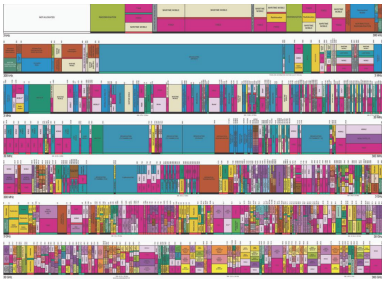
Outline:

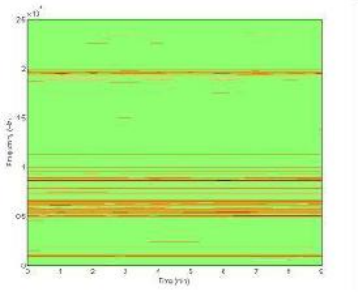
Anant Sahai, Mubaraq Mishra, Rahul Tandra, Kristen Woyach, George Atia, and Venkatesh Saligrama

Prospects and challenges for spectrum sharing by cognitive radios.
EE Seminar at Harvard, Feb 19, 2009.

Slides available: www.eecs.berkeley.edu/~sahai/Presentations/Harvard09.pdf
This Handout: .../~sahai/Presentations/Harvard09.H.pdf
Further discussion and references can be found in the papers.

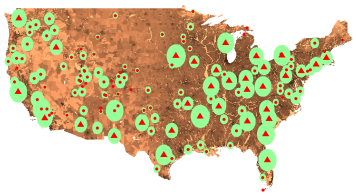
1 How much useable whitespace is out there?

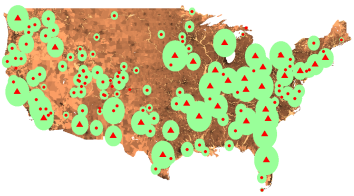




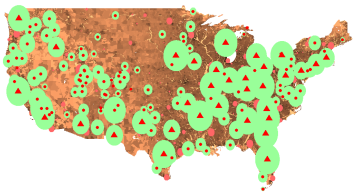
The traditional approach to spectrum is to assign each band of spectrum to a particular use and then to license channels in that band to geographically distributed users. From the FCC's perspective, there is not very much unassigned spectrum. However, many groups have made recent measurements showing that in any given location for most of the time, the spectrum looks very underutilized. **The central problem of the field is to devise a new regulatory environment with far less regulatory overhead.**

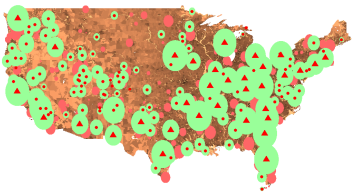
To understand what is happening, we took the USA's 2000 population data per zip-code, the FCC's database of all TV towers, the national topographic database, and used the FCC's wireless propagation models to estimate how much whitespace is there in the TV bands (since the FCC approved cognitive radios in the TV bands on Nov 4th 2008). The figures below all use TV channel 39.





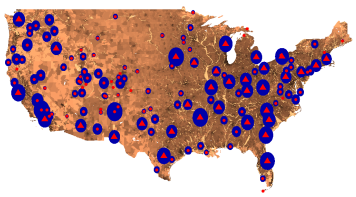
The "pollution" perspective says that locations very close to the television towers are not attractive for cognitive radio operation because of interference from television. The signal decays as we move away and so the amount of area excluded depends on how much the cognitive radio cares about the noise-floor going up. On the left, a 15dB increase in noise is o.k. On the right, only a 5dB rise is tolerated.

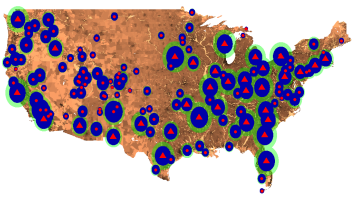




For realistic radios, pollution does not only come from within the band. Strong transmissions from adjacent channels also leak in. On the left, we assume that we can reject adjacent channel interference by 55dB and are willing to tolerate only 5dB of net pollution within our channel. On the right, the adjacent-channel rejection is only 35dB.

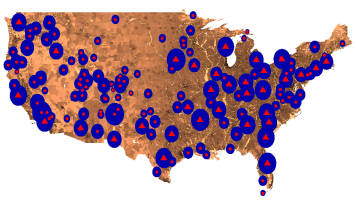
The primary user's perspective is different and demands "protection" from harmful interference.

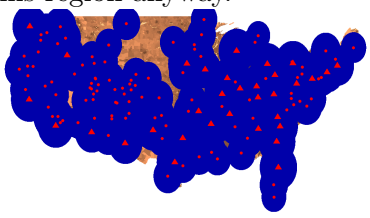




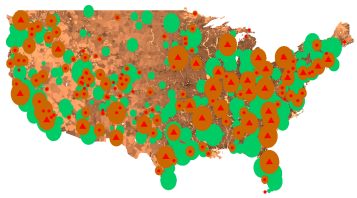
On the left, we see the primary footprint with more than 1dB of nominal fading margin. Even a low-powered secondary should not transmit here because it might be next to a primary receiver. On the right are the exclusion zones for louder 4W secondaries.

The interesting fact is that this protection region is *within* the pollution region for 5dB. So self-interested secondaries would not want to use this region anyway.

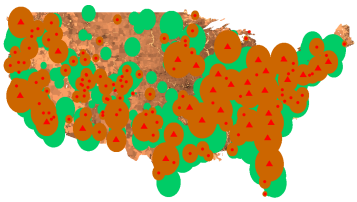




Increasing the permissible secondary power to 20W and then on to 100kW results in dramatically more area being excluded. So there is very little geographical room for loud users like new non-interfering television stations on channel 39.

