BROOKINGS

QUALITY. INDEPENDENCE. IMPACT.

The Technical Basis for Spectrum Rights: Policies to Enhance Market Efficiency

March 3, 2011

Robert Matheson
Retired U.S. Department of Commerce/NTIA[1]

Adele C. Morris
The Brookings Institution[2]

The views expressed in this paper are the views of the authors and do not necessarily represent the views of the U.S Department of Commerce or Brookings.

[1] Contact: rjmatheson@att.net, 303-494-7590
[2] Contact: amorris@brookings.edu

Table of Contents

EXECUTIVE SUMMARY

The inefficiencies inherent in the traditional command-and-control spectrum regulatory system are increasingly costly as demand for spectrum-dependent services explodes. This paper describes a conceptual framework to articulate clear rights of access to spectrum in a way that fosters a market-based allocation of the resource. We also offer simple rules that reasonably account for imperfect receivers and challenging physical properties of radiowaves. The key features of the system we propose are:

- Regulators construct an initial partition of spectrum rights across the dimensions of space, time, frequency, and direction of propagation. Each partition is called a licensed electrospace right (LER). Regulators devolve these rights to LER owners.

- Licensees may buy, sell, aggregate, and subdivide their LERs at will.

- Licensees must keep all signals within their respective LER, including its frequency band, geographical area, angle of propagation range, and authorized time of operation. In particular, all signals must have a power level of less than a regulated limit (E_0) outside the LER, with exceptions allowed with a probability no greater than an amount specified by regulators (such as one percent).

- Licensees must limit transmitter power or field strength within their LER to below a regulator-set level for the band in which they operate (E_{max}).

- Regulators or other parties must establish and maintain a detailed database and propagation model that facilitates transactions and enforcement.

In this system, regulators set up the rights database and establish a few core parameters for each band. Thereafter their role is limited to enforcing compliance with the simple set of rules on signal strength. Importantly, this system includes no protection of, or constraints on, receivers, so it does not directly control interference. Rather, through transactions and negotiations between LER owners, the system we outline here would induce an efficient level of interference in which the costs of controlling interference are balanced by the benefits.

1. INTRODUCTION

In the United States, the federal government has traditionally regulated access to radio spectrum with "command-and-control" licensing. Regulators divide the radio spectrum into frequency ranges, or bands, and typically allocate each band for a single type of radio service with very specific rules about its operation. These rules define the system of rights for licensees to provide the specified services for each band.

This traditional command-and-control spectrum regulation has sought to optimize the technical rules for different radio services, generally with good success. However, rapidly evolving technology and increasing demands for wireless services mean this rigid regulatory structure is increasingly poorly suited to optimizing the economic efficiency of the allocation of frequencies across different kinds of radio services. That is because the spectrum rights conveyed by a traditional license are usually exactly sufficient to provide only the services the regulator intends. The licensee can transmit with a specified power, at a specific location, using a specific antenna and tower, employing a specific modulation and bandwidth as needed to provide the specified service. This has generally provided a good "recipe" for operating the intended service with an acceptable quality, but it relies heavily on the regulator, rather than the market, to determine which spectrum resources are used for which applications. This fails to accommodate new services and applications, particularly for cell phones and other wideband wireless services, and it fails to provide efficient incentives to develop and deploy new technology. In most cases, to accommodate new uses regulators have moved the existing users of allocated bands to other suitable bands. Because of the time and expense of such reallocations, regulators need other ways to accommodate a wide range of services without requiring reallocation to each specific new use.

This paper offers a conceptual way forward. It explores the technical fundamentals of establishing rights to access spectrum, including the institutional, scientific, and engineering considerations important to policymakers. It describes how regulators could articulate rights to spectrum access such that rights holders could transfer, subdivide, aggregate, and protect their rights in an economically efficient market that accommodates evolving demands for the resource. We examine current approaches to expressing rights to access spectrum, their advantages and disadvantages, and how they may lead to economically inefficient underutilization of spectrum resources.

The technical basis for spectrum rights, meaning the way the regulatory system articulates rights to access spectrum, is key to achieving an efficient allocation of the resource through markets. If spectrum resource owners can subdivide or transfer their rights in a competitive market, then they perceive the full opportunity cost of holding the resource as revealed by market prices for spectrum rights. In an efficient spectrum market, incumbents would have an incentive to adopt technologies that optimize their use of the resource and to devolve underutilized spectrum to others with higher and better uses. Thus, in a flexible rights regime, market forces and available technologies would determine the efficient degree of partition of rights in any particular band or application, and at the same time market forces would induce investors to develop new technologies that use highly valued spectrum more efficiently and exploit low cost spectrum for new applications.

The paper proceeds as follows. Section 2 outlines the basics of band allocations, frequency assignments, and spectrum licenses. This provides the foundation to examine the details of exclusive spectrum rights and assess the advantages and disadvantages of the current system with an eye toward identifying approaches that could work better. Section 3 describes the physics of radio signal propagation that underlie any spectrum applications and introduces the seven-dimension "electrospace" approach to describing radio signals and the rights to emit them. We argue that increased exploitation of these dimensions will be central in improving spectrum capacity.

Section 4 describes the traditional command-and-control approach to regulating airwaves that is the basis for most management of radio use in the United States and the rest of the world, and Section 5 briefly discusses several frequency management alternatives to command-and-control, including low-power commons and some opportunistic dynamic spectrum sharing techniques. Section 6 describes how the electrospace approach can be the technical basis of flexible-use spectrum rights. We present a way to express the rights to use spectrum that is not tied to any specific service or technology, allowing market forces to allocate spectrum such that new radio technologies and applications can be rapidly accommodated with minimal regulatory oversight. Section 7 describes some challenges with selecting and enforcing exact rules to regulate flexible-use bands. We describe relatively simple rules that can simultaneously prevent interference and allow substantial flexibility in use, and we discuss what we see as the most promising applications flexible-use frequency bands. Section 8 concludes.

2. ALLOCATION AND LICENSING OF SPECTRUM

In the U.S. and most other countries, regulatory authorities parcel out radio frequencies to users in a two-stage process.[3] The first stage is *allocation*, whereby regulators divide the radio spectrum into frequency bands of differing rules. The rules in a band typically specify what services users can provide and the technical and operational parameters that apply to those services. Figure 1 below shows the 2003 frequency allocation chart that applies to the United States.[4] It shows that every frequency between 9 kHz and 300 GHz falls into a frequency band and illustrates the extensive partitioning of the radio regulatory system.

The executive branch of the U.S. government, through the Department of Commerce's National Telecommunications and Information Administration (NTIA), manages some bands for the benefit of federal agencies. Federal missions that require spectrum include law enforcement (e.g., the FBI, Coast Guard, and Secret Service), national infrastructure management (e.g., air traffic control by the Federal Aviation Administration, the Global Positioning System, and

[3] Details of the U.S. regulations for the Federal Communications Commission can be found in Part 47 in the Code of Federal Regulations (47 CFR). Similar regulations for the National Telecommunications and Information Administration (NTIA) can be found in the *Manual of Regulations and Procedures for Federal Radio Frequency Management*, available here: http://www.ntia.doc.gov/osmhome/redbook/redbook.html.
[4] This chart is available at: http://www.ntia.doc.gov/osmhome/allochrt.pdf.

weather forecasting by the National Weather Service), and the safety of life and property (e.g., forest fire monitoring and patrolling of federal lands). A large amount of NTIA-managed spectrum is used by the military.

The Federal Communications Commission (FCC), an independent federal agency, manages the spectrum for all non-federal uses, including state and local government missions, broadcasting, private sector uses, and unlicensed applications (e.g. Wi-Fi and cordless phones). Finally, a few frequency bands are managed jointly by the FCC and NTIA.

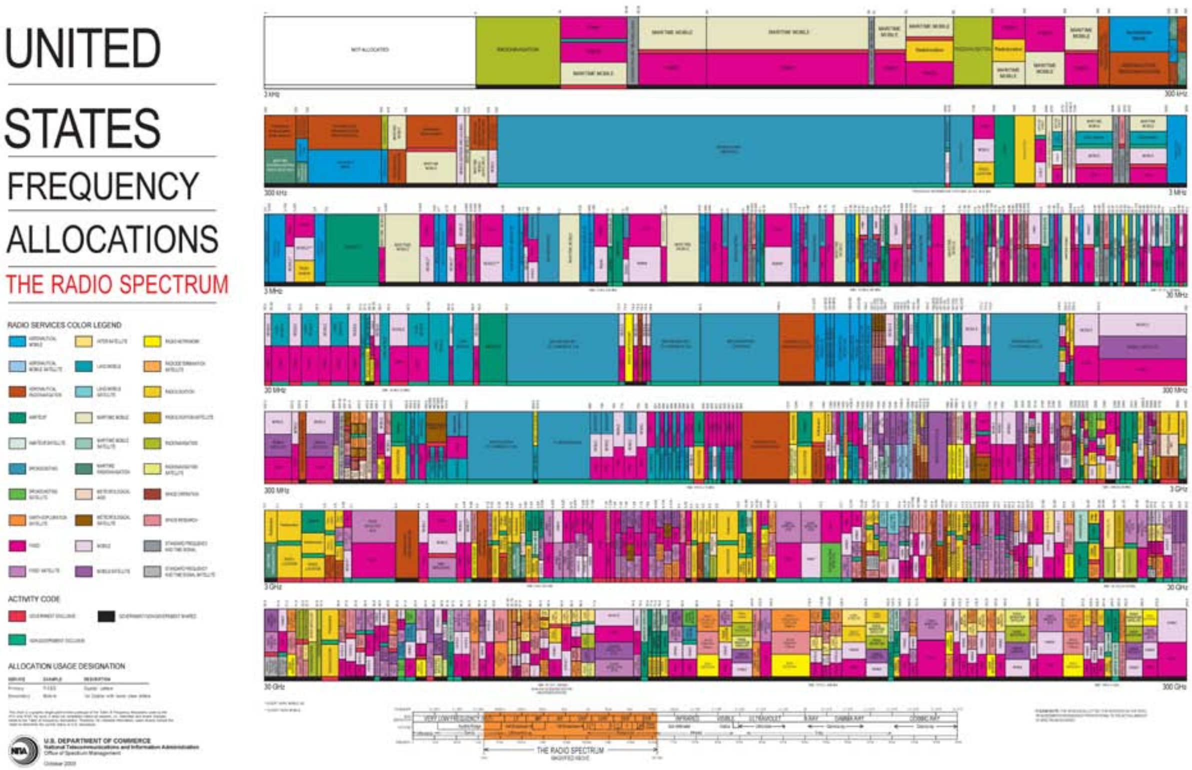

Figure 1. U.S. Frequency Allocation Chart

Some spectrum use, both federal and non-federal, is governed at least in part by international agreements. The U.S. makes two types of radio agreements with other countries. With immediate neighbors (chiefly Canada and Mexico), bilateral treaties describe how frequencies are to be shared in the areas adjacent to shared national borders. For example, the Canadian government must not issue spectrum access rights near the border with the U.S. that conflict with rights on U.S. soil, and vice versa. More broadly, the International Telecommunications Union (ITU) of the United Nations coordinates the use of radio frequencies on a worldwide basis. The ITU standardizes band allocations within three different regions; the U.S. is in Region 2, which includes North and South America. The ITU regulates frequencies only in bands where the usage is inherently international. Such bands include many frequencies below 20 MHz (where ionospheric reflections carry signals around the world) and frequencies used by satellites.

The allocation table in Figure 1 does not specify which individuals may use a given frequency at a given location. That determination occurs in the second stage of governance, through *licenses* and *assignments*. NTIA uses the term “assignment” to denote the rights to access radio spectrum given to federal users, while the FCC uses the term “license” to denote such rights for non-federal users. In this paper, we will use the term “license” to mean either license or assignment. We will use the term “exclusive rights” to denote the rights contained in either a license or an assignment.

The term “exclusive rights” is importantly distinct from “exclusive use.” Indeed, spectrum that is licensed exclusively to a particular entity, such as a wireless communications provider, may be used by a great number of people. We use the term “exclusive rights” here to mean the legal right to exclude users who are not authorized by the spectrum rights holder. Exclusive rights stand in contrast to unlicensed spectrum, which can be used by anyone whose signals conform to the relevant rules for power levels and other technical parameters.[5]

3. PHYSICS OF SPECTRUM USE

This section describes the physical properties of electromagnetic radio waves and explains the “electrospace” approach, a way of expressing the quantitative presence of those waves in the environment. We believe this approach is the best foundation for a flexible and technologically neutral articulation of exclusive spectrum rights.

Immutable laws of physics govern the opportunities and limitations for all current and future spectrum uses, so this section seeks to make clear which constraints to spectrum use derive from radio physics and which constraints derive only from the existing regulatory structure or current technology. This distinction is critical to evaluating the current policy framework for spectrum management and identifying beneficial and practical options for improvement.

3.1. Electrospace Description of Spectrum

The term “spectrum” is used colloquially to mean several things, including a given frequency, a frequency band, or a set of rights to access a set of frequencies at a given time and location. In this paper, we use the alternative term “electrospace” to express the full potential for the extent and coverage of radio signals through frequency, time, space, direction, and other dimensions.[6] The electrospace describes the radio field strength at a given electrospace “location,” defined by

---

[5] A system of “opportunistic” rights would allow certain kinds of non-harmful incursion into otherwise exclusive rights. In this case, an “exclusive right” becomes the right to exclude all use of the licensed spectrum that is inconsistent with the rules for opportunistic access.

[6] R. J. Matheson, “The electrospace model as a tool for spectrum management,” addendum to the Proceedings of the International Symposium on Advanced Radio Technologies, March 2003, NTIA Special Publication SP-03-401.

a number N of independent electrospace dimensions. This means the electrospace represents an N-dimensional hyperspace.

The appropriate value of N depends on the number of characteristics of radio signals that current feasible receiver technologies can reasonably process independently from one another. Table 1 shows seven key characteristics that today's technology can usefully exploit, suggesting that at least for now, a seven-dimensional hyperspace could be sufficient for articulating a useful bundle of spectrum rights.

Table 1. Electrospace Dimensions

QUANTITY	UNITS	# OF DIMENSIONS
Frequency	kHz, MHz, or GHz	1
Time	seconds, hours, or years	1
Spatial location	latitude, longitude, altitude	3
Direction-of-travel	azimuth, elevation angle	2

We argue in this paper that the electrospace approach is particularly useful for describing flexible, exhaustive, exclusive spectrum rights because it provides a straightforward and unique basis for specifying exact regions of spectrum (called "electrospace volumes") to which an individual can hold rights of access. That is not to say that any particular rights holder would wish to subdivide his or her rights along every dimension. Rather, the value of enumerating these dimensions is to provide a flexible underpinning for an economically efficient, market-based allocation of spectrum across all its possible applications.

To describe the seven electrospace dimensions and the overall rights regime we propose, we must first develop some terminology. The three spatial dimensions identify the physical location of a hypothetical receiver at which location and time all the other dimensions (the field strength) are defined. We posit that a license for rights to an electrospace volume would establish a limit we call $E_{0,}$ expressed in volts per meter for example, on the field strengths caused by the licensee outside the licensed regions. In practice, a de minimis statistical exemption may be necessary, as we discuss in Section 4.5 below. We say a given signal "occupies" an electrospace volume consisting of all "locations" in the seven-dimensional hyperspace wherever and whenever the strength of its field is greater than the de minimis level.

We define an "ideal receiver" as a theoretically perfect receiver that can separate any two radio signals that differ in at least one of their electrospace dimensions – even if they are present in the same geographic location. For example, two co-located ideal radio receivers could function without interference if the signals were at different frequencies, or if the signals occurred at different times, or if the signals came from different directions. Likewise, ideal receivers can separate two otherwise identical radio signals (same frequency, operating time, and direction) if they are present at different locations.

A key feature of the electrospace model we develop below is that the rights to access an electrospace volume (the set of licensed electrospace rights, or LERs) describe only the features of transmitted radio signals and the "locations" they occupy. An electrospace volume includes no consideration of receivers and their performance characteristics, and therefore no explicit definition of, or limit to, interference. This is critical to the technology-neutral feature of our proposed rights system. Interference is a degradation of receiver performance caused by unwanted signals, meaning it is a function of specific technologies, applications, and operational factors, especially receiver capabilities. Section 3.6 describes in more detail a dual-space spectrum rights model that includes a separate transmitter space and receiver space and outlines situations in which it offers strengths over the approach we propose here.

In earlier literature, some researchers suggested a somewhat different set of electrospace dimensions than the seven we propose here.[7] For example, many of the earlier investigators excluded the two "direction-of-travel" dimensions of azimuth and elevation, possibly because the technology at the time did not lend itself to useful exploitation of those propagation properties. It was difficult to construct efficient directional receiving antennas for the much lower frequencies in use at the time, and adaptive directional receiving antenna systems and Multiple Input Multiple Output (MIMO) technology had not yet been invented. In the electrospace model, the transmitter includes the directional characteristics of the associated transmitting antenna, and the receiver includes the characteristics of the associated receiving antenna.

