Spectrum Sensing: Fundamental Limits

Anant Sahai, Shridhar Mubaraq Mishra and Rahul Tandra

Abstract Cognitive radio systems need to be able to robustly sense spectrum holes if they want to use spectrum opportunistically. However, this problem is more subtle than it first appears. It turns out that real-world uncertainties make it impossible to guarantee both robustness and high-sensitivity to a spectrum sensor. A traditional time-domain perspective on this is relatively straightforward, but to really understand what is happening requires us to think more deeply about the role of fading. In particular, this demands that we look at the spatial perspective as well. We show how to set up reasonable approximate metrics that capture the two desirable features of a spectrum sensor: safety to primary users and performance for the cognitive radios. It is the tradeoff between these two that is fundamental. Single-user sensing turns out to have fundamental limits that require access to more diversity to overcome. Cooperative sensing can provide this diversity, but it too has its own limits that come from the degree to which the model can be trusted.

1 Introduction and overview

Cognitive radios must have the ability to sense for spectrum holes. Philosophically, this is a decision problem: is it safe to use the spectrum where we are or is it unsafe? This is a question with a binary answer and so it seems natural to encapsulate the entire problem of spectrum sensing by mathematically casting it as a binary hypothesis testing problem [1]. This chapter shows that while such formulations are seemingly natural, if we are not careful, they can blind us to many of the true fundamental limits involved in spectrum sensing. In particular, the core spatial dimension to the problem and its interaction with fading in wireless channels introduces tradeoffs that need a better formulation to understand.

All three authors are with the University of California, Berkeley CA 94720, e-mail: sahai@eecs.berkeley.edu

This chapter tells this story in phases. First, the traditional binary hypothesis testing story is recapitulated with the conceptually central role being played by the traditional detection metric of *sensitivity*[1]. It is well known that the sensitivity of detectors can be improved by increasing the sensing time and so the sample complexity gives us a natural way to compare different spectrum sensors. However, one must also consider the impact of real-world uncertainties on the performance of detectors since robustness is important. Doing this reveals that the sample complexity blows up to infinity as the detector sensitivity approaches certain critical values — called *SNR walls* [1, 2, 3].

A closer look at the receiver operating characteristic (ROC) reveals why. Below these SNR walls, it is completely impossible to robustly distinguish the two hypotheses. The location of the walls themselves depends on what is known about the signal being sensed as well as the size of certain critical uncertainties in the noise distribution and fading process [3].

These SNR wall limits suggest that it is impossible to design very sensitive detectors. However, why does one need sensitive detectors? The main reason is because the cognitive radio needs to be sure that it is far away from any primary user before using the channel. The strength of the primary signal received at the cognitive radio is just a proxy to ensure that we are 'far enough' from the primary transmitter. If there were no wireless fading, there would be a single 'right level' of sensitivity. It is the reality of fading that makes us demand additional sensitivity.

Because fading can effect different detectors differently, this reveals that a head-to-head comparison of the sensitivity of two detectors can be misleading. Instead, the possibility of fading has to be incorporated into the "signal present" hypothesis itself. The logical tradeoff is then between the effective probability of missed detection (it matters not whether the miss is due to an unfortunate fade or a quirk in the noise realization) and the sample complexity. Here, we show a surprising example. Whereas traditionally, the coherent detection of a pilot tone is considered to have better asymptotic sample complexity than an energy detector, this need not be true when fading is considered. The coherence bandwidth matters. If the primary signal is sufficiently wideband, then the simple energy detector can have *better* asymptotic sample-complexity behavior than a coherent pilot-tone detector!

The bigger conceptual challenge comes in trying to understand "false alarms." The traditional hypothesis-testing formulation would say that a false alarm consists of when the detector says that we should not use the channel when the primary is truly absent. But this is not the problem actually facing a cognitive radio. It wants to avoid saying that we are close to the primary when we are indeed far enough away. The "signal absent" hypothesis needs to be modified in some reasonable way.

At this point, a spatial perspective is essential and while the resulting formulation can fit into the traditional binary hypothesis testing framework, it is useful to reconceptualize the problem in terms of two new metrics, first introduced in [4]. The first metric, namely the *Fear of Harmful interference* F_{HI}, captures the safety to the primary users. This is largely the fading-aware probability of missed detection intro-

[1] The sensitivity of a detector is defined as the lowest SNR for which a given target probability of error can be met.

duced earlier, with some modifications to allow easier incorporation of system-level uncertainty. The second metric, namely the *Weighted Probability of Area Recovered WPAR*, captures the performance of spectrum sensing by appropriately weighting the probability of false alarm (P_{FA}) across different spatial locations. These metrics give a unifying framework in which to compare different spectrum-sensing algorithms. We show how to obtain reasonable metric parameters (most crucially, something which can be interpreted as a spatial discount factor) from real-world data. The tradeoff between *WPAR* and F_{HI} is thus the correct *ROC* curve for spectrum sensing. However, the probabilistic uncertainty underlying the hypotheses is non-ergodic and so the tradeoffs are interesting even if we allow an infinite number of samples.

The new metrics show that fading uncertainty forces the *WPAR* performance of single-radio sensing algorithms to be very low for desirably small values of F_{HI}, even with infinite samples. The physical reason for such a poor performance is that a single radio cannot distinguish whether it is close to the primary user and severely shadowed, or if it is far away and not shadowed. Furthermore, these metrics shed a new perspective on the impact of noise uncertainty on the sensing performance of spatial spectrum holes as well as on the comparison of different detectors. We show that under noise uncertainty, there exits an F_{HI} threshold beyond which the *WPAR* vanishes to zero, i.e., if we need to guarantee protection to the primary below this threshold, then one cannot robustly recover any spectrum holes in space. In addition, a head-to-head comparison is made between the energy detector and the coherent pilot detector. This reveals that even with an infinite number of samples, the energy detector will do better for high values of F_{HI} and it is only noise-uncertainty that allows the coherent pilot detector to do better at low F_{HI}. This fact would be invisible without using the right metrics.

The inherent spatial advantage of the energy detector over the coherent pilot-tone detector comes from its ability to exploit frequency diversity. More diversity helps. Cooperation among cognitive radios allows them to exploit spatial fading diversity and hence get much higher *WPAR*. Even here, the new spatial metrics bring new insights into the fundamental tradeoffs involved. Consider the question of how to fuse 1-bit tentative decisions from individual radios into a single decision for a group of cooperating radios. A traditional sensitivity-oriented hypothesis testing framework would suggest that the best rule is the OR-rule that gives every radio a veto over using the channel [5]. The diversity allows each individual radio to have relaxed sensitivity requirements and this keeps traditionally understood false alarms very rare. However, the spatial metrics perspective reveals that a majority-vote-rule actually works significantly better while also being quite robust to uncertainty.

The fundamental limits of spectrum sensing by cognitive radios cannot be understood unless we properly recognize what exactly cognitive radios are trying to do. Carefully incorporating the spatial nature of the problem into the formulation is critical in doing so. This allows us to see the critical role that diversity plays.

2 Spectrum Sensing: time-domain perspective

We first consider the traditional formulation of the spectrum sensing problem as a binary hypothesis test [1]. The reader is encouraged to read [3] for more details.

