Spectrum Markets:
Motivation, Challenges, and Implications

Randall Berry*, Michael L. Honig*, and Rakesh Vohra†
*EECS Department, Northwestern University, Evanston, IL 60202
{mh, rberry}@eecs.northwestern.edu
† CMS-EMS, Kellogg School of Management
Northwestern University, Evanston, IL 60208
{r-vohra}@northwestern.edu

Abstract

It is widely recognized that the centralized approach to spectrum management currently used in most countries has led to highly inefficient allocations. It is also recognized that more efficient allocations could be achieved through spectrum markets; however, most discussions have so far focused on secondary markets, which are managed by licensees. Here we take a more expansive view, and discuss some challenges and implications of implementing extensive spectrum markets across locations, time, and diverse sets of applications. The discussion is motivated by first examining the fundamental question: Is spectrum scarce or abundant? Given that spectrum is indeed scarce, and that spectrum property rights are appropriately defined, we speculate on the emergence of a two-tier market; the upper tier consists of spectrum owners that trade spectrum assets analogous to land rights, and the lower tier consists of spot markets for limited-duration rentals of spectrum assets from owners at particular locations. The changes such spectrum markets could bring to the provision of wireless services and wireless network design are discussed along with methods for addressing related interference management issues.

I. Introduction

The continued growth of wireless networks and services depends on the availability of adequate spectrum resources. Accelerating demand for those resources, due to the popularity of portable data-intensive wireless devices, are testing the limits of current commercial wireless networks, underscoring the need for changes in current spectrum allocations. This has prompted the Federal Communications Commission (FCC) in the United States to consider ways to increase the supply of spectrum allocated to broadband access and to introduce techniques for improving the utilization of existing allocations [1, Ch. 5].

Spectrum allocations generally fall into one of two categories: a *licensed* allocation gives exclusive use rights to the licensee, whereas an *unlicensed* allocation corresponds to the commons model in which the band can be shared by different applications and service providers [2]. Licensed spectrum typically carries restrictions on how it can be used, and is generally not transferable. Although these restrictions have been alleviated to some extent by the introduction of secondary spectrum markets [3], existing rules still inhibit the reallocation of spectrum to more efficient uses.

In contrast to the current "command and control" method for licensing spectrum, a spectrum *market* is based on a notion of spectrum property rights, which can be traded among buyers and sellers. The potential benefits of spectrum markets for increasing the efficiency of spectrum allocations is widely acknowledged. Thus far related discussions have focused on secondary markets, which allow service providers with licensed spectrum to lease their spectrum to other service providers. Transactions must be filed with the FCC for approval (which are automatic in some scenarios), introducing delays that increase transaction costs [3].

Here we reconsider the spectrum allocation problem without existing regulatory constraints. We start by providing general motivations for introducing spectrum markets. That is, a basic policy choice is whether to define and enforce spectrum property rights. From a social welfare point of view, this choice ultimately depends on whether spectrum is scarce, that is, if demand for it exceeds supply when it is free. If spectrum is abundant, then it can be made freely available (subject to appropriate power constraints), as in the commons model. If spectrum is scarce, then an allocation mechanism becomes necessary to mitigate interference and avoid a "tragedy of the commons."

To determine whether spectrum should be viewed as a scarce resource, we estimate in Section III the achievable rate per user assuming that spectrum assigned to non-government services between 150 MHz and 3 GHz is available for mobile broadband access. The calculation assumes a cellular infrastructure with a fixed density of Access Points (APs), and accounts for interference between adjacent cells using standard large-scale propagation models. Although the answer depends on assumptions concerning frequency reuse in different parts of the band, the power constraint, and the distance of the user from the cell boundary, we conclude that extensive spectrum sharing in the range considered (with a managed infrastructure) could provide a few Mbits/sec per user. While this is a relatively large number for many types of services, it is small enough that some distributed spectrum management is likely to be necessary to provide for a wide range of future services.

Spectrum markets are subsequently described along with implications for wireless services and networks. A challenge in creating such markets is how to define the spectrum assets being traded. While there has been a great deal of discussion about this in the legal literature (e.g., see [4]–[7]), here our emphasis is on how spectrum markets may affect the provision of wireless services. Specifically, we speculate that the distinction between owned and leased spectrum assets would give rise to a two-tier market: in the upper tier spectrum property rights at particular locations (APs) are traded among spectrum owners (as in a commodities market); in the lower tier spectrum owners rent or lease their spectrum to service providers at particular APs via a spot market run by spectrum brokers. (Spectrum spot markets have been previously proposed assuming a given supply of spectrum, e.g., assigned by a regulator [8]. A key difference here is that the supply of spectrum at the lower tier is determined by the upper-tier spectrum market.)

An important feature of two-tier spectrum markets is that the market for spectrum is separated from the market for wireless services. That would allow efficient and flexible allocation of spectrum, while lowering entry barriers for wireless service providers. We conclude with a discussion of related interference management issues, and implications for wireless system design.

II. The Motivation for Spectrum Markets

From an economic perspective a common objective of any resource allocation is to maximize efficiency, meaning the total utility (summed over all agents requesting the resource) derived from the allocation. Determining an efficient spectrum allocation is complicated by propagation characteristics, which can vary substantially across frequency and locations, and variations in application requirements (e.g., voice, internet, broadcast, emergency, etc). Hence the portion of the spectrum most suitable for a particular application can change over location and time. The relative value of the applications to consumers can also change across locations, times, and user groups. Moreover, the mapping of application requirements to spectrum depends on available technologies and their costs, which change over time. It is well known that properly designed markets are an effective approach for solving these types of problems. After highlighting the main problems with centralized (command and control) allocations, we compare market solutions with other proposed methods for making more efficient use of spectrum.

A. Problems with Centralized Allocations

Allocations of spectrum to different applications by government agencies, such as the FCC, are typically static, i.e., they apply for many years. Hence changes in traffic demands, potential applications, user preferences, and available technologies over time and locations have led to inefficient use of spectrum resources. *Dynamic* spectrum allocation that adapts to these variations over time- and geographic-scales of interest is, of course, extremely difficult to accomplish via a centralized allocation scheme due to the overwhelming amount of information and computation required.

