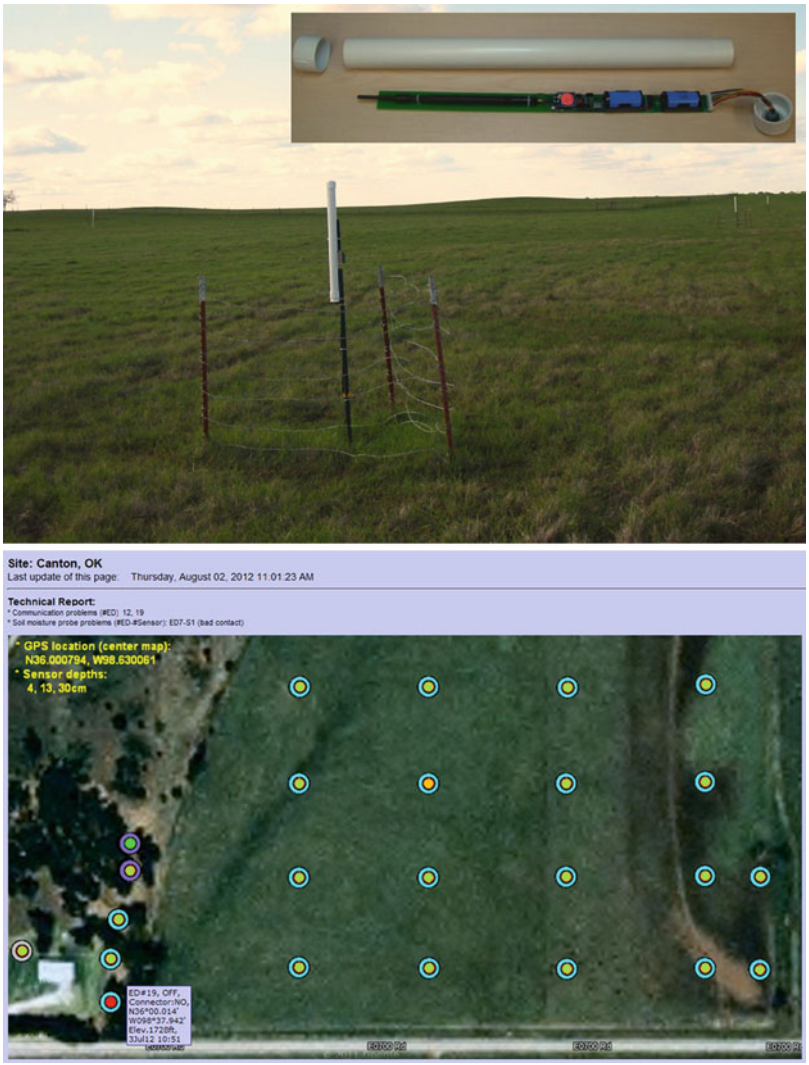
Under an open-loop approach, recent results from the theory of compressive sensing can be applied. This technique exploits an underlying sparsity feature of the measured signal and is able to reconstruct the soil moisture process from a very small number of samples to great accuracy. Advantages of this approach include the following: (1) it does not require training or a priori statistical knowledge of the soil moisture process, (2) the sampling sequence (measurement times) can be completely determined offline and in its simplest form can be a periodic sequence, therefore making implementation very easy, and (3) the same signal reconstruction technique can be used if we augment the sampling sequence with exogenous information like rainfall (e.g., take more samples during a rainfall event and less during dry periods). The figures below show the recovery accuracy of this approach under three types of measurement schedules (US, uniform/periodic sampling; RS, random sampling; GS, Gaussian sampling; TV, true value) (Wu and Liu, 2012b) (Figure 3).

Energy management schemes

Flexible energy management solutions are needed to meet the lifetime requirements of remote sensing missions such as SMAP. In doing so, increased reliability and reduced cost of wireless sensor network operation, extensibility to a large variety of target environments, and therefore an open and generalized architecture solution are needed.

Past experience has shown a significant number of battery failures caused by extreme temperatures as shown in Figure 4. Therefore, for many realistic deployment

Emerging Technologies, Sensor Web, Figure 4 Sensor nodes based on solar panels and rechargeable batteries typically have problems associated with sub-zero temperatures.