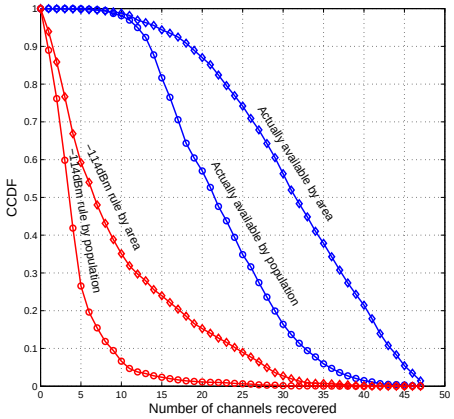


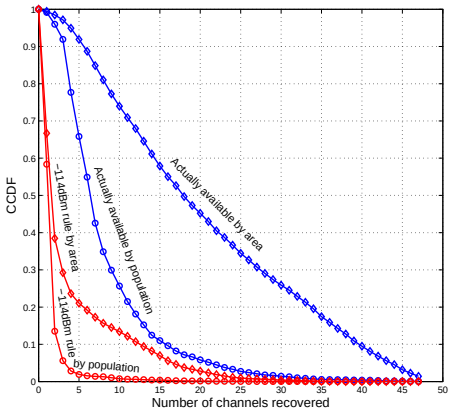
There is also the need to protect TV receivers watching adjacent channels (area excluded shown in green). Even if they can reject 25dB of interference from cognitive radios, uncertainty about the adjacent-channel TV receivers' positions rules out secondary operation within their protected radius itself.



If the received power level is used to decide whether the band is safe to use, then wanting 99% confidence forces us to budget for 1 in a 100 fading events. This eliminates a lote more area for secondary users of channel 39. The FCC requires even more confidence than that for purely sensing-based approaches.

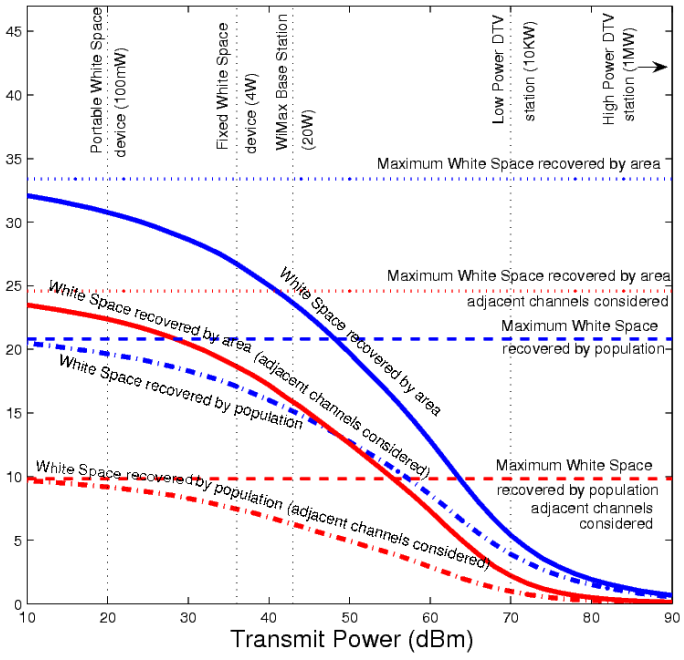
Nationwide, we can see the CDFs of the whitespace channels for a 4W secondary from the protection perspective.





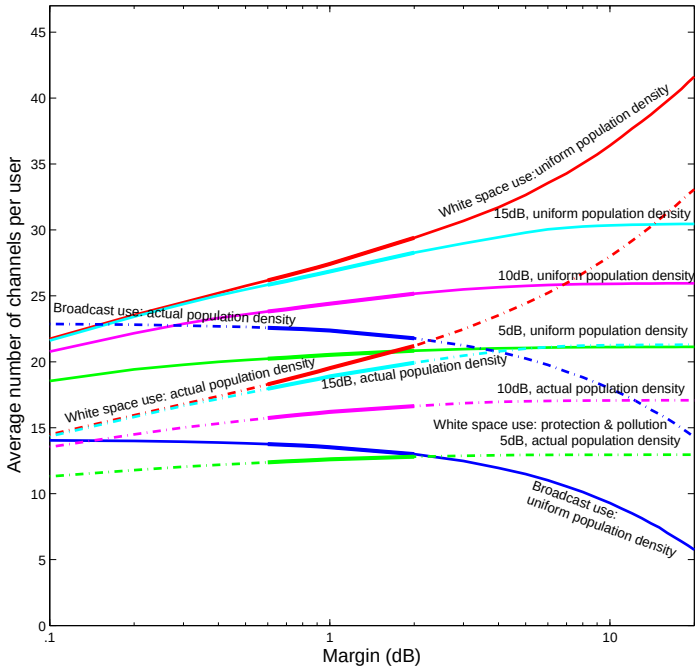
The red curves show what FCC-mandated sensing would be able to recover while the blue show what is available for secondaries who know exactly where they are. The CDF is worse by population as compared to by area. On the left, adjacent channel protection is ignored while it is considered on the right. It makes a big difference.