Let $X(t)$ denote the band-limited signal we are trying to sense, let H denote the fading process, and let the additive noise process be $W(t)$. The discrete-time version is obtained by sampling the received signal at the appropriate rate. The two hypotheses are:

$$\begin{aligned} &\text{Signal absent } \mathscr{H}_0 : Y[n] = W[n] && n = 1,2,\cdots,N \\ &\text{Signal present } \mathscr{H}_1 : Y[n] = \sqrt{P}\cdot HX[n] + W[n] && n = 1,2,\cdots,N \end{aligned} \tag{1}$$

Here P is the received signal power, $X[n]$ are the unattenuated samples (normalized to have unit power) of the primary signal, H is a linear time-varying operator, $W[n]$ are noise samples and $Y[n]$ are the received signal samples. We assume that the *signal is independent of both the noise and the fading process*. Random processes are traditionally assumed to be stationary and ergodic unless otherwise specified.

2.1 Traditional metrics, Sample Complexity and SNR walls

Consider the detection problem in (1). The goal is to design a detection algorithm that minimizes the number of samples required (N) to distinguish between the two hypotheses subject to constraints on the probability of false-alarm and the probability of missed-detection. For concreteness, we consider test-statistic/threshold based detection algorithms.

Let the detector be given by $T(\mathbf{Y}) := \frac{1}{N}\sum_{n=1}^{N}\phi(Y[n]) \underset{\mathscr{H}_0}{\overset{\mathscr{H}_1}{\gtrless}} \gamma$, where $\phi(\cdot)$ is a known deterministic function and γ is the detector threshold. The detector threshold γ must be chosen such that

$$\begin{aligned} \mathscr{P}_{\mathscr{W}}\left(T(\mathbf{Y}) > \gamma | \mathscr{H}_0\right) &\leq P_{FA}, \\ \mathscr{P}_{\mathscr{W}}\left(T(\mathbf{Y}) < \gamma | \mathscr{H}_1\right) &\leq P_{MD}. \end{aligned} \tag{2}$$

The lowest signal to noise ratio, $SNR := \frac{P}{\sigma_n^2}$ (σ_n^2 is the nominal noise power) for which the constraints in (2) are met is called the *sensitivity* of the detector. Furthermore, eliminating γ from (2) we can solve for N as a function of the SNR (sensitivity), P_{FA}, and P_{MD}. Hence, we can write

$$N = \xi(SNR, P_{FA}, P_{MD}). \tag{3}$$

This is called the *sample complexity* of the detector. The traditional metrics triad of sensitivity, P_{FA}, and P_{MD}, are used along with the sample complexity to evaluate the performance of detection algorithms. For reasonable detectors, $\xi(SNR, P_{FA}, P_{MD})$ is

a monotonically decreasing function of SNR, P_{FA} and P_{MD}. In particular, when the probabilistic uncertainty is ergodic, arbitrarily low sensitivities can be achieved by increasing the number of samples. For instance, the sample complexity of an energy detector scales as $N = O(SNR^{-2})$, and the sample complexity of a matched filter scales as $N = O(SNR^{-1})$ [6].

2.1.1 Noise uncertainty model

All three aspects (W,H,X) of the problem in (1) admit statistical models. So far we have assumed that these statistical models are completely known. However, it is unrealistic to assume complete knowledge of their parameters to infinite precision. To understand the issue of robustness to uncertainty, we assume knowledge of their distributions within some bounds and are interested in the worst case performance of detection algorithms over the uncertain distributions.

We describe the noise uncertainty model here, but the reader is referred to [3] for a detailed description of the uncertainty models for noise and fading processes, their motivation, and our modeling philosophy. A bounded-moment uncertainty model is used to capture the idea of approximately Gaussian noise. A white noise distribution $W_a \in \mathscr{W}_x$ if:

- The noise process is symmetric $\mathbb{E}W_a^{2k-1} = 0, \forall k = 1,2,\ldots$.
- Even moments of the noise must be close to the nominal noise moments in that $\mathbb{E}W_a^{2k} \in [\frac{1}{\rho^k}\mathbb{E}W_n^{2k}, \rho^k\,\mathbb{E}W_n^{2k}]$, where $W_n \sim \mathscr{N}(0,\sigma_n^2)$ is a nominal Gaussian noise random variable and $\rho = 10^{x/10} > 1$.

The parameter x is used to quantify the amount of non-probabilistic uncertainty in the noise power, i.e., we allow for x dB of uncertainty in the noise variance and allow the other moments to have commensurate flexibility.

Now, both hypotheses do not actually specify a unique probability model. A reasonable interpretation of this is that the probability of false alarm and the probability of missed-detection constraint must be met for all possible noise distributions in the uncertainty set. That is,

$$\begin{aligned} \sup_{W\in\mathscr{W}_x} \mathscr{P}_W\left(T(\mathbf{Y}) > \gamma | \mathscr{H}_0\right) &\leq P_{FA}, \\ \sup_{W\in\mathscr{W}_x} \mathscr{P}_W\left(T(\mathbf{Y}) < \gamma | \mathscr{H}_1\right) &\leq P_{MD}. \end{aligned} \tag{4}$$

2.1.2 Impact of uncertainty on sample complexity

Under the noise uncertainty model given in Section 2.1.1, the sample complexity of detection also depends on the parameter $\rho = 10^{x/10}$, i.e., $N = \xi(SNR, P_{FA}, P_{MD}, \rho)$. From [3] the sample complexity of the radiometer (energy detector) with noise uncertainty is

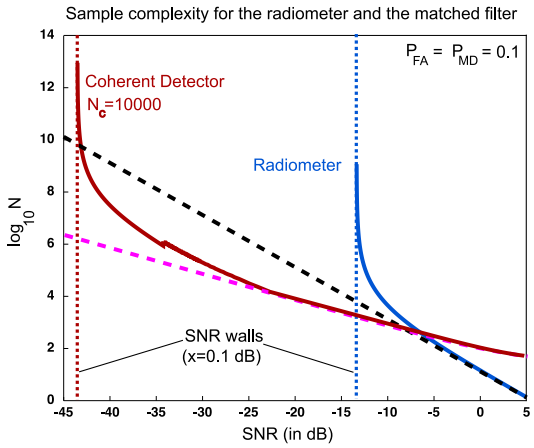

Fig. 1 Under noise uncertainty the sample complexities for both the radiometer (blue curve) and the matched filter (red curve) blow up to infinity as the SNR approaches the corresponding wall. The dashed lines show the sample complexity without noise uncertainty. In the plot, the pilot tone contains a fraction $\theta = 0.1$ of the total signal power.

$$N^{radiometer} \approx \frac{2[\mathcal{Q}^{-1}(P_{FA}) - \mathcal{Q}^{-1}(1-P_{MD})]^2}{\left[SNR - \left(\frac{\rho^2-1}{\rho}\right)\right]^2}, \tag{5}$$

where $\mathcal{Q}^{-1}(\cdot)$ is the inverse of the Gaussian tail probability function.