These problems, based primarily on economic considerations, are not unique to spectrum. (They apply to land assets as well.) That prompted an early critique of the command and control model for spectrum allocation, and a proposal for spectrum property rights by the economist R. Coase [9]. More recently, spectrum markets have been proposed and discussed in [2], [4], [5], [10].

Inefficient spectrum use is one consequence of the current policy. A second is that it erects formidable entry barriers to the market for wireless services. This is due in part to the high degree of complementarity among spectrum licenses. To offer a wireless service over a broad coverage area, a potential entrant must acquire a package of associated spectrum licenses. If the service is tied to a given spectrum block, then the entrant must bid for that block across different geographic regions. Therefore, the value a provider obtains from a license is contingent on the bundle of licenses already owned. The resulting high cost of spectrum combined with the high infrastructure investment makes it difficult to enter the market on a small scale (e.g., within a small geographic region). Hence the current cellular market is confined to service providers that can make a huge initial investment (i.e., several billion dollars). The limited amount of competition means that service providers can potentially exert considerable influence over related markets (e.g., third-party hardware and software). (This also creates incentives for "rent-seeking" behavior [4].)

We consider instead a scenario in which spectrum is made available for sharing among many different applications across a large geographic region. Our main assumption is that the spectrum is partitioned into a set of spectrum assets that can be allocated among agents (service providers) at different locations. Spectrum markets might therefore be associated with a network of APs, which includes the current cellular infrastructure of base station towers. Each AP would have a set of particular spectrum assets, which are allocated among agents by a spectrum broker. Our main focus is on the scenario where spectrum is used to provide network access via these APs. Such connections could provide the commercial wireless services available today including voice, internet access, and broadcast radio/television.

B. Spectrum Sensing and Harvesting

In addition to the preceding criticisms of command and control allocations concerning *economic* efficiency, there are also the following engineering criticisms that pertain to *spectral* efficiency (i.e., bits per second per Hz):

1) Static assignments cannot exploit statistical multiplexing of traffic across different applications over shorter time scales. Even if the spectrum assignments are able to match *average* long-term demand, there are typically large fluctuations in demand, which lead to inefficient allocations over shorter time periods.
2) Centralized allocations often hinder the introduction of new technologies and services.

The first criticism can be addressed through the introduction of cognitive radios that seek out and exploit (or "harvest") idle bands in real-time [11]. This is the basis for the primary/secondary model for sharing vacant broadcast television bands (IEEE 802.22 standard). While schemes for harvesting idle spectrum would help to increase the spectral efficiency associated with particular bands, they do not address the previous issues concerned with static allocations of large blocks of spectrum to particular applications. Also, applications supported by spectrum harvesting are limited by the interference constraints with primary users of the band. Hence without the flexibility of reallocating this spectrum to higher-utility applications via a market, social costs due to inefficient allocations are still incurred.

Ideally, spectrum sensing and harvesting could be *combined* with spectrum markets. For example, the spectrum owner or licensee could negotiate a fixed usage fee for secondary users subject to acceptable interference constraints. (In contrast to the development of the IEEE 802.22 standard, interference constraints imposed on both secondary and primary users could vary substantially depending on the usage fee.) That would increase spectral efficiency while allowing markets to determine economically efficient allocations of spectrum to applications along with interference levels between primary and secondary users.

C. Markets and the Spectrum Commons

The motivation for spectrum markets is predicated on the assumption that spectrum is a scarce resource. That in turn depends on propagation characteristics, the transmitted power, which determines range and interference levels, and the nature of the traffic demands. For applications requiring *short-range* communications over links spanning no more than a few meters, there is an abundance of spectrum above 3 GHz that can be used. Interference is unlikely to be a major concern in these scenarios unless the density of wireless devices becomes very large. Hence the demand for short-range applications can be satisfied with a spectrum commons at high frequencies.

Longer-range communications (e.g., over 50 meters) requires lower frequencies, where interference becomes problematic. There are two primary concerns with using a spectrum commons for these types of applications. First, the propagation range becomes more difficult to predict at lower frequencies, since depending on the environment, signals may propagate much farther in certain directions than in other directions. This complicates interference management, especially without restrictions on where APs can be deployed. (To alleviate this concern, at lower frequencies the commons model could be combined with a cellular infrastructure, which restricts AP locations, or restricts the use of particular frequencies at certain locations. Subject to those constraints, spectrum could otherwise be freely available.)

The second more fundamental concern is that as the demand for wireless services grows, demand may eventually exceed supply, creating excessive interference and lowering overall utility. Although the rate calculation in the next section gives a rough indication of whether spectrum is scarce or abundant, ultimately the value of a particular spectrum asset can be determined only through a market. If spectrum is truly abundant, then the prices of all spectrum assets will fall to zero, in which case the spectrum market reduces to the commons model [5]. (In practice, the price may not be zero, but rather large enough to cover any costs required to police the spectrum for violations of power constraints.) Of course, in general prices of spectrum assets should vary across frequencies and locations according to variations in demand and interference levels.

The preceding discussion implies that the frequencies at which the spectrum market transitions to a commons model can be automatically determined by the spectrum market (see also [12]). Namely, at high enough frequencies the price of the spectrum assets should be zero, since the propagation range is highly confined, and therefore useful only to a small number of devices.

D. Spectral Efficiency, Cost, and the Supply Curve

The cost of providing a wireless service includes the cost of spectrum plus the cost of the associated devices and systems. Efficient use of spectrum should balance these two costs. This is reflected in the current design of wireless equipment and standards, which have been developed under the assumption that spectrum scarcity poses a major limitation on system capacity and revenues. The high cost of spectrum has led to the development of cellular standards with sophisticated (expensive) air interfaces, which attempt to maximize spectral efficiency.

Inexpensive spectrum, obtained through a spectrum market, would likely motivate the deployment of inexpensive devices and systems, which operate at much lower spectral efficiencies, compared with current cellular standards. (In addition, less expensive devices may not be able to support as wide a range of frequencies, further limiting the amount of spectrum that can be effectively shared.) In economic terms this implies that if a market is used to allocate spectrum, then the “supply” of spectrum is not inelastic.

