Seeing the bigger picture: context-aware regulations

Kate Harrison and Anant Sahai Wireless Foundations, EECS, UC Berkeley {harriska, sahai}@eecs.berkeley.edu

Abstract-The TV whitespaces represent an incredible opportunity for innovation. Along with this opportunity come significant challenges: the whitespaces are very heterogeneous in terms of channel availability as well as quality. This poses a challenge to anyone wishing to achieve uniform or even similar quality of service nationwide. In this paper, we consider using heterogeneity in the emissions limits to counteract the inherent heterogeneity of the whitespaces, ultimately resulting in a more homogeneous quality of service for secondary users. However, heterogeneity cannot be added to regulations in a haphazard manner. Rather, it must be carefully crafted as a result of looking at the bigger picture and fully exploiting the capabilities of databases. Databases should not be seen as a necessary evil but rather as an exciting opportunity for improvement and innovation. In particular, rather than being viewed as a simple repository of data, the "database" can be viewed as a cloudbased entity that reports the result of a default kind of Coasianbargaining that could be expected to occur among frequencyagile radios. We conclude by showing a few small examples of positive heterogeneity as it applies to real-world data for the TV whitespaces in parts of the United States.

I. INTRODUCTION

In 2008, the FCC approved regulations allowing wireless transmissions by unlicensed devices within the spectrum reserved for over-the-air TV broadcasts¹ [3]. The regulations define a *protected region* surrounding each TV tower (a.k.a. primary transmitter, in reference to its priority) inside of which TV reception is theoretically guaranteed. Unlicensed devices (a.k.a. secondary transmitters) may transmit once they are safely outside the protected region but they are subjected to a maximum power constraint intended to keep aggregate interference at the primary's receivers (i.e. television sets) at acceptable levels. Since the rules are location-dependent, a secondary transmitter must² contact a database to determine which channels are available for use at its location. The databases are operated by neutral third parties and issue coordinating commands to the secondary transmitters.

The 2008 FCC rulemaking was groundbreaking. It brought to the table a new method of spectrum utilization which provokes many interesting technological and regulatory questions. It will revolutionize the use of spectrum as surely as did the original spectrum assignment that first took place in the late 1920s. Not only do the whitespaces open up large portions of viable spectrum but they also hint at novel and interesting ways to think about spectrum regulations in this new era.

The dynamic nature of the whitespaces made databases all but necessary. We already know from previous studies that the use of databases drastically increases the whitespace opportunity [4]. This is a revolutionary way of thinking of wireless spectrum because we can now *depend* on having ubiquitous Internet connectivity through which to contact a database to determine critical parameters for establishing a physical-layer connection. This dependence simply was not conceivable two decades ago.

A. The power of databases

Discussions of whitespaces typically revolve around the technical challenges of building frequency-agile and cognitive radios and ensuring that they operate safely. However, the TV whitespaces have introduced a far more exciting element into the wireless regulatory ecosystem: databases. The whitespaces represent our chance to fully understand the power of *dynamic* spectrum access: rules which are aware of the bigger picture and evolve over time without the need to re-certify devices. Databases also give us a tool by which we can adapt regulations after deployment if they are found to be unsafe or too conservative [5]. With the rise of cloud computing, software-as-a-service providers can now rapidly test deploy new features; databases provide the analogous functionality in the whitespaces and spectrum management generally³.

Furthermore, databases can be used to shift trust away from the devices and to the databases themselves. No longer does the regulator have to trust that all the cognitive radios will correctly interoperate to carry out a distributed inference regarding a newly updated piece of policy language that they have downloaded [6]. This is accomplished by computing the implications of the current policies in the database itself and issuing only basic commands to the devices (e.g. allowed frequencies and emissions limits). This simultaneously makes certification both simpler and more robust, thus increasing assurances that whitespace devices will operate safely⁴.

The President's Council of Advisors on Science and Technology (PCAST) recently recognized that the database approach easily and naturally scales. Implicit in their proposal

³Under the current paradigm, devices are certified to comply with a particular set of rules, e.g. a maximum emission limit. Traditionally these rules have been determined at certification time and were implemented on the device. However, databases give us the opportunity to certify only that a device complies with a set of simple instructions and allow the database to compute the relevant instructions given the current rules. This would mean that policy updates could happen on a much shorter time scale as they would not require any coordination with the devices, only the databases.

⁴This also makes foreign operation trivial: devices need not be aware of the underlying policy engine and only the basic interface must be standardized.

¹These rules were subsequently updated in 2010 [1] and again in 2012 [2].

²The regulations include provisions for sensing-only devices; however, we do not consider such devices in this work.

for use of federal spectrum is the idea that databases (and the related administrative software) can be trusted more than the devices themselves and that databases should do the heavy lifting with regards to policy implementation:

The heart of the proposed SAS is a database that holds information about what spectrum is occupied for a a given location and time; the parameters of the signal, such as power and bandwidth; constraints for specific locations, such as no transmission in blasting zones or along international boarders; and the price for accessing the spectrum. The Radio Access Coordination and Management and Optimization function provides frequency assignments and authorizations. It may work to optimize overall spectrum efficiency over a given region, but above all will insure that legacy Federal retain priority access to spectrum. [7, §2.4]

The authors of [5] have discussed some of the potential that databases hold (e.g. incorporating rules which allow secondary transmitters to increase their power as they increase their distance from the primary receivers) but few have truly explored the implications of this technology. Databases mean that regulators are no longer tied to rules made at auction-time: we can now have truly dynamic spectrum access. Furthermore, rules can change with location, time, and other variables. This functionality has been partially explored by the FCC in that they require secondary transmitters to reveal their location to the database in order to obtain a list of channels that are available at that location, but there is no reason or need to stop there. We advocated strongly in [8] that it is imperative to also include some sort of information regarding safe transmit power levels which vary with location: any one-size-fits-all rule would be either far too conservative or far too dangerous.

There are many ways to choose safe power levels; the real trick is to find *good* power levels which foster creativity and innovation in new devices. It is clear that these power-level rules should be application-agnostic and that they should support nationwide-applicable business models. With only these constraints in mind, we can already see a multitude of problems: unlike typical allocations, the whitespaces are incredibly heterogeneous in nature. This is a large obstacle in terms of nationwide coverage because it means that we need to overcome the drastic heterogeneity in the whitespaces in a way that results in a relatively homogeneous quality of service nationwide. Of course, spatial and temporal heterogeneity is a generic feature of the new world of spectrum sharing.