Emerging Technologies, Sensor Web, Figure 5 A Ripple-2 node powered by non-rechargeable batteries (*top*) installed in SoilSCAPE field locations such as Canton, OK, USA (*bottom*).

scenarios, the estimated lifetime of two or more years for a sensor node cannot be realized with rechargeable batteries because of their temperature sensitivity. Non-rechargeable batteries are the best option in this regard because of their tolerance to both high and low temperatures. If the usage of the network is consistently maintained low, such an approach could be very cost-effective and robust. For instance, assuming measurements every 15 min, the batteries need only be replaced every 2–4 years (depending on the model/type of the battery). No external part, such as a solar panel, is required to be exposed to the environment. This solution has been implemented in the "Ripple-2" system, developed by the authors (Figure 5) as part of the Soil moisture Sensing Controller And oPtimal Estimator (SoilSCAPE) project.

However, it is not always possible to maintain a low duty cycle. For instance, one could expect to periodically (and temporarily) modify the schedule of the sensor nodes to make measurements with the full node-density capacity of the network to optimize the scheduling and estimation processes. The main drawback in this case is the high energy cost. By adopting an energy-harvesting model, such as solar panels, one could compensate the additional loss of energy. Accordingly, two additional versions of Ripple-2 node are developed. The Ripple-2B is based on supercapacitors and a solar panel, and the Ripple-2C is a hybrid solution with supercapacitors and non-rechargeable batteries.

The usage of non-rechargeable batteries is not the preferred option for conventional sensor web solutions, in particular for outdoors. The main reason is the high cost involved in frequent battery exchanges (Weddell et al., 2008). Also, typically the energy consumption of the nodes is not uniform among them, and multiple

maintenance trips are necessary. Under the Ripple-2 development, the reasons behind this scenario are investigated, and it is concluded that cooperation among nodes is the main factor for the quick battery depletion. In other words, the average energy spent by a node *serving* the network node is multiple times the energy spent by a node to periodically take measurements and transmit data to the base station (Silva et al., 2012a).

The authors envisioned a scenario where the network is segmented into physical clusters and each segment has a data collector node (LC) that can communicate with all the sensor nodes in that segment by means of a single hop (direct link) (Silva et al., 2012b). By carefully assigning the location of the LC in relation to the sensor nodes of that area, the collaboration among sensor nodes due to multi-hopping is not necessary. As a result, the network overhead is drastically reduced from typical 3–10 % to less than 1 %. Also, a sensor node does not need to periodically wake up to *serve* the network (i.e., relaying messages). Therefore, the sleep scheduling of a Ripple-2 nodes is only governed by the application duty cycle. In other words, the energy consumption of a node has a very deterministic behavior, and energy balance among nodes is finally achieved no matter the size of the network.

Because a Ripple-2 node can potentially have a very long and continuous sleeping period (i.e., hibernation mode), it is possible to turn off many modules, such as the radio transceiver, rather than just put these modules into standby mode. The energy savings achieved by following this approach can be as high as 50 % (Silva et al., 2012a). Combined with the savings related to a reduced network overhead, the lifetime of a non-rechargeable battery of a Ripple-2 node can be realistically extended by multiple folds assuming a low duty-cycle regime, such as soil moisture measurements every 20 min (Menachem and Yamin, 2004).

However, there is an important trade-off in relation to non-rechargeable batteries: the *pulse current effect* which is the nonlinear and drastic energy capacity reduction of a battery when it is discharged by a slight higher current, such as 50 mA (Silva et al., 2012b). No matter how quick is this pulse current (e.g., a very fast radio transmission), the energy capacity of the battery is impacted and its lifetime can be as small as 10 % of the nominal/expected lifetime (Silva et al., 2012a). In order to solve this issue, Ripple-2 adopts supercapacitors as *power-matching* components. That is, these capacitors are slowly charged by a low current from the battery and quickly discharged by a high current (radio). Such approach protects the battery against the pulse current effect but also creates additional data latency issues when the node needs to transmit data very frequently. However, it is not a problem for a low duty-cycle web sensors used in SMAP.