(Here "spectrum supply" loosely refers to the amount of services that can be provided with a particular spectrum resource.) Even if the physical amount of spectrum available for a particular service is fixed, the spectrum supply effectively increases with the price since more expensive spectrum justifies the use of more expensive equipment that achieve higher spectral efficiencies. (Hence the same amount of physical spectrum can provide more services.)

This discussion is illustrated in Fig. VIII, which shows supply and demand for a particular (fictitious) spectrum resource versus price. (The meaning of "spectrum resource" is discussed in Section IV.) The supply curve has a positive slope, and an equilibrium price p^* is shown, which determines whether or not the spectrum should be used as a commons. If p^* is sufficiently small, then the transaction costs incurred from running a market exceed the cost of interference (or loss in utility) with a commons model. This is represented by the boundary $p = p_0$ in the figure. Since p^* is shown to the right of the boundary, a spectrum market is needed to coordinate spectrum usage, whereas if p^* were instead to the left of the boundary, the commons model would be more efficient.

III. Is Spectrum Scarce?

We now examine the assumption that spectrum is a scarce resource. This depends in part on regulatory and economic considerations, but is ultimately a technical problem of determining if network architectures can scale effectively as their applications grow [13]. Here we address this question for a specific network architecture, namely an infrastructure of APs, which represents the primary network architecture deployed today for commercial wireless services. (Moreover, this architecture has better scaling properties than alternatives, such as mesh networks, given a sufficient number of APs.)

Our objective is to give a rough estimate of what rates could be provided with more extensive spectrum sharing. We therefore assume that all spectrum between 150 MHz and 3 GHz is pooled for commercial services, excluding spectrum currently assigned for military and government use. The particular bands used in the calculation are shown in Table I. Note that broadcast television bands are included in this list. Demand for those services might be satisfied by a combining wire-line cable services with the wireless infrastructure assumed here. (Current broadcast services, such as television, could be offered as multicast services over this type of cellular architecture. That would make more efficient use of spectrum since the signal would be transmitted in a given cell only when someone in the cell requests it.)

We also assume that the APs are deployed with particular frequency reuse patterns for interference mitigation. This should give an optimistic indication of what rates are achievable with full coordination among service providers; a relatively low rate per user indicates that spectrum is scarce, and needs to be carefully managed, whereas a very high rate indicates that simple spectrum management schemes (e.g., the commons model) are likely to be adequate.

A. Cellular Model

To compute an achievable rate per user, a cellular topology with hexagonal cells is assumed over which both APs and users are uniformly spread with densities ρ_{ap} and ρ_u, respectively. Each AP serves the same number of users. Here we focus on the achievable rate for the *downlink*, i.e. communication from the AP to each user. We expect similar results for the uplink.

We make the following assumptions:

1) The entire set of available frequencies in Table I is quantized into 1 MHz pieces, which are allocated across the APs according to a standard frequency reuse pattern.
2) Each AP transmits with uniform power spectral density over the set of assigned channels.
3) Each AP applies Time-Division Multiple Access (TDMA) to multiplex users within the cell.
4) Interference from neighboring cells only is taken into account. Also, we assume that the signal attenuation is determined according to large-scale propagation models, and do not account for random fluctuations (fading).

For a particular 1 MHz channel at frequency f, the rate for a particular user at distance d from the AP is assumed to be the Shannon rate

$$R(d, f) = \frac{1}{n} \log\left(1 + \mathsf{SINR}(d, f)\right), \tag{1}$$

where $n = \rho_u/\rho_{ap}$ is the number of users per cell, and SINR is the Signal-to-Interference-Plus-Noise Ratio given by

$$\mathsf{SINR} = \frac{P_r(d, f)}{N_0 + \sum_{i \in \mathcal{I}} P_r(d_i, f)}. \tag{2}$$

Here, $P_r(d, f)$ is the received power for a user at distance d at frequency f, $\mathcal{I}$ is the set of interfering APs, and $P_r(d_i, f)$ is the received interference power from the i-th interfering AP.

This rate assumes optimal coding and delay-tolerant applications and so gives an optimistic estimate of the rate that can be obtained. (However, we ignore the possibility of using multiple antennas and cooperative techniques to increase the achievable rate per mobile.) In principle, we can account for channel variations (fading), practical coding schemes with delay constraints, interference from more distant cells, etc. by adding an appropriate margin to the SINR.

The achievable rate for a particular user depends upon where they are in the cell. The lowest rate corresponds to a user at the corner of a cell, as illustrated in Fig. 2. It is then straightforward to compute the SINR in (2) based on the hexagonal geometry. To determine the received power $P_r(d, f)$ we use Hata's outdoor propagation model for the frequency range 150 MHz to 1.5 GHz, and its extension to PCS for $f > 1.5$ GHz (see [14, Ch. 4]). The rate is then obtained by quantizing the spectrum bands shown in Table I into 1 MHz pieces, and summing the rate function over those bands. To account for losses expected in practice (e.g., due to other channel impairments) a 6 dB margin is subtracted from the SINR. Also, we assume a fixed power per unit area, i.e., the total power across the area covered does not scale with the density of APs. That also constrains the background interference level.

B. Achievable Rate Results

Results from the preceding calculation are shown in Fig. 3. For these plots the transmit power density is $P_t = -40$ dBm/Hz for all APs, the noise power spectral density is N_0 = -174 dBm/Hz, the base station antenna height is 30 m, and the receiver antenna height is 1 m.

Figure 3(a) shows achievable rates for the worst-case user at a corner point of the cell versus user density. Different curves are shown for different values of the cell radius r. For these plots the frequency reuse factor N is chosen such that the rate at each frequency is maximized. (Possible values are 1, 3, 4, or 7.) Hence the rate per user decreases as the cell radius increases. (This is not necessarily true if N is fixed, since the interference decreases with the cell radius.) As a specific example, for a large city like Chicago, which has a population density of approximately 4000 people per square km, the worst-case rate per user increases from about 0.3 to 2 Mbps as the cell radius shrinks from 500 m to 200 m. (For this example, the spectrum efficiency is about 0.7 bps/Hz per cell, which is less than that expected for Long-Term Evolution cellular systems. This is because the user is located on a cell boundary and we assume omni-directional antennas without cell sectorization.)