B. Heterogeneity in the whitespaces

The heterogeneity in the whitespaces is two-fold: protected regions near primaries cause black-out regions for secondary transitters *and* signals from the primary transmitters practically need to be treated as noise by the secondaries. This heterogeneity is not an unknown phenomenon: all authors of whitespace papers are aware of this reality and a few have attempted to quantify it using real-world data [4], [9]–[12] for the United States and European countries. We see from Figure

1 that the distribution of the number of whitespace channels in the United States has a long tail, indicating large amounts of heterogeneity. This poses a challenge for anyone wishing to offer services that can scale to nationwide coverage.

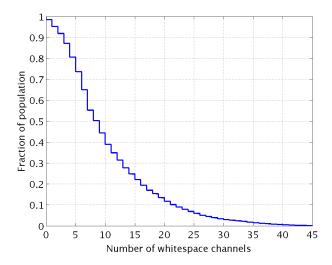


Fig. 1. Complementary cumulative distribution function (CCDF) by population of the number of channels available to secondary devices in the United States (adjacent channel exclusions included).

Some authors attempt to counteract this heterogeneity via intelligent channel allocation algorithms for use within secondary systems [13]–[15]. For example, [14] considers link allocations in the presence of channel heterogeneity (both in quantity and in reward). These authors recognize that it is important to avoid self-interference among secondaries and therefore often advocate against frequency-reuse-1 schemes. Self-interference is often modeled through an "interference radius": transmitters that are inside the interference radius cannot operate on the same frequency without effectively jamming one another's transmission.

One can imagine using similar rudimentary interference models to understand the interaction among primaries and secondaries in the whitespaces. For example, one could use the existing frameworks but alter them slightly by declaring a second class of "secondaries" (i.e. primaries) which are automatically allocated any spectrum they request. Indeed, this seems to be the approach of the FCC⁵.

However, secondary-to-primary aggregate interference has been shown to be a significant effect [8], [16]. Intuitively, aggregate interference matters in the primary-and-secondary situation — but not as much in the secondary-and-secondary situation — for the following two reasons:

⁵From the FCC's 2008 ruling in which they develop the separation distance requirements: "In developing the table of separation distances, we believe it is desirable to minimize complexity for compliance. In this regard, we have balanced this goal of simplicity with the need to provide assurance that TV services will be adequately protected. ... We find that a transmit antenna at a height of 30 meters transmitting with 4 watts EIRP could cause co-channel interference to a TV receiver with an antenna 10 meters above ground at a distance of 14.4 kilometers and adjacent channel interference at 0.74 kilometers." [3, ¶181]

- The distances between secondary transmitters are generally small which means that the aggregate interference felt by a transmitter is dominated by his nearest neighbor. On the other hand, primary receivers are somewhat far from the secondary transmitters and therefore the aggregate interference to the primary is not dominated by any single secondary transmitter.
- 2) Unlicensed devices are designed to have resilience to fluctuating noise levels. The primary network was designed specifically to have to deal with only its own interference and therefore we cannot assume that it is tolerant of high noise levels.

C. Existing heterogeneity in the rules

The 2008 version of the FCC regulations [3] included provisions for licensed wireless microphones which would allow operators to register the wireless microphones as temporary primaries. This would allow operators of such licensed devices to reserve spectrum for events utilizing wireless microphones. However, Carlson Wireless, Motorola, and WISPA (Wireless Internet Service Providers Association) argued that there should be a portion of the whitespaces reserved for unlicensed wireless microphones at all times and at all locations $[1, \P 26]$. As a result, the FCC amended its regulations in 2010 with additional provisions for wireless microphones [1]. However, adding nationwide spectrum for the wireless microphones is not as simple as defining a few channels as off-limits to secondaries regardless of primary presence. Any reasonablysized set of channels would have regions of unavailability due to the presence of TV transmitters (which clearly have priority) and thus would not suffice. Instead, it was necessary to introduce rules which vary based on location:

All TVBDs [TV bands devices, a.k.a. secondary devices] are permitted to operate [in] available channels ... subject to the interference protection requirements in 15.711 and 15.712, except that operation of TVBDs is prohibited on the first channel above and the first channel below TV channel 37 (608-614 MHz) that are available, i.e., not occupied by an authorized service. If a channel is not available both above and below channel 37, operation is prohibited on the first two channels nearest to channel 37. These channels will be identified and protected in the TV bands database(s). [1, §15.707a]

This rule essentially reserves the first two non-primary-inhabited channels above and below channel 37 for wireless microphones. Because of the variability of whitespaces due to the variety of primary locations and channels, the specific channel chosen will vary depending on location. For example, in Berkeley these two channels are 15 and 16 whereas near Chicago they are 24 and 41. From this example, we see that the FCC has recognized the need for heterogeneity in order to provide homogeneous quality of service. It is necessary to use databases to implement these rules but the same databases can also be used to implement other heterogeneous regulations.

D. Context-aware rules

The FCC's provisions for wireless microphones are an excellent example of what we will call a *frequency-aware rule*: that is, a rule whose outcome for a given channel depends on the characteristics of other channels available at a particular location. The remaining rules for secondary devices in the TV whitespaces are minimally frequency-aware: while cochannel exclusions are independent, adjacent-channel exclusions⁶ do induce some dependence. This means that these rules are frequency-aware in only a limited way.

Rules can also be spatially aware. We gave examples of such rules⁷ in [8] where we considered methods for scaling the maximum power limit based on the power limits in other locations. The power limits were coupled across locations because we enforced an aggregate interference constraint. Indeed any rule that has such a constraint will be spatially aware: there is a finite "budget" (the amount of interference a primary receiver will tolerate) and each transmitter uses a nonzero portion of the budget which affects the decisions made for devices at other locations. In contrast, the FCC regulations are *not* spatially aware because the transmit power of one device does not affect the permissible transmit power of another device; however, these regulations are locationaware. In this paper we differentiate between location-aware and spatially-aware. The former refers to the cognizance of a device's own location whereas the latter refers to the awareness of the potential locations of all devices.

E. Frequency agility

In our work, we assume that devices do not have a preference for a particular channel or set of channels: devices using the spectrum for data communications will view channels as substitutable. Furthermore, the amount of variability in the whitespaces means that any device which has a narrow operating range will not be well-suited to take advantage of whitespaces regardless of the rules. A service without a wide range of acceptable frequencies will suffer from preventable blackouts. In this way frequency agility is essentially a requirement for any marketable device operating in the whitespaces. Any service which cannot be frequency agile (e.g. radio astronomy) should be allocated its own band.

Furthermore, we assume that devices do not have a preference for a contiguous block of channels. Technology has advanced to the point where frequency agility is not uncommon. In most cases, devices will not transmit on all available channels. For example, consider the case in which there are many secondary devices at the same location. Each device will receive a fraction of the total rate available at that location and each device will be able to achieve its rate using only a subset of the available channels.