Detection Scheme/Rule	*By Area*				*By Population*			
	LVHF 2,5,6	HVHF 7-13	LUHF 14-51	HUHF 52-69	LVHF 2,5,6	HVHF 7-13	LUHF 14-51	HUHF 52-69
Pollution (5dB,45dB adj.)	1.6	1.63	**15.6**	15.8	1.62	0.729	**6.63**	14.8
Geolocation	1.52	2.86	**22.3**	16.2	1.69	2.09	**14.8**	15.8
Geolocation with adj.	1.24	1.63	**14.1**	14.6	1.25	0.703	**5.36**	13.1
Sense -114dBm	0.985	0.409	**7.7**	13.8	1.13	0.167	**2.57**	13.6
-114dBm,-110dBm adj.	0.515	0.0635	**2.63**	9.83	0.576	0.008	**0.284**	8.87



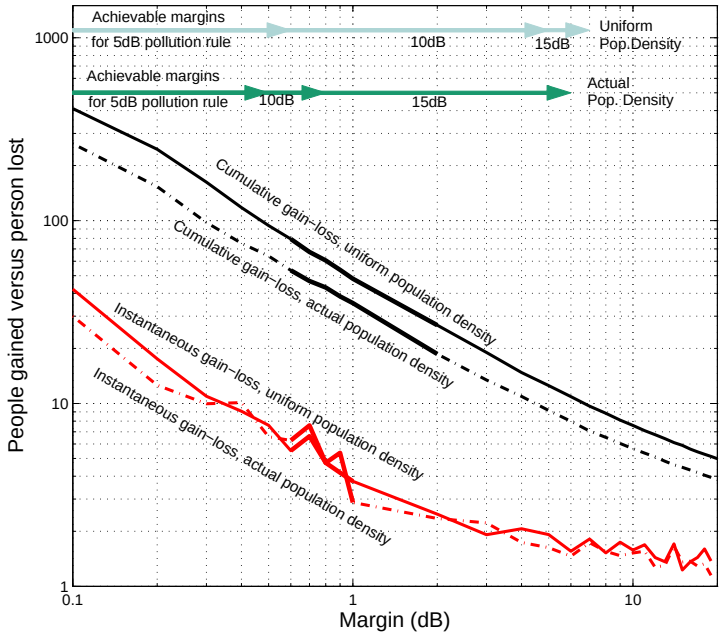
On average, the most attractive whitespace channels in the short-term are the ones in LUHF. The 18 HUHF channels are largely shifting after the DTV transition to public safety and cellular-style uses, but this shows that on average, the accessible TV whitespace is in principle no more than about $\frac{1}{2}$ that of the 700MHz bands of HUHF. Sensing alone is likely to be unable to recover any economically viable amount of spectrum.

The choice of acceptable secondary power can be made by considering how the average amount of whitespace would be reduced. Here the red curves include adjacent channel considerations and so the FCC limit at 4W can be seen as the reasonable value where there are still about seven whitespace channels on average per person.



The core political tradeoff is between potential white-space device users and users of broadcast television. The key parameter is the fading-margin of TV receivers that is potentially eroded. (These curves ignore adjacent-channel effects.) As more of the fading margin is eroded, the TV channels per viewer (blue) drop. Meanwhile, the channels available for for whitespace users increase. Assuming a uniform population density understimates TV channels and overestimates whitespace per person. The differently colored increasing curves show how the pollution-tolerance impacts the number of white-enough channels.

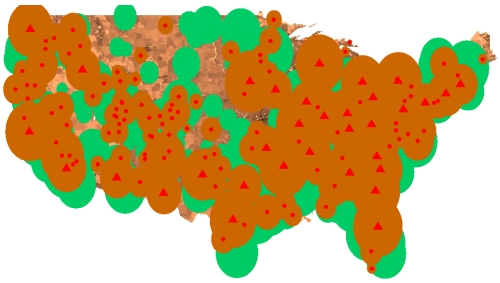
The political tradeoff is made more clear by looking at the ratio of the number of people gaining whitespace channels to the number of people whose broadcast channel is now unreliable.



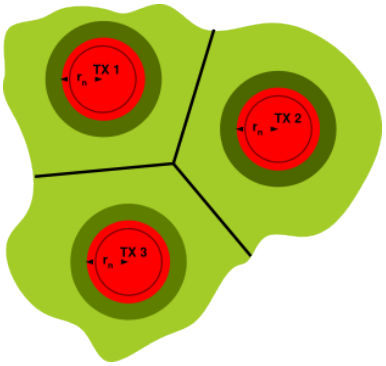
The black curves show the cumulative ratio. The Nov 4th FCC ruling corresponds to a margin around 0.5 to 1dB (depending on which TV band). This corresponds to about 50-20 people gaining access to whitespace channels for every single person who might lose reliable TV reception. If only 15% of people watch over-the-air, we can adjust this ratio to 133-333. But if the market-penetration of whitespace devices is only 5%, then the number drops from 133-333 down to about 7-17.

The red curves show the tradeoff at the margin. For an infinitesimal rise in the margin erosion, how many whitespace users do we gain for every TV user who is now too far eroded? These numbers are around 3-6 white-space users gained for every TV user lost at the margin. With 15% broadcast TV users and 5% whitespace, this would mean a 1-1 tradeoff at the margin.

2 How can we understand sensing?



The current FCC rules for single-user sensing inadvertently eliminate most of the whitespace that is available. So better approaches to sensing are needed. One problem is that traditional metrics for sensing (sensitivity, probability of false alarm, and probability of missed detection) only indirectly reflect the underlying tradeoffs between safety – *protecting primary users* – and performance – *recovering white-space for secondary users*. Simulating the performance of a sensing rule against the full census-data and FCC database does not allow us to develop insight. A simpler perspective is needed.