Figure 3(b) shows how the achievable rate varies with the distance from the AP. The achievable rate per cell at different distances from the AP is shown versus the cell radius. (The rate per user is then obtained by dividing this rate by the user density.) These results indicate that the rate increases by about 50% when moving from the edge of the cell to distance $r/2$, and more than doubles if the distance decreases to $r/4$.

C. Interpretation

The preceding results indicate that if a cellular infrastructure with cell radii less than 200 m has access to all of the bandwidth in Table I, then rates well above 1 Mbps could, in principle, be made continuously available to every member of a dense urban population. Furthermore, the achievable rate increases substantially with the density of APs, and with the fraction of inactive users (as opposed to assuming all users are active).

Since the range of rates indicated here are sufficient to support a wide range of near-term mobile services, one might conclude that simple allocation schemes, such as those based on a commons model, may be adequate for spectrum allocation. However, the rates shown in the preceding section are optimistic in that the availability of a managed infrastructure with coordinated frequency reuse has been assumed. Also, we have assumed an extreme case in which a large amount of spectrum currently assigned to many different applications is pooled for shared use. If less spectrum is available, the rates decrease accordingly.

As discussed in Section II-D, another reason why spectrum may still be scarce even with extensive sharing is that the additional spectrum would encourage the use of simpler systems having lower spectral efficiency, e.g., by using modulation and coding schemes that operate at rates well below the Shannon limit. Additional bandwidth also enables a reduction in transmit power and associated interference. Finally, although the rates reported here may seem large (especially when divided among a relatively small set of active users), it is possible that in the long-term new applications may arise that require rates on the order of (or beyond) what are indicated here.

Hence, we conclude that even with extensive spectrum sharing and with coordination of spectrum resources across APs, the demand for spectrum may exceed supply as users, applications, and systems proliferate. We therefore discuss how spectrum markets might be defined that achieve efficient allocations and benefit consumers of wireless services.

IV. The Challenge: Spectrum Property Rights

A basic requirement for any market is to define clearly the asset being traded. The purpose of property rights in spectrum is to limit the amount of interference an owner (or licensee) receives from transmitters operated by other owners (or licensees). The definition of these assets influences the technical constraints under which wireless networks operate, the valuation of spectrum by market participants, and the resulting market mechanism. This has prompted extensive discussions of how such rights should be defined in the legal literature [4]–[7], [9]. We briefly highlight some of the key issues, and subsequently propose a definition based on transmitted power mask that will be used as a basis for the subsequent discussion of market features.

From an economic point of view, the definition of a spectrum property right (or asset) should satisfy the following criteria:

1) It should be *clear* and *easily enforced.*

2) It should be *transparent*, meaning that it is straightforward to determine the benefit of ownership (or rental).
3) It should facilitate efficient allocations (e.g., avoid "natural" monopolies).
4) It should be *flexible*, so that it can be applied to different radio environments with varying traffic and propagation characteristics.

Satisfying all of these criteria is difficult due to the fundamental problem of interference management in wireless networks. For example, to satisfy the second criterion a property right would ideally guarantee the owner of a spectrum asset that received interference power from transmitters using other spectrum assets will not exceed a given value. Due to random propagation characteristics, such a right would be difficult if not impossible to enforce, failing the first criteria. Moreover, for many applications, attempting to approximate such a right (e.g., by enforcing large frequency re-use distances) would be overly conservative and lead to inefficient use of spectrum, failing the third property. (Such a conservative approach has been common practice and has contributed to the current inefficiencies in spectrum use.) Hence in practice the preceding criteria must be relaxed when defining spectrum property rights.

A. *Power Mask*

In its most general form a spectrum property right can be defined in terms of constraints on transmitted and/or received power over frequencies, time, and space. This definition should depend on the propagation environment, traffic characteristics, and application requirements, which determine "acceptable" interference levels. For example, power and interference constraints for a community broadcast type of service in a rural area are clearly different from those for a high-speed data service in a city center.

Attempting to divide regions of frequency/time/space into a set of spectrum property assets, which *a priori* accounts for all of the previous factors is clearly impractical. Rather, an initial (perhaps coarse) definition of spectrum asset can be provided, which is subsequently refined through negotiations among spectrum owners [7]. In general, there is a tradeoff between the *front-end* cost of defining an initial set of spectrum property rights to minimize potential interference, and the *back-end* cost of negotiating subsequent changes to those rights once a market is introduced [6]. A challenge is to provide a set of definitions for spectrum property rights that balance those costs.

Because of the difficulties associated with constraining received power, we will assume that a spectrum asset is defined in terms of a *transmitted* power mask, which limits radiated power in a particular band (and outside the band) in a particular geographic region for a given time duration. This can serve as the basis for the spectrum markets described in the next section, provided that the power masks can vary over locations, times, and frequencies. Specifically, the power mask may depend on factors such as the antenna height and the distance of the AP or mobile to the boundary of the given geographic area. (Such restrictions currently exist in some bands.) Similarly, the power limit may increase with frequency (due to higher attenuation) and also increase when a neighboring system is expected to be lightly loaded.

In practice, the variations of spectrum power masks over time/frequency/space should be negotiated by neighboring service providers and spectrum owners, depending on their intended applications. (Rather than change the definition of power mask, the negotiations could instead settle upon monetary compensation for high interference levels.) Defining spectrum property rights in this way provides great flexibility, since the spectrum assets can be adapted to the environment and applications, and can evolve according to user demands and changes in technology. However, it does not give hard guarantees about received interference. Instead constraining transmit power would provide reasonable (statistical) expectations about neighboring interference levels. Moreover, if a spectrum owner or service provider requires a stronger guarantee about the level of interference, it would have the option of acquiring the neighboring spectrum assets to prevent other transmitters from using them.