⁶Cochannel exclusions refer to the ban on transmission using channel *c* when a device is located inside the protected region of any primary operating on channel *c*. Adjacent-channel exclusions are similar but consider primaries operating in adjacent frequency bands.

⁷Technically these rules were also somewhat frequency-aware because we enforced adjacent-channel exclusions as well.

Rules	Spatial awareness	Frequency awareness
Current FCC regulations	None: rules do not depend on activity in other locations	Minimal*
Power scaling rules**	Devices obey aggregate interference constraints which induces location dependency	None: channels are treated independently
SRASC method***	 Devices obey aggregate interference constraints Secondary quality-of-service guarantee for all locations 	Locations may opt to leave fallow available channels to improve the systemwide utility

^{*}The specific channels chosen for wireless microphone exclusions depend on which other channels are available at that location.

Adjacent-channel exclusions also create short-range inter-channel dependencies.

Fig. 2. Comparison of context-awareness in proposed rules. "Power scaling rules" refer the rules presented in [8].

F. Overview of paper

In this paper, we argue that good rules will need to be both frequency- and spatially-aware. To make our point, we show that either of these qualities alone is insufficient and that in combination they are quite potent. We will begin with two motivating examples which show the various consequences of regulations and contrast these outcomes with that of a hypothetical secondary market. We then formulate and discuss several optimization problems which represent the potential goals of regulations. Then we test our hypotheses using a simple one-dimensional world but real TV tower data [17]. Finally, we conclude by showing that in our model it is important to be both frequency- and spatially-aware and we quantify the gains yielded by these two qualities.

II. MOTIVATING EXAMPLES

In this section we will discuss two examples which motivate our argument for context-aware rules. The model for each example is shown in Figures 3 and 4, respectively. There are two channels, red and blue, and primary transmitters on each of these channels. The protected region⁸ is represented by colored circles around the primary transmitters and the no-talk region⁹ (where applicable) is shown with black semi-circles. Secondary transmitters are shown in discrete locations as green dots whose size indicates their relative power. We assume that a secondary transmitter will use all available channels (i.e. channels with nonzero secondary power). For simplicity, we do not consider the effects of self-interference among secondaries. Here we do not consider adjacent-channel or microphone-related exclusions.

Each of the subfigures shows the power allocation chosen under different regulations. The first follows the current FCC regulations. The second follows a power scaling rule such as that in [8]. The third is a rudimentary context-aware rule — termed the "SRASC method" — which will be further defined in Section III-C3. The properties of the three rules are given in Figure 2 and an analysis is presented below.

A. Existing FCC-style rules

Consider the FCC-style rules shown in Figures 3(a) and 4(a). We note that the separation distance is fixed (specifically, it is 14.4 km) and that secondaries outside of the no-talk region operate with a fixed transmit power. As we argued in [8], there are two main problems with these rules:

- 1) The fundamental idea of a per-device power limit is based on an expected device density. If there are too many secondaries operating simultaneously, the aggregate interference may cause some primary receivers in the protected region to experience an outage. This is a big threat to the future of whitespaces: services which feel threatened or made vulnerable by the regulations may advocate strongly for more conservative regulations in order to allay their fears.
- 2) The choice of separation distance is a political one: since urban areas tend to be near protected regions, a large separation distance unfairly discounts urban areas. On the other hand, increasing the separation distance means that a higher maximum transmit power can be chosen; this would greatly benefit areas which are able to transmit. This tradeoff makes it difficult to give the same quality of service to everyone.

B. Power scaling rules

Based on the arguments above, we developed candidate power scaling rules that attempted to "blur" the no-talk region

^{**} These power scaling rules were developed in our DySpAN 2011 paper.

^{***} These rules will be developed later in this paper.

⁸The protected region is the space where secondaries are not allowed to transmit under any rules. Primary receivers inside this protected region are theoretically guaranteed reception.

⁹This is the region is the space where secondaries are not allowed to transmit. The no-talk region necessarily includes the protected region.

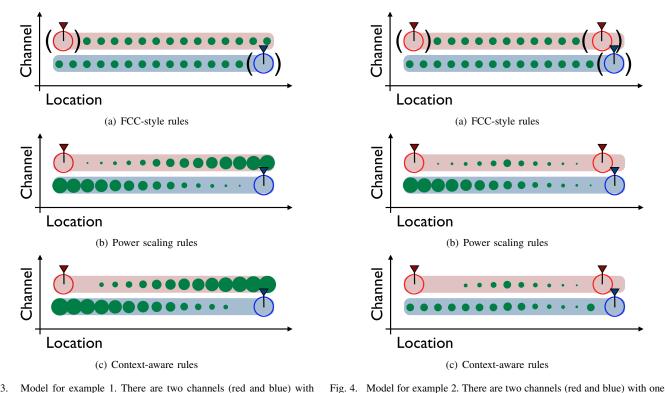


Fig. 3. Model for example 1. There are two channels (red and blue) with one primary tower each. The protected regions are shown as circles around the primary transmitters and the no-talk regions are shown using black semicircles. Secondaries are shown in discrete locations and their sizes indicate their relative transmit powers. We assume that a secondary transmitter will use all available channels (i.e. channels with nonzero secondary power). In this example we see a large difference between the FCC-style rules and the power scaling rules but minimal difference between the power scaling rules and the context-aware rules. Note that the FCC-style rules do not guarantee that the primary's aggregate interference constraint is obeyed.

primary tower on the blue channel and two primary towers on the red channel. The protected regions are shown as circles around the primary transmitters and the no-talk regions are shown using black semi-circles. Secondaries are shown in discrete locations and their sizes indicate their relative transmit powers. We assume that a secondary transmitter will use all available channels (i.e. channels with nonzero secondary power). In the context-aware rules we allocate power to a point near the protected region on the blue channel because it is unable to use the red channel. Note that the FCC-style rules do not guarantee that the primary's aggregate interference constraint is obeyed so their powers are not comparable to the power scaling and context-aware rules.

and create a graceful degradation of power — and hence data rate — as one neared the protected region [8], as shown in Figures 3(b) and 4(b). These rules had the advantage of being safe for primaries while being flexible for secondaries. Notice that there is no explicit no-talk region in these rules.

Because the rules enforced the aggregate interference constraint for the primary, they were *spatially aware*: powers were coordinated across locations to ensure that the primary's interference constraint was obeyed.

The important thing to notice here is that secondaries which are very near to the protected region would cause much more interference to the primary. For example, see Figure 5 which shows the relative impact of a secondary's transmission on the aggregate interference based on his distance to the protected region. We see that the impact at a distance of 1 km is almost four orders of magnitude higher than the impact at 10 km: this means that his power would have to be 10,000 times lower to cause the same interference to the protected primary receiver. Because of this fact, we can think of secondary power as being more "expensive" near the protected region.