To capture performance in a single-primary way, we use

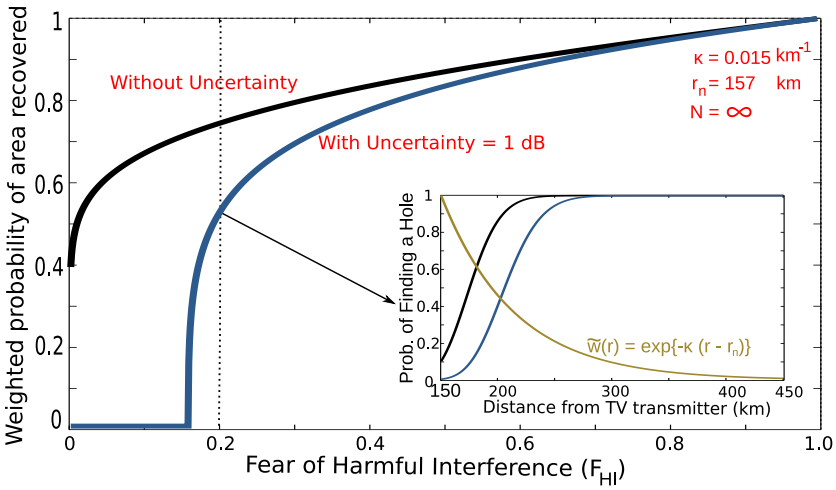
$$WPAR = \int_{r_n}^{\infty} P_{FH}(r)\ w(r)\ rdr$$

which represents the fraction of actual whitespace recovered using a discounting function $w(r)$ to capture both the fact that people tend to be around primary transmitters and that far enough from one primary transmitter, we will get close to another one. The P_{FH} – *probability of finding a hole* – is evaluated for the sensing rule.

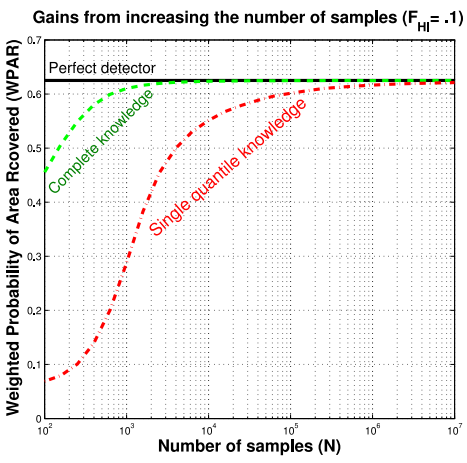
For sensing's safety, the probability of missed detection is the natural candidate metric. However, here an asymmetry is important to capture. The primary user has no reason to completely trust the probability model for the environment or the deployment model proposed by the secondary.

$$F_{HI} = \sup_{0 \le r \le r_n} \sup_{F_r \in \mathbb{F}_r} \mathcal{P}_{F_r}(D = 0 | r_{actual} = r)$$

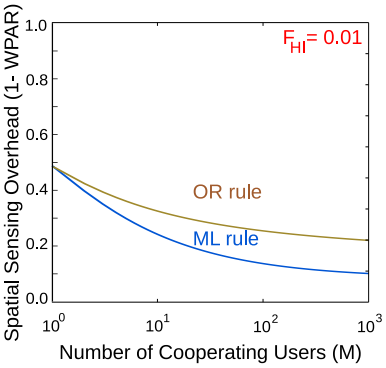
where $\mathbb{F}_r$ is the uncertainty about the environmental probability distribution F_r relevant to sensing algorithm D.



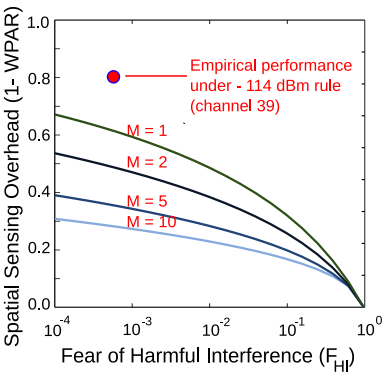
We see how the tradeoff is calculated for sensing in the presence of fading. Even without uncertainty, getting to very low F_{HI} requires a bigger fading margin. This consumes a lot of area. For 1dB of uncertainty in the noise, there is an SNR Wall which is *a fundamental limit beyond which a detector cannot robustly detect a primary transmission without any specified structure.* This means that it is impossible to rule out the presence of the signal for rare-enough fades. So for low enough target F_{HI}, it is incapable of recovering *any* part of the hole.

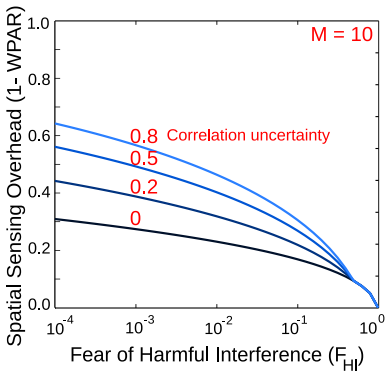


The effect of a finite number of samples is to require more conservatism in setting thresholds. This translates into area loss. If the fading model is uncertain (known only to within a single quantile that we can choose), the number of samples required grows significantly to achieve the same level of safety and performance. This shows that the spatial overhead behaves in a qualitatively similar manner to the time-overhead (traditional sample complexity).



Cooperative sensing can exploit the diversity of fading across different secondary users and thereby effectively reduce the fading margin required. While traditional sensitivity makes no sense, the new metrics are fine. $1 - WPAR$ is *the amount of whitespace area that is not recoverable*. The ML rule to combine sensing information from multiple secondary radios needs to have a reliable full fading model across all users, while the OR rule does not depend on as much. However, even with the ML rule it is hard to drop below 10% spatial overhead without thousands of cooperating users.





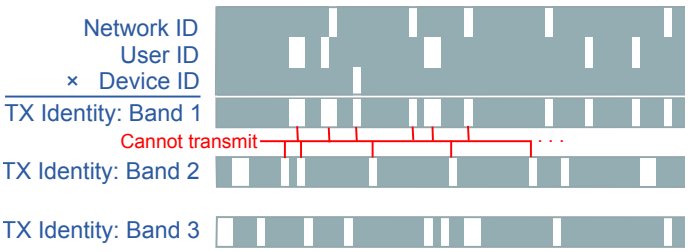
Cooperation is most significant when trying to achieve very low values of F_{HI}. However, it relies upon the independence of fading across spectrum sensors. While reasonable to assume for multipath, it is questionable for shadowing. A feared correlation uncertainty rapidly destroys much of the cooperative gain. This can be overcome with multiband sensing that *exploits the fact that physical shadowing has a very large coherence bandwidth*: being indoors for one band also implies you are indoors for others!