Summary

To respond to the challenge of validation of coarse-scale remote sensing retrieval products such as soil moisture, a generalized wireless sensor network architecture has been developed. This architecture can be scaled to large-scale outdoor wireless sensor webs with flexible placement, scheduling, and power management schemes. The latest implementation of this architecture has been termed "Ripple-2" (with Ripple-1 being the first generation of this architecture). Due to its advantages, this architecture can be extended not just for soil moisture but for other sensing applications by making it flexible enough for other processor platforms and wireless technologies. Such applications can include any environmental monitoring application that has an extensive network deployed over a large area.

The high degree of robustness, energy efficiency, and reliability of Ripple-2 are achieved under the assumption of low duty cycles (e.g., sensor measurements every 10–20 min) and data latencies from seconds to minutes.

The Ripple-2 architecture can be considered a milestone in wireless sensor networks (WSNs) because of its specialization for low duty cycle and data-centric applications, breaking well-established concepts for WSNs. Without increasing the costs, the energy performance of Ripple-2 nodes is significantly superior compared to any similar WSN/telemetry solution. In fact, even non-rechargeable batteries can now be considered as a cost-effective option. However, technological enhancements can provide the path to turn Ripple-2 into a *generic* WSN solution.

Bibliography

Entekhabi, D., Njoku, E., O'Neill, P., Crow, W., Entin, J., Jackson, T., Johnson, J., Kimball, J., Koster, R., McDonald, K., Moghaddam, M., Moran, S., Reichle, R., Shi, J. C., and Tsang, L., 2010. The soil moisture active and passive (SMAP) mission. *Proceedings of the IEEE*, **98**(5), 704–716 (featured on cover).

Menachem, C., and Yamin, H., 2004. High-energy, high-power pulses plus battery for long-term applications. *Journal of Power Sources*, **136**, 268–275.

Moghaddam, M., Entekhabi, D., Goykhman, Y., Li, K., Liu, M., Mahajan, A., Nayyar, A., Shuman, D., and Teneketzis, D., 2010. A wireless soil moisture smart sensor web using physics-based optimal control: concept and initial demonstration. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **3**(4), 522–535.

Shuman, D., Nayyar, A., Mahajan, A., Goykhman, Y., Li, K., Liu, M., Teneketzis, D., Moghaddam, M., and Entekhabi, D., 2010. Measurement scheduling for soil moisture sensing: from physical models to optimal control. *Proceedings of the IEEE*, **98**(11), 1918–1933.

Silva, A. R., Liu, M., and Moghaddam, M., 2012a. Power Management Techniques for Wireless Sensor Networks and Similar Low-Power Communication Devices Based on Non-Rechargeable Batteries. *Journal of Computer Networks and Communications*, 2012 (article ID 757291), 10 pages, doi:10.1155/2012/757291.

Silva, A., Liu, M., and Moghaddam, M., 2012b. Ripple-2: A non-collaborative, asynchronous, and open architecture for highly-scalable and low duty-cycle WSNs. In *Proceedings ACM International Workshop on Mission-Oriented WSN (MiSeNet' 12)*. Istanbul.

Weddell, A. S., et al., 2008. Alternative energy sources for sensor nodes: rationalized design for long-term deployment. In *Proceedings IEEE International Instrumentation and Measurement Technology Conference (IMTC' 08)*. Victoria, British Columbia, pp. 1370–1375.
Wu, X., and M. Liu, 2012a. In-Situ Soil Moisture Sensing: Optimal Sensor Placement and Field Estimation. *ACM Transactions on Sensor networks*, vol. 8, no. 4, pp. 33:1–33:30.
Wu, X., and M. Liu, 2012b. In-situ soil moisture sensing: measurement scheduling and estimation using compressive sensing. *International Conference on Information Processing in Sensor Networks (IPSN)*. Beijing.

ENVIRONMENTAL TREATIES

Alexander de Sherbinin
Center for International Earth Science Information Network (CIESIN), Columbia University, Palisades, NY, USA

Synonyms

Multilateral Environmental Agreements (MEAs)

Definition

Environmental treaty: an environmental treaty is an agreement under international law entered into by two or more states addressing natural resource management or environmental protection. A treaty may also be known as an agreement, protocol, covenant, convention, exchange of letters, accord, exchange of notes, or a memorandum of understanding.

