The Electrospace Model as a Frequency Management Tool

Robert J. Matheson[*]
NTIA/ITS.M
Phone: (303) 497-3293, fax: (303) 497-3680
matheson@its.bldrdoc.gov

In this tutorial paper, the electrospace is described as a theoretical hyperspace occupied by radio signals, which has dimensions of location, angle-of-arrival, frequency, time, and possibly others. Because these dimensions are independent, a given radio signal has a unique descriptor in the electrospace. Signals having different electrospace descriptors can theoretically be separated by a suitable receiver. The electrospace model provides a good framework to define spectrum user rights that divide licensed spectrum into parcels that can be flexibly used in an independent and non-interfering manner, while allowing complete freedom to divide and aggregate spectrum parcels via a secondary market. Disadvantages of the electrospace model are that it assumes ideal receivers and it allows the specification of spectrum parcels that cannot practically be used in the real world. Additional rules can be added to account for non-ideal receivers.

1. A General Description of Radio Systems

For many decades, radio regulators have realized that multiple radio systems could be used without interfering with each other if the radio systems were operated at different frequencies, or at different locations, etc. Ways to squeeze in more non-interfering radio users have been a major concern of radio regulators for almost a century, though the specific techniques have changed greatly with the technologies in use at various times. Formal concepts of how the radio spectrum could be divided among users included the development of the electrospace model by Hinchman in 1969. Others have developed similar models, synonymously using terms like "spectrum space" or "spectrum utilization." This tutorial paper describes electrospace concepts and their application to spectrum management environments based on flexible-use user rights.

All radio systems are characterized by a transmitter that emits a radio signal, a transmission path, and a receiver that examines the signal at the receiving location. In communications systems, the purpose of the system is to move information (e.g., TV programming, cellphone conversations, or credit card data) from the transmitting location to the receiving location. In the case of sensing systems like radars, the purpose is to compare the received signal with the transmitted signal to gain information about the transmission path (inferring the presence of airplanes, tornadoes, or hidden underground pipelines). Although the above characterization is useful, it is significantly incomplete, because it omits interactions between different radio systems.

The most common form of interaction between different radio systems is interference. "Interference" is defined in this paper as any degradation of radio system performance caused by the presence of extraneous radio signals. The radio system being degraded is known as the "victim" system. The source of the extraneous radio signal is known as the "interferer." All interference occurs within receivers; no interference occurs outside of receiver. The term "receiver" means the whole receiving system, including the receiving antenna(s). By definition, interference is a classic example of an "externalized cost." It is always done to somebody else–at their expense–often without the slightest awareness on the part of the interferer that someone else is being caused interference. As an externalized cost, mechanisms must be established to control it, since there is no benefit to the interferer to control it.

Lacking a specific regulatory limit, there may be only a vague opinion as to when interference

---

[*] This tutorial paper represents the author's understanding of electrospace concepts and should not be construed to reflect current or future NTIA policies or procedures.

becomes harmful. All interference is unwelcome, even if the only cost is to consume some of the system gain margin or forward error correction budget. The victim receiver is always culpable in cases of interference, since a sufficiently capable receiver (possibly outrageously complex and expensive) could always be constructed to operate adequately in the presence of any unwanted signal that was radiated from a different antenna than the desired signal. Any instance of interference is *prima facie* evidence that the owner of the receiver didn't provide a good enough receiver. There is no technical basis to say that some interference is caused by inadequate receivers, while other interference is not. Regulations that protect against interference operate by allowing interference-free service using simpler or less expensive receivers. One major function of spectrum management is to maximize the value of spectrum by designing frequency allocations so that less expensive equipment can be used effectively.

In a command-and-control regulatory environment, the regulatory agency takes responsibility for the whole design of radio systems, licensing only specific services, over given service areas, using specific technical parameters. Receiver performance can be strictly enforced to ensure that interference will be caused only by higher-than-expected unwanted signal amplitudes.

In a flexible-use environment, the user has wide latitude to choose the type of service provided and many technical parameters. It is sufficient to provide receiver users with an expectation (guarantee) concerning the maximum levels of unwanted signals. Depending on specific details of the user's operation, the receiving system could achieve rejection of unwanted signals using antenna gain patterns, frequency separation, bandpass filtering, receiver location, forward error correction, etc. In a given set of circumstances, some of these solutions would be much more cost effective than others. Economical industry solutions would presumably be developed for many standard circumstances.