For a matched filter looking for a sinusoidal pilot tone, the story is a bit more involved [3]. If the coherence-time N_c could be finite, then an unmodified matched-filter could fail miserably if it tries to coherently integrate across multiple coherence times. Instead, the coherent processing gain must be limited to the shortest possible coherent block, and then information from multiple blocks combined in a non-coherent manner. For low enough SNRs, we get

$$N^{mf} \approx \frac{2N_c[\mathcal{Q}^{-1}(P_{FA}) - \mathcal{Q}^{-1}(1-P_{MD})]^2}{\left[\theta \cdot N_c \cdot SNR - \left(\frac{\rho^2-1}{\rho}\right)\right]^2} \tag{6}$$

where θ is the fraction of the total power in the pilot tone.

Figure 1 plots the sample complexity of the radiometer and the matched filter with/without noise uncertainty. Notice that at high *SNR*, the radiometer has better sample-complexity performance than the matched filter. This is because the radiometer uses the total power in the signal for detection, but the matched filter uses only a small fraction of the total signal power. The most prominent feature of the figure is that under noise uncertainty, the sample complexity curves blow up to infinity as the *SNR* approaches a critical value SNR^T_{wall}.

$$\lim_{SNR \downarrow SNR^T_{wall}} \xi(SNR, P_{FA}, P_{MD}, \rho) = \infty. \tag{7}$$

Notice that the sample-complexity curves for the radiometer and the matched filter (dashed lines) without uncertainty have slopes of $-\frac{2}{10}$ and $-\frac{1}{10}$ respectively.

With noise uncertainty, the sample-complexity curves deviate from their nominal (no uncertainty) behavior as the SNR approaches the wall. The matched-filter's curve (blue) has an interesting intermediate phase. For moderately low SNRs, the number of samples required for the matched filter is on the order of multiple coherence times and here the slope transitions to $-\frac{2}{10}$ before ramping to $-\infty$ near the SNR wall.

Equations (5) and (6) show that the *SNR* wall for the radiometer and a matched filter are given by $SNR_{wall}^{radiometer} = \frac{\rho^2-1}{\rho}$ and $SNR_{wall}^{mf} = \frac{1}{\theta \cdot N_c}\frac{\rho^2-1}{\rho}$. The matched filter gets coherent processing gain, and hence the effective *SNR* for the matched filter is $N_c \theta SNR$, and this helps greatly when N_c is large.

2.1.3 Absolute SNR walls

So what happens on the other side of the wall? An understanding can be obtained by looking at a detector's *Receiver Operating Characteristic* (ROC) curves. Figure 2 plots them with/without (solid/dashed) noise uncertainty. The noise uncertainty leads to the ROC curves shifting away from the $(0,0)$ corner. If the SNR is above the SNR wall (plots on the right in Figure 2), the performance degradation due to noise uncertainty can be compensated for by increasing the number of samples. The sets of test-statistic means under both hypotheses do not overlap and hence ergodic averaging is helpful. However, if the SNR is below the SNR wall, then the sets of test-statistic means under both hypotheses overlap. The ROC curves are worse than those of the random coin-tossing detector! (dotted straight line in Figure 2)

It turns out that this effect is not limited to the radiometer's test statistic. This SNR wall limitation holds generally true for any non-coherent detector.

Theorem 1. *Consider the robust hypothesis testing problem defined in (1) and the above noise uncertainty model with $\rho = 10^{x/10} > 1$. Assume that there is no fading and that the primary signal $X[n]$ is known to satisfy*

- *$X[n]$ are independent of the noise samples $W[n]$.*
- *$X[n]$ takes values from a known bounded set $\mathscr{X}$ (signal constellation)*

Consider a detector that must robustly sense any primary signal satisfying the above properties. In particular, the detector must allow for primary signals that satisfy

1. *$X[n]$ are independent and identically distributed as a random variable X.*
2. *All the odd moments of X are zero, i.e., $\mathbb{E}[X^{2k-1}] = 0$ for all $k = 1,2,\cdots$*

Define $SNR_{peak} = \frac{P \sup_{x\in\mathscr{X}} x^2}{\sigma_n^2}$. Under these assumptions, there exists a threshold SNR_{wall}^* such that robust detection is impossible if $SNR_{peak} \leq SNR_{wall}^*$.

Furthermore, there are easy-to-compute bounds on this SNR wall:

$$\rho - 1 \leq SNR_{wall}^* \leq \frac{\rho^2 - 1}{\rho}. \tag{8}$$

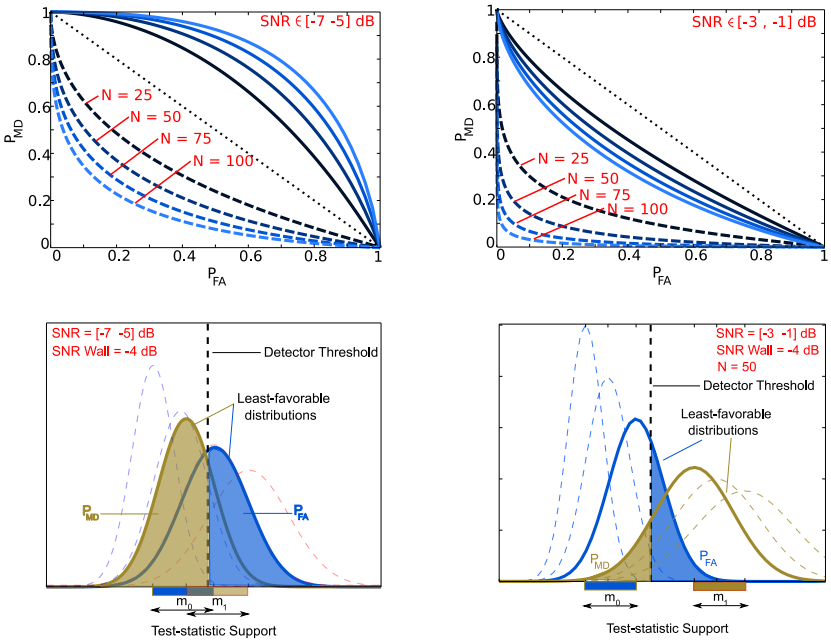

Fig. 2 The solid ROC curves correspond to the case with noise uncertainty and the dotted ROC curves correspond to the case without noise uncertainty. The plots on the right correspond to the case where the operating SNR is above the SNR wall. The plots on the left correspond to the case when the operating SNR is below the SNR wall.

Proof. See proof of Theorem 1 in [3].

Remarks: The theorem says that knowledge of the signal constellation does not significantly improve the robustness of non-coherent detection. In fact, the bounds in (8) tell us that knowledge of the constellation gives atmost a 3 dB improvement in robustness as compared to a radiometer.

2.2 How much sensitivity do we really need?

At this point we have to ask the question – *At what level should we set the detector sensitivity?* We have seen that is easier to detect signals at high SNR (we need fewer samples and can avoid the problem of SNR walls). As we move away from the transmitter, the signal becomes weaker. How far away do we still have to say "don't use this channel?" If we knew that, then the wireless propagation loss model would reveal a baseline requirement for sensitivity.