Introduction

Over the past 40 years, there has been a parallel increase in the number, range, and complexity of environmental agreements and the number and sophistication of remote sensing instruments. There were 172 environmental treaties in 1970, and that number has risen to more than 475 today (Mitchell, 2008), covering almost every conceivable environmental system. In terms of civilian or commercial remote sensing instruments, in 1970 there were only a handful of meteorological satellite sensors, whereas today there are more than 100 operational sensors with a wide range of spatial, spectral, and temporal resolutions.

Until the 1990s, the two communities – the international environmental law community, on the one hand, and remote sensing scientists, on the other – largely operated in isolation from one another. True, in the 1980s, images from the Total Ozone Mapping Spectrometer (TOMS) instrument onboard Nimbus-7 were used to document the seasonal depletion of ozone over the Antarctic, a visualization capability that significantly contributed to the Montreal Protocol of the Vienna Convention on Ozone Depleting Substances. But no one had asked the question: What kind of remote sensing data products are required by the environmental treaty community? In the late 1990s, the two communities began to assess – through a series of workshops and dialogues – the ways in which remote sensing could contribute to international environmental policy making. This entry addresses a number of the opportunities and constraints for the application of remote sensing data to environmental treaties that have been identified through recent dialogues and documented by research projects and publications (cf. CIESIN, 2001; de Sherbinin, 2005; Johnston, 2006; Strand et al., 2007).

Opportunities

One of the greatest attributes of remote sensing is its synoptic view with wall-to-wall coverage. This, coupled with the fact that the data are "objective" and consistent over time, at least compared to the highly variable in situ monitoring between countries and over time, makes remote sensing a potentially ideal source of data for a range of environmental issues of international concern. These include habitat loss, biodiversity conservation, desertification, transboundary air pollution transport, eutrophication of coastal waters, and greenhouse gas emissions from land-based sources, among others. The data are also applicable for regions of widely varying scales, from global down to transboundary resource management contexts.

Another advantage is that, unlike most other data gathering techniques, collection of remote sensing imagery does not infringe national sovereignty. Furthermore, it enables data collection in remote, inaccessible, or war-torn locations at much lower costs than ground-based methods and without risk to personnel.

In terms of treaty processes, remote sensing can be applied in a number of areas. Perhaps the most common use is in the context of environmental assessments. Both the Intergovernmental Panel on Climate Change (IPCC) and the Millennium Ecosystem Assessment relied heavily on remote sensing data to reach science-based conclusions about the scope of climate change and the loss of ecosystem services, respectively. Information from scientific assessments often feed directly into treaty negotiations, leading to new protocols or amendments to treaty texts. Another area of great potential is in national reporting, an often data-intensive process that can be greatly facilitated by remote sensing. For example, in the second national reports of the Convention on Biological Diversity, contracting parties were asked if they were using rapid assessment or remote sensing techniques. Thirteen parties reported a lot of use, 58 reported some use, and 33 reported no use or that they were exploring the possibility.

Perhaps the most important role of remote sensing is its support for the broader political process. Images capture the significance of a problem in a way that the written word does not. Remote sensing before and after images of deforestation or deglaciation has served to bolster public support for stronger environmental agreements

and the national laws that contribute to treaty implementation. Remote sensing has illustrated in powerful ways alterations to a wide range of ecosystems such as the Everglades in the USA, the Aral Sea, and the Amazon basin (UNEP, 2005). These same images can also serve as wake-up calls to political leaders. Political support at the highest levels for the Mesoamerican Biological Corridor, an agreement under the Central American Commission for Environment and Development (CCAD), can be partly attributed to a Landsat image of the Mexico-Guatemala border that portrayed the extent of deforestation in the Mexican state of Chiapas as compared to the neighboring Guatemalan Petén (de Sherbinin et al., 2002).

Traditionally, one of the problems of remote sensing was its high cost. The costs of imagery, software packages, computer processing, and trained technicians were beyond the reach of many developing countries and represented significant barriers even for many developed countries. Fortunately, this is beginning to change. In the case of imagery, costs have come down considerably for all except the highest resolution (1–4 m) commercial imagery. MODIS data are free of charge, Landsat imagery will soon be available at no cost, and data from many other instruments, such as ASTER, are freely available for scientific uses. In addition, there has been a proliferation of instruments with widely ranging characteristics – including radar instruments – that cover a much wider spectrum of the Earth and environmental processes than the few land and meteorological satellites of the 1980s. The growth in availability of imagery with different characteristics has inevitably brought costs down, even for commercial imagery.