2. The Electrospace

The electrospace is a formalized description of the radio signal environment, as it applies to all types of radio systems and all regulatory environments. It can be used to describe the ways in which the radio environment can be shared among multiple radio systems. It is particularly useful in flexible-use environments, where it provides a straightforward basis for aggregating and dividing multiple electrospace regions.

The electrospace describes radio signals, which means that it describes the domain of transmitters and transmission paths. The domain of receivers is totally separate from the electrospace description. Although any real radio system must consider all system components – i.e., the electrospace and the receiver– it is appropriate to divide the two components for regulatory purposes. The crucial regulatory difference is that the electrospace describes the ability of radio signals to cause interference to others–which is an externalized cost that must be strongly regulated. The receiver domain includes only system components that do not cause interference to others (i.e., no associated externalized costs that need to be controlled by regulations).

The use of an electrospace model in no way diminishes the role of receivers in the design and implementation of systems. In fact, it overtly recognizes the fundamental role that receivers play in the design of any radio system, and it completely frees the engineers and the businessmen to optimize receiver performance without any regulatory entanglements. In Section 6, various assumptions about the limitations of real receivers to reject unwanted signals will be used to modify electrospace management rules (discussed later), giving interference rights that are more appropriate and efficient for real-world use.

The electrospace describes the radio field strength at a given electrospace "location" that is defined by the 7 electrospace dimensions. These seven dimensions are all independent of each other, which means that the electrospace can be considered to be a 7-dimensional hyperspace. A "location" in the electrospace can be described by assigning specific values to several independent variables. It should be noted that different investigators have sometimes included somewhat different variables in the electrospace. The set shown in Table 1 is a useful starting point, and probably no great harm is done by including or omitting some marginal variables. The cases for two of these marginal variables–polarization and modulation–are discussed in the next section.

Table 1 - Electrospace Dimensions

Quantity	**Units**	**# Dim**
Frequency	kHz, MHz, or GHz	1
Time	seconds, hours, or years	1
Spatial location (geography)	latitude,longitude, elevation	3
Angle-of-arrival	azimuth, elevation angle	2

The physical location of a test point or hypothetical receiver is defined by the three spatial dimensions. The field strength at that location is described by the remaining variables, including the frequency, time of occurrence and direction of arrival. In a frequency band whose licensing is based on the electrospace, a numerical limit will typically be established, e.g., X FV/m, such that field strengths in excess of X are considered to be signals and are not permitted outside of the user's licensed regions of the electrospace. A given signal is said to occupy an electrospace "region" consisting of all locations in the 7-dimension hyperspace where field strength is greater than X.

One characteristic of the electrospace is that receivers can theoretically separate any radio signals that differ by at least one of their seven electrospace dimensions. For example, two co-located 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. Radio signals using the same frequency, operating time, and angle-of-arrival could be separated without interference if they were present at different locations.

3. Electrospace Dimensions

Frequency. The frequency dimension of the electrospace has the standard meanings of the word, namely a description of the frequency or range of frequencies (pass band) at which field strength is being characterized. Frequencies can be divided over a wide range of increments, typically matching the channelization of particular services.

Time. The time dimension can be subdivided over a wide range of increments. Useful time divisions might include the several-year duration of a licence, 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-ms time slot that would be available once every 20 ms, or transmit data during the vertical blanking interval of an NTSC television signal 30 times every second.

Spatial location. The spatial dimensions represent physical locations. They can be problematic, because there is no practical way to confine radio signals within a desired region. In typical hilly terrain, there are many distant locations that have higher signal amplitudes than many closer locations. Therefore, although one might easily select an arbitrary spatial region, the selected region might be extremely inconvenient to use efficiently. In order to prevent excessive signal levels (larger than "X") outside the boundaries of the selected spatial region, it might be necessary to greatly diminish signal amplitudes at many useful locations within the spatial boundaries. Transmitter power, details of the terrain, and the use of directional transmitting antennas are operative in establishing the spatial boundaries of the electrospace associated with a given transmitter.

A geographical area could be subdivided into smaller geographical regions or aggregated as part of a larger region, keeping in mind that radio waves propagate according to the laws of physics. Sub-dividing regions may have to be done carefully, so that the new users can actually make use of the new smaller regions, without creating or suffering interference with their neighbors. A useful technique might be to license the spatial dimensions by areas matching real coverage regions, determined in advance by measurements or modeling. For very-short-range systems, the geometry could possibly even be subdivided vertically. For example, the same frequency might be used on every tenth floor of a skyscraper (e.g., floors 7, 17, 27, ...).