The power scaling rules presented in [8] are far from the only example of variable power limits. Indeed, the Electronic Communications Committee (ECC) in Europe noted that

"location specific output power seems to be better from a spectrum usage view" [18, §9.1]. The UK's Ofcom has also included provisions for scaling the transmit power [19].

By examining Figure 3(b), we notice that locations which are near the protected region on the red channel are far from the protected region on the blue channel and vice versa. This suggests that taking into account the location-specific alternatives may yield considerable gains over assuming it to be an all-or-nothing game on each channel.

C. How would a secondary market behave?

We are now at a point where we understand the competing interests of the secondary transmitters at each location. We assume for simplicity that each location is equally interested in increasing its achievable rate.

In a simple aggregate-interference model we can think of primary receivers inside the protected region¹⁰ as having an "interference budget" which reflects the amount of interference from secondary transmitters that they are able to tolerate. Due to the difference in pathloss shown in Figure 5, we know that

¹⁰Since signal attenuation is monotonically decreasing with distance, we can take the worst-case viewpoint of a primary receiver which is at the edge of his respective protected region.

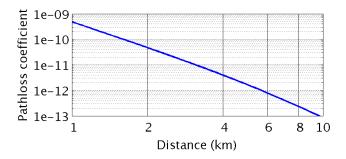


Fig. 5. Pathloss coefficients as a function of distance on TV channel 21. The pathloss coefficient is also the interference weight of a secondary transmitter as a function of distance to the protected region. Notice that the weights decrease by almost four orders of magnitude over a distance of 9 km, making power much "cheaper" at 10 km than it is at 1 km.

secondary power used near the protected region is "expensive" (i.e. it uses more of the budget per unit of secondary power) while secondary power used at far-away locations is "cheap."

Using this knowledge, we can describe the relative prices in the example of Figure 3:

- Power used at locations toward the left is expensive in the red channel but cheap in the blue channel.
- Power used at locations toward the right is expensive in the blue channel but cheap in the red channel.
- Power costs are roughly the same near the center.

Taking the power scaling rules as a starting point, the Coasian-bargaining behavior of the participants is easy to predict: they will "sell" some or all of their expensive power on one channel and use the "profits" to "buy" more cheap power on the other channel. For example, a secondary located near the left edge of the model will relinquish his right to transmit in the red channel in exchange for increasing his power in the blue channel. The secondary does this because the increase in rate in the blue channel exceeds his loss in the red channel.

In Figure 4 we see a similar example which is asymmetric. The point on the right-hand side of the model now has the blue channel as his only option and thus "buys" power from other locations on the blue channel in order to achieve a comparable rate. In both cases, we see that frequency-agile or wide-band secondaries are making decisions based on the characteristics of the channels available to them; in other words, they are demonstrating frequency awareness.

Unfortunately, an actual money- and transaction-based secondary market such as this may turn out to be very complex due to the sheer number of participants. The advantage of the whitespace-style regulations has always been that they provide a reasonable default way to access spectrum without engaging in complex transactions, for example in the provisions for wireless microphones. Given that frequency awareness can deliver value, it is worth seeing if that can also be done in a good default way. The power of databases is that we can simulate trading without needing actual trades to occur.

D. Context-aware rules

The power scaling rules shown earlier in Figures 3(b) and 4(b) are safe but they are ultimately too conservative. In making the rules, we assumed that shutting someone out on one channel was the worst thing you could do because it would give them no rate at all. However, many locations have alternative channels on which they are not so close to the relevant protected region. The Coasian-bargaining solution has these locations selling their right to transmit in the "expensive" channels and increasing their rate in the alternative channels. For those locations which don't have alternatives — such as the secondaries on the right-hand side of Figure 4(c) — we need *frequency-aware* regulations which allow them to transmit near the protected region if that is their only option.

In this particular example, we see in Figures 3(b) and 4(b) that most locations have both the red and blue channels available. We can therefore judiciously restrict some locations to only one channel if it benefits the greater good¹¹. The choice of who to "kick" from which channels then depends on the definition of "greater good." We consider three options below.

- 1) Maximize total power used: Given the primary's interference constraint, we could consider trying to maximize the total amount of power available for secondary use. The solution to this problem turns out to be equivalent to maximizing the amount of power that can be used at the point furthest from the protected region since that is where power is "cheapest." Intuitively we know that this is not a good allocation of resources because it favors only one location, i.e. it is not fair. There is no fairness incentive because of the linearity of the objective function. In reality, we know that the utility of power is not linear but rather logarithmic due to the nature of the information-theoretic capacity bound. In order to incentivize fairness we consider maximizing the average utility in the next section.
- 2) Maximize average rate: It is tempting to think that maximizing the average rate to all secondaries will maximize the individual utility of each secondary. If all secondaries are subjected to the same amount of noise (with or without self-interference), have the same amount of spectrum available, and there is only a maximum system-wide power constraint i.e. in homogeneous spectrum it is optimal in terms of fairness to give each secondary an equal amount of power because of the concave nature of the rate function 12.

However, the properties of the whitespace environment mean that we are not operating in such a simple world. First of all, the utility of one unit of power depends heavily on the location since the noise level is now location-dependent.

 $^{^{11}}$ Note that is always better to use power P in each of two identical channels than power 2P in one channel. This is because rate (Shannon capacity) is linear in bandwidth but logarithmic in power. However, the choice given here is more complex so sometimes it is better to use fewer channels in one location so that other locations will be able to use more channels and/or power.

 $^{^{12}}$ For example, suppose the maximum power limit is 10P and there are 10 secondaries in the system, each with a noise floor N. We could obtain rate $\log(1+P/N)$ on each of 10 secondaries or rate $\log(1+10P/N)$ for one secondary and 0 for all the rest. We know that $10 \cdot \log(1+P/N) > \log(1+10 \cdot P/N)$ and therefore the first scheme is optimal.

Locations closer to the primary transmitter will experience higher noise levels. Secondly, we no longer have a sum power constraint but instead a *weighted* sum power constraint. Our only constraint is that the primary receiver (without loss of generality we consider one at the nearest edge of the protected region) observes aggregate noise from the secondaries which is no greater than some fixed value N_s . As we saw in Figure 5, the relative impact of a secondary's transmit power varies greatly depending on his distance to the primary receiver. Thus we can think of power as "cheaper" for secondaries which are far from the protected region and more "expensive" for those near the protected region.