The cost of commercial software packages is still fairly high, though a number of shareware packages are now available with some impressive capabilities. The American Museum of Natural History's Biodiversity Informatics web site lists 12 free software packages (AMNH, undated). There are others, such as TerraLook, which prepackages ASTER imagery for use in assessing land cover change in and around protected areas. Although they lack analytical capabilities, web-based applications such as Google Earth and NASA World Wind are popularizing remote sensing imagery in ways that can only enhance their application in the environmental policy realm.

Computer processing power has increased exponentially, and costs have come down in that domain as well. Probably, the only area that has not seen significant cost declines is the salaries for trained technicians. We return to the issue of training in the next section.

In terms of institutional support for the application of remote sensing to environmental treaties, a great stride forward has been made with the establishment of the Group on Earth Observations' (GEO) Global Earth Observation System of Systems (GEOSS). GEOSS is facilitating exactly the kind of dialogue between data users in the environmental policy realm and remote sensing scientists that began in the 1990s, and the dialogue is now being expanded to Internet technologists who are dreaming up new ways to combine remote sensing imagery with in situ data to better understand and manage the environment.

Constraints

As mentioned in the previous section, a number of former barriers to the use of remote sensing for international environmental policy – such as the high cost of imagery and software – have fallen. Nevertheless, there remain some technical hurdles that may never be fully overcome. One is the basic trade-off between spatial, spectral, and temporal resolution that is governed by basic laws of physics. High-resolution sensors of less than 5 m typically have narrow fields of view with small footprints and long repeat times, making them less useful for broadscale analyses. Though this could be considered a constraint, fortunately most environmental applications do not require such high-resolution imagery.

Another basic fact of life is that the data from passive sensors measure at sensor radiance, not what is actually on the ground. If space-based passive sensors were able to accurately, precisely, and repeatedly capture the actual reflectance from a feature on the ground, regardless of the time of day, season, or weather conditions, much of the hard work of image processing would be eliminated. But the reality is that the atmosphere scatters radiation that is reflected back out to space. Smoke, haze, clouds, and humidity exacerbate the problem and can block reflected energy entirely. This leads to gaps in the archival record. Radar or so-called "active" sensors can see through clouds, but the data are often difficult to interpret given the complicated reflectance patterns of different surface types. Where change detection is needed, there are further problems in discriminating actual on-the-ground changes from apparent changes due illumination, view angle, or seasonality that affects ground cover.

Another issue is the fact that many remote sensing instruments are experimental, not operational, and this impacts the continuity of the data streams. Many of the most widely used sensors, such as MODIS or SPOT VEGETATION, are from experimental programs and the sensor characteristics will not be replicated in future missions. Even some operational sensors such as Landsat, which was until recently the most widely used sensor for environmental applications, have had data gaps due to technical failures and delays in planning continuity missions.

Beyond the constantly varying data streams, as new experimental sensors come online or older sensors fail, standard algorithms for many environmental applications have yet to be developed. The conclusion of one survey of remote sensing methods applied to biodiversity assessment and conservation found that there is great promise for their application, especially as new remote sensing instruments become available but that the field lacks a set of standard methodologies that could move

remote sensing applications from an experimental to an operational stage of implementation (de Sherbinin, 2005). In the realm of international law, replicability of observations and the reliability and consistency of measurements are of paramount importance (GOFC-GOLD, 2012). Algorithms also need to be validated through field studies or "ground truthing" that can be many times more expensive than the initial image processing.

As mentioned in the previous section, training is still somewhat of a barrier to wider application of remote sensing in the treaty realm, and this affects two communities. The first is the science community, which needs to ensure that adequate training programs are developed at undergraduate and graduate levels that ensure that students are grounded in the basic science behind remote sensing so that scientific findings are robust. The second is the policy community. Without a basic literacy in the science behind remote sensing, many contracting parties and treaty secretariats lack the ability to critically evaluate scientific findings. This results in either the uncritical acceptance of findings which may in fact be wrong (which is probably the more likely scenario) or the discounting or rejection of findings because of a lack of understanding of what the imagery tells us. This suggests that there is a need for increased literacy among the policy community concerning the science upon which remote sensing is based so as to make informed decisions.