Angle-of-arrival. This factor describes the angle-of-arrival or direction of radio signals at a given location, including the possible effect of multipath components scattered from many objects in many different directions from the receiver location. Note that this factor is not created by physical antenna pointing angles. The pointing direction of transmitting antennas primarily affects the spatial dimensions of the electrospace, i.e., the geographical areas where signals are strongest. No aspect of receivers–including the pointing angle of receiving antennas–ever has any effect on the electrospace. Therefore, neither transmitting nor receiving antennas influence the angle-of-arrival factor. On the other hand, receivers that exploit the angle-of-arrival will often have directional antennas.

In traditional radio systems, the useful angle-of-arrival is limited to the direct path between transmitter and receiver, e.g., point-to-point, free-space propagation between high-gain, narrow-beam antennas furnishes the technical basis for terrestrial microwave networks. The multipath signals arriving from other directions are often a major detriment to traditional receivers, causing frequency-selective fading and other problems. Directional receiving antennas can very efficiently exploit the electrospace direction-of-arrival dimension, separating out individual signals at the same frequency from multiple microwave towers or geostationary satellite orbital slots. This technique is easily scalable; additional angular subdivision is possible by using more-directional receiving antennas to separate out additional signals arriving from new transmitters at different angles at the receiving site.

If there were an easy way for a transmitter to send a signal to the vicinity of a distant receiver and make it appear that the signal came from a different direction than from the transmitter, the distant receiver could use a directional antenna to preferentially receive that signal. If the transmitter could somehow generate multiple independent signals in the vicinity of the receiver that appeared to be coming from different directions, the receiver could use multiple directional receiving antennas to receive multiple independent signals from the transmitter. This would remain true even if all of the multiple signals were at the same frequency. Thus, the direction-of-arrival dimension could (theoretically) be used to increase the traffic that could be carried to a given receiver–using the general principle that a receiver can separate any signals having different electrospace descriptions. Unfortunately, there is no practical method for a transmitter to generate multiple signals that appear to a distant receiver to be coming from different directions.

Recently developed Bell Labs Layered Space-Time (BLAST) technology exploits multipath reflections and multiple transmitting and receiving antennas to generate independent transmission channels, somewhat like the postulated multiple beams with different apparent angles-of-arrival described in the previous paragraph. Instead of using a combination of properly-phased antennas to produce directional beams, however, BLAST processing represents a much more generalized approach to the vector addition of multipath signals received on multiple omnidirectional antennas. Under certain conditions, the signals received on the multiple receiving antennas can be mathematically processed to separate the independent signals that were transmitted by a set of multiple transmitting antennas.BLAST technology uses a mathematical combination of multipath signals reaching the receiver site from different directions to synthesize multiple radio channels between a pair of transmitter and receiver sites.

Polarization. Some investigators have included polarization as one of the electrospace variables, since radio waves can be polarized using two orthogonal polarizations–e.g., vertical and horizontal or clockwise and counterclockwise. Many satellite systems transmit two separate signals, one on each polarization. However, a variable that has only two possible values seems to provide a very meager basis for a generalized dimension, so polarization was not included in Table 1.

Modulation. Some investigators have suggested that the electrospace should include a parameter for modulation. They reason that a suitable receiver could separate out many independent signals based on orthogonal coding, e.g., CDMA codes. The argument against using a "modulation" dimension in the electrospace is mainly a practical consideration, namely that there are no obvious codes/modulations that are orthogonal to all other general user codes. This prevents users from independently choosing a code and knowing that it will remain unaffected by signals from other users employing different codes and modulations. Even if all other users are carefully constrained to use codes belonging to a specific set of codes, the other users will substantially affect the available S/I ratios. Thus, even the best choices of codes or modulations currently available do not guarantee independence among the corresponding signals.

4. Spectrum Use Rights from the Electrospace

The electrospace model can be directly applied to a flexible-use, market-based frequency management environment, since the model describes a way that all aspects of the use of the radio spectrum can be unambiguously divided (shared) among multiple users. The only significant regulatory principle is that a licensee has the right to radiate a signal within a licensed electrospace region. Outside the region, signals must be kept below a specified field strength limit, X FV/m. There are no restrictions on type of service, transmitter power, bandwidth, modulation, antenna height, number of sites, etc., as long as the signal is kept lower than X in all regions outside the licensed electrospace boundaries.