In addition, some researchers recommend a rights regime that includes polarization and/or modulation as possible electrospace dimensions[8] and we do not. Both signal properties are valuable tools in system design, but do not lend themselves to robustly distinct electrospace volumes. Although polarization is surely a useful method of separating radio signals, we exclude it here because polarization involves only two possible orthogonal values, compared with an arbitrarily large number of different values for some of the other electrospace dimensions. Moreover, whenever two polarizations are in use at a given location, it is likely that a single user will coordinate their use to ensure that the signals remain orthogonal to each other. Similarly, we have excluded modulation because any two differently-modulated signals must be distinguished from one another in a coordinated fashion. Of course, specific families of coded modulations exist that are orthogonal to other members of the same family (e.g., CDMA), but receivers cannot necessarily reject unknown modulations. The high degree of necessary cooperation between users of different polarities and modulations suggest that establishing regulatory boundaries between access rights across those dimensions would be problematic, at least with current technology.

---

[7] See, e.g., FCC, Spectrum Policy Task Force, Report of the Spectrum Efficiency Working Group at 8 (Nov. 15, 2002), available at http://www.fcc.gov/sptf/files/SEWGFinalReport_1.pdf.

[8] Leslie A. Berry, Spectrum Metrics and Spectrum Efficiency: Proposed Definitions, IEEE Transactions on Electromagnetic Compatibility, Vol. EMC-19, No. 3, 254 (1977).

Now we consider each of the seven dimensions in more detail and explain our reasoning for including them in a flexible spectrum rights regime.

3.2. **Frequency**

The frequency dimension of the electrospace has the standard meaning of the word, namely a description of the frequency or range of frequencies (bandwidth or pass band) at which one characterizes field strength. The term "frequency" is often shorthand for the range of frequencies (bandwidth) within which a system operates. A spectrum rights regime can divide frequencies over a wide range of increments, and regulators typically divide particular services into channels. For example, regulators divide federal mobile radio bands into multiple adjacent channels that are 12.5 kHz wide.

The physics of electromagnetic waves and the history of technology have influenced how different frequencies are used today. For example, the ionosphere often reflects frequencies less than 20 MHz back down to earth, which made these frequencies especially valuable for long distance communications in the era before satellites and terrestrial/undersea fiber optic cables.[9] In addition, vacuum tubes that formed the basis for early radio equipment worked best at the lower frequencies. Thus early radio devices operated mostly at lower frequencies, and their signals filled the bands below 20 MHz. After WWII, improved vacuum tubes (and later, transistors) raised the frequencies at which consumer devices could operate, but for many years the Ultra High Frequency (UHF) television bands (500-800 MHz) were at the top edge of usable consumer frequencies. As recently as the 1980's, 1 GHz was the effective upper frequency limit for consumer devices.

From the 1950's through the 1970's, some specialized military, industrial, and scientific electronic devices operated at much higher frequencies (up to 10 GHz) to provide point-to-point microwave links and radars. These expensive and exotic electronic systems needed those higher frequencies because shorter wavelengths can focus into tighter beams of energy. For example, radar systems needed narrow beamwidths to send energy long distances and make high resolution radar images. Point-to-point microwave systems take advantage of narrowly-focused radio beams and their compatible antennas to pass communications efficiently between fixed relay stations in a chain of microwave links.[10] In contrast to the antennas designed to receive narrowly directed beams (called "high-gain" antennas), omnidirectional antennas work less effectively at high frequencies. Mobile and cellular systems that use omnidirectional antennas at higher frequencies typically compensate for this with shorter-range radio links.

The commercial exploitation of higher frequency bands continues as technology improves. For example, the personal communication systems (PCS) in the frequency band 1850-1990 MHz are highly successful. Consumer Wi-Fi systems now operate at 5 to 6 GHz and consumer satellite TV receivers operate in the 12 GHz band. The development of cheap and powerful consumer

---

[9] John M. Vanderau, Robert J. Matheson and Eldon J. Haakinson, "A Technological Rationale to Use Higher Wireless Frequencies," NTIA Report 98-349, 3 (1998).
[10] Id. at 6.

electronics for the higher frequencies has greatly expanded the amount of electrospace that is suitable for consumer and other systems, opening new "frontiers" of usable higher frequency bands, as shown in Figure 2 below.

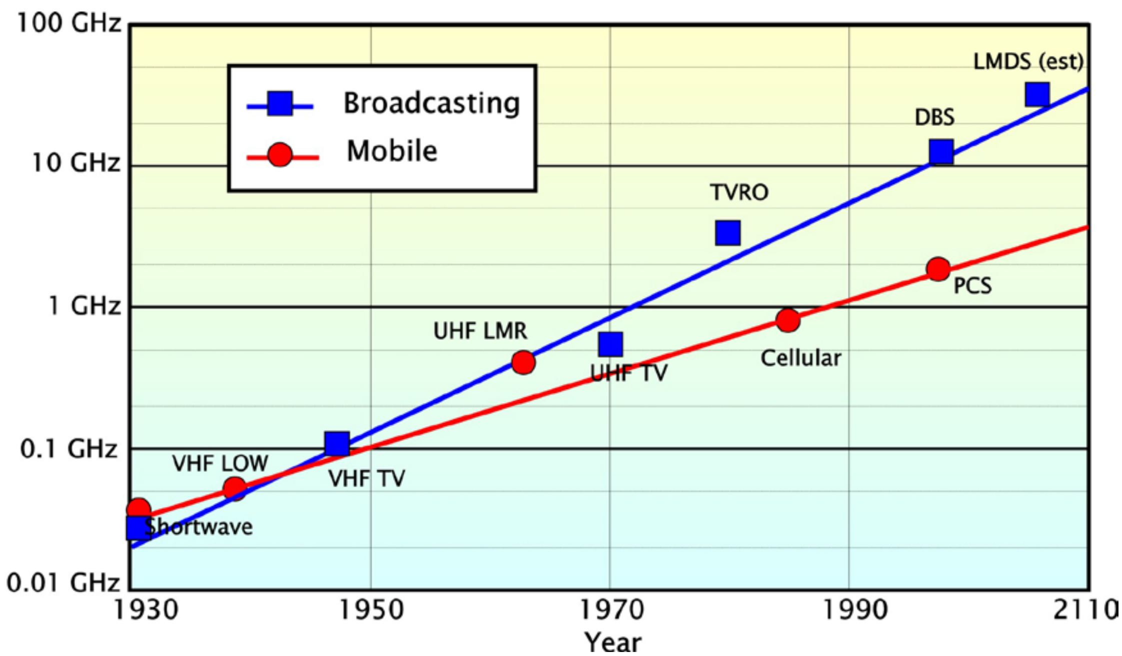

Figure 2. Frequency versus year-of-introduction for new consumer radio systems

3.3. Time

A spectrum rights regime can subdivide time over a wide range of increments. Useful time divisions might include the several-year duration of a license, an agreement to allow a particular user to transmit regularly during the midnight-to-5 AM time block (when bandwidth would be inexpensively available to update computer files for the following day), or a one-time use during a 3-hour special events broadcast. On a much smaller time scale, a user could use a particular time slot on a TDMA system, to broadcast for a 2.5 millisecond time slot once every 20 milliseconds, or transmit data during the vertical blanking interval of an NTSC television signal 60 times every second. Many proposed cognitive radio systems presume that future radios will find and use frequencies that are temporarily unused in a specific time and place – possibly for free or by paying a fee. In its 2003 secondary market rules, the FCC encouraged the resale (or leasing) of unused spectrum by licensees, including temporal subdivisions.[11]

3.4. Physical Location

[11] Promoting Efficient Use of Spectrum Through Elimination of Barriers to the Development of Secondary Markets, *Second Report &Order, Order on Reconsideration, and Second Further Notice of Proposed Rulemaking*, 18 FCC Rcd. 20,604, 20,609 (2003).

The three spatial dimensions describe the physical (geographical) locations where radio energy is present. Figure 3 illustrates how a signal at 825 MHz may propagate over a geographic area, where the different colors show different signal levels (white being the strongest signal and pink the lowest). Although spectrum management would be more convenient if the signal smoothly covered a circular area, Figure 3 shows that the world is typically more “messy.” In typical hilly terrain or within a city, many distant locations could have higher signal amplitudes than many closer locations.

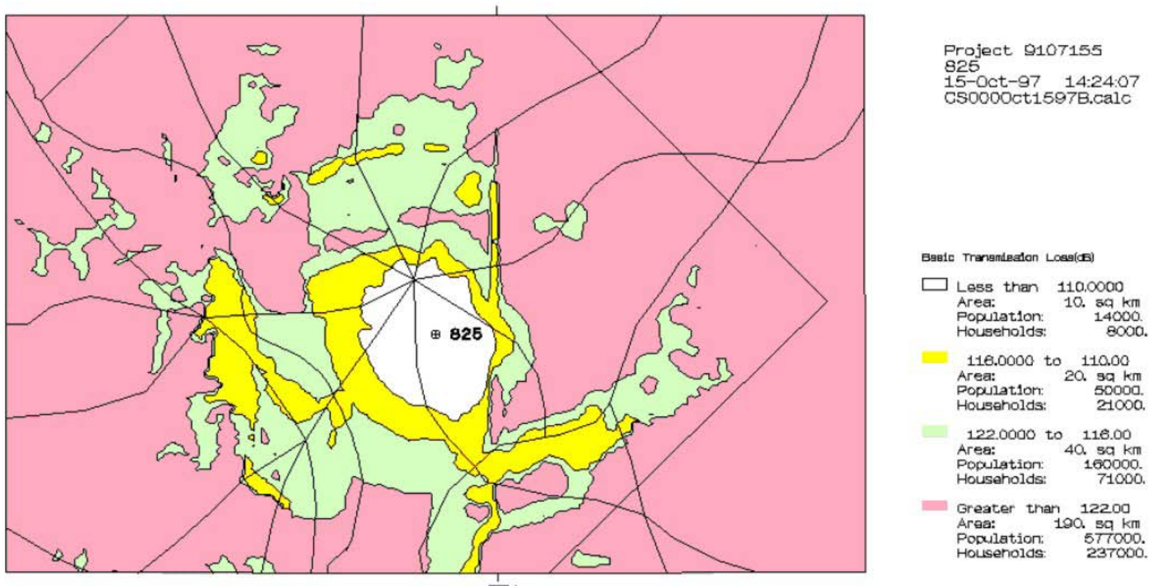

Figure 3. Illustrative Propagation Map at 825 MHz

This figure also illustrates that the coverage area of a transmitted signal might include holes and drippy spills into outlying areas. In order to prevent excessive signal levels (larger than a de minimus limit) outside the boundaries of a selected spatial region, users may have to greatly diminish signal amplitudes at useful locations within their spatial boundaries. Thus keeping a transmitter within specified spatial boundaries relies greatly on setting the right transmitter power, choosing a transmitter location, and using the appropriate directional transmitting antennas given the details of the terrain.

Along with longitude and latitude, the three spatial dimensions of an electrospace volume include height above the terrain, absolute altitude, or some other measure to indicate a vertical dimension. Figure 3 shows the predicted signal strength at about six feet above the level of the local terrain (possibly the typical height of a mobile antenna on a car or a handheld carried by a walking person). The area in white in Figure 3 would tend increase in size and become more circular as altitude increases and obstacles diminish. Eventually the height reaches a point at which propagation is line-of-sight, and the signal strength decreases quite slowly with distance. Often blockage by the curvature of the Earth is the major limit on signal coverage in these cases.

3.5. **Direction of Propagation**

Receivers, including their pointing angle, have no effect on the electrospace volume a signal occupies because they don't affect the strength of a signal as it propagates through the air. Rather, receivers with directional antennas can exploit the direction of propagation (for example, for point-to-point microwave systems, geosynchronous satellites, and radars) such that two signals with otherwise identical electrospace characteristics may be distinguished from one another by the different directions of propagation of the signals.

Specialized applications can usefully exploit different directions of signal propagation, but it involves particular technical challenges. Although a transmitting antenna may produce a beam of energy that propagates in a single direction, the signal may ultimately propagate in many different directions as the original beam is reflected in many directions by objects in its path. At a given location close-in behind a directional transmitting antenna, the strongest signal may come from the back side of the antenna or possibly from reflection off some object illuminated by the main beam. Therefore, the apparent direction of a signal received at any given location does not necessarily correspond to the direction of the transmission. However, the directional characteristics of the transmitting antenna do affect where signals are strongest.

In traditional radio systems, the most useful angle-of-arrival is usually the direct path between transmitter and receiver. For example, terrestrial microwave networks rely on point-to-point, free-space propagation between high-gain, narrow-beam antennas. These directional receiving antennas can very efficiently exploit the direction dimension by separating out individual signals at the same frequency from multiple microwave towers or geostationary satellite orbital slots. Figure 4 shows the directions of signals radiated from two omnidirectional transmitter sites. A receiver located within the small dashed circle would experience signals traveling in two different directions and could separate these signals using directional receiving antennas (even if the signals were at the same frequency). As noted above, the signal directions at the receiving site allow this discrimination, not their direction at transmitter sites. Directional technology is easily scalable by narrowing the beamwidth of the receiving antennas to separate out signals having smaller differences in direction at the receiving site.

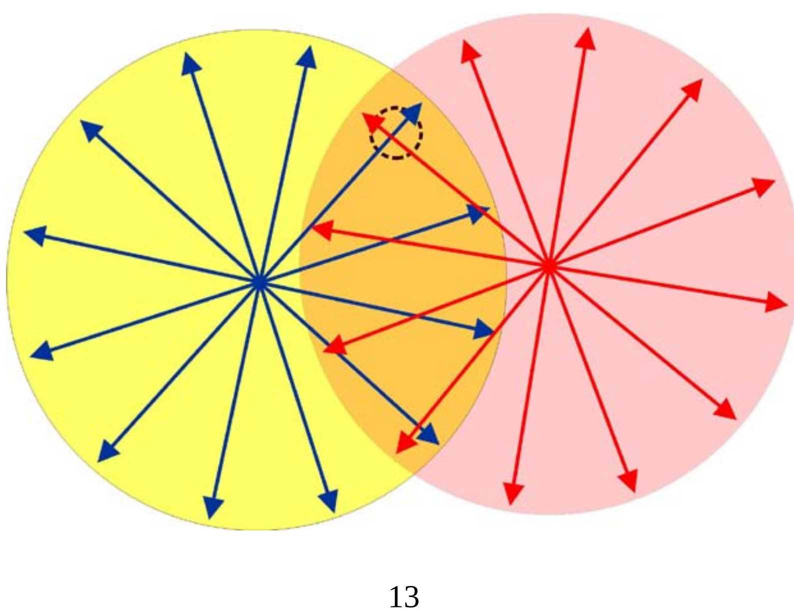

Figure 4. Directional antennas separate signals with different directions of propagation

In many applications, including most mobile and cellular systems, the directional attributes of the radio signals vary over time. Such applications cannot exploit the direction of arrival using simple directional receiving antennas. However, cellular/PCS systems are beginning to use adaptive receiving antenna systems that continuously adjust receiving antennas to track the directional characteristics of the received signals.[12] Such systems can substantially increase the ability of base stations to re-use frequencies without increasing interference, and illustrate the importance of including direction of propagation as a dimension of spectrum rights.

In the right environment, a transmitter can generate multiple independent signals (at the same frequency) and strategically scatter them such that they come from different directions at the receiver. In that case the receiver can use multiple directional receiving antennas to receive multiple independent signals from the transmitter site at a single frequency. Figure 5 shows how this could be done.

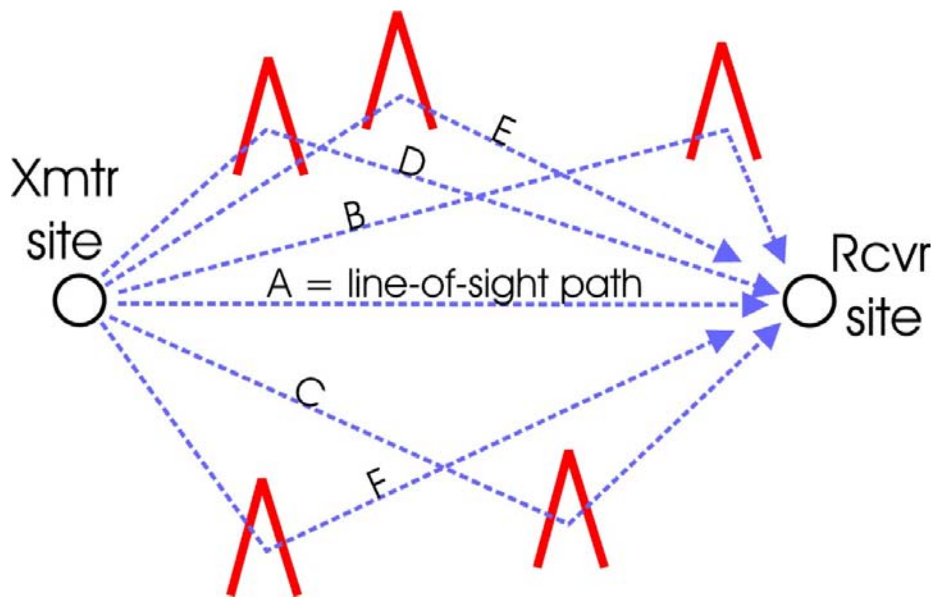

Figure 5. Multiple independent signals by strategic scattering

In a standard line-of-sight system, only Path A (the direct path between the transmitter and receiver) is normally used. However, if the transmitter site transmitted a higher power directional beam (at the same frequency) along Path B towards a selected scattering object (shown by the inverted "V") that was mutually line-of-sight to both the transmitter and the receiver, a usable amount of signal would be scattered from the object to the receiver. Part of this signal would arrive at the receiver site coming from the direction of the scattering object.