B. Owning versus Leasing

For the spectrum markets to be described there is an important distinction between owning and renting (or leasing) a spectrum asset. The definition of the spectrum asset depends on this distinction. Namely, an *owned* spectrum asset has a long (perhaps unlimited) duration, and is traded as a land right. Spectrum assets can be *rented* or leased by the spectrum owner. The duration of the spectrum asset being rented can vary across frequencies, agents, and locations, and determines market dynamics. A short duration (say, less than a day) may be associated with a spot market for short-term commercial use (analogous to electricity markets [15]), whereas a long duration (e.g., years) may be associated with broadcast services that require continual use of spectrum. Note, in particular, that a spectrum owner could conceivably decide to switch applications (e.g., migrate from broadcast to cellular), or sell spectrum rights to another owner once a rental agreement has expired.

V. Two-Tier Spectrum Market

Allowing spectrum property rights to be flexibly defined and traded would produce major changes in markets for wireless services. The most visible of those changes would be the separation of spectrum ownership from the provision of wireless services. One reason a service provider may prefer to rent or lease spectrum, rather than trade it as an owner, is that leasing carries relatively low transaction costs compared to buying and selling the spectrum asset. This becomes especially important for services which require intermittent use of spectrum. (Examples may include emergency or monitoring, and delay-tolerant applications, such as video downloads, that can exploit periods of light usage.) A second reason is that leasing avoids “maintenance” costs, such as negotiating power levels with neighboring spectrum owners along with associated policing functions. Hence leasing or renting spectrum on a short-term basis allows for flexible and efficient allocations that vary over time and locations.

The distinction between owned and rented spectrum assets would lead to the two-tier spectrum market shown in Fig. VIII. The upper tier consists of spectrum owners that buy and sell spectrum rights with unlimited duration. Owners could choose to rent or lease their spectrum to service providers at the lower tier market through a spectrum broker, which manages spectrum assets at particular locations. The upper- and lower-tier markets would likely operate at two different time-scales: owned spectrum assets at the upper tier might be traded on a relatively slow time scale (months or years), whereas rented spectrum assets at the lower tier could be negotiated over short time scales (e.g., hours or even minutes, depending on the application).

We next describe additional features of two-tier spectrum markets. They are organized as properties associated with the spectrum owners, service providers and the spectrum broker.

1) Spectrum Owners: Spectrum assets at a particular AP, or geographic region, would be traded according to a conventional market mechanism, as in a commodity market. An issue, which arises with a spectrum asset, is that its value depends on the interference generated by nearby mobiles and APs, which can change over time. This may encourage aggregation of spectrum assets across neighboring locations. Strong interference between nearby APs with different owners may have to be resolved through additional negotiations.

Spectrum owners would have an incentive to rent their assets to service providers with applications that generate the most revenue. (That could, of course, vary with the time of day.) Allowing owners to trade spectrum assets implies that each asset could be reassigned to applications that generate higher revenue, or alternatively, to groups that want to purchase spectrum for non-commercial purposes (e.g., community broadcast). Furthermore, with extensive spectrum sharing, as assumed in Section III, many spectrum assets would be available at each AP and the transaction costs for trading spectrum would presumably be low. Hence the way spectrum is used would be determined by market supply and demand, and the price of a particular spectrum asset would be tied to the long-term expected revenue it is expected to generate.

2) Service Providers: A service provider offers a set of wireless services to end customers through a particular pricing scheme. With spectrum markets a service provider could purchase (rent) spectrum on a short-term basis. As a consequence, a service provider need not build out a national footprint of APs, which use the same spectrum. The spectrum could be rented via the spectrum spot market at desired locations according to customer demand. This, of course, assumes that the customers have frequency-agile radios that are able to switch to the assigned band and use the appropriate modulation and coding format. (Notifying the end-user what particular band to use would also require some signaling overhead, although that is likely to incur a relatively small cost.)

The service provider could also conceivably rent the necessary equipment at an AP from an equipment manufacturer. (That cost would also account for the cost of the tower on which it is mounted.) Hence the combination of spectrum markets and equipment rentals could dramatically lower the entry (sunk) costs for a service provider, potentially increasing competition along with service options.

Service providers may also provide an arbitrage function for customers. Namely, large fluctuations in the demand for particular wireless services may cause large price fluctuations for spectrum rental. Since end customers typically prefer predictable (e.g., flat-rate) pricing plans, a service provider may provide such an option, but with a premium, which accounts for statistical fluctuations in the price of spectrum. (Alternatively, the arbitrage function may be performed by third-party resellers.) A service provider may also choose to negotiate longer-term contracts for spectrum with the spectrum owner to provide more reliable Quality of Service. (Another possibility is to create a spectrum "futures" market, analogous to current electricity futures markets [15], in which rights to use particular spectrum assets at future times are traded.)

3) Spectrum Broker: The lower-tier spot market for spectrum assets at each AP could be managed by a spectrum broker, which determines how spectrum assets are allocated among service providers, and how much each service provider pays for each spectrum asset. The allocation method, or mechanism, must balance efficiency with complexity, a topic discussed in the literature on *algorithmic mechanism design* [16].

For example, the allocation could be determined through an auction mechanism in which the broker collects bids to buy from the service providers, bids to sell from the spectrum owners, and subsequently determines the allocation along with the price for each spectrum asset. The auction would then be repeated as spectrum assets become available (i.e., as they are released by service providers).

Alternatively, the spectrum broker could announce a set of prices for the available spectrum assets, and adjust the prices over time to maximize expected revenue or to clear the market periodically. This approach is generally simpler, and requires less overhead (information exchange) than an auction mechanism. However, a well-designed auction mechanism can achieve either a higher efficiency or more revenue (whichever is the objective). The choice between these two approaches should depend on the "thickness" of the market; with relatively few buyers and sellers (a "thin" market) an auction mechanism becomes simple to implement, so may be preferred. With many buyers and sellers the loss in efficiency (or revenue) with the pricing scheme becomes small, so that the pricing scheme may be preferred.

With either approach the protocol for information exchange (bids for assets or price adjustments) could be automated and run on a spectrum server. (See also [8], which proposes a related type of spectrum server.) Hence the lower-tier market for renting spectrum assets could operate on a very fast time scale with small transaction costs.