To answer the question we have to take a spatial perspective and define how much potential coverage loss the primary user must tolerate along with the cognitive radio's maximum allowed power and the cognitive-to-primary propagation

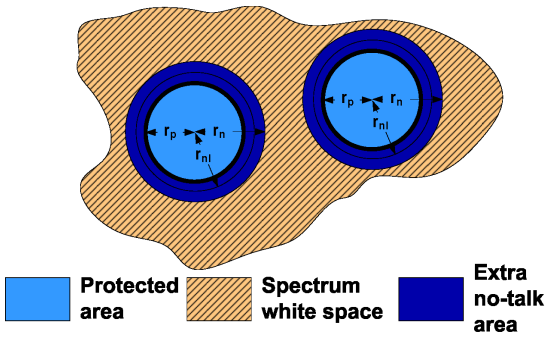

Fig. 3 Spectrum 'white space' from a primary protection perspective. The primary receivers between r_p and r_{nl} are potentially sacrificed. The 'extra no-talk area' is the extra space where there are no protected primary receivers and yet cognitive radios are not permitted to transmit. The problem of spectrum sensing is to correctly answer the question of whether or not the cognitive radio is within the whitespace.

model [1, 7, 8]. Corresponding to the these, there is a no-talk (r_n) radius around the primary transmitter within which the cognitive radio is forbidden to transmit. Figure 3 illustrates this.

The aim of sensing is to determine whether the cognitive radio is inside or outside this no-talk radius. To protect the primary users, it is important to maintain an appropriately low probability of mis-declaring that we are outside whenever we are actually inside[2]. If there were no fading, the required sensitivity would immediately follow from the path-loss model. However, multipath fading and shadowing exist.

We may hope to average over the multipath fading since it changes every coherence time. However, the coherence time is itself uncertain since it depends on physical motion — there is a real possibility of an infinite coherence time since the transmitter and the cognitive radio may both be stationary. This is thus potentially a nonergodic uncertainty, even though it presumably has a probabilistic model. In effect, we must take the worst-case coherence time while calculating the ROC for a detector. Furthermore, behavior of different detectors may be effected differently by the details of the fading distribution.

In Figure 4 we revisit the issue of sample complexity that had previously appeared in Figure 2. To see the role of fading, we suppress the noise uncertainty but incorporate instead the fading distribution at r_n into $\mathscr{H}_1$ (signal present). The threshold of each detector is set so that the P_{FA} is met. Notice that so far, fading has no effect on the "signal absent" hypothesis since there is nothing to fade! The average P_{MD} is calculated as: $\int_{-\infty}^{\infty} P_{MD}(p)\, dF_{r_n}(p)$ where $F_{r_n}(p)$ is the cumulative distribution function (CDF) of the received signal strength at the no-talk radius.

The performance of the radiometer depends on the amount of multipath averaging it can count on (for example Digital TV occupies a band of 6MHz and the coherence bandwidth is significantly smaller ([9]) hence it can count on a frequency diversity gain [10]) (the diversity order specifies the number of taps in the channel

[2] This probability is the equivalent of the missed-detection probability in standard binary hypothesis testing.

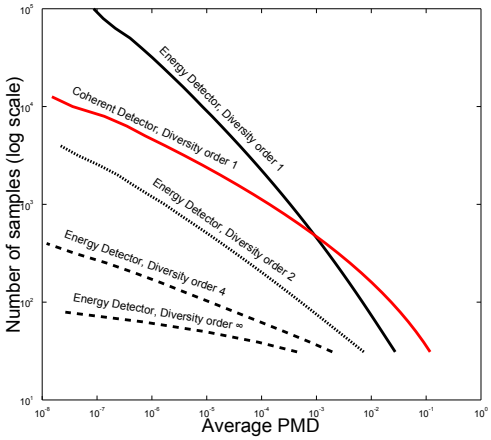

Fig. 4 Number of samples versus the target value of the average P_{MD} while holding $P_{FA} = 0.1$. We assume a DTV transmitter with a 1MW transmit power and a r_n of 156km. The average received power is -90dBm with a noise power of -106 dBm. With a diversity order of just 2 the energy detector performs better than the coherent detector. The model assumes log-normal multipath and no shadowing.

filter). With a diversity order of two[3] or more the radiometer performs better than the coherent detector at all desired P_{MD}. This example illustrates a major point – taking the fading distribution into account is important since it impacts the choice of the detector to be used.

2.3 Defining a spectrum hole in space

P_{MD} averaged over fades better captures safety for the primary and reveals issues that the sensitivity metric alone does not. We now turn our attention to rethinking $\mathscr{H}_0$. Traditionally, this has been viewed as the "signal absent" hypothesis and modeled as receiving noise alone. However, that does not reflect what we actually care about for cognitive radio systems. We only want to verify that the local primary user is absent: it is perfectly fine for there to be some distant primary transmission if we are beyond that tower's no-talk radius.

How we set our detector's threshold impacts how much area we can recover for cognitive radio operation [8]. Figure 5 illustrates the difference between where it is safe to transmit and what space can be recovered by the -114dBm sensing rule for transmitters on DTV channel 39. Notice how significant additional area is lost by setting the sensitivity this low.

How can we see capture this very fundamental tradeoff in our mathematical model? The need for asymmetry becomes clear. The true position of the cognitive radio is uncertain. For $\mathscr{H}_1$, it was natural to take the worst-case position of being just within r_n and then evaluate P_{MD} averaging over fading. That is the most challenging

[3] Since there is also shadowing (common to all taps) in real life, the net diversity order can easily be a fractional value and does not have to be larger than two.

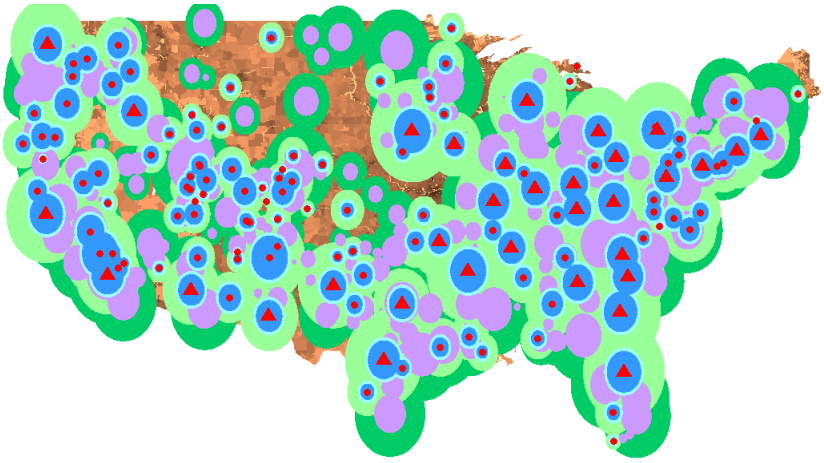

Fig. 5 Location of all transmitters for TV channel 39 and the associated protected (dark blue)and no-talk (light blue) areas. The no-talk area induced by the need to protect adjacent channels is shown in purple. The additional area lost due to very sensitive co-channel and adjacent channel sensing are shown in light and dark green respectively.

point in terms of sensitivity. Suppose we took the same approach to $\mathscr{H}_0$. We would then evaluate P_{FA} immediately outside r_n. After all, if we can recover this location then we can recover all the area greater than r_n. This approach is fatally flawed since the distribution of the signal strength just within r_n and just outside of r_n are essentially the same. No interesting tradeoff is possible. Such a formulation would miss the fundamental fact that we must give up some area immediately outside of r_n to avoid having a cognitive radio use the channel within r_n.