[12] Swales, S.C. & Beach, M.A., "A spectrum efficient cellular base-station antenna architecture," Sixth International Conference on Mobile Radio and Personal Communications, 1991, 9-11 Dec 1991, pp:272 – 279.

The receiver could use a high-gain receiving antenna that would pick up only the Path B scattered signal, giving a Path B signal that is independent of the original Path A signal. Similarly, the transmitting site could use addition directional antennas to bounce signals off other objects to generate additional independent paths C, D, E, and F. The scattering objects must be mutually line-of-sight to the transmitter and receiver, and they must all be at different angles as seen by the transmitter and receiver.

The direction of propagation is likely to be an increasingly important dimension across which to partition rights to access spectrum. Advanced technologies, including adaptive antennas and MIMO systems exploit the angle of arrival to allow two or more times greater communication capacity and increased range with no greater transmitter power than conventional systems.[13] Recently developed Multiple Inputs and Multiple Outputs (MIMO) technology illustrates the significant and growing potential to create independent, distinguishable signals at the same frequency in the same location. It exploits multipath reflections and multiple transmitting and receiving antennas to generate mathematically independent transmission channels,[14] somewhat like the multiple directional beams with different apparent angles-of-arrival shown in Figure 5. However, instead of using a combination of directional antenna beams to produce multiple independent paths, MIMO uses multiple omnidirectional receiving and transmitting antennas, whose signals are bounced off various objects in their path, resulting in multiple signals with slightly different directional compositions.[15] Under certain (fairly common) conditions, mathematical processors in the receivers can separate out the independent signals, including those at the same frequency.[16] Compared to the use of directional antennas as in Figure 5, MIMO technology is potentially much cheaper and requires no painstaking aiming of directional antennas. Recent standards bolster the potential of MIMO by including techniques that substantially increase the distance and bits per second of transferable data, relative to single-path Wi-Fi techniques.[17]

3.6. **Dual-Space Spectrum Usage Models**

The electrospace model of spectrum usage we have described so far recognizes only radio signals and ignores any aspects of receivers. This contrasts with the "dual-space" spectrum rights approach employed by most regulators for the past century"[18] Traditional spectrum regulations implicitly contain two interacting "spaces" having the same nominal seven dimensions as the electrospace.[19] One of these spaces is known as the "receiver-denied-space" or "receiver-space." The other is known as the "transmitter-denied-space" or "transmitter-space."

[13] See Ezio Biglieri et al., MIMO Wireless Communications 1; Andrea Goldsmith et al., Capacity Limits of MIMO Channels, IEEE Journal on Selected Areas in Communications, Vol. 21, No. 5, 689 (2003).
[14] Id. at 684.
[15] Claude Oestges and Bruno Clerkx, MIMO Wireless Communications 30, 66 (2007).
[16] Id. at 1.
[17] See Allen Fear, Catch the new wave in wireless networking: 802.11n, CNet, Sept. 26, 2005, at http://reviews.cnet.com/1990-3243_7-5124418-1.html. Also see IEEE standards 802.11n for Wi-Fi.
[18] See Berry supra note, at 255.
[19] D. L. Ewing, L. E. Berry, "Metrics for spectrum-space usage," OT Report 73-24, November 1973

Understanding this traditional dual-space approach is important to appreciating the novelty of the electrospace volume approach we propose here.

In the dual-space approach, the transmitter-denied-space consists of the seven-dimension region where signals from transmitters are present. A given location is "occupied" if the presence of transmitted signals at that location (geographical, temporal, frequency, and direction) would deny that space to a new receiver, because the new receiver would get interference from the existing transmitted signals.[20] Regulators must analyze transmitter-space before they license a new receiver, for example using terrain-based propagation models, existing transmitter characteristics, and the characteristics of a typical receiver in that band and service (a "reference receiver"). If, say, the reference receiver can't operate properly within 200 meters of the transmitter over a 100-MHz range of frequencies, the transmitter-space is deemed occupied over the 100-MHz range for geographical locations within 200 meters of the transmitter location. On the other hand, if a proposed new receiver used a directional receiving antenna, the regulators would have to modify the transmitter-space to indicate the lack of occupancy in locations where the new receiver antenna pattern would adequately reject interfering signals.

Receiver-denied-space is the electrospace that is "occupied" by receivers, and regulators must analyze it before they can license a new transmitter. Regulators consider receiver-space occupied at all locations where a new transmitter would cause interference to the existing receivers. Therefore, the existing receivers, regardless of how inefficient they may be, deny the use of that spectrum-space to new transmitters.

At first glance, the transmitter-space may seem almost identical to the electrospace model we described above, but they are different. The transmitter-space and the receiver-space do not contain signal-strength numbers like the electrospace does. Instead, the transmitter-space and receiver-space contain occupancy states, whose values are either "yes" or "no," i.e., "occupied" or "not occupied." For example, transmitter-space occupancy in systems with omnidirectional antennas will tend to surround transmitter locations, while receiver-space occupancy will surround receiver locations.

Different bands could have considerably different versions of the receiver-space and transmitter-space, depending on the details of services provided in them. For example, point-to-point microwave bands would probably involve an 8-dimension space that includes an additional dimension for polarization. Mobile radio bands would use five dimensions, including neither direction nor polarization. If a band includes two quite different sets of equipment, such as a typical mobile band with fixed-location base stations and mobile units that communicate with them, then regulators must determine separately how to protect mobile units and base stations from interference.

The dual-space model is typical of traditional command-and-control frequency bands, where each band contains a uniform type of transmitters and receivers and where licensees must protect others' receivers from interference. The requirement to protect receivers means that the

[20] See Berry supra note, at 255.

spectrum rights system must include information on receivers and what it takes to protect them. This approach creates ambiguity and conflict over whether interference is caused by a faulty transmitter or by a faulty victim receiver, for example, and can require adjudication over which system was licensed first at a specific location. This complicates determining responsibility for correcting interference problems, especially if the regulatory record does not enumerate all details of the technical performance of the devices. Thus the complex dual-space regulatory system creates uncertainty, costly disputes, dampened investment, and impeded innovation.

3.7. **Exploiting More Electrospace Dimensions Expands Capacity**

The amount of information carried by wireless systems is growing rapidly across the globe. Ten to twenty years ago, some were concerned that usable spectrum would soon be fully exploited, precipitating a major communications crisis. Today analysts view spectrum capacity, the ability of spectrum to carry communications, as a more elastic function of technology. Here we consider how improvements in technology and spectrum productivity require flexibly partitionable spectrum rights.

A great variety of transmitters and receivers offer a wide range of information carrying capacity. For example, a 6-MHz broadcast television channel can carry one standard-definition TV (SDTV) program using old analog NTSC technology, or five SDTV programs using digital TV with MPEG-2 compression techniques, or possibly ten SDTV programs using future MPEG-4 or similar compression techniques. Thus, the number of TV programs that can be carried in a single 6-MHz-wide channel depends on the technology and infrastructure deployed.

This illustrates how spectrum capacity can be a function of advanced image-compression processing. Many other spectrum capacity improvements result from more intensive exploitation of electrospace dimensions. For example, short-range systems allow more frequent re-use of a frequency than longer range systems. Thus, smaller cells in PCS or cellular systems, or short-range Wi-Fi systems increase spectrum capacity by geographically partitioning the electrospace.[21] Trunked radio systems statistically partition a channel by time across many users, allowing the system to carry much greater amounts of traffic on each frequency.[22] The number of frequencies available to carry the rapidly-growing amounts of consumer wireless traffic has increased almost five–fold over the last decade or two as the upper frequency limit for economical consumer systems grew from 1 GHz to about 5 GHz. And Section 3.5 discussed, the directional dimension is now exploited much more intensely in systems using MIMO technology to generate multiple paths at the same frequency.

Since every electrospace dimension can allow additional independent information paths between the transmitter and receiver, more complex systems can in theory greatly multiply the amount of traffic carried in a given band. For example, a new system that might combine a MIMO technique that increases direction-of-arrival capacity by a factor of K, a small-cell architecture that improves spatial reuse by L times, and a modulation/compression technology that increases

[21] P. Mohana Shankar, Introduction to Wireless Systems 2, 3 (2002).
[22] Id. at 156.

the traffic in a given bandwidth by M times could produce a total improvement in system capacity by a factor, F = K x L x M.

In 1994, Matheson forecasted the importance of many of these techniques to improve spectrum capacity. A review of an excerpt from that paper validates his prognosis:[23]

> "SUMMARY OF SPECTRUM CAPACITY FACTORS
>
> The previous sections have described some of the factors that are expected to produce additional spectrum or spectrum capacity. There are more factors that could generate additional spectrum capacity, such as trunking technologies, more precise modeling in frequency management decisions, the use of active interference avoidance in low-power bands, and market-based redistribution of under-used frequencies. In this section, however, I will consider only the cumulative effect of spectrum capacity factors discussed earlier:
>
> a. Fiber optical spectrum reclamation 1.2
> b. Federal/military frequency reclamation 1.1
> c. Availability of higher frequencies 5
> d. Frequency reuse/short-range systems 25
> e. Digital compression techniques 3
>
> **Total increase in spectrum capacity = 1.2 x 1.1 x 5 x 25 x 3 = 495**
>
> The product of these factors gives an overall increase in spectrum capacity of 495. This means that within the next 10-15 years these factors are expected to give us the equivalent of 495 times the spectrum capacity that we have at present using traditional technologies. It is worth noting that the factors that make more frequencies available (a,b,c) give an increase of 6.6, while the factors making better use of frequencies (d,e) give an increase of 75. The reclamation-based factors (a,b) increase the spectrum by 1.3, while the technology-based factors (c,d,e) increase spectrum capacity by 375. This suggests that technology improvement has much more payoff than a rigorous reclamation of frequencies from lower-priority users."

This illustrates that improved technology and investment can greatly increase the amount of available spectrum capacity (somewhere around 375 - 495 times, according to the above summary). In the 16 years since the paper was written, many of the predicted improvements have been realized or are recognizably progressing. Forecasts made today would emphasize other factors, for example the two- to three-fold potential capacity gain from MIMO techniques, even greater frequency reuse from picocells, and the potential gains from dynamic sharing.

However, for new technologies to incorporate these improvements, the radio regulatory framework must allow their rapid and flexible deployment. History suggests that technological improvement flourishes in the most flexible regulatory environments. Over the past 10 to 15

---

[23] R.J. Matheson, "Spectrum stretching: adjusting to an age of plenty," IVHS Journal, 1994, Vol. 1(4), pp. 397-407.

years, a disproportionate share of improvements in spectrum capacity occurred in the unlicensed bands (via short-path systems – especially Wi-Fi – and higher frequencies), the cellular/PCS bands (via short-path systems, compression, and fiber optic reclamation), and the satellite broadcasting (TV and audio) bands (via compression, directional discrimination, and higher frequencies). The traditional command-and-control bands, which we discuss in the following section, did not typically experience similar improvements. Licensed usage in many of the huge federal and non-federal point-to-point microwave and MMDS bands actually dropped, as the traditional services provided by these bands were replaced by more modern alternatives, especially optical fiber.[24] One possible exception is the broadcast TV digital transition which has provided important spectrum capacity improvements through digital image compression.

The next 10 to 15 years could bring unknown new systems with efficiency improvements analogous to the past decade or so, and our ability to foresee these new systems is likely no better than savvy observers 15 years ago who failed to foresee the amazing smart phones we have today. Thus policymakers should set up a spectrum management system now that enables these potentially enormous improvements whatever they may be.

4. COMMAND-AND-CONTROL REGULATION OF THE AIRWAVES

4.1. The Framework: Band Allocations

This section describes how the current command-and-control approach of the FCC and NTIA works, with special attention to the advantages and disadvantages of command-and-control regulations for specific kinds of spectrum-dependent applications. The FCC and NTIA reserve specific frequency bands for specific services, covering all currently useful frequencies.[25] The range of frequencies they regulate lies between 9 kHz and 275 GHz (the “radio frequency range”), although currently no devices operate above about 100 GHz.[26] Optical devices (with frequencies 10,000 times higher than 100 GHz) also use electromagnetic waves, but at this time these are specifically exempted from regulation as radio devices.

Within the radio frequency range, every frequency falls into exactly one allocated band. In general, a single set of rules applies to each frequency in a band, although the U.S. allocation table annotates some individual frequencies with special footnoted rules.[27] The broadly applicable rules within a frequency band dictate what services the user can offer, by whom, using what signal characteristics. For many years, these rules were even more specific, for example specifying exactly which type of mobile radio user (such as forest products firm, film producer, or ground transportation service) could use a given frequency. Over the past few decades, regulators have eased band allocation rules, but important strictures remain. For example, limits

[24] R. J. Matheson, “Spectrum usage for the fixed services,” March 2000. NTIA Report 00-378
[25] See Figure 1 supra.
[26] See 47 C.F.R. § 2.102(a).
[27] See 47 C.F.R. § 2.106.

remain on the ability of users to change the services they offer and to transfer, subdivide, aggregate or otherwise flexibly manage their spectrum rights.

The most important constraints in a command-and-control band are the one type of radio system it allows in the band and the highly specific rules about how the systems must be configured. In a mobile radio band, for example, a spectrum license usually specifies the exact frequency and bandwidth of operation; the location, height, and gain pattern of the base station transmitting antenna; the maximum transmitter power and the modulation, the function of the communications link, and more. In some bands, a user may face a menu of options that allow adaptation across rural or urban operation, heavy traffic versus low traffic, long range versus short range, and other conditions. The regulatory agency typically adopts well-engineered system designs for each allocated band, with the expectation that a radio system built with standardized designs will function without causing or experiencing interference. These rules also set typical or worst-case geographical separation distances. “Channelization bandwidths” are the width of frequency bands that regulators supply to each user, and “emission masks” describe how much energy users can leak into adjacent frequencies. In most cases, readily available electronic equipment can meet the technical requirements in the band rules.

Regulators occasionally adjust rules for specific types of service as technology, experience, and market demands evolve, often at the instigation of an industry or user committee and sometimes via an FCC-organized series of public hearings. These hearings typically start with an FCC notice of inquiry, followed by a notice of proposed rule-making, and end with a final ruling with a modified set of band allocation rules. Such proceedings can take a year or more, depending on whether opponents attempt to delay or modify the changes. Opposition often arises from fear of interference to existing systems, but interference concerns also serve as cover for competition-related motives. A more flexible rights system would allow firms to compensate one another for potential losses and accommodate new entrants without lengthy bureaucratic procedures.

4.2. Band Allocation Technology Examples

To illustrate the wide range of technical problems that band allocation rules must solve, this section describes a few technology examples.

4.2.1. Land Mobile Radio

Land Mobile Radio (LMR) bands use extensive division of the electrospace time and frequency dimensions, moderate division of the spatial dimension, and little division of the directional dimension. LMR transmits signals between base stations and users in cars or walking with handheld radios, for example for police and taxi services. The range of the service of most systems is typically 20 miles, but higher antennas and more transmitter power can greatly extend it. LMR systems can take advantage of lower frequencies to improve the potential distance over which the radios can transmit useful signals with omnidirectional antennas. The current band plan includes many LMR bands, ranging from 30-50 MHz to higher than 900 MHz.

The LMR communications channel is primarily designed to carry speech, so it can be relatively narrowband. Analog LMR radio channels are as narrow as 15-25 kHz, while LMR radios that use digital compression can be as narrow as 5-12 kHz. Since a typical LMR radio user might use the radio channel for only a few minutes or tens of minutes per day, the efficient use of a radio channel means that many LMR systems may be licensed to share a single channel among multiple independent users. Strong "courtesy" conventions require a user to listen for current traffic among other users and wait until the current conversation has finished before transmitting.[28]

Special challenges for licensing LMR systems arise from the wide range of potential transmitter locations, including hilltops and rooftops (where their radios have a greatly increased range) and inside basements (where signals will have an especially decreased range) and the wide range of potential distances from the transmitter to the receivers. These factors, along with the random timing of usage, challenge service providers' ability to plan for specific guaranteed levels of performance. But because of the high demand for LMR service, regulators have channeled the LMR as narrowly as possible, reducing channel widths from 50 kHz historically to 12.5 kHz today (with prospects of soon dropping to 6.25 kHz). [29]

More recently, trunked radio technology (which automatically switches users to unused channels) has begun to replace single-channel radio technology. Accordingly, the FCC set new LMR rules that allowed and encouraged licensees to assemble trunked systems from multiple single channels.[30] Unfortunately, service providers did not always manage the transition to trunked radio systems correctly, and newly-built trunked systems in the 800 MHz band caused interference to other LMR radios in the band. LMR services are also facing major competition from cellular/PCS systems, which will probably result in the demise of many LMR systems.

Cellular/PCS systems are essentially trunked LMR systems that have evolved to handle the very dense user environment in large metropolitan areas. By using smaller cells and more flexible modulation formats, cellular/PCS systems provide a wider range of services and much more total traffic per MHz than traditional LMR services. The smaller cells allow greater frequency reuse and wider data bandwidths for a given transmitter power, but they require more costly infrastructure with more base stations. In addition, cellular/PCS systems provide simultaneous voice channels in both directions (full-duplex), whereas LMR usually provides a one-direction, push-to-talk voice channel.