VI. Interference Management

For the spectrum markets considered here interference management at a basic level is accomplished by the power masks corresponding to spectrum property rights. However, fixed power masks (as defined today) would lead to inefficiencies due to changes in the deployment of wireless networks (e.g., density of APs), demand for spectrum across time and locations, and propagation characteristics. Hence depending on the location and time, some power masks may be too stringent (lowering the value of the services provided), and some may permit excessive interference (lowering the value of services at neighboring locations). We next describe ways in which spectrum markets may address this issue. Much of the following discussion applies to both owners and service providers, which we will refer to as agents.

A. Local Cooperation

Rather than fixing power masks across time and locations, power masks could be *adapted* through negotiations, or by means of particular protocols. For example, adjacent owners may negotiate cross-rental agreements for the same spectrum, or alternatively, agree to *cross-payments* for reducing or increasing interference. Specifically, an agent A at a particular AP could offer to pay agent B at a neighboring AP to reduce interference by imposing a smaller power mask on its customers. Alternatively, agent A could pay B to accept more interference, allowing A to increase its power mask.

To maximize efficiency (i.e., total utility for both service providers), the cross-payment for interference should match the loss in utility (or externality) incurred by the neighboring agent that agrees to reduce its power, or accept more interference. To find this cross-payment, the agents would need to exchange information about their utilities and cross-channel gains. This may be worthwhile, provided that the signaling overhead is manageable (i.e., occurs on a sufficiently slow time scale), and the agents are willing to exchange this information. For example, that might be the case if the same agent is managing spectrum assets across neighboring APs, as in current cellular systems. (The APs may then exchange "interference prices" to adjust powers, taking into account externalities due to interference [17].)

Competing agents at neighboring APs may not be willing to exchange information about actual utilities (e.g., expected revenue), but could negotiate cross-payments through a bargaining procedure. Moreover, there would be the possibility of negotiating a *protocol* for adapting powers that specifies how power limits and cross-payments are determined as a function of measured interference at particular locations. In this way, power masks can adapt to changing application requirements, network topology (e.g., if an agent wishes to add or remove an AP), and anticipated traffic.

We also point out that local cooperation can be applied to the commons model (e.g., 802.11 networks). An example of this, which requires minimal signaling overhead, is distributed dynamic channel allocation. As the traffic and network load increase, more extensive local cooperation, such as interference pricing and cross-payments, might be introduced. For example, an AP may offer to carry traffic from neighboring locations, rather than accept the additional interference from another AP. (The effect of such negotiations on the density of APs is considered in [18].) Referring to Fig. VIII, by mitigating interference through local cooperation, the threshold price p_0 increases, thereby extending the range of frequencies over which the commons model is more efficient than a spectrum market.

B. Asset Aggregation

To avoid negotiating interference levels with neighboring agents, an agent may try to purchase or rent similar spectrum assets at neighboring locations. In that way, the agent would have more control over interference within a particular region. (It may then be willing to allow more interference at the boundaries from different APs.) Also, the agent may wish to purchase or rent similar spectrum assets at neighboring APs to ensure adequate coverage. An agent's valuation for "bundles" of spectrum would then exhibit *complementarities*, i.e., the value an agent places on a particular asset may be greater if the agent owns (or rents) neighboring assets. (More generally, the value of a bundle of assets may be greater than the sum of the values of each asset alone.)

The existence of complementarities implies that spectrum markets should allow spectrum assets to be *aggregated*. Agents could then bid for bundles of assets across different locations. (The power constraints within the region could be relaxed provided that the total received power at the boundary stays the same.) This complicates the design of spectrum markets, since a particular asset may appear in many bundles, each containing a different combination of assets. Finding an efficient allocation in general can be computationally difficult and may require excessive information exchange (bids and asks) [19]. These costs could be reduced by adopting a less efficient mechanism.

Finally, complementarities for spectrum assets may also exist over adjacent frequency bands. For example, an agent may wish to own or rent those bands to manage adjacent-channel interference, e.g. to allow for simpler receiver filters. Hence spectrum assets might be bundled across both frequency and spatial locations.

VII. Implications for Wireless System Design

Because spectrum has been viewed as a scarce resource, wireless systems engineering has put a premium on spectral efficiency (bits/sec/Hz). As spectrum becomes more abundant through extensive sharing, the importance of spectral efficiency may diminish. Instead, the design objective with extensive sharing may shift toward lowering equipment cost and/or increasing power efficiency. That is, power could be reduced to provide longer battery life and reduce interference. Hence an abundance of spectrum may encourage the use of less expensive, low-power, wideband (spread spectrum) systems.

Another consequence of more abundant spectrum is that the economic benefit from spectrum use becomes limited by transaction costs for repeated spectrum (re)allocations. Hence an objective should be to minimize those costs. This suggests developing standards for broker mechanisms for the lower-tier spot market that can be included as a core part of cognitive radios (in addition to standard air interfaces). Service providers could then build applications, which exploit standard dynamic protocols for acquiring spectrum.

Alternatively, the commons model might be used in some bands with distributed interference management schemes, such as interference pricing and local negotiations. Those schemes effectively *introduce* transaction costs to mitigate interference, which may be needed as the density of nodes increases. The choice between these two approaches (market or commons) would again depend on the price of spectrum and associated transaction costs.

VIII. Conclusions

Allowing large parts of the radio spectrum to be traded and rented across geographic locations and time would provide incentives for more efficient use, and encourage alternative models for dynamically sharing spectrum. A key consequence is that spectrum ownership could be separated from the provision of wireless services. That would lower entry barriers, and thereby facilitate the introduction of more diverse sets of services. More abundant spectrum may also motivate the deployment of different types of radio systems, which operate at lower spectral efficiencies.

We have highlighted some of the main technical features and issues associated with spectrum markets. Additional issues not discussed include networking functions, such as handoff, and spectrum policing to enforce property rights. Perhaps more difficult to resolve are the policy issues associated with the transition away from current allocations. It remains to be seen whether the potential benefits of spectrum markets can overcome the obstacles to the necessary spectrum policy reforms.

Acknowledgment: The authors thank Hang Zhou for generating the plots in Figs. 3(a)-3(b), and Junjik Bae, Eyal Beigman, Hongxia Shen, and Hang Zhou for their contributions to an earlier version of this paper presented at DySPAN 2008. This work was supported by the U.S. National Science Foundation under grants CNS-0519935 and CNS-0905407.