Simply averaging over R (distance from the primary transmitter) also poses a challenge. The interval (r_n, ∞) is infinite and hence there is no uniform distribution over it. This mathematical challenge corresponds to the physical fact that if we take a single primary-user model, the area outside r_n that can be potentially recovered is infinite. With an infinite opportunity, it does not matter how much of it that we give up! We need to come up with probability distribution over r or in other words, we need to weight/discount area outside r_n appropriately. Weighting area by "utility" is a possibility, but as discussed in [4], this would tightly couple the evaluation of sensing with details of the business model and system architecture. It is useful to find an approximate utility function that decouples the evaluation of the sensing approach from all of these other concerns.

Two discounting approaches can be considered:

- We want to use an overtly single-primary user model to approximately capture the reality of having many primary users reusing a particular frequency. As we move away from any specific tower, there is a chance that we may enter the no-talk zone for another primary tower transmitting on the same frequency. As discussed in [4], this can be viewed as a spatial analogy to "drug-dealer's dis-

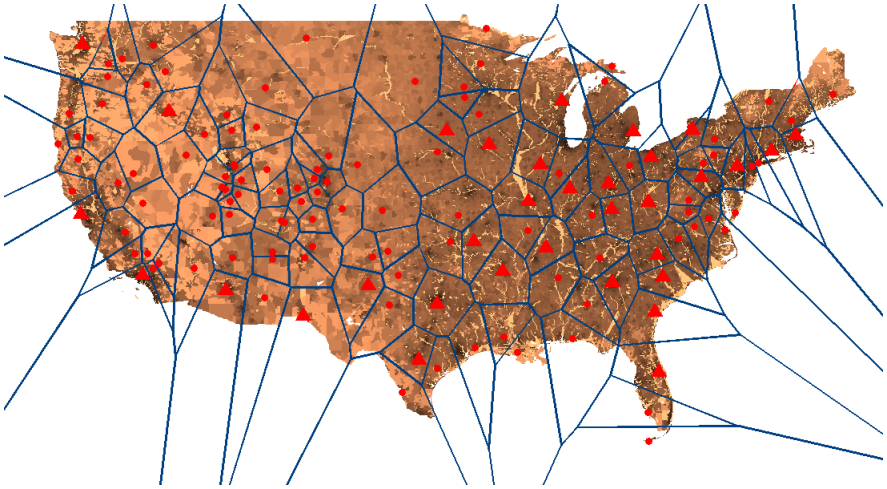

Fig. 6 Voronoi regions of the transmission towers for TV channel 39.

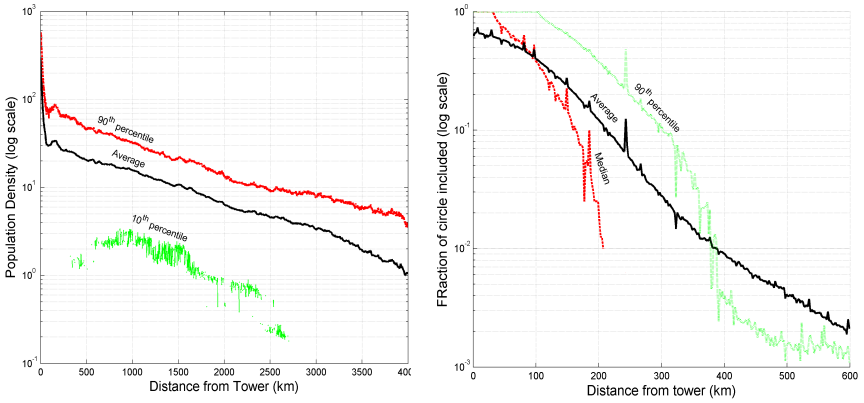

Fig. 7 The figure on the left shows the decay of population density with distance from a Digital TV transmitter. The average, 90^{th} and 10^{th} percentile are plotted. The decay rate is around 8×10^{-4}people/km^2 per km. The figure on the right shows the fraction of a circle (of a given radius) that is included in a tower's Voronoi region. The average, 90^{th} percentile and median are plotted (over towers). The decay rate is around 110×10^{-4} per km.

counting" in which money in the future is worth less than money in the present because it is uncertain whether the drug dealer will survive into the future because of the arrival of the police or a rival gang [11].

Figure 6 shows the Voronoi regions[4] for the transmitters on channel 39. The graph on the right in Figure 7 illustrates the result. The X-axis is the distance from the tower while the Y-axis is the natural logarithm of the percentage of the circle (of that radius) that is included in the transmitter's Voronoi region. Beyond 400km, the mean is dominated by rare large values. The natural logarithm of the included fraction decays linearly with distance *i.e.* the included fraction decays exponentially with distance. The decay rate is roughly 110×10^{-4} per km.

- Presumably, we want cognitive radios to be useable by people. Since TV towers are often located to serve areas of high population density, areas around the no-talk region are more valuable than areas far away. As discussed in [4], this can be viewed as a spatial analogy to "banker's discounting" in which money in the future is worth less in present units.
  Population densities can be modeled as decaying exponentially as one moves away from the central business district [12]. To validate this, we calculated the distribution of the population density at a given distance for all TV towers (taking into account both the high power and low power DTV databases as discussed in [8]). If Q_i^r is the number of people in a radius r around tower i, the average population density at a given distance r is given by:

$$\frac{\sum_{i=1}^{N_T} Q_i^{r+\delta} - Q_i^r}{2\pi r \delta N_T} \tag{9}$$

  where N_T is the total number of transmitters.
  The left graph in Figure 7 shows the average population at a given distance from the tower. The decay rate is roughly 8×10^{-4}people/km^2 per km. We see that this is much smaller than the discounting induced by the frequency reuse.

Now we have a way of discounting the area recovered and presenting the cognitive radio's ability to recover area as a single number.

3 Spatial Metrics

The discussion so far has resulted in a new hypothesis-testing problem. In both of the hypotheses, the received signal $Y[n] = \sqrt{P(R)}HX[n] + W[n]$ but the two hypotheses potentially differ in $P(\cdot)$ (the path-loss and transmit power function), R (the distance from the primary user to the cognitive radio), H (the fading process), and W (the noise process).

Both hypotheses could agree on common models for $P(\cdot)$, H, and W, but there is guaranteed to be a difference in R.

[4] Ideally we would like to construct the received-signal-strength Voronoi region – a received signal strength Voronoi around a transmitter would be all points where the F(50, 50) signal strength from that transmitter is higher than from any other transmitter. Such a Voronoi region is hard to compute.

$$\begin{aligned} \text{Safe to use } \mathscr{H}_0 &: R \sim w(r)r \\ \text{Unsafe } \mathscr{H}_1 &: R \in [0, r_n] \end{aligned} \tag{10}$$

where $w(r)$ must satisfy $\int_{r_n}^{\infty} w(r)\ r\ dr = 1$ and $w(r) = 0$ if $r < r_n$. Following the discussion in the previous section, the numerical results in this paper have been computed using an exponential weighting function, $w(r) = A\exp(-\kappa r)$.

From the above, the asymmetry between the two hypotheses is clear. $\mathscr{H}_0$ is a well-defined probability model and so P_{FA} can be calculated for a detector. Meanwhile, $\mathscr{H}_1$ has R as a non-probabilistic uncertainty and so we would have to require something like $\sup_{r\in[0,r_n]} \mathscr{P}_{R=r}(T(\mathbf{Y}) < \gamma|\mathscr{H}_1) \leq P_{MD}$. The resulting "mixed" ROC curve for a detector reveals the fundamental tradeoffs.