4.2.2. Point-to-point Fixed Microwave Services

[28] See e.g., 47 C.F.R. § 15.711.

[29] See, e.g., The Development of Operational, Technical and Spectrum Requirements for Meeting Federal, State and Local Public Safety Agency Communication Requirements Through the Year 2010, 17 FCC Rcd 14999, 15000 (2002) (adopting a migration path to a 6.25 kHz channel for the 764-776 MHz and 794-806 MHz band General Use and State License channels).

[30] See 1998 Biennial Regulatory Review--47 C.F.R. Part 90 - Private Land Mobile Radio Services, 15 FCC Rcd 16673, 16685 (2000).

Point-to-point microwave services are in many ways orthogonal to LMR service in that they heavily exploit the directional dimensions of the electrospace and usually transmit continuously on single channels. Typical microwave systems use a series of fixed base stations to relay wideband data from one end of the chain to the other, generally carrying traffic in both directions.[31] The stations are all fixed in location and typically communicate only with the two adjacent stations on a line-of-sight basis using highly directional antennas. Microwave systems typically carry continuous wideband signals that may combine a large number of independent narrower bandwidth signals.

Point-to-point microwave links use relatively high frequencies (1.7-24 GHz) that are well suited to directional antennas and less desirable for many other applications.[32] Many users of microwave links are large corporations and federal agencies with the capital and expertise necessary for such systems. Because of the narrow antenna beamwidth (e.g., 1 to 5 degrees), frequencies can be re-used in multiple directions within a single area, and special multi-level modulation techniques can expand the communications capacity further.[33]

The major technical challenges in allocation rules for point-to-point microwave bands concern how well users can control the directional properties of radio waves. State-of-the-art propagation models with accurate terrain and obstruction data, along with certain details on climate and precipitation data, are vital for adequate planning of microwave systems. Directional antenna technology is highly important, and rules may require the use of antennas with narrower beamwidths in metropolitan areas where the bands are relatively crowded and higher frequency reuse is valuable.

Several technological changes have gradually eroded the optimality of the FCC's band allocation and related rules for point-to-point microwave. The first is the development of optical fiber as the chief conduit for the densest data streams, largely replacing point-to-point microwave in the role for which point-to-point microwave frequency bands were so generously allocated in the past.[34] The more-than-adequate availability of microwave spectrum is underlined by recent FCC rule changes in the 10 GHz band and proposed changes in the 11-GHz band that allow smaller antennas (one of the rare recent instances of regulators reducing spectrum efficiency).[35] On the other hand, the rapid expansion of consumer wireless services has created large new demand for relatively short-range, wide-bandwidth microwave links to connect many new base stations. The slow but inexorable improvement in high-frequency electronic devices moves inexpensive consumer electronics into continually-higher frequency bands (possibly up to a 5-GHz upper limit today). Microwave bands that previously required exotic and expensive ultra-professional-grade electronics (thus limiting their usefulness to a narrow class of applications) are now becoming attractive for many other services. Point-to-point bands that were previously free of

[31] Roger L. Freeman, *Fundamentals of Telecommunications* 197, 198 (2d ed. 2005).
[32] See Matheson, supra, 93
[33] For example, quadrature phase shift keying has four state possibilities and quadrature amplitude modulation has sixteen state possibilities. Id. at 208-09.
[34] See Matheson, supra, at 3-4.
[35] FCC 07-38. WT Docket No. 07-54. NPRM for Amendment of Part 101 of the Commission's rules to modify antenna requirements for the 10.7-11.7 GHz band, March 27, 2007.

competition with consumer services may become new "beachfront" properties for consumer services. Finally, geostationary satellite services now share many terrestrial microwave bands. Although existing terrestrial and satellite services share quite well with each other, many of the newer terrestrial services may not share well with the satellite services. This may substantially limit the scope for the FCC to reallocate terrestrial microwave bands.

4.2.3. Radars

Radar systems are non-communications (sensing) applications for various navigation and military purposes. They make extensive use of the electrospace dimensions of direction and time, with less exploitation of frequency and location. Ground-based fixed radars detect and locate aircraft, ships, storm clouds, and various terrain features. Radars in ships and airplanes detect the vehicle location relative to terrain, as well as other vehicles docks, runways, and other relevant objects. Radar antennas often rotate azimuthally (typically a complete 360 degree rotation every five to ten seconds). Any radio-reflective objects in the antenna main beam reflect (echo) some of the pulse energy back towards the radar transmitting antenna, which switches to receiving mode immediately after transmitting the pulse.[36] Radars determine the distance from the radar to the object by carefully measuring the exact timing of the received echo.[37] The amount of energy received in the echo gives some indication of the potential size of the object.[38] As the antenna rotates, the system completes a picture of the location of all objects surrounding the radar transmitter within a range of 50 to 250 miles every 5 to10 seconds.

The first radars used very high power pulsed signals (such as a one million watt pulse lasting one microsecond in duration and repeated 1000 times each second) with a very high gain (30-40 dB) directional antenna.[39] They generated the short powerful pulses using a "magnetron," which tended to radiate energy across a wider-than-necessary band of frequencies. These relatively "dirty" signal sources caused nearby radars to interfere with each other, initially requiring a fairly large frequency separation between radars. More recently, the magnetrons and other high-power devices have become cleaner, and smarter radar receivers can eliminate much of the interference from other radars by adopting different rates of pulsing. Some newer radars use longer duration coded pulses with a series of phase reversals. These long-pulse radars often use solid-state power amplifiers (instead of magnetrons), have much cleaner spectral outputs, provide additional ways to reject interfering radar pulses, and decrease the potential of high-power radar interference to other radio systems.

Some radar applications have demanding military and security performance requirements. In particular, military radars may need to change frequencies, waveforms, signal processing, and beam-scanning rapidly and adaptively.[40] This can help radars escape jamming and spoofing,

[36] Byron Edde, Radar, Principles, Technology, Applications 33 (1993).
[37] *Id.* at 10.
[38] *Id.* at 33.
[39] S.S.Swords, Technical History of the Beginnings of Radar 88 (1986)
[40] See, e.g., Merrill I. Skolnik, Radar Handbook 24.38-24.40 (3d E. 2008).

detect “stealth” targets, and avoid incoming missiles homing on the radar signal.[41] These mission requirements may limit how technically efficient these applications can be.

Radar bands may seem quite empty most of the time even when they are fully in use. The very strong radar transmitter signal is typically present for only one part in 100,000 of the time at a given location. The very weak radar echoes (which constitute the important part of the radar signal) are present a much larger portion of the time, but they are usually too weak to be seen by normal receivers. Regulators have selected some radar bands near 5 GHz to be shared with unlicensed wideband wireless data devices using a sharing technique called “dynamic frequency selection” (DFS). DFS requires the unlicensed wireless device to continually search for radars and to change to another frequency if it detects one.

Radars use roughly the same frequencies as point-to-point microwave systems for the same reasons: bandwidth availability and the need for highly directional antennae. Higher frequencies are especially important for radars integrated into aircraft, missiles, and other smaller vehicles. Therefore, regulators often place radar bands and point-to-point microwave bands adjacent to each other in the frequency chart. Radars cannot generally operate in the same bands with point-to-point microwave links and satellite systems because their high power pulses can overload microwave receivers.

The degree to which radar bands can accommodate other users is an uncertain function of military demands. On one hand, many radar bands are relatively empty (in peacetime, at least) and global positioning system technology often provides much cheaper, more accurate positioning information than radars. Further, new coded-pulse radar technologies could make it possible to squeeze multiple radars into a much smaller set of frequencies than they now use. On the other hand, a large inventory of important military radars may operate most effectively when they can change frequencies over a wide range, an intrinsically spectrum-intensive application.

4.2.4. Broadcasting

Broadcasting is one of the earliest common uses of radio. It is a highly efficient method of sending a common one-way signal to many different users, and it very efficiently exploits the electrospace dimensions of frequency and location, with little division of access by time or direction. A broadcast signal is typically high power to reach a large number of potential users and to minimize the cost of the many individual receivers. Broadcasting today is tightly regulated by a number of technical broadcast standards and many non-technical constraints such as public service announcements and warnings, political campaigning rules, station ownership limitations, and obscenity censorship.[42]

---

[41] See, e.g., Id. at 24.40.
[42] See, e.g., Enforcement of Prohibitions Against Broadcast Indecency in 18 U.S.C. § 1464, 8 FCC Rcd. 704 (1993); Political Primer, 100 F.C.C.2d 1476 (1984); 47 C.F.R. § 73.3555.

Two facts bear most critically on the future of broadcast spectrum. First, most viewers get TV via cable or satellite, with a minority receiving TV via a broadcast signal.[43] Second, with the switch to digital TV (DTV), a standard broadcast signal no longer serves most mobile or portable receivers. TV broadcasters are locked in a losing competition for viewers with cable, fiber, DSL, and satellite, yet they abandoned the huge market of mobile and portable users whose cellphones already have color video screens. These portable and mobile TV watchers are increasingly served by 3G and 4G providers, using much more expensive infrastructure that requires broadband connections to each user. Moreover, it will probably be the growth of 3G and 4G video service that will create the pressure to take additional TV spectrum from broadcasters and give it to cellphone providers. The DTV conversion will reduce the number of TV channels from 59 to 43, but each DTV channel can transmit as many as five standard definition TV (SDTV) programs. Therefore, the switch to digital TV could actually *increase* the number of broadcast TV programs from 59 to 215 (with five SDTV programs per channel).

4.2.5. Summary of Command-and-Control Allocations

In many ways, the command-and-control management of frequencies has served spectrum users very well. The rules that apply within each of the band allocations we described above result from the particular technical requirements for those respective services. These band allocations provide distinct technically optimized radio services and allow vastly different services to operate in appropriate electrospace volumes. This includes a wide range of frequencies – from the lowest of ocean-penetrating 20 kHz submarine communications to the 60 GHz band where huge bandwidths are available over short line-of-sight paths (where spatial re-use is aided because signals are strongly absorbed by the oxygen in the atmosphere). For each allocated band, the command-and-control approach allows carefully tailored rules and standards for systems with good spectrum technical efficiency and moderate cost. The set of regulations and standards for many band allocations constitute virtual blueprints for the design of a wide variety of complex radio systems.

With new technologies and spectrum demands, however, the command-and-control approach could give way to a future policy portfolio that allocates spectrum under a variety of regulatory approaches, including flexible-use property rights, opportunistic or dynamic sharing, unlicensed spectrum, and hybrid approaches that capture some features of licensed bands and some features of unlicensed bands. The flexible-use spectrum rights could accommodate competitive new services and allow more efficient market allocations of rights across exclusive rights holders. Opportunistic sharing arrangements in licensed bands could accommodate transient spectrum demand, and unlicensed bands allow low power transmissions from very-short-range devices without a license. Unlicensed spectrum is particularly useful for applications in which the transaction costs of licensing users would far exceed the value of the small quantity of electrospace that they consume. Each of these approaches provides an advantage over traditional radio-system-specific rules.

[43] See Annual Assessment of the Status of Competition in the Market for the Delivery of Video Programming, *Thirteenth Annual Report*, MB Docket No. 06-189, 2009 FCC LEXIS 556 at *9 (Jan. 16, 2009).

4.3. **Licensing**

This section explores the means by which the FCC decides whether, how, for what purpose, and to whom it will convey spectrum rights with a license.

4.3.1. FCC Licenses, Allotments, and Auctions

The history of FCC licensing illustrates the pressures towards market allocation of spectrum rights as demand for spectrum access shifts out.[44] The original licensing approach allowed all eligible applicants to obtain one, usually for a small registration fee. Frequencies were often "allotted" to specific types of users, such as oil companies and bus services. This led to inefficient allocations of resources because in a given geographical area all of the licenses in one allotment might be fully devolved, while frequencies in other allotment categories remained unused. Most of these narrow categories have been combined into more general categories, but some bands retain special categories of eligible licensees.

As frequencies became more heavily used and licenses became more valuable, the excess demand prompted the FCC to develop a means to select which eligible applicants receive a license. The FCC set up comparative hearings, an expensive and time-consuming process, to determine which one of several competing applicants would provide the greatest social benefits.[45] The comparative hearings were sometimes called "beauty contests" or "liar's contests," depending on one's point of view.

When increasing demand for licenses finally made comparative hearings unworkably complex, in the 1980s the FCC began license lotteries. Predictably, individual applicants hired "application mills" to fill out hundreds of license applications in hopes of being awarded a valuable license. Although this process did work in the sense of expeditiously awarding licenses, it had very few other redeeming virtues. In particular, it raised the issue of whether the lottery winners could immediately sell their licenses to others (not nominally permitted by law) who could actually use the licenses productively.

The last major FCC change in granting licenses was to auction licenses, beginning in the mid 1990s. Auctions have several major advantages, including raising substantial federal revenue, being transparent, and efficiently allocating the resource (at least initially) to companies that would put it to the highest-valued uses. In general, the auction process has worked well.

Despite the improvement, auctions have not addressed important inefficiencies in the structure and substance of that which is auctioned. First, the FCC has not applied auctions to all spectrum blocks. Second, transactions costs still impede the secondary markets for spectrum licenses, despite the good intentions of the FCC's Secondary Market Initiative in 2003, which permits the

---

[44] For more on the history of license allocation policy, see Jonathan E. Neuchterlein and Philip J. Wieser, *Digital Crossroads, American Telecommunications Policy in the Internet Age*, MIT Press, Chapter 7.
[45] Stuart Minor Benjamin et al., *Telecommunications Law & Policy* 127 (2d ed. 2006).

open redistribution of licenses on a secondary market. Third, even robust secondary markets can only facilitate the distribution of licenses in the existing allocated bands. There is no systematic market-driven way for a licensee to respond to changing demands for its services or for the FCC to provide more spectrum for services in high demand.

4.3.2. Spectrum Scarcity and Electrospace Volumes

In some frequency bands and applications, little spectrum scarcity exists. To obtain a license, one may need only apply for it. For example, in some of the fixed services (point-to-point microwave bands), the typical way to obtain a license is to hire a competent consultant to design the propagation path, select a frequency, check the existing licenses for possible interference, and publish the tentative new license details. If no one objects within a certain number of days, the license goes into effect. Likewise, in the 70-, 80-, and 90-GHz bands, licenses are essentially free for the taking and only require registration of the desired frequencies. However, virtually no commercial equipment or components using such high frequencies yet exist, so the only demands are speculative. These frequencies are likely to be used commercially for short-range, high-bandwidth, directional links. Thus when they are eventually in demand, the useful electrospace volumes are likely to be extensively reused geographically.

In bands that are already heavily in demand, the ways in which regulators initially partitioned the electrospace vary greatly. In some bands, such as the original very high frequency (VHF) TV band, the FCC determined the total number of licenses in each city at the time that the entire band was allocated. These initial channel assignments ensured that co-channel licenses had sufficient separation to prevent any harmful interference between them given the technology at the time. The FCC conducted studies on adjacent-channel assignments, assuming that consumer TV receivers would have certain adjacent-band rejection characteristics. In the UHF TV bands, channel protection studies involved as many as three or four channels on either side of each licensed frequency. The FCC determined every possible transmitter license, and applicants applied for a license for a specific transmitter. This made it easy for the FCC to identify available licenses by noting which ones had not yet been issued.

In most frequency bands, however, the FCC does not predetermine the total number of licenses and their locations. Instead, the FCC uses algorithms to determine whether a requested license can be granted based mainly on whether it would interfere with the existing licenses that it has already granted. Based on engineering studies of the specific service and frequency band, the rules might say, for example, that the requested land mobile radio station must be at least 60 miles from all existing licensed stations using the same center frequency and at least 25 miles from any existing station using adjacent-frequency channels. The FCC may adjust these minimum separation distances up or down, depending on the details of the proposed radio site, antenna height and gain, and transmitter power. For point-to-point microwave systems, the FCC would also include antenna beamwidth, pointing angle (azimuth), and detailed topographic data. Using these rules, information about incumbents, and the information in the license application, the FCC determines whether or not to grant the new license.

Importantly, in this approach the partition of the electrospace into licensed volumes is a function of the chronological order and exact specifications for which new licenses are requested. For example, assuming the rule requires a minimum distance of 60 miles between stations, the FCC could grant a new license for a station midway between two stations separated by 180 miles. Granting this single license would block granting any additional licenses between the original two stations. However, if the requested new station had been only 60 miles from one of the original stations, the remaining 120-mile gap would have permitted the licensing of an additional station in the middle of the gap. Therefore, the order in which the FCC receives license requests affects the number of stations that fit into the 180-mile space between the existing stations. This procedure makes it difficult to know in advance how many licenses the FCC can grant in a band. It is also difficult to know the degree to which this procedure creates an economically optimal number and location of licenses.