On the policy side, there are other constraints. One of the areas which remote sensing scientists have been keenest in promoting is the potential for remote sensing to be used for treaty monitoring and enforcement. Although the remote sensing does hold out the promise of more rigorous regimes based on real monitoring, the reality is more complex. Because most environmental treaties are considered "soft law" based on consensus approaches to environmental management, treaty secretariats and Parties have typically shied away from even suggesting the use of remote sensing in a compliance context. Until more environmental treaties are crafted in such a way as to have enforceable requirements, the full promise of monitoring and enforcement applications will not be met.

Still, there have been research projects exploring the potential of remote sensing for compliance enforcement among the few agreements with requirements. One was in the case of the Bonn Agreement, officially known as The Agreement for Cooperation in Dealing with Pollution of the North Sea by Oil and Other Harmful Substances (1983). In that case, researchers were exploring the use of SAR images to detect oil slicks on the high seas (Jones, 2001). It was found that SAR gave an unacceptably high level of false-positives, and even if it could identify oil slicks, establishing the connection between a slick and a vessel would need to be corroborated by photographic evidence according to the terms of the convention. Another research project in Bolivia sought to apply a combination of ground-based measurements and aerial videography to afforestation, reforestation, and deforestation (ARD) monitoring in the context of the Kyoto Protocol (see Noel Kempff Mercado Climate Action Project). Ultimately, to be applicable at broader scales, ARD monitoring will require a wider array of satellite sensors (including LIDAR) and more automated techniques than are currently available (Rosenqvist et al., 1999; CIESIN, 2001; GOFC-GOLD, 2012).

Other policy constraints include a lack of expertise on the part of treaty secretariats to assimilate remote sensing imagery or its results in their standard practices, and a fear by some, particularly developing country parties, of a "big brother" approach to compliance monitoring. Related to the former, it could be said that until recently there was a "technology push" on the part of space agencies and scientists rather than a genuine acceptance of remote sensing in the treaty world, lending the appearance of data developers in search of applications. An example was the European Space Agency's Treaty Enforcement Services using Earth Observation (TESEO) project, later subsumed under the Data User Element for MEAs. This technology push, however, was not necessarily a bad thing. TESEO was successful in engaging a number of treaty secretariats such as that of the Ramsar Convention on Wetlands and the Convention to Combat Desertification (CCD) and informing their staff. One impact metric is that references to remote sensing or satellite imagery in Ramsar Conference of Party (COP) decision documents increased from no references in all up until 1999 and then three references at COP7 (1999), three references at COP8 (2002), and four references at COP9 (2005). In the case of UNCCD, there were no references to either term until COP8 (2007) (SEDAC, 2010).

For its part, the US National Aeronautics and Space Administration (NASA) engaged the Convention on Biological Diversity (CBD) secretariat as well as a number of partners through its NASA-NGO Biodiversity Working Group. Working with the CBD and the World Conservation Monitoring Centre, the working group published the *Sourcebook on Remote Sensing and Biodiversity Indicators* (Strand et al., 2007), which provides a good balance of technical and applied information concerning the application of remote sensing that is useful to the CBD community.

Conclusion

In conclusion, although a number of technical and political obstacles remain, in all likelihood remote sensing will play an increasing role in international environmental politics, including in the treaty realm. Treaty-specific applications are likely to be enhanced as treaty secretariats or those familiar with treaty provisions actually participate in instrument design teams. The GEOSS 10 year action plan, established in 2005, makes reference to a number of application areas that are central to environmental treaties such as the UN Framework Convention on Climate Change, the CBD, the CCD, and Ramsar, including:

- Improving management of energy resources
- Understanding, assessing, predicting, mitigating, and adapting to climate variability and change
- Improving water resource management through better understanding of the water cycle
- Improving the management and protection of terrestrial, coastal, and marine ecosystems
- Supporting sustainable agriculture and combating desertification
- Understanding, monitoring, and conserving biodiversity

The document specifically states that GEOSS will "further the implementation of international environmental treaty obligations." An early workshop for the GEO Biodiversity Observation Network (Potsdam, Germany, April 2008) included significant representation of remote sensing scientists, and treaty secretariats have been engaged in the Network's development.

Although remote sensing is not a panacea for the data and information needs of environmental treaties, and though even the best data does not guarantee action on the growing list of environmental problems, the data do represent a significant resource that will undoubtedly contribute to environmental assessments and serve to strengthen treaty mechanisms well into the future.