These properties combined mean that it is both "cheaper" and more beneficial in terms of the average rate to allocate power to secondaries which are far from the protected region. Note that if there were multiple secondaries at the same location, we would still split the power equally among them (as in the previous example). However, this type of objective function does not favor fairness among locations.

- 3) Maximize minimum rate: We saw in the previous example that it may be important to explicitly incorporate fairness into our objective function. As a result, we will look at the most fair objective function: the max-min rate. This objective function seeks to provide all locations with a maximal quality-of-service guarantee. The optimal solution will have the following properties:
 - Locations which are inside of the protected region on all channels will not be able to transmit on any whitespace channel. This is unfortunate but unavoidable because the protected region is a hard constraint. We will not consider these locations to be within the feasible set of the minimization function because any rules which would allow them to use the whitespaces would necessarily violate the primary's interference constraints.
 - In general, each location should try to use the channels with the minimum interference weight, thus reducing its impact on other locations. This lower interference weight also allows it to use a higher power if necessary.
 - Locations which are near the protected region on all channels are still allowed to transmit. These locations are "expensive" but it is unfair to deny them service.

We see an example of this allocation in Figure 3(c). The locations near the protected region on the red channel have graciously moved to the blue channel where they can increase their power to make up for the rate they lost by vacating the red channel. The same thing happens on the other side with locations near the protected region on the red channel. In Figure 4(c), we see the main difference between this method and the other rules: the point on the far right is allowed to transmit despite being near the protected region.

This system has two advantages:

- The no-talk region is not fixed which allows leniency for devices with few or no alternatives.
- By removing as many people from the neighborhood of the protected region as possible, we have increased the

total amount of power we can use in our system and therefore the rates.

Like the power density rules, this approach is spatially aware because it enforces an aggregate interference constraint and thus induces a dependency between locations. However, this solution is also inherently *frequency aware*. That is, in order to make these decisions, we really need to know what other options exist for a given location. We will see examples later which demonstrate why we cannot achieve this performance without frequency awareness.

The answer is relatively obvious for the toy examples discussed above but how does it work in the real world? In the next section, we will use an example one-dimensional world within the United States and check our hypothesis there.

III. ONE-DIMENSIONAL TEST IN THE UNITED STATES

In this section, we apply our hypothesis to the US using actual tower data [17] and the ITU propagation model [20].

A. Assumptions

For simplicity, we restrict ourselves to a one-dimensional line¹³ connecting Berkeley and New York City, pictured in Figure 6. In future work, we hope to extend our results to two dimensions. The main difficulty lies in the computational complexity and we do not expect the qualitative results to change. We also make several other simplifying assumptions:

 We refer to the theoretical achievable capacity (a.k.a. the Shannon capacity) as the "rate." If the power level is P, the noise level is N, the bandwidth is B, and the pathloss coefficient is γ we can write the resulting rate as

$$R = B \cdot \log_2 \left(1 + \frac{\gamma \cdot P}{N} \right)$$

- The distance between the secondary transmitter and the secondary receiver is constant (which implies that γ is constant). This implies that noise, power, and bandwidth are the only factors in the rate calculation.
- In this particular example, there are no primary towers on channel 5 whose protected regions intersect the line. This implies that there are no interference constraints on channel 5. As a result, there is no power limit and therefore no rate limit. We exclude secondaries from using channel 5 in our calculations in order to make our results meaningful and interesting.

B. Heterogeneity

We have previously seen statistics showing the extreme heterogeneity found in the whitespaces. Here, we wish to show the reader the exact amount of heterogeneity present along our one-dimensional line. The line and nearby TV towers are shown in Figure 6; this illustration informs Figures 7 and 8.

In Figure 7, we see the number of channels available under the FCC rules for secondary use along the same line. The

¹³We only worry about protecting the TV towers whose protected regions include some portion of the line. This excludes towers with protected regions adjacent to but not overlapping with the line so this is only a toy.

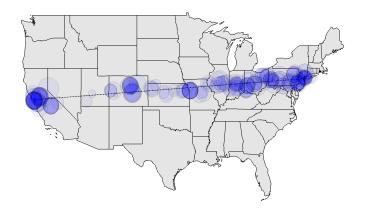


Fig. 6. Trip across the United States from Berkeley to New York City. Blue circles indicate the protected regions of nearby TV towers.

difference between the two lines in this graph underscores the effect of adjacent-channel exclusions¹⁴ In Berkeley, there are only a few channels available due to the preponderance of TV stations on TV towers in the area (e.g. Sutro Tower and the San Bruno Tower). As we pass through other less populous western states, we see the number of available channels increases. In the eastern United States, the population density and consequently the TV tower density increases, leading to a decrease in the number of available channels. Finally, upon reaching the east coast we once again see a drastic decrease in the number of available channels since New York City has many local TV stations. We see the locations of these towers also reflected in the average noise floor, shown in Figure 8.

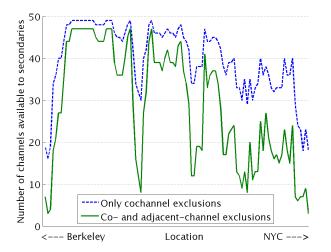


Fig. 7. Number of channels potentially available to secondaries in the TV whitespaces along the line shown in Figure 6.

C. Methods for whitespace power allocation

In this section, we look at several algorithms for allocating power to each secondary (location, channel) pair. We present

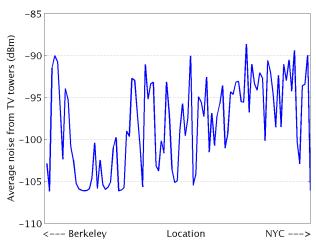


Fig. 8. Average noise level for secondaries operating in the TV whitespaces along the line shown in Figure 6.

the details of each algorithm and compare the results. A comparison of the algorithms can also be found in Figure 9. Note that at optimality, all locations achieve the same rate R.

Nomenclature and variables: we will use the following nomenclature and variable names:

- l is an index indicating the discrete location of a secondary. We will refer to "location" and "secondary transmitter" interchangeably.
- c is an index indicating the channel (i.e. frequency).
- ullet N_c is the total number of channels in the model.
- $N_c(l)$ denotes the number of channels available at location $l.\ 1 \le N_c(l) \le N_c$ for all l.
- $R(l,c) \ge 0$ is the rate achieved by the secondary at location l on channel c.
- R denotes the total rate achieved at every location. That is, $\sum_{c=1}^{N_c} R(l,c) = R(l) = R$ for all l.

Our goal is to maximize (over power allocations) the minimum rate (over all locations) subject to the aggregate interference constraints of the primary receivers in the protected regions. We can write this as

$$\max_{powers} \min_{l} \sum_{c=1}^{N_c} R(l, c) \tag{1}$$

subject to: aggregate interference constraints

We will now examine several algorithms which attack this optimization problem with different sets of knowledge.