As market demand and technology changes, the FCC can revise its algorithms for granting additional licenses. For example, in some mobile radio bands, the high demand for additional licenses and the availability of audio compression technology drove the FCC to change the bandwidth of channels in the band, from 30 kHz channels to 15 kHz channels to additional 15 kHz channels offset by 7.5 kHz to fit between the original 15 kHz channelization. In addition, the FCC may decrease the minimum required separation distances on the basis that communications with a little interference is better than no communications at all, or it may assign a given frequency to multiple sets of users. Improved receivers with better adjacent-channel rejection and more-efficient digital modulations have also allowed regulators to decrease the bandwidth of channels and to squeeze the narrower channels closer together. In addition, improved propagation modeling allows planners to move from very approximate rule-of-thumb algorithms (e.g., 60-mile minimum separation distance) to more exact parameters based on topographical data. Thus, the combination of increased market pressure and improvements in technology has (in some bands) prompted the FCC to increase the number of grantable licenses, but the process has been piecemeal, slow, unpredictable, and ad hoc.

4.4. **Interference**

A radio system design must ensure that the application works and doesn’t violate the rights of others. Indeed, the threat of interference is the main motivation for regulating rights of spectrum access, even in unlicensed environments. This section examines different kinds of interference and how command-and-control and flexible spectrum rights systems treat interference differently.

Interference is the degradation of the performance of a receiver by the presence of an extraneous radio signal. As we discussed above, interference is a phenomenon associated with the performance of a receiver, although the extraneous signal that causes the interference is often (but inaccurately) called “interference” or “interfering signal.” Importantly, the mere fact of interference does not imply anything about who is responsible for its mitigation; that is the role of the spectrum rights regime.

There are two distinct types of interference, co-channel and out-of-band. Co-channel interference comes from unwanted signals at the same frequency as the desired signal. Out-of-band interference comes from (usually strong) signals at frequencies that are different from the desired signal.

4.4.1. Co-channel interference

Co-channel interference can come from an unwanted signal that is centered on the tuned receiver frequency, as well as spurious radiation from signals whose main energy is at other frequencies. Since the receiver is most sensitive to signals at its tuned frequency, relatively small unwanted signals can cause co-channel interference. Typically, interference will occur whenever the ratio between the received power in the desired signal, S, and the received power in the unwanted (interfering) signal, I, is less than K, where K(dB) = 10 log (S/I). Depending on the modulation, K is typically 4 to 20 dB, meaning that co-channel interference might occur whenever the desired signal is larger than the unwanted signal by less than a factor of 4 to 20 dB.

Co-channel interference often occurs near the edge of the coverage area of the radio system, where the desired signal is relatively weak and the unwanted co-channel signal coming from transmitters in adjoining areas are relatively strong. Co-channel interference can also arise when two mobile users on the same frequency try to transmit at the same time. In general, co-channel interference is relatively well understood, and the FCC licensing process prevents licenses for the same frequency from being assigned “too close” together geographically. The FCC generally assumes that receivers will have imperfect rejection of signals transmitting on adjacent channels (because of unwanted transmitter sidebands, as well as imperfect receiver bandpass filters). Therefore, it adopts similar rules regarding how close adjacent channels can be licensed. However, as bands become more crowded, the FCC may decrease the spacing between users licensed to use the same (or adjacent) frequency to accommodate more users, causing some additional co-channel interference in the absence of improved receivers.

4.4.2. Out-of-band interference

Out-of-band interference occurs when one or more strong signals at a frequency different from the tuned frequency of the receiver cause some of the electronic components in the receiver to partially overload, impairing its performance. In particular, intermodulation (IM) products derive from combinations of strong signals that enter the very-wide-bandwidth first radio frequency (RF) amplifier or mixer stages in the receiver. If these signals are sufficiently strong, they will cause overloading and distortions, which generate numerous spurious signals at additional frequencies. If any of these additional frequencies happen to coincide with the frequency to which the receiver is tuned, IM interference can occur. IM interference is harder to predict than co-channel interference because it is caused by combinations of signals not directly associated with the desired signal. Moreover, by assumption, these other signals are usually operating within the conditions in their licenses. Therefore, usually less remedial action can address the signals causing IM interference. Fortunately, IM interference typically occurs only in very strong signal environments, usually near transmitter sites.

Engineers apply two main approaches to reduce IM interference. The first is to construct receivers that are more resistant to IM interference by designing RF amplifiers and mixers that can handle more total signal power without distortion. However, such techniques use more electrical power, and this can substantially compromise the utility of battery-operated portable equipment. Another approach is to reduce the probability that strong signals will get to the first RF amplifier by incorporating a narrower bandwidth RF filter to protect the RF amplifier. A narrower RF bandwidth proportionately reduces the chance of strong signals within the bandwidth. However, narrower RF filters reduce the range of frequencies over which one can tune the receiver, possibly decreasing the usefulness of the receiver while increasing its size or manufacturing costs.

Spectrum planners can mitigate IM interference by carefully planning the radio environment so that unwanted radio signals are unlikely to cause out-of-band interference. Long-term spectrum planning can help separate the applications involving high transmitter power from applications requiring large numbers of battery-operated receivers. However, historic band allocations are generally quite difficult to change. Moreover, it may be impractical for band allocation rules to keep up with the rapid changes in future technology that might improve or degrade out-of-band interference. Worse, the rigid centralized planning needed to minimize IM problems may discourage the rapid development and innovation needed to deploy new services. In many IM interference situations, the problems are caused by specific combinations of local (i.e., co-sited) transmitter and receiver frequencies. In these cases, regulators or licensees can examine the frequency combinations and adopt a small change in transmitter or receiver frequencies that will often eliminate the IM issues.

4.4.3. Interference protection sets the boundaries of spectrum rights

We noted above that an ideal receiver can successfully separate any two signals occupying distinct electrospace volumes. A sufficiently good receiver can eliminate almost all types of interference. For example, a suitable directional antenna can reduce an unwanted signal while increasing the desired signal at the same frequency. Likewise, a poor receiver can experience interference in almost any environment. This means that the amount of unwanted signal that causes interference is highly dependent on the performance of the receiver.

In theory, two parties could negotiate the boundaries of their respective electrospace volumes such that a cost effective level of receiver performance obtains. This classic Coasian bargain would be the natural result of internalizing the externality of unwanted emissions by establishing clear property rights. However, the FCC's band allocation rules (either explicitly or via other assumptions) assume an expected level of receiver performance and thus dictate the electrospace boundaries, including the separation distance between two systems using the same frequency, the bandwidth of the signal relative to the frequency spacing, and the transmitter power.

Under such an approach, as long as a receiver maintains a minimum required level of interference rejection, any interference that actually degrades system operation is a violation by those who are emitting the interfering signal. On the other hand, if the receiver does not meet the required interference-rejection specifications, then the receiver is at fault. This system provides less incentive for receivers to be any better than the rule minimum, because much of the benefit from higher performance receivers goes to potentially interfering transmitters.

"Receiver-caused" interference and "interfering signal" interference may be technically identical, with the only difference being the party whose rights prevail, often a function of the specific rules for each band. For example, in a mobile radio system, co-channel interference usually faults the "interfering signal" instead of the receiver. However, this assignment of blame assumes that receivers in a mobile radio band are not required to have directional antennas to separate out multiple signals on the same frequency. A future mobile radio system might require adaptive "interference-nulling" antennas, and under these conditions the electrospace rights could configure differently. In contrast to the LMR scenario, a modern fixed microwave receiver (and associated receiving antenna) must be able to reject same-frequency signals on the basis of direction of propagation.

4.4.4. Interference disputes and tenant's rights

The FCC designs its rules regarding emission masks, transmitter separation criteria, and other licensing rules to prevent harmful interference. If all parties follow license rules, systems usually operate without interference. But what if all parties are in compliance with all of the applicable rules and specifications, and interference still arises? Under these circumstances, the principle of *tenant's rights* applies. This principle states that earlier-licensed stations at a site have precedence and the newest station must take the necessary action to eliminate whatever interference occurs as the result of its presence. Another way to put it is "first-in-time is first-in-rights."

Both the NTIA and the FCC (in 47 CFR) support general tenant's rights in their rules. For example, Section 2.3.7 of the NTIA Manual states:

> "In principle, spurious emissions from stations of one radio service shall not cause harmful interference to stations of the same or another radio service… Providing appropriate spectrum standards in Chapter 5 are met, an existing station is recognized as having priority over a new or modified station. Nevertheless, engineering solutions to mitigate interference may require cooperation of all parties involved in the application of reasonable and practical measures to avoid causing or being susceptible to harmful interference."

And Section 5.0.1 of the NTIA manual states:

> "In any instance of harmful interference caused by nonconformance with the provisions of this chapter, the responsibility for eliminating the harmful interference normally shall rest with the agency operating in nonconformance."

These two sections outline a procedure for fixing interference. First, if either system doesn't follow the applicable standards, it must be brought into conformance. Second, if the interference persists, the new or modified station must eliminate the interference, either by modifying its own system or others'. This flexibility is useful because the least cost solution may be modifying one of the earlier systems.

The FCC rules follow a similar outline within rules for some specific services. For example, they define, typically in paragraphs (a) and (b), an applicable transmitter emission mask, and then stipulate:

> "Should harmful interference be caused to the reception of other ... stations by out-of-band emissions, the licensee may be directed to achieve a greater degree of attenuation than specified in paragraphs (a) and (b) of this section."[46]

However, the regulation fails to say how much additional attenuation the FCC may require or how it should decide.[47]

Thus, despite supporting the general principle of tenant's rights, neither the NTIA Manual nor FCC regulations offer details on how to enforce them, not least because their rules have prevented most interference. However, allowing licensees to negotiate levels and kinds of interference would likely necessitate more clarity in interference liabilities and the procedures to resolve conflicts.

However, clarity in rules alone isn't enough to ensure efficient outcomes. Rules that protect existing systems from interference by newer systems can reduce the incentive for incumbents to invest in superior receiver performance. Indeed, they can create perverse incentives for early entrants and deter new investment. If new entrants must protect incumbents, then the first movers into a band have the incentive to expand their implicit rights by adopting poor receivers. If the burden of preventing interference falls only on newer systems, this raises the cost of new investment and holds innovators hostage to a contentious process that protects obsolete systems.

4.5. **Receiver Performance Standards**

If better receivers can reduce interference, why not require higher-performance receivers? Such rules would surely decrease interference, allow more users per band, and generally improve spectrum technical efficiency and capacity. This logic is consistent with a command-and-control regulatory approach because without price signals, market forces, and flexible rights, regulators can't ensure an efficient level of receiver performance. However, mandating receiver technologies is generally economically inefficient, and we argue here that they are unnecessary and possibly counterproductive even in a command-and-control context.

---

[46] 47 CFR §73.44 (c). This regulation applies to AM broadcast radio. Similar regulations can be found in §73.687 (television broadcasting, particularly channels 14 and 69 which are adjacent to mobile radio bands).

[47] Other rules in certain services extend the authority of the FCC to require adjustments in effective radiated power, antenna construction, and antenna polarization. See e.g., 47 CFR §21.106, §21.107, §21.109, and §21.110

Since the engineering of an allocated band (including channel spacing and frequency re-use distances) assumes a set of receiver performance factors, licensed systems can normally operate as planned if the receiver performs at least as well as the design criteria for the band. In addition, as discussed in the section above, one criterion for deciding who is responsible for fixing interference problems is whether the victim receiver meets the expected performance specifications. Further, 47 CFR suggests that the victim receiver must be at least "average" in its performance to qualify for protection against interference. Therefore, users may be able to provide satisfactory service with a cheap receiver, but choose more expensive receivers just to be eligible for the regulatory protection from interfering signals. This is inefficient.

Another reason a specific performance level is undesirable is that optimal receiver performance can vary considerably across different users in the same band, depending on factors like the distance between the receiver and nearby transmitters and the strength of signals on adjacent channels. In most cases, receiver owners are aware of their emissions environments, including the presence of strong signals and the priority given to interference-free operation of the system. Moreover, if a receiver owner deploys an inadequate receiver, he or she bears the full cost of that problem. Unlike non-conforming transmitters (which can cause interference to other users), poorly performing receivers create no external costs to others. Further, a receiver owner may rationally choose to deploy a low quality, low cost receiver if the risk to service quality is worth the savings in receiver costs. At best, regulating receivers in this context inefficiently protects users from themselves. At worst, receiver performance standards may make matters worse; regulators have less information about users' specific circumstances than the users themselves, and users' receiver performance needs are heterogeneous so the standards will be inefficient for at least some of them.

Another reason to avoid receiver performance standards is that ordinary technical standards are too simplistic for today's technologies. Typical receiver performance standards set fairly simple hardware specifications like bandwidth and adjacent channel rejection (possibly including a selectivity mask), noise figure or sensitivity, intermodulation specifications, and maximum local oscillator radiation. However, modern systems can deploy more advanced interference rejection technologies. For example, a modern radio system may be smart enough to choose another frequency or frequency band where the signal environment is less challenging or possibly to use complex signal processing or adaptive antennas to avoid the interference. In addition, users should be able to trade off receiver performance with other aspects of their system design. For example, it might be more efficient to use cheaper mobile receivers in combination with higher signal levels broadcast by a more elaborate fixed infrastructure. Policymakers should not preclude these creative and sophisticated approaches with command-and-control regulation.

Our objection to performance standards applies strictly in the policy domain. Where receivers are part of a system where the faulty operation of one receiver can cripple the performance of other parts of the system, the other parts of the system have a legitimate interest in the performance of every receiver. Public safety, defense, and many other large radio networks are examples of such systems. However, in these cases, the system designers, not regulators, determine the performance of receivers (along with all other system components).

4.6. **Duration of rights**

The electrospace model we presented in Section 2 is fully explicit in the time period over which a user controls the rights to access spectrum; indeed, time is one of the key dimensions of the electrospace. The rights that FCC licenses for non-federal entities and NTIA assignments to federal agencies convey are much less clear on this point than an efficient market requires. Most FCC licenses have a specified multi-year duration, but carry the strong expectation (but not a guarantee) that the FCC will renew the license when it expires. This renewal process could repeat indefinitely, approximating permanent rights but introducing important risks that regulators could encumber the renewal process with reviews and extra requirements, including termination. These unnecessary risks can depress spectrum prices, capital investment, and innovation by introducing uncertainty in the time period over which investors can obtain a return on their investments.

NTIA regulations require it to review spectrum assignments to federal agencies every five years, and it requires agencies to renew licenses every five years, as needed.[48]

In the past, the lack of well-defined temporal rights has served mainly to ensure that spectrum that is no longer needed for its original purpose will remain unused, instead of being promptly re-used for other purposes. Until recently, FCC licensees had little ability (or reason) to transfer those spectrum rights to other users. Therefore, the only choice they had was between keeping possession of the unused spectrum rights or returning the license to the FCC. Recent FCC policy changes have allowed and encouraged licensees to rent or lease unused licenses to other users.[49] However, this policy change tends to repeal the claim that licenses convey no property rights, but only convey the right to a specific user to use a frequency for a while for a single, very specific purpose. The debate over whether or not to provide new flexibilities or more permanent rights often convolves the efficiency and distributional effects of proposed change. Here, without taking a position on who should receive the benefit of more flexible or longer-duration rights, we merely say that *certain* rights are likely to support a more efficient spectrum market than *uncertain* rights.

In general, FCC licensees and federal agencies appear to have lived with ambiguous temporal rights reasonably well, but one can never know the opportunity cost of vague rights in terms of foregone resource productivity. Clarity in the temporal dimension of spectrum rights is likely to become increasingly important in the future, particularly in the context of high spectrum demand, dynamic spectrum access, market driven resource allocation, and spectrum leasing and other secondary market activity between and among federal and non-federal entities.

4.7. **Command-and-control conclusions**

The command-and-control approach has many advantages. It allocates frequency bands with optimal rules for diverse specific services, such as LMR, broadcasting, radars, and point-to-point

[48] NTIA, Manual of Regulations and Procedures for Federal Radio Frequency Management, Annex F.1 (2009).
[49] FCC Report and Order and Further Notice of Proposed Rulemaking (FCC 03-113). May 15, 2003

microwave. Most of these allocated bands remain well- designed from a technical perspective, allowing reliable services for many users with a minimum of interference. Further, detailed allocation rules provide a template for radio system designs, and users can be assured a good degree of performance since incumbents are well-protected from new entrants.

However, the command-and-control approach incurs significant, but uncertain opportunity costs. Even when command-and-control bands are well designed—having technical and economic efficiency—they are designed to support only a single specific service in a given band. In no way do the rules guarantee the efficient allocation of spectrum resources across and within allocated bands to provide the proper mix of desired services. A command-and-control approach substitutes the judgments of regulators for market-driven outcomes, and explicitly prohibits Pareto-improving transactions and resource reallocations. Despite the FCC's Secondary Market Report and Order, burdensome rules still apply to many license transactions, and among the things *not* included in most licenses are rights to change the use of the license in any way, including changing transmitter location, receiver location (if there is a specific receiver location), bandwidth, modulation, and services provided. These inefficiencies likely incur profound dynamic costs by discouraging investment and innovation by increasing the risk that new technologies cannot find sufficient accommodation in the established band plan.

Although regulators eventually match most proposed spectrum uses with a frequency band in which they can operate, it can take years and require millions in legal and consulting fees. In some cases, regulators can slightly modify existing band allocation rules to accommodate the new technology, for example by changing the bandwidth/channelization of a mobile channel from 25 kHz to 12.5 kHz. In other cases, regulators could relocate incumbent licensees in a relatively unpopulated band and redevelop the band with new allocation rules. Such a process converted the old 1850-1990 MHz point-to-point microwave bands to the new Personal Communications Systems (PCS) bands.[50] However, current approaches to accommodating new services are cumbersome, costly, and result in great delay and forgone consumer and producer surplus.