An electrospace region is permitted unlimited aggregation or subdivision along all of its dimensions. This means that electrospace can be freely repackaged and resold as a market-based commodity, distributing spectrum without requiring approval by a regulator. A 1-MHz bandwidth could be subdivided into 40 channels of 25 kHz or augmented with additional adjacent frequencies to make a 5-MHz bandwidth. A given channel could be subdivided into TDMA time slots of 10 ms occurring every second and rented to a hundred separate transmitters. A statewide geographic coverage area could be divided into much smaller geographical cells and rented to short-range neighborhood wireless ISPs. Multiple fixed transmitters could be allowed to radiate signals into a common receiver location, if the transmitters are arranged to provide signals that have different angles-of–arrival.

Although the electrospace model is critically based on a field strength value, X FV/m, which cannot be exceeded outside the licensed region, it is not obvious what value to choose for X. Presumably X will be chosen so that systems licensed outside the region will not receive interference from the signal. However, the minimum level of interfering signal for various types of systems varies over a wide range–perhaps 50-60 dB–depending on the system. Since all types of systems are assumed to operate in an electrospace-based band, which type of system should X protect? One answer is that the selection of a specific value for X might make that band particularly suitable or unsuitable for various types of services; multiple bands could use different values of X to efficiently accommodate various services.

5. Practical Limitations on Electrospace

Although the electrospace model is conceptually powerful and potentially very useful, there are a few important problems with its application to the real world. The major problem, non-ideal receivers, will be discussed in the next section.

The division of the electrospace along any selected dimensions–while theoretically possible– may or may not produce a useful division in the real world. Arbitrary spatial regions, for example, may not match easily achievable propagation/coverage areas. A more useful spatial division technique may be to determine easily achievable coverage areas and divide the electrospace regions in a corresponding way. The angle-of-arrival dimensions may be compromised by unintended scattering from the terrain or by lack of sufficiently narrow beamwidth receiving antenna performance (especially at lower frequencies). Division into very narrow time slots may produce systems that are difficult to synchronize properly. Division into very narrow frequency slots may produce unreasonable requirements for frequency stability and Doppler shift.

The spatial dimensions pose some other problems. The field strength at a particular location is often the vector sum of many multipath signals. These multiple signals can occasionally add up to a field strength that is larger than the average field strength in the general vicinity. Therefore, it may be desirable that the field strength limit contain a statistical description, which would allow the occasional presence of signals above the limit. The inclusion of a statistical limit might make it much more difficult to show that a licensee had violated the electrospace limits, since a single instance of excess field strength might not be sufficient proof of a violation.

One obvious application of spatial coordinates is to describe licensed regions using latitude and longitude. For many applications, radio signals will be attenuated by buildings, terrain, and the earth's curvature, which tends to give the greatest attenuation at ground level. Raising a receiving antenna farther above ground will usually increase the received signal level. Therefore, a transmitted signal that meets the field strength limit at ground level may not meet that limit at higher elevations above ground. Many radio systems have receiving antennas located on tall buildings, towers, or mountaintops. Therefore, the success of a given application may depend on an understanding of how field strength changes with all three of the spatial electrospace dimensions.

The frequency dimension can also cause problems. Although a transmitter can radiate any amount of power inside the licensed frequency band, the field strength outside the licensed band must be less than X. Presumably this condition must be met at all locations–even very close to the transmitting antenna, where the in-band field strength is very high. To meet this condition near a transmitter may require unreasonable amounts of filtering. Therefore, the out-of-region limit, X, may need to be suspended in the immediate vicinity of the transmitter.

6. Receiver Effects in the Electrospace Model

The most serious limitation on the practical application of the electrospace model to flexible-use spectrum management is that the electrospace model assumes that all receivers are "ideal." In this context, "ideal" means that the receiver has infinite rejection of unwanted frequencies (i.e., frequencies outside of the intended receiver band-pass) and infinite dynamic range (strong out-of-band signals will not cause intermodulation products or gain compression). Unfortunately, none of the receivers that are actually available to users are ideal. Even worse, the most popular and rapidly growing class of receivers–handheld multi-band cellphones–are especially non-ideal, with performance constrained by small size, low cost, and limited battery power.