1) Maximize each channel's rate ("MECR"): One natural approach to this problem is to maximize the achievable rate on each channel independently. This solution ignores the dependence between channels and is therefore frequency-unaware. However, we will enforce the aggregate interference constraint so that it is spatially aware.

¹⁴For all results in this paper we enforce adjacent-channel exclusions.

Method	Spatial awareness	Frequency awareness
SRASC Split rate among a subset of channels	Devices obey aggregate interference constraints QoS guarantee for <i>all</i> locations	Locations may opt to leave fallow available channels to improve the systemwide utility
SREAC Split rate equally among channels	Devices obey aggregate interference constraints QoS guarantee for <i>all</i> locations	Minimal: only considers the number of available channels
MECR Maximize each channel's rate	Devices obey aggregate interference constraints QoS guarantee for <i>all</i> locations	None: each channel is maximized independently
FPE Fixed power everywhere	Does not obey aggregate interference constraints	None: power does not change with the number of available channels
FPMCQ Fixed power, minimum channel quality	Does not obey aggregate interference constraints	Transmits only on channels which are of sufficient quality

Fig. 9. Comparison of the properties of the methods presented in Section III-C.

In this method, the power is allocated such that the rate R(l,c) is R_c if location l can transmit on channel c and 0 otherwise. Note that each location will have a different set of available channels which causes spatial variation in the total rates achieved. The max-min rate problem can be written as

$$\min_{l} \quad \sum_{c=1}^{N_c} R(l, c)$$

s.t. R_c is the maximum achievable rate on channel c

Note that the maximization already occurred in the choice of R_c . The method presented in the next section adds a minimal amount of frequency awareness.

2) Split rate equally among channels ("SREAC"): This method is motivated by the concept of a service at location l which desires a fixed rate R and blindly attempts to achieve this rate by splitting it equally among its $N_c(l)$ channels, thus achieving rate $R/N_c(l)$ on each. The max-min rate R is chosen such that the aggregate interference constraint is obeyed.

This method is spatially aware since it obeys the aggregate interference constraint but it is only minimally frequency-aware since it does not consider the relative quality of channels. The SREAC method can be viewed as the starting condition for the SRASC method discussed in the next section.

3) Split rate among a subset of channels ("SRASC"): This method is a heuristic approach rather than an exact solution to the max-min problem¹⁵ of (1). It uses the SREAC method as a starting point but greedily "bans" locations from specific channels (i.e. imposes additional constraints). This has the effect of creating a flexible separation margin: locations which

have a "cheaper" channel available will vacate "expensive" channels (thus making space for others) while those who have only the "expensive" channels are still allowed to use them¹⁶. Note that this is the method shown in Figures 3(c) and 4(c). This method is both frequency- and spatially-aware.

4) Fixed power everywhere ("FPE"): These rules are inspired by the FCC's regulations which allow transmitters located outside of the no-talk region to operate at 4 Watts. Likewise, we allow each location to transmit at 4 Watts on each channel. In this model, we found that a separation margin of less than 26 km did not protect receivers inside the primary's protected region. Unlike the previous methods we will not enforce the aggregate interference constraint. For these reasons the FPE method is both frequency- and spatially-unaware.

5) Fixed power, minimum channel quality ("FPMCQ"): These rules were inspired by the idea of a device using the FCC rules but which has a minimum channel quality constraint. This reflects the cost-benefit analysis that will be done by transmitters when considering whether or not to transmit on an additional channel.

Since these rules do not obey the aggregate interference constraint, they are considered spatially-unaware. However, since they take into account the relative quality of the available channels they are frequency aware.

Regardless of the channel quality threshold these rules will never perform better than the FPE rules since the powers used will be less than or equal to those from the FPE rules. We have included these rules for completeness but we will make no further mention of them since they are obviously strictly

 $^{^{15}\}mbox{Although}$ the problem stated in Equation 1 is one of concave maximization over a convex set, conventional solvers such as CVX and the Matlab Optimization Toolbox do not seem to have the necessary precision. The constraints define the interior of a polytope and R(l,c) is clearly concave due to the logarithm. Summation and pointwise minimization preserve concavity, thus the problem is concave.

¹⁶There are situations in which this is not an entirely selfless behavior. For example, consider the example given in Figure 3(c): locations near the protected region on one channel used the other channel exclusively. This left more of the "budget" for far-away locations, meaning that they received a boost to their powers. Since the example is symmetric, vacating one channel meant receiving a much higher power in the remaining channel.

inferior to the FPE rules for all metrics considered here 17.

D. Evaluation

In this section, we discuss the relative performance of the methods described in the previous section. We begin by looking at the rate achievable at all locations¹⁸:

Method	Max-min rate R	Safe for primaries?
SRASC	588 Mbps	Yes
SREAC	0.13 Mbps	Yes
MECR	97 Mbps	Yes
FPE	254 Mbps	No
FPE, 26 km	197 Mbps	Yes

The rates in the table above are much higher than one might expect to see in the TV whitespaces. However, we wish to remind the reader that these results hold for a *one-dimensional* world and thus reflect a drastic decrease in the number of constraints. This allows some locations to use a much larger power than would be safe in the two-dimensional case¹⁹. We expect that investigating a two-dimensional example will confirm the same basic picture revealed here.

We see that the SRASC method is the clear winner, achieving over twice as much as the FPE method. This is especially impressive because the power allocation used in the FPE method is *not* "safe" for primary receivers (i.e. it does not obey the primary's aggregate interference constraint) while the power allocation from the SRASC method *is* safe. This improvement is due to the judicious use of available channels employed in the SRASC method. The MECR and SREAC methods perform poorly for the same reason. In an attempt to give these other methods a fighting chance, we consider two variations of the problem.

1) In terms of separation distance: Here we consider increasing the separation distance to determine if it improves performance. The separation distance is the enforced distance between the protected region and the nearest transmitting secondary²⁰. The existence of a separation distance blindly bans secondary transmitters from channels without considering the alternatives available to them. As the separation distance increases, secondaries will need to increase the power used on remaining channels in order to compensate for a decreasing number of available channels. We can see the resulting maxmin rate in Figure 10 as a function of the separation distance. We discuss the general trend of each method in turn.

SRASC: This method already considered "banning" secondaries from technically-available channels. Although it generally increased the distance between secondaries and the

protected region, this was not a hard constraint. Introducing this as a hard constraint decreases the max-min rate. To see why, consider a secondary with two available channels which is near (but not inside) the protected region on each channel. As the separation distance increases, he loses one of these two channels. In order to maintain the same rate R, he would need to increase his power on his remaining channel. However, he cannot always increase it enough to maintain rate R due to the interference constraint. This accounts for the downward jumps as the separation distance increases. At a separation distance of about 140 km, at least one location no longer has any channels available and so the minimum rate drops to zero for all methods.