Arguably, NTIA's approach must remain more command-and-control than the FCC's. NTIA has some frequency bands with numerous unique systems designed for military or continuity-of-government missions, so NTIA must engineer each one into a band independently. In addition, certain frequencies must be available in the event of a national emergency with a certainty exceeding any commercial service quality requirements. However, new policy and technological approaches could possibly allow productive use some of the "reserved" spectrum during the long periods between emergencies.

Fortunately, regulators with command-and-control authority can use their discretion to implement band rules that are quite different from traditional exclusive licensed bands. In the subsequent sections we explore workable approaches.

---

[50] Redevelopment of Spectrum to Encourage Innovation in the Use of New Telecommunications Technologies, *Second Report & Order*, 8 FCC Rcd 6495, 6496 (1993).

5. ALTERNATIVES TO COMMAND-AND-CONTROL EXCLUSIVE RIGHTS

One can characterize spectrum policy approaches along two key features: the flexibility users have to design and use radio systems how they please and the degree to which users can exclude others in the same electrospace. The ideal portfolio of policy regimes would maximize the net social welfare that spectrum-using services provide.

Figure 6 below places alternative spectrum management techniques along these two axes. The horizontal axis of Figure 6 represents the degree of flexibility, going from the least flexibility on the left to complete flexibility on the right, and the vertical axis represents the exclusivity of user rights, increasing from bottom to top.

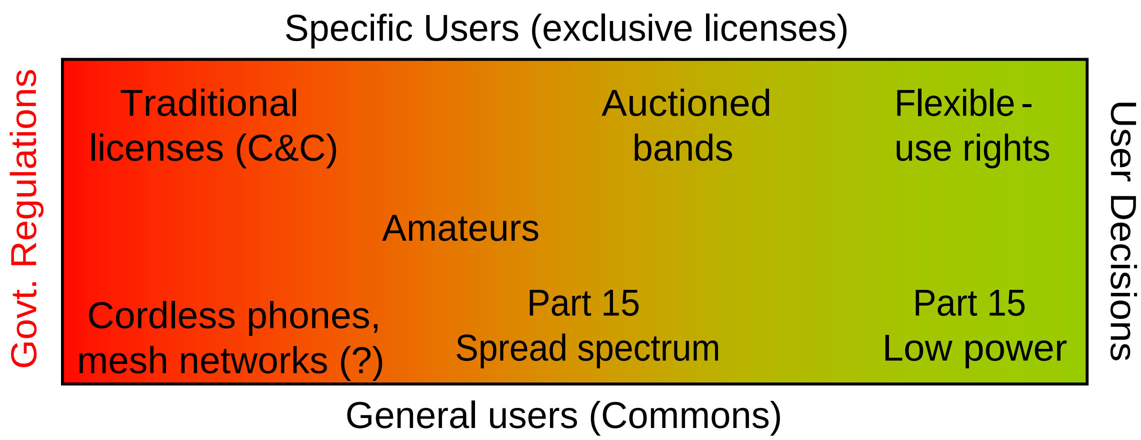

Figure 6. Spectrum management approaches along two continua

Traditional exclusive licenses lie in the upper left corner of Figure 6. As we discussed in Section 4, these licenses involve detailed federal rules and a high degree of interference protection. The lower right-hand corner includes the unlicensed low-power spectrum allocation (47 CFR Part 15), which allows users to operate low power devices without a license. [51] Proliferating unlicensed devices currently offer countless short-range wireless services, including garage door openers, wireless routers, cordless phones, and remote control toys.

In the lower left corner of Figure 6 lies a class of systems that, although not licensed to specific users, follows very specific federal regulations. For example, Family Radio Service (FRS) is a service intended to provide families and other non-commercial users an unlicensed medium-range mobile radio service on designated frequencies/channels. Regulations tightly specify the parameters of the service. Spread spectrum systems also operate in unlicensed bands. More specific regulations govern their use than low-power systems, including specifications on frequency-hopping rates and number of frequency steps. The upper right corner includes flexible-use spectrum rights. These systems use frequencies that are exclusively licensed to specific users, but licensees have broad discretion about how they employ and share their frequencies.

[51] We use the common term "unlicensed" to refer to spectrum in which FCC does not require users to have a license.

In the following sub-sections, we consider the advantages and disadvantages of three alternatives to command-and-control traditional licenses: unlicensed use (Section 5.1), opportunistic use (Section 5.2), and flexible spectrum rights (Section 5.3).

5.1. **Unlicensed spectrum**

Low power, highly localized services are not well-suited to traditional exclusive licenses. Each device occupies such a small electrospace volume that the cost of individual licensing would greatly exceed the value of the electrospace used. Accordingly, the FCC sets two major conditions for devices that operate in unlicensed bands. First, the FCC limits transmitters to very low power and it must approve the equipment for use in the band.[52] Second, users have no presumed protection from interference.[53] One can think of unlicensed spectrum as the regulator granting bubbles of individual electrospace volumes to unspecified nameless users of authorized devices, bubbles which together form a virtual foam of user rights. The power limit of the devices dictates the radii of the individual bubbles. Thus an unlicensed low-power band is not truly a commons in which "anyone can do anything," but rather an *a priori* partitioning of the electrospace into small parts that are unlikely to intersect, at least until devices proliferate to the point that congestion arises. However, given the low power limits, the small bubbles of electrospace occupied by low-power unlicensed devices pose little risk of interference to systems used by anyone else.

As electronics technology has improved, the FCC has allowed somewhat higher power devices in the unlicensed electrospace, provided that the devices include other means to reduce interference. For example, some Part 15 devices include "listen-before-talk" hardware and software that searches a frequency range for existing signals and moves to operational frequencies that are not already in use. As technology matures and congestion increases, more elaborate interference-avoidance capabilities in unlicensed spectrum may become cost effective, including "listen-before-talk" protocols and dynamic frequency selection.

The unlicensed space is a kind of command-and-control approach in that it mitigates interference through low power or other technical constraints. A number of inefficiencies can arise from these constraints. First, unlicensed devices mitigate the risk of interference by severely restricting the operating range, even when there are no other users around. For example, the range of a cordless phone is the same (bounded by regulation) no matter whether one operates it in the center of a 100 acre farm or in a studio apartment in the city. This lack of context-specific protocols leaves spectrum inefficiently underutilized in some instances. Second, different technologies may be better suited to different sets of unlicensed rules, and there is no way for the FCC to ensure that its portfolio of different rules in different unlicensed bands is efficient.

For example, mesh network advocates have proposed various types of unlicensed self-organizing wireless mesh networks as a possible new class of radio systems that could automatically

[52] See 47 C.F.R. §§ 15.201(b), 15.209(a).
[53] *See* 47 C.F.R. § 15.5.

recognize and connect to one another in adjacent areas. This self-organized network would allow users to pass packet data from system to system to the intended destination. Thus, a mesh network could operate over long ranges while being as spectrum-efficient as short-range systems. Although the large-scale feasibility of such systems remains to be demonstrated, their potential is promising, particularly if regulators establish protocols to prevent interference from other unlicensed uses. The FCC's quandary is that it's unclear how to optimally allocate spectrum and set rules for technologies that are not yet mature, but without spectrum and rules the technology won't develop.

Unlicensed policy approaches also raise the question of whether unlicensed devices, because they can operate without charge in a non-licensed band, benefit from an implicit subsidy relative to radio services using licensed spectrum. One issue is whether the difference in pricing distorts services inefficiently towards unlicensed systems and another is whether pricing spectrum would be a relatively efficient source of revenue. The degree to which the government should charge for spectrum access depends, among other things, on the transactions costs of collecting revenue and whether pricing spectrum improves or worsens distortions in the allocation of resources.[54] Licensing low-power systems to individual users is probably not practical, but if unlicensed spectrum becomes inefficiently congested, it might be feasible to tax the sale of unlicensed devices.

5.2. **Opportunistic Access**

Almost all spectrum is idle periodically as the demand for radio communications fluctuates over time, so the question arises whether it's possible to make idle electrospace available for other users. Frequently idle spectrum includes geographic "white spaces" around the edges of licensees' service areas and unused frequencies in bands allocated to out-of-date uses. Further, some bands are filled in crowded urban areas, but not in rural areas.

In this context, some argue that dynamic-use systems could legitimately operate as opportunistic overlays, able to access any spectrum so long as they produce no harmful interference to the licensed incumbents.[55] In theory, this is consistent with current policy because unlicensed spectrum use carries no guarantee of interference protection, and current exclusive licenses guarantee a level of freedom from interference, but not full control over the electrospace volume.

At least two policy challenges arise with allowing opportunistic spectrum access. One is to ensure that opportunistic uses don't harm licensed users. The second is to determine who controls opportunistic access, the licensee or the opportunistic user. We consider these in turn.

[54] For a detailed discussion of the economics of raising government revenue from spectrum access, see "Spectrum Auctions: Distortionary Input Tax or Efficient Revenue Instrument?" by Adele C. Morris, *Telecommunications Policy* 29 (2005), pp. 687-709.

[55] An "overlay" refers to opportunistic spectrum use by a system using relatively high transmitter power. Overlay devices must employ suitable techniques to determine that they will not cause unacceptable interference to existing systems before they can transmit.

The higher the opportunistic transmitter power, the greater the risk it will cause interference. That is because higher power signals have greater range, and it's harder to infer the presence and the technical operating parameters of more distant and numerous potential victim receivers. Scholars have proposed many ways opportunistic users could protect distant victim receivers, and policymakers could require or encourage any number of them:

- Include more sensitive search receivers in opportunistic devices.
- Use networked opportunistic devices to search for distant devices.
- Require licensed receivers desiring protection to transmit their status to a real-time database.
- Synthesize information on licensed devices from license databases, geographical information systems, and real-time status updates from the devices themselves.
- Embed the information in opportunistic systems.
- Use the information to enable propagation modeling.
- Let a band manager with the proper information grant or deny opportunistic access.
- Have potential victim receivers measure and report on test signals from opportunistic users.
- Conduct interference temperature calculations at victim receivers before allowing opportunistic use.

None of these techniques is foolproof or easy. Some approaches would require substantial cooperation from licensees, and most involve gathering and disseminating information and other costs. Policymakers would have to decide who bears these costs and how to incentivize compliance. For example, one option would be to set penalties that rise with interfering power levels.

If regulators allow licensees to exclude opportunistic use, then to promote the efficient allocation of the resource regulators should allow licensed users to invite opportunistic sharing under mutually agreeable conditions. In this approach the market, rather than regulators, would determine the tradeoff between the value to licensees of exclusivity and the value of spectrum access by opportunistic users.

A clash of interests is virtually inevitable in a transition to robust opportunistic access. Licensees may be uncomfortable with unauthorized "trespass" on their licensed rights, even if they cannot prove they have experienced harmful interference. Potential opportunistic users can argue that license provisions don't guarantee exclusivity. If regulations are vague or ambiguous, an efficient resolution to these competing interests through market forces is nearly impossible. This underscores the primary importance of establishing regulatory clarity, one way or the other, about the degree to which licensees can exclude all other signals. Given the intensity of interests on all sides, this will require strong leadership.

Opportunistic systems face other challenges as well. They must compete with services provided with licensed spectrum, while potentially requiring more complex infrastructure and user equipment to provide less reliable service. They must also find their market niche. The greatest demand for opportunistic services could be where unused spectrum is most scarce, but these could be the same areas with the highest cost of service. Opportunistic services may find their comparative advantage with lower-priority, interruptible communications, similar to services provided in nonlicensed bands.

Despite these challenges, finding pragmatic ways to allow some form of opportunistic use is likely to be increasingly important. Present trends point to enormous growth in wireless services that require flexibility in location and services. Under these conditions, it is increasingly unlikely that traditional static licensed rights will efficiently match the communications requirements of the marketplace. Section 6 below argues that a regulatory system that employs an electrospace approach to expressing spectrum rights would allow consenting licensees to permit opportunistic use of their spectrum and meet dynamic market demands with the minimum of transactions costs.

5.3. **Flexible Spectrum Rights**

Regulators could delineate spectrum access rights that are flexible, exhaustive, and economically efficient. In our suggested regulatory regime, licensees would acquire and use rights to access spectrum that are expressed as disjoint electrospace volumes. We draw much of the material to this topic from earlier papers by Matheson.[56,57] The ideas here are also very consistent with Kwerel and Williams[2006].[58]

In this system of "ideal spectrum rights," regulators govern interference and the allocation of the resource by the following fundamental rule:

> 1. Licensees must keep all signals within their respective licensed electrospace volumes (including frequency band, geographical area, and authorized time of operation). In particular, all signals must have a power level of less than E_0 outside their electrospace region.

The rule simply reiterates the boundaries of the electrospace volume as the rights of spectrum access. Section 6 addresses the practical setting of $E_{0.}$ This idealized approach is best explained by the rules it does *not* include. Signals within the electrospace volume have no limitations on power or field intensity. No regulatory constraints on receivers or other technologies apply. No regulatory constraints on the services offered with the spectrum rights apply, including no requirements for public interest services. In particular, if the government wishes to ensure licensees undertake certain public interest activities, then the government would have to achieve its goals more transparently than through encumbrances on spectrum rights, for example through direct grants. No specific protection from interference (other than the electrospace parameters on

---

[56] R. J. Matheson, "Flexible-use spectrum rights - a tutorial," Proceedings of the International Symposium on Advanced Radio Technologies, March 2003. NTIA Special Publication SP-03-401.
[57] R. J. Matheson, "Principles of flexible-use spectrum rights," Journal of communications and networks, Volume 8, Number 2, June 2006, pp. 144-150. (ISSN 1229-2370)
[58] *Defining Spectrum Rights,* a presentation at the NTIA Workshop on Improving Spectrum Management Through Economic and Other Incentives, Feb. 28, 2006.

other licensees) applies. Licensees may aggregate, divide, lease, or transfer their rights via secondary markets and private contracts as they see fit without prior approval by regulators, and they may partition their rights along all electrospace dimensions. Licensees will presumably design their radio systems so that they can operate in the presence of small, unwanted foreign signals (less than or equal to E_0).

In this approach, the role of the government is to devolve initial allocations of electrospace volumes, to manage a public database of spectrum rights, to enforce those rights, and to mediate disputes when necessary. This approach allows licensees to efficiently manage interference with the maximum of flexibility of spectrum use, and it would allow robust and liquid markets in electrospace volumes to emerge. The issues we discuss here concern only the delineation of spectrum rights and obligations. We do not address the means by which the user obtains those rights.

As long as all licensed users of non-overlapping electrospace regions follow the fundamental rule, each licensed user can operate without interference, no matter what the other compliant licensed users are doing within their own non-overlapping electrospace regions. Since we have made no assumptions about the actual sizes of the various users' electrospace regions, there is no reason to suppose that changing the size or number of electrospace regions would cause any additional interference. One must ensure that any partition of an electrospace volume creates truly disjoint sub-volumes, but other than that there is no reason to regulate secondary market transactions.

6. PRACTICAL ELECTROSPACE SPECTRUM RIGHTS

There are several practical issues in implementing an electrospace model of flexible-use spectrum management. The first is to set the emissions limits outside licensed electrospace volumes and ensure that the rules prevent such signals from combining in ways that produce inadvertently strong signal environments. Another challenge is to accommodate real-world receivers that cannot reject all signals at unwanted frequencies and from unwanted directions. Finally, we have to recognize that signal propagation varies by altitude, weather, and other conditions, so it's hard for users to constrain their signals to their electrospace volumes in all times in all places with total certainty. The following sections review these challenges and suggest pragmatic policy approaches to deal with them.

6.1. Flexible Use Rules: Setting Emissions Limits

Rules for E_0

The first fundamental rule of the electrospace approach to spectrum rights is that outside the licensed electrospace region all signals must have a power level of less than E_0. It is not obvious what numeric value to choose for E_0, although its units would logically be in W/MHz/m^2. E_0 would not meaningfully vary as a function of the geographic size of a licensee's electrospace value, or its bandwidth, duration, or angle-of-arrival boundaries. When the bandwidth of a signal rises, the bandwidth of the signal leaking across geographical boundaries will rise as well, but the value of E_0 at any particular frequency should remain the same.

Regulators could choose different values of E_0 in different bands since the minimum level of interfering signal for different systems varies over a wide range – perhaps 40 to 50 dB – depending especially on the gain of the receiving antennas and the required signal-to-interference ratio. This could make different bands particularly suitable for different types of services. In principle some variation in E_0 across bands makes sense, but, again, the purpose of flexible rights is to allow the market to determine the use of the spectrum as much as possible. It is important that the setting of flexible use parameters doesn't become another way to regulate which services are provided in which bands.

One complication in allowing free market aggregating or dividing electrospace regions is ensuring signal strengths don't accumulate much higher than E_0 across adjoining electrospace volumes. For example, the owner of a single electrospace region could in theory game the system to partition his electrospace region across one of its dimensions and with each partition gain a separate allowance for E_0. Regulators may find that the E_0 limit should apply to the overall emissions outside a licensee's aggregate holdings of electrospace volume rights rather than emissions outside a particular licensed electrospace volume.