The important characteristic of real (i.e., non-ideal) receivers is that they can generate interference even when no interfering signal is actually present at the tuned receiver frequency. Strong signals at close-in frequencies or very strong signals at frequencies further away from the tuned frequency will cause receiver distortions that are seen as interference. Therefore, the electrospace model may need to be supplemented with additional rules, if real receivers are going to be protected from interference while keeping a market-based, flexible-use environment.

The additional rules are needed mainly to control the presence of strong signals at frequencies near the receiver tuned frequency. Receivers will benefit from frequency bands that are carefully engineered into duplex band architectures, where base station receive frequencies are systematically separated from base station transmit frequencies. Therefore, some frequencies may be designated for base or mobile use. Maximum limits on transmitted power must be observed, to protect nearby receivers, even when increased power would not violate the geographical boundaries of the license. It may be desirable to limit the maximum field strength at ground level (where the receivers usually are) in the vicinity of transmitters. When bandwidths are aggregated or divided, the maximum transmitter power should aggregate or divide proportionally to the bandwidth. It may be useful to specify limits on how power can be distributed within the aggregated bandwidth, so that the total transmitter power cannot be shifted to one edge of the aggregated bandwidth, posing an impossible adjacent-band rejection problem for a receiver at the adjacent frequency. These additional receiver-based rules can be added to the electrospace rules, while still maintaining almost complete market-based flexibility with well-defined rights.

7. The Role of Receivers

Although this paper is primarily about the electrospace–which by definition does not include receivers–it may be useful to discuss the role of receivers in an electrospace environment.

In the command-and-control model, the receiver's role is well defined, usually having a required performance that is based on the technology that was available when the regulations were originally defined. Typically, the receiver is designed to operate in the absence of co-channel interference, often even requiring reduced amounts of adjacent channel interference. Operation under these specifications is expected to guarantee a given level of performance over a licensed area of service. Because service is guaranteed, there are tight definitions of harmful interference and receiver performance specifications. When service is guaranteed, using a specified receiver, the presence of a licensed receiver can constrain the operation of additional nearby transmitters. Therefore, both receivers and transmitters may need to be licensed and regulated.

The role of the receiver is completely controlled by the receiver owner in bands regulated by electrospace-based flexible-use rules. Although the maximum permitted levels of interfering signals are known, almost nothing else is. Nothing in flexible-use bands can be construed to specify any particular type of service, level of performance, or service area. However, since a receiver cannot cause interference to anyone else except the receiver user, and the receiver user is the only person who will benefit or suffer from the selected receiver, the user is the only party motivated to make proper decisions about receiver performance. Therefore, the receiver user can be entrusted to make all decisions about optimizing receiver performance for the job at hand. In bands where receiver limitations have been included in the electrospace rules, the expected receiver signal environments would be presumed to be somewhat more benign, allowing the use of less-expensive receivers.

The preceding paragraph should not be construed to mean that mandatory receiver standards are never desirable in a flexible-use environment. User groups may need to establish minimum receiver performance standards to assure interoperability or other shared aspects of system performance. Consumer groups may want to set standards by which the performance of receivers can easily be judged and compared. Such groups should be able to set whatever receiver standards seem appropriate to them. The point is that such receiver standards would be a proper concern of these groups, not spectrum managers.

In the flexible-use model, interference is not prohibited; instead, it is a system parameter that can be consciously optimized by the user to serve a particular function. The receiver must work well enough to provide the needed service, but not better (because better receivers will cost more, but provide only additional unneeded benefits). Since the flexible-use environment provides a maximum guaranteed level of interfering signals from external users, the whole system (including receiver performance) can be realistically engineered and optimized.

8. Summary

The electrospace model has been shown to have direct application to a market-based, flexible use spectrum management environment. In this role, it provides licensees very great freedom in providing services and distributing spectrum, while unambiguously providing rules for preventing interference to other users. The electrospace model becomes even more useful for frequency management when it is modified to account for normal receiver limitations.

As a conceptual tool, the electrospace directly suggests various methods to squeeze additional communications capacity into a given frequency band, since one can imagine subdividing any of the electrospace dimensions to allow extra users. Modern cellular systems have provided huge initial gains in spectrum capacity by dividing the spatial dimensions to get smaller cells and much higher geographical frequency reuse. The use of cellular base station adaptive antennas further subdivides the spatial dimension by the beamwidth of a base station transmitting antenna array and exploits the different angles-of-arrival from individual mobile users.