3 Light-handed regulation: identity

Single-user sensing has the advantage of being easily certifiable on a per-device basis (like Part 15 unlicensed devices today). Allowing cooperative sensing would require something different. The currently dominant proposed approach is to aim for provably correct protocols whose implementation in each device is then verified by the regulator. However [Hatfield and Weiser '06] point out that wireless environments are uncertain and so any *a priori* proof of correctness is going to be tied to models that will be hard to validate or overly conservative.

So, some form of *a posteriori* enforcement is going to have to be a part of the mix. Here, [Faulhaber '05] points out that this requires a way to identify who is violating the rules otherwise "hit and run radios" will be entirely uncatcheable. **This is a problem for all approaches to dynamic spectrum use, even those that involve markets.** Prior work on identity has either considered explicit beacon signals or relying on implicit radio features [Hall, Barbeau, Kranakis '03], [Rasmussen and Capkun '07], [Brik, Banerjee, Gruteser, Oh '08].

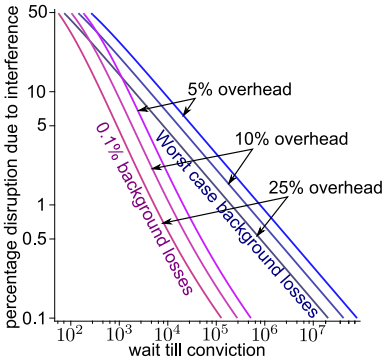
It is unclear if implicit features allow the level of confidence in identification that is needed without assuming very high SNRs, and neither approach can distinguish between truly guilty radios or innocent bystanders. *Our idea is to imbue radios with explicit identity fingerprints in terms of assigning them timeslots during which they are physically prevented from transmitting signals.* This is an easily-verified property of an individual device.

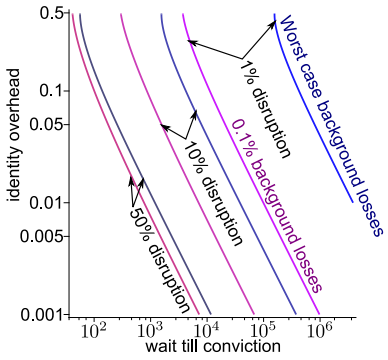


The idea of a taboo-based identity itself is inspired by [Rossmo '95] work on the "geographical profiling" of serial-killers: they have taboo zones close to their homes where they do not tend to kill anyone. The idea of giving radios a "temporal profile" is also connected to the SEEDEX protocol of [Rozovsky and Kumar '01] which was designed to help MAC performance in ad-hoc wireless networks and our own recent work in [Tandra and Sahai '08] on how to make signals that are robustly detectable even at low SNRs.

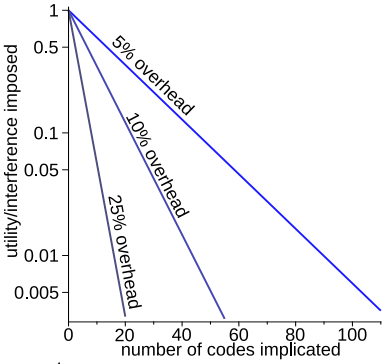
The length of the timeslot should be long enough to contain a useful data packet as well as longer than the delay-spread of any channel between two nodes that might interfere with each other. A few ms seems appropriate. Different layers of identity can be overlaid by allowing each to veto the timeslot. The identity overhead is *the fraction of timeslots that are taboo* since these represent lost opportunities for spectrum use. To support very low-latency links, the identity code should be different in bands that are far apart so that systems can rapidly hop between them to avoid large latency.

The taboo slots act as experimental controls to allow a legitimate user to reliably determine whether its degraded performance is due to natural causes like fading, its own malfunction, or due to interference from an illegitimate user. The other user does not need to be able to demodulate or decode any special signal, it just needs to be able to monitor and record the quality of its own link — no additional burden. *It doesn't matter if the interference is being generated from adjacent channels or co-channel.*

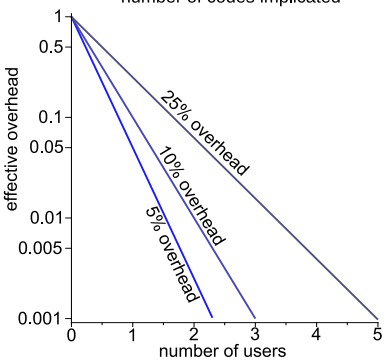




Achieving confidence in an accusation requires a significant wait till conviction: thousands of slots. More overhead helps and it is easier to convict more disruptive radios. A subtle issue is the "princess and the pea" effect: *it is easier to identify radios that interfere with "princess" systems that are used to a very low background level of loss*. Even a small amount of interference is easily caught. A wait of a thousand slots translates into only a few seconds if each slot is a few ms long. This suggests that an identity overhead of around 10% is required.



Cognitive users might want to cooperate with each other at the physical layer and coordinate transmissions. This effectively gives each of the cooperating users' codes a veto over a timeslot. The performance penalty of the identity system then grows rapidly as more codes are implicated in the transmission. This is one reason to allocate a network-level identity. This same effect has a positive implication as well. If a malicious cognitive user wants to "frame" other radios, it will have to voluntarily stay silent during timeslots for which the other radios must be silent. This reduces its own utility as well as the interference it causes to others.