However, such a formulation mixing worst-case and Bayesian uncertainties in different ways across the two hypotheses is novel. Using the traditional names P_{FA} and P_{MD} in this context is likely to lead to confusion. Therefore, in [4] we gave them new and more descriptive names that better reflected their roles.

3.1 Safety: Controlling the Fear of Harmful Interference

The idea behind the safety metric is to measure the worst-case safety that the cognitive radio can assure the primary (the worst case is calculated over the fading distribution negotiated between the cognitive radio and the primary – for example in Figure 4 the primary and cognitive radio agree on a single distribution of fading).

We call the safety metric the *Fear of Harmful Interference*. This is the same as the average P_{MD} in traditional formulations, but takes into account all uncertainties in location and fading.

Definition 1. *Fear of Harmful Interference* (F_{HI}) metric is

$$F_{HI} = \sup_{0\leq r\leq r_n} \sup_{F_r\in\mathbb{F}_r} \mathscr{P}_{F_r}(D = 0|R = r). \tag{11}$$

where $D = 0$ is detector's decision declaring that the cognitive radio is outside the no-talk radius and $\mathbb{F}_r$ is the set of possible distributions for $P(r), H, W$ at a distance of r from the primary transmitter. The outer supremum is needed to issue a uniform guarantee to all protected primary receivers and also reflects the uncertainty in cognitive radio deployments. The inner supremum reflects the non-probabilistic uncertainty in the distributions of the path-loss, fading, noise, or anything else.

3.2 Performance

Next we consider a metric to deal with the cognitive radio's performance — its ability to identify spectrum opportunities. From a traditional perspective, this is ba-

sically a weighted P_{FA}. Every point at a radial distance $r > r_n$ is a spectrum opportunity. For any detection algorithm, there is a probability associated with identifying an opportunity there, called the probability of finding the hole $P_{FH}(r)$:

$$P_{FH}(r) = \mathscr{P}_{F_r}(D = 0|R = r), \quad r > r_n. \tag{12}$$

where F_r represents the propagation, fading, and noise models believed by the cognitive radio designer. A worst-case perspective could also be used here if needed, but it is reasonable to believe that the designer's uncertainty about this could be placed into a Bayesian prior. As mentioned in [4], they have no reason to lie to themselves.

Definition 2. The *Weighted Probability of Area Recovered* (WPAR) metric is

$$WPAR = \int_{r_n}^{\infty} P_{FH}(r)w(r)\ rdr, \tag{13}$$

where $w(r)$ is a weighting function that satisfies $\int_{r_n}^{\infty} w(r)\ r\ dr = 1$.

The name *WPAR* reminds users of the weighting of performance over spatial locations that is fundamental to the cognitive radio context. $1 - WPAR$ is the appropriate analog of the traditional P_{FA} and quantifies the sensing overhead from a spatial[5] perspective [14].

3.3 Single radio performance

Consider a single radio running a radiometer to detect whether the frequency band is used/unused. As the uncertainty in the fading can be non-ergodic, the F_{HI} vs *WPAR* tradeoff for a single-radio detector is interesting even when the number of samples is infinity. In the rest of the paper we assume that the detectors have infinite samples. The test-statistic for a radiometer with infinite samples is $T(\mathbf{Y}) := \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} |Y[n]|^2 = 10^{\frac{P(r)H^2}{10}} + \sigma_w^2$, where $P(r)H^2$ (in dBm) is the average received signal power at distance r. Therefore, the perfect radiometer decides whether the band is used/unused according to the following rule

[5] The designer is free to extend the integral into the time-dimension as well, but that is not the only way to deal with time. Since the $(F_{HI}, WPAR)$ metrics correspond to ROC curves, sample-complexity is still available as a complementary metric. However, even this does not completely capture the relevant design tradeoffs since when time is considered, there is potentially a difference between the startup transient and steady-state performance. The binary hypothesis-testing framework is more startup oriented since it is inherently one-shot. Repeating one-shot hypothesis tests is only one way to operate in the steady state. A more event-oriented perspective is also possible [13]. However, the community's current conceptual understanding there is far more limited.

$$T(\mathbf{Y}) = 10^{\frac{P(r)H^2}{10}} + \sigma_w^2 \underset{D=0}{\overset{D=1}{\gtrless}} \lambda$$

$$\Leftrightarrow P(r)H^2 \underset{D=0}{\overset{D=1}{\gtrless}} 10\log_{10}(\lambda - \sigma_w^2) =: \widetilde{\lambda}. \quad (14)$$

The black curve in Figure 8 shows the F_{HI} vs $WPAR$ tradeoff for a single user running a radiometer. The $WPAR$ performance at low F_{HI} is bad even for the perfect radiometer. This captures the physical intuition that guaranteeing strong protection to the primary user forces the detector to budget for deep fading events. Unlike in traditional communication problems where there is no harm if the fading is not bad, here there is substantial harm since a spectrum opportunity is left unexploited. The reader is encouraged to read [4, 15] to understand the $(F_{HI}, WPAR)$ tradeoffs for detectors limited to a finite number of samples as well as those for uncertain fading models.

3.4 Impact of noise uncertainty: SNR walls in space

In the analysis of the perfect radiometer (see Eqn (14)) we assumed that the noise power σ_w^2 is completely known. We now see the impact of uncertainty in the noise power on the F_{HI} vs $WPAR$ tradeoff for the perfect radiometer.

Theorem 2. *Consider a perfect radiometer, whose test-statistic is defined in (14), where $P(r)H^2$ (in dBm) is the received signal power, $\widetilde{\lambda} := 10\log_{10}(\lambda - \sigma_w^2)$ is the detection threshold, σ_w^2 is the nominal noise power, and $\mathbb{F}_r$ is the set of possible distributions for P at a distance of r from the primary transmitter. Assume,*

- *The received power distribution is completely known ($\mathbb{F}_r$ is a singleton) and is given by $P(r)H^2 \sim \mathcal{N}(\mu(r), \sigma^2)$, where $\mu(\cdot)$ is a known monotonically decreasing function.*
- *The noise power is uncertain, and is known only within a certain range given by $\sigma_w^2 \in [\frac{1}{\rho}\sigma_n^2, \rho\sigma_n^2]$, where σ_n^2 is the nominal noise power, and ρ is a parameter that captures the uncertainty in the noise power.*

Then, there exists an F_{HI} threshold $F_{HI}^t := 1 - \mathcal{Q}\left(\frac{10\log_{10}([1-\frac{1}{\rho}]\sigma_n^2) - \mu(r_n)}{\sigma}\right)$, below which the area recovered is zero, i.e., $WPAR = 0$.