References

[1] Federal Communications Commission, "The national broadband plan," March 2010. [Online]. Available: http://www.broadband.gov/plan/5-spectrum

[2] J. M. Peha, "Approaches to spectrum sharing," *IEEE Communications Magazine*, vol. 43, no. 2, pp. 10–12, 2005.

[3] FCC, "The development of secondary markets - report and order and further notice of proposed rule making," Federal Communications Commission Report 03-113, 2003.

[4] T. W. Hazlett, "The wireless craze, the unlimited bandwidth myth, the spectrum auction faux pas, and the punchline to Ronald Coase's 'big joke': An essay on airwave allocation policy," *Harvard Journal of Law & Technology*, vol. 14, no. 2, pp. 335–567, 2001.

[5] G. R. Faulhaber and D. J. Farber, "Spectrum management: Property rights, markets, and the commons," *Rethinking rights and regulations: institutional responses to new communication technologies*, pp. 193–226, 2003.

[6] P. J. Weiser and D. Hatfield, "Spectrum policy reform and the next frontier of property rights," *George Mason Law Review*, vol. 15, p. 549, 2007.

[7] T. W. Hazlett, "A law & economics approach to spectrum property rights: A response to Weiser and Hatfield," *George Mason Law Review*, vol. 15, no. 4, 2008.

[8] O. Ileri and N. Mandayam, "Dynamic spectrum access models: toward an engineering perspective in the spectrum debate," *IEEE Communications Magazine*, vol. 46, no. 1, pp. 153–160, January 2008.

[9] R. H. Coase, "The Federal Communications Commission," *The Journal of Law and Economics*, vol. 2, no. 1, p. 1, 1959.

[10] E. R. Kwerel and J. R. Williams, "Moving toward a market for spectrum," *Regulation: The Review of Business and Government*, 1993.

[11] Q. Zhao and B. M. Sadler, "A survey of dynamic spectrum access," *IEEE Signal Processing Magazine*, vol. 24, no. 3, pp. 79–89, May 2007.

[12] M. Cave and W. Webb, "Spectrum licensing and spectrum commons: where to draw the line," in *International Workshop on Wireless Communication Policies and Prospects: A Global Perspective*, 2004.

[13] D. P. Reed, "How wireless networks scale: the illusion of spectrum scarcity," in *International Symposium On Advanced Radio Technologies (ISART)*, 2002.

[14] T. S. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Prentice Hall, 1996.

[15] S. Hunt and G. Shuttleworth, *Competition and Choice in Electricity*. John Wiley & Sons, May 1996.

[16] N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, Eds., *Algorithmic Game Theory*. Cambridge Univ. Pr., 2007.

[17] J. Huang, R. Berry, and M. Honig, "Distributed interference compensation for wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 5, pp. 1074–1084, May 2006.

[18] J. Bae, E. Beigman, R. Berry, M. Honig, and R. Vohra, "Incentives and resource sharing in spectrum commons," *IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN 2008)*, pp. 1–10, Oct. 2008.

[19] H. Zhou, R. Berry, M. Honig, and R. Vohra, "Complementarities in spectrum markets," in *Allerton Conference on Communications, Control and Computing*, Monticello, IL, 2009.

Randall Berry [S'93-M'00] (rberry@eecs.northwestern.edu) received the B.S. degree in Electrical Engineering from the University of Missouri-Rolla in 1993 and the M.S. and PhD degrees in Electrical Engineering and Computer Science from the Massachusetts Institute of Technology in 1996 and 2000, respectively. He is currently an associate professor in the Department of Electrical Engineering and Computer Science at Northwestern University. In 1998 he was on the technical staff at MIT Lincoln Laboratory in the Advanced Networks Group. He is the recipient of a 2003 NSF CAREER award.

Michael L. Honig [F'97] (mh@eecs.northwestern.edu) received the B.S. degree in electrical engineering from Stanford University in 1977, and the M.S. and Ph.D. degrees in electrical engineering from the University of California, Berkeley, in 1978 and 1981, respectively. He subsequently joined Bell Laboratories in Holmdel, NJ, and in 1983 he moved to the Systems Principles Research Division at Bellcore. Since fall 1994, he has been at Northwestern University where he is a Professor in the Electrical and Computer Engineering Department. He is the recipient of a Humboldt research award for senior U.S. scientists, and the co-recipient of the 2002 IEEE Communications Society and Information Theory Society joint paper award and the 2010 IEEE Marconi prize paper award.

Rakesh Vohra (r-vohra@kellogg.northwestern.edu) was educated at University College London, the London School of Economics and the University of Maryland, where he received a Ph.D. in applied mathematics. He is presently the John L. & Helen Kellogg Professor of Managerial Economics and Decision Sciences at Northwesterns Kellogg School of Management. He also serves as the Director of the Center for Mathematical Studies in Economics and Management Science. His primary research interests are auction theory and game theory.

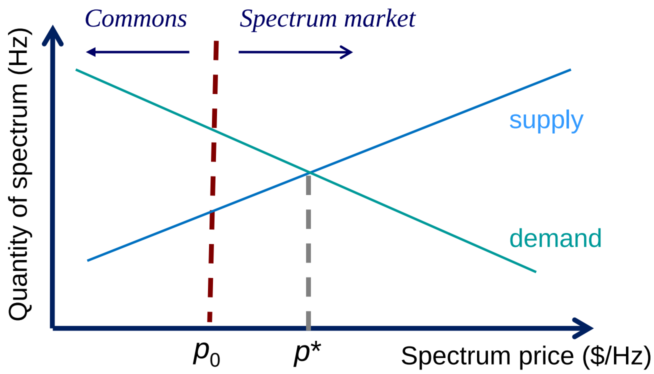


Fig. 1. Comparison of equilibrium spectrum price p^* with the transition price p_0. For the case shown $p^* > p_0$ implies that a spectrum market is more efficient than a commons model.

TABLE I

FREQUENCIES USED TO CALCULATE ACHIEVABLE RATES.