It is also important to prevent the wireless equivalent of looting in that *the presence of one bad or malfunctioning radio should not make other radios feel free to interfere with impunity.* A coalition of a few radios can effectively reduce the number of taboo slots from the perspective of the victim. This reduces the victim's ability to finger the culprit and shows up as a reduction in the effective overhead — leading to a longer wait for conviction. This argues for more identity overhead (around 25%) to make sure that coalitions of at least two bad users are easily caught. On the positive side, anywhere spectrum is scarce and there are multiple systems vying for use, this suggests that the identity system will not be truly imposing a large overhead.

Identifying Culprits	MAC channel
N Distinguishable Secondary Users	N Distinct messages
K Coalition size	K different users
T_c Time-to-identification	Codeword length T_c
γ Taboo-fraction	γ average cost-constraint on codewords
Users may/may-not cheat/interfere	MAC channel model

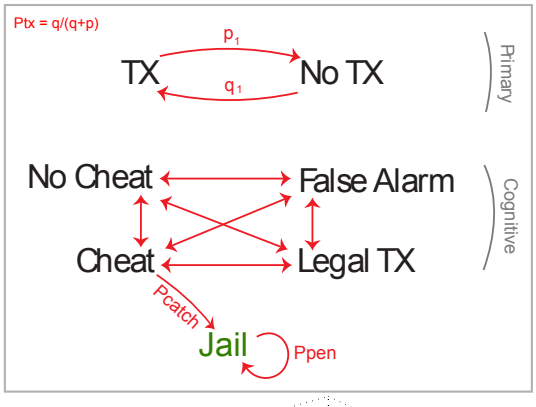
There is an interesting parallel between the information-theoretic multiple-access channel and the problem of identifying culprits. So one can get formal results like $\lim_{T_c\to\infty} \frac{\log N}{T_c} \leq \min_{k\leq K} \frac{I(X_1^k;Y|X_{k+1}^K)}{k}$. These all make sense in the noiseless case where interference is guaranteed to harm the victim. However, there is a subtlety involved. The true incentives for a user are local: it is important to be able to be caught to deter bad behavior and it is important to not be wrongfully convicted too often. Any individual radio system does not care about what happens to others. This makes the true problem in the spirit of [Ahlswede and Dueck, '89] "identification" rather than Shannon-esque communication.

In particular, the idea of individual identity randomization is reasonable from this perspective. The number of identity codes out there only needs to be large enough to support a low enough probability of wrongful conviction based on aliasing.

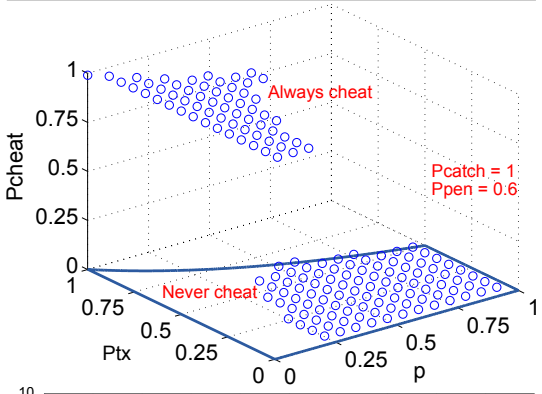
4 Light-handed regulation: deterrence

The key principle of light-handed regulation is to move from mandates to incentives so good behavior is in a user's own self-interest. In wireless, the issue is interference and the idea of iterative waterfilling has been considered in game-theoretic models of multiuser interaction [Rose, Ulukus, Yates '01], [Popescu and Rose '04]. External prices are considered [Huang, Berry, Honig '04], but these are usually considered as signaling devices to help achieve good system behavior among compliant nodes as opposed to a true system of incentives. In the economics/law literature, they are quite serious about using real money in these markets. However, for prices to act as true incentives, there would have to be a completely certified unified system of payments and spectrum access. While seeming light-handed on the surface, the details suggest otherwise.

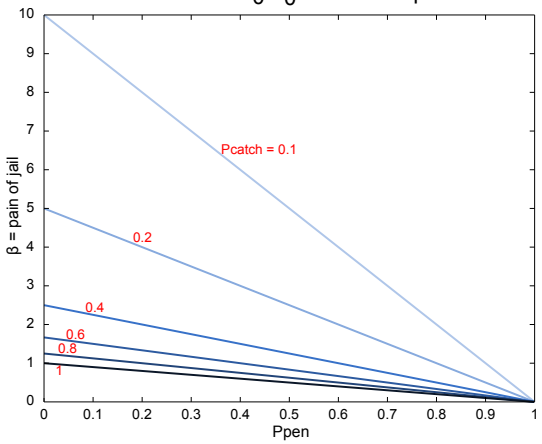
The alternative to monetary prices is to use in-kind incentives. Here [Etkin and Tse '05] reveal that radios can use the threat of punishment through interference to enforce a range of equilibria. However, they point out that this fails when there is fading asymmetry and one system can cause harmful interference to the other but not vice-versa. [Xu, Kamat, Trappe '06] point out that any system of cognitive radios probably needs a remote kill-switch built into it in case something goes wrong. This prompts us to ask, why not take a criminal-law perspective instead of a civil-law perspective? Instead of having capital-punishment as the only form of punishment, why not add a jail? Conceptually, this would eliminate the regulator's fear of asymmetric fading as a barrier to deterrence.