The potential for emissions leakage across electrospace volumes is important in the frequency dimension, too. Although a transmitter can radiate high power inside the licensed frequency range, the signal strength outside the licensed band must be less than E_0. The way we've discussed this so far, this condition must be met at all locations – even very close to a transmitting antenna, where the field strength is very high. At such locations, very high field strength inside the licensed frequency range would need to drop below E_0 immediately outside the licensed frequency range (possibly within a 12.5 kHz LMR channel) – requiring a very rapid decrease in signal strength over a small change in frequency. To allow licensees to ensure their compliance in high field strength locations such as very close to transmitters, regulators could offer licensees a "safe harbor." For example, regulators could hold licensees harmless if they adopt a specified emission mask technology.

Controlling Out-of-band Interference: The Case for E_{max}

So far we have developed the electrospace spectrum rights model in a way that assumes that receivers can separate any two signals that differ in at least one of their electrospace dimensions. In practice, this means that interference arises because the receiver is not good enough. Unfortunately, a "sufficiently-good" receiver might be very complex and expensive. Thus, a trade-off arises between the benefit to licensees of allowing them to access the full range of rights in their electrospace volume and the possibly large costs to others of rejecting unwanted signals. Where parties can negotiate, the optimal approach is to establish clear property rights and allow the parties to resolve the interference problem themselves. Later in this section we discuss where we think this would work. However, in some cases the number of potentially affected parties is likely too large to allow efficient negotiations. Thus regulators may have to set rules that broadly strike a reasonable balance of interests, recognizing that they will have to update their policy given the rapid technological change in this field.

Out-of-band interference is a particular problem for receivers. If receiver technology was perfect, everyone would just have to worry about others trespassing on their electrospace volumes. Unfortunately, real-world (i.e. practical) receivers can experience interference even there's no signal at the tuned frequency. Strong signals at close-in frequencies or very strong signals at more distant frequencies can cause distortions called "out-of-band" interference, which Section 4 discussed. Policymakers could either ignore the practical problem of out-of-band interference and stick to a pure electrospace model or limit the power level of signals that could cause out-of-band interference. That is, they could limit signal power from all licensees *even within their licensed electrospace volumes*.

With current technology, the most cost effective regulatory approach in most cases is to supplement the ideal flexible-use rules we described above with a second rule, a limit on transmitter power customized to the band. Regulators would optimally choose the transmitter power limit (we'll call it E_{max}) within and across bands to maximize overall net benefits, balancing the benefits of less-expensive receivers with the disadvantages of limits on transmitter power. Although a blanket limit on transmitter power would not eliminate all out-of-band interference, it would generally confine interference to situations in which parties are likely to be able to negotiate resolution. This is the relatively unusual situations when the victim receiver is very close to a transmitter tuned to a nearby frequency.

This rule doesn't necessarily have to be a limit on maximum transmitter power per se. Out-of-band signals only matter if receivers are present, and the interference they cause is not only a function of total transmitter power but also of the transmitter's antenna pattern. So a rule to limit the out-of-band signals that are most likely to cause problems could limit signal field strength in certain specified locations where receivers are most likely to be, for example at ground level in public areas. This approach would allow much more flexibility for radio systems design than a strict transmitter power limit, and it would probably protect receivers better.[59]

The optimal design of an E_{max} rule is a direct function of receiver technology, and regulators should update it accordingly. The recent development of various receiver-on-a chip technologies have made receivers much smaller and cheaper, but not necessarily more resistant to strong unwanted signals. Future changes in receiver performance may result from any number of developments: much smarter receivers that can move to a better frequency or a better modulation; from receivers using digital RF or IF processing (where optimum bandpass filters can be synthesized); from room-temperature superconductors (producing very-narrow-band, very-high-Q, tunable RF filters that could reject many of the signals that otherwise would cause

59 Note that the interference to public safety LMR in the 800-MHz band in the US was caused in part by allowing apparently reasonable changes in antenna locations, without requiring changes in transmitter power. In this case, E_{max} limits at ground level would have provided better protection from interference, while allowing more flexibility in use.

out-of-band interference in today's receivers); or from adaptive antenna technology (that could null out strong unwanted signals). On the other hand, software-defined radios (SDRs) and cognitive radios (CRs) may use receivers that are inferior in some ways to current receivers. The requirement to operate in many different frequency bands may curtail the use of high-performance passive RF bandpass filters, increasing the susceptibility to out-of-band interference.

It would not generally be useful to vary E_{max} by time, space, or angle-of-arrival, nor should it be a direct function of the number of transmitters. Although more transmitter sites within a region can result in more total power radiated at a given frequency, out-of-band interference depends on strong unwanted signals irrespective of how many transmitters are involved. Since out-of-band interference depends on the signal strength within the very-wide-bandwidth electronic circuits at the receiver front end, it is likely that all of the energy from a transmitter contributes to the out-of-band interference. Therefore, E_{max} should not rise with transmitter bandwidth. Ideally, E_{max} would apply to the power from multiple transmitters that cumulatively could violate the field strength limit.

Interference resolution

Assuming licensees obey the two flexible-use rules (E_0 and E_{max}), receiver owners would be completely responsible for solving their own interference problems. "Victim" receiver owners have a number of options. First, they could try to show that a specific transmitter is violating one of the applicable flexible-use rules and pursue compliance. Second, they could improve their systems to eliminate the interference. Such changes might involve improving the victim receiver, increasing desired transmitter power, or adding better error correction. They could tolerate the interference, for example by changing operating procedures, restricting the operation to areas where interference is not a problem, or adapting to lower quality service. Finally, they could negotiate with the interferer to redraw the boundaries of their respective electrospace volumes, change interfering transmitter power, and/or seek or offer compensation.

6.2. The Problem of the Height Dimension

The spatial component of any electrospace region is a 3-dimensional solid volume, including height (or altitude), as well as latitude and longitude. Although it is straightforward to articulate these dimensions in a rights database, the way signals propagate at different heights poses a practical challenge to how to design the spectrum rights system. Most wireless applications' radios are close to ground level, carried by people or cars or fixed in homes and offices. For many of these applications, ground level radio signals are attenuated by buildings, terrain, and the earth's curvature. However, at higher altitudes radio waves encounter fewer obstacles than they do at ground level, so they travel further before they eventually decrease to E_0, the edge of an occupied electrospace region. This means that neighboring transmitters that don't encroach on each other at lower heights (meaning both signals attenuate to below E_0 at their ground-level geographic borders) could do so at higher elevations.

Figure 7 illustrates the predicted coverage areas for a single transmitter, where the respective coverage areas are defined by a common field strength, but differ by height above the ground (2m, 10m, and 50m). The large difference in these coverage areas shows that the geographical area in which signals remain below any particular level is quite dependent on the height at which the signal is measured.

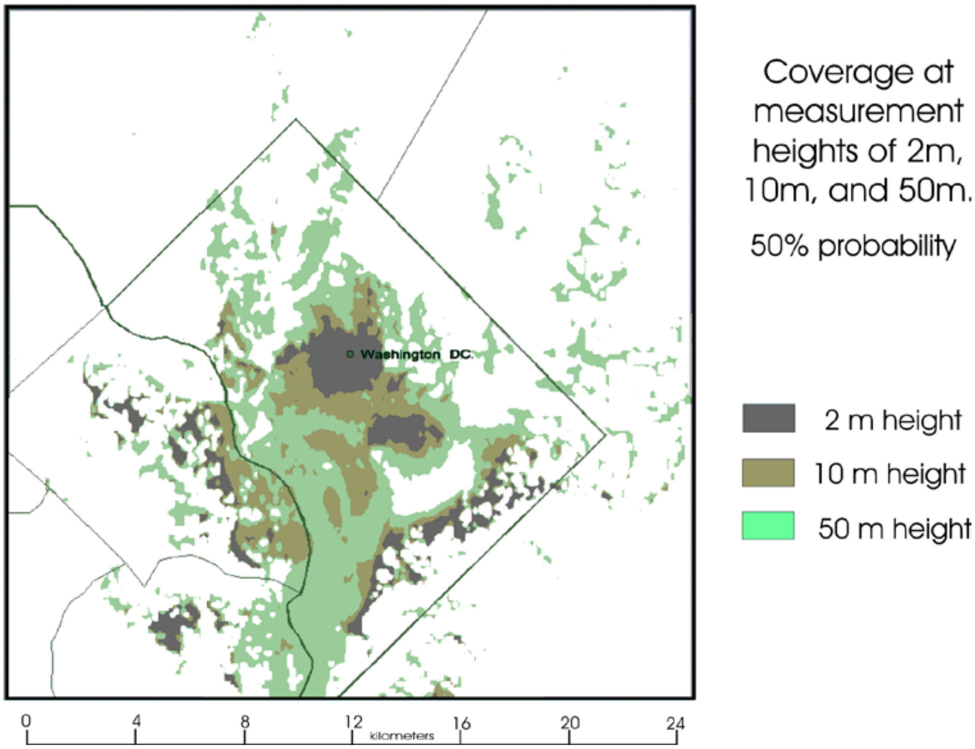

Figure 7. Coverage as a function of measurement height

One way regulators could handle this would be to set a standard measurement height (SMH).[60] Below the SMH, regulators would enforce an out-of-volume signal limit of E_0, and above the SMH they wouldn't. If most radio systems operate with antennas at about the same height above ground, regulators could set the SMH consistent with that height. However, flexible-use bands will likely contain a large variety of systems with antennas deployed over a wide range of heights. One possible reasonable (initial) value for an SMH could be 10 meters above ground. This would strike a compromise between lower heights where propagation is greatly affected by transient and small structures and greater heights that are well above most users and where measuring signals is more difficult. Setting an initial SMH could allow regulators to license non-overlapping, contiguous electrospace regions, and then side agreements between neighbors could adjust the electrospace volumes thereafter. All of this would need rich databases for propagation models to predict the occupied volumes and the necessary rights to go with them. Of course,

[60] This is also sometimes called the "statutory measurement height," but this is a little misleading since the value would be set by regulators, not in law.

regulators could insist that the electrospace regions not intersect at greater heights, but that would greatly restrict the highly valuable use of the electrospace regions on the ground.

A related approach would set the altitude boundaries of the licensed electrospace rights (LERs) with two height parameters. Regulators define the top of LERs, say at an altitude of 30 km, as the boundary below which no others can transmit; satellites and other non-terrestrial transmitters could operate above the top of the LER. The other height parameter governing LER boundaries is the SMH, below which regulators enforce the E_0 limit. Figure 8 illustrates this approach. It shows signals from a transmitter in one LER propagating into an adjacent LER via three paths and different heights. The owner of LER #2 must hold Path A to signal strength less than E_0 outside the pink area because Path A propagates below the SMH limit. The higher Path B signals do not need to be less than E_0 outside the pink region, but their strength is likely constrained since the Path A signal is constrained. Path C shows that the owner of LER #2 can legally put a receiver inside the blue area to receive transmissions from the pink area. That's because there are no limitations on the locations of receivers in this system. The owner of LER #2 could even put a narrow-beam, high-gain antenna to receive signals along Path A, although it would have to receive them at power levels under E_0.

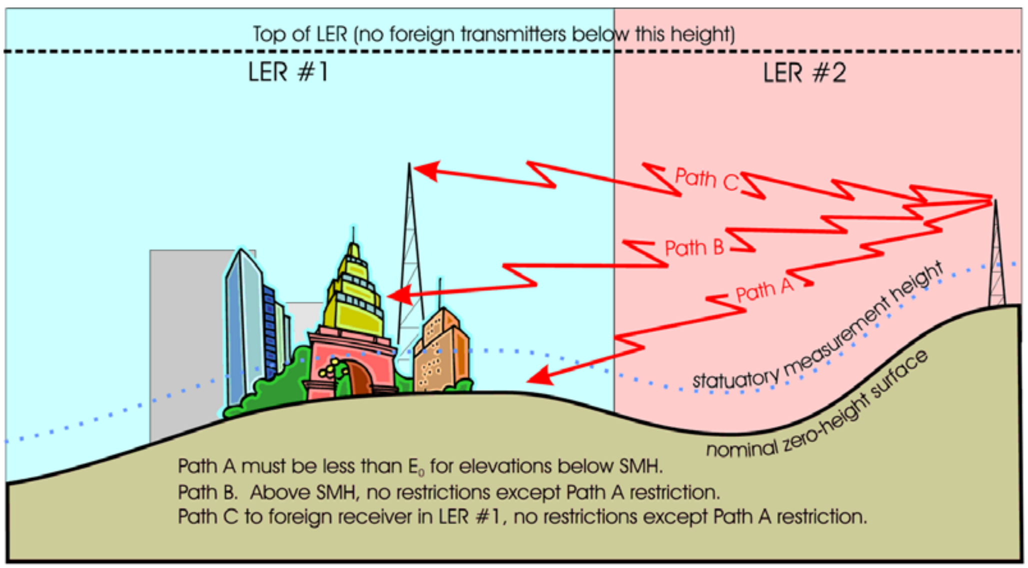

Figure 8. Flexible-use height considerations

An SMH of around 10 m could lead to interference for wireless users in tall buildings, where many overlapping distant signals might arrive at field strengths greater than E_0. In this case, wireless operators might need to acquire nearby electrospace regions at a given frequency in order to control that frequency at greater heights above ground. As in Path A in Figure 8, those who transmit from the top of tall buildings need to make sure their signals attenuate to below E_0 by the time they reach the SMH outside the footprint of their license. Conversely, licensees in a valley must ensure that their signals don't encroach on owners of rights on a mountaintop.

6.3. The Probabilistic Nature of Propagation

We noted above that signal propagation varies with changing conditions, so it's hard for users to constrain their signals to their electrospace volumes with total certainty. Although the ideal flexible-use rule states that the signal should *never* be greater than E_0 outside the LER, propagation models and vast experience show that unintentionally high signals may occur rarely at substantial distances. Through no fault of a spectrum user, airplanes, special atmospheric conditions, and other transient factors can produce a reflected line-of-sight radio signal greater than E_0. Such events are rare, but they exist and it seems reasonable to allow some de minimus exception to the E_0 limit.

An illustration helps make the case. Figure 9 below shows the same propagation model results as Figure 7, except that it shows results only at 10m height and the different colors show different proportions of a certain time period the signals are expected to exceed a specified strength threshold.

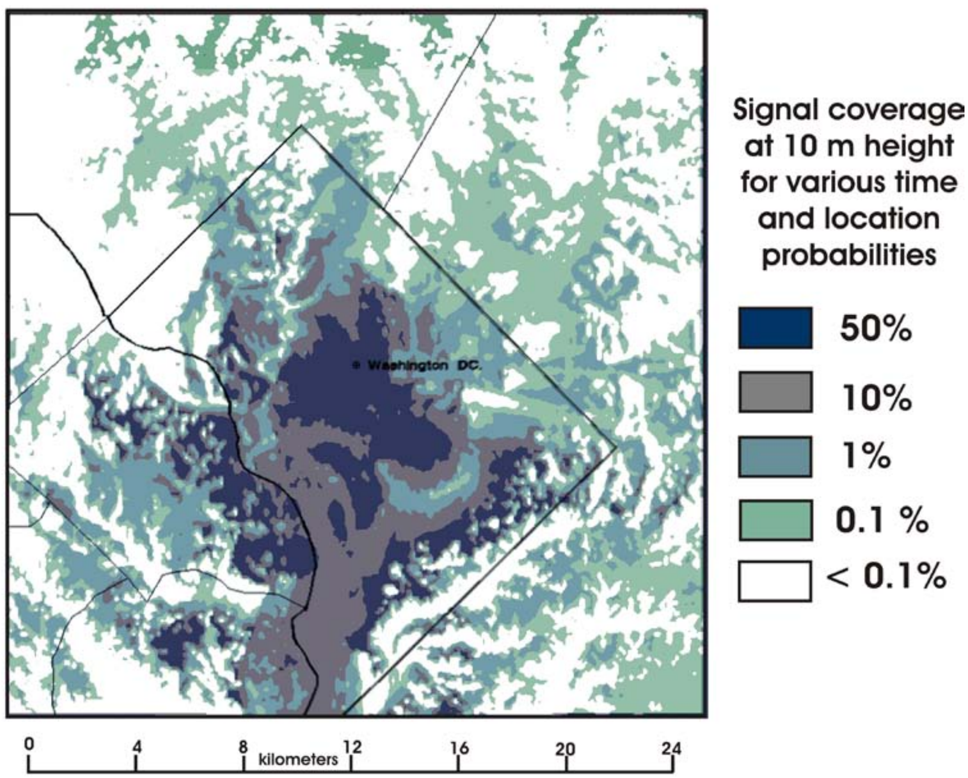

Figure 9. Coverage as a function of probability

Figure 9 shows the expected propagation of one transmitter's signal as a function of five different probabilities: 50%, 10%, 1%, 0.1%, and < 0.1%. A probability of 50% (the dark blue area) means that the signal is expected to be present at the selected height at least 50% of the time.[61] In the white areas on the graphs, the signal is received at the selected signal strength for less than 0.1% of the time. The coverage areas become much larger as the probability specification gets smaller. This is an "edge" problem of specifying the boundaries of the LER to cover propagation that may be ragged geographically and variable temporally. One partial solution to this problem is for regulators to license large electrospace regions, so the ratio of interior space to "edge" is larger. If the geographical size of the electrospace is sufficiently large, the edge problem dissipates greatly.