Bibliography

American Museum of Natural History (AMNH), undated. *Biodiversity Informatics*. Available from World Wide Web: http://biodiversityinformatics.amnh.org/. Accessed June 24, 2008.

SEDAC (Socioeconomic Data and Applications Center), 2010. *Environmental Treaty and Resource Indicators (ENTRI) COP Decision Search Tool*. Available from World Wide Web: http://sedac.ciesin.columbia.edu/gsametasearch/cop_start.jsp. Accessed September 12, 2010.

Center for International Earth Science Information Network (CIESIN), Columbia University. 2001. *Remote Sensing and Environmental Treaties: Building More Effective Linkages*. Report of a workshop held 4–5 December 2000 at the Woodrow Wilson Center, Washington, DC. Available from World Wide Web: http://sedac.ciesin.columbia.edu/rs-treaties/. Accessed July 2, 2008.

de Sherbinin, A., 2005. *Remote Sensing in Support of Ecosystem Management Treaties and Transboundary Conservation*. Report prepared for the project on Remote Sensing Technologies for Ecosystem Management Treaties funded by the United States Department of State, Bureau of Oceans and International Environmental and Scientific Affairs initiatives program (OESI). Available from World Wide Web: http://sedac.ciesin.columbia.edu/rs-treaties/RS&EMTreaties_Nov05_screen.pdf. Accessed June 26, 2008.

de Sherbinin, A., Kline, K., and Raustiala, K., 2002. Remote sensing data: valuable support for environmental treaties. *Environment*, **44**(1), 20–31.

GOFC-GOLD, 2012. *A Sourcebook of Methods and Procedures for Monitoring and Reporting Anthropogenic Greenhouse Gas Emissions and Removals Caused by Deforestation, Gains and Losses of Carbon Stocks in Forests Remaining Forests, and Forestation*. Available at http://www.gofcgold.wur.nl/redd/sourcebook/GOFC-GOLD_Sourcebook.pdf. Accessed June 19, 2012.

Johnston, S., 2006. *Space in Environmental Diplomacy: Exploring the Role of Earth Observing Satellites for Monitoring International Environmental Agreements*. PhD thesis, The George Washington University. Available from http://search.proquest.com/docview/305336614/abstract. Accessed June 7, 2013.

Jones, B., 2001. A comparison of visual observations of surface oil with synthetic aperture radar imagery of the sea empress oil spill. *International Journal of Remote Sensing*, **22**(9), 1619–1638.

Mitchell, R., 2008. *International Environmental Agreements Database Project (Version 2007.1)*. Available from World Wide Web: http://iea.uoregon.edu/. Accessed July 2, 2008.

Noel Kempff Mercado Climate Action Project web site. Available from World Wide Web: http://www.forestcarbonportal.com/project/noel-kempff-mercado-climate-action-project. Accessed June 7, 2013.

Rosenqvist, A., Imhoff, M., Milne, A., and Dobson, C., 1999. *Remote Sensing and the Kyoto Protocol: A Review of Available and Future Technology for Monitoring Treaty Compliance*. Report of an ISPRS workshop, October 20–22, 1999, Ann Arbor, MI. Available from http://sedac.ciesin.columbia.edu/rs-treaties/Remote_Sensing_Kyoto.pdf. Accessed June 7, 2013.

Strand, H., Hoft, R., Strittholt, J., Miles, L., Horning, N., and Fosnight, E. (eds.), 2007. *Sourcebook on Remote Sensing and Biodiversity Indicators*. CBD Technical Series No. 32. Montreal: Convention on Biological Diversity. Available from World Wide Web: http://www.cbd.int/doc/publications/cbd-ts-32.pdf. Accessed June 26, 2008.

United Nations Environment Programme (UNEP), 2005. *One Planet Many People: Atlas of Our Changing Environment*. Nairobi: UNEP/UNEP Division of Early Warning and Assessment.

Cross-references

Air Pollution
Atmospheric General Circulation Models
Climate Monitoring and Prediction
Coastal Ecosystems
Fisheries
Global Climate Observing System
Global Earth Observation System of Systems (GEOSS)
Global Land Observing System
Rainfall
Rangelands and Grazing
Water Resources
Wetlands