SREAC: Unlike the other methods, we see that the trend for this method is not monotone decreasing nor is it monotone increasing. We explain these two behaviors as follows:

- *Increases:* As the separation distance increases, secondaries are banned from using the channels which are near the protected region (therefore "expensive;" see Figure 5). By using cheaper channels, the net power usage increases and consequently the rates increase.
- *Decreases:* Occasionally the separation distance increases so much that a location which really depended on the use of a particular channel is now "banned." This location is now unable to achieve his former rate without affecting the power allocation for other locations.

The interplay between these two phenomena causes the function to follow an interesting path. Ultimately, though, this method performs worse than the SRASC method because it makes a hard decision about channel availability without considering the bigger picture.

MECR: The max-min rate for this method is strictly increasing with the separation distance. The rates are very low for small separation distances due to allocation of power to locations which are "expensive." As we increase the separation margin, we start using cheaper locations and the expensive (location, channel) pairs now receive their rate using cheaper channels. This method still performs worse than the SRASC method at every point because it is forced to serve all locations outside of the no-talk region on each channel: anything less runs the risk of under-serving the minimizing location. In contrast the SRASC method can intelligently leave fallow certain (location, channel) pairs if this yields an overall gain.

FPE: This method simply did not have the power to be able to compete with the SRASC method. Even though it is unsafe for separation distances smaller than 26 km, it still uses a lower power than the SRASC method when far from the protected regions (i.e. it doesn't employ power scaling).

Conclusion: Ultimately the goal is to maximize the minimum rate. We have seen by examining Figure 10 that this maximum is achieved by using the SRASC method with no minimum enforced separation distance (i.e. maximum flexibility). Recall that the SRASC method is a heuristic approach to solving the max-min problem and furthermore the algorithm itself is not fully optimized. However, the solutions for all

¹⁷The only potential advantage these rules have is that they will require a smaller separation margin to be safe; however, since they are not being tested for safety this is irrelevant.

¹⁸In our example one-dimensional world, all points are outside of the protected region and the adjacent-channel protected region on at least one channel. This can be verified using Figure 7.

¹⁹However, we saw little change in the number after artificially capping the secondary transmit power at 100 W.

²⁰For simplicity, we assume that the separation distance is the same for cochannel and adjacent-channel exclusions. The FCC rules use different separation distances.

other methods are exact so the advantage of the context awareness is actually understated.

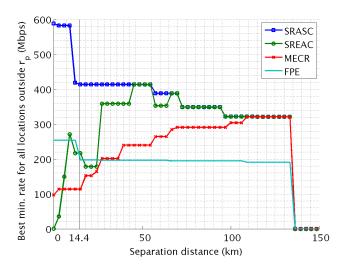


Fig. 10. Comparison of max-min rates achieved with four power-allocation methods while varying the separation distance. We see that the SRASC method is strictly superior to the other three methods.

2) In terms of coverage: We saw in Figure 7 that some points — especially those near the coasts — suffer from a severe lack of channels. Nonetheless, we have been requiring our algorithms to accommodate these locations. To check that there aren't a few "trouble locations" which are giving an unfair advantage to the SRASC method, we consider making the conscious decision to deny service to some locations.

Practically, we did this by removing locations from the feasible set in the minimization problem. The algorithm is as follows for each method (individually):

- 1) For each location as l, consider removing l from the feasible set in the minimization. Evaluate the maxmin optimization problem to get the potential rate as a function of the removed point, l.
- 2) Find the location l' which, when removed, gives the greatest potential rate. Note that l' need not be the same point for each method.
- 3) Remove the point l' from the feasible set and return to the first step.

By construction, the max-min rate will monotonically increase as we deny coverage to more and more locations. Indeed we see this behavior in Figure 11. For ease of interpretation, we do not consider varying the separation distance in this exercise and instead assume that a secondary is eligible to transmit if it is outside of the protected region²¹.

From these results we see that not only is the SRASC method outperforming all other methods but in fact it does about three times better than the SREAC method and 2.5 times better than the MECR method even when we remove the most bothersome points. This demonstrates the important gains made possible by full context awareness.

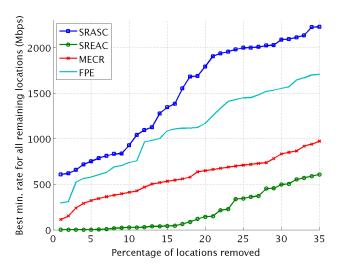


Fig. 11. Comparison of the max-min rates achieved with four powerallocation methods while varying the desired amount of coverage. Notice that the SRASC method is strictly superior to the other three methods.

3) Discussion: We have seen that under all examined conditions, the SRASC method outperforms the other four methods. From Figure 9 we see that it is the only method which has true spatial and frequency awareness. This suggests that these qualities are necessary in regulations which seek to provide a fair quality of service to all constituents.

IV. CONCLUSIONS

The creation of a spectrum regulatory agency in 1927 and the subsequent rearrangement of the spectrum brought about a brand new way of utilizing the spectrum. It reduced interference between competing services and improved performance. This method worked well for almost a century until spectrum became artificially scarce, at which point the FCC recognized the need to utilize the whitespaces. The dynamic nature of the whitespaces coupled with the need to keep the primaries safe really makes databases essential to the use of the whitespaces. However, databases should not be seen as a necessary burden but rather as an exciting opportunity for exploration and improvement. In particular, we explored the potential gains of including a variable secondary power limit which is disseminated via databases.

We have argued that the extreme heterogeneity in the whitespaces means that context-aware rules can provide huge gains over simpler rules. We first showed these improvements using rudimentary examples including a thought-experiment about the behavior of a secondary power market. The lesson was that there was a powerful incentive to trade which raised the question: why not let the database do some of the obvious trades itself? In fact, the only situation in which the database cannot compute the optimal Coasian-bargaining solution is when the secondary devices have some sort of specialized or local knowledge not available to the database ahead of time. When considering the advantages of secondary markets we

²¹As always, adjacent-channel exclusions are enforced.

should therefore compare them to the optimal default databasereliant rules rather than the current one-size-fits-all rules.

We then considered a more complex example which included real-world data on TV transmitters in the United States. The gain here from using a context-aware method was substantial and remained so even after modifying the example to remove the apparent disadvantage of the other methods.