Another approach would be to specify the rule for E_0 probabilistically. In that approach LER owners must ensure that signals outside their LER be more than E_0 with a probability (p) no higher than, say, one percent. This would reasonably constrain the level of signal outside the LER, while allowing productive use inside the LER. The efficient value of the probability would take some analysis, as would the way in which it should be calculated, including the duration and area over which to average potentially excessive signals. Efficient enforcement proceedings would require such officially sanctioned methodologies.

7. DOCUMENTING AND ENFORCING SPECTRUM RIGHTS

7.1. Rights Database

LERs describe the "shape" of the 7-dimension electrospace volume within which a user can emit radio signals. One role of government in a flexible rights regime is to ensure that all LERs and any future modifications are accurately described in an up-to-date database. This database is vital for the government and users to be able to enforce rights. Some or all of this data should be public to allow licensees to identify parties who may be violating their rights and to help spectrum market participants more easily identify licensees with which to negotiate.

Assuming parties stay jointly within their electrospace rights, there should be no limit to bilateral or multilateral contracts among individual electrospace neighbors. Such agreements might allow one party to encroach on the rights of another, for example by emitting signals outside the geographic boundaries at powers over E_0. An agreement might require that interfering signals be suppressed at heights above the SMH or require suppression at powers different from E_0. On the other hand, the value for E_{max} could not be increased by private agreements, since this value protects many unrepresented users from possible increased out-of-band interference. Regulators would have to decide whether private contracts in spectrum rights should also be included in the public database, particularly if the agreement may result in higher levels of interference to a third party.

Two basic types of electrospace boundaries will be especially important in the spectrum

[61] The probability given for these predictions includes both location probability and time probability. The location probability reflects the statistical characteristics of the signal across a geographical area. The time probability concerns the changes over time in weather, multipath propagation, terrain, buildings, moving vehicles, and the like. For simplicity, Figure 9 sets both probabilities equal to the reported probability and averages the results over specified time periods and geographical areas.

management process: the boundaries of the LERs and the boundaries of radiated signals at various amplitudes (especially E_0). Let us consider how the boundaries of the electrospace volumes should be expressed in the database. Because these rights may be complex, including all seven electrospace dimensions, regulators must give considerable thought to their technical representation, including how the information can be made available for a range of public needs. New dynamic/opportunistic spectrum applications may need to query the rights database very quickly, so regulators may face a tradeoff between a very finely scaled map of boundaries and the ease with which one may determine whether a point falls inside or outside the boundaries. A map created by a propagation model with appropriately detailed geographical and physical databases (electronically represented) could draw the boundary of geographic rights. The most important signal boundaries for interference purposes will be the predicted signal levels from licensed transmitters at the SMH altitude, showing the geographical locations where the signal is greater than E_0 with probabilities greater than 1% (or whatever probability criteria is selected re the discussion in 6.3). This model will immediately identify areas of predicted signal encroachment into neighboring locations and frequencies.[62]

Regulators must decide the appropriate minimum range for each dimension of a licensed electrospace volume. LERs that are geographically too small may not provide an area of usable signals without encroaching on neighboring LERs, but small LERs may be necessary as part of larger sets of rights in areas where distant "islands" of higher signal strength may propagate.

7.2. Rights Protect Against Encroachment, Not Interference

Interference is importantly distinct from encroachment. "Encroachment" means that a signal is illegally present within an LER at a level higher than E_0. "Interference" means that an unwanted signal is degrading the performance of an operating radio system. Sometimes encroachment will cause interference, and sometimes it won't. Likewise, some interference will be the result of encroachment, and some will not. Under flexible-use rules, the transmitter operator is required to eliminate encroachment, while the receiver owner is responsible for addressing interference.

Recalling Figure 9, the uncertain and variable nature of radio propagation and the strong effects of terrain, structures, and other physical features could make it very difficult for licensees to construct their rights holdings to totally cover all their possible signal propagations. Some researchers have noted that if a licensee does not control all of the small bits and pieces of the rights their system uses, others who own those rights could hold licensees hostage. This potential "encroachment troll" problem would induce licensees to acquire and use their rights holdings very cautiously, including potentially foregoing beneficial uses near the boundaries of their LER. Hatfield and Weiser [2005] discuss the problematic properties of radio propagation and rightly note that much of the spectrum policy literature fails to grapple with how such

[62] Another approach could be to express the results of a propagation model with a mathematical solid model, which represents three dimensional geographic boundaries in a way that allows analysts to ensure easily that adjacent rights are non-overlapping. Such a mathematical approach would allow clear regulatory delineation of even the most complex propagation predictions.

properties can be squared with simple clear rights of access.[63] Given that a licensee's signals might on occasion inadvertently encroach into areas to which they do not hold rights, a challenge arises in crafting a system of rights that distinguishes between trivial encroachment and violation of enforceable rights. Exposing licensees operating in good faith to unexpected and costly litigation is problematic, but so is denying legitimate rights holders a way to redress encroaching signals.

Expressing the E_0 limitation in a statistical sense, as described in Section 6.3, is one way to help ensure that de minimis emissions above E_0 do not trigger litigation. Licensees could also protect themselves by controlling a sufficiently large geographic area within their electrospace volumes such that they lower the risk of signals spilling into others' territories. In addition, licensees could enter into private contracts with those upon whom they may potentially encroach so that they can resolve transgressions before the fact and obviate litigation.

In addition to ways private actors can minimize the risk of encroachment trolls, regulators could require those who wish to claim their rights are encroached on to show evidence that they are harmed or potentially harmed by the encroachment, similar to showing legal standing for a lawsuit. Regulators might require proof of interference to operational systems instead of mere encroachment as a minimum basis for damages. Another approach could be to require claimants to go through binding arbitration in which expert arbiters can expeditiously adjudicate rights and mitigate unreasonable hold ups. In any case, such rules must be crafted carefully so as not to blur the system of clear and enforceable rights that is fundamental to a functioning market for spectrum access.

7.3. **Licensed rights and transmitters**

Let us consider how the spectrum regulatory system should enforce compliance. In general, new or modified systems should be run through some kind of propagation model to see whether predicted signals would lie within the LERs of the system owner. Key questions for the regulatory system are who should undertake that analysis and what protections the results should afford the licensee. In addition, the question arises whether the process should allow an opportunity for public comment and *a priori* objection to the potential new signals. The policy challenge is to establish a system that balances the risks to incumbents of new uses with the benefits of flexible, dynamic spectrum rights. Strong protection from the risk of interference would help prevent costly damage to incumbent systems but could incur delays, add paperwork burdens, and possibly invite gaming by potential market rivals.

Here we consider four approaches to striking this tradeoff: a federal operating license, an expert third party certifier, a band manager, and a self-certifying approach. Other approaches are possible, but we offer these examples to illustrate a range of options. In all of these approaches we envision broadly available high-quality propagation analysis tools and the associated terrain and structure data bases. These tools, no matter who would be responsible for using them, would

[63] Weiser, Phil and Hatfield, Dale, "Policing the Spectrum Commons," Fordham Law Review, Vol. 74, No. 2, pp. 663-694, 2005.Available at SSRN: http://ssrn.com/abstract=704741

make the required coverage analysis relatively easy, technically neutral, and consistent across users. For example, electrospace neighbors could use the same modeling tools to predict their own system coverage and to scrutinize their neighbor's coverage. High quality propagation tools should obviate the role of expensive and time-consuming environmental measurements of field strength in the licensing process.

The first approach, establishing federal operating licenses, may provide the strongest protection from interference, but could incur the highest regulatory burden. It would require each licensee to obtain a separate operating license or other certification for each transmitter from the regulatory authorities before operation.[64] To allow owners of neighboring LERs to have a limited (but informed) opportunity to raise objections, the process could require disclosure of technical information about the licensee's intended transmitter and a brief public comment period. The disclosure would include sufficient information to assess potential encroachment (such as propagation model results), but exclude confidential business information (such as data rates and service quality). Public technical information about transmitters would hold system operators accountable as they design their systems and make it easier for authorities to identify culprits in encroachment situations. Lacking credible objections, the regulatory process would automatically grant operating rights.

The second approach substitutes third party expert certification in place of a federal operating license. This makes sense if federal authorities have no particular comparative advantage in undertaking the requisite engineering analysis. For example, Australia requires that all proposed transmitters be "certified" by specially-licensed engineers who check the system designs with suitable propagation models to give assure that the proposed signals would not encroach into non-licensed areas.[65] Certification largely protects licensees from challenge by other incumbents and encroachment trolls.

In the third approach, federal authorities would cede spectrum regulatory authority over relatively large geographic and frequency ranges to band managers. Such entities would devolve legal rights to access spectrum to individual users as the band manager sees fit. Band managers could tailor rules to prevent risks of interference and encroachment trolling to the environment in their bands. The relatively large LERs would ensure that most interference complaints will be among users who are mutually associated with the same band manager (instead of between different band managers, where interference control may be legally more complex). However, the question arises exactly what advantage band managers confer relative to federal regulators, who could presumably impose band-specific rules themselves. Band managers may provide more efficient rules if they have better information or are more nimble to respond to changing market or technical conditions.

[64] Licensing an individual transmitter before operation is a two-stage spectrum licensing process quite analogous to the real estate transactions. Owning an LER is like owning a parcel of land. A transmitter license is like a building permit for a specific structure on the land.

[65] "Flexible Radio Spectrum Access: Moving from device-centric to space-centric management." Michael Whittaker, FuturePace Solutions, 2006.

A fourth approach would require LER owners to test their pending systems with approved propagation models and retain records that show that the predicted signals lie entirely within their LERs. They could do this test themselves or hire a third party. In the event of encroachment disputes, the licensee would have to show the records indicating their compliance with the modeling requirements.

None of these approaches fully eliminates the risk of encroachment trolls or actual harmful interference. In addition, all of these approaches require possibly costly and expensive propagation analysis, although good analytical tools could reduce these transactions costs. Finally, all of these options raise the question of how to allow for dynamic secondary markets in flexible spectrum rights when different users operate systems with different transmitters and other technical characteristics. Indeed, such dynamic spectrum markets raise a lot of questions that we have not fully addressed here.

The optimal approach might be to adopt different rules in different frequency bands or other subdivisions of the electrospace. This would allow users to choose the regulatory environment best suited to their applications and tolerance for risk. However, any efforts to customize regulatory environments should recall that the whole objective in creating flexible-use spectrum rights is to avoid an inefficient, specialized, Balkanized, command-and-control spectrum management system.

7.4. **Implementing Flexible-Use Rights: Issues for Regulators and Users**

Issues for Regulators

A number of implementation issues arise, both for regulators and for users. For example, we have not addressed how flexible rights would evolve from the current system of command-and-control rules. Would flexible and tradition rights co-exist in a given band for a time, or would a disruptive conversion technique like band-clearing be needed? Likewise, although the flexible-use model we've outlined includes only a very small number of rules and limits, in most cases we have not offered numerical values for these limits. Setting these limits involves complex and difficult technical trade-offs that will require considerable analysis. For example, although LER owners could establish any rules inside their licensed electrospace region, they still need to observe the established E_0 and E_{max} limits, since these limits protect other LER owners. These limits might prevent some high power spectrum applications from being developed within flexible-use bands. This is probably useful, though, since high power applications need special regulatory attention.

In addition to the power limits within (i.e., E_{max}), regulators may need to place additional limits in certain frequency bands to allow certain kinds of applications. For example, most LMR and cellular/PCS services will benefit from duplex band architectures, where base station receiver frequencies are systematically separated from base station transmitter frequencies. In this important instance, a rigorous separation of base station transmit and receive frequencies would be very beneficial when even a single non-conforming frequency could greatly complicate base station implementation. Therefore, although “maximum-flexibility-of-use” remains a key principle, some applications may benefit from some limitations on flexibility.

One option for regulators is to apply the simple flexible-use rules to large scale LERs and let the LER owners manage access to their spectrum according to their own rules. This would allow flexible-use rules to serve as a mechanism to “allocate” the use of larger pieces of spectrum. Indeed, even if regulators don’t initially allocate large LERs, market activity could consolidate those rights. Owners of large LERs may, for example, divide their bands for different uses in rural and urban areas, reselling “franchised” portions of the electrospace subject to rules of the owners’ design. LER owners could support nearly any service, including a “semi-unlicensed” band with specialized rules to support a subscription-based mesh network service, a new urban mobile WiMAX band, and a new rural band using higher-power Wi-Fi with directional antennas to access a paid internet service, possibly with all of the bands having a real-time band manager to control real-time opportunistic spectrum access. The key here is that private owners of such bands have every incentive to maximize the economic value of the resource. Thus, regulators could make large LERs part of a broad strategy to shift spectrum from federal command-and-control regulation to a market-based system.

Regulators will necessarily have to adopt a more hands-off approach to interference resolution in the flexible use context we’ve described, and in some cases this could be difficult. The flexible-use system does not protect receivers from interference, and so some users will experience harmful interference and have no recourse. However, to preserve the full rights of LER owners and provide market certainty of those rights means that however tempting, regulators must not intervene when there is no encroachment. Because this is fundamentally different than command-and-control approaches, simply grafting flexible-use freedoms into established command-and-control bands may run afoul of incumbents’ expectations and create other implementation problems. In particular, anyone who benefited from the ambiguity of the command-and-control system of spectrum access could be made worse off when that ambiguity is eliminated, risking a protracted legal process as regulators attempt to clarify rights. Thus regulators must choose flexible use bands and transition approaches with care.

Issues for Users

One potential disadvantage of flexible use spectrum could be that what some view as “freedom” might be a “lack of needed guidance and prescribed practices” to another licensee. Users of flexible spectrum may need greater system design sophistication than they would in a command-and-control environment, and any lack of expertise might create more interference in a flexible-use system than it would in a command-and-control system.

The demands on receivers could be more intense in flexible-use bands. Receivers must withstand interference from a much wider variety of possible interferers, and this might require more expensive receivers and more conservative system design than would be necessary with only one potential type of interference. Out-of-band interference requirements might be particularly difficult to anticipate. Not only would there likely be many more types of systems at moderately spaced frequencies, but the radio environment will likely change more rapidly, possibly requiring ongoing adaptation.

The requirement that users suppress radiated signal levels below E_0 immediately outside their LERs means that many types of radio systems will need to leave a small unused buffer zone at the frequency edges of the LER. In flexible use bands, users won't know what type of radio system occupies the immediately adjacent frequencies. This is in contrast with command-and-control bands that provide technical details about the systems on adjacent frequencies and can therefore minimize the separation between channels. Thus, it is likely that N independent channels of a given service would require less total spectrum in a suitable command-and-control band than they would in a flexible-use band. However, the mere fact that command-and-control bands can be more technically efficient in some types of spectrum use doesn't tell us anything about whether they're more economically efficient. Moreover, the inefficiency from additional buffer space needed in flexible-use bands disappears when larger flexible-use bands are engineered for a common use. It may well be that the overall cost of the buffer spectrum necessary to prevent encroachment in a flexible use system is worth the value of that flexibility. Thus regulators must analyze the tradeoffs associated with creating flexible-use bands and as much as possible allow market forces to allocate the resource.

8. CONCLUSIONS AND NEXT STEPS

The inefficiencies inherent in the traditional command-and-control spectrum regulatory system are increasingly costly as demand for spectrum-dependent services explodes. This paper describes a conceptual framework to articulate clear rights of access to spectrum in a way that fosters a market-based allocation of the resource. We also offer simple rules that reasonably account for imperfect receivers and challenging physical properties of radiowaves. The key features of the system we propose are:

- Regulators construct an initial partition of spectrum rights across the dimensions of space, time, frequency, and direction of propagation. Each partition is called a licensed electrospace right (LER). Regulators devolve these rights to LER owners.

- Licensees may buy, sell, aggregate, and subdivide their LERs at will.

- Licensees must keep all signals within their respective LER, including its frequency band, geographical area, angle of propagation range, and authorized time of operation. In particular, all signals must have a power level of less than a regulated limit (E_0) outside the LER, with exceptions allowed with a probability no greater than an amount specified by regulators (such as one percent).

- Licensees must limit transmitter power or field strength within their LER to below a regulator-set level for the band in which they operate (E_{max}).

- Regulators or other parties must establish and maintain a detailed database and propagation model that facilitates transactions and enforcement.

In this system, regulators set up the rights database and establish a few core parameters for each band. Thereafter their role is limited to enforcing compliance with the simple set of rules on signal strength. Importantly, this system includes no protection of, or constraints on, receivers, so it does not directly control interference. Rather, through transactions and negotiations between LER owners, the system we outline here would induce an efficient level of interference in which the costs of controlling interference are balanced by the benefits.

The preceding discussion as background suggests an extensive research agenda. A logical next step would be to develop a detailed proposal for a flexible-use spectrum rights system, including identifying specific bands and developing values for E_0, E_{max}, SMH, and encroachment probabilities for them. Research is also necessary for policy options such as the initial partitioning of spectrum into LERs, specific ways to define, detect, and resolve E_{max} and encroachment problems, and the design of institutions and databases necessary for enforcement and secondary markets. A related exercise would explore how to induce incumbents in command-and-control bands to relinquish their rights so regulators can make more flexible-use spectrum available, for example through incentive auctions.