A simple Markov chain is used to model primary and cognitive user behavior, employing a "spectrum jail" system to punish cognitive user misbehavior. The idea is to model the incentives involved in cheating, and to evaluate what an enforcement mechanism has to do in order to deter it. The top chain defines p and q which *characterize primary usage*. Cognitive users move in response to the primary users, choosing P_{cheat}, *the probability of cheating*, to maximize their utility —*the average time spent transmitting*. If in the cheating state, P_{catch} is *the probability of being caught and sent to jail*. Once in jail, P_{pen} *determines how long* ($\frac{1}{1-P_{pen}}$ *on average) the cognitive user must wait before rejoining the game*. The regulator can adjust P_{pen} and P_{catch} to deter cheating.



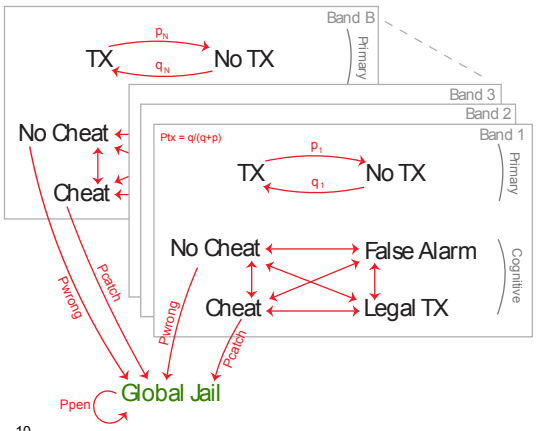
Typical cognitive user cheating behavior, when the primary is allowed to choose only P_{catch} and P_{pen}. The cheating behavior falls into one of two regimes: always cheating or never cheating, depending on the enforcement parameters. Note that if punishment is simply a time-out, there is no way to deter cheating when the primary is nearly always active (high P_{TX}=$q/(q+p)$) because the cognitive user will not be transmitting whether being honest or sitting in jail.



The regulator needs to set the enforcement parameter P_{pen} at certification time, and so it must work for any values of p and q, particularly the worst case when P_{TX} is very high. As noted above, time-outs alone are insufficient to deter cheating in this case. Therefore, we introduce β as *the additional pain of jail* set by the regulator to deter cheating regardless of primary transition characteristic. This figure shows just how high β must be (above the line, cheating is not in the cognitive users best interest)

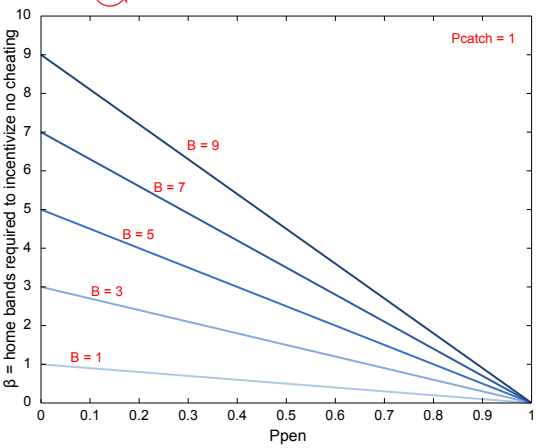
Band:	Home	Expansion Band 1	Exp. Band 2	...	Exp. Band B	Total
Utility:	β	1	1	...	1	β + B

Alternatively, we can think of β as *the utility of a dedicated clean home band* the cognitive user has to stake as a guarantee against cheating.

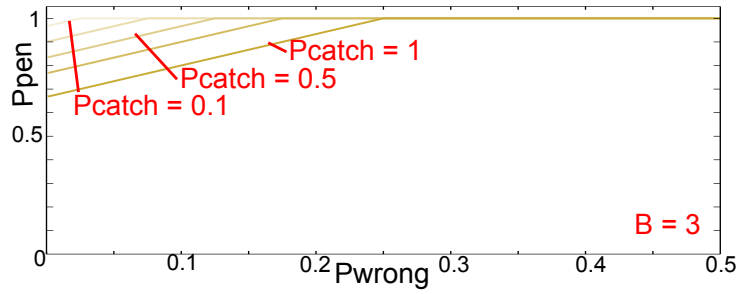


Cognitive use allows potential bandwidth expansion, *the maximum ratio of the number of bands a cognitive user may expand into relative to its own dedicated home band*. If the cognitive user is caught cheating in any band, it goes to a Global Jail where *the cognitive user is barred from using not only all of the expansion bands, but its home band as well*.

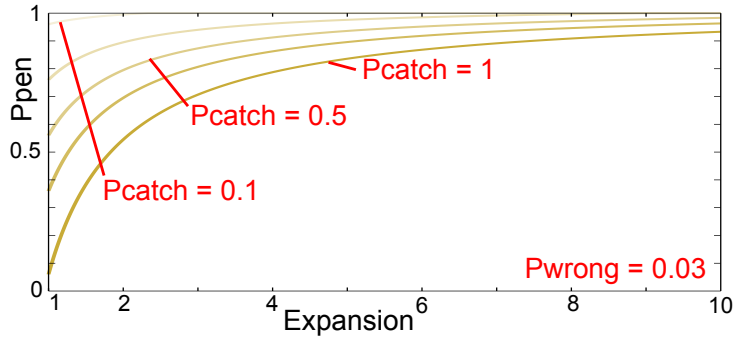
Furthermore, the model is augmented to allow imperfections in the regulator's ability to catch and punish the correct cheater. P_{wrong} models *the probability of being wrongfully punished* (occurs when the wrong cognitive user is accused, the regulator employs collective punishment, or the cognitive user does not detect the primary).