All of our results suggest that context-aware regulations perform better with all metrics, so what sort of situation would not benefit from the use of context-aware rules? If we look at context-aware rules as the steady-state result of Coasian-bargaining, the answer is clear: context-aware rules provide no gain when resources cannot be traded. For example, radioastronomy applications rely on the use of a specific frequency and cannot use any other frequency no matter how high the incentive. However, whitespaces are not meant to be used for such selective applications but rather as a breeding ground for innovative devices and applications. Well-chosen context-aware rules will foster this innovation, ensuring that the full potential of the whitespaces is harnessed.

V. FUTURE WORK

We have identified several directions for future work:

- We expect our results to generalize from a onedimensional line to a two-dimensional map. The main difficulty is in the computational complexity that this problem presents with numerical precision issues.
- We suspect that our results hold for general risk-averse (i.e. coverage-inclined) objective functions but we have not yet worked out these results.
- Heterogeneity in the whitespaces is not confined to the variability of the noise floor and the number of available channels. We fully expect that a wide variety of devices will utilize the whitespaces and it is extremely important to ensure that regulations are application-agnostic [15]. Anything less than this undermines the notion of the whitespaces as a place for discovery and growth. The authors of [13] began an exploration of this space when they considered assigning a range to each channel.
- Given physical limitations of current cognitive radio technology, we may need to add constraints that address the maximum frequency agility of a device. For example, devices may not be able to transmit simultaneously at 60 MHz and 600 MHz.
- It may also be useful to impose a "minimum utility" constraint (e.g. minimum spectral efficiency) on a channel: if a channel offers too little, it will go unused and the resources would have been better allocated elsewhere.

Finally, as important as it is to choose good regulations, we do not necessarily need the full solution at the time of deployment: databases afford us new regulatory flexibility that can be used to our advantage.

VI. ACKNOWLEDGEMENTS

We thank the U.S. NSF (CNS-403427, CNS-0932410 and a Graduate Research Fellowship) for their support. We also

wish to thank the reviewers for their insightful comments.

REFERENCES

- [1] "In the Matter of Unlicensed Operation in the TV Broadcast Bands: Second Memorandum Opinion and Order," Federal Communications Commision, Tech. Rep. 10-174A1, Sep. 2010. [Online]. Available: http://hraunfoss.fcc.gov/edocs%5Fpublic/attachmatch/FCC-10-174A1.pdf
- [2] "In the Matter of Unlicensed Operation in the TV Broadcast Bands: Third Memorandum Opinion and Order," Federal Communications Commision, Tech. Rep. 12-36A1, Aug. 2012. [Online]. Available: http://hraunfoss.fcc.gov/edocs%5Fpublic/attachmatch/FCC-12-36A1.pdf
- [3] "In the Matter of Unlicensed Operation in the TV Broadcast Bands: Second Report and Order and Memorandum Opinion and Order," Federal Communications Commission, Tech. Rep. 08-260, Nov. 2008. [Online]. Available: http://hraunfoss.fcc.gov/edocs%5Fpublic/attachmatch/FCC-08-260A1.pdf
- [4] M. Mishra and A. Sahai, "How much white space is there?" EECS Department, University of California, Berkeley, Tech. Rep. UCB/EECS-2009-3, Jan, 2009.
- [5] D. Gurney, G. Buchwald, L. Ecklund, S. Kuffner, and J. Grosspietsch, "Geo-location database techniques for incumbent protection in the tv white space," in *New Frontiers in Dynamic Spectrum Access Networks*, 2008. DySPAN 2008. 3rd IEEE Symposium on. IEEE, 2008, pp. 1–9.
- [6] M. Kokar and L. Lechowicz, "Language issues for cognitive radio," Proceedings of the IEEE, vol. 97, no. 4, pp. 689 –707, april 2009.
- [7] [Online]. Available: http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast%5Fspectrum%5Freport%5Ffinal%5Fjuly%5F2012.pdf
- [8] K. Harrison and A. Sahai, "Potential collapse of whitespaces and the prospect for a universal power rule," in *New Frontiers in Dynamic Spec*trum Access Networks (DySPAN), 2011 IEEE Symposium on. IEEE, 2011, pp. 316–327.
- [9] K. Harrison, S. Mishra, and A. Sahai, "How much white-space capacity is there?" in *New Frontiers in Dynamic Spectrum*, 2010 IEEE Symposium on. IEEE, 2010, pp. 1–10.
- [10] J. van de Beek, J. Řiihijarvi, A. Achtzehn, and P. Mahonen, "UHF white space in Europe - a quantitative study into the potential of the 470–790 MHz band," in New Frontiers in Dynamic Spectrum Access Networks (DySPAN), 2011 IEEE Symposium on. IEEE, 2011, pp. 1–9.
- [11] R. Jantti, J. Kerttula, K. Koufos, and K. Ruttik, "Aggregate interference with FCC and ECC white space usage rules: Case study in Finland," in New Frontiers in Dynamic Spectrum Access Networks (DySPAN), 2011 IEEE Symposium on. IEEE, 2011, pp. 599–602.
- [12] M. Nekovee, "Quantifying the availability of TV white spaces for cognitive radio operation in the UK," in *Communications Workshops*, 2009. ICC Workshops 2009. IEEE International Conference on. IEEE, 2009, pp. 1–5.
- [13] M. Ma and D. Tsang, "Impact of channel heterogeneity on spectrum sharing in cognitive radio networks," in *Communications*, 2008. ICC'08. IEEE International Conference on. IEEE, 2008, pp. 2377–2382.
- [14] H. Zheng and C. Peng, "Collaboration and fairness in opportunistic spectrum access," in *Communications*, 2005. ICC 2005. 2005 IEEE International Conference on, vol. 5. IEEE, 2005, pp. 3132–3136.
- [15] M. Kaplan and F. Buzluca, "A dynamic spectrum decision scheme for heterogeneous cognitive radio networks," in *Computer and Information Sciences*, 2009. ISCIS 2009. 24th International Symposium on. IEEE, 2009, pp. 697–702.
- [16] K. Woyach, P. Grover, and A. Sahai, "Near vs far field: interference aggregation in TV whitespaces," 2011.
- [17] See transition.fcc.gov/mb/video/tvq.html.
- [18] "Technical and operational requirements for the possible operation of cognitive radio systems in the white spaces of the frequency band 470-790 MHz, ECC Report 159," http://www.erodocdb.dk/docs/doc98/official/Pdf/ECCRep159.pdf.
- [19] [Online]. Available: http://stakeholders.ofcom.org.uk/binaries/ consultations/geolocation/statement/statement.pdf
- [20] "Method for point-to-area predictions for terrestrial services in the frequency range 30 MHz to 3 000 MHz," International Telecommunications Commission (ITU), Recommendation ITU-R P.1546-3, 2007.