Broadcasting TV (total: 348 MHz)	174-216 MHz, 470-608 MHz, 614-764 MHz, 776-794 MHz
Fixed, Mobile, Satellite, Amateur (total: 669.7625 MHz)	150.8-157.0375 MHz, 157.1875-162.0125 MHz, 173.2-173.4 MHz, 450-460 MHz, 764-776 MHz, 794-902 MHz, 928-932 MHz, 935-941 MHz, 944-960 MHz, 1390-1395 MHz, 1427-1429 MHz, 1850-2025 MHz, 2110-2200 MHz, 2300-2310 MHz, 2385-2417 MHz, 2450-2483.5 MHz, 2500-2655 MHz
The table is based on U.S. Frequency Allocation Table as of October 2003, and includes all non-Federal Government exclusive spectrum between 150 MHz and 3 GHz. The total bandwidth shown in the table is 1.018 GHz.	

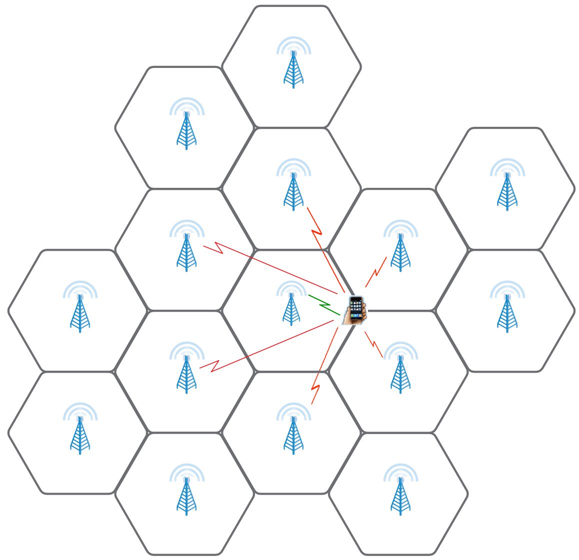
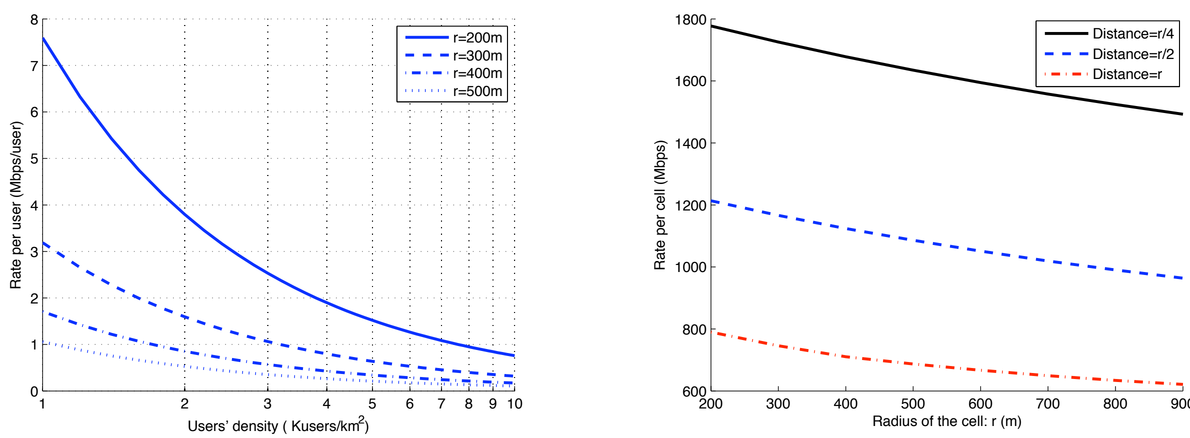


Fig. 2. Cellular system used to calculate achievable rates. The location of the worst-case user with the lowest rate is shown. Only interference from the neighboring cells is taken into account.



(a) rate/user vs ρ_u

(b) rate/cell vs radius

Fig. 3. Achievable rates with a cellular infrastructure assuming all frequencies shown in Table I are available for sharing: (a) worst-case rate per user versus user density; (b) rate per cell versus cell radius.

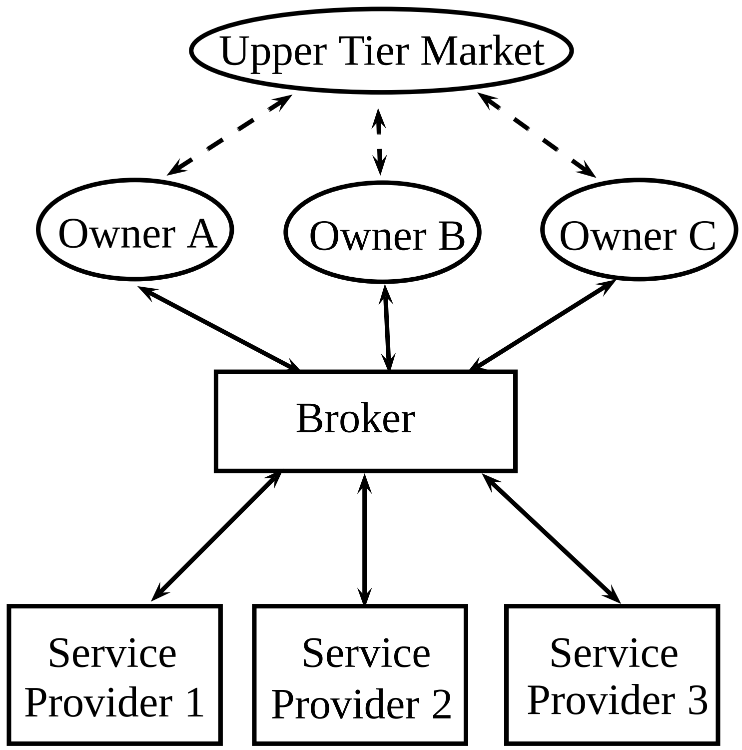


Fig. 4. Two-tier spectrum market: Spectrum assets corresponding to particular locations are traded by owners at the upper tier; those assets are then rented or leased to service providers via lower-tier spot markets at the APs. The spot market at a particular AP (or set of APs) is managed by a spectrum broker.