Proof. From the definition, we have

$$
\begin{aligned}
F_{HI} &= \sup_{\sigma_w^2 \in [\frac{1}{\rho}\sigma_n^2, \rho\sigma_n^2]} \mathscr{P}\left(P(r)H^2 < \widetilde{\lambda} | r = r_n\right) \\
&= \sup_{\sigma_w^2 \in [\frac{1}{\rho}\sigma_n^2, \rho\sigma_n^2]} \left[1 - \mathscr{Q}\left(\frac{10\log_{10}(\lambda - \sigma_w^2) - \mu(r_n)}{\sigma}\right)\right] \\
&= \left[1 - \mathscr{Q}\left(\frac{10\log_{10}(\lambda - \frac{1}{\rho}\sigma_n^2) - \mu(r_n)}{\sigma}\right)\right]
\end{aligned}
$$

From the above expression the threshold λ can be computed to be

$$
\lambda = \frac{1}{\rho}\sigma_n^2 + 10^{\left(\frac{\mu(r_n) + \sigma\mathscr{Q}^{-1}(1-F_{HI})}{10}\right)} \tag{15}
$$

Now, the probability of finding a hole is given by

$$
P_{FH}(r) = \mathscr{P}_{\mathbb{F}_r}\left(P(r)H^2 < 10\log_{10}(\lambda - \sigma_n^2)\right) = \mathscr{P}_{\mathbb{F}_r}\left(10^{\frac{P(r)H^2}{10}} < \lambda - \sigma_n^2\right) \tag{16}
$$

Substituting the expression for λ from (15) in (16), we get

$$
P_{FH}(r) = \mathscr{P}_{\mathbb{F}_r}\left(10^{\frac{P(r)H^2}{10}} < 10^{\left(\frac{\mu(r_n) + \sigma\mathscr{Q}^{-1}(1-F_{HI})}{10}\right)} - (\sigma_n^2 - \frac{1}{\rho}\sigma_n^2)\right) \tag{17}
$$

Since $10^{\frac{P(r)H^2}{10}} > 0$, $P_{FH}(r) = 0 \; \forall r \geq r_n$

$$
\begin{aligned}
&\Leftrightarrow 10^{\left(\frac{\mu(r_n) + \sigma\mathscr{Q}^{-1}(1-F_{HI})}{10}\right)} - (\sigma_n^2 - \frac{1}{\rho}\sigma_n^2) \leq 0 \\
&\Leftrightarrow 10^{\left(\frac{\mu(r_n) + \sigma\mathscr{Q}^{-1}(1-F_{HI})}{10}\right)} \leq (\sigma_n^2 - \frac{1}{\rho}\sigma_n^2) \\
&\Leftrightarrow F_{HI} \leq 1 - \mathscr{Q}\left(\frac{10\log_{10}([1-\frac{1}{\rho}]\sigma_n^2) - \mu(r_n)}{\sigma}\right),
\end{aligned} \tag{18}
$$

This implies, $WPAR = 0$, for all $F_{HI} \leq F_{HI}^t$.

Theorem 2 gives an F_{HI} threshold such that a safety guarantee to the primary beyond this threshold will force the cognitive radio to lose all the recoverable area ($WPAR = 0$). In order to guarantee very low F_{HI} for the primary, the threshold must be set such that the primary is protected against extremely deep fading events. In traditional terms, the resulting sensitivity requirement is beyond the SNR wall. Recall that the traditional $\mathscr{H}_0$ corresponds to $R = \infty$ since an infinitely far away primary transmitter might as well not exist at all.

Figure 8 shows the F_{HI} vs $WPAR$ performance for the radiometer and the matched filter under fading and noise uncertainty. The noise uncertainty model used is de-

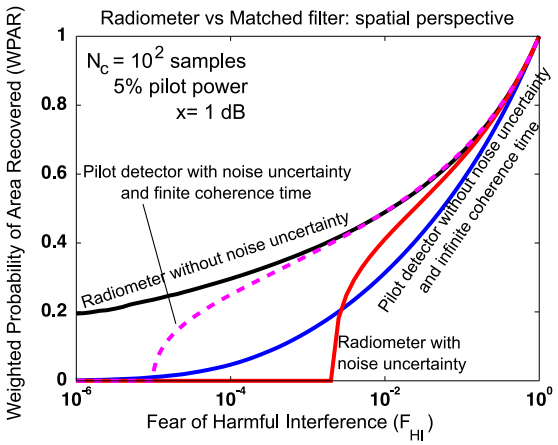

Fig. 8 The impact of noise uncertainty from a spatial perspective is illustrated in this figure. Under noise uncertainty, there is a finite F_{HI} threshold such that if the cognitive radio needs to guarantee protection below this threshold, the area recovered by a radiometer is zero ($WPAR = 0$). The coherent detector (modified matched filter) has a more interesting set of plots discussed in the text.

scribed in Section 2.1.1 and the fading is assumed to be uncertain, and only knowledge of the minimum length of the uncertain channel coherence time, N_c is assumed. The black curve in Figure 8 is the tradeoff for the radiometer with fading uncertainty and no noise uncertainty, whereas the red curve is the tradeoff for the radiometer with both fading and noise uncertainty. Notice how the noise uncertainty introduces a F_{HI} threshold below which the $WPAR$ is zero.

3.5 Dual detection: how to exploit time-diversity

A diversity perspective is interesting to consider. Since the number of samples N is infinite, one could exploit time diversity for multipath if we believed that the actual coherence time is finite $N_c < \infty$. However, for the radiometer all the thresholds must be set based on the primary user's fear of an infinite coherence time. So, the radiometer cannot do anything even if the cognitive radio designer actually believed that the coherence time is finite.

Now consider the matched filter trying to sense for a sinusoidal pilot tone. As the sinusoidal pilot is narrowband the matched filter suffers from the lack of frequency diversity as compared to the radiometer: effectively, fading is more variable. This means that we have to be more conservative, which can cost us in area. However, the matched filter gives us an opportunity to deal with the situation where the coherence-time is uncertain. We can run two parallel matched filters, one assumes that the coherence time is infinity and the other assumes that the coherence time is N_c, with the thresholds set according to their respective assumptions on the coherence time. If either of these detectors declares that the band is used then the decision is also the band is used (this is an "OR" rule among the two sub-detectors). This ensures that the F_{HI} constraint is met irrespective of the actual coherence time.

From a $WPAR$ point of view, the matched filter assuming an infinite coherence time has no SNR wall (as infinite coherent processing kills the uncertainty in the noise), but is susceptible to multipath (no time-diversity to exploit) and shadowing. So, as Figure 8 shows, this detector loses a lot of area. The other matched filter runs using a coherence time of N_c. This enjoys time-diversity that completely wipes out multipath and so has better performance. However, this detector still has an SNR wall due to noise uncertainty. This SNR wall shows up as the $WPAR$ crashing to zero at an appropriately low F_{HI}.

The dual-detector approach leads to two different F_{HI} vs $WPAR$ curves depending on what the mix of underlying coherence times is (stationary devices or moving devices). The good thing about the dual-detector approach is that the F_{HI} is met even when the primary is uncertain, simultaneously guaranteeing the best possible $WPAR$ based on the true channel coherence time. Figure 8 shows this for a very short coherence time $N_c = 100$. For any realistic coherence time, the SNR wall effect would becomes negligible at all but extremely paranoid values for F_{HI}.

So Figure 8 shows an interesting effect. In the case when the actual coherence time is infinite, the radiometer (red curve) has a better $WPAR$ performance than the matched filter for $F_{HI} \geq 2 \cdot 10^{-3}$, even under noise uncertainty! This suggests that diversity is very important, and the lack of it can lead to poor performance. This effect of the radiometer outperforming a matched filter at high F_{HI} is analogous to the time-domain effect of the radiometer sometimes having a better sample complexity than the matched filter (see Figure 1, and Figure 4).