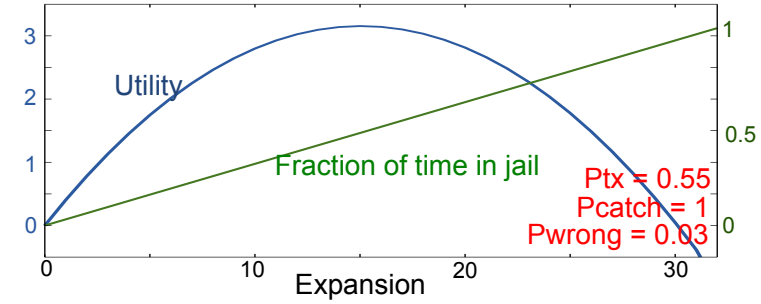
With multiple bands, the new β required to deter cheating even when all of the primaries are transmitting is shown here for the case of no wrongful convictions. Notice that this is simply B times the values before — when there are B bands to potentially gain by cheating, there is B times the temptation. In this idealized model, setting P_{pen} high enough would allow the cognitive user to expand into a great many bands even if β were small.



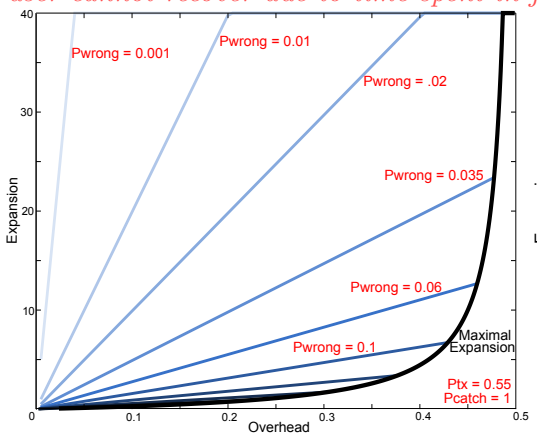
If the regulator can make mistakes, this increases the β required to deter cheating when the primaries are all active all the time. It is easy to see that when P_{wrong} is close to P_{catch}, no amount of extra punishment is sufficient to deter cheating since you will be wrongfully sent to jail at the same rate anyway. But because β is the number of home bands a cognitive user has, it is presumably an unchangeable quantity. Therefore, it is better to think about the minimum P_{pen} for a given β necessary to deter cheating. Notice that as either P_{wrong} or the possible expansion B grows, the necessary P_{pen} approaches 1. It equals 1 (death-sentence) when that expansion cannot be safely allowed.

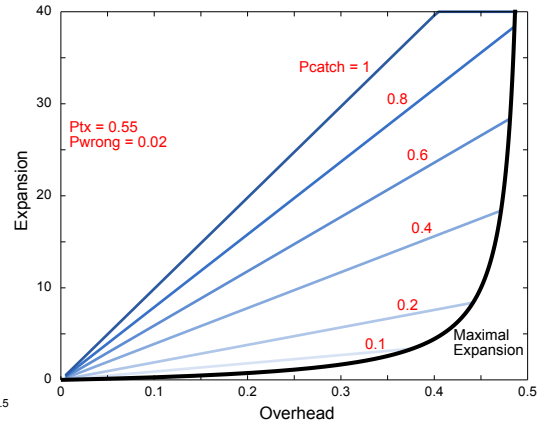


B and the resulting P_{pen} interact for a particular P_{TX} to determine how much utility the cognitive user will get. If choosing whether to expand at all, the cognitive user cares about the zero-crossing of this utility function. If it can choose its expansion, it will try to maximize the utility function shown here. Notice that the maximum utility occurs around the point where the cognitive user is spending half of its time wrongfully in jail.



$P_{wrong} > 0$ dictates that the cognitive user will spend some time in jail despite being honest. We define the overhead as *the average amount of available bandwidth (spectrum the primaries are not currently using) the cognitive user cannot recover due to time spent in jail.*





Notice that the overhead is never larger than 0.5 and for small overheads with large expansions, P_{wrong} must be extremely small. Reducing P_{catch} is not so bad. This makes precise why it is more important that the innocent not be wrongfully convicted than for the guilty to be punished.

Conceptually, there are many different approaches to reforming the currently inefficient approach to spectrum:

	Interference management is primary's responsibility	Interference management not primary's responsibility
Secondary has permission	**Markets**	**Spectrum Monitors**
Secondary must take care	**Denials**	**Opportunistic**

The bottom left corner denials – meaning *a system in which the primary can shoo away secondary users if they encroach upon its spectrum* – is the one that has received the least amount of study in the literature. However, from the technological perspective of light-handed regulation, it seems to be required to support any of the other three quadrants.

5 Summary

This talk intends to convey the following ideas:

- Although fully allocated, there exist many "holes" in spectrum that opportunistic use can help fill. However, the adjacent-channel effects are significant because there is a limit to how well potential victims can filter out strong interference from nearby.
- Single-user sensing to detect whitespaces ends up being conservative to protect against deep asymmetric fades. The right metrics for sensing show the tradeoff between safety for primary users and the ability of cognitive radios to find whitespaces.
- Cooperative approaches for sensing can exploit diversity to get much better tradeoffs, but even so it is hard to safely recover all of the whitespace. About 10% spatial-waste seems inevitable.
- Enforcement with only *a priori* device-level certification seems difficult. Interfering radios need a robust way to be identified and then punished to deter bad behavior.
- Light-handed identity can be given by radios having their transmissions obey a "temporal profile" in which each radio has an individualized sequence of band-specific time-slots that are taboo to it. The fraction of taboo timeslots must be substantial (around 10%) and so this is also an overhead that must be paid.
- Radios can be forced to obey a "go directly to jail command" in which they are blocked from any wireless transmissions for a certain sentence that depends on the current capability of the radio for bandwidth expansion as well as the value of the home band they are able to stake. The fraction of time that innocent radios might spend in jail represents the overhead of this approach.
- **Freedom isn't free.** The different overheads need to be understood and balanced together to get the right approach to wireless spectrum reform.