3.6 Cooperation: getting spatial diversity

Several groups have proposed cooperation among cognitive radios as a tool to improve performance. Table 3 in [4] lists the major research themes in the area of cooperative spectrum sensing and representative references. We believe that the most significant gains from cooperation (from the standpoint of recovering spatial holes) are diversity gains. Hence we look at cooperation as a tool to increase WPAR. It allows us to exploit diversity of shadowing.

We assume that a group of M cognitive radios are listening to the primary signal on a given frequency band. For simplicity, we assume that the radios are close enough to each other be essentially the same distance away from the primary transmitter, and yet far enough apart to experience diverse shadowing. Each gets a perfect estimate of the received primary power $P_i(r)$ (in dB) $i = 1, \cdots, M$.

One approach to combine the observations from the cooperating radios is to average the received powers and compare it to a threshold. This is also called the Maximum Likelihood (ML) rule [4], and the test-statistic is $\frac{\sum_{i=1}^{M} P_i}{M}$. Figure 9 shows the performance gains from cooperation. With ten cooperating radios and an $F_{HI} = 5 \cdot 10^{-3}$, the performance of ML sensing rule is within 70% of what would be possible if the radios knew r exactly. Figure 9 also compares the performance of the

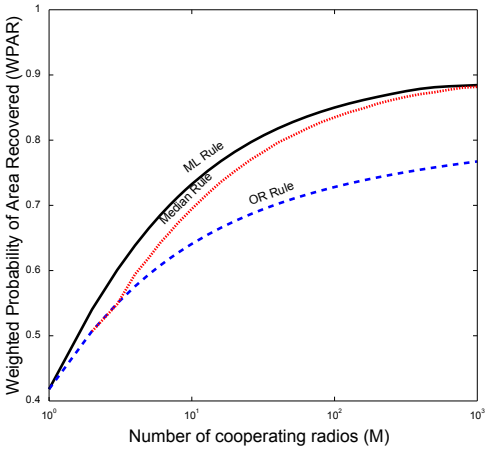

Fig. 9 Performance of infinite-sample cooperation using different fusion rules. The ML Rule performs the best while the OR rule performs the worst. The Median rule (majority-vote) has the best performance among the hard-decision rules.

Maximum Likelihood (ML) rule with the "median" and OR rules. Both of these are k-out-of-M hard-decision combining[6] rules with $k = \lfloor \frac{M}{2} \rfloor$ and $k = 1$ respectively.

From a traditional perspective, it was believed that the OR rule is the optimal rule among the k-out-of-M rules for recovering purely time-domain holes [16]. Interestingly, when the spatial perspective is incorporated, the median (majority-rule) rule performs better than the OR rule. The reasons are explained in [17], but the heart of this effect is the tighter concentration of the median relative to other quantiles whenever the fading distribution is appropriately central[7]. This behavior is different from the traditional purely-time-domain perspective with $\mathscr{H}_0$ being truly signal absent. In such cases, there is nothing to concentrate since the wireless channel has nothing to fade! Instead, the OR rule is preferred from a sample complexity point of view because it permits the detection threshold to be set higher and thereby lower false alarms for the same missed detection [5].

It should be noted that cooperation also suffers from uncertainties — chief among which are unreliable users, shadowing correlation uncertainty and lack of knowledge of the complete fading distribution. The impact of correlation uncertainty and lack of complete fading-distribution knowledge is discussed in [4] while [17] discusses the impact of unreliability coming from improper installation, misconfiguration or outright maliciousness.

[6] In a k-out-of-M rule the fusion center declares the band used is k or more of the radios declare the band used.

[7] The OR rule would benefit from tighter concentration if the fading were uniformly distributed on a bounded interval.

4 Concluding remarks

A careful examination of the problem of spectrum sensing reveals that it takes care to cast it correctly as a binary hypothesis-testing problem. Both of the hypotheses are different than what they are traditionally considered to be, and even the nature of the uncertainty is different between the two. Because of this, it is useful to label the axes of the ROC curve with the new names F_{HI} and $WPAR$. These metrics reveal the fundamental tradeoffs in the problem and illuminate the critical role that diversity plays in improving performance.

References

1. A. Sahai, N. Hoven, and R. Tandra, "Some fundamental limits on cognitive radio," in *Forty-second Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Oct. 2004, pp. 1662–1671.
2. R. Tandra and A. Sahai, "Fundamental limits on detection in low SNR under noise uncertainty," in *International Conference on Wireless Networks, Communications and Mobile Computing*, June 2005, pp. 464–469.
3. ——, "SNR walls for signal detection," *IEEE Journal on Selected Topics in Signal Processing*, vol. 2, pp. 4 – 17, Feb. 2008.
4. R. Tandra, S. M. Mishra, and A. Sahai, "What is a spectrum hole and what does it take to recognize one?" *Proc. IEEE*, vol. 97, pp. 824–848, May 2009.
5. S. M. Mishra, A. Sahai, and R. W. Brodersen, "Cooperative sensing among cognitive radios," in *IEEE International Conference on Communications*, vol. 4, June 2006, pp. 1658–1663.
6. R. Tandra, "Fundamental limits on detection in low SNR," Master's thesis, University of California, Berkeley, 2005.
7. N. Hoven, "On the feasibility of cognitive radio," Master's thesis, University of California, Berkeley, 2005.
8. S. M. Mishra and A. Sahai, "How much white space is there?" Department of Electrical Engineering, University of California, Tech. Rep. UCB/EECS-2009-3, Jan. 2009.
9. G. Hufford, "A characterization of the multipath in the HDTV channel," *IEEE Transactions on Broadcasting*, vol. 38, no. 4, pp. 252–255, Dec 1992.
10. D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, 1st ed. Cambridge, United Kingdom: Cambridge University Press, 2005.
11. J. Hirshleifer and J. G. Riley, *The Analytics of Uncertainty and Information*, ser. Cambridge Surveys of Economic Literature Series, 1992.
12. H. Niedercorn, "A Negative Exponential Model of Urban land use densities and its implications for Metropolitan Development," *Journal of Regional Science*, pp. 317–326, May 1971.
13. A. Parsa, A. A. Gohari, and A. Sahai, "Exploiting interference diversity for event-based spectrum sensing," in *Proc. of the 3rd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks*, Chicago, IL, Oct. 2008.
14. A. Sahai, S. M. Mishra, R. Tandra, and K. Woyach, "Cognitive radios for spectrum sharing," *IEEE Signal Processing Mag.*, vol. 26, no. 1, pp. 140–145, Jan. 2009.
15. R. Tandra, S. M. Mishra, and A. Sahai, "What is a spectrum hole and what does it take to recognize one: extended version," University of California, Berkeley, Berkeley, CA, Tech. Rep., Aug. 2008.
16. J. Ma, G. Y. Li, and B. H. Juang, "Signal processing in cognitive radio," *Proceedings of the IEEE*, vol. 97, no. 5, pp. 805–823, May 2009.
17. S. Mishra and A. Sahai, "Robust cooperation for area recovery," *IEEE Trans. Wireless Commun.*, To be